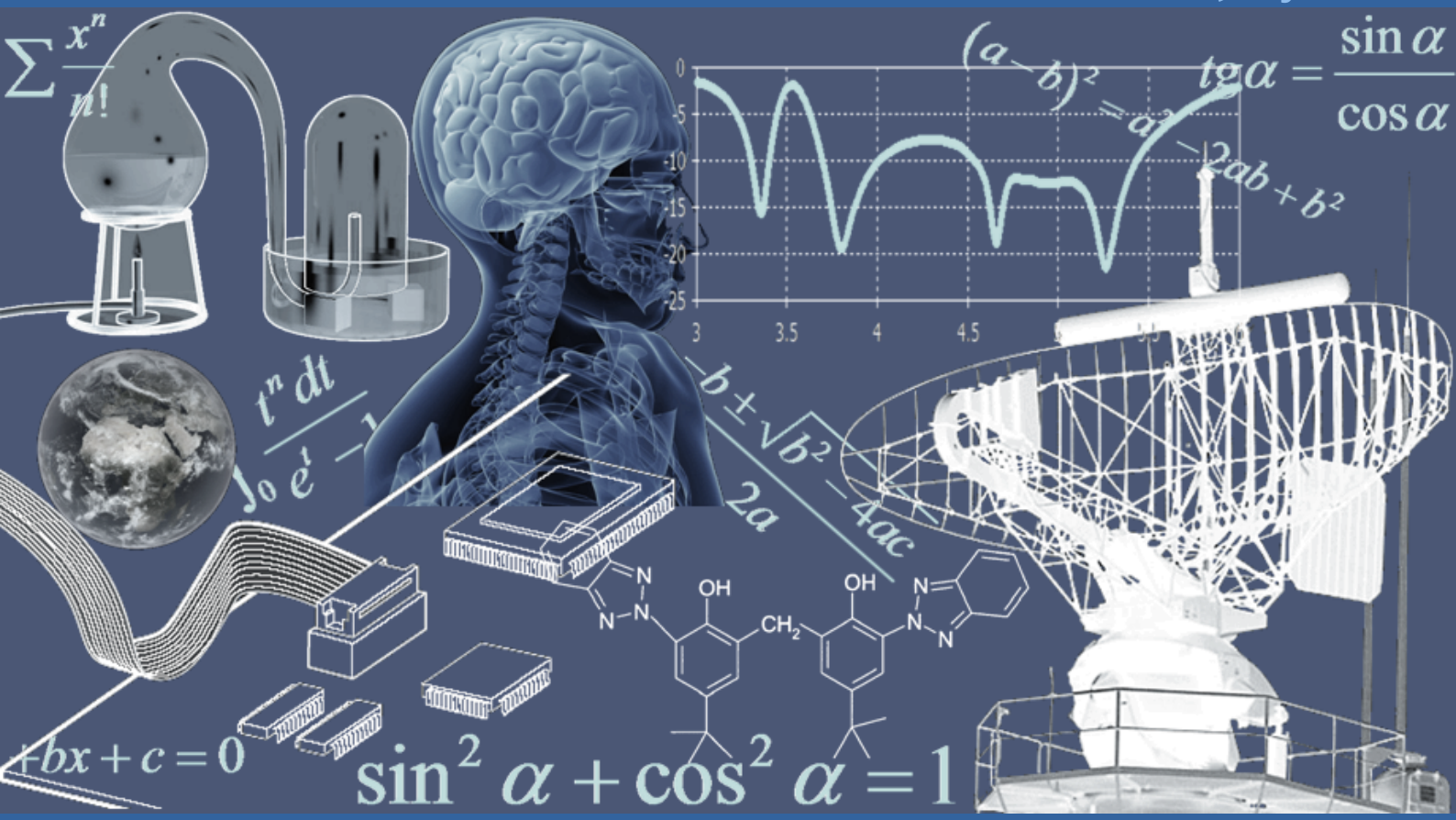
ISSN: 2028-9324
CODEN: IJIABO
Impact Factor: 3.456

INTERNATIONAL JOURNAL OF INNOVATION AND APPLIED STUDIES

Vol. 3 N. 3 July 2013



International Peer Reviewed Monthly Journal

Innovative Space of Scientific Research Journals
http://www.issr-journals.org/

International Journal of Innovation and Applied Studies

International Journal of Innovation and Applied Studies (ISSN: 2028-9324) is a peer reviewed multidisciplinary international journal publishing original and high-quality articles covering a wide range of topics in engineering, science and technology. IJIAS is an open access journal that publishes papers submitted in English, French and Spanish. The journal aims to give its contribution for enhancement of research studies and be a recognized forum attracting authors and audiences from both the academic and industrial communities interested in state-of-the art research activities in innovation and applied science areas, which cover topics including (but not limited to):

Agricultural and Biological Sciences, Arts and Humanities, Biochemistry, Genetics and Molecular Biology, Business, Management and Accounting, Chemical Engineering, Chemistry, Computer Science, Decision Sciences, Dentistry, Earth and Planetary Sciences, Economics, Econometrics and Finance, Energy, Engineering, Environmental Science, Health Professions, Immunology and Microbiology, Materials Science, Mathematics, Medicine, Neuroscience, Nursing, Pharmacology, Toxicology and Pharmaceutics, Physics and Astronomy, Psychology, Social Sciences, Veterinary.

IJIAS hopes that Researchers, Graduate students, Developers, Professionals and others would make use of this journal publication for the development of innovation and scientific research. Contributions should not have been previously published nor be currently under consideration for publication elsewhere. All research articles, review articles, short communications and technical notes are pre-reviewed by the editor, and if appropriate, sent for blind peer review.

Accepted papers are available freely with online full-text content upon receiving the final versions, and will be indexed at major academic databases.

Table of Contents

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 592-607

http://www.issr-journals.org/ijias/

Interrogational Neuroimaging: The Missing Element in Counter-Terrorism

Farhan Hyder Sahito

Institute for Software Technology,
Graz University of Technology,
Graz, Austria

ABSTRACT: Following the September 2001 terrorist attacks in New York, governments have waged a global campaign against terrorists groups in order to ensure national security. A crucial part of this campaign has been intelligence gathering with different methods of interrogation in order to extract allegedly necessary information from suspected terrorists. Similarly, it is not surprising that intelligence personnel have started recognizing that neuroimaging technologies—in particular, functional Magnetic Resonance Imaging (fMRI) addresses this fundamental lack within the realm of scientific scrutiny. The current research introduces a first step towards developing a novel experimental interrogation paradigm that aims to apply a number of reliable and practical applications of fMRI within a rule of law and human rights framework. This prototype is applied in such a way that implications of interrogative methodologies will become a reality for mining of knowledge from potential suspects. The ultimate goal of our innovative methodology is the implementation of fMRI in real life situations that may serve the cause of human rights by providing an innocent person the means to scientifically prove his/her innocence. This truth verification tool has potential to replace torture and aggressive existing interrogation strategies. However, we discuss that there are still human rights and privacy concerns that must be addressed prior to moving this technology to real-world application. Similarly, this paper will recommend best practices and guidelines to address scientific, social, ethical, privacy and general public concerns. The future of law enforcement agencies may very well be under construction with this new line of attack that could revolutionize police work and likely to provide significant benefits to society.

KEYWORDS: fMRI, Interrogation, Counter-terrorism, Human rights, Law enforcement.

1 INTRODUCTION

The fear about investigating suspects has reached new heights and put pressures upon the law enforcement agencies for reliable interrogative methodologies to find out whereabouts of terrorist groups and a prior knowledge of their practices. Towards this end, efforts are invested for centuries by investigators, intelligence officers and psychiatrists to accurately identify perpetrator [1]. Methodologies include torture, polygraph and tools like PSE (Psychological Stress Evaluator), VSA (Voice Stress Analysis), EEG (Electroencephalography) and SVA (Statement Validity Analysis'). Unfortunately, none of these techniques have yielded optimal results and are not entirely suitable for detecting deception and concealed information in the brain of the suspect [2].

According to [3] "Every generation has attempted to develop objective and reproducible methods to discover the truth". Similarly, it is not surprising that security officials have started turning to neuroimaging technologies such as fMRI - a part of a growing trend of technological innovation. It has an emerging potential to revolutionize the investigatory landscape and rapidly coming to the forefront in today's heightened level of security [4] [5] [6] [7]. Proponents of this brain scanning technology argue that arena where fMRI is perhaps most likely to be employed in the future is counterterrorism efforts - from identifying potential terrorists to obtained intelligence information [4] [6] [8]. The brain never lies, and if the information is stored in the individual brain, fMRI can objectively detected this record regardless of the honesty or dishonesty of the subject [10] [11] [12] [13]. This paper is also focusing on the use of fMRI in interrogation room context. It may play a vital role in combating terrorism and assessing the veracity of investigation responses in criminal justice system.

Corresponding Author: Farhan Hyder Sahito (fsahito@ist.tugraz.at)

We argue that this tool has a potential to provide a number of applications ranging from lie detection and to reveal recognition in the brain [14]. This work proposes a methodology in interrogation framework and applies the scientific procedure of fMRI analysis to determine objectively whether or not the person is aware with the information contained in the probes. These stages are: Examination, Interviewing and Scanning that will be applied on the dynamics of the pre-Incident and post incident phase of terrorist related activity. The ultimate goal of our innovative methodology is the implementation of fMRI in real life situations. That will enable intelligence operatives to detect suspicious behavior indicators to provide real-time decision support.

This paper recommends that our methodology will emerge as more promising practice as various researches have reported reliability and reasonably high accuracy rates for fMRI studies. However, critics of brain scanning have leveled several different concerns regarding possibility that counterterrorism agencies may deploy this technology prematurely [15] [16. This research will counter these claims by arguing that because of the reliability and novelty of the physiological parameters being measured; it would be a human rights violation to deny access to fMRI scanning. This technology has potential to replace torture and aggressive existing interrogation strategies. However, we discuss that there are still human rights and privacy concerns that must be addressed.

The prevention of inappropriate application of this technology in counterterrorism context is as important as highlighting its potential usefulness. Similarly, this paper will recommend best practices and guidelines to address scientific, social, ethical, privacy and general public concerns. This research will not discuss the legal issues; instead the focus will be on addressing the research challenges and related issues to elucidate our methodology. We argue that law enforcement agencies and neuroscientists who are engaged in interrogation process must be held to account for their actions. Specially, if the public trust in the integrity of these professions is being seriously compromised or violated international human rights laws. Our efficient internal measures will also propose an elite interrogation unit and a call for greater partnership between major stake holders to engage the general public. It is a major step forward in the action to our defense needs and essential to the national security. The objective of this research is to break old orthodoxies that have confined our mission on war on terrorism.

2 THE ART OF INTERROGATION IN THE WAR ON TERROR: GETTING INSIDE THE TERRORIST MIND

Interrogation is the interview of a suspect to uncover tiny bits of the truth to collect evidence that can be used as evidence at the trial [17]. However, the objective of interrogation in counterterrorism context is to reveal information concerning threats to national security [18]. The benefits of interrogation can be enormous such as preserve important institutions, protect our own population, prevention of a nuclear explosion, maintain civic order and stability etc. [19]. Poor performance in the interrogation function results in the loss of these benefits and can also impose other costs [19]. This process is different from a conversation and the suspect being interrogated is often not willing to divulge information or comfortable with this practice [20]. This procedure can take different forms, but all methods have a similar aim to control the individual in such a way that provides the information being asked for [21]. Interrogations officers are trained to extract this information with different methods. However, it is reasonable to assume that terrorists (guilty suspects) enter examination room with different counter-interrogation strategies [22].

The honesty of reports offered by suspects concerning their doubtful activities and relations with the specific organizations must all be assessed to inform and guide decision making process. The aim is to also establish accurate connections between characteristics of the terrorism related activity and features of the terrorists (e.g. bomb making knowledge, terrorist training). For instance, a suspect may leave traces of permanent feature of criminal on the crime scene, such as fingerprints and DNA [23]. Similarly, members of a terrorist cell share a particular body of information and the revealing of such knowledge could help in the identification of a suspect. The other fundamental task is to keep the number of both false-negative and false-positive errors as low as possible to conclude the decision [24]. This conclusion may hold important consequences for the subject (e.g., continued custody versus release). Law enforcement agencies have attempted for centuries to determine the accuracy of statements and truth with different interrogative methodologies [25]. Torture is one of the traditional investigating methods used by police officers to orchestrate another possible future attacks [26].

3 TORTURE AND ETHICS IN THE WAR ON TERROR: NO PAIN, NO GAIN?

Security agencies sometimes characterize torture as an indispensable interrogation technique for gathering strategic intelligence [26]. However, torture is obsolete, or at least obsolescent. Over the past few decades the tools of psychological and physical torture have been also refined [27]. Furthermore, the consequences of coercion techniques are numerous, not just for individual victim or suspect but also for the world as whole [26]. Several international human rights bodies have also

highlighted the risk of discrimination in this methods presented by the investigators [28]. This process undermines fair procedures that would otherwise safeguard against miscarriages of justice [29]. Torture also offends ethical concerns and skirts the rule of law that shocks the conscience and violates international and domestic laws [30]. These practices may weaken state's long struggle against terrorism, undermine the legitimacy of its action and eventually limit the government's ability to act [30].

Adapting a quotation from the classic book Front-Line Intelligence the interrogators' purpose is to "facilitate the accomplishment of the mission, and to save lives. When they fail, all the wrong people are hurt1." It is necessary that there should be a standard that keeps interrogators' away from gray areas that might be considered as torture [31]. Police officers are now trying to obtained information without leaving a physical trace of the trauma of torture [32]. The goal is to determine the innovative technologies and reliable methods to shape future interrogation policies [33].

4 PROVIDING NEW TOOLS TO FIGHT TERRORISTS: IS THIS A COMPREHENSIVE RESPONSE TO TERRORISM?

Novel interrogation technologies often bring with them new opportunities to influence criminal justice system. Few technologies, however, have captured the imagination of law enforcement agencies, policy makers, and the general public, like Polygraph. However, this technique has failed. Polygraph relies on physiological manifestations of anxiety and measures changes in skin conductance, respiration and heart rate to detect deception and makes it intrinsically susceptible to producing erroneous results [34]. This technology is not effective when the subject has learned to suppress these manifestations or when suspect is a sociopath [35]. In addition, serious doubts can be raised when questions and answers are translated to and from the suspect's native language [36]. Other technique such as PSE was invented for the detection of emotional stress in the voice to identify deception with infrasonic frequency modulation that varies between 5 Hz and 20 Hz [37]. It is controlled by the central nervous system and disappears during emotional stress of the suspect. However, the popularity of the PSE was declined as this method is criticized as invalidate. Experimental research also failed to found its validity in various studies and today it is rarely used [37].

SVA is another tool based on the analysis of statements made by suspects to specify what he/she did or what he/she saw is investigated [38]. This technique assumes that false statements differ from true ones in both quality and content, and considered to be a highly effective as a police interrogation technique. However, critics argue that this method is theoretically vague with little or no empirical evidence in its favor. It encourages interrogators to presume a subject as deceptive and affirm a presumption of guilt before the interrogation process even starts [38]. The SCAN (Scientific Content Analysis) is also a technique used to recognize deception assumes that suspects will use more deviations in pronoun usage [39]. For instance, replacing I with you and use many unnecessary connectors, such as 'after I left', and 'and then' [33]. Unfortunately, there is no scientific evidence on the validity of the SCAN and it is not especially effective beyond its ability to generate admissions in interrogations [33] [39].

Thus, scientifically sound studies have concluded that above technologies and methods are too inaccurate to be used in practical settings and alternative methods ought to be investigated [35] [37] [38]. These strategies are geared toward getting subjects to comply and talk. It may lead innocent subject to make incriminatory testimony against his own self-interest and possibly resulting in conviction and incarceration [33] [39] [40]. By and large, such tactics have been shown to be unreliable and counterproductive.

5 WHAT IS NEEDED TO START TO WIN THE WAR ON TERROR: DEFINING THE PROBLEM?

The treatment of prisoners at Abu Ghraib and at Guantanamo Bay have shocked the public and provoked the collective conscience of the society [41] [42]. Civil society, theologians and scholars have condemned the harsh and traditional methods used by the Intelligence operatives that include both physical and psychological torture [42] [43]. The law enforcement agencies need to know whether any pragmatic technique of obtaining information actually works to avoid the creation of more enemies and maintaining the integrity of the state [41]. However, there has been no objective interrogation method and scientific way to discriminate between truthful and deceptive statements [44]. It is also very difficult to uncover concealed information in terrorist's mind with above tools. The objective of the criminal justice system is implement

[1] *Front-line intelligence / [by Stedman Chandler and Robert W. Robb]. (Washington, D.C. : U.S. Marine Corps, 1986)*

technologies that can identify the line of attack as we can better prepare ourselves to extend our defenses and protect the nation [45]. However, this article supports the possible uses of fMRI, based on their current validity and the availability of peer-reviewed public data. This tool may prove to be an effective deterrent for extremists and avoiding the prosecution of innocent subjects, thus, freeing up governmental resources [46]

6 FUNCTIONAL MRI AS A COUNTERINSURGENCY STRATEGY: TOWARDS A NEW MANIFESTO IN FIGHT AGAINST TERROR

The functional MRI is widely known and accepted in the scientific community as it does have a significant amount of scientific research behind its claims and validity. There have been hundreds of articles published and studies conducted on it over the past two decades. An fMRI is an increasingly popular psychiatric neuro-imaging technique that was developed in the 1990s and has since become the preferred method for studying the functional anatomy of the human brain [34]. This is a hybrid technique that provides real time and ultra-high resolution, computer-generated models of brain activity [47]. It produces a series of ultra-high resolution structural scans of the brain, which expose brain activity [27]. This technique relies on the fact that cerebral blood flow and neuronal activation are coupled. It involves placing the subject in a donut-shaped magnetic technology, which can identify subtle changes in electromagnetic fields [36]. The subject must remain still during the scanning process. During scanning, when an area of the brain is in use, blood flow to that region also increases [36]. Thus, its responses associated with neuronal activation correlated with cognitive tasks and various behavioral functions [48]. This is how this machine detects this physiological change due to the blood oxygen-level-dependent, or BOLD, effect. The changes are represented onto a three-dimensional, computer-generated image of the person's brain [36].

The introduction of fMRI is considered to be technologically superior to any another comparable imaging method such as PET (Positron Emission Tomography) [5]. In contrast with PET, functional Magnetic Resonance Imaging does not need the injection of radioactive labels into the subject. Secondly, fMRI has better temporal resolution (down to 2 to 3 seconds with rapid event-related) and superior spatial resolution (down to 1 mm3) than PET [49]. According to the director of the fMRI Research Center at Columbia University, this novel technology really opening the black box as it signifies a "quantum leap" over any previous technology for imaging the brain [49]. An fMRI has already had a major impact on neuroscience and in clinical settings. It has been applied ranging from language comprehension to treatment of neurological impairment disease, psychiatric illness, aesthetic judgment and justification of cognitive enhancing drugs in educational settings [14]. With these rapid developments many researchers claimed this technology to be useful outside the laboratory settings. For instance, economics contexts, investing personality traits, mental illness, religious extremism, racial prejudice, suicidal thoughts aggressive or violent tendencies and lie detection [14].

Proponents of this neuro-imaging technology hailed this machine as a next truth meter [6]. They conclude that because of the novelty of the physiological parameters being measured, this technology may be more accurate than other traditional methods [34]. According to Spence, this ground breaking research proves that fMRI has the potential to reduce the number of miscarriages of justice and capacity to address the question of guilt versus innocence. Since the first publication by [50] on deception detection by fMRI, various papers and studies [8] [12] [51] [52] have reported different experiments in which subjects were asked to respond deceptively in some blocks and truthfully in others. In these two studies, subjects were instructed to say yes when the truth is no and vice versa [50] [53]. In another study, the task paradigm included spontaneous lies [54], for instance, the subject was instructed to say Chicago when the truthful answer is Seattle. Similarly [51] [52] studies were included feigning memory impairment tasks. In addition, lying about having a play card [4] [10] [11] [55] and lying about having fired a gun [12] revealed that particular spots in the brain's prefrontal cortex become more active when a subject is suppressing the truth or lying. In some of the other experimental tasks, subjects were motivated by monetary incentives as they were told that they would double their reward money if they were able to deceive the fMRI machine. For example, lying about having taken a ring or a watch [56] and lying about the place of hidden money [57] [58].

In above studies, subjects were asked to conceal their information by lying and press buttons to respond 'no' or 'yes' to specific questions. Though the answers varied from trial to trial but it was possible to determine brain activity in response to specific pieces of information. In spite of different paradigms employed in the laboratory settings and the content of the questions, brain activation was compared in response to deceptive answers to truthful ones. It proved that lying involve more efforts than truth and expose that specific brain areas respond strongly in generating deceptive responses. As with lying, several brain regions show significant increases and light up on during scanning when a person sees a familiar object or image or during deception compared to truth telling [59]. For instance, dorsolateral prefrontal cortex (DLPFC), anterior cingulate cortex (ACC), ventroletral (VLPFC) and left and right cerebral hemispheres increases activity when people tell lies [16] [55]. Similarly, during the interrogation phase, if a suspect is asked a question, the information to which is unknown then the specific regions of the brain is unusually active and it is presumed that suspect is lying; if, however, the same areas are no

more active it may presumed that subject is telling the truth [59]. Thus, this technology has potential to reveal recognition regardless of whether the suspect speaks or attempts to conceal the recognition.

In above experiments, this technology has been claimed to be 80 - 90% accurate by the researchers. Apart from above laboratory experiments, Sean Spence, who has pioneered the use of this groundbreaking technology, carried out a real-life experiment in 2008 [6]. He investigated the potential innocence of a woman who had been convicted of intentional inducing illness in a child (and later was sentenced to four years in prison, see [6]). Brain imaging technologies has also captured the attention of legal system to influence criminal justice system. For instance, in September 2008, a court in India allowed to use brain scan images in a criminal case. Aditi Sharma was convicted by a court for the murder of her former fiancé, Udit Bharati [60]. However, for the first time, a brain scan was used as evidence of a criminal defendant's guilt. This case marked the dawn of a new era for the use of brain scan technology in criminal prosecution. The court found that the brain scan proved that Aditi Sharma had experimental knowledge of having murdered Udit Bharati herself [60].

A variety of recent advances in neurological research and the development of this new technology claims to be a more accurately deception revealing tool for screening. It can be effective in distinguishing truth tellers from liars and to determine hidden conscious states of an individual, with accuracy greater than chance [7]. Thus, unlike polygraph—which detects a person's emotional response to deception—fMRI measures person's decision to lie, as subjects cannot control their cerebral activity to avoid detection [27]. Thus fMRI can be used as a tool warranted in interrogation techniques in this era of terrorism that is creating an all-pervasive fear. This technology can be considered as a magic bullet in the war on terror [6] [7]. Not only has this neuro-imaging technology taken the attention of scientific communities and law enforcement agencies but it has also attracted interest of corporate world [61]. Two private firms: No Lie MRI and Cephos Corp trying to make the dream of perfect truth verification into a reality and have begun marketing since 2006. They offer high-tech lie detection services based on research comparing neuronal activation patterns [36].

7 FUNCTIONAL MRI: A NEW WAY FORWARD IN THE CRIMINAL JUSTICE SYSTEM

This technique has emerged as more promising technology that aims to directly reveal if a suspect's brain displays particular responses: Specially, when it is deal with specific information that could only be known to the criminal or terrorist. This tool has potential to directly reveal deception and read out the contents of suspects' mind, including their intentions and memories to reveal recognition. Ruben Gur, a neuropsychologist at the University of Pennsylvania, states that fMRI scans can reveal cognitive tasks when a subject recognizes a familiar picture, face or place, no matter how hard he or she tries to conceal it [62]. This cognitive analysis technology could function as a hyper-accurate lie detector that is nearly impossible to deceive [8] [56]. For instance, an interrogator could present a suspect with images of terrorist leaders, potential targets, or specific information that could generate neural responses if the subject were known with that pictured information [56]. This scientific technique provides intelligence operatives to focus their investigations on the suspects who actually commit terrorism and to determine if he or she has been to any specific place before. If a person was in any terrorist training camp, you can actually determine that [12] [46].

On the other side, an information absent will provide support for the claims of innocence that individual is not guilty of committing any crime and has no knowledge specific to any particular group [6]. The imaging results can be used against the suspect at trial and prevent future tragedies. Thus, this machine is capable of witnessing the brain in action by tracing the way blood flowed and takes pictures that highlight specific areas of the brain activated during certain tasks. Similarly, the primary goal of the current research is to develop a novel experimental paradigm with fMRI based interrogation techniques. The purpose is to maximize the likelihood of a true confession of a terrorist activity.

8 OUR METHODOLOGY: APPLYING THE CONCEPT TO LAW ENFORCEMENT

One of the most important aspects of security agencies is the prevention of terrorist attacks with a prior knowledge of terrorist practices and mindsets regarding preparation and implementation of attacks – so called Pre-Incident Phase. Secondly, to establish accurate and reliable connections between features of the terrorist attacks the one hand and features of the perpetrator or witnesses related to the terrorist activity on the other - Post-Incident Phase. Our research is also focusing on the dynamics of the pre-Incident and post incident phase of terrorist attacks. Examples of the knowledge of these phases include weapons details, information regarding specific locations, time, key personnel, source of funding, recognition of false identities for group members, acquisition of supplies, the deployment of assets and other related information. The future of law enforcement may very well be under construction with this new approach that is becoming a reality for mining of knowledge from potential terrorists to assess potential threats rapidly. This methodology proposes three paradigm of using fMRI in interrogation process. These phases are: Examination, Interviewing and Scanning.

8.1 EXAMINATION

This process is research-intensive as it consists of the designee that will determine the significant features of terrorist activity. The interrogator must be careful to select stimuli in such a manner that a subject who is innocent would find them as equally plausible as the irrelevant chosen. Thus, no physiological response is expected on fMRI. However, this information must be present in the brain of suspected terrorist. The probes selected in terrorist related activities must be included the landscape that the terrorist ran through while planning or committing the act. Interrogator must formulate the actual event of terrorist related activity with the features and background information in two ways. Firstly, about known terrorists whose suspicious activities or relation with terrorist organization is reasonably certain because of the evidence available. Secondly, the suspects whose guilt is doubtful or uncertain because of lack of essential evidences or because of weaknesses in the available facts. It is important to note that some suspects cannot be placed precisely in either of these two groups.

The accuracy of interrogator's efforts to classify a suspect depends upon their experience, ability, availability and accuracy of the information [63]. For instance, the questioning must be designed to develop a detailed account of the suspect's activities before, during, and after the action was committed [2]. Information that is certainly known to interrogator and if suggest the suspect's activities, then these details should be used in formulating questions to determine her/his reactions and to test whether the suspect is inclined to lie. An inaccurate classification may lead to an unsuccessful interrogation or innocent person can be punished [2]. Specially, if the questioning technique based on the original classification is not skillfully modified or changed during the examination.

8.2 INTERVIEWING

Once this information has been collected and probes are prepared, interrogation officer must interview the suspect prior to the fMRI scanning process. This procedure is necessary to determine exactly what the person knows, why he/she has knowledge of certain information relevant to the examination. This phase is also useful to find out about subject's innocence and non-suspicious explanation [2]. Moreover, questioner also observes the verbal and non-verbal behavioral symptoms of deception in the subject [2]. When evidence is weak, interviewer must proceed cautiously by different questions or pictures. The purpose is to place the suspect in a position where he/she will be forced to alter facts that are definitely known to him. It will lead the potential terrorist to believe that answers are already known to the police officer. However, when evidences are strong and when suspect whose relation with the radical organization is reasonably certain, interviewer should assume an air of confidence. He should stress the evidence to analyze the relation of the suspect with terrorists and strive with WHY the suspect committed a terrorist activity rather than IF the suspect committed the terrorism. This process would also help interrogators to remove those stimuli that are significant, not related and independent of the suspicious activity at issue. This process may serve as baseline for security officials. It will ensure that person informed about the targets that will be shown to him in fMRI machine will render a scientific conclusion regarding guilt or innocence.

8.3 SCANNING

After the interviewing, the investigator must select stimuli that is collected through interview process (known to the suspect) to apply the scientific procedure of fMRI. This phase is a scale moving from overt conscious evaluation of stimuli accompanied by response selection of the subject, to unconscious perception constructing meaningful and measurable brain activity [64]. Test administrator must also select irrelevant targets and placed a subject in MRI scanner to analyze scanning parameters by showing a series of words and picture to detect recognition. For example, for the deception task, different types of questions can be visually displayed to the subject with control questions because of the different imaging site. The button-press paradigm will be used to investigate brain activity associated with deception. This task would be designed in a way that the subject would consciously evaluate the stimuli presented and decide whether to press "yes" or "no (i.e. "Yes, I know him" or "No, I do not know him").

The subject is told to press one button to confirm a fact and another button to deny the information as each image is shown. The subject will click a pad button to advance to the next stimuli to keep his/her attention on the scanning test itself. Interviewer must present each question in a way that it is easy to identify the category of the stimuli (e.g., one of the following is the terrorist weapon…) to observe a subject's neural response with. Investigator may also present a suspect with pictures of potential terrorist targets, suspected terrorists, recognition of key people or places and watches movies (e.g. a digital reconstruction of the terrorist scene). These images would generate certain neural responses if the suspect were already familiar or to reveal different information (such as, where a suspect had been or what he/she had seen with another suspect). Thus counterterrorism agencies would be able to distinguish whether the subject was lying based on the BOLD

signal change associated with the response to a particular question [10]. The subject's response can be classified by complex mathematical algorithms recently created by various researchers that are able to analyze imaging data of deception [10].

These responses can be recorded with a software tool such as E-Prime (Psychology Software Tools, Inc. Pittsburgh, PA) and a custom built apparatus such as MRA Incorporated, Washington, PA . Functional MRI analysis can be carried out with different software such as using SPM (Statistical Parametric Mapping software, Welcome Department of Cognitive Neurology, London, UK), BV (BrainVoyager), AFNI (Analysis of Functional NeuroImages), PLS (Partial Least Square), FSL (FMRIB Software Library), AIR (Automated Image Registration) or MIPAV (Medical Image Processing, Analysis, and Visualization).

Security agencies must ensure that subjects are examined medically to ensure that they are safe to enter the MR environment and all the necessary actions are taken. The exclusion criteria that are generally preferred for an MRI Investigation must be used (e.g., claustrophobia, not MR-compatible Transplants, Pacemaker, Insulin pump, Middle ear implants, cardiorespiratory and orthopedic disorders and neuropsychological or sensory impairments, etc.). Usually the subjects in the MRI scanner have an emergency ball in one of their hands with which they can signal at all times that they want to stop the scanning. Subjects should be immediately taken out of the scanner after pressing this ball or at any other sign of emergency, claustrophobia or discomfort. The finding of information present or information absent will recommend a scientific determination of whether the suspect has knowledge of the probe stimuli tested or not. Difference in brain responses among individuals can be used as a baseline for comparison. The results of fMRI analysis will educate law enforcement agencies and judiciary in rendering their verdict about the subject [35].

9 FUNCTIONAL MRI: STRENGTHENING THE CRIMINAL JUSTICE SYSTEM AGAINST TERRORISM

This development will lead to speculations about the development of this neuro-imaging technology that could directly examine the terrorist's memories, intentions and its mind. Interrogators will be able to confidently say that the fMRI told us this detainee lied about X or that he recognizes terrorist Y or fMRI picked him out as a terrorist. This confidence that intelligence operatives will have in this neuroscience technique will be based on aura of infallibility, scientific validity and objectivity [64]. Secondly, the uncritical acceptance of stimulus recognition and lie-detection will be recognized in light of the graphic images that functional MRI is capable of producing. Thirdly, there will be no similar chilling effect like polygraph - when fMRI will be used in high-pressure environments as a part of counterterrorism operations [64]. For instance, the mistreatment of detainees in Abu Ghraib and Guantanamo Bay is arguably the result of worst excesses confidence in the reliability of devices like polygraph.

Researchers are skeptical of claims that an fMRI have potential to identify innocent subject from abuse at the hands of intelligence operatives. However, those who doubt the deployment of this machine in scenarios for good or ill should judge the statement from a United States intelligence officer. Who explained how he and his team "once put a suspects hand on the Xerox machine, turned it on, and told him it was a truth detector and would administer a massive shock if he lied" [64]. Interestingly, the result was positive and the subject "was bluffed into a good confession" [64]. Similarly, such kind of output is also more likely to achieve successfully by functional MRI and we can expect that this tool will successfully be applied in interrogation context with a high degree of confidence [64].

This technology is of course expensive requires extensive support facilities and highly trained staff. However, this mechanism may therefore be most useful in national security scenarios due to the security clearance and complexity of this equipment. This machine is hard to cheat unlike polygraph. It is very easy to deceive polygraphs with a simple internet search that can reveal many ways how to mislead the interviewer. One former polygrapher charges $59.95 for his manual plus DVD offering information on beating this device [65].

10 NEURO-TECHNOLOGY AND NATIONAL SECURITY CONCERNS: REMOVING OBSTACLES TO INVESTIGATING TERRORISM

Although above arguments and studies reported reliability and reasonably high accuracy rates for fMRI studies, there are still significant concerns must be addressed prior to moving this technology to real-world application [14]. In addition to the scientific challenges, advances in fMRI identify numerous social, legal and general public concerns to the process of and the science behind it [65]. Some state that this tool isn't reliable enough to be used outside of a laboratory setting [67]. Thus, these challenges require further investigations to assess its relevance capabilities to national security.

According to some critics, variation in experimental design, situational variables, subject characteristics and the preliminary nature of the existing data are the key scientific challenges in fMRI studies [15] [35]. Large numbers of replicate scans under extremely controlled conditions are needed to accommodate for interscan and intersubject variability [15] [35].

However proponents counter that certain methods and techniques have been developed to overcome for the inbuilt physical limits of fMRI machine [59]. Though, the only feasible technique of increasing the subject's signal is by repeating the scanning several times in order to reach a level where the signal can be heard over the noise and to get the meaningful data [59]. Secondly, critics explore that so far fMRI studies have been conducted in artificial laboratory environments with small numbers of normal (drug free, non-criminal) volunteers to maximize positive results. The criminals and experienced liars were not included and their effects on interrogation are unknown as none of the protocol studies applied to the actual criminal investigations [36].

Other common line of criticism deals with the mental capacity of subjects to record information either during the alleged commission of the terrorism or prior to the fMRI scanning itself. For instance, what if suspect is under extreme emotional distress, under the influence of narcotics or intoxicated or so forth? [65]. According to William Iacono "we don't know enough of how memories are formed during crimes." However, such a criticism is unfounded, as the human brain is always recording information regarding of whether we realize it [68].

Next, various opponents criticize that suspect could attempt to be deliberately deceptive and it is possible for well-prepared terrorist to cheat the fMRI [19]. However, supporters of this technology suggest that fMRI scanning is different from other lie detection tests and it is hard to beat this neuroscience technology [6]. The responses are evaluated by the neural activity and the presence of certain information in their brain suspect not merely for their truth or falsity. Self-deception will have no effect on fMRI testing. A terrorist or a criminal who has convinced himself that he is not guilty; he/she still has recorded information and knows the salient feature of the crime [68]. Thus, our suggested method can measure a brain response at the moment of recognition or if suspect lie. It is equally effective if this test is given to hardened terrorist or pathological liar. A number of opponents claim that fMRI scanning is based on bias [69]. Interrogator could potentially impact the analysis result through the decision process in choosing the specific stimulus [71]. However, [70] has responded that bias is impossible to insert in fMRI scanning process because its responses associated with neuronal activation. The determination of information present or absent is directly revealed by suspect's brain displays made by the fMRI machine and not by the interrogator.

Furthermore, critics of fMRI points out that this technique is not reliable enough to be used as a lie detector in interrogation course of action. It could lead to further abuse of prisoners and human rights violations in the form of torture. It may also allow interrogators to believe more justified in using whatever painful method they use in investigations to extract the information they are looking for [19]. However, advocates strongly recommend that fMRI has potential to minimize the torture dilemma by monitoring involuntary responses and indicating when such fabrications occur. [6] also counter this argument and saying that, pain would appear to be a necessary condition for any kind of physical torture in interrogation but functional MRI is not painful and uncomfortable. It certainly not represents any physical or mental torture and no foreseeable direct health risks associated with its use [27]. The only possible pain that this scan could inflict is to keep the individual motionless. The subject's head is immobilized with foam pillows while inside this machine as suspect's movement could compromise the quality of the scan result. These restraints would almost certainly not inflict to even uncooperative subject and anything near the level of pain that would rise to torture. This practice is contradictory to so-called stress positions that international tribunals have considered to be a torture (for instance, subject being hung by the legs or arms). In fMRI machine, the individual is required only to remain lying down for an extended period that has no relation to the extreme stress positions [10] [12].

It can be concluded that the suspect's body is not physically compromised by this piece of equipment as fMRI is passive, in the sense that it does not enter the body. Now the concern is only with the mental, rather than physical intrusion. Various critics have highlighted that this technology erode the right of fundamental liberty interest in private thoughts. However, [72] is rejecting this assessment and says that this tool does not provide any precise conclusion about a person's thought or what a person is thinking. It can only show a difference across time, across location and across tasks. An fMRI is very good at discovering when brain tissues are active during different cognitive tasks. Thus, regardless of the technological particulars of this tool, it is strongly suggested that fMRI does not violate the right to internal mental privacy [72]. It must be regarded as intruding upon the fundamental liberty interest in private thoughts. Though it can be argued that, a claustrophobic subject might undergo mental suffering in the scanner. However, this distress must have been the result of the use of mind-altering substances or procedures. This psychological condition inherent in the subject is probably not sufficiently severe to rise to the level of torture such as threatened dismemberment or castration. It is also important to note that this scanning cannot be operated on an unconscious person - unlike blood tests, so some form of considerable restraint will be required to ensure that an unwilling suspect remains virtually motionless [27].

Despite the subject's mental condition, he/she would almost certainly not suffer prolonged mental harm because of fMRI, and certainly not constitute torture under any International Human Rights Law [27]. One other objection to the fMRI as a lie

detection raises worrying questions regarding civil liberties [69]. However, we argue that it would be a human rights violation to deny access to fMRI scanning as it serves the cause of human rights and provide scientific means to prove subject's innocence. The potential forensic uses of fMRI also reflect the fact that outrage attending the news about Abu Ghraib and Guantanamo Bay would have been different if prisoners had been examined by fMRI instead of hooded, naked, sexually posed by hostile interrogation. The images featured provoked shock and anger in the society and turned into emblems of degradation and humiliation [26]. Being lying in the fMRI scanner is neither the moral equivalent of being deprived of sleep for 36 hours in a cold torture cell nor legal equivalent of being forced to strip naked and simulate sex with another prisoner.

Opponents also argue that use of this novel method as reliable lie detection will raise different unanswered legal questions [69]. These controversies can be elevated under legal regimes and the international law about privacy and government power. However, science always moves forward, not backward. Hank Greely, Professor of Law at Stanford Law School support this theme and saying that fMRI evidence is certain to be accepted by the courts in future [60]. "The easier, the cheaper, the more pleasant a technique is, the more likely it is to be used in the legal system [60]." [73] also state that, "courts usually seem willing to consider brain imaging evidence under the same standards that they apply to other scientific evidence".

Putting aside all the arguments, but more to the point, how much precision could be increased for fMRI? How accurate should it be to be widely accepted in legal and security settings? Scientists are unable to accurately predict how much the error rate might be reduced? It is also unclear that whether this technology needs to reduce the error rate from 10 percent to something comparable with the billions-to-one accuracy (such as DNA) - will be useful or not? However, given the mechanics of the scientific research involved, it is difficult to conceive of this claim as the legal system has also issues concerning unreliability and repeatability in many procedures [36]. For example, fingerprints experts have sometime claimed perfect accuracy, but a number of pragmatic studies have revealed misidentification rates of about 5% [74].

Critics have also argued against the effort and length of time that would be requisite to acquire an adequate number of probes in interrogation and also in judicial process [27]. However, this problem can be solved by putting more logistic support and by offering more trained staff to the interrogation process to run the fMRI scan and analyze the results effectively. Finally, one of the most common concerns of fMRI scanning may involve the portability of this tool (weighing 20,000 lbs or so). For example, what if intelligence community wants to carry out fMRI scanning on a large number of people or if subjects live in tribal areas, such as Afghanistan? However, we argue that Mobile MRI is the counter strategy that can be used to rebut this objection. Mobile MRI is housed in a highly specialized trailer and a great way to have access to this equipment. It is useful to have access to this facility if a hospital's or imaging center's MRI is not currently functional or not available. This unit adheres to the same strict procedures a fixed MRI unit at imaging center must meet. There is no cause for concern by the subjects if this unit is used for MRI imaging.

Advantages of mobile MRI systems are shorter installation times, lower initial investment and rapid response. Similarly, mobile fMRI will be particularly useful if government has to scan a number of people in remote areas. This service may allow law enforcement agencies to carry out identity checks on suspects in large public occasions and sporting events that could be targeted by the terrorist attacks. It is also useful to access crucial data in challenging environments such as national border control areas military and nuclear power plant zones. However, security officials must make absolutely certain that they are scanning only when they suspect an individual of an offence and can't establish his identity. This action will reduce the number of errors and will rapidly improve security reaction times. Furthermore, fMRI mobile scanning could help police performance with decrease the number of arrest significantly and hasten the speed of criminal investigations.

The potential of fMRI in attempt to transfer this technology outside the research context poses several challenges in the context of national security. However, apart from many challenges, critiques have to bear in mind that functional MRI is just two decades old. Scientists reviewing the ability of photography 20 years before could not visualize the idea that one day this device would be able to determine images of planets orbiting other planets and resolve images less than a fraction of a second long at micrometer scales – which has now been done [75]. According to Vanderbilt's Frank Tong, "If brain scans were admissible in court, and became popular enough, then even if they were not mandatory they would become in a sense obligatory. Because if you didn't voluntarily undergo it, then there would be the question, `Why didn't you take the test? '

Secondly, for decades, polygraph has been widely used in interrogation by law enforcement agencies and has long been rife in the courtroom despite their flaws. Even supporters of this device confess a 10 percent failure rate [76]. Brandon L. Garrett, the law professor of the University of Virginia analyzed 200 cases in his published study in which innocent people were wrongly convicted by the courts. He found that in 55 percent of these cases, courts had been presented with faulty forensic evidence such as DNA and polygraph [77].

Further pointed out that fMRI is ethically acceptable in the market to the same extent as traditional polygraphs. If suspects are permitted to undergo a traditional polygraph examination, the argument is equally strong concerning fMRI scans as it is superior to the polygraph in accuracy and reliability [78]. The involuntary information extract from subject's mind should be considered as fundamental liberty interest. This right must be "deeply rooted in this Nation's history and tradition" and "implicit in the concept of ordered liberty". In the eyes of society and international law, fMRI based interrogation would be less objectionable than interrogation based on torture and physical beatings of naked hooded bodies. By contrast, fMRI is less invasive and harmful that can be legally defended by law and by the society [27]. However, this is true that ethical conflicts and criticism often arise when clinical technology is used for non-clinical purposes. There is a need to build an elite interrogation unit and a call for a greater partnership to employ policies and to counter above threats.

11 DEVELOP THE NEW ELITE INTELLIGENCE UNIT: A NEW ALLIANCE ARCHITECTURE

It is vital to establish an elite interrogation unit that must be filled with skilled intelligence professionals, neuroscientists, neuro-ethicists and other qualified individuals who must be knowledgeable about the application of fMRI and understand the limitations. The unit members must continue to uphold principles of medical ethics and the development of interrogation strategies must be addressed to protect public interest. For instance, involvement of neuroscientists in this unit for intelligence gathering is necessary as they can perform physical and mental assessment of subjects to provide medical care and to disclose the limit of access to the medical record. Secondly, the role of neuro-ethicists in the panel is essential to inform policy discussions about setting up the necessary infrastructure to protect the privacy of suspects. Thirdly, this elite interrogation panel will be helpful to inform the general public about the ethical, legal and social implications of this technology and permissible interpretations of test results without contributing to technology hype.

12 A NATIONAL STRATEGY & COALITION IS NEEDED TO GUIDE OUR PREPAREDNESS EFFORTS

Advances in fMRI have necessitated discussions on the ways this neuroscience tools could be used as a weapon in the war on terror. However, among the many challenges to this application, a central one is the partnership between major stakeholders. The primary reason is of course a lack of neuroscience expertise and the frequent unwillingness of the scientific community itself to engage and in dialogue with high levels of government level. Secondly, to work against political agenda that promotes that tools like fMRI perceived as wrong, misguided or even dangerous for general public. These people are those whose finances or status depends on the old means of doing things, and this group of people often resists progress because they see it as a threat for their own ways. To work against the resistance to fMRI application requires the full commitment and engagement of experts that resides only within the scientific community [79]. We recommend that a four-way partnership is needed between intelligence officials, neuroscientists, neuro-ethicists and policy makers to serve the national security interests. This goal can only be achieved by the concerted efforts, imaginative thinking, planning, coordination and participation of each of these groups.

13 EVIDENCE-BASED POLICIES AND GUIDELINES: A RELIABLE RESPONSE TO PUBLIC CONCERNS

This paper recommends that investigators involved in screening process must be made aware of the issues raised by fMRI to develop best practices and efficient internal measures. Neuroscientists may contribute in developing effective interrogation strategies for general training purposes for investigators that must be humane and respect the rights of individuals. Secondly, training of interrogators is one of the major challenges for the implementation of this technology. This training is necessary for the evaluation of interrogation centers to appropriately protect subjects while allowing for scanning. Thus, only trained experts will be required to evaluate subjects and conduct the scan. Furthermore, this education will help to establish proactive and defensive knowledge of scientific and technological capabilities of fMRI analyses. More important, it will assist new elite interrogation unit to identify what systems, methods, or processes of interrogation are best to protect the nation's security. This guidance will also address the ethico-legal and social issues and the vulnerabilities they exploit. The principal benefit of this training is to obtain knowledge from suspects that will increase investigator's understanding of terrorist adversaries and may assist them in developing potential countermeasures [80].

Thirdly, Neuroscientists are ethically obligated to report to the appropriate authorities when they have reason to believe that interrogation is coercive and violating human rights. They must ensure that if experts do not detect any abnormal behavior, the subject is not harmed. However, if an abnormality is detected, the results of the scan should be analyzed by other highly trained neuroscientists and possibly rectified. Fourthly, it is also important that professionals involved in interrogation will be required to acquire security clearances. This shield will make it impossible for them to share the findings with colleagues in unclassified settings. Fifthly, the "Certificate of Confidentiality and Privacy" issued by the new elite

interrogation unit can provide additional protection and can make a difference in the interrogational context. This certificate will allow the members who have access records to refuse to disclose identifying information at the civil, criminal, legislative, federal, state, or local level if the subject is not guilty. Disclosure of sensitive information could have adverse consequences on innocent person's reputation, employability as well as financial standing. The revelation of such knowledge could reasonably lead to social stigmatization or discrimination. This credential is necessary to protect data relating to persons' sexual attitudes, genetic information, use of alcohol, drugs and other different practices and preferences. This document will particularly encourage subjects (e.g., in employee screening) to participate in scanning process. In sum up, the Certificate of Confidentiality and Privacy will ensure that informed consent is appropriate, risks are minimized and protections are adequate [80].

Sixthly, in case of suspicious employee screening (e.g., Nuclear power plant), employee's right must be protected by Article 8 of the European Convention on the Protection of Human Rights and Article 12 of the Universal Declaration of Human Rights. The interrogation process must implement the United Nations International Labour Organisation (ILO) code of practice on the Protection of Workers' Personal Data (1996) as well as European Union Guidelines 95/46 and 97/66 on data protection. The access to the results should be restricted for interrogators in order to prevent the misuse of these preliminary data. It is important that counterterrorism agencies must ensure the safety of the subjects through the systematic monitoring of the international law and human rights – including the United Nations Conventions Against Torture, the International Covenant on Civil and Political Rights, and the Universal Declaration of Human Rights. The state must also consider the nuances of the Geneva Conventions as applied to suspected terrorists. Finally, uniformed personnel's and medical experts who are engaged in interrogation panel using fMRI must be held to account for their actions if they have violated human rights laws. Innocent subjects or victim of this technology must be offered compensations, health care services and a formal apology to address ethical violations caused by this technology or by the professionals. A comprehensive federal investigation is required if the public trust in the ethical integrity of the security and medical profession being seriously compromised. If interrogators dismiss a subject for failing an fMRI scan test, they must be able to justify the action against him/her under the influence of a Human Rights Act, such as the European Convention on Human Rights (ECHR) or the UK Human Rights Act 1998 [80].

Innovation in technology has been a key driver of change - the defense and security arenas are no exception [87] [88]. Similarly, members of elite interrogation unit should be well aware of current knowledge, novel literature, latest technologies, valuable processes and services about fMRI scanning for the purpose of developing image analysis to improve investigating methods. It is a major step forward in the action to our national interests that will continue to play a key role in the effectiveness of fMRI as a counterterrorism tool. We also recommend that government must push promising research on fMRI as they could meet our defense needs through collaboration with research sectors and universities to ensure a strong research base in this area. This action must be vibrant, inventive and innovative that looks most promising in interrogational neuroimaging. Investigators and neuroscientists must grasp the opportunities and adapt them quickly and effectively as this benefit is critical to our security and sovereignty [80].

14 THE COSTS AND BENEFITS OF INTERROGATIONAL NEUROIMAGING IN THE STRUGGLE AGAINST TERRORISM

Detecting deception and intelligence gathering from human resources is increasingly important to protect vital national security interests. We argued that fMRI has a potential to detect the neuro-circuitry involved in deception. This technology can support state's struggle against terrorism by understanding the brain basis of deception, so that the means for dealing with terrorists are developed in a timely manner. However, in a democracy, the legitimacy of state's action is important to maintain support for what the government does in the war on terror with in rule of law to protect public interest [80]. Knowledge is power, and certainly advance security measure government could take is to ensure that general public is aware of its benefits and threats. In this regard, it is important to pay explicit attention on adopting fMRI with cost and benefit analysis. Thus, the decision might be improved with great understanding and confidence that it is widely believed to be the right thing to do.

Firstly, apart from the challenges we described above, fMRI has several other disadvantages as a tool for lie deception. For instance, it is time-consuming and expensive process [81]. Secondly, fMRI need a separate control room (magnetically shielded) filled with computers, power supplies and data storage devices that require a significant capital investment from state [82]. Thirdly, the noise level during examination is uncomfortably high that needs protective ear coverings for subject. Fourthly, a relatively minor head or body motion during the scan can spoil the analysis [82]. Unfortunately, these movements could be effective countermeasures for resistant terrorist. In addition to these challenges there are also some safety hazards associated with this scanning. People with claustrophobia and pregnant women are generally not scanned for obtaining information [19].

On the other side, the use of vulnerable populations (such as prisoners) for the new interrogation techniques has a long and disturbing history filled with misguided unethical experimentation. Those who designed, control, monitor and supervised these alleged practices – whether security officials or health professionals claims to be in the service of national security objectives. Though, sometime this practice faces conflict with the interests of those whom they are monitoring such as suspects of crime. However, given the attacks on 9/11 in New York and those in 2004 in Madrid and London in 2005, the public is well aware of a heightened threat of terrorism to national security. As a result, the public has generally accepted the government's new steps, new tools and new ways of thinking to fight terrorism [83]. Nonetheless, this war requires constant vigilance and the commitment of resources on all fronts. For instance, the perpetrators of 9/11 used commercial airplanes as a major weapon. This attack exposed major weaknesses in the existing immigration system and border security. In this regard, air travelers have adjusted to the need for more intensive passenger screening on airport [83]. Incredibly, there is a willingness from general public to provide, iris scan, fingerprints, and other biometric screening methods to acquire secure identification [83].

The cost of liberty is high, but it is a price people always have been, always will be and willing to pay. Public support is a strategic instrument and a vital component and there is a great deal of complacency amongst the public in the war on terror. Similarly, there is a great expectation among scientists and counterterrorism agencies that public will also realize the urgency of the threat and the significance of interrogational neuroimaging application to national defense. We have proposed our novel experimental methodology and guidelines that introduce a first step toward developing reliable and practical interrogation applications. This paradigm will provide counter terrorism agencies with cost effective approaches that could have a profoundly beneficial impact on society. It will allow interrogators to focus their investigations on the suspects who actually commit terrorism. Thus, innocents can be treated with the dignity befitting human beings. It would appear from this research judgment and assessment that fMRI has a greater probability for success to identify recognition and lie detection during interrogation procedure.

We also argued that some of the claims are unfounded such as concerns about privacy, confidentiality and torture. It has potential to truly deliver what its advocates such as cognitive liberty and potential to replace torture and aggressive existing interrogation strategies that inevitably violate the core human rights obligations. The goal is to create an environment where neither torture nor coercive interrogation is permissible. Thus, implementation of fMRI may render the dark art of interrogation unnecessary in the Global War on Terrorism. Armed with this neuro-imaging technology, investigators will no longer feel the need to torture or use 'torture-lite' interrogation tactics. An fMRI is compatible with human rights law and information can now be achieve without leaving a physical trace of the trauma of torture [13].

More significant, consideration must also be given to the government's purpose in subjecting the suspect to fMRI scan. It is important that state's interest in interrogating high-value terrorists may be justifiable and most likely not to rise to the conscience shocking level and will not injure substantial liberty interest. Whether or not policy makers or civilized society can or should allow brain scanning is a matter that will continue to be debated for years to come. However given only the terrible choice of permitting the death of many innocent people OR scanning an individual ,who can possibly prevent mass causalities, the state have to make sensible decisions what is necessary to save lives.

REFERENCES

[1] Redlich, A. D. & Meissner, C. A. (2009). Techniques and controversies on the interrogation of suspects: The artful practice versus the scientific study. In J. Skeem, K. Douglas, & S. Lilienfeld (Eds.), Psychological science in the courtroom: Controversies and consensus (pp. 124-148). New York: Guilford Publications, Inc.
[Online] Available: http://psycnet.apa.org/?&fa=main.doiLanding&uid=2009-11373-006

[2] Gordon, N. J., & Fleisher, W. L. (2006). Effective Interviewing and Interrogation Techniques, Third Edition by Academic Press, Burlington, MA. ISBN 0123694906
[Online] Available: http://www.amazon.com/Effective-Interviewing-Interrogation-Techniques-Fleisher/dp/0122603818

[3] Eck, M. (1970). Lies and truth. New York: McMillan. Eriksson, Anders. & Lacerda, Francisco. (2007). Charlatanry in forensic speech science: A problem to be taken seriously, International Journal of Speech, Language and the Law [formerly Forensic Linguistics], 2007, Vol. 14, No. 2, pp. 169 – 193.
[Online] Available: http://books.google.at/books/about/Lies_and_truth.html?id=QRVwAAAAMAAJ&redir_esc=y

[4] Langleben, D. D., Loughead, J. W., & Bilker, W. B. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. Hum Brain Mapp, 26:262–72.
[Online] Available: http://www.med.upenn.edu/langleben/neuroimage15_2002.pdf

[5] Khoshbin, L. S., & Khoshbin, S. (2007). Imaging the mind, minding the image: An historical introduction to brain imaging and the law. American Journal of Law and Medicine, 33, 171-192.

[Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/17910156

[6] Spence, S. A., Hughes, C. J., Brook, M. L., Lankappa, S. T., & Wilkinson, I. D. (2008). 'Munchausen's syndrome by proxy' or a 'miscarriage of justice'? An initial application of functional neuroimaging to the question of guilt versus innocence. European Psychiatry, 23: 309-314. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/18029153

[7] Faulkes, Z. (2011). Can brain imaging replace interrogation and torture? Global Virtue Ethics Review 6(2): 55-78. [Online] Available: http://www.spaef.com/article.php?id=1266

[9] Kozel, F. A., Johnson, K. A., Grenesko, E. L., Laken, S. J., Kose, & X. S. Lu. (2009). Functional MRI detection of deception after committing a mock sabotage crime. Journal of Forensic Sciences, 54(1), 220-231. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/19067772

[10] Haynes, J. D, & Rees, G. (2006). Decoding mental states from brain activity in humans. Nat Rev Neurosci 7:523-534. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/16791142

[11] Davatzikos, C., Ruparel, K., & Fan, Y. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. Neuroimage, 28:663– 8. [Online] Available: http://www.sciencedirect.com/science/article/pii/S1053811905005914

[12] Phan, K. L., Magalhaes, A., and Ziemlewicz, T. J. (2005). Neural correlates of telling lies: a functional magnetic resonance imaging study at 4 Tesla. Acad Radiol 12:164 –72. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/15721593

[13] Mohamed, F. B., Faro, S. H., & Gordon, N. J. (2006). Brain mapping of deception and truth telling about an ecologically valid situation: functional MR imaging and polygraph investigation: initial experience. Radiology 238:679–88. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/16436822

[14] Marks, J. H. (2007). Interrogational Neuroimaging in Counterterrorism: A "No-Brainer" or a Human Rights Hazard? American Journal of Law and Medicine, 33, 483-500. [Online] Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1005479

[15] Garnett, A., Whiteley, L., Piwowar, H., Rasmussen, E., & Illes, J. (2011). Neuroethics and fMRI: Mapping a Fledgling Relationship. PLoS ONE 6(4): e18537. doi:10.1371/journal.pone.0018537 [Online] Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0018537

[16] Bizzi, E., Hyman, S. E., Raichle, M. E., Kanwisher, N., Phelps, E. A., Morse, S. J., Sinnott- Armstrong, W., Rakoff, J. S., & Greely, H. T. (2009). Using Imaging to Identify Deceit: Scientific and Ethical Questions. Cambridge, Mass: American Academy of Arts and Sciences. [Online] Available: http://www.amacad.org/pdfs/deceit.pdf

[17] Wolpe, P. R., Foster, K., & Langleben, D. (2005). Emerging neurotechnologies for lie detection: Promises and perils. American journal of Bioethics 5(2):39-49. [Online] Available: http://ajobonline.com/journal/j_articles.php?aid=728

[18] Taylor, M. K., Horning, D. S., Chandler, J. F., Phillips, J. B., Khosravi, J. Y., Bennett, J. E., Halbert, H., Fern, B. J., & Gao, H. (2011). A comparison of approaches to detect deception (Report No. ADA537848). [Online] Available: http://www.au.af.mil/au/awc/awcgate/navy/approaches-to-deception-detection.pdf

[19] Moore, W. H. (2010). Incarceration, Interrogation, and Counterterror: Do (Liberal) Democratic Institutions Constrain Leviathan? PS: Political Science & Politics, vol. 43, pp. 421-424. [Online] Available: http://mailer.fsu.edu/~whmoore/garnet-whmoore/research/Moore2010PS.pdf

[20] Robert, A., Fein, P. L., Bryan, V. (2006). Educing information: Interrogation-science and art: foundations for the future: phase 1 report" published by the Center for Strategic Intelligence Research. [Online] Available: http://www.fas.org/irp/dni/educing.pdf

[21] Kassin, S. M., Drizin, S. A., Grisso, T., Gudjonsson, G. H., Leo, R. A., & Redlich, A. D. (2010). Police-induced confessions: Risk factors and recommendations. Law and Human Behavior, 34, 3_38. doi:10.1007/s10979-009-9188-6. [Online] Available: http://web.williams.edu/Psychology/Faculty/Kassin/files/White%20Paper%20-%20LHB%20(2010).pdf

[22] Allen, S., Chaffee, D., & Hashemian, F. (2007). Leave No Marks: Enhanced Interrogation Techniques and the Risk of Criminality. Cambridge (MA): Physicians for Human Rights; and New York (NY): Human Rights First. [Online] Available: http://physiciansforhumanrights.org/library/documents/reports/leave-no-marks.pdf.

[23] Granhag, P. A. & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mindreading. Psychology, Crime & Law, 14, 189-200. [Online] Available: http://www.tandfonline.com/doi/abs/10.1080/10683160701645181#preview

[24] Farwell, L. (2012). Brain fingerprinting: a comprehensive tutorial review of detection of oncealed information with event-related brain potentials,Cognitive Neurodynamics 6 , no. 2, 115-154. [Online] Available: http://www.citeulike.org/article/10376056

[25] Maschke, G. W. & Scalabrini, G. J. (2005). The Lie Behind The Lie Detector. 16-17 (4th digital ed. 2005). [Online] Available: http://www.antipolygraph.org/lie-behind-the-liedetector.pdf.

[26] Wang, G., Chen, H., & Akabashch, H. (2004). Automatically detecting deceptive criminal identities. Communications of the ACM, 47 (3), 71-76. [Online] Available: http://dl.acm.org/citation.cfm?id=971618

[27] Costanzo, M. & Gerrity, E. (2009). The effects and effectiveness of using torture as an interrogation device –using research to inform the policy debate, Social Issues and Policy Review, Vol. 3, No. 1 pp. 179-210.
[Online] Available: http://dl.acm.org/citation.cfm?id=971618
[28] Thompson, S. (2005). The legality of the use of psychiatric neuroimaging in intelligence interrogation. Cornell Law Rev 90:1601–37. [Online] Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=656841
[29] Mayerfeld, J. (2007). Playing by Our Own Rules: How U.S. Marginalization of International Human Rights Law Led to Torture, 20 HARV. HUM. RTS. J. 89, 107, 135-36.
[Online] Available: http://faculty.washington.edu/jasonm/12-Mayerfeld.pdf
[30] Cole, S. (2009). "Cultural Consequences of Miscarriage of Justice," Behavioral Sciences & the Law, vol: 27 no: 431 – 449.
[Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/19402029
[31] Coulam, R. (2006). Approaches to Interrogation in the Struggle against Terrorism: Considerations of Cost and Benefit. In: SWENSON, R. (ed.) Educing information, Interrogation: Science and Art, Foundations for the Future.
[Online] Available: http://www.fas.org/irp/dni/educing.pdf
[32] Stone, J. A., David, P. Shoemaker, Nicholas R. Dotti (2008). Interrogation: World War II, Vietnam, and Iraq, National Defense Intelligence College Washington, Dc., ISBN 978-1-932946-23-9, p.31-34.
[Online] Available: http://www.ndic.edu/press/12010.htm
[33] Lasson, K. (2008). Torture, Truth Serum, and Ticking Bombs: Toward a Pragmatic Perspective on Coercive Interrogation, Loyola University Chicago Law Journal, vol: 39; no: 2, pages 329-360.
[Online] Available: http://www.luc.edu/law/activities/publications/lljdocs/vol39_no2/lasson.pdf
[35] Galit, N., Vrij, A. & Fisher, R. P. (2012). Does the truth come out in the writing? Scan as a lie detection tool. Law and Human Behavior, Vol 36(1), 68-76. doi: 10.1037/h0093965
[Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/21253849
[36] Bruni, T. (2012). Cross-Cultural Variation and fMRI Lie-Detection. Technologies On The Stand: Legal And Ethical Questions In Neuroscience And Robotics, pp. 129-148, B. Van den Berg, L. Klaming, eds., Nijmegen: Wolf Legal Publishers, 2011. [Online] Available: http://ssrn.com/abstract=1983536
[37] McCabe, D. P., Castel, A. D. & Rhodes, M. G. (2011). The influence of fMRI Lie Detection Evidence on Juror Decision-Making', Behavioral Sciences and the Law, 29: 566-577.
[Online] Available: http://castel.bol.ucla.edu/publications/McCabe%20Castel%20Rhodes%20BSL%20in%20press.pdf
[38] Simpson, J. R. (2008). Functional MRI lie detection: too good to be true? J Am Acad Psychiatry Law 36:491– 8.
[Online] Available: http://www.jaapl.org/content/36/4/491.abstract
[39] Elaad, E., (2004). "Detection of Deception," in Encyclopedia of Forensic Sciences, A. S. Editor-in-Chief: Jay, Ed., ed Oxford: Elsevier, pp. 550-556.
[Online] Available: http://onlinelibrary.wiley.com/doi/10.1002/9780470061589.fsa495/pdf
[40] Zhou, L., Judee, K., Burgoon, J. F., Nunamaker, JR. & Doug, T. (2004). Automating Linguistics-Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communications, Group Decision and Negotiation, vol. 13, pp. 81-106, 2004/01/01. [Online] Available: http://www.iula.upf.edu/materials/050603vazquez.pdf
[41] Armistead, T. W. (2011). Detecting deception in written statements: The British Home Office study of scientific content analysis (SCAN), Policing: An International Journal of Police Strategies & Management, Vol. 34 Issue: 4, pp.588 – 605.
[Online] Available: http://www.emeraldinsight.com/journals.htm?articleid=17003361
[42] Rose, D. (2004). Operation Take Away My Freedom: Inside Guantanamo Bay on Trial. Vanity Fair.
[Online] Available: http://www.vanityfair.com/politics/features/2004/01/guantanamo200401
[43] Lewis, M. W. (2010). A Dark Descent into Reality: Making the Case for an Objective Definition of Torture, 67 Wash. & Lee L. Rev. 77, 121–25. [Online] Available: http://law.wlu.edu/deptimages/Law%20Review/67-1Lewis.pdf
[44] Brandon, S. E. (2011). Impacts of psychological science on national security agencies post-9/11. American Psychologist, Vol 66(6), 495-506. doi: 10.1037/a0024818 [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/21823770
[45] Pustilnik, A. C. (2012). Pain as a fact and heuristic: how pain neuroimaging illuminates moral dimensions of law. University of Maryland School of Law, USA. Cornell Law Rev May; 97(4):801-48.
[Online] Available: http://www.lawschool.cornell.edu/research/cornell-law-review/upload/Pustilnik-final.pdf
[46] Kassin, S. (2010). Interviewing Suspects: Practice, Science, and Future Directions. Legal & Criminological Psychology, 15(1), 39-55. [Online] Available: http://onlinelibrary.wiley.com/doi/10.1348/135532509X449361/abstract
[47] Granhag, P. A., & Vrij, A. (2012). Eliciting cues to deception and truth: What matters are the questions asked, Journal of Applied Research in Memory and Cognition, vol. 1, pp. 110-117.
[Online] Available: http://www.sciencedirect.com/science/article/pii/S2211368112000320
[48] Sahito, Farhan.; Slany, Wolfgang. (2012), "Functional Magnetic Resonance Imaging and the Challenge of Balancing Human Security with State Security", Human Security Perspectives 1 (European Training and Research Centre for Human Rights and Democracy (ETC), Graz, Austria) 2012 (1): 38–66.

[49] Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. Nature, 453(7197), 869-878. [Online] Available: http://www.nature.com/nature/journal/v453/n7197/full/nature06976.html

[50] Fenton, A. & Baylis, F. (2009). Ethical challenges and interpretive difficulties with non-clinical applications in pediatric fMRI. American Journal of Bioethics (AJOB Neuroscience); 9(1):3–13.

[51] Robinson, R. (2004). fMRI beyond the clinic: Will it ever be ready for prime time? PLoS, 2, 715-717. [Online] Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC423132/

[52] Spence, S. A., Farrow, T. F. and Herford, A. E. (2001). Behavioral and functional a natomical correlates of deception in humans. Neuroreport 12:2849 –53. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/11588589

[53] Lee, T. M. C., Liu, H. L., & Tan, L. H. (2002). Lie detection by functional magnetic resonance imaging. Hum Brain Mapp 15:157– 64. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/11835606

[54] Lee, T. M. C., Liu, H-L., & Chan, C. C. (2005). Neural correlates of feigned memory impairment. Neuroimage 28:305–13. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/16165373

[55] Nunez, J. M., Casey, B. J., and Egner, T. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. Neuroimage 25:267–77. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/15734361

[56] Ganis, G., Kosslyn, S. M., & Stose, S. (2003). Neural correlates of different types of deception: an fMRI investigation. Cerebral Cortex 13: 830–6. [Online] Available: http://cercor.oxfordjournals.org/content/13/8/830.abstract

[57] Langleben, D. D., Schroeder, L., & Maldjian, J. A. (2002). Brain activity during simulated deception: an event-related functional magnetic resonance study. Neuroimage 15:727–32. [Online] Available: http://www.med.upenn.edu/langleben/neuroimage15_2002.pdf

[58] Kozel, F. A., Johnson, K. A., & Mu, Q. (2005). Detecting deception using functional magnetic resonance imaging. Biol Psychiatry 58:605–13. [Online] Available: http://www.musc.edu/pr/fmri.pdf

[59] Kozel, F. A., Revell, L. J., & Lorberbaum, J. P. (2004a). A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men. J Neuropsychiatry ClinNeurosci 16:295–305. [Online] Available: http://neuro.psychiatryonline.org/article.aspx?articleid=101888

[60] Kozel, F. A., Padgett, T. M., & George, M. S. (2004b). A replication study of the neural correlates of deception. Behav Neurosci 118:852–6. [Online] Available: http://www.personal.psu.edu/krm10/PSY597SP07/Kozel%20neural%20correlates.pdf

[61] Law, J. R. (2012). Cherry-Picking Memories: Why Neuroimaging-Based Lie Detection Requires a NewFramework for the Admissibility of Scientific Evidence under FRE 702 and Daubert (Winter). 14 Yale Journal of Law & Technology 1. [Online] Available: http://ssrn.com/abstract=1582262

[62] Brown, T. & Murphy, E. R. (2009). Through A Scanner Darkly: The Use of fMRI As Evidence of Mens Rea. Journal of Law & Health 22.2: 319-341. [Online] Available: http://develdrupal. law.csuohio.edu/currentstudents/studentorg/jlh/documents/gMurphyBrownTranscript.pdf

[63] Tim, B. (2011). Mindreading: From Neuroimaging to the Philosophy of Mind, Humanities Research Showcase, University of Oxford.

[65] Marks, D. H. (2008). Interrogation Using Functional MRI And Cognitive Engrams. Journal of the Institute of Justice and International Studies. University of Central Missouri, Institute of Justice & International Studies. [Online] Available: http://www.highbeam.com/doc/1P3-1621658721.html

[66] Inbau, F. E. & Reid, J. E. (1962). Criminal interrogation and confessions. Baltimore, MD: Williams and Wilkins. [Online] Available: http://www.amazon.com/Criminal-Interrogation-Confessions-Fred-Inbau/dp/0763747211

[67] Matthew, H. (2008). One Image, One Thousand Incriminating Words: Images of Brain Activity and the Privilege against Self-Incrimination, 27 TEMP. J. SCI. TECH. & ENVTL. L. 141, 144-50. [Online] Available: http://www.temple.edu/law/tjstel/2008/spring/v27no1-Holloway.pdf

[68] Greely, HT., and Illes, J. (2007). Neuroscience-based lie detection: The urgent need for regulation. American Journal of Law and Medicine 33(2 and 3):377–431. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/17910165

[69] Edersheim, J. G., Rebecca, W. B., Bruce, H. P. (2012). Neuroimaging, Diminished Capacity and Mitigation," in Neuroimaging in Forensic Psychiatry, ed: John Wiley & Sons, Ltd, pp. 163-193. [Online] Available: http://onlinelibrary.wiley.com/doi/10.1002/9781119968900.ch10/summary

[70] Nugent, K. M. (2012). Practical Legal Concerns, in Neuroimaging in Forensic Psychiatry, ed: John Wiley & Sons, Ltd, , pp. 253-273. [Online] Available: http://onlinelibrary.wiley.com/doi/10.1002/9781119968900.ch15/summary

[71] Taylor, E. (2006). A New Wave of Police Interrogation? "Brain Fingerprinting," The Constitutional Privilege Against Self-Incrimination, and Hearsay Jurisprudence, U. ILL. J.L. TECH. & POL'Y 287. [Online] Available: http://www.jltp.uiuc.edu/archives/Taylor.pdf

[72] Miller, G. (2009). Brain Scans of Pain Raise Questions for the Law, 323 Science 195. [Online] Available: http://www.sciencemag.org/content/323/5911/195.summary

[73] Marks, D. H., Adineh, M., & Gupta, S. (2006). Determination of Truth from Deception Using Functional MRI and Cognitive Engrams. The Internet Journal of Radiology. Volume 5 Number 1. DOI: 10.5580/1847. [Online] Available: http://www.ispub.com/journal/the-internet-journal-of-radiology/volume-5-number-1/determination-of-truth-from-deception-using-functional-mri-and-cognitive-engrams.html

[74] Shen, F. X. & Jones, O. (2011). Brain Scans as Evidence: Truths, Proofs, Lies, and Lessons. Mercer Law Review, Vol. 62, p. 861, Vanderbilt Public Law Research Paper No. 11-2.[Online] Available: http://ssrn.com/abstract=1736288

[75] Jones, O. D., Buckholtz, J., Schall, J., & Marois, R. (2009). Brain Imaging for Legal Thinkers: A Guide for the Perplexed. Stanford Technology Law Review, Vol. 5, Vanderbilt Public Law Research Paper No. 10-09. [Online] Available: http://ssrn.com/abstract=1563612

[76] Pettit, M. (2007). FMRI and BF Meet FRE: Brain imaging and the federal rules of evidence. American Journal of Law and Medicine, 33, 319-340. [Online] Available: http://www.aslme.org/index.php/fmri-and-bf-meet-fre-brain-imaging-and-the-federal-rules-of-evidence.html

[77] Saks, M. J. & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. Science, 309 (5736), 892-895. [Online] Available: http://www.sciencemag.org/content/309/5736/892.full

[78] Marois, C., Macintosh, B., Soummer, R., Poyneer, L., & Bauman, B. (2008). In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7015, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series.

[79] Grubin, D. MD. (2010). The Polygraph and Forensic Psychiatry. J Am Acad Psychiatry Law 38:4:446-451. [Online] Available: http://www.jaapl.org/content/38/4/446.full

[80] Sahito, F.; Slany W., (2013). "Advanced Personnel Vetting Techniques in Critical Multi-Tennant Hosted Computing Environments" International Journal of Advanced Computer Science and Applications (IJACSA). Vol. 4, No.5, 2013.

[81] White, A. E. (2010). The lie of fMRI: an examination of the ethics of a market in lie-detection using functional magnetic resonance imaging. HEC Forum 22(3):253–266. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/20730595

[82] Kitcher, P. (2003). Science, truth, and democracy. Oxford, UK: Oxford University Press. [Online] Available: http://www.amazon.com/Science-Democracy-Oxford-Studies-Philosophy/dp/0195145836

[83] Kroenig, M. & Pavel, B. (2012). How to Deter Terrorism, The Washington Quarterly. [Online] Available: https://csis.org/files/publication/TWQ_12Spring_Kroenig_Pavel.pdf

[84] Bennett, C. M. & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? Annals of the New York Academy of Sciences. [Online] Available: http://www.ncbi.nlm.nih.gov/pubmed/20392279

[85] Heckman, K. E. & Happel, M. D. (2006). Mechanical Detection Of Deception: A Short Review, in: Swenson, Russell (ed.), Educing Information: Interrogation: Science And Art Foundations For The Future, National Defense Intelligence College Press, Washington, D.C, pp. 63-93. [Online] Available: http://www.newtactics.org/en/node/2617

[86] Turhan, C., Susan, B., William, C., Philip, J., Crowley, Don DuRousseau, Henry T. Greely, Alvaro Pascual-Leone. (2007). Neuroethics and National Security, American Journal Of Bioethics: Neuroscience 7:(5) 3-13. [Online] Available: http://www.tuftsgloballeadership.org/files/Neuroethics-Argument.pdf

[87] Sahito, Farhan Hyder, Wolfgang Slany, and Syed K. Shahzad. "Search engines: The invader to our privacy—A survey." Computer Sciences and Convergence Information Technology (ICCIT), 2011 6th International Conference on. IEEE, 2011.

[88] Sahito, Farhan, Atif Latif, and Wolfgang Slany. "Weaving twitter stream into linked data a proof of concept framework." Emerging Technologies (ICET), 2011 7th International Conference on. IEEE, 2011.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 608-615

http://www.issr-journals.org/ijias/

Dynamic Channel Allocation in Cellular Networks

Arafat Abu Mallouh[1], Christian Bach[2], and Abdullah A. Abdullah[1]

[1]Computer science Department,
University of Bridgeport,
Bridgeport, CT, USA

[2]Technology management Department,
University of Bridgeport,
Bridgeport, CT, USA

ABSTRACT: The wireless technology and its application growing faster and faster in last decades. Mobile network is one of the fastest growing technologies in wireless network. This headed to some challenges that face mobile network such as how to serve the big number of users, efficiently of frequencies is scarce and interferes with each other. One of the solutions to deal with such challenges is Cellular Networks which is used to divide a geographical area in to cells so that we can reuse the scarce frequencies in order to support more users and also to decrease interference. This paper introduces the importance of dynamic channel allocation in cellular networks and how much gain could be utilized by this technique. The Methodology depend on an intensive reading of what other research has been done in the field, then the model factors and the goal was built according to the main importance issues in this field. In order to realize the complications and limitations of the topic and to have comprehensive understanding many work in the literature have been revised. The mechanism was tested in two different scenarios, with uniform and non-uniform load distribution. For the findings: A new mechanism was introduced to overcome the previous limitations and to gain more efficient results. Also it utilizes artificial intelligence approach to make the allocation process optimal. Moreover, the new mechanism depends on four factors cell size, coordination, frequency reuse, and hand over to make the allocation process efficient and reliable.

KEYWORDS: Channel Allocation, Cellular, Cooperation, Frequency Reuse, Dynamic.

1 INTRODUCTION

The development in wireless and mobile applications in the last decade was huge and it's expected to grow more in the future. The domain of applications for wireless networks includes communications, military, media, medical, and much more. This led to challenging problems to deal with, like how to serve this big number of users efficiently given that frequencies are scarce and interfere with each other. One of the solutions to deal with such challenges is Cellular Networks which is used to divide a geographical area in to cells so that we can reuse the scarce frequencies in order to support more users and also to decrease interference (Fig 1). When cellular approach is applied frequencies could be allocated in several ways, one is called fixed channel allocation, in this approach channels are allocated for each cell and do not change then, the number of channels depends on the number of users in each cell and the allocated channels for neighboring cells, since cells that are close in distance will suffer from interference. [1] [2] [3]

Another approach is dynamic channel allocation in which the allocation of channel for each cell is dynamic which means the for each cell the number of channels depends on the varying numbers of users inside that cell, if the number of user increases the allocated channel will increase too without causing interference or system failure, in this approach if a cell needs more channels then the system will borrow idle channels from cells that have unused channels. The last type is hybrid channel allocation where each cell has a fixed number of allocated channels and there is a pool of dynamic channels that

Corresponding Author: Arafat Abu Mallouh (aabumall@bridgeport.edu)

could be allocated dynamically for any cells that needs more channels. For the previous types of channel allocation approaches there are four important factors that affect their application, coordination between cells in cellular networks, frequency reuse, cell size, and handover. [4] [5] [6] [7]

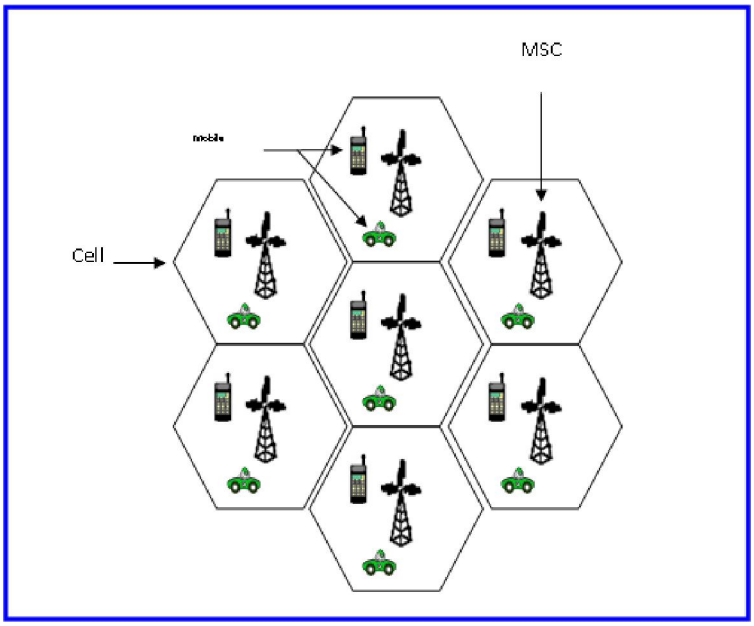


Fig. 1. Cells over a Geographical Area

2 RESEARCH METHODOLOGY

The research methodology is one of the most main sections in any research. In research method we identify the explanation and the conclusion of the problem, in this paper we develop a model that is derived from articles which used to allocate channels dynamically in cellular networks. As shown on the following model there are four main factors that used to increase the efficiency of channel allocation in cellular networks in a dynamic way (Fig 2). These factors were listed after a deep review in the literature; many researchers have talked about this topic due its importance in now a day's technology. [8, 9] [10] Most researches emphasized on these four factors as will be shown in the next discussion. The following figure shows the main four factors along with the goal of this research methodology.

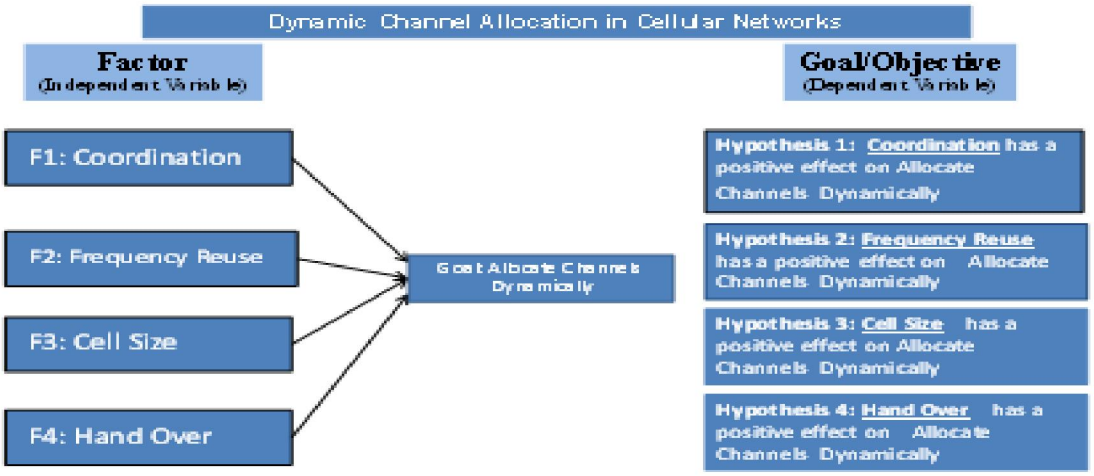


Fig. 2. Proposed Model

2.1 FACTOR 1

Coordination “There are several possible degrees of cooperation, offering a trade-off between performance gains and the amount of overhead placed on the backhaul and over-the-air feedback channels. Multi-cell MIMO cooperative networks [11] p. 3". The author here mentions the importance of cooperation between base stations and what complications we may suffer for achieving the intended degree of cooperation.

2.2 FACTOR 2

Hand over “If the user moves to another cell, then the problem of hand-off must be considered [12] p. 10”. The author here specifies when the hand over process should occur and exactly between two base stations.

2.3 FACTOR 3

Frequency Reuse “the FRS are used for improving the QoS of the cell edge MS and the BS usually focus on the central MS. [13] p. 1”. Here, the frequency reuse is used for improving the quality of service between multiple base stations.

2.4 FACTOR 4

Cell Size “Cell size is also important in facilitating intelligent CAH strategies involving CDMA beacons. [14] p. 4”. The authors focused on relation of cell size and intelligent techniques in cellular networks.

2.5 GOAL

“The available bandwidth for cellular communication is limited and the number of subscribers is increasing day by day, therefore it became necessary to find an optimal frequency assignment scheme that uses frequencies as efficiently as possible [15] p. 9”. The authors focused on the importance of how to allocate the rare available channels in an optimal way so that we can support the largest number of users as we can, and we should take care of the harsh competition on those scarce channels.

3 EXPLANATION/DISCUSSION OF MODEL

As mentioned before there are many factors that affect channel allocation in cellular networks, here we will focus on four of the most importance. Understanding these factors and how they relate to channel allocation; will help to analyze the problem and try to weight each part in any possible solution. In addition each part of these factors has a major role in the proposed dynamic solution and will determine to what extent we can deploy the dynamicity. [16] [17]

3.1 F1: COORDINATION

Base station coordination means that each base station should contact the neighboring cells base stations to know about used frequencies so that each base station is aware of interference that may be caused by close frequencies in the spectrum. Also each base station which has a network load that could not be served by the currently allocated channels may borrow other frequencies from neighboring cells that have unused allocated channels. In [18] emphasized on the importance of coordination between base stations in cellular networks to decrease the interference and as a result to increase the capacity of the network. They found that interference could be reduced by coordinated base stations when it compared with non-coordinated system. [19] [7] [20]

In [21] investigated the importance of coordination between cellular networks base station and found that it has a positive effect in the boundary users and also decreases interference, also they applied their technique using coordination between sectors in different base station. In [22] presented a resource allocation technique that utilize coordination between base station, they used a local dynamic strategy for a cell that varies depending on traffic load, and resulted interference.

3.2 F2: HAND OVER

The hand over term in cellular networks is related to situations where users whom belong to one cell moves to another cell during an active call, in this case the call will be blocked for a period of time until the new cell reserve a time slot on a

new channel for this call that differs from the original channel. This call blocking is a measure of how efficient the channel allocation technique is. In dynamic channel allocation we present a new technique to deal with hand over problem, the proposed solution depends on reserving a number of special channels for each cell that used only for users located on the boundaries of each cell such that these channels will be overlapped between the cell and its neighbors, and it's used only for the boundary users who have active calls, when such users moves from one cell to another during an active call we guarantee no interrupt in the call since the same channel is shared between different cells and its usage is exclusive for boundary users, and also when a cell reserve a time slot in one of the shared channels it will inform the neighboring cells so that no conflict occurs.

Types of Hand Over:

There are two common types of hand over the soft and hard hand over.

- Soft hand over: in this technique when the mobile user moves from one cell to another, it can keep the old channel in the first cell until it reserve a new channel in the new cell so that it guarantee no cut in service.
- Hard Hand Over: Here the mobile user should release the old channel in the old cell before a new channel is allocated for it in the new cell.

In [23] they introduced the importance of hand over process in cellular networks and also into femto cells, they proposed a new mechanism to handle the handover between regular base station and femto cells, their solution was fast but the cost was more computation resources will be consumed.

In [24] they investigated the effect of mobile IP on the hand over process in cellular network and they found that the link layer could be deployed in parallel between two cell to achieve more efficient and robust handover process that will increase the throughput of the dynamic channel allocation. In [3] they used the handover process inside and at the edge of the cells and they found that the performance of the multi-hop cellular network is much better than single hop, also the use of the relay stations has a positive effect on the handover process.

3.3 F3: Frequency Reuse

In cellular networks, a geographical area is divided into a number of cells where the whole frequency bands are divided equally between these cells and no cell is allowed to share any frequency between other cells in the same pattern. This technique is used to prevent any interference that may occur by giving each cell in the pattern unique subset of frequencies. To cover other areas the same pattern is repeated and by this way we guarantee no interference will happen since the distance between the same reused channels is far enough to cause any interference. In our proposed solution where the distribution of channels is dynamic and no fixed channels are reserved during the network life time, we propose to make the initial distribution to pattern wise and then during the run of the network we could reuse channels in the same pattern on condition that it's not used in the same time inside the pattern and also there is enough distance between the cell and any other cells outside the pattern that is currently using the same frequency.

By applying frequency reuse we guarantee more users will be able to be served by the network operator which means more users with same number of frequencies and this leads to decrease cost. In [25] they proposed a novel technique for dynamic frequency reuse that will lead to more efficient utilization of the shared spectrum, they said that dynamic reuse of channels will enable a base station to reuse all of the allocated spectrum. In [26] focused on the usage of end-to-end devices as a cellular network to achieve tremendous multimedia load by improving cell capacity by efficiently and intelligently allocation available channels. In [27] a novel way for self-organizing and adaptive frequency reuse was introduced, they found that three factors play a major role for enhancing frequency reuse and one of them is the frequency reuse factor with number of sectors in a cluster.

3.4 F4: Cell Size

When we divide a geographical area to cells; each cell has one base station that is responsible for providing the required channels for service. Usually the base station is located in the middle of the cell and should be located in a high location to support all of the coverage area of the cell (Fig3). The size of the cell has a direct impact on the capacity of the base station such that when the cell size is small the capacity of the base station is high and when the cell size is big the capacity of the base station is low. For the relation between the cell size and dynamic channel allocation, it was found that when the size of the cells is small then it's better since more cells can cooperate to find the optimal channel allocation among the available frequencies. On the other hand smaller cells means more communication and computation overhead in the network which will make the allocation process complicated even the allocation is optimal, so the degree of the smallness of the cells is a

relative value that is correlated with a threshold value, and this threshold value depends on the number of intended customers to support and the area to be covered. As a result a smaller cell is better but it's limited to a threshold value.

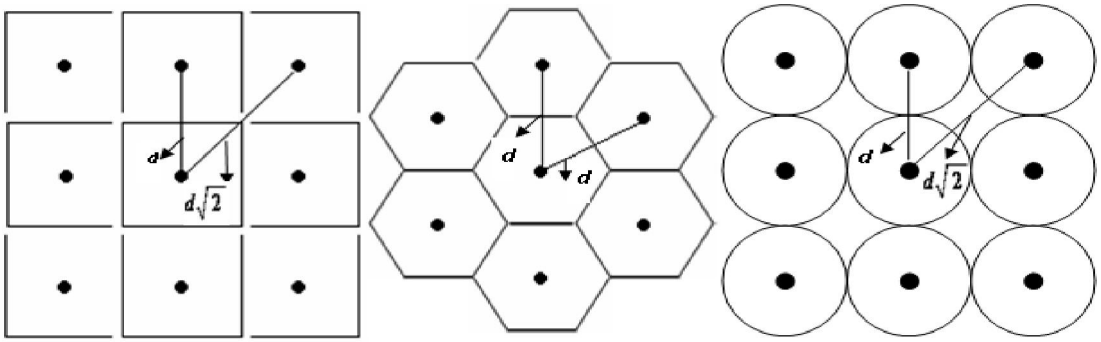


Fig. 3. Cell Shapes

In [28] they found that the cell size should be calculated using a heuristic function depending on the cell capacity and the coverage area to reach a balanced throughput. While [29] argued that the cell size should adapt to the distribution of the users and their requests for channels to reach the optimal cell size. In [30] they discussed the effect of dynamic cell sizing on the overall performance of the cell capacity, they found that by dynamic cell sizing they can increase the cell transmit power and reduce the radii and this is consistent with dynamic channel allocation.

4 WORKING ENVIRONMENT METHODOLOGY

In cellular networks, the traffic load of each cell varies from one cell to another and from one pattern to another; the variance comes from the number of users in each cell Which is differs from one location to another and from time to time. As a result, to make sure that the dynamic channel allocation technique is efficient then it should be tested under different traffic loads. The traffic load could be divided into two types:

- Uniform Traffic Load Distribution: Here the number of users in each cell is equal to the number of users in other cells in the pattern, also it is supposed that in case there is an increase in this number in once cell then the same ratio should increases on their cells in the pattern during the simulation time and vice versa if it is decreases.
- Non Uniform Traffic Load Distribution: In this model we assume that the number of users inside one cell is different from any other cells even with big variance; also this difference may change during the simulation time. In addition, there are cells that suffer from hot spot effect where the number of users during most time of the simulation is steady but in some times it grows rapidly and the cell needs urgent channel to serve the new coming users, this may occur in hotels, stadiums, downtowns. From the previous, any dynamic channel allocation technique should works efficiently in such different simulation scenarios.

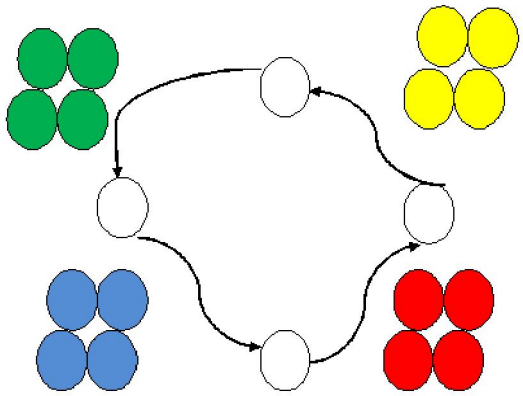


Fig. 4. ResMes Between Cells

Here we classify the system into groups of base stations, the grouping her depend on the distance between the base stations, for each group of base stations that are close enough to be affected by their interference we introduce a special type of message called 'ResMes', this message will circulate between the base stations in each group (Fig 4), when a base station receive the ResMes message it can know what the neighboring cells currently allocate, also the base station which hold the ResMes has the exclusive right to allocate new channels, when it's done the base station will bypass the ResMes to the next base station in the group. With each base station in each group aware of what other base stations have then it can allocate new channels without any conflict.

In this new model, the four important factors that were mentioned before are used; the cell size here has a major role on the number of needed channels and number of other cells in the group. Frequency reuse is the core of this new technique since each base station can allocate any channel when it needs it on condition that it does not violate the interference restrictions, so one channel could be reused several times in different base stations and different groups. For the hand over problem, in our technique it will be solved perfectly since the allocation of channels is done intelligently and each cell has an accurate view of what neighbors has so it can predict the channels needed to be kept free for boundary users. Finally, for coordination, we can obviously see how the ResMes message shows the coordination between the base stations and also between the groups; actually coordination is the heart of our technique.

5 IMPORTANCE OF MODEL

Since the available frequencies for channel allocation in cellular communication networks are scarce, whatever way to improve the utilization of these resources is important to support better service. In this work our model helps to optimally allocate the scarce channels and thus provide better service. The strength of our proposed work comes from the dynamic and intelligent way of allocation channels.

6 CONCLUSION

In this work we found how much important the dynamic allocation process in cellular networks. There are four factors that affect the efficiency of this technique, cell size, coordination, hand over, and frequency reuse. A new dynamic channel allocation method based on artificial intelligent was proposed, the base stations here should be intelligent to communicate with other base stations in order to allocate channel intelligently to reach the optimal balance between the available frequencies and the number of users.

REFERENCES

[1] J. Ming-Hui, *et al.*, "A channel allocation algorithm for large scale cellular networks," in *Parallel and Distributed Systems, 2002. Proceedings. Ninth International Conference on*, 2002, pp. 465-470.

[2] L. Pan, *et al.*, "Capacity scaling of multihop cellular networks," in *INFOCOM, 2011 Proceedings IEEE*, 2011, pp. 2831-2839.

[3] C. Sunghyun, *et al.*, "Handover in multihop cellular networks," *Communications Magazine, IEEE,* vol. 47, pp. 64-73, 2009.

[4] D. D. Dimitrijevic and J. Vucetic, "Design and performance analysis of the algorithms for channel allocation in cellular networks," *Vehicular Technology, IEEE Transactions on,* vol. 42, pp. 526-534, 1993.

[5] L. Du, *et al.*, "Intelligent cellular coverage control according to geographic call traffic distribution," in *3G Mobile Communication Technologies, 2002. Third International Conference on (Conf. Publ. No. 489)*, 2002, pp. 423-427.

[6] F. P. Garcia, *et al.*, "Indexing Mobile Objects: A Strategy for Supporting Dynamic Channel Allocation in Cellular Networks," in *Wireless and Mobile Computing, Networking and Communications, 2006. (WiMob'2006). IEEE International Conference on*, 2006, pp. 284-291.

[7] S. Mandal and D. Saha, "An efficient technique for dynamic channel allocation (DCA) in mobile cellular networks," in *Personal Wireless Communications, 2005. ICPWC 2005. 2005 IEEE International Conference on*, 2005, pp. 470-473.

[8] I. Spyropoulos and J. R. Zeidler, "Supporting Asymmetric Traffic in a TDD/CDMA Cellular Network via Interference-Aware Dynamic Channel Allocation and Space–Time LMMSE Joint Detection," *Vehicular Technology, IEEE Transactions on,* vol. 58, pp. 744-759, 2009.

[9] E. Del Re, *et al.*, "Performance analysis of a dynamic channel allocation technique for terrestrial and satellite mobile cellular networks," in *Global Telecommunications Conference, 1993, including a Communications Theory Mini-Conference. Technical Program Conference Record, IEEE in Houston. GLOBECOM '93., IEEE*, 1993, pp. 1698-1702 vol.3.

[10] A. Pattavina, *et al.*, "Reuse partitioning in cellular networks with dynamic channel allocation," in *Global Telecommunications Conference, 1995. GLOBECOM '95., IEEE*, 1995, pp. 1543-1548

[11] D. Gesbert*, et al.*, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference," *Selected Areas in Communications, IEEE Journal on,* vol. 28, pp. 1380-1408, 2010.
[12] A. Hac, "Cellular network model with hand off delays," in *Communications, 1995. ICC '95 Seattle, 'Gateway to Globalization', 1995 IEEE International Conference on*, 1995, pp. 1834-1838
[13] L. Min*, et al.*, "A Novel Frequency Reuse Scheme for OFDMA Based Relay Enhanced Cellular Networks," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, 2009, pp. 1-5.
[14] R. Ganesh and H. Stellakis, "Impact of cell size on dual-mode CDMA cellular networks," in *Personal, Indoor and Mobile Radio Communications, 1998. The Ninth IEEE International Symposium on*, 1998, pp. 298-302 vol.1.
[15] S. Pinagapany and A. V. Kulkarni, "Solving channel allocation problem in cellular radio networks using genetic algorithm," in *Communication Systems Software and Middleware and Workshops, 2008. COMSWARE 2008. 3rd International Conference on*, 2008, pp. 239-244.
[16] G. Xiaohu*, et al.*, "Capacity Analysis of a Multi-Cell Multi-Antenna Cooperative Cellular Network with Co-Channel Interference," *Wireless Communications, IEEE Transactions on,* vol. 10, pp. 3298-3309, 2011.
[17] X. Yanfang*, et al.*, "Interference-aware channel allocation for Device-to-Device communication underlaying cellular networks," in *Communications in China (ICCC), 2012 1st IEEE International Conference on*, 2012, pp. 422-427.
[18] F. Boccardi and H. Huang, "Limited Downlink Network Coordination in Cellular Networks," in *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, 2007, pp. 1-5.
[19] J. Vucetic*, et al.*, "Implementation and performance analysis of multi-algorithm dynamic channel allocation in a wideband cellular network," in *Communications, 1996. ICC '96, Conference Record, Converging Technologies for Tomorrow's Applications. 1996 IEEE International Conference on*, 1996, pp. 1270-1274 vol.3.
[20] F. Delli Priscoli*, et al.*, "Application of dynamic channel allocation strategies to the GSM cellular network," *Selected Areas in Communications, IEEE Journal on,* vol. 15, pp. 1558-1567, 1997.
[21] G. Wunder*, et al.*, "Self-organizing distributed inter-cell beam coordination in cellular networks with best effort traffic," in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2010 Proceedings of the 8th International Symposium on*, 2010, pp. 295-302.
[22] B. Golkar and E. Sousa, "Adaptive Localized Resource Allocation with Access Point Coordination in Cellular Networks," in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1-5.
[23] A. Rath and S. Panwar, "Fast handover in cellular networks with femtocells," in *Communications (ICC), 2012 IEEE International Conference on*, 2012, pp. 2752-2757.
[24] A. Taghizadeh and W. Tat-Chee, "A network-based IP mobility scheme for fast and global handover in cellular networks," in *Communications (APCC), 2011 17th Asia-Pacific Conference on*, 2011, pp. 307-312.
[25] L. Jian*, et al.*, "A Novel Dynamic Full Frequency Reuse Scheme in OFDMA Cellular Relay Networks," in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, 2011, pp. 1-5.
[26] C. Hyang Sin*, et al.*, "Radio resource allocation scheme for device-to-device communication in cellular networks using fractional frequency reuse," in *Communications (APCC), 2011 17th Asia-Pacific Conference on*, 2011, pp. 58-62.
[27] A. Imran*, et al.*, "A novel Self Organizing framework for adaptive Frequency Reuse and Deployment in future cellular networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, 2010, pp. 2354-2359.
[28] Y. H. Tam*, et al.*, "Effective Cell Size Scheme in Multi-Hop Cellular Networks," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, 2009, pp. 1-6.
[29] Y. H. Tam*, et al.*, "Optimal Cell Size in Multi-Hop Cellular Networks," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 2008, pp. 1-5.
[30] A. G. Spilling and A. R. Nix, "Performance enhancement in cellular networks with dynamic cell sizing," in *Personal, Indoor and Mobile Radio Communications, 2000. PIMRC 2000. The 11th IEEE International Symposium on*, 2000, pp. 1589-1593

ARAFAT ABU MALLOUH

Arafat is a PhD candidate in computer science and engineering at University of Bridgeport. Arafat research interest includes artificial intelligence, data mining, and network security. Arafat currently works on his PhD dissertation.

CHRISTIAN BACH

Christian Bach holds academic honors as: PHD in information science and was executive MBA at the university at Albany/SUNY. Now he is professor in technology management department in university of Bridgeport. His multidiscipline research approach is aimed to integrate in particular marketing research and information effectiveness.

ABDULLAH A. ABDULLAH

Abdullah is M.Sc. Student at University of Bridgeport Computer Science Department. Interesting research topics for Abdullah are mobile network, parallel processing, and web development.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 616-625

http://www.issr-journals.org/ijias/

An Empirical Study of Advertised Reference Price on Consumers' Behaviors

Yasser Rayes[1] and Christian Bach[2]

[1]Technology Management, University of Bridgeport,
Bridgeport, Connecticut, USA

[2]Technology Management and Biomedical Engineering,
University of Bridgeport,
Bridgeport, Connecticut, USA

ABSTRACT: This study is directed to the business community regarding the planning of advertisements and more specifically implementing their reference prices. The aim is to investigate the effects of the presentation of reference prices of the consumers' behaviors. It is an empirical study in which the researcher has used literature reviews and survey to investigate the issue. A survey is delivered to 116 randomly selected respondents, where 86 people have replied. The researcher has investigated different areas related to reference prices such as the importance of advertised reference prices for firms, the effects of advertised reference prices to consumers purchase behaviors; the effects on the consumers purchase decisions; the products characteristics that influence consumers purchase behaviors, and finally the impact of the firms' variables to the consumers' behaviors. Reference prices may be defined as the cost that customers consider reasonable to pay or expect paying for a typical product. The market department of the organizational consistently engages in assessing the product's reference prices in order to set the level of prices affordable to customers or target markets. Today's customers are more knowledgeable than the customers in the past. This research will help in providing positive and easily acceptable tool to launch a new product in the market. Furthermore, it might be of great benefit to the business communities in the industries.

KEYWORDS: Reference Price, Advertising Effectiveness, Consumers' Purchase behaviors, reference-pricing system, Pricing Strategy.

1 INTRODUCTION

Reference pricing can be defined as the mechanism that a third party payer will use to determine the maximum price that will reimburse the suppliers price or reimbursement price, [1]. The presence of a company being able to present the marketing perspective to its clients all depends on the standards and types of claims that it presents with the help of reference pricing. The advertising reference pricing is affected by the level of a consumer's knowledge [2] in to the perspective that it will be able to conduct as a company. The reference advertising pricing is supposed to be fair and not deceptive [3]. The claims that are offered for the pricing reference advertising are seen to be controlled by government bodies that are seen to look at the reference of the documentation [4]. The Bureau of consumer protection has a responsibility to monitor the claims and advertising that the disclaimer will be seen to provide. This is permitted through the federal trade commission oversight [5].

The consumers buying behaviors are seen to be the decision process that people make for the contribution in by and using of products. The marketers are required to understand the buying behaviors of the ultimate consumer [6]. The consumers are seen to undergo the following stages for the purpose of acquiring a commodity or service. These stages are the problem recognition, evaluation and information search of alternatives, purchase decisions, purchase action and post purchase outcome [1].The reference pricing advertising on consumers' behaviors is affected by the marketing mix The marketers are required to analyze how, when, where and what questions with regard to the products prices provision

Corresponding Author: Yasser Rayes (yasser.rayes@hotmail.com)

the market. The buyers' reaction to the pricing strategy will be analyzed with respect to the impact on the buyers. reference prices portray a sense of fairness in the prices that are offered for a product [2].

The internal aspect of the reference prices is seen to be the overall interpretations of the price expectations company which the consumers experience, while the external aspect of the reference prices is seen to be the external consumers' price expectations [7]. Reference prices may be defined as the cost that customers consider as reasonable to pay or expect to pay for a typical product. The market department of the organizational consistently engages in assessing the product's reference prices in order to set the level of prices that are affordable to customers or target markets. Today's customers are more knowledgeable than the customers in the past. The effects of reference prices on empirical generalization in marketing [8], and such price promotions [4] and product quality [9] create the master project about the advertised references prices on consumers' behaviors.

Therefore, it is necessary for marketers and organizations to assess the needs of their consumers, and produce the products and services according to these needs and wants [10]. They should also provide the products, which satisfy them. The topic of consumers buying has been remained a debatable for many decades in all over the world. In previous studies, researchers have investigated this topic and found out only exploratory findings. It is important for marketers to relate between the buying behavior processes of consumers and their social and psychological concerns [11].

1.1 SCOPE OF THE STUDY

This study is an empirical observatory research which has used literature reviews and survey analysis to investigate the issue. A survey is delivered to 116 randomly selected respondents. The focus of this research will be on the reference pricing influences on the behaviors that consumers are seen to portray [1]. The researcher will concentrate on the importance's of advertised reference prices of firms, the effects of advertised reference prices to consumer purchase behaviors; the effects on the consumers purchase decisions; the products characteristics that influence consumers purchase behaviors and the impact of the firms' variables to the consumers' behaviors.

The researcher has found out that the use of reference prices in the advertisements can show a positive impact on the consumers' purchase behaviors. The buyers' reactions to the pricing strategy will be analyzed with respect to the impact on the buyers. This research will help the management in the analysis and prediction of the marketing strategies that can be used for the purpose of reference pricing advertising.

1.2 HYPOTHESIS OF THE STUDY

Advertised reference prices increase the firms' sales level by motivating the customers purchase behaviors towards the firms' products and services.

1.3 RATIONALE OF THE STUDY

This research will be presented with respect to the marketing terms. It will base its analysis and the data collection through the use of the personal experiences that affect the company's activities [12]. This study is imperative due to the need of understanding the effects that human consumers' behaviors influence the running of the company. The basis of this study is related to the reason of increasing the sales of companies with regard to the marketing strategy that they will present. Therefore, the results and findings of this study will be a credible due to the analysis of 119 marketing professionals.

1.3.1 PROCESS REVIEW OF LITERATURE

Marketing a company is seen as a challenging task that involves the scrutiny of the issues that affect the stakeholders. Stakeholders are seen to be the customers, competitors, [12] market and industry [3]. The advertisement of the reference prices is the presentation of the cheapest prices to the clients for the purpose of attracts the customers to contract the company. The consumers' behaviors are seen to look at the overall procedures in which a company will be able to seek the highest paying job opportunities. The consumers' behaviors can be seen to be the overall analysis of the way customers will behave with respect to a change in the markets strategy.

Advertised reference prices and consumers' behaviors are seen to affect the stakeholders [13]. The approach provides sense of transparency to the consumers. The presence of sticking with the facts of the prices helps customers in the acknowledgement of opinions, which might make people feel uncomfortable. The approach is seen to allow the existence

consumers' accommodation to the making of failures by the company. This is seen to allow shared responsibilities by the company and the stakeholders. That's why this approach helps in the alienation of secrets.

This makes the company to be able to search for the best quality goods for the purpose of service provision. The customers' spending is seen to increase with the increase of the disclosure that the company provides. This enables the building up more trust between the company and the customers. Moreover, it provides a competitive edge in the acquisition of results by the company. The presence of this approach gives a sense of control to customers, which is seen to allow the overall control of the company's processes. The customers will be proficient to choose for the company whether to keep a product brand in the market or remove it. This power of reject can be evaluated by the profit contribution that a company will be seen to likely produce.

According to previous studies, the advertised reference a price was seen to increase the bid value of the perceived believable results with strong effect on the sales for the firm. This indicates that advertised reference prices were seen to be the best procedure in harness-exaggerated expectations of the consumers [14]. In the present study, the advertised reference prices on consumers purchase behaviors are seen to lead to an increased pricing scenario. In the post price scenario consumers are seen to have replication in the amount of purchases that they perform. Advertised reference prices have been seen to reduce the search for information by the consumers. This increases decision-making effectiveness and provides the overall increase in the firms operation. It is found also that the growth of the Internet programs has led to the increase of the firms operation, (refer to figure 3 Advertised Reference Prices framework).

1.3.2 REVIEW AND PROPOSITIONS

Looking ahead, there is a need to conduct future studies that would help expand the sample of respondents who are exposed with the purchase of professional services. Given the case, there is a need to conduct further study on the relationship between demographic variable and accuracy of reference prices; factors like age and income have substantial impact in the study with actual consumers to help determine the difference in accuracy of recall [15].

The value of understanding reference prices in consumers' price perceptions and evaluation has been the subject of numerous studies. As consumers possess reference prices because of their previous experiences, their effects redound to varying effects on advertised prices or contextual effects. Considering that consumers behave using reference prices, this has effects on how consumers make their decisions. Moreover, the predictive power of choice models will help to generate a common consensus by the consumer's knowledge. This will allow the reference prices to have a significant impact on the product evaluation and purchase decision. Most likely, consumers prefer prices of items that are low hence they are willing to accept such items at those cost levels [16].

As consumers' behaviors are varied and are changing because of many factors. That's why there is a need to continuously study their buying behavior patterns and the factors that influence their decision making processes to purchase items using reference pricing or something else. The key is that by understanding customers or consumers, there is a better edge for marketing strategies to be developed, which will influence consumers to a higher degree in patronizing products, or services that are offered to consumers. For reference pricing to really prosper as indication for successful price management and right buying behavior patterns, reference prices are formed, retrieved and used as reference prices.

Given this case, research studies will still be pursued, and the external and internal environments have to be taken into consideration in determining buying behavior patterns of consumers. Furthermore, there is also a need to understand the many reference price categories that consumers use in evaluating product value and its effects because these will have impact on how one prepares to understand reference pricing in the future and for effective future use.

2 IMPORTANCE OF THE REFERENCE PRICES CONCEPT

The reference prices concept is seen to be important due to the following facts. The advertising reference prices raises the customers' familiarity with respect to the acquisition of the audience of the products service provision [13]. Reference prices are standards where the purchase prices of a product is judged [17]. The reference prices advertising helps in the formation of the good image for the company with the customers. The firms minimize the information leakages to the legal authorities that can lead to the revocation of licenses and information leakages [18]. The reference prices advertising is seen to help in the mitigation of the markets impact cost. The competitors will be unable to attract customers due to the provision of the reference prices. The reference prices is seen to allow the control of order pricing or revenues [19].

The management is seen to be able to control the interactions of the programs that will be able to seek the best for the company. The reference prices are seen to reduce the occurrence of the unequal levels of transparency in pricing.

customers will be able to search for the provision of the highest achievers of activities with respect to quality service provision for the clients [20]. The customers are seen to form their loyalty with respect to the provision of the amount that the company will be able to search for the provision of the quality skills. The companies will be able to earn a good return the provision of the advertising reference prices. The advertising is seen to provide a new outlook to the company about market due to the presence of the consumers' behaviors, which is always changing. The advertised reference pricing is to allow the use of memorable words for products by the customers. This will enable high overall growth of the company**.**

2.1 RESEARCH METHODOLOGY

The research is an empirical study, which focuses on the reporting of results of the advertised reference prices on consumers' behaviors. The research sample is seen to be involving marketing specialist of multinational companies and consumers of the firms, which use advertised reference prices. The expected informants of this study are 119, but actual number of the informants that presented the questionnaires are 89 specialist and 15 customers. The questionnaires are sent through email addresses where the participants are required to answer the questions. The data collected is seen to be basically observatory remarks and analysis that the marketing specialists have observed for consumers' behaviors. There is also another tool for the data collection, which is through the use of interviews and observatory survey of companies. The informants are asked questions with respect to the qualitative concept of marketing. The research questions are seen top look at the numerous success factors that are seen to control the functioning of the company's operations. The concept of advertised reference prices is seen to be the latest information, which is used for the provision of the quality skills for service provision.

2.2 ANALYSIS AND RESULTS

The questionnaire sample is looking at the overall pertinent issues that have been seen to affect the marketing aspect of the firms. The questionnaire will focus on advertised reference prices and their effect on the business. The questions are expected to be answered within semi structured questionnaire pattern.

Consumers' Purchase Behaviors and Advertised Reference Prices:

Marketing Strategy and Consumers' Choices:

- What are the most competent promotion strategies for advertised reference prices? Why are they efficient?
- Does the promotion strategy applicable in the implementation of advertised reference prices approach?
- How does the advertising project affect?
- How does the advertised reference prices affect the brand choices and purchase quantities?

Stimulus and Competition:

- How does the market stimulus affect the consumers' purchase behaviors?
- What is the impact of the price promotion stimuli towards the products quality?
- How do the products characteristics affect the consumers' purchase behaviors?
- How does competition affect the products demand from the consumers?
- How can advertised reference prices be used to control the competition industry?

The advertised reference prices are seen to provide numerous alternatives for the firms purchase decision and product availability.

- How do these alternatives affect the organizations?

Part 2: Linker Scale

The analysis will incorporate the analysis of the observations that are done by the informants. The results be analyzed in respect to the advertised reference prices. 1 (For Strongly Disagree), 2 (Disagree), 3 (Neither Agree nor Disagree), 4 (Agree) and 5(Strongly Disagree) are the ranking scale for the analysis.

- Consumers are seen to engage in planned purchases and not unplanned
- Store premium moderator and affect the promotion of the firm's activities.
- The consumers are conversant with the choices of looking at the overall pricing schemes.
- Consumers are bound to do informational search for the purpose of purchasing commodity.
- The management of advertised reference prices is dependent on the organization's staff unity.

The principle of this study is to investigate the overall efficiency and effectiveness of the advertised reference prices implementation with respect to consumers' behaviors. The importance of advertised reference prices is highlighted and seen to increase, confidence and comfort for the customers. The lack of advertised reference prices for products is raised as a concern where customers are seen to regret on the prices they are used to purchase certain products. The advertised reference prices reduce the customers' flexibility in the search for products and hence maintain products purchase decisions, [21]. Customers are seen to address the need for choice on acceptance and rejection of prices in the cases of bidding. It is found out that the lack of reference prices information affects the production of goods. The analysis indicates that advertised reference prices lead to poor buying behaviors for the services and products. This enables the firms to benefit with respect to the prices that they present to the market. Advertised reference prices control competing firms through the use of pricing strategies.

The advertising strategy is presented as high exposure to the implementation of the advertising strategy. The strategy of being a challenger firm is seen as the use of the integral part of the advertised reference prices. The specialists point out the fact that the diversification of the strategies has to be implemented in order to increase the growth of the firm. The advertising exposure is a marketing variable that is virtually purchased for products [22].The advertising exposure of the patients indicates the provision of the highest quality services. The newspapers, magazine, television and radio are considered as important tools of the advertised reference prices communication. The sales situation is affecting the manner in which the management will engage the advertised reference prices. This interprets the overall influence of the concurrent validity of the firms. The advertised reference prices influence the amount of cash the organization which is seen to be likely to run during the implementation of the product [23]. Moreover, the slogan of an offered discount for the new customers is seen as a major concern for the provision of the quality goods and services.

The product price is seen to be explicit to customers with the way in which the customers compare the product prices. The efficiency and effectiveness of advertised reference price are implemented by the overall analysis of the prices that the customers conduct, [24]. This has made the customers to look at the overall procedure of the acquiring goods through the use of the advertised reference prices. The quality of products is an important need that is demanded by the customers to use a product.

The status or preference of the customers is seen to be subject to use advertised reference prices. The high priced products quality is the main influence, which makes the perception of a good necessary to be of high quality. The under pricing of a product is seen by customers to present a scenario where the customers will recognize the firms output to be of low quality [25]. The brand image is seen to be affected by the preference of the customers. This affects the products and service delivery of recognizable brands.

The products characteristic needs to be identified in order to acknowledge the reasoning for the running of the firms. The demographics of the customers affect the efficiency of advertised reference prices in the sale of a product [3]. The advertised reference prices approach allows the overall increase in the investment of the product services for clients. The frequency, investment and expansion strategies affect the running of the firms with respect to the quality service provision for the client’s management. The wrong target audience is seen to influence the efficiency of using advertised reference prices. The competition industry looks at financial performance that requires the transcending individual product lines.

The marketers use new features, product, change of distribution, and packaging to new promotional offers. Online environment has presented the communication for the production of goods with respect of listings, which are done through the bidding. The terms of sales of the products affect the running of the firms' performance. The billing, pricing and sellers delivery of the integral part of the sales transaction and efficiency of advertised reference prices are the product differentiation of the low cost leadership. The customers are concerned by the post purchase of the sales transaction. The dynamism of the production of the goods is influenced by the buying behaviors of the consumers.

3 FUTURE WORKS

The present study has a limitation on the carrying out of experiments. The analysis is focused on the observatory characteristics of consumers and management of the firms' activities for the basis of elicitations of the emotions of individuals. The use of the analysis of the marketing specialists for many firms is seen to allow the overall implementations of the advertised reference prices. The presence of using a study sample of 89 informants who are market specialists has led to the fact that the research is valid. The overall generalization of facts is done for the purpose of eliciting the consumers' reaction with regards to their purchasing decisions.

Therefore, the researcher recommends a new study that will analyze the emotions and elicitations of the consumers' purchase decisions with respect to the acquisition of the goods and services. Current shopping status of the consumers

should be eliminated with the use of the activities that is presented for the firms. Researchers should analyze the internal references and external reference prices.

The advertised reference prices should be studied under the respect of the manipulation of the reference prices, which is limited in the present study. Researchers are advised to investigate other means of manipulating the market in order to elicit and analyze the relations of people in the market. The product involvement and familiarity can be analyzed through the use of market brands by firms in order to establish the consumers' purchase behaviors in the running of the firms. The multiple reference prices might affect consumers' search behaviors. Another field of study that might be investigated is the impact of multiple advertised reference prices in the market. Researchers can look at analyzing more than one product with respect to the provision of the quality standards. Moreover, the negative feelings of the poor post purchase regret or dissatisfaction can be investigated. The consumers' behaviors are seen to be affected by the use presence of the implementation strategies that are different as compared to others. The polar opposites of good advertised reference prices approach in marketing and bad advertised reference prices approach can be also investigated with respect to customers' experiments. The analysis might look at the positive outcomes of the firms' management.

4 TABLES AND FIGURES

4.1 TABLES

Table 1. Marketing variables that are affected by advertised reference prices

Marketing Strategy	"Therefore, a major source of competitive advantage in the global market is the ability to produce high quality, low-price products (Levitt, 1983). To attain a low-cost position, the optimum global marketing strategy is to sell standardized products using standardized marketing programs [3].
Nature of Product	"Developing profit functions that depend on the partition of companies into joint ventures and the nature of product competition between venture partners" [20], [26].
Competition	"Product competition patterns have risen and, in some cases, become central in markets for computers, financial services, consumer electronics, and a growing number of products"[27].
Consumer Purchase Experience	"A consumer who is familiar with the purchase of an embarrassing product will experience little difference in the amount of embarrassment that is felt when there is a real or imagined social presence"[28].

4.2 FIGURES

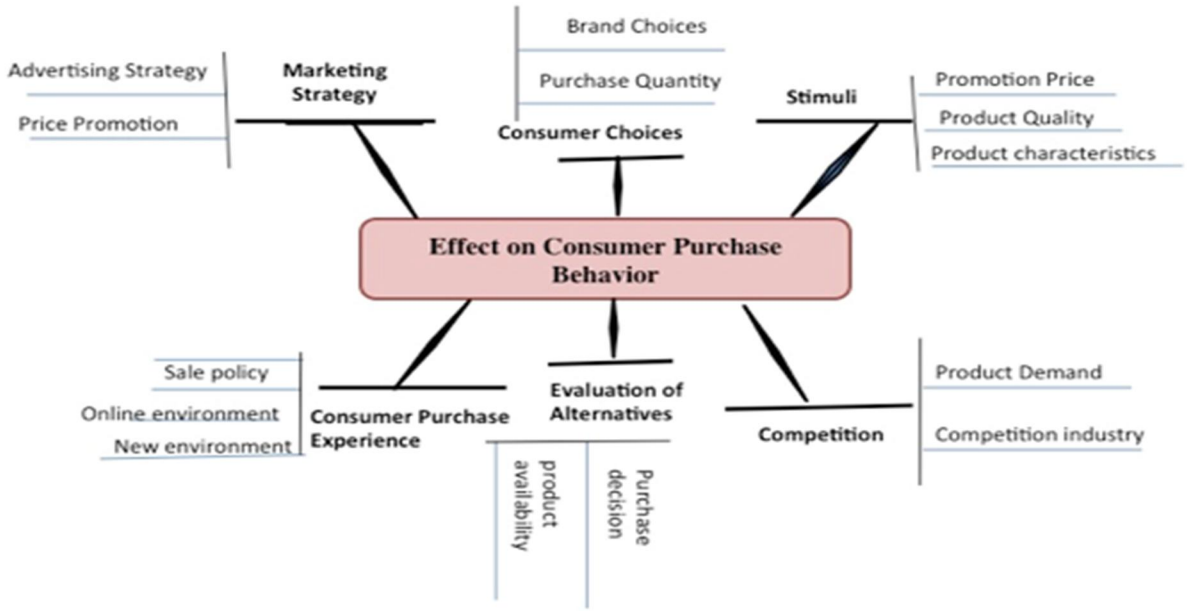


Fig: Factors of: An empirical study of Reference price on consumer behavior

Fig. 1. Factors Affecting of The advertised reference prices on consumers' Purchase behaviors Model

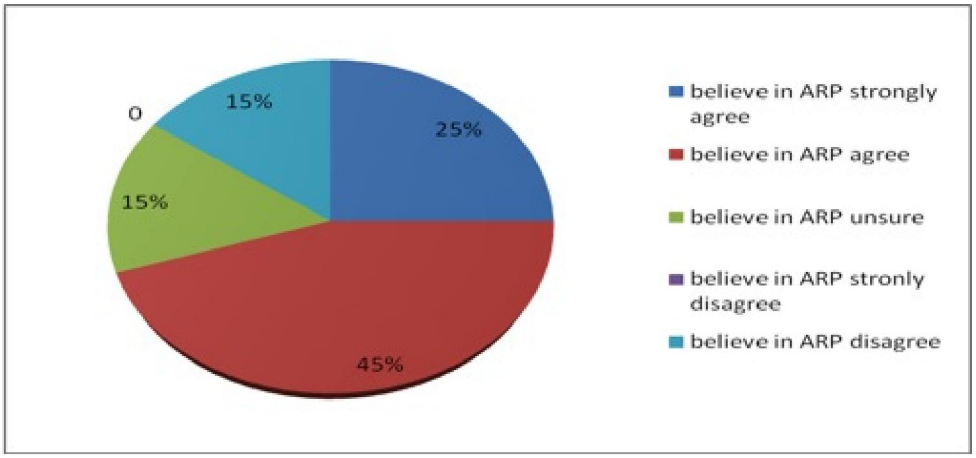


Fig. 2. Customer Belief In Advertised Reference Prices (This figure is according to the researcher's results)

The customer's belief has affected the consumers' behaviors with respect to the quality service requirements. The data collected indicate that many customers are optimistic about the advertised reference prices and they believe on them. In the meanwhile the disagreement of the population of the individuals is low.

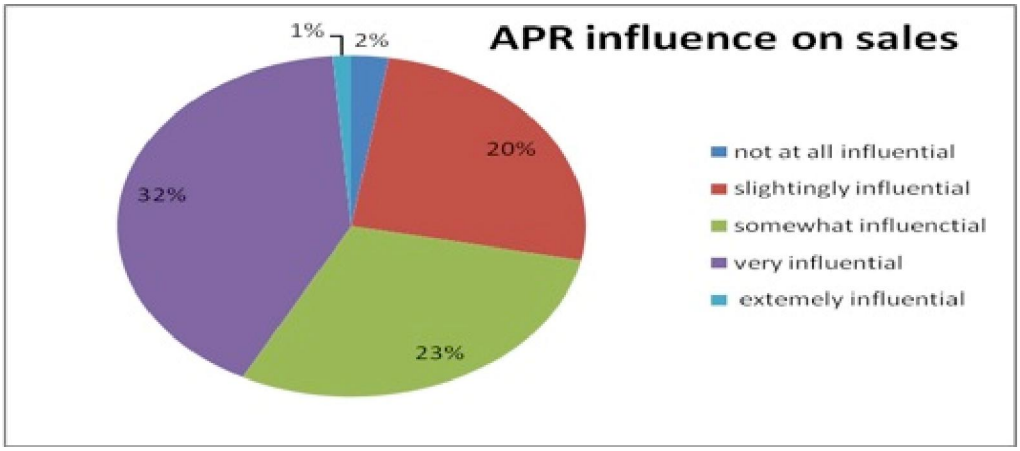


Fig. 3. Market Share Influences on Advertised Reference Prices (This figure is according to the researcher's results)

The market share is seen to be influenced with respect to the acquisition of the highest quality standards. Increased advertised reference prices are the strategy that would remove the competitors from investing in certain labor markets. Advertised reference prices affect the implementation of the strategies. The nature of the product remains as the same, while the market share influences in respect to the nature of the product. Competition for the market share affects the running of the firms with respect to the nature of the product. The purchase quantity for the products and services is increased due to the implementation of the advertised reference prices approach. The sales affect the influence of the market share sales to a great extent. Moreover, the market specialists affect the overall growth of the company.

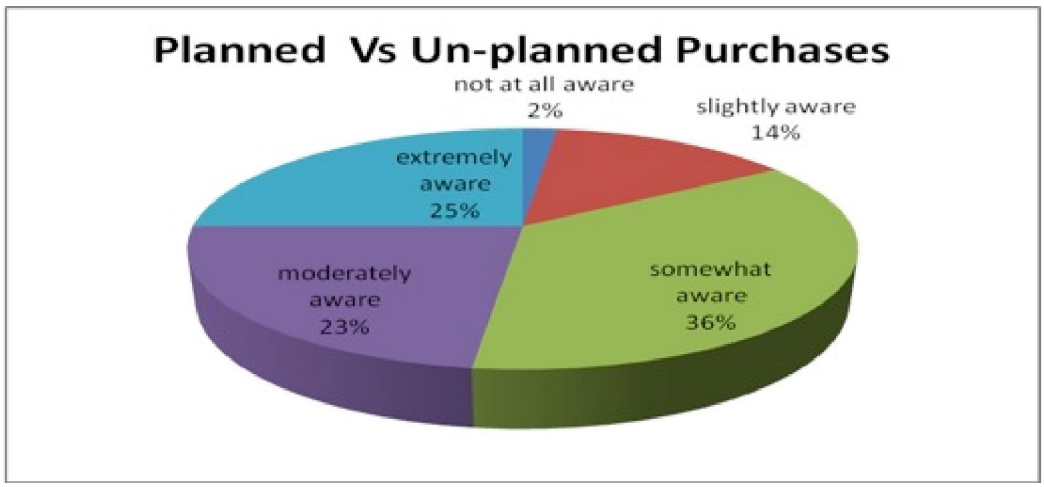


Fig. 4. Unity Dependence (This figure is according to the researcher's results)

The specialists address the fact that the unity provision of the advertised reference prices is required to increase the quality service provision for the clients. The unity dependence indicates at the dependence of quality standards. The point of view of both customers and specialists show that the unity dependence of the marketing advertised reference prices and other departments of the business are poor.

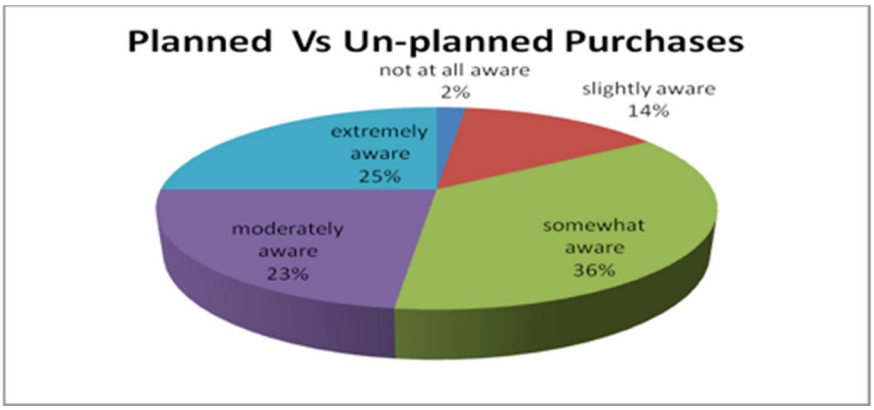


Fig. 5. Planned Vs. Un-planned Purchases (This figure is according to the researcher's results)

The customers have information search of the products before doing a transaction. They address the presence that most of the purchases they conduct are unplanned. Thus, the customers use the ready information that they are provided with. The lack of awareness of the product that the customers buy is seen to be as a major concern for the provision of the best services and goods. The findings indicate that the presence of the firms, which are able to present advertised reference prices communication to their customers through the Internet, increases the companies' productivity. The research findings indicate that consumers do not decrease their search for information for products in the presence of advertised reference prices.

5 CONCLUSION

In conclusion, advertised reference prices increase the firms' sales level by motivating the customers purchase behaviors towards the firms' products and services. The importance and effects of advertised reference prices to the firms raise the customers' familiarity, increase companies selling standards, give a good image for company, information leakages to the legal authorities, mitigation of the markets impact cost, control of order pricing or revenues, manage the interactions of the programs, reduce the occurrence of the unequal levels of transparency, and finally provide the company a new outlook of the market.

The advertised reference prices affect the consumers' purchase behaviors in the following ways. Increased communication makes the eliciting of goods and services efficient and effective. Consumers do not decrease the search for information. The unity of the firms and other stakeholders are enhanced through the use of advertised reference prices. Advertised reference prices increase the likelihood of growth for the company. The marketing variables are discussed as marketing strategies; products characteristics, competition and consumers purchase expectation. These variables affect the firms in many ways.

REFERENCES

[1] J. Roberts, "Environmental Segmentation alternatives," *Journal on Consumer Marketing,* vol. 16, pp. 265- 278, 1999.
[2] T. Mazumdar, S. P. Raj, and I. Sinha, "Reference price research: review and propositions," *Journal of Marketing,* vol. 69, pp. 84-102, 2005.
[3] K. L. Anderson, "Terms of Sale," *Journal of Marketing,* vol. 11, pp. 250-257, 1947.
[4] L. James and R. E. Bucklin, "Reference Effects no Price and Promotion on Brand Choice " *journal of Marketing Research,* vol. 26, pp. 299-310, 1989.
[5] Daniel, Howard, and R. A. Kerin, "Broadening the Scope of Reference Price Advertising Research: A Field Study of Consumer Shopping Involvement," *Journal of Marketing,* vol. 70, pp. 185-204, 2006.
[6] I. Popescu and Wu, "Dynamic Pricing Strategies with Reference Effects," *Operations Research,* vol. 55, pp. 413-429, 2007.
[7] P. K. Kopalle, A. G. Rao, and J. L. AssunSao, "Asymmetric Reference Price Effects and Dynamic Pricing Policies," *Journal of Marketing,* vol. 15, pp. 60-85, 1996.

[8] Kalyanaram and R. S. Winer, "Empirical Generalizations from Reference Price Research," *journal of Marketing Science,* vol. 14, pp. 161-169, 1995.
[9] Hardie, Eric J Johnson , and P. Fader, "Modeling Loss Aversion and Reference Dependent Effects on Brand Choice " *journal of Marketing Science,* vol. 12, pp. 378-94, 1993.
[10] D. J. Howard and R. A. Kerin, "Broadening the scope of reference price advertising research: a field study of consumer shopping involvement," *Journal of Marketing,* vol. 70, pp. 185-204, 2006.
[11] B. Rosenbloom and B. Dimitrova, "The marketing mix paradigm and the Dixonian systems perspective of marketing," *Journal of Historical Research in Marketing,* vol. 55, pp. 53-66, 2011.
[12] Martin, "Consumer behavior in Tourism," *Journal of Marketing,* vol. 21, pp. 155-168, 2002.
[13] G. George and K. Zoe, "Spring air Gabbler Publication," *Journal on Enterprising Communities,* vol. 3, pp. 152-175, 2009.
[14] J. Giessen, " Collectivism and individualism," *Journal on Business Venturing,* vol. 12, pp. 367- 384, 2013.
[15] M. Brucks, V. Zeithaml, and G. Naylor, "Price and Brand Name as Indicators of Quality Dimensions for Consumer Durables," *Journal of the Academy of Marketing Science,* vol. 28, pp. 359-374, 2000.
[16] V. A. Zeithaml and K. L. Graham, "The Accuracy of Reported Reference Prices For Professional Services," *Journal of Marketing,* vol. 10, pp. 607-611, 1983.
[17] K. B. Monroe, "Buyers' subjective perceptions of price," *journal of Marketing Research,* vol. 1, pp. 70-80, 1973.
[18] R. A. Briesch, L. Krishnamurthi, T. Mazumdar, and S. P. Raj, "A Comparative Analysis of Reference Price Models," Journal of Consumer Research," *Journal of Consumer Research,* vol. 24, pp. 202-214, 1997.
[19] C. Grönroos, "Keynote paper From marketing mix to relationship marketing - towards a paradigm shift in marketing," *Journal of Marketing,* vol. 35, pp. 322-339, 1997.
[20] C. Hemingway, "Behavior: Spring ear's Publishers," *Journal on Personal Values as A catalyst for CSR,* vol. 60, pp. 233-249, 2005.
[21] R. Miriam and M. Ames, *Journal on Corporate Governance,* vol. 16, pp. 443-459, 2008.
[22] H. Morris and W. Allen, "Consumer Behavior Palgrave Macmillan," *Journal on international Business studies,* vol. 25, pp. 65-89, 2012.
[23] S. Puffer, "Academy of management," *Journal on Business Strategies and Employee development,* vol. 15, pp. 16-121, 2005.
[24] E. Richard and S. Wanda, "Examination of Diversity of Corporate Stakeholder roles," *Journal on the Business Society,* vol. 40, pp. 266- 294, 2008.
[25] B. Shawn and R. Peters, "Consumer behavior," *Journal on International Enterprise development and Management,* vol. 6, pp. 38- 54, 2009.
[26] P. Greenlee and B. Cassiman, " Managerial and Decision Economics,," *Journal of Marketing,* vol. 20, pp. 115-130, 1999
[27] R. Sanchez, "Strategic Flexibility in Product Competition," *Journal of Strategic Management* vol. 16, pp. 135-159, 1995.
[28] D. W, R. Dahl, V. Manchanda, and J. J. Argo, "Embarrassment in Consumer Purchase: The Roles of Social Presence and Purchase Familiarity," *Journal of Consumer Research,* vol. 28, pp. 473-481, 2001.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 626-650

http://www.issr-journals.org/ijias/

Vers un dispositif d'appréciation de la pertinence des IFRS dans un contexte pré-implémentation

[Towards a platform for evaluating the relevance of IFRS in a prior-implementation context]

Raoudha Trabelsi

Montpellier Recherche en Management,
Université de Montpellier 1,
Montpellier, France

ABSTRACT: This paper proposes a platform for evaluating the relevance of IFRS in a prior-implementation context that may be of interest to any country concerned by the international accounting harmonization. Thus, in one hand, a tool is provided to measure the level of voluntary harmonization of local accounting practices with IFRS (voluntary de facto harmonization). In the other hand, the theoretical foundations that might explain such a measure are presented. In fact, the contingency theory and the neo-institutional approach helped us establish the explanatory variables, hypothesis and research models, related to two distinct local populations, involving the commercial firms and the accounting firms. Finally, we test the proposed platform for evaluating the relevance of IFRS in a prior-implementation context on the Tunisian case. We found that this emergent country is not yet prepared for an immediate transition to IFRS. Overall, the potential level of de facto voluntary harmonization is rather due to foreign institutional pressures that require IFRS as a reassuring and seducing label, than to a real local need.

KEYWORDS: IFRS, harmonization, relevance, device, contingency theory, neo-institutional approach.

RESUME: Cet article propose un dispositif d'appréciation a priori de la pertinence des IFRS pour tout pays désireux de se décider quant au passage ou non à l'harmonisation comptable internationale. C'est ainsi que, un outil est proposé pour mesurer le niveau d'harmonisation volontaire des pratiques comptables locales avec les pratiques recommandées par les IFRS. Egalement, cet article pose les fondements théoriques susceptibles d'expliquer ce niveau évalué d'harmonisation internationale de facto volontaire, par des facteurs contingents et/ou institutionnels. Les variables, hypothèses et modèles de recherche, propres à chacune de deux populations locales distinctes, les cabinets comptables et d'audit et les entreprises du pays, sont ainsi posés à partir de la théorie de la contingence et de l'approche néo-institutionnelle. Enfin, une application directe du dispositif proposé d'appréciation a priori de la pertinence des IFRS est réalisée, en se basant sur le cas de la Tunisie, et montre que ce pays n'est pas encore prêt pour un passage immédiat aux IFRS. En effet, pour l'ensemble des entités économiques du pays, le niveau d'harmonisation de facto est plus dû à des pressions institutionnelles étrangères nécessitant l'affichage d'un « label IFRS », qu'à un réel besoin local.

MOTS-CLEFS: IFRS, harmonisation, pertinence, dispositif, théorie de la contingence, approche néo-institutionnelle.

Corresponding Author: Raoudha Trabelsi (raoudha.trabelsi@gmail.com)

1 INTRODUCTION

Aujourd'hui, la mondialisation financière crée une polémique autour de ses effets sur le monde. Prônée par certains et responsable de tous les maux pour d'autres, elle se trouve au cœur des débats politiques et économiques [1].

Ceci est également le cas de l'harmonisation comptable internationale, un processus qui, bien que faisant l'objet de la réunion des efforts de plusieurs pays, ne fait pas l'unanimité. En effet, cette aspiration à une comptabilité universelle est loin d'être facile à concrétiser, en raison des différences entre pays du monde, souvent assez importantes pour empêcher la compréhension entre cultures. D'ailleurs, on reproche souvent aux normes comptables internationales un type de culture trop marqué et peu objectif. Or, selon [2], la réussite du partenariat entre ces pays engagés dans une collaboration internationale ne peut et ne doit sous-estimer l'importance de la variable culturelle. C'est ainsi que plusieurs pays et organismes nationaux de normalisation comptable perçoivent l'idée de l'harmonisation comptable comme un risque de perte de leurs cultures nationales, et refusent donc de se soumettre à un système doté d'une identité générique.

C'est pourquoi, souvent les réformes vers l'harmonisation comptable internationale sont « forcées », induites par des pressions d'institutions influentes, telles que la banque mondiale ou le Fonds Monétaire International qui exigent le « label IFRS » [3], plutôt que par un réel besoin et une profonde conviction émanant de l'intérieur du pays. Dans ce contexte, cette recherche tente de répondre à la problématique suivante :

« **Comment peut-on apprécier, au préalable, la pertinence ou non des IFRS pour un pays donné?** »

Ainsi, l'objectif de cette étude est de tenter de construire un dispositif d'aide à la décision d'adoption ou non des IFRS pour un pays donné. Or, vu que l'harmonisation *de jure* (ou des règles comptables) est beaucoup moins compliquée à s'accomplir que l'harmonisation *de facto* (ou des pratiques comptables), puisque, pour la première, tout se joue au niveau de la « formulation » des règles comptables pouvant même être « copiées » directement sur les IFRS, le problème se pose donc notamment au niveau du processus d'harmonisation comptable de facto. En effet, l'accomplissement de ce dernier nécessite une évolution culturelle comptable, et donc l'abandon, par tous les acteurs, de leur culture comptable d'origine, pour une adhésion à l'esprit des IFRS.

De ce fait, il conviendrait d'abord de faire appel à un outil permettant d'évaluer quantitativement, au sein des entités économiques du pays, le niveau d'harmonisation volontaire des pratiques comptables locales avec les pratiques recommandées par les IFRS. Toutefois, cette étape ne suffit pas pour conclure si une éventuelle adoption des IFRS par le pays en question est pertinente ou non. En effet, le fait que les pratiques des professionnels soient volontairement en harmonie ou en divergence avec celles recommandées par les normes IFRS peut, certes, donner une idée sur le degré de diffusion de la culture des IFRS dans un pays donné, mais ne peut affirmer la pertinence de ces normes pour ce même pays.

C'est pourquoi, une fois le niveau d'harmonisation volontaire de facto évalué, il conviendrait de pousser l'analyse en tentant d'expliquer ce dernier par deux soubassements théoriques, la théorie de la contingence et l'approche néo-institutionnelle. Autrement dit, deux types d'hypothèses de recherche sont émises ; celles qui se basent sur la théorie de la contingence, et dont la validation éventuelle confirme que le niveau d'harmonisation du pays est dû à des facteurs de contingence traduisant un réel besoin local de passer aux IFRS ; et celles qui se basent sur l'approche néo-institutionnelle, dont la validation éventuelle rattache le niveau d'harmonisation atteint à des pressions institutionnelles, notamment étrangères, poussant les entités à afficher le « label IFRS » [3].

Ainsi, cette recherche tente de sélectionner un outil d'évaluation de l'harmonisation comptable de facto au sein des entités d'un pays donné (1), pour présenter ensuite les soubassements théoriques à emprunter pour déceler la pertinence ou non des IFRS pour ledit pays : l'approche contingente et l'approche néo-institutionnelles (2). Enfin, ce dispositif d'appréciation de la pertinence des IFRS sera appliqué aux entités économiques d'un pays émergent, la Tunisie, et ce afin d'évaluer, a priori, l'éventuel choix de passage aux IFRS dans ce pays (3).

2 LA SÉLECTION D'UNE GRILLE DE MESURE DU NIVEAU D'HARMONISATION DE FACTO DANS UN PAYS N'AYANT PAS ENCORE ADOPTÉ LES IFRS

A notre connaissance, la littérature existante ne propose pas de grille permettant la mesure, au sein d'une entité, du niveau d'harmonisation de facto ou des pratiques comptables convergentes avec les IFRS. En effet, une telle rationalité instrumentale est généralement proposée par les différents acteurs de la profession comptable, tels que les organismes de normalisation, les associations ou organisations relatives à la profession comptable et d'audit, certains cabinets ou réseaux de cabinets d'expertise comptable et d'audit, etc. En l'occurrence, l'étude documentaire réalisée dans le cadre de cette recherche a permis de repérer diverses grilles de mesures, élaborées par les professionnels, notamment dans les pays où l'on

procédait à court terme à un passage obligatoire aux IFRS. Ceci est le cas du Canada où, à partir de l'année 2011, les entités canadiennes ayant une obligation publique de rendre des comptes, y compris les sociétés cotées en Bourse, ont été tenues d'adopter les IFRS, en remplacement des Principes Comptables Généralement Reconnus (PCGR) du Canada.[1] Dans le cadre de cette réforme comptable de 2011, le cabinet canadien Samson Bélair, membre du réseau international Deloitte Touche Tohmatsu Limited, l'un des quatre plus grands cabinets d'audit et de conseil au monde (Big Four), avait établi en avril 2008 une grille d' « évaluation de l'état de préparation à la conversion aux IFRS », un outil mis à dispositions des entités canadiennes, afin d'apprécier leur niveau de préparation pour le passage aux normes IFRS [4]. Ainsi, étant donnée la renommée du réseau d'expertise comptable et d'audit Deloitte, émanant notamment de son appartenance aux quatre plus grands cabinets au monde (Big Four), et face à l'absence de tout autre outil similaire dans la littérature scientifique, la grille de Deloitte sera adaptée pour mesurer le niveau d'harmonisation de facto volontaire, dans tout pays n'ayant pas encore décidé d'adopter les IFRS.

Dans ce qui suit, nous exposerons la grille de Deloitte (2.1), nous justifierons le choix du recours à cette dernière (2.2), pour en proposer enfin une adaptation aux pays n'ayant pas encore procédé à l'harmonisation (2.3.).

2.1 PRÉSENTATION DE LA GRILLE D' « ÉVALUATION DE L'ÉTAT DE PRÉPARATION À LA CONVERSION AUX IFRS » DE DELOITTE

Malgré la grande similitude établie à plusieurs niveaux entre les IFRS et les PCGR canadiens, le cabinet Deloitte canadien considère que « les sociétés ouvertes canadiennes ainsi que les investisseurs, prêteurs et conseillers devront se familiariser avec les nouvelles normes et adapter leurs processus. Grâce à une planification attentive et à une stratégie de mise en œuvre bien réfléchies, les sociétés devraient être en mesure de se convertir aux IFRS en douceur et de manière efficiente » [4].

2.1.1 COMPOSANTES DE LA GRILLE

Cette grille prend en compte douze principaux facteurs clé de succès pour réaliser une conversion aux IFRS réussie. Ci-dessous, les douze éléments sur lesquels porte la grille d' « évaluation de l'état de préparation à la conversion aux IFRS » sont intégralement retranscrits. Cette grille se base sur des échelles de mesure [5], les répondants devant évaluer l'accomplissement de chacun des douze éléments en attribuant une note qui varie sur une échelle allant de « 1 » (Début du processus) à « 5 » (Mise en œuvre terminée).

Table 1. Synthèse de la grille d'évaluation de l'état de préparation à la conversion aux IFRS de Deloitte

		Échelle de l'état de préparation à la conversion				
Les douze principaux facteurs clé de succès pour réaliser une conversion aux IFRS réussie selon Deloitte		Début du processus (1)	(2)	(3)	(4)	Mise en œuvre terminée (5)
Élément 1 Sensibilisation	Comprendre la stratégie du Conseil canadien des normes comptables relative au passage aux IFRS ainsi que le calendrier de leur mise en œuvre.					
Élément 2 Surveillance	Élaborer un plan pour le passage aux IFRS qui comporte un calendrier détaillé et une attribution claire des responsabilités.					
Élément 3 Compréhension	Mettre au point un programme de formation sur les IFRS pour l'ensemble des membres du conseil et des employés qui touchent à l'information financière.					

[1] Les entreprises à capital fermé qui présentent leur information financière selon les PCGR canadiens ont eu le choix d'adopter, pour les exercices ouverts à compter du 1er janvier 2011, soit les Normes Comptables pour les Entreprises à Capital Fermé (NCECF) ou les Normes internationales d'information financière (IFRS).

Élément 4 Avantage concurrentiel	Élaborer, pour se démarquer par rapport aux concurrents, des lignes directrices relatives à la communication de l'information sur les IFRS ; Avoir fourni des renseignements clairs et transparents au sujet du plan de conversion aux IFRS et des conséquences sur ses documents de présentation de l'information financière, y compris des renseignements fournis sur le site Web de l'entité.					
Élément 5 Contrôles internes	Passer en revue les systèmes de contrôle interne de la société dans le but de repérer toutes les modifications exigées par le passage aux IFRS (c.-à-d. les renseignements supplémentaires provenant des services non liés à la présentation de l'information financière).					
Élément 6 Vérification	Participation active des vérificateurs internes et externes au plan de conversion aux IFRS afin de s'assurer que toutes les différences significatives entre les normes comptables actuelles et les IFRS sont traitées adéquatement.					
Élément 7 Systèmes des Technologies de l'Information (TI)	Examen des systèmes de TI qui appuient le processus de présentation de l'information financière dans le but de repérer toutes les modifications exigées par le passage aux IFRS (l'information à fournir supplémentaire exigée et les principes comptables différents).					
Élément 8 Appréciation du risque	Réalisation d'une appréciation globale du risque, y compris des risques liés à l'information à fournir, et conception de contrôles et mesures d'atténuation afin de réduire les risques identifiés.					
Élément 9 Incidence sur le financement	Évaluation et traitement des conséquences sur les clauses restrictives, les ratios financiers, le fonds de roulement etc. de la société, conjointement avec les créanciers de la société.					
Élément 10 Incidence sur les impôts	Évaluation et traitement de l'incidence, dans la stratégie fiscale de la société, de la conversion aux IFRS (montants et information à fournir) sur les impôts sur les bénéfices, exigibles et futurs.					
Élément 11 Incidence sur les parties intéressées	Les conséquences sur l'information financière de la société ont été analysées et seront clairement communiquées dans le rapport de gestion conformément au calendrier proposé par les Autorités canadiennes en valeurs mobilières.					
Élément 12 Incidence sur les ressources	Les heures à consacrer et la formation requise pour le passage aux IFRS ont été évaluées pour la haute direction, le service des finances, les TI, le service juridique, la trésorerie, les opérations et les autres services et fonctions.					

2.1.2 INTERPRETATION DES RESULTATS

Les analystes de Deloitte interprètent les réponses des entités canadiennes selon le barème présenté ci-dessous.

Table 2. Barème d'interprétation des scores obtenus par la grille de Deloitte

Réponses situées dans le bas de l'échelle (1 ou 2)	La société doit probablement accélérer son plan de conversion afin de s'assurer de respecter le calendrier de mise en œuvre.
Réponses situées au milieu de l'échelle (3)	La société a encore du travail à faire pour s'assurer de respecter le calendrier de mise en œuvre et d'effectuer le passage aux IFRS de manière contrôlée.
Réponses situées dans le haut de l'échelle (4 ou 5)	La société semble bien en selle pour effectuer le passage en temps opportun et de manière soignée et mesurée.

2.2 JUSTIFICATION DU CHOIX DE LA GRILLE DE DELOITTE

La grille d'évaluation de Deloitte a pour objectif d' « aider les entités économiques canadiennes à évaluer le degré de préparation au passage aux IFRS de l'organisation et à repérer les secteurs qui méritent un examen plus approfondi » [4]. Sans ignorer les limites de l'application d'une grille de mesure qui n'est pas issue de la littérature, et établie par un cabinet de comptabilité et d'audit, nous avons opté, dans ce travail de recherche pour l'application de la grille d'évaluation de l'état de préparation à la conversion aux IFRS de Deloitte, après son adaptation aux pays n'ayant pas encore décidé d'adopter les IFRS. La logique qui a conduit à ce choix, malgré ses limites, se base sur plusieurs justifications.

D'un point de vue théorique, à notre connaissance, la littérature comptable existante ne propose pas de grille permettant la mesure, au sein d'une entité, du niveau d'harmonisation de facto ou de familiarisation avec les IFRS, conformément aux besoins de notre étude. En effet, les études que nous avons pu repérer grâce à notre revue de la littérature, présentent des travaux d'évaluation de l'harmonisation comptable poursuivant, pour la plupart, une finalité différente de celle à laquelle nous aspirons dans notre travail de recherche. Parmi ces études, les travaux [6] - [8] introduisent l'indice de comparabilité « C », un indicateur des effets de l'harmonisation, qui mesure le degré de comparabilité de l'information comptable diffusée par les sociétés. Malgré sa pertinence, cet indice ne peut être exploité dans notre étude, puisqu'il constitue un modèle de comparaison interentreprises, en mesurant les pratiques de différents groupes d'entreprises d'un même pays ou de pays différents, à un moment donné, pour un aspect comptable particulier, permettant ainsi la mesure de la proportion de comparaisons faisables par paires d'entreprises, compte tenu des méthodes comptables adoptées par ces dernières. Or, la variable que nous cherchons à mesurer dans notre étude est, pour chacune des entités économiques constituant l'échantillon, le score de familiarisation avec les IFRS, ou encore le score de préparation au passage à ces normes, de manière générale et non pas par rapport à un élément particulier des états financiers. Ceci est donc bien différent de ce que propose Van Der Tas, avec son indice « C », et donc de ce que proposent les études ultérieures qui s'en inspirent, telles que [9], [10]. Ceci explique donc le recours que nous avons dû faire à une grille de mesure établie par un groupe de professionnels.

D'un point de vue pratique, parmi les cabinets ayant proposé des grilles de mesures ou « check-lists » relatives à l'évaluation de l'application des IFRS, le réseau international Deloitte Touche Tohmatsu Limited est le plus ancien. Leader mondial dans le domaine de la comptabilité et d'audit, il a réalisé en 2010 un chiffre d'affaires de 26,6 milliards de dollars, et dispose, depuis 2009, de la plus grande masse salariale comptant 169 000 employés [11]. De plus, ce réseau de cabinets dispose de centres de recherche, réunissant, en plus des professionnels les plus compétents, des doctorants et docteurs en comptabilité, contrôle audit, garantissant un certain niveau de rigueur et pertinence aux diverses publications et études produites par le groupe de cabinets. De ce fait, se basant sur le savoir-faire du cabinet Deloitte, la grille d' « évaluation de l'état de préparation à la conversion aux IFRS » peut être considérée comme une bonne alternative, face à l'absence, dans la littérature comptable, d'outil de mesure du niveau d'harmonisation de facto approprié à notre étude.

D'un point de vue méthodologique, la grille que nous proposons est une grille de suivi de l'avancement de la procédure de passage aux IFRS. Or, dans le domaine des contrôles de procédures, les cabinets d'audit ont un savoir-faire et une expertise reconnue, y compris par les académistes. En conséquence, les grilles de mesure de la connaissance ou de l'application des IFRS dans les entités sont généralement établies par les grands cabinets comptables et d'audit, notamment dans les pays où l'on procède à court terme à un passage obligatoire aux IFRS. En l'occurrence, ceci a été le cas du Canada, ayant engagé une réforme comptable pour l'adoption, à partir de 2011, les IFRS pour les entités ayant une obligation publique de rendre des comptes, en remplacement des Principes Comptables Généralement Reconnus (PCGR) du Canada. Dans le cadre de cette réforme, le cabinet canadien Samson Bélair, membre du réseau international Deloitte Touche Tohmatsu Limited, a établi en avril 2008 une grille d' « évaluation de l'état de préparation à la conversion aux IFRS », un outil mis à dispositions des entités canadiennes, afin d'apprécier leur niveau de préparation pour le passage aux normes IFRS.

Au même titre, dans un pays donné, avant de prévoir une parfaite harmonisation comptable internationale, avec une adoption intégrale des IFRS, il est important qu'à l'instar des entités canadiennes, les entités planifient rigoureusement un tel passage, prévoyant toutes les étapes nécessaires à la réussite du processus de conversion, et prenant en considération toutes les ressources nécessaires permettant un passage efficient et sans qu'il y ait « indigestion » des nouvelles normes. En l'occurrence, l'utilisation d'une telle grille, pour apprécier le niveau de préparation des entités canadiennes au passage aux normes IFRS, peut être considérée comme pertinente pour tout autre pays cherchant à évaluer la familiarisation actuelle de ses entités avec l'esprit des normes IFRS, avant de décider ou non de passer à ses normes.

Toutefois, si cet outil est élaboré par Deloitte pour s'appliquer spécialement au cas canadien dans le cadre d'une réforme vers un passage obligatoire et imminent vers les IFRS, le fait de l'emprunter pour l'appliquer à un contexte qui relève d'une simple volonté d'évaluation du niveau de préparation des entités pour une adoption éventuelle des IFRS, sans qu'il y ait une échéance fixée pour une conversion imminente, peut ne pas être aussi pertinent. De ce fait, il conviendrait de procéder,

avant d'exploiter la grille de Deloitte, à son adaptation, en prenant en considération le décalage qui existe entre les deux niveaux d'avancement du processus d'harmonisation (harmonisation imminente et harmonisation éventuelle).

2.3 VERS UNE GRILLE D'ÉVALUATION DE L'HARMONISATION DE FACTO DANS TOUT PAYS N'AYANT PAS ENCORE ADOPTÉ LES IFRS

L'objectif étant d'évaluer le niveau d'harmonisation comptable de facto dans un pays donné, pour tenter ensuite d'expliquer ce niveau par des variables contingentes ou institutionnelles, la grille devrait être susceptible d'attribuer un score d'harmonisation de facto par entité économique. Pour ce faire, les douze éléments retenus par Deloitte comme principaux facteurs clé de succès pour la conversion aux IFRS sont repris, adaptés à un contexte où le pays n'a pas encore pris la décision d'adopter ou non les IFRS. Ces éléments seront soumis aux professionnels exerçant dans les entités du pays, pour évaluation.

L'adaptation de la grille de Deloitte a nécessité la prise en considération des différences existant entre le contexte de la réforme comptable canadienne pour une adoption définitive des IFRS, et celui de pays où le passage aux IFRS est encore incertain. En effet, le besoin des entités canadiennes était d'effectuer une évaluation de leur niveau de préparation à un passage imminent et obligatoire aux IFRS, ce qui renferme entre autres l'analyse du degré d'avancement de toutes les étapes du processus de conversion aux IFRS, comme la tenue d'un calendrier de planification, la communication aux diverses parties prenantes, l'estimation des ressources nécessaires à la conversion aux IFRS, etc. Certaines de ces étapes ne peuvent concerner, du moins dans l'immédiat, le cas d'un pays dont la décision de passage aux IFRS n'est pas encore prise. En revanche, ce qui est pertinent pour ce dernier, c'est d'évaluer le niveau de familiarisation avec la culture des IFRS, c'est-à-dire le niveau d'harmonisation comptable internationale de facto au sein des entités économiques, et ce pour tenter ensuite d'établir un lien entre ce niveau d'harmonisation et certaines variables propres à l'entité.

Le tableau 3 présente l'adaptation de la grille originale destinée aux entités canadiennes, aux spécificités d'un pays n'ayant pas adopté les IFRS, afin de pouvoir l'exploiter pour évaluer le niveau d'harmonisation de facto au sein dudit pays.

Table 3. Adaptation de la grille originale de Deloitte, destinée aux entités canadiennes, aux pays n'ayant pas encore pris la décision de passer aux IFRS

Les douze principaux facteurs clé de succès pour réaliser une conversion aux IFRS réussie selon Deloitte		
	Version originale destinée aux entités canadiennes	**Version adaptée**
Élément 1 Sensibilisation	Comprendre la stratégie du Conseil canadien des normes comptables relative au passage aux IFRS ainsi que le calendrier de leur mise en œuvre.	Avoir un niveau minimum de connaissance des normes IFRS, de points de vue définition et contenu.
Élément 2 Surveillance	Élaborer un plan pour le passage aux IFRS qui comporte un calendrier détaillé et une attribution claire des responsabilités.	Estimer le temps nécessaire pour que l'entité bascule sans problèmes aux IFRS.
Élément 3 Compréhension	Mettre au point un programme de formation sur les IFRS pour l'ensemble des membres du conseil et des employés qui touchent à l'information financière.	Identifier, au sein de l'entité, la part des cadres comptables et financiers compétents en matière d'IFRS par rapport au nombre total de cadres comptables et financiers.
Élément 4 Avantage concurrentiel	Élaborer, pour se démarquer par rapport aux concurrents, des lignes directrices relatives à la communication de l'information sur les IFRS ; Avoir fourni des renseignements clairs et transparents au sujet du plan de conversion aux IFRS et des conséquences sur ses documents de présentation de l'information financière, y compris des renseignements fournis sur le site Web de l'entité.	Estimer l'importance de la prise en compte de la concurrence dans la décision d'adopter définitivement les IFRS.
Élément 5 Contrôles internes	Passer en revue les systèmes de contrôle interne de la société dans le but de repérer toutes les modifications exigées par le passage aux IFRS (c.-à-d. les renseignements supplémentaires provenant des services non liés à la présentation de l'information financière).	Estimer l'importance de la conception du référentiel international comme solution palliative aux lacunes du système comptable local du pays étudié.

Élément 6 Vérification	Participation active des vérificateurs internes et externes au plan de conversion aux IFRS afin de s'assurer que toutes les différences significatives entre les normes comptables actuelles et les IFRS sont traitées adéquatement.	Connaitre les principales spécificités des normes IFRS par rapport aux normes locales du pays étudié.
Élément 7 Systèmes des Technologies de l'Information (TI)	Examen des systèmes de TI qui appuient le processus de présentation de l'information financière dans le but de repérer toutes les modifications exigées par le passage aux IFRS (l'information à fournir supplémentaire exigée et les principes comptables différents).	Cet élément ne peut être évalué dans un pays n'ayant pas encore décidé d'adopter les IFRS, vu qu'il s'agit d'une étape très avancée du processus d'harmonisation.
Élément 8 Appréciation du risque	Réalisation d'une appréciation globale du risque, y compris des risques liés à l'information à fournir, et conception de contrôles et mesures d'atténuation afin de réduire les risques identifiés.	Évaluer la perception des professionnels quant à l'importance d'un passage aux IFRS.
Élément 9 Incidence sur le financement	Évaluation et traitement des conséquences sur les clauses restrictives, les ratios financiers, le fonds de roulement etc. de la société, conjointement avec les créanciers de la société.	Cet élément ne peut être évalué dans un pays n'ayant pas encore décidé d'adopter les IFRS, vu qu'il s'agit d'une étape très avancée du processus d'harmonisation.
Élément 10 Incidence sur les impôts	Évaluation et traitement de l'incidence, dans la stratégie fiscale de la société, de la conversion aux IFRS (montants et information à fournir) sur les impôts sur les bénéfices, exigibles et futurs.	Cet élément ne peut être évalué dans un pays n'ayant pas encore décidé d'adopter les IFRS, vu qu'il s'agit d'une étape très avancée du processus d'harmonisation.
Élément 11 Incidence sur les parties intéressées	Les conséquences sur l'information financière de la société ont été analysées et seront clairement communiquées dans le rapport de gestion conformément au calendrier proposé par les Autorités canadiennes en valeurs mobilières.	Pour un pays n'ayant pas encore décidé d'adopter les IFRS, on ne peut pas encore parler de communication aux parties prenantes pour les informer du déclenchement du processus de conversion aux IFRS. Par contre, il est possible d'exploiter cet élément pour s'interroger sur la principale partie prenante privilégiée volontairement par l'entité, en matière de communication financière.[2]
Élément 12 Incidence sur les ressources	Les heures à consacrer et la formation requise pour le passage aux IFRS ont été évaluées pour la haute direction, le service des finances, les TI, le service juridique, la trésorerie, les opérations et les autres services et fonctions.	Évaluer les difficultés appréhendées par l'entité quant à un éventuel passage aux IFRS, et ce afin de déduire le niveau de préparation et la disponibilité des ressources nécessaires pour réaliser un tel passage.

Ainsi, dans le cadre de l'adaptation de la grille de Deloitte, seulement neuf sur les douze principaux facteurs clé de succès pour la réalisation d'une conversion aux IFRS réussie sont retenus. En effet, l'élément 7, portant sur la mise à jour des systèmes des technologies de l'information pour s'adapter aux normes IFRS, ainsi que les éléments 9 et 10, portant sur l'appréhension de l'incidence de l'adoption des IFRS, respectivement sur le financement et sur les impôts, correspondant à une étape avancée de la conversion aux IFRS, ne peuvent être pris en considération.

Par ailleurs, la grille initiale, destinée aux entités canadiennes, appelait ces dernières à attribuer à chacun des douze éléments facteurs clé de succès de la conversion un pointage sur une échelle allant de « 1 » (Début du processus) à « 5 » (Mise en œuvre terminée), afin de décrire au mieux la situation de la société quant à l'élément en question. L'adaptation de la grille a nécessité l'apport de quelques modifications à la méthode de cette notation, tout en respectant l'échelle à cinq

[2] Si l'investisseur s'avère être la partie prenante privilégiée, il s'agit là d'un signe de présence de culture comptable anglo-saxonne, et donc d'un niveau important d'harmonisation de facto.

niveaux pour l'évaluation de l'importance de chacun des neuf éléments à évaluer. De ce fait, pour simplifier le calcul des scores de familiarisation avec les IFRS en les rendant compris entre « 0 » et « 36 » au lieu d'un intervalle allant de « 9 » à 45 », l'échelle d'évaluation du niveau de préparation à la conversion en IFRS allant de « 1 » à « 5 », est remplacée par une échelle allant de « 0 » à « 4 », le niveau « 0 » correspondant à un niveau faible de l'élément évalué, et le niveau « 4 » de l'échelle correspondant à un niveau très avancé.

Le tableau 4 présente l'adaptation des items évalués par la grille originale destinée aux entités canadiennes, ainsi que notre proposition de deux grilles, afin d'évaluer le niveau d'harmonisation de facto au sein des cabinets comptables et d'audit et au sein des entreprises, tous localisés dans le pays à étudier.

Table 4. Proposition de deux grilles adaptées aux entités locales (Cabinets et entreprises)

Les facteurs de réussite d'une conversion aux IFRS selon Deloitte	**Version adaptée aux cabinets comptables et d'audit**	**Version adaptée aux entreprises**
Élément 1 Sensibilisation	Comment évaluez-vous votre niveau de connaissances en IFRS ? ✓ Excellent (4) ✓ Très fort (3) ✓ Fort (2) ✓ Moyen (1) ✓ Aucun niveau (0)	Comment évaluez-vous votre niveau de connaissances en IFRS ? ✓ Excellent (4) ✓ Très fort (3) ✓ Fort (2) ✓ Moyen (1) ✓ Aucun niveau (0)
Élément 2 Surveillance	Combien de temps cela vous prendrait-il pour basculer sans problèmes aux IFRS ? ✓ Moins d'1an (4) ✓ Entre 1an et 2ans (3) ✓ Entre 2ans et 3ans (2) ✓ Plus de 3ans (1) ✓ Jamais (0)	Combien de temps cela vous prendrait-il pour basculer sans problèmes aux IFRS ? ✓ Moins d'1an (4) ✓ Entre 1an et 2ans (3) ✓ Entre 2ans et 3ans (2) ✓ Plus de 3ans (1) ✓ Jamais (0)
Élément 3 Compréhension	Au sein de votre cabinet : Part des cadres comptables et financiers compétents en matière d'IFRS par rapport au nombre total de cadres comptables et financiers ✓ [80 , 100%] (4) ✓ [60% , 80%[(3) ✓ [40% , 60%[(2) ✓ [20% , 40%[(1) ✓ [0 , 20%[(0)	Au sein de votre entreprise : Part des cadres comptables et financiers compétents en matière d'IFRS par rapport au nombre total de cadres comptables et financiers ✓ [80 , 100%] (4) ✓ [60% , 80%[(3) ✓ [40% , 60%[(2) ✓ [20% , 40%[(1) ✓ [0 , 20%[(0)
Élément 4 Avantage concurrentiel	Selon vous, si un jour vous en êtes à adopter définitivement les IFRS, quelles seraient les raisons de cette adoption ? Classez les propositions suivantes de 1 à 5, 1 correspondant à la raison d'adoption la plus importante et 5 à la moins importante ✓ Le cabinet tête du réseau vous oblige, tout comme les autres cabinets membres du réseau, à l'adoption des IFRS ou du moins à l'implication en ces normes ✓ L'État vous y oblige, comme touts les autres cabinets répondant à certains critères ✓ Pour essayer de suivre la tendance internationale ✓ Vous essayez de faire comme les organisations ayant déjà adopté les IFRS	Selon vous, si un jour vous en êtes à adopter définitivement les IFRS, quelles seraient les raisons de cette adoption ? Classez les propositions suivantes de 1 à 5, 1 correspondant à la raison d'adoption la plus importante et 5 à la moins importante ✓ La société mère vous y oblige, comme toutes ses autres filiales (si vous faites partie d'un groupe) ✓ L'État vous y oblige, comme toutes les autres entreprises répondant à certains critères ✓ Vous remarquez que les consultants suggèrent l'utilisation des IFRS ✓ Pour essayer de suivre la tendance internationale ✓ Vous essayez de faire comme les organisations ayant déjà adopté les IFRS

	✓ **Vos concurrents principaux adoptent les IFRS [1 : (4), 2 : (3), 3 : (2), 4 : (1), 5 et plus ou pas de classement : (0)]** ✓ Vous remarquez que les autres cabinets du même groupe auquel vous appartenez ont adopté les IFRS ✓ Vous remarquez que les plus grands cabinets locaux ou internationaux exerçant localement ou ailleurs, notamment les Big Four suggèrent l'utilisation des IFRS	✓ La société mère adopte les IFRS même si elle ne vous y oblige pas (si vous faites partie d'un groupe) ✓ **Vos concurrents principaux adoptent les IFRS [1 : (4), 2 : (3), 3 : (2), 4 : (1), 5 et plus ou pas de classement : (0)]** ✓ Vous remarquez que les autres filiales du même groupe auquel vous appartenez ont adopté les IFRS
Élément 5 Contrôles internes	Selon vous, quelle solution pourrait-on envisager pour pallier aux lacunes du système comptable local ? ✓ Le système comptable local est parfait et ne comporte pas de lacunes (0) ✓ Le laisser tel qu'il est à condition que le normalisateur local produise des documents détaillés explicatifs des cas particuliers qui posent problème. (1) ✓ L'abroger carrément et produire un nouveau système comptable propre aux entreprises locales et à leurs particularités. (2) ✓ Apporter à ce système certaines modifications inspirées du référentiel comptable international (IFRS). (3) ✓ L'abroger carrément, et tant qu'à faire adopter directement le référentiel comptable international (IFRS). (4)	Selon vous, quelle solution pourrait-on envisager pour pallier aux lacunes du système comptable local ? ✓ Le système comptable local est parfait et ne comporte pas de lacunes (0) ✓ Le laisser tel qu'il est à condition que le normalisateur local produise des documents détaillés explicatifs des cas particuliers qui posent problème. (1) ✓ L'abroger carrément et produire un nouveau système comptable propre aux entreprises locales et à leurs particularités. (2) ✓ Apporter à ce système certaines modifications inspirées du référentiel comptable international (IFRS). (3) ✓ L'abroger carrément, et tant qu'à faire adopter directement le référentiel comptable international (IFRS). (4)
Élément 6 Vérification	Quatre questions évaluées à 0 ou 1 selon les réponses pour obtenir un score total entre 0 et 4 : - Si vous aviez le choix, quelle est la méthode de comptabilisation que vous privilégieriez personnellement ? ✓ Coût historique (0) ✓ Juste valeur (1) ✓ Autre réponse (0) - Pouvez-vous indiquer une spécificité des IFRS par rapport aux normes locales ? ✓ Non (0) - Oui (1) - Pouvez-vous expliquer les changements apportés par les IFRS par rapport aux IAS ? ✓ Non (0) - Oui (1) - Avez-vous déjà eu l'occasion de parcourir le référentiel comptable international ? ✓ Non (0) - Oui (1)	Quatre questions évaluées à 0 ou 1 selon les réponses pour obtenir un score total entre 0 et 4 : - Si vous aviez le choix, quelle est la méthode de comptabilisation que vous privilégieriez personnellement ? ✓ Coût historique (0) ✓ Juste valeur (1) ✓ Autre réponse (0) - Pouvez-vous indiquer une spécificité des IFRS par rapport aux normes locales ? ✓ Non (0) - Oui (1) - Pouvez-vous expliquer les changements apportés par les IFRS par rapport aux IAS ? ✓ Non (0) - Oui (1) - Avez-vous déjà eu l'occasion de parcourir le référentiel comptable international ? ✓ Non (0) - Oui (1)
Élément 7 Systèmes des Technologies de l'Information (TI)		
Élément 8 Appréciation du risque	Que pensez-vous de l'adoption des IFRS par les entités locales ? ✓ C'est indispensable (4) ✓ Ce serait une bonne chose (2)	Que pensez-vous de l'adoption des IFRS par les entités locales ? ✓ C'est indispensable (4) ✓ Ce serait une bonne chose (2)

	✓ Cela ne sera pas une bonne idée (0)	✓ Cela ne sera pas une bonne idée (0)
Élément 9 Incidence sur le financement		
Élément 10 Incidence sur les impôts		
Élément 11 Incidence sur les parties intéressées	Quand vous préparez les états financiers de vos clients, quel(s) utilisateur(s) de l'information financière privilégiez-vous ? Classez les propositions suivantes de 1 à 5, 1 correspondant à l'utilisateur le plus privilégié et 5 au moins privilégié. ✓ Les dirigeants ✓ **Les investisseurs (propriétaires)** **[1 : (4), 2 : (3), 3 : (2), 4 : (1), 5 et plus ou pas de classement : (0)]** ✓ L'Administration fiscale ✓ La Société Mère si l'entreprise est une filiale de multinationale ✓ Les bailleurs de fonds ✓ Autres	Quand vous préparez vos états financiers, quel(s) utilisateur(s) de l'information financière privilégiez-vous le plus ? Classez les propositions suivantes de 1 à 5, 1 correspondant à l'utilisateur le plus privilégié et 5 au moins privilégié. ✓ Les dirigeants ✓ **Les investisseurs (propriétaires)** **[1 : (4), 2 : (3), 3 : (2), 4 : (1), 5 et plus ou pas de classement : (0)]** ✓ L'Administration fiscale ✓ La Société Mère si l'entreprise est une filiale de multinationale ✓ Les bailleurs de fonds ✓ Autres
Élément 12 Incidence sur les ressources	Selon-vous quelle serait la principale difficulté à craindre lors du passage aux IFRS ? ✓ Votre cabinet n'est pas prêt pour ce changement, les cadres n'ont pas eu la formation nécessaire et cela va être difficile et long de les recycler (0) ✓ Votre cabinet n'est pas prêt pour ce changement, les cadres n'ont pas eu la formation nécessaire mais il sera possible de les recycler assez rapidement (2) ✓ Il n'y aura pas de grandes difficultés, les IFRS ne sont pas très différentes des normes locales (4)	Selon-vous quelle serait la principale difficulté à craindre lors du passage aux IFRS ? ✓ Votre entreprise n'est pas prête pour ce changement, les cadres n'ont pas eu la formation nécessaire et cela va être difficile et long de les recycler (0) ✓ Votre entreprise n'est pas prête pour ce changement, les cadres n'ont pas eu la formation nécessaire mais il sera possible de les recycler assez rapidement (2) ✓ Il n'y aura pas de grandes difficultés, les IFRS ne sont pas très différentes des normes locales (4)

A l'aide des grilles d'évaluation du niveau d'harmonisation de facto au sein des entreprises et des cabinets comptables et d'audit du pays, un score de familiarisation avec les IFRS, compris entre 0 et 36, sera établi pour chaque entité faisant partie de l'échantillon sélectionné.

Il est à rappeler que l'objectif principal de l'étude est d'établir un lien entre le niveau de familiarisation avec les IFRS de chaque entité, avec des variables propres à cette dernière. Ainsi, en fonction des variables qui seraient identifiées comme responsables d'accroître le niveau d'harmonisation de facto au sein d'une entité, il serait possible de déterminer les déterminants de l'harmonisation de facto dans le même pays. En effet, s'il s'avère que le score obtenu dépend plus de variables contingentes, l'harmonisation de facto serait pertinente pour le pays, et expliquée donc par la théorie de la contingence. Par contre, si le score de familiarisation avec les IFRS est lié à des variables institutionnelles, l'harmonisation de facto serait tributaire de pressions institutionnelles, et expliquée par l'approche néo-institutionnelle.

Dans ce qui suit, les deux soubassements théoriques explicatifs du niveau d'harmonisation de facto sont présentés.

3 LES ORIGINES DE L'HARMONISATION DE FACTO : CADRE THÉORIQUE ET HYPOTHÈSES DE RECHERCHE

En se basant simultanément sur le cadre d'analyse de la théorie de la contingence et celui de l'approche néo-institutionnelle, il est possible d'expliquer le niveau d'harmonisation de facto volontaire dans un pays donné, et d'en déduire la pertinence des IFRS pour ce même pays.

Ainsi, si une corrélation importante est établie entre le niveau d'harmonisation de facto au sein des entités économiques, et les variables de contingence qui s'y rapportent, ceci montrerait qu'il serait pertinent pour le pays de se lancer dans le processus d'harmonisation comptable internationale. L'harmonisation de facto serait donc le résultat d'un réel besoin local, déterminé par des variables issues de la théorie de la contingence. Par contre, s'il existe une forte corrélation entre le niveau d'harmonisation de facto au sein des entités économiques, et des variables issues de l'approche néo-institutionnelle traduisant le contexte social de l'entité, les IFRS ne seraient pas réellement pertinentes pour le pays, puisque leur adoption viendrait juste répondre à des pressions institutionnelles, notamment étrangères, exercées sur les entités locales. L'harmonisation serait dans ce cas une simple étiquette, ou « Label » [3], à afficher par le pays pour gagner la confiance des différentes parties intéressées.

3.1 LES ORIGINES DE L'HARMONISATION DE FACTO : SOUBASSEMENTS THÉORIQUES

Afin d'expliquer le lien établi entre le niveau d'harmonisation de facto et ses déterminants potentiels, deux approches qui placent l'acteur au centre de l'analyse sont appliquées : l'approche contingente et l'approche néo-institutionnelle. Il est à rappeler que l'acteur est celui dont l'action contribue à construire des régulations [12]. La construction des règles (le construit social) est expliquée à partir du jeu des acteurs empiriques, calculateurs et intéressés. Ils sont dotés de rationalité, certes limitée, ils sont autonomes, intervenant dans un système qui est l'organisation, qui s'adapte à de diverses contingences et mutations. Ceci étant, l'ajustement ne peut se faire automatiquement, mais dépend de l'existence d'une intentionnalité de la part des acteurs [13].

3.1.1 HARMONISATION DE FACTO VOLONTAIRE RESULTANT DES CONTINGENCES DE L'ENTITE

S'inscrivant dans la théorie des organisations [14], la théorie de la contingence a été formulée en 1958, notamment à partir des travaux [15], [16]. En stipulant qu'il n'existe pas une forme idéale du système de gestion, et que les circonstances particulières ou les contingences dictent le meilleur choix de ce système, l'école de la contingence abandonne définitivement le préalable, longtemps dominant, du « the one best way » des Ecoles classiques et des relations humaines. Ainsi, l'influence exercée par des facteurs dits contingents structure l'organisation. Cette dernière est conçue comme un système ouvert, où la structure est adaptée aux contraintes de l'environnement et où les déterminants structurels sont cohérents avec les variables organisationnelles, de manière à concilier l'organisation avec son milieu [17].

3.1.1.1 PRINCIPAUX FACTEURS DE CONTINGENCE

La référence [18] montre que, face à l'instabilité de l'environnement et à ses mutations rapides, il n'existe pas une meilleure structure de l'organisation, mais différentes structures qui sont les meilleures dans différentes conditions. Ainsi, une organisation est bien contingente à des variables externes et internes. La référence [19] a consolidé, dans les années 80, l'école de la contingence, en établissant une typologie des différentes organisations en fonction des différents types de contingence. La référence [19] a identifié quatre facteurs de contingence ; l'âge et la taille de l'organisation, son système technique de production, son environnement, et son système de pouvoir. D'autres auteurs ont enrichi la théorie en proposant et testant empiriquement d'autres facteurs de contingence, tels que la stratégie adoptée par l'organisation, l'environnement culturel national du pays où se trouve cette dernière, etc. [17]. Ainsi, comme il est expliqué dans [20], la théorie de la contingence n'admet pas l'existence de facteurs clés de succès universels et valables en toute circonstance, mais c'est la combinaison d'un ensemble de facteurs environnementaux [21], [22], organisationnels [23], financiers [24], humains [25], qui peut être efficace ou non selon le contexte.

- **Les facteurs de contingence d'ordre environnemental**

Les références [21], [22] ont souligné l'importance de l'environnement en tant que facteur ayant un impact sur l'organisation. En effet, étant donné les effets que peuvent avoir des facteurs internes ou externes rattachés à l'organisation sur une ou plusieurs de ses composantes [26], la théorie de la contingence conçoit l'entreprise comme un système ouvert, formé de plusieurs sous-systèmes en perpétuelle interaction, dont la survie dépend de l'adaptation à l'environnement [27]. En d'autres termes, l'entreprise est tenue de s'adapter aux contraintes de l'environnement, qui ne sont ni statiques, ni homogènes, en vue de faire face aux incertitudes et de les contrôler au mieux [28]. De ce fait, l'environnement est perçu par la théorie de la contingence comme un facteur prépondérant affectant considérablement l'établissement et la réalisation des objectifs de la firme [29]. En ce sens, l'évolution de l'environnement détermine aussi bien les facteurs de contingence auxquels l'entreprise doit faire face [28], que la capacité d'apprentissage et de performance de cette dernière [30].

- **Les facteurs de contingence d'ordre organisationnel et financier**

En plus des effets de l'environnement sur la structure de l'organisation, d'autres variables d'ordre organisationnel [23] et financier [24] ont été identifiées comme ayant un impact sur cette dernière. Parmi ces variables, on trouve l'âge [31], la taille [32], le pouvoir [33], les conditions stratégiques [34], la technologie [35] et le système technique [19].

- **Les facteurs de contingence d'ordre humain**

Des facteurs d'ordre humain, relatifs aux différents acteurs, affectent l'organisation à différents niveaux [25]. On cite l'âge [31], les conditions socioculturelles [36], [37], la compétence en gestion [38], ainsi que l'éducation de base, la famille, l'entourage et la personnalité de l'acteur [20].

3.1.1.2 APPLICATION AU DOMAINE DE RECHERCHE COMPTABLE

La théorie de la contingence a été largement appliquée dans le domaine de la recherche en comptabilité, notamment pour l'établissement de liens de causalité entre certaines pratiques comptables et des variables de contingence d'ordre environnemental, organisationnel, financier et humain.

La référence [39] tente de déterminer les facteurs contingents ayant un impact sur la publication d'informations financières volontaires. Selon cet auteur, des facteurs tels que la visibilité internationale de la firme, la structure de son actionnariat, la culture de ses auditeurs, la culture et usages du pays où se trouve l'entité économique ainsi que son secteur d'activité, exercent une influence sur cette dernière de manière à accroître le contenu, le détail et la diffusion de ses informations financières.

D'autres études se sont concentrées sur l'impact de certains facteurs de contingence sur la qualité d'audit. Dans ce propos, les caractéristiques organisationnelles de l'entité auditée, à savoir la taille de l'entreprise, sa dispersion géographique, son statut de cotation et son âge, ont été identifiées comme les principaux déterminants de la demande d'une meilleure qualité de l'audit [40].

La référence [41] explique le choix de la politique de transition aux IFRS pour les entreprises françaises cotées par certains facteurs structurels de contingence, dont l'appartenance à un secteur d'activité. En effet, « les principaux ajustements financiers résultant du passage aux IFRS sont étroitement liés au secteur d'activité de l'entreprise, c'est-à-dire : l'immobilier, les technologies, les banques et assurances et l'industrie » [41]. Aussi, la structure financière de l'entreprise constitue un facteur qui influence ses choix comptables, et en l'occurrence, sa politique de transition aux IFRS. En effet, « les sociétés fortement endettées et/ou affichant un gearing défavorable privilégient des options et des choix comptables permettant la maximisation de leurs fonds propres, agrégats d'importance en matière d'analyse financière » [41]. Par exemple, le contenu informatif du résultat net est considéré dans [42] comme moins discriminant en référentiel IFRS (approche bilancielle) qu'en contexte français. Ceci est confirmé dans [43], [44], où la variabilité ou volatilité des résultats est potentiellement accrue en référentiel IFRS du fait notamment des effets de la juste valeur. Par ailleurs, les préoccupations fiscales constituent un autre déterminant de la politique comptable d'une entité [41]. Il s'agit en effet de vérifier, avant de procéder à des choix comptables, l'absence de risques fiscaux potentiels, et de maintenir autant que possible le principe de la connexité entre fiscalité et comptabilité, afin d'éviter les retraitements extracomptables. Nous soulignons enfin que les auditeurs financiers spécialisés en IFRS jouent un rôle considérable dans la facilitation du processus de transition aux IFRS, et ce en conseillant les préparateurs des comptes de l'entreprise sur les choix et options possibles dans le cadre du référentiel IFRS, et en envisageant des solutions en cas d'absence de traitement spécifique prévu par le référentiel IFRS [41].

Ainsi, le niveau d'assimilation de la culture comptable des IFRS, importée dans un pays donné, et véhiculée par l'harmonisation internationale, peut dépendre des contingences et des caractéristiques propres à chaque entreprise et à ses différents acteurs.

3.1.2 HARMONISATION DE FACTO VOLONTAIRE RESULTANT DE PRESSIONS INSTITUTIONNELLES

Les courants institutionnalistes, initiés par Commons et Selznick montrent le rôle important que jouent les institutions dans les diverses activités et les différents choix ayant trait à l'économie [45]. A l'instar de ces travaux, « l'approche néo-institutionnelle revendique l'importance des institutions pour expliquer les faits sociaux et économiques, tout en se focalisant sur une analyse inter-organisationnelle » [46]. Par ailleurs, s'intégrant dans le cadre de la théorie des organisations, et occupant une place importante dans cette dernière depuis la fin des années 1980 [47], l'approche néo-institutionnelle montre l'insuffisance de la recherche de la rationalité et de l'efficience dans la prise des décisions [48]. En effet, les acteurs

ne doivent pas seulement prendre leurs décisions sur la base des critères économiques, mais aussi sur la base des relations sociales, liant l'entité aux diverses institutions de son environnement. A ce propos, une pratique institutionnalisée est considérée dans [49] comme une pratique introduite pour des motifs autres que la rationalité économique. En effet, les outils de gestion sont considérés comme des pratiques institutionnalisées qui ne résultent pas uniquement de la rationalité économique, mais aussi de l'intervention d'industriels, de l'État, d'organisations patronales, etc. [49]. De ce fait, les choix de l'entreprise sont expliqués beaucoup plus par le processus social que par l'objectif de la maximisation des profits et des contraintes économiques [50]. Ainsi, Dans le cadre de l'approche néo-institutionnelle, la rationalité se base notamment sur la légitimité organisationnelle, dont la quête conditionne et influence les choix émis par les organisations [51].

3.1.2.1 LA LEGITIMITE ORGANISATIONNELLE COMME FONDEMENT DE LA RATIONALITE

L'approche néo-institutionnelle s'applique dans divers domaines de recherche, et est particulièrement pertinente pour analyser la diffusion et la légitimation de nouvelles pratiques dans les organisations [52]. Ainsi, le fondement de base de l'approche néo-institutionnelle est cette quête continue de légitimité des organisations. En effet, toujours à la recherche de soutien externe et de stabilité managériale, les organisations peuvent aller jusqu'à l'adoption de toute sorte d'éléments structurels, sans même en vérifier la compatibilité avec l'environnement et les structures en place [51]. Ainsi, ce courant de recherche s'intéresse particulièrement aux processus entrepris par les organisations pour apporter une légitimité à leurs actions [46]. En d'autres termes, les actions et les choix d'une entité économique ne sont pas seulement motivés par une recherche d'efficience mais aussi par une quête de légitimité de l'organisation aux yeux de ses diverses parties prenantes. L'originalité du cadre de l'approche néo-institutionnelle est donc perçue dans sa conception de la rationalité organisationnelle, fondée sur la quête d'adéquation avec les attentes de la société, en plus de la quête classique de maximisation des profits.

Par ailleurs, la notion de légitimité organisationnelle s'intègre dans une vision socio-politique des relations entre les entreprises et leur environnement, et est donc souvent rattachée à des contraintes environnementales leur imposant une conformité à leur contexte [46]. Ainsi, seules les organisations qui se soumettent à ces contraintes environnementales sont légitimes. On parle dans ce cas du respect, par ces organisations, d'un isomorphisme institutionnel [53].

3.1.2.2 L'ISOMORPHISME INSTITUTIONNEL DE DIMAGGIO ET POWELL (1983)

La référence [54] montre que la quête de rationalité organisationnelle ne se limite pas à l'adaptation des organisations à leur environnement, mais nécessite également leur acquisition d'une légitimité à partir de l'adoption et l'application des règles en vigueur dans l'environnement institutionnel, considérées comme des mythes « rationnels ». C'est suite à ces travaux que l'approche de l'isomorphisme institutionnel a été développée et proposée par [53].

La référence [55] décrit l'isomorphisme comme un processus qui contraint une unité appartenant à une population de ressembler à d'autres unités qui connaissent le même ensemble de conditions environnementales. L'isomorphisme est considéré dans [53] comme le concept qui traduit le mieux le procédé d'harmonisation. Dans ce sens, la théorie de l'isomorphisme institutionnel s'intéresse aux convergences des comportements et pratiques entre les structures organisationnelles. Il existe par ailleurs trois types de convergence ou d'isomorphisme institutionnel [53] :

- **L'isomorphisme coercitif provenant de l'influence politique et du problème de légitimité**

On parle d'isomorphisme coercitif lorsque des organisations ou institutions faisant partie de l'environnement exercent des pressions formelles et informelles sur l'entité, de manière à faire converger ses actions vers des pratiques harmonisées. Il est expliqué dans [53] que dans certains cas, le changement organisationnel constitue une réponse directe à une directive gouvernementale. Ceci étant, le fait que ces changements soient essentiellement formalistes, ne veut pas dire qu'ils sont sans importance. En effet, ces changements poussent les cadres à réagir, afin de préserver leurs fonctions, ce qui peut altérer les rapports de force au sein des organisations sur le long terme [56]. L'existence d'un environnement légal commun affecte donc plusieurs aspects des pratiques et de la structure d'une organisation. La référence [57] montre que les environnements politiquement construits présentent deux caractéristiques, à savoir : les décideurs politiques n'expérimentent souvent pas directement les conséquences de leurs actions ; et les décisions politiques sont appliquées dans tous les domaines à des classes entières d'organisations, ce qui rend de telles décisions moins adaptables et moins souples. Ainsi, les organisations sont de plus en plus homogènes dans les différents domaines, et de plus en plus organisées autour de rituels de conformité aux plus larges institutions [53]. L'imposition directe de procédures opérationnelles standard et de règles et structures légitimées se produit également en dehors de l'autorité gouvernementale. En effet, pour les groupes multinationaux, lorsque

ces derniers augmentent de taille et de portée, les critères « standard » de performance ne sont pas nécessairement imposés aux filiales, mais ces dernières sont souvent soumises à des mécanismes de reporting standardisés [58].

- **L'isomorphisme normatif associé à la professionnalisation**

Cette forme de convergence est fondée non plus sur des règles institutionnelles mais sur des normes plus ou moins reconnues. Ces dernières n'exercent pas de contraintes sur le comportement organisationnel mais sont utilisées en tant que critère de décision, pour orienter les pratiques des acteurs [46]. Plus la reconnaissance des normes est répandue à travers les organisations, plus ces normes acquièrent un pouvoir contraignant les autres organisations à y adhérer. La référence [53] se base sur d'autres travaux (par exemple [59], [60]) pour montrer que l'isomorphisme normatif provient principalement du phénomène de professionnalisation. Ce phénomène est défini comme « la lutte collective des membres d'une profession, pour définir les conditions et méthodes de travail, et établir ainsi une base cognitive et une légitimation de leur autonomie professionnelle » [53]. Ainsi, une grande importance dans la diffusion des pratiques organisationnelles est ainsi attribuée aux réseaux professionnels [53]. En outre, dans de nombreux cas, le pouvoir professionnel est d'autant plus approuvé par l'Etat lorsqu'il est créé par les activités des professions libérales [53]. Par ailleurs, un mécanisme important qui incite à l'isomorphisme normatif est le filtrage du personnel. Souvent, l'accès à certaines professions est tellement sélectif que l'on y retrouve des professionnels très peu différenciables. Ceci est particulièrement le cas des secteurs financiers ou de service, où la qualification est très importante [60]. En effet, dans ces domaines, les principaux cadres sont généralement issus des mêmes universités et sélectionnés selon un ensemble commun de critères. C'est pourquoi, ces derniers auront tendance à traiter les problèmes de manières similaires, à appliquer les mêmes politiques, procédures et structures considérées normativement certifiées et légitimées, et à adopter des approches décisionnelles globalement identiques [53]. En outre, l'échange d'informations entre les professionnels contribue à construire une hiérarchie communément reconnue, qui devient une matrice de flux d'information et de mouvements de personnel entre les organisations. De plus, la reconnaissance de quelques grandes entreprises dans un secteur donné en tant qu'agents de négociation clé dans le domaine, peut attribuer à ces entreprises une position centrale dans d'autres égards [53]. En l'occurrence, la reconnaissance par le gouvernement des entreprises ou organisations clés peut donner à ces dernières plus de légitimité et de visibilité et peut conduire les entreprises concurrentes à copier certains aspects de leurs structures ou procédures d'exploitation dans l'espoir d'obtenir les mêmes avantages. Le personnel des organisations très reconnues peut également avoir sa position renforcée, entre autres, par sa consultation par les organismes du gouvernement [61]. Les politiques et structures de ces organisations centrales servent donc de modèles, qui seront copiés à plus d'un niveau. Ainsi, les pratiques qui régissent une profession sont orientées par une quête de conformité avec les normes produites par les organisations ou entreprises leadeurs dans le domaine [46]. En effet, c'est la reproduction de ces normes qui engendrerait finalement une homogénéisation des structures organisationnelles [46].

- **L'isomorphisme mimétique résultant des réponses standards face à l'incertitude**

L'isomorphisme mimétique est observé lorsque les organisations, confrontées à des situations d'incertitude, cherchent à imiter d'autres modèles réussis, apparaissant comme les plus légitimes vis-à-vis de l'environnement [46]. En effet, lorsque les technologies organisationnelles ne sont pas maitrisées [62], lorsque les objectifs sont ambigus, ou encore lorsque l'environnement crée une incertitude considérable, les organisations sont amenées à copier le modèle d'autres organisations, optant ainsi pour une solution efficace et peu coûteuse [53]. Par ailleurs, il est affirmé dans [46] que ce mécanisme mimétique peut être observé même en l'absence d'incertitude. En effet, le fait qu'un nombre significatif d'organisations adopte une innovation, celle-ci devient admise et s'étend aux autres organisations qui la copient. La diffusion d'innovation semble donc être mieux expliquée par une tendance des organisations à l'isomorphisme mimétique que par une recherche de la performance économique. Par ailleurs, plus la population des clients ou du personnel d'une organisation est large, plus cette dernière sera incitée à l'isomorphisme mimétique [53]. C'est pourquoi, la présence massive de certains types de modèles structurels est souvent mieux expliquée par l'universalité des processus mimétiques que par la capacité des modèles adoptés à augmenter l'efficience de l'organisation. Parmi les exemples les plus frappants d'isomorphisme mimétique, on cite l'imitation organisationnelle étatique pratiquée par les pays émergents [63]. En effet, il est facile de prévoir l'organisation de l'administration d'un pays émergent, sans rien savoir de la nation elle-même, juste en se basant sur l'isomorphisme mimétique. Ces pays sont décrits comme « beaucoup plus isomorphes au niveau de la forme administrative et du modèle économique que ce que n'importe quelle théorie du système économique ou organisationnel mondial serait à même de prévoir » [63].

3.1.2.3 APPLICATION AU DOMAINE COMPTABLE

Les recherches comptables qui s'inscrivent dans le cadre la théorie néo-institutionnelle ont émergé autour des années 1990, avec des travaux portant sur l'ensemble du domaine comptable (en l'occurrence, [64] - [66]).

Par ailleurs, l'approche néo-institutionnelle permet d'expliquer les choix comptables des entités économiques et est particulièrement adaptée au domaine de la comptabilité internationale [67]. En effet, le processus d'harmonisation comptable internationale, entrainant l'adoption ou l'adaptation de nouvelles pratiques comptables et d'audit inspirées du référentiel international, correspond au cadre néo-institutionnel, en raison des pressions institutionnelles et des effets de modes inhérents à ce domaine [68]. En effet, constituant d'importantes règles en vigueur dans l'environnement institutionnel comptable international, les normes comptables internationales (IFRS) correspondraient à ces mythes « rationnels », définis dans [54] comme ceux dont l'adoption et l'application répond à la quête de rationalité organisationnelle et acquiert aux organisations qui y adhèrent une légitimité organisationnelle.

Par ailleurs, les entités qui se soumettent aux contraintes d'un contexte internationalisé, et s'imprègnent de ces mythes « rationnels » qui sont les IFRS, acquièrent une certaine légitimité, respectant un isomorphisme institutionnel [53]. D'ailleurs, l'isomorphisme est considéré dans [53] comme le concept qui traduit le mieux le procédé d'harmonisation (comptable pour notre cas), présentant une convergence des pratiques et règles comptables.

L'harmonisation comptable internationale peut correspondre, en fonction de la situation, à chacun des trois types de convergence ou d'isomorphisme institutionnel définis dans [53].

En effet, on parle d'**isomorphisme coercitif** dans le cas d'un pays qui impose le passage obligatoire, à partir d'une date donnée, aux IFRS. Egalement, l'isomorphisme coercitif concerne les groupes multinationaux qui augmentent de taille et de portée, et dont les filiales sont soumises à des mécanismes de reporting standardisés [58], et doivent, en l'occurrence, adopter des pratiques comptables compatibles avec les politiques de la société mère.

La référence [69] démontre l'importance de **l'isomorphisme normatif** dans le domaine comptable, en citant l'exemple de la profession comptable aux États-Unis d'Amérique, marquée par l'influence qu'exercent les associations professionnelles dans l'adoption des principes comptables (USGAAP) par certains États américains. Un deuxième exemple consiste en la réforme comptable de 1996 en Tunisie, qui a été menée, pour le compte du Conseil National de Comptabilité, par le cabinet international BIG Four, PricewaterhouseCoopers, capitalisant les bonnes pratiques en matière de communication financière et gérant une base de donnée systématiquement mise à jour en matière d'IFRS et de US GAAP. Ce cabinet avait par ailleurs publié un livre portant sur une comparaison entre les Normes Comptables Internationales (IFRS), Américaines (US GAAP) et Tunisiennes (NCT), préparé par son groupe de recherche Corporate Reporting Group (CRG). Cet exemple illustre le précepte expliqué dans [61], selon lequel la reconnaissance par le gouvernement des entreprises ou organisations clés peut donner à ces dernières plus de légitimité et de visibilité et peut conduire les entreprises concurrentes à copier certains aspects de leurs structures ou procédures d'exploitation dans l'espoir d'obtenir les mêmes avantages. Cette reconnaissance par le gouvernement est d'autant plus renforcée par la consultation du personnel de PricewaterhouseCoopers-Tunisie par un organisme du gouvernement tunisien, le Conseil National de Comptabilité. Ainsi, la référence [46] montre que les politiques et structures de ces organisations centrales (ici PricewaterhouseCoopers-Tunisie) servent de modèles, qui seront copiés à plus d'un niveau. Ainsi, pour cet exemple précis, les pratiques qui régissent la profession comptable sont orientées par une quête de conformité avec les normes produites ou adoptées par les cabinets leadeurs dans le domaine.

Enfin, parmi les exemples les plus frappants **d'isomorphisme mimétique**, l'imitation organisationnelle étatique pratiquée par les pays émergents est indiquée dans [63]. En effet, la référence [63] explique qu'il est facile de prévoir l'organisation de l'administration d'un pays émergent, sans rien savoir de la nation elle-même, juste en se basant sur l'isomorphisme mimétique. Ces pays émergents sont alors considérés comme « beaucoup plus isomorphes au niveau de la forme administrative et du modèle économique que ce que n'importe quelle théorie du système économique ou organisationnel mondial serait à même de prévoir » [63]. Dans ce sens, l'isomorphisme mimétique peut être perçu dans le domaine de l'harmonisation comptable internationale, lorsque les entités d'un pays, émergent ou non, intègrent les IFRS, ou du moins la culture IFRS, par imitation aux entités des pays développés. Là aussi, comme il est expliqué dans [63], le recours à ces normes est souvent mieux expliqué par l'universalité des processus mimétiques que par la capacité du référentiel international adopté à augmenter l'efficience de l'organisation.

Ceci étant, quel que soit l'isomorphisme constaté, comme le postule l'approche néo-institutionnelle, le souci de légitimité des organisations conduit souvent ces dernières à adopter des pratiques qui sont peu efficientes [68]. De ce fait, l'adoption ou l'adaptation des normes internationales par les entités économiques d'un pays donné peuvent ne pas être motivées par l'utilité intrinsèque de ces normes, en constituant plutôt une forme d'isomorphisme [53].

3.2 PERTINENCE DES IFRS POUR UN PAYS DONNÉ : HYPOTHÈSES, VARIABLES ET MODÈLES DE RECHERCHE

Au cours de cette démarche, il s'agit de tester l'impact potentiel de certains facteurs de contingence ainsi que du contexte social sur le niveau d'harmonisation de facto au sein d'une entité économique. Par ailleurs, portant sur l'harmonisation de facto, dans les cabinets d'expertise comptable et d'audit d'une part (2.2.1), et dans les entreprises commerciales et industrielles d'autre part (2.2.2), cette étude pose des hypothèses, des variables et un modèle explicatif d'analyse, propres à chacune de ces deux populations d'un pays donné.

3.2.1 PERTINENCE DE L'ADOPTION DES IFRS POUR LES CABINETS LOCAUX : HYPOTHESES, VARIABLES ET MODELE DE RECHERCHE

L'étude des facteurs qui favorisent l'harmonisation de facto au sein d'une entité ne pouvant pas être exhaustive, le choix des hypothèses et variables de recherche relatives à l'approche contingente et à l'approche néo-institutionnelle se basera sur la littérature existante et l'observation empirique. Aussi, seules les variables pour lesquelles il est possible de trouver une mesure pertinente ont été retenues.

3.2.1.1 HYPOTHESES ISSUES DE L'APPROCHE CONTINGENTE

Comme il a été précédemment expliqué (cf. section 3.1.1), la littérature existante identifie de nombreuses variables de contingence environnementales, financières, organisationnelles et humaines, ayant une influence sur diverses pratiques comptables, dont l'harmonisation de facto avec les pratiques préconisées par les IFRS.

Dans le cadre de cette recherche, les variables et hypothèses suivantes, issues de la théorie de la contingence, sont retenues :

Hypothèse 1 : Les caractéristiques propres au cabinet d'expertise comptable et d'audit ont un impact sur l'harmonisation de facto volontaire avec les IFRS.

Variable 1-1 : Chiffre d'affaires

Variable 1-2 : Effectif

Hypothèse 2 : Les caractéristiques propres au cadre exerçant au sein d'un cabinet de comptabilité et d'audit ont un impact sur son application de la culture des IFRS.

Variable 2-1 : Age

Variable 2-2 : Date d'obtention du diplôme

3.2.1.2 HYPOTHESES ISSUES DE L'APPROCHE NEO-INSTITUTIONNELLE

Comme il a été précédemment expliqué (cf. section 3.1.2), étant basée sur l'adoption par l'organisation de pratiques en réponse à des attentes qui lui sont externes [51], et très utilisée dans le domaine de la comptabilité, l'approche néo-institutionnelle est pertinente pour la détermination de la relation entre le niveau d'harmonisation de facto dans une entité, et certains facteurs institutionnels.

Issues de l'approche néo-institutionnelle, les variables et hypothèses suivantes, sont retenues :

Hypothèse 3 : La dimension internationale du cabinet favorise l'harmonisation de facto volontaire avec les IFRS.

Variable 3-1 : Appartenance un réseau international ou non

Variable 3-2 : Appartenance à un réseau Big 4 ou non

Hypothèse 4 : La dimension internationale des entreprises clientes favorise l'harmonisation de facto volontaire avec les IFRS.

Variable 4-1 : Part de dossiers de filiales de multinationales et/ou purement étrangères par rapport au nombre total des entreprises clientes du cabinet

Variable 4-2 : Part d'entreprises totalement ou partiellement exportatrices par rapport au nombre total des entreprises clientes du cabinet.

Ci-dessous le modèle explicatif d'analyse lié à la première population, avec les variables et hypothèses de recherche.

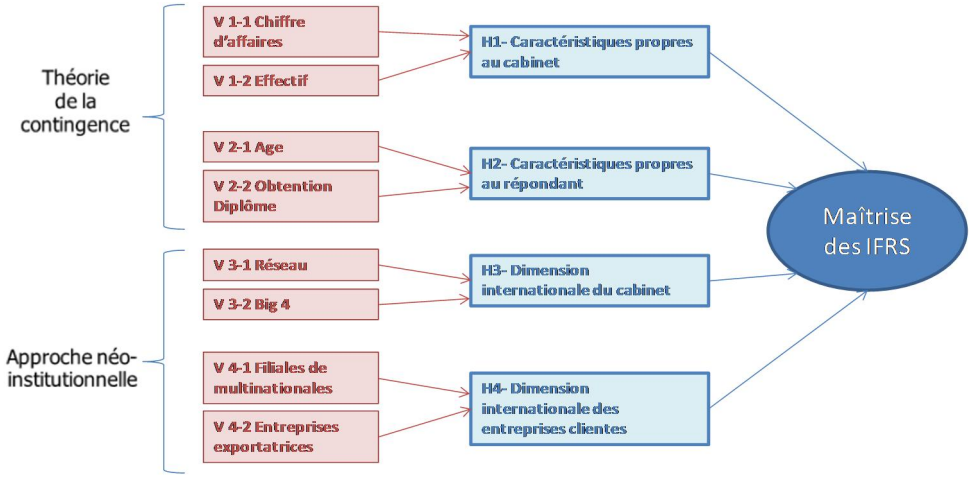


Fig. 1. Modèle explicatif d'analyse (population 1)

3.2.2 PERTINENCE DE L'ADOPTION DES IFRS POUR LES ENTREPRISES LOCALES : HYPOTHESES, VARIABLES ET MODELE DE RECHERCHE

Pour effectuer le choix des facteurs relatifs au niveau d'harmonisation de facto au sein des entreprises, et donc aux hypothèses de recherche relatives à l'approche contingente et à l'approche néo-institutionnelle, les déterminants cités dans la littérature ou observés empiriquement, relatifs à l'application volontaire des IFRS par une entité, sont utilisés. Aussi, nous avons sélectionné des variables pour lesquelles il est possible de trouver une mesure pertinente.

3.2.2.1 HYPOTHESES ISSUES DE L'APPROCHE CONTINGENTE

Comme il a été précédemment expliqué (cf. section 3.1.1), la littérature existante identifie plusieurs variables de contingence environnementales, financières, organisationnelles et humaines, ayant une influence sur diverses pratiques comptables, dont l'harmonisation de facto avec les pratiques préconisées par les IFRS.

Dans le cadre de cette recherche, ces variables et hypothèses, issues de la théorie de la contingence, sont retenues :

Hypothèse A : Les caractéristiques des entreprises ont un impact sur l'harmonisation de facto volontaire avec les IFRS.

Variable A-1 : Chiffre d'affaire

Variable A-2 : Effectif

Hypothèse B : Les caractéristiques propres au cadre comptable et/ou financier des entreprises ont un impact sur son application de la culture des IFRS.

Variable B-1 : Age

Variable B-2 : Date d'obtention du diplôme

3.2.2.2 HYPOTHESES ISSUES DE L'APPROCHE NEO-INSTITUTIONNELLE

Analogiquement aux hypothèses posées pour la première population constituée des cadres des cabinets d'expertise comptable et d'audit, l'approche néo-institutionnelle est également pertinente pour expliquer la relation qui existe entre les pratiques comptables harmonisées avec les IFRS des cadres comptables des entreprises, et certains facteurs institutionnels.

Ainsi, issues de l'approche néo-institutionnelle, les variables et hypothèses suivantes, sont retenues :

Hypothèse C : La dimension internationale de l'entreprise favorise l'harmonisation de facto volontaire avec les IFRS.

Variable C-1 : Filiale de multinationale ou non

Variable C-2 : Chiffre d'affaires réalisé à l'export

Hypothèse D : La dimension internationale du cabinet comptable et d'audit engagé par l'entreprise favorise l'harmonisation comptable de facto volontaire avec les IFRS.

Variable D-1 : Appartenance un réseau international ou non

Variable D-2 : Appartenance à un réseau Big 4 ou non

Ci-dessous, le modèle explicatif d'analyse lié à la deuxième population est présenté, mettant en évidence les variables et hypothèses de recherches.

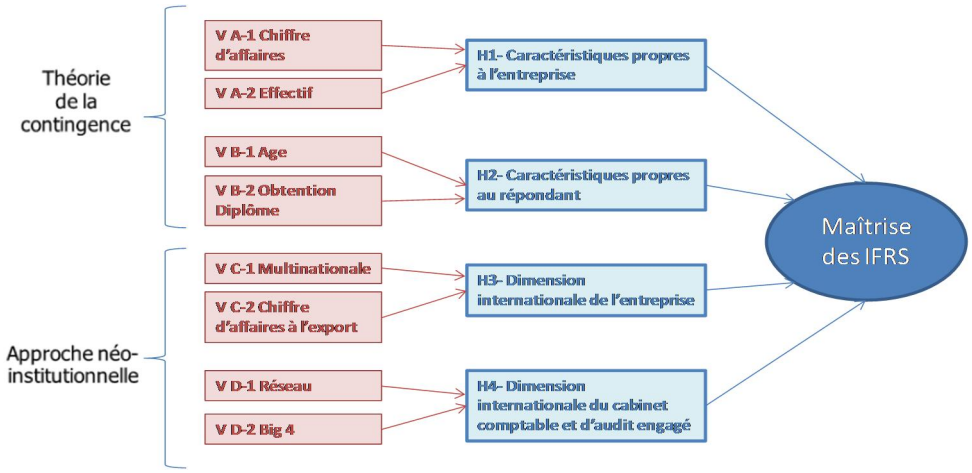


Fig. 2. Modèle explicatif d'analyse (population 2)

4 APPLICATION DU DISPOSITIF D'APPRÉCIATION A PRIORI DE LA PERTINENCE DES IFRS : CAS DE LA TUNISIE

En 1997, la Tunisie, un pays à tradition comptable continentale, a remplacé son Plan Comptable Général (1968), hérité de la période coloniale française, par le nouveau Système Comptable des Entreprises [70], un référentiel de type anglo-saxon, largement inspiré du référentiel comptable international (IAS[3]) de l'époque. Malgré cette réforme, selon le RRNC (Rapport sur le Respect des Normes et Codes) établi par une équipe de la banque mondiale le 23 juin 2004 [71], l'harmonisation comptable internationale de jure n'a pas été totale en Tunisie, entre autres parce que les nouvelles Normes Comptables Tunisiennes (NCT) sont « basées sur une version antérieure des normes IFRS qui était moins complexe que la version actuelle ». Ainsi, au fil des années, un gap s'est progressivement creusé entre normes internationales et normes tunisiennes.

Si la Tunisie avait réellement souhaité rattrapé le processus d'harmonisation comptable internationale depuis 1997, pourquoi fait-on face aujourd'hui à un référentiel tunisien non tenu à jour par rapport aux évolutions du référentiel international?

Cette situation **interroge la pertinence des IFRS pour le cas tunisien**, en laissant entendre que la réforme pour l'harmonisation internationale en Tunisie aurait davantage été induite par des pressions d'institutions étrangères, telles que la banque mondiale, le Fonds Monétaire International, ou même des entités économiques étrangères, telles que les sociétés appartenant à des groupes multinationaux ou les réseaux comptables et d'audit internationaux, plutôt que par un réel besoin et une profonde conviction émanant de l'intérieur du pays.

Dans ce qui suit, le dispositif d'appréciation de la pertinence des IFRS proposé est appliqué au cas de la Tunisie, afin d'établir le niveau général d'harmonisation comptable internationale de facto volontaire, et de l'expliquer par les facteurs contingents et/ou institutionnels précédemment définis.

4.1 POPULATIONS ET MÉTHODE DE RECHERCHE

Cette étude s'est déroulée de juin 2009 à septembre 2009, et de juin 2010 à septembre 2010. Elle s'intéresse à deux polulations distinctes ; les entreprises et les cabinets d'expertise comptable et d'audit, en Tunisie.

[3] *En 1997 les normes comptables internationales étaient connues sous l'appellation « IAS : International Accounting Standards ». Ce n'est qu'en 2001 que ces normes auront été dites « IFRS : International Financial Reporting Standards ».*

4.1.1 PREMIERE POPULATION : CADRES COMPTABLES ET/OU FINANCIERS DES ENTREPRISES

La première population mère comporte les cadres comptables et/ou financiers travaillant au sein d'entreprises ou groupes qui répondent aux conditions suivantes réunies :

- localisés en Tunisie,
- disposant d'un service (département) financier et/ou comptable (afin de garantir l'existence d'un interlocuteur informé au sein de l'entreprise),
- dont l'activité est commerciale et/ou industrielle, et ;
- dont l'effectif est supérieur ou égal à 10 salariés.

Pour des considérations de coûts, de faisabilité et de pertinence, cette étude s'est exclusivement focalisée sur les entreprises situées dans les gouvernorats de la région du Grand Tunis, à savoir : Tunis (capitale de la Tunisie), l'Ariana, la Mannouba et Benarous ; et dans le gouvernorat de Sfax ($2^{ème}$ grande ville de Tunisie). Ceci réduit la taille de la population à 2130 individus en 2011 [72]. Ce choix est justifié par la forte activité économique qui caractérise ces gouvernorats ainsi que par leur emplacement géographique qui en facilite l'accès. Pour le prélèvement de l'échantillon, la méthode des itinéraires a été empruntée, et a permis d'approcher 400 individus. A l'issue de la période de collecte de données, **147 réponses** ont été obtenues, traduisant un pourcentage de réponse de l'ordre de 37 %.

4.1.2 DEUXIEME POPULATION : CADRES DES CABINETS D'EXPERTISE COMPTABLE ET D'AUDIT

La deuxième population mère est constituée des personnes exerçant dans des cabinets d'expertise qui peuvent être :

- Des experts comptables inscrits à l'OECT (Ordre des Experts Comptables de Tunisie) : titulaires du certificat de la révision comptable (validation de trois modules), puis titulaires, suite à un stage de trois ans dans un cabinet d'expertise, et à l'élaboration et la soutenance d'un mémoire, du diplôme d'expertise comptable.
- Des experts-comptables mémorialistes : titulaires du certificat de la révision comptable (validation de trois modules), et en cours d'élaboration d'un mémoire après avoir validé un stage de trois ans dans un cabinet d'expertise.
- Des experts-comptables réviseurs (ou stagiaires) : titulaires du certificat de la révision comptable (validation de trois modules), et en cours de validation d'un stage de trois ans dans un cabinet d'expertise.

Ainsi, la deuxième population est composée de 438 experts comptables et de 450 réviseurs stagiaires et mémorialistes, ce qui donne un total de 888 individus en 2011 [73]. Pour le prélèvement de l'échantillon, toutes les adresses mail publiées sur le site de l'Ordre des Experts Comptables de Tunisie (personnes physiques et morales) ont été exploitées. Ainsi, nous estimons que la grille d'évaluation du niveau d'harmonisation de facto volontaire a pu atteindre environ 600 individus. A l'issue de la période de collecte de données, **112 réponses** émanant de cadres comptables, financiers et/ou auditeurs travaillant dans des cabinets d'expertise comptable, ont été obtenues, traduisant un taux de réponse de l'ordre de 19 %.

4.2 NIVEAU GENERAL D'HARMONISATION COMPTABLE INTERNATIONALE DE FACTO VOLONTAIRE

Après avoir adapté la grille de Deloitte au cas tunisien, un score d'harmonisation de facto volontaire avec les IFRS a été établi pour chacune des deux populations approchées ; les entreprises et les cabinets.

Les individus sont alors classés selon que le niveau du score obtenu soit "Très faible" (scores compris entre 0 et 8 inclus, "Faible" (scores compris entre 9 et 17 inclus), "Bon" (scores compris entre 18 et 26 inclus), ou "Très bon" (scores compris entre 19 et 36 inclus).

4.2.1 REPARTITION DES CADRES DES ENTREPRISES EN FONCTION DES SCORES D'HARMONISATION DE FACTO VOLONTAIRE

Pour les entreprises tunisiennes, le niveau général d'harmonisation de facto volontaire, calculé à partir de la grille de Deloitte adaptée est de **10,14 sur 36**.

La figure 3 montre la répartition des individus qui travaillent dans des entreprises tunisiennes, en fonction des scores d'harmonisation de facto volontaire obtenus par ces derniers. Il apparait que 43% des individus présentent un score de niveau "Faible", 41% de niveau "Très Faible", seulement 16% de niveau "Bon", et aucun individu appartenant à l'échantillon n'a obtenu un score de niveau "Très bon".

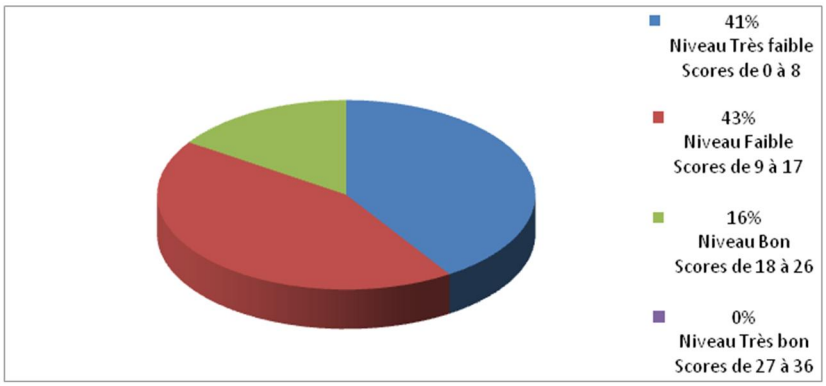


Fig. 3. Répartition des cadres des entreprises en fonction des scores d'harmonisation de facto volontaire obtenus

4.2.2 Repartition des cadres de cabinets en fonction des scores d'harmonisation de facto volontaire obtenus

Pour les cabinets tunisiens, le niveau général d'harmonisation de facto volontaire, calculé à partir de la grille de Deloitte adaptée est de **18,38 sur 36**.

La figure 4 présente la répartition des individus composant l'échantillon des cadres exerçant dans les cabinets tunisiens en fonction des scores obtenus. Il en ressort que 63% des individus présentent un score de niveau "Bon", 29% de niveau "Faible", et seulement 5% et 3% des individus ont respectivement obtenu des scores de niveau "Très faible" et "Très bon".

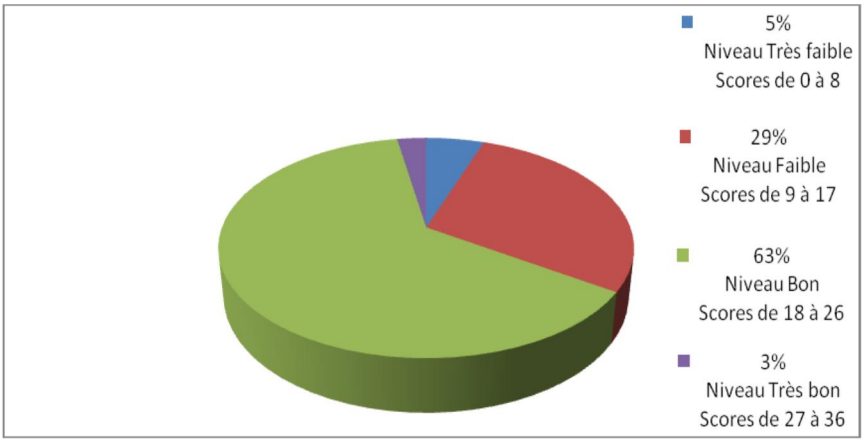


Fig. 4. Répartition des cadres de cabinets en fonction des scores d'harmonisation de facto volontaire obtenus

4.2.3 Un niveau moyen d'harmonisation de facto volontaire faible en Tunisie

Avec des scores moyens de 18,38 sur 36 et de 10,14 sur 36, respectivement pour l'échantillon des cabinets et celui des entreprises, et en respectant la pondération de ses résultats en fonction du nombre d'individus composant chacun des deux échantillons, le niveau moyen d'harmonisation de facto volontaire pour les entités tunisiennes concernées par l'étude, est estimé à **13,70 sur 36**, un score bien au dessous de la moyenne de 18 sur 36, traduisant donc une faible existence de pratiques comptables en Tunisie harmonisées avec les IFRS.

4.3 Contingence ou Pressions institutionnelles : Pertinence des IFRS pour le cas tunisien

Comme il a été précédemment établi, afin d'appliquer le dispositif d'appréciation de la pertinence des IFRS, il convient de vérifier l'existence ou non d'une relation significative linéaire, négative ou positive, entre la variable « score d'harmonisation de facto volontaire », et chacune des variables contingentes et institutionnelles sélectionnées par le dispositif, et ce pour chacune des deux populations étudiées (entreprises et cabinets). Cette relation est vérifiée par les corrélations de Pearson.

4.3.1 PERTINENCE DES IFRS POUR LE CAS TUNISIEN : RESULTATS RECUEILLIS AUPRES DES CADRES DES ENTREPRISES

En se référant au tableau des corrélations de Pearson présenté ci-dessous, il est établi que seulement quatre variables sur les huit étudiées sont identifiées comme significativement corrélées avec la variable « Score d'harmonisation de facto volontaire » (dans un ordre de significativité décroissant : « appartenance à un réseau international du cabinet engagé par l'entreprise », « appartenance de l'entreprise à une multinationale », « date d'obtention du diplôme du cadre comptable », « le fait que le cabinet engagé par l'entreprise soit un Big Four »). De ce fait, trois sur ces quatre variables significativement corrélées avec la variable « Score d'harmonisation de facto volontaire » sont issues de l'approche néo-institutionnelle.

Ainsi, il apparait qu'au sein des entreprises tunisiennes, les pratiques comptables volontairement harmonisées avec les IFRS, lorsqu'elles existent, sont dues à des pressions institutionnelles et non à un réel besoin local d'appliquer les IFRS.

Nous pouvons donc conclure que « **les normes IFRS ne sont pas pertinentes pour la Tunisie, au niveau des entreprises** ».

Table 5. Corrélations de Pearson entre toutes les variables et la variable « Score » : Population des entreprises

		CA	Effectif	Age	Date Diplome	Multi nationale	CA Export	Réseau	Big4	SCORE
SCORE	Corrélation de Pearson	,095	,151	-,155	,170*	,179*	,122	,195*	,167*	1
	Sig. (bilatérale)	,256	,068	,061	,047	,030	,141	,018	,043	
	N	146	146	147	137	147	147	147	147	147

*. La corrélation est significative au niveau 0.05 (bilatéral).

4.3.2 PERTINENCE DES IFRS POUR LE CAS TUNISIEN : RESULTATS RECUEILLIS AUPRES DES CADRES DES CABINETS

Le tableau des corrélations de Pearson présenté ci-dessous montre que les variables, contingentes et institutionnelles utilisées dans l'étude, sont toutes identifiées comme corrélées avec la variable « Score d'harmonisation de facto volontaire », avec un sens et une significativité de la relation linéaire différant d'une variable à l'autre. En effet, ce résultat ne précise pas la nature des facteurs dominants de cette relation, et ne permet donc pas de rattacher, précisément, l'harmonisation de facto volontaire en Tunisie à la contingence ou aux pressions institutionnelles.

Toutefois, les corrélations significatives établies entre la variable « Score d'harmonisation de facto volontaire » et les variables de contingence suffissent pour nous permettre de conclure que « **les normes IFRS sont pertinentes pour la Tunisie, au niveau des cabinets d'expertise comptable et d'audit** ».

Table 6. Corrélations de Pearson entre toutes les variables et la variable « Score » : Population des cabinets

		CA	Effectif	Date Diplôme	Age	Réseau	Big4	Clientes Filiales Multi.	Clientes Export.	SCORE
SCORE	Corrélation de Pearson	,218*	,297**	,324**	-,259**	,382**	,206*	,232*	,262**	1
	Sig. (bilatérale)	,021	,001	,000	,008	,000	,030	,022	,010	
	N	112	112	112	102	111	111	97	97	112

**. La corrélation est significative au niveau 0.01 (bilatéral).

*. La corrélation est significative au niveau 0.05 (bilatéral).

5 CONCLUSION

Cet article propose un dispositif d'appréciation de la pertinence des IFRS dans un contexte pré-implémentation. Il établit en effet les fondements de toute enquête à mener, interrogeant la pertinence des IFRS pour un pays n'ayant pas encore pris la décision d'adopter ces normes.

C'est ainsi que, pour tout pays désireux de se décider quant au passage ou non à l'harmonisation comptable internationale, un outil est proposé pour mesurer le niveau d'harmonisation volontaire des pratiques comptables locales avec les pratiques recommandées par les IFRS. Cet outil n'est autre qu'une adaptation de la grille d'évaluation de l'état de préparation à la conversion aux IFRS, élaborée par le cabinet canadien Samson Bélair, membre de Deloitte & Touche, et initialement destinée aux entités canadiennes [4].

Egalement, cet article pose les fondements théoriques susceptibles d'apprécier la pertinence des IFRS pour ledit pays, et ce en essayant d'expliquer le niveau évalué d'harmonisation internationale de facto volontaire dans le pays, par des facteurs contingents et/ou institutionnels. Les variables, hypothèses et modèles de recherche, propres à chacune de deux populations locales distinctes, les cabinets comptables et d'audit et les entreprises, sont ainsi posés à partir de la théorie de la contingence et de l'approche néo-institutionnelle.

La dernière partie de cet article présente une application directe du dispositif proposé d'appréciation de la pertinence des IFRS, en se basant sur le cas de la Tunisie, et montre que ce pays n'est pas encore prêt pour un passage imédiat aux IFRS. En effet, pour l'ensemble des entités économiques du pays, le niveau d'harmonisation de facto est plus dû à des pressions institutionnelles étrangères nécessitant l'affichage d'un « label IFRS » [3] qu'à un réel besoin local.

Les apports de cette recherche sont de nature théorique, méthodologique et pratique. En effet, sur le plan théorique, cette étude présente l'originalité de comparer deux théories complémentaires, la théorie de la contingence et l'approche néo-institutionnelle. Elle dévoile une insuffisance de la théorie de la contingence et confirme l'intérêt de l'approche néo-institutionnelle pour l'explication de l'harmonisation de facto dans un contexte où les IFRS ne sont pas encore adoptées. Les apports méthodologiques résident notamment dans la proposition d'une grille de lecture de l'évolution du processus d'harmonisation, dans un pays émergent. Au niveau pratique, à l'issue de cette étude, la pertinence des IFRS pour le cas tunisien est mise en question, notamment en ce qui concerne les entreprises tunisiennes dont les activités s'arrêtent au niveau local. Ceci met donc en avant les limites de ce référentiel qui se veut universel, et relativise l'intérêt qu'il présente en tant que vecteur d'un langage comptable commun. Par conséquent, face aux éventuelles lacunes du cadre comptable local d'un pays n'ayant pas encore adopté les IFRS, et pour lequel ces dernières ne s'avèrent pas pertinentes, nous préconisons aux régulateurs des mesures palliatives qui ne relèvent pas d'une adoption « forcée » des IFRS, mais éventuellement d'un système hybride, entre un modèle continental traditionnel, et un modèle anglo-saxon, qui serait adapté aux spécificités du pays et aux besoins des différents acteurs et utilisateurs.

Par ailleurs, bien que porteuse, cette recherche présente toutefois des limites, qui constituent autant de prolongements. En effet, la sélection des facteurs explicatifs de l'harmonisation de facto ne pouvant pas être exhaustive, il est indéniable que certains facteurs additionnels, notamment de contingence, mériteraient d'être intégrés à notre modèle d'analyse. En particulier, nous envisageons, pour des recherches futures, la considération de la structure du capital des entreprises, qui pourrait apporter un éclairage supplémentaire quant aux résultats obtenus. Aussi, présentant juste une application de démonstration du dispositif proposé, l'étude empirique s'est contentée de traitements statistiques ciblés. De ce fait, un prolongement consisterait à explorer, pour le même cas, d'autres méthodes de traitement plus approfondies. Par ailleurs, vu la dimension politique de la question de l'harmonisation, il serait intéressant de prolonger l'étude en recueillant et en analysant la vision des régulateurs comptables quant au sujet. Enfin, basée sur le cas d la Tunisie, il est possible de reprocher à l'étude empirique un horizon géographique réduit. Il serait donc intéressant de tester, dans des recherches futures, ce dispositif d'appréciation de la pertinence des IFRS proposé, sur d'autres pays où la question de l'adoption des IFRS est encore posée. Ainsi, au-delà de l'aide à la décision de l'adoption ou non des IFRS, tester le dispositif sur plusieurs pays permettrait d'établir une typologie ou une étude comparative des pays n'ayant pas encore adopté les IFRS.

REFERENCES

[1] H. Rey, « La Mondialisation Financière », *Conférence de l'Université de tous les savoirs*, Paris, 2003.

[2] M. Fouda, « Différences culturelles et relations d'affaires entre pays d'Afrique et pays émergents d'Asie », *Revue Française de Gestion*, vol. 32, n°. 167, pp. 65-84, 2006.

[3] H. Daske, L. Hail, C. Leuz, and R. Verdi, "Adopting a Label: Heterogeneity in the Economic Consequences of IFRS Adoptions", *30th congress of the European Accounting Association*, Lisbonne, 2007.

[4] Collectif, *Turning away from German, GAAP − Accounting Transition Experiences, from German GAAP to IFRS or US−GAAP*. Deloitte, 2004.

[5] R. Likert, "A Technique for the Measurement of Attitudes", *Archives of Psychology*, n°. 140, pp. 5-55, 1932.

[6] L. G. Van der Tas, "Evidence of EC Financial Reporting Practice Harmonization: The Case of Deferred Taxation", *European Accounting Review*, vol. 1, n°. 1, pp. 69-104, 1992.

[7] L.G. Van Der Tas, "Measuring International Harmonization and Standardization: A Comment", *Abacus*, vol. 28, n°. 2, pp. 211-216, 1992.

[8] L.G. Van der Tas, "Measuring harmonization of financial reporting practice", *Accounting and Business Research*, vol. 18, n°. 70, pp. 157-169, 1988.

[9] S. Archer, P. Delvaille, and S. McLeay, "A Statistical Model of International Accounting Harmonization", *Abacus*, vol. 32, n°. 1, pp. 1-29, 1996.

[10] S. Archer, P. Delvaille, and S. McLeay, "The Measurement of Harmonization and the Comparability of Financial Statement Items: Within-Country and Between-Country Effects", *Accounting and Business Research*, vol. 25, n°. 98, pp. 67-80, 1995

[11] Deloitte, [Online] Available: http://www.deloitte.com (2011)

[12] M. Crozier et E. Friedberg, *L'acteur et le système*. Éditions du Seuil, 1977.

[13] P. Bernoux, *La sociologie des organisations*, Éditions du Seuil, 1995.

[14] S. Marquet-Pondeville, *Le contrôle de gestion environnemental: une approche théorique contingente Et Une Étude Empirique Du Cas Des Entreprises Manufacturières Belges*. Presses Universitaires De Louvain, 2003.

[15] J. Woodward, *Management and Technology*. Her Majesty's Stationery Office, 1958.

[16] T. Burns and G.M. Stalker, *The Management of Innovation*, Tavistock Publications, 1961.

[17] R. Aïm, *L'essentiel de la théorie des organisations*. Editions Gualino Publication, 2006.

[18] P.R. Lawrence and J.W. Lorsch, *Organizations and environment*. Harvard University Press, 1967.

[19] H. Mintzberg, *Structure et dynamique des organisations*. Éditions d'organisation, 1982.

[20] Boukar H., « Les facteurs de contingence de la croissance des micro et petites entreprises camerounaises », *La Revue des Sciences de Gestion : Direction et Gestion des Entreprises*, vol. 44, n°. 237/238, pp. 75-83, 2009.

[21] D. Katz and R.L. Kahn, *The social psychology of organizations.* John Wiley & Sons Inc, 1966.

[22] P.Callot, « Intelligence Économique et PME : Le paradigme de l'intelligence perceptuelle », *La Revue des Sciences de Gestion : Direction et Gestion des Entreprises*, n°. 218, pp. 61-71, 2006.

[23] B. Mazouz, « Décider autrement, évoluer différemment : Une étude empirique sur la diversité des styles de décision et des trajectoires d'évolution des organisations », *La Revue des Sciences de Gestion : Direction et Gestion des Entreprises*, n°. 199, pp. 15-50, 2003.

[24] O. Colot et M. Croquet, « La contribution de variables propres aux PME et à leur dirigeant dans l'explication de la structure financière des PME », *La Revue des Sciences de Gestion : Direction et Gestion des Entreprises*, n° 228, pp. 61-72, 2007.

[25] J.R. Feudjo, « Homogénéité, hétérogénéité de l'équipe dirigeante, performance de l'entreprise et enracinement du dirigeant », *La Revue des Sciences de Gestion : Direction et Gestion des Entreprises*, n°. 219, pp. 115-128, 2006.

[26] B. Fabi, D.J. Garand, et N. Pettersen, « La GRH : contingences davantage qu'universalité ? Opérationnalisation d'un modèle de contingence », *4ème congrès de l'AGRH*, Jouy-en-Josas, 1993.

[27] D. Miller and P.H. Friesen, "Strategy-making environment: The third link", *Strategic Management Journal*, n°. 4, pp. 221-235, 1983.

[28] J.D. Thompson, *Organizations in action, social science bases of administrative theory*. Mc Graw-Hill, 1967.

[29] R.A. Thiétart, *Précis de direction*. Les Éditions d'organisation, 1980.

[30] S. Terreberry, "The evolution of organization environments", *Administrative Science Quarterly*, vol. 12, n°. 4, pp. 590-613, 1968.

[31] D. Bellemare, L. Poulin-Simon, et D.G. Tremblay, *Le paradoxe de l'âgisme dans une société vieillissante : enjeux et défis de gestion*. Éditions Saint-Martin, 1998.

[32] P.M. Blau, "A formal theory of differentiation in organizations", *American Sociological Review*, vol. 35, n°. 2, 1970.

[33] J. Child, "Organization structure, Environment and Performance, the role of strategic choice", *Sociology*, pp. 1-22, 1972.

[34] D. J. Hickson, C. R. Hinings, C. A. Lee, R. E. Schneck, and J. M. Pennings, "A Strategic Contingencies' Theory of Intraorganizational Power", *Administrative Science Quarterly*, vol. 16, n°. 2, pp. 216-229, 1971.

[35] J. Woodward, *Industrial Organization: Theory and Practice*. Oxford University Press, 1991.

[36] R. Pérez, « Quelques réflexions sur le management responsable, le développement durable et la responsabilité sociale de l'entreprise », *La Revue des Sciences de Gestion : Direction et Gestion des Entreprises*, n°. 211-212, pp. 29-46, 2005.

[37] M. Brasseur, « Le rôle des stéréotypes dans le management de la diversité culturelle : le cas de l'Afrique », *La Revue des Sciences de Gestion : Direction et Gestion des Entreprises*, n°. 230, pp. 61-67, 2008.

[38] G. Verna, « Croissance ou survie ? Quelle sera l'attitude la plus probable pour une petite entreprise légale du tiers-monde? », *Gestion 2000*, vol. 4, pp. 93-113, 1994.

[39] F. Pourtier, « La publication d'informations financières volontaires : un essai de synthèse », *Cahiers électroniques du CRECCI - IAE*, Cahier 08, 2004.

[40] P.A. Copley, J.J. Gaver, and K.M. Gaver, "Simultaneous estimation of the supply and demand of differentiated audits: evidence from the municipal audit market", *Journal of Accounting Research*, vol. 33, pp. 137-155, 1995.
[41] E. Tort, « La contingence de la politique comptable des sociétés cotées : le cas de la transition aux normes IFRS », *Comptabilité Contrôle Audit*, tome 13, vol. 3, pp. 171-194, 2007.
[42] R. Janin, « Les contenus informationnels respectifs du résultat net et des indicateurs de flux de trésorerie dans le contexte français », *Comptabilité Contrôle Audit*, tome 8, vol. 2, pp. 87- 107, 2002.
[43] T. Jeanjean, « Incitations et contraintes à la gestion de résultat », *Comptabilité Contrôle Audit*, Tome 7, vol. 1, pp. 61-76, 2001.
[44] Y. Mard, « Les sociétés françaises cotées gèrent-elles leurs chiffres comptables afin d'éviter les pertes et les baisses de résultats ? », *Comptabilité Contrôle Audit*, tome 10, vol. 2, pp. 73-98, 2004.
[45] D. Masson, *Femmes et politiques: l'État en mutation*. Les Presses de l'Université d'Ottawa, 2005.
[46] F. Déjean et M. Saboly, « La profession comptable française : entre mimétisme et rivalité ? », *Cahiers de recherche, IAE de Toulouse*, n°. 175, 2006.
[47] M.S. MIZRUCHI and L.C. FEIN, "The Social Construction of Organizational Knowledge: A Study of the Uses of Coercive, Mimetic and Normative Isomorphism", *Administrative Science Quarterly*, vol. 44, pp. 653-683, 1999.
[48] H. Moalla, « Les mécanismes de diffusion, d'adoption et de rejet de la méthode ABC dans l'environnement tunisien », *28ème congrès de l'Association Francophone de Comptabilité*, Poitiers, 2007.
[49] F. Le Roy et Y. Lemarchand, « L'introduction de la comptabilité analytique en France: de l'institutionnalisation d'une pratique de gestion », *Finance Contrôle Stratégie*, vol. 3, n°. 4, pp. 83-111, 2000.
[50] C. Oliver, "The Antecedents of Desinstitutionalization", *Organization Studies*, vol. 13, pp. 563-588, 1992.
[51] J.W. Meyer and B. Rowan, "Institutionalized organizations: formal structure as myth and ceremony", *American Journal of Sociology*, vol. 83, n°. 2, pp. 340-363, 1977.
[52] M.T. Dacin, J. Goodstein, and W.R. Scott, Institutional Theory and Institutional Change: Introduction to the Special Research Forum, *Academy of Management Journal*, vol. 45, n°. 1, pp. 45-57, 2002.
[53] P.J. DiMaggio and W.W. Powell, "The iron cage revisited: institutional isomorphism and collective rationality in organizational fields", *American Sociological Review*, vol. 48, n°. 2, pp. 147-160, 1983.
[54] Meyer, J.W. and Scott, W.R., *Centralization and the legitimacy problems of local government*, In: J.W. Meyer, and W.R. Scott (Eds.), *Organizational environments: ritual and rationality*, Beverly Hills: Sage Publications, pp. 199-215, 1983.
[55] A. Hawley, *Human ecology*, In: *International encyclopedia of the social sciences (Sills)*, New-York: Ed. Mcmillan, pp. 328-337, 1968.
[56] R.R. Ritti and F.H. Goldner, "Professional pluralism in an industrial organization", *Management Science*, vol. 16, n°. 4, pp. 233-246, 1969.
[57] J. Pfeffer and G. Salancik, *The External Control of Organizations: A Resource Dependence Perspective*. Harper & Row, 1978.
[58] L. Coser, C. Kadushin, and W.W. Powell, *The Culture and Commerce of Book Publishing*. Basic Books, 1982.
[59] M. S. Larson, *The Rise of Professionalism: A Sociological Analysis*. University of California Press, 1977.
[60] R. Collins, *The Credential Society*. Academic Press, 1979.
[61] M. Useem, "The social organization of the American business elite and participation of corporation directors in the governance of American institutions", *American Sociological Review*, vol. 44, pp. 553-572, 1979.
[62] J. G. March and J.P. Olsen, *Ambiguity and Choice in Organizations*. Universitetsforlaget,1976.
[63] M. Meyer, "Persistence and change in bureaucratic structures", 76th *meeting of the American Sociological Association*, Toronto, 1981.
[64] M.A. Covalesky and M.W. Dirsmith, "An Institutional Perspective on the Rise, Social Transformation and the Fall of a University Budget Category", *Administrative Science Quarterly*, vol. 33, n°. 4, pp. 562-587, 1988.
[65] T. J. Fogarty, "Financial Accounting Standard Setting as an Institutionalized Action Field: Constraints, Opportunities and Dilemmas", *Journal of Accounting and Public Policy*, vol. 11, issue 4, pp. 331-335, 1992.
[66] S. Mézias, "An institutionnal model of organizational practice: financial reporting at the fortune 200", *Administrative science quarterly*, vol. 15, n°. 1990, pp. 431-457, 1990.
[67] P. Touron, « Apports et limites de la théorie institutionnelle des organisations, Étude de trois cas d'adoption de normes comptables internationales en France », *21ème congrès de l'Association Francophone de Comptabilité*, Angers, 2000.
[68] O. Boiral, « La certification ISO 14001 : une perspective néo-institutionnelle », *Management International*, Vol. 10, n°. 3, pp. 67-79, 2006.
[69] V.L. Carpenter and E.H. Feroz, "Institutional theory and accounting rule choice: an analysis of four US state governments' decisions to adopt generally accepted accounting principles", *Accounting, Organizations and Society*, vol. 26, pp. 565-596, 2001.

[70] Loi N° 96-112 du 30 décembre 1996 relative au système comptable des enterprises, [Online] Available: http://www.cmf.org.tn/pdf/textes_ref/reglementations/Version_FR/sys_compt_entrep/loi_96-112_301296_fr.pdf (2011)
[71] World Bank, *Rapport sur le Respect des Normes et Codes, Report on the Observance of Standards and Codes*, 2004. [Online] Available: http://www.worldbank.org/ifa/rosc_aa_bdi_fre.pdf (2004)
[72] Agence de Promotion de l'Industrie et de l'Innovation, [Online] Available: http://www.investintunisia.tn (2011)
[73] Ordre des Experts Comptables de Tunisie, [Online] Available: http://www.oect.org.tn (2011)

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 651-661

http://www.issr-journals.org/ijias/

Using Geographic Information System to Maintenance and Upgrading Public Utility Networks Using Technology AM / FM

AUSAMA ALTADMORY

Department of PHD studies,
Technical University of Civil Engineering,
Bucharest, Romania

ABSTRACT: Many applications using Geographic Information Systems technology exist today. From facility mapping to inventory management, from land use analysis to trash collection routing, GIS technology is enhancing the ability of government agencies to provide services to its citizenry. Public utility networks comprise the most important and infrastructure in any city, state, or country. They provide essential support for running a society on uninterrupted daily basis, in most countries. The service area of a single utility company could vary from several hundred to many thousands square miles, and may serve anywhere from a few thousand to several millions customers. With many distributed facilities remotely located, utility systems become more sophisticated to manage and maintain. Traditionally, there are three major challenges facing the utility industry: large service areas, many distributed customers, and remotely distributed aging facilities. Recent deregulation and increasing market competition also have imposed more sophisticated difficulties. Utilities are seeking new technologies to tackle these challenges, and automated mapping/facilities management integrated with emerging technologies, provides vital solutions. (AM/FM) applications consist of water and wastewater, electricity, cable television, telephone, gas, telecommunication. The scope of this article is to find the best solution for managing the public utilities networks, and performing monitoring and maintenance process by combine automated maps which are useful for organizing data in layers overlay with facility management systems, which is responsible for the public utilities network management, reporting, maintenance. The combination of two systems called automated mapping/facilities management which will achieve a better improve, work-order, management, better integrating, and inventory control.

KEYWORDS: GIS, AM/FM, maintenance, management, map.

1 INTRODUCTION

Utilities have used 'intelligent computerized mapping' systems, also known as automated mapping/facilities management (AM/FM) or geographic information systems (GIS), which have been defined as "a geographic information system that integrates non-graphic facilities management information into a database that is tied to facilities maps". In most cases, the topological data structures employed by modern AM/FM software also include sufficient information to store and update connectivity relationships between facilities, thus simplifying utility network analysis applications.

Public utilities have many uses of AM/FM, most of which focus on the need to provide cost-effective services to consumers and better management of resources. Companies across utility industries have demonstrably accepted GIS as a necessary tool for their business. When one takes a look at the nature of information with witch utilities deal, it becomes clear that AM/FM can play a critical role for them.

Public utilities networks are unique in that they have distribution networks that must be maintained, and the locations and selected attribute information on such networks must be shared with other utility providers working in the same areas, and are grouped into two categories:

Corresponding Author: AUSAMA ALTADMORY (ausamat@hotmail.com)

- pipeline-based systems: gas, water, and wastewater.
- Cable-based systems: electricity, telecommunications, cable TV, and the Internet.

These two categories have historically been the most powerful users of AF/AM technology. Although the components of a public utility vary from one utility to another, with respect to facilities and services, utilities do have the following common characteristics:

- All utilities possess a physical network infrastructure with facilities/plants locally or remotely attached. This network must be maintained and its information (especially location data) shared with other utilities that share the same land use.
- They usually provide services in a similar way within a regulatory framework.
- They require similar geographic information to support their operations, including: (1) property/land use maps, (2) locations of pipelines and cable systems, (3) street/road networks, and (4) locations of other municipal facilities.
- They use similar spatial datasets and have similar work order management requirements with three basic processes: production, transmission, and service.
- Their routine operations are supported by the same types of spatial data management, analysis, and output functions, including: (1) load and network analysis, (2) records keeping and reporting, (3) facility mapping, (4) outage analysis, (5) maintenance and inventory, and (6) market analysis and customer service [1].

2 AM/FM/GIS BASICS

The term AM/FM/GIS mostly refers to GIS software that allows utility users to digitize, manage and analyze their utility network data. This data is stored in an underlying GIS database which also maintains the associations between the graphical entities and the attributes.

AM/FM/GIS is a combination of four separate systems:

- Automated Mapping (AM)
- Facilities Management (FM)
- Automated Mapping/Facilities Management (AM/FM)
- Geographic Information System (GIS)

Table 1. The differences between (AM/FM/GIS) systems

Feature	AM	FM	AM/FM	GIS	AM/FM/GIS
Layers	Y	N	Y	Y	Y
Topology	N	N	N	Y	Y
Network definition	N	Y/N	Y	Y	Y
Lines	Y	N	Y	Y	Y
Nodes	N	Y/N	Y	Y	Y
Polygons (areas)	N	N	N	Y	Y
Attributes	N	Y	Y	Y	Y
Actual locations	Y/N	N	Y	Y	Y
Map intelligence	N	N	N	Y	Y
Facilities management	N	Y	Y	N	Y

Notes: Y = Yes; N = No; and Y/N = Both Yes and No.

2.1 AUTOMATED MAPPING (AM)

Automated mapping (AM), also known as computer-aided mapping (CAM), is a CAD application for producing maps. It can be considered an alternative to the traditional manual cartographic maps. Data are organized in layers that are conceptually like registered film overlays. Layers organize data by theme (streams vs. roads) and type (linework vs. text). There are no spatial relations (topology) among data elements except orientation.

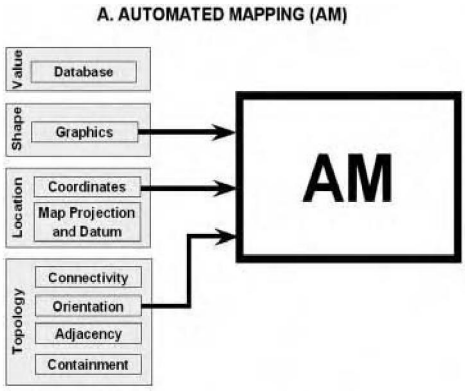


Fig. 1. AM capabilities

2.2 FACILITIES MANAGEMENT (FM)

Facilities management (FM), also referred to as asset management, is a CAD technology for managing utility system data. FM consists of such activities as inventory, inspection, and maintenance performed by cities, utilities, and government agencies. Organizations incur considerable expenses and resources as these functions are performed on a routine basis. FM includes an infrastructure management database.

Compared with AM, there is less emphasis on graphical detail or precision and more emphasis on data storage, analysis, and reporting. Relationships among utility system components are defined as networks. Because FM systems have network definitions, an FM system "knows" the pipes connected upstream or downstream of a given pipe. As shown in figure 2, FM systems generally do not have full topology; they offer connectivity and orientation only.

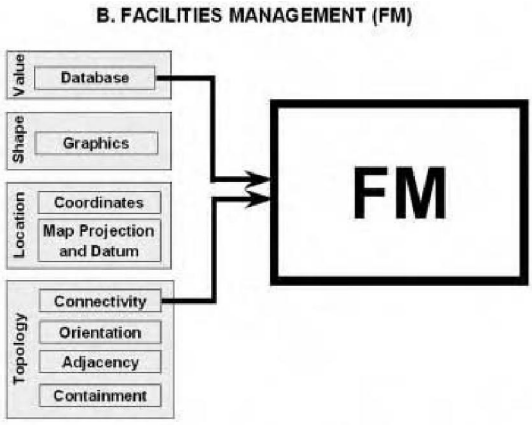


Fig. 2. FM capabilities

2.3 AUTOMATED MAPPING (AM)/FACILITIES MANAGEMENT (FM)

AM/FM is a combination of AM and FM technologies. AM/FM software is used to automate maintenance. It allows the integration and automation of maintenance management.

In the past, when a map was needed, a crew of surveyors, draftspersons, and geographers would combine their resources and develop a map on paper. This map was created by hand, updated by hand, and reproduced by a professional printer.

AM/FM systems have both orientation and network definitions as

shown in figure 3. The benefits of AM/FM are improved work-order management, better integrated inventory control, capture of maintenance data and costs, and allocation of costs [2].

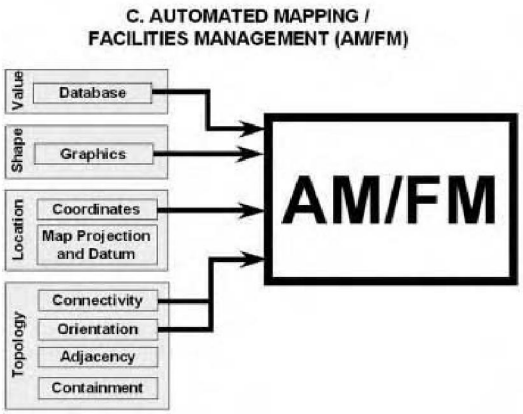


Fig. 3. AM/FM capabilities

AM/FM examples

- locating pole or facility item by street address
- generate reports about puplic utility netowrks
- generate maps of electrical circuits or feeders at prescribed scale
- produce continuing reports on property
- provide reports for tax purposes[3].

2.4 AM/FM/GIS SYSTEMS

GIS and AM/FM are different systems but each has its own advantages and applications. A GIS can help locate the worst pipe. An AM/FM can help users prioritize the work required to bring their worst pipes up to a minimum operating standard. For many years, people have used the GIS and AM/FM systems separately. Developing and maintaining two different systems is expensive and inefficient. Thanks to the latest advances in computer hardware and software, integrated AM/FM and GIS systems called AM/FM/GIS systems are now available. These systems are especially useful for asset inventory, inspection and maintenance, and work management. They are especially popular among visual learners who prefer maps over tables and databases. Visual learners like to click on a manhole in a GIS map to initiate a work order rather than locate it by querying a database. As shown in figure 4. AM/FM/GIS systems have more capabilities than AM/FM systems, such as orientation, network, database, and topology [2].

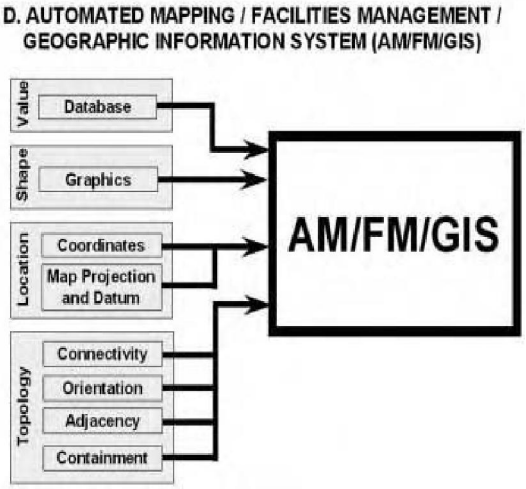


Fig. 4. AM/FM/GIS capabilities

2.5 AM/FM APPLICATIONS

AM/FM technology has been applied to many areas in public utilities. These include automated mapping and map maintenance, infrastructure sitting, records keeping and management, planning/routing, customer information service, network analysis, work order management, and decision making.

As integrated systems become more powerful and alternative spatial datasets become more widely available, more important value-added applications have been realized, including market analysis, decision support, and work order management. Now most utilities are focusing on reducing cost, attracting and keeping customers, and remaining competitive in the market. Under the continuing pressure of deregulation and distribution, utilities are even facing more challenges in using AM/Fm to support new application such as collaborative decision-making and distributed map services.

Spatial analysis is the most outstanding capability in AM/FM systems which support facilities design and planning, operations management, and customer service applications. Because the topological nature of utility networks (connectivity relationships in particular) may be modeled using fairly generic and well-developed tools and data structures, network analysis has become one of the workhorse applications' underline AM/FM usage in most utilities. The applications of AM/FM technology in the utility workplace are numerous. The next Table lists specific applications of spatial analysis capabilities in selected utility sectors.

The spatial analysis application areas common to all utilities are :

- Engineering planning and design
- Network planning, modelling, and simulation
- Trouble spots/plants identification
- Emergency/service crew dispatching
- Market and needs pattern analysis.

In addition, AM/FM systems have proven their principal worth with regard to basic mapping and inventory functions. However, information integration and modelling capabilities of AM/FM systems make them even useful for strategic planning and support of day-to-day routine activities. Strategic planning may include expansion of existing facilities, planning of new services, and sitting of new transmission lines.

The AM/FM technology permits the integration of base mapping, ownership information, property and political boundaries, and existing/proposed land use information to identify potential opportunities and constraints to development. Another important application area is work order management that includes issue work orders, dispatch service crews, schedule equipment (e.g. transmission and distribution equipment), inspection programs, and control workflow [51].

Scattered geographic service areas require work order management systems that identify costs related to specific activities by location. AM/FM helps provide the data necessary to evaluate efficient operations and integrate data into work orders and tasks. In addition, many tools within AM/FM systems provide routine operational support, to other real-time data acquisition applications, for trouble call/outage analysis, distribution automation, customer service, leak detection,

maintenance, automated meter reading, and SCADA. For example, there are three benefits that could be gained by using AM/FM to perform an outage analysis:

- Helping operators of a trouble-call system to quickly pinpoint the problems
- Providing valuable information online for dispatching trouble crews, such as site maps, routing, and maintenance records of the faulty facilities
- Providing vital data for the infrastructure replacements through cumulative analysis of outage records.

Table 2. Specific spatial analysis applications for various utilities

Utility	Spatial analysis applications
Electricity	Outage analysis (dealing with trouble calls) Transmission line sitting Load pattern and growth analysis Impact analysis in facilities sitting
Water/wastewater	Breakage and leakage diagnosing Water network flow analysis Modelling damage to water distribution systems Water resource planning and management Simulation of ground water mass destruction Prediction of runoff rates Determination of pressure zone when planning new water distribution facilities
Telephone/cable TV	Network/cable routing Facilities sitting and location optimization Outage and performance problem analysis Black spot/zone analysis in cable television
Telecommunications	Radio propagation and area coverage analysis Optimum antenna heights and locations using DTM information Optimal design of a broad band network layout Analysing tower coverage areas and service accessibility of a mobile communications network Network traffic analysis by combing demography.

3 AM/FM/GIS SOFTWARE

There are currently a variety of computer applications on the market, which were developed for specific components of facility management. In applying automation to these diverse functions, vendors have taken approaches that originate from two different technical directions: CAD or relational database management systems (RDBMS).

GIS-based facilities management requires AM/FM software extensions that can be run from within the GIS software. These add-on programs are also referred to as AM/FM/GIS software. Representative AM/FM/GIS software is listed in Table 3. According to this table, there are two types of AM/FM/GIS software:

1. AM focus: These software provide more mapping (AM) capabilities. ArcFM is an example of this type of software.
2. FM focus: These software provide more database (FM) capabilities. Cityworks is an example of this type of software.

- Software with an AM focus provide better data editing and mapping capabilities and generally require an FM add-on to provide work-order management functions.
- Software with an FM focus provide maintenance management functions (e.g., work-order management) but may lack the map maintenance functions.
- Projects that require both strong AM and FM capabilities may have to use two software products.

For example, ArcFM's strong suite of CAD-like map-editing capabilities can be supplemented by Cityworks' work-order management functions [4], [5]

Table 3. AM/FM/GIS Software Examples

Software	Company	Web Site
	AM Focus	
ArcFM	Miner and Miner	www.miner.com
ArcFM Water	ESRI	www.esri.com
FRAMME	Intergraph	www.intergraph.com
GeoWater and GeoWasteWater	MicroStation	www.bentley.com
	FM Focus	
Cityworks	Azteca Systems	www.azteca.com
GBA Water Master and Sewer Master	GBA Master Series	www.gbamasterseries.com
GeoPlan	Regional Planning Technologies	www.rpt.com
WATERview, SEWERview, STORMview	CarteGraph Systems	www.cartegraph.com
	Proprietary Systems	
CASS WORKS	RJN Group, Inc.	www.rjn.com
IMS-AV and IMS-ARC	Hansen Information Technologies	www.hansen.com

4 MAINTENANCE APPLICATIONS

GIS can be used to prepare inspection or maintenance work orders simply by clicking on a sewer pipe or manhole. This approach simply takes just a few minutes compared to the conventional method of locating and copying maps and typing the work order forms, which usually takes several hours.

4.1 ASSET MANAGEMENT

At present our water and wastewater infrastructure, especially in the older cities, is in critical stages of deterioration and has started to crumble. Nationally and internationally, aging water and wastewater infrastructure is imposing enormous costs on local communities.

For water and wastewater systems, asset management can be defined as managing infrastructure capital assets to minimize the total cost of owning and operating them while delivering the service levels customers. A typical asset management system has five components :

1. Facilities inventory: Describes each system element in an asset group. GIS can be very useful in completing this task.

2. Condition assessment: Classifies each asset according to its capability to perform the intended function.

3. Valuation: Assigns a financial value to inventoried assets consistent with Generally Accepted Accounting Principles (GAAP).

4. Operations, maintenance, repair, and replacement management: Arguably the heart of a management system, this component tracks and records data about work orders and customer complaints, issues and tracks preventive and predictive maintenance schedules, and generates crew assignments and work-site maps. GIS has extensive capabilities to fulfill this part.

5. Analysis and evaluation: Considered as the brains of an asset management system, this component prioritizes work effort, analyzes cost-effectiveness, and optimizes asset performance.

An asset management system helps predict the future condition of assets and major rehabilitation costs for planning purposes. An effective asset management system can reduce the cost of system operation and maintenance (O&M). Every successful maintenance program should begin with an accurate system map because it is difficult to maintain a system if the

users do not know where the water or sewer lines are. A well-constructed GIS should be used to create the system map. Historical maintenance data should also be linked with the GIS because it is difficult to schedule maintenance when you do not know the last time it was done.

4.2 WET WEATHER OVERFLOW MANAGEMENT APPLICATIONS

Management of wet weather overflows is a fertile field for GIS technology. By using geographic information in mapping, facilities management, and work order management, a wastewater system manager can develop a detailed capital improvement program or operations and maintenance plan for the collection system.

Broken and damaged sewers, laterals, and manholes usually contribute significant amounts of wet weather inflow and infiltration (I/I) to a wastewater collection system. This contribution often results in combined sewer overflows (CSO) from combined sewer systems and sanitary sewer overflows (SSO) from sanitary sewer systems. This systems require implementation of a Capacity, Management, Operations, and Maintenance (CMOM) program.

CMOM refers to the entities responsible for the administration and oversight of the sewer system and its associated staff (in either a municipal or industrial context); capacity evaluation, management, operation, and maintenance programs; equipment; and facilities. The owner and operator may be two different entities. For example, the owner may own the infrastructure and be responsible for its maintenance while it designates responsibility for the day to day operation of the system to the operator [6].

- CMOM requires the system owners/operators to identify and prioritize structural deficiencies and rehabilitation actions for each deficiency.
- CMOM requirements offer a dynamic system management framework that encourages evaluating and prioritizing efforts to identify and correct performance-limiting situations in a wastewater collection system.
- CMOM is a combination of planning tools and physical activities that help communities optimize the performance of their sewer systems.
- CMOM requirements mandate that the system owner/operator properly manage, operate, and maintain, at all times, all parts of the collection system.
- The owner/operator must provide adequate capacity to convey base flows and peak flows for all parts of the collection system.
- The owner/operator must provide adequate capacity to convey base flows and peak flows for all parts of the collection system.
- CMOM requirements include “maintaining a map”, which is the simplest application of GIS.

4.3 CCTV INSPECTION OF SEWERS

As sewer system networks age, the risk of deterioration, blockages, and collapses becomes a major concern. As a result, municipalities worldwide are taking proactive measures to improve performance levels of their sewer systems. Cleaning and inspecting sewer lines are essential to maintaining a properly functioning system; these activities further a community’s reinvestment into its wastewater infrastructure.

Inspection programs are required to determine current sewer conditions and to aid in planning a maintenance strategy. Ideally, sewer line inspections need to take place during low flow conditions. If the flow conditions can potentially overtop the camera, then the inspection should be performed during low flow times between midnight and 5 AM, or the sewer lines can be temporarily plugged to reduce the flow. Most sewer lines are inspected using one or more of the following techniques: *Closed-circuit television (CCTV), cameras, visual inspection, lamping inspection*.

Visual inspections are vital in fully understanding the condition of a sewer system. Visual inspections of manholes and pipelines are comprised of surface and internal inspections. Operators should pay specific attention to sunken areas in the groundcover above a sewer line and areas with ponding water. In addition, inspectors should thoroughly check the physical conditions of stream crossings, the conditions of manhole frames and covers or any exposed brickwork, and the visibility of manholes and other structures. For large sewer sewer inspection company, has deployed light-linebased and sonar-based equipment that measures the internal cross-sectional profile of sewer systems [7].

Fig. 5. TV inspection footage is hyperlinked for multimedia viewing within the mapping software

4.4 VIDEO MAPPING

Today, GIS map features can be linked to any kind of file to document inspection of a location through pictures, field notes, audio, streaming video, etc. When digital video, GPS, and GIS maps are combined, the possibilities are limited only by the user's imagination.

Geographic Information Systems are extremely useful for digitally mapping sanitary sewer defects. By locating defects that are spatially referenced to other GIS layers, one may make inferences about the cause of the defect, the type of repair, and an assessment of its repair priority and whether it is hazardous to public safety or the environment. Using field data collection computers, field crews are able to collect defect features with digital maps in the background. After locating the defect either manually according to background layers or by using GPS, the user attributes the defect with a predefined set of attributes. See figure 6 to view the application. Tabular attributes for defect features generally include the following:

- Date Tested
- Inspected By
- Ground Condition (wet, dry, moist)
- Precipitation Type (none, rain, snow)
- Leak Type (main line, manhole, public service line, private service line
- Leak Category
- Degree of Smoke (low, medium, high)
- Leak Cover
- Comments
- Defect Picture(s)

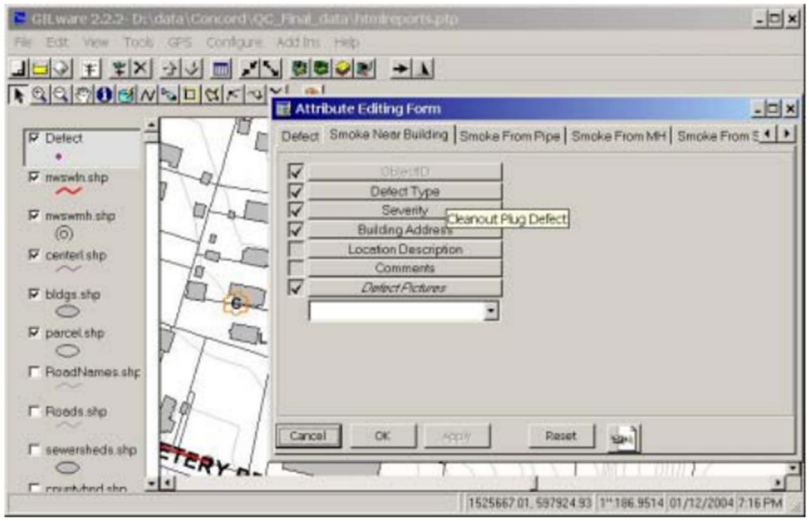


Fig. 6. A GIS application that allows field crews to attribute defects with both location and tabular data. Also, the application relates one or more digital pictures to the defect at the time of capture

Real-time mapping using GPS can accurately locate defects and sanitary sewer structures. By using GPS, field crews may locate defects more accurately to better relate them to a sanitary sewer layer or other layers. Field crews use GPS equipment with accuracy varying from sub-meter to 3 meter. It is important that defects be located with relatively high accuracy in order to be able to relate the defects to other features.

Digital picture documentation greatly helps one determine the type and magnitude of the defect. It also cues repair crews to what they can expect and the type of equipment and materials they will need to make the repairs. The application relates digital photographs to the feature at the moment the image is captured so as to maintain data integrity. Digital cameras are mounted onto the field computer. Additionally, pipe cameras (PipeCam) are special cameras mounted at the end of long poles that are rugged for entering sewer pipes and are equipped with their own lighting. All cameras feed live video to the GIS application where field crews may capture a single frame from the video stream to use as the digital documentation [8].

5 CONCLUSIONS

AM/FM/GIS Systems have been in existence long enough to have a proven track record for a variety of information management and facility management applications. From simple manual file and/or map consolidation uses for eliminating redundancy and increasing speed of reporting, to complicated infrastructure management programs for facility condition inventories, preventive maintenance programs, predictive modeling and real-time data from remote sensors, properly implemented and maintained AM/FM/GIS mapping and database management systems can improve an organizations quality of services delivered while reducing the costs of those services delivered. It is becoming easier for smaller organizations to take advantage of this information management technology due to the reduced cost of hardware and software, and the fact that a higher percentage of the work force is computer literate. Implementation costs can be minimized by sharing the costs of basemaps development with other organizations, and by taking a detailed look at what uses you actually want to implement through a full scale Pilot Study.

Public utilities have realized that AM/FM can be a valuable tool to improve efficiency and productivity. It is not only just for facility mapping, but also for better data integration, decision making, infrastructure management, work order management, and market/customer services. With the use of AM/FM, utility companies also have opportunities to change their industry images and strengthen their ability to compete in the market.

The AM/FM benefits do not come without a price. Typically, organizations incur high up-front costs associated the initial purchase of hardware and software. Further, any significant financial paybacks through saving and fees applications have been developed. Since neither may occur during the first 3-5 years of a program, the discounted B-C ratios for some periods may appear unfavorable to managers unfamiliar with the dynamics of AM/FM investments. As well, even with top-level support for the initiative, building consensus on data sharing at the corporate level can be very time-consuming and may deliver only limited results over the short term.

Even though GIS and AM/FM systems grew from different roots, their mission was the same to streamline map drafting and automate previously manual functions to achieve bottom-line savings. Older systems were implemented on a project or departmental basis. Today's systems tend to be interdepartmental with many different classes of users. The real payback in geospatial systems comes from the applications they provide in addition to the mapping of facilities. But access to spatial data is the key to achieving payback across the enterprise. True enterprise GIS for utility companies means access by virtually every employee through every corporate information system that deals with geographically dispersed assets or customers.

ACKNOWLEDGMENT

A special thanks for Professor, PhD, Eng. MOLDOVEANU Constantin, Technical University of Civil Engineering Bucharest for his effort on the scientific review of this article.

REFERENCES

[1] Said Easa and Yupo Chan, Urban Planning and Development Applications of GIS, American Society of Civil Engineers, 2000 / 304 pp.
[2] U.M.Shamsi, *GIS Applications for Water, Wastewater, and Stormwater Systems,* CRC Press, USA, 2005.
[3] Warren Ferguson, Ferguson Cartotech, and San Antonio, "NCGIA Core Curriculum in GIS," *National Center for Geographic Information and Analysis*, University of California, Santa Barbara CA, USA, 1997.
[4] Telvent, *Arcfm™ Solution Overview*, USA. [online] Available: http://www.telvent-gis.com/products/ (2013)
[5] Cityworks, "Water, Storm, & Wastewater," *Azteca Systems, Inc.*
[Online] Available: http://www.cityworks.com/products/industries/water-storm-wastewater/ (2013)
[6] Office of Enforcement and Compliance Assurance, *Guide for Evaluating CMOM Programs at Wastewater Collection Systems*, Environmental Protection Agency EPA, USA, 2005.
[7] Office of Water Washington, *Collection Systems O&M Fact Sheet: Sewer Cleaning and Inspection*, Environmental Protection Agency EPA, USA, 1999.
[8] Gil Inouye, *Using GIS and GPS for Sanitary Sewer I/I Detection,* Water Management Department, Woolpert LLP, USA, 2008.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 662-671

http://www.issr-journals.org/ijias/

Analysis of Psychological Well-being and Turnover intentions of Hotel Employees: An Empirical Study

Zulkarnain Amin and Kharissa Pratiwi Akbar

Department of Industrial and Organizational Psychology,
Faculty of Psychology, University of Sumatera Utara,
Medan, Indonesia

ABSTRACT: Turnover intentions can be reduced if the employees feel the organization have met their needs, through improving of psychological well-being. Employees' psychological well-being must be fulfilled, to get their best performance. This study aims to find out the relations between psychological well-being (i.e. autonomy, environmental mastery, positive relations with others, purpose in life, self-acceptance and personal growth) and turnover intention of hotel employees. The subjects were 212 hotel employees in Medan, Indonesia. Data were collected using turnover intention scale and psychological well-being scale. The result showed negative correlations between psychological well-being and turnover intentions towards hotel employees. To find out the determinants of turnover intention, a stepwise regression method was used. Based on the stepwise method used, the two predictor variables were found to be of significance in explaining turnover intentions. The two predictor variables were autonomy and positive relations with others. The result of this study will contribute to the guide lines for the policy makers for implementation of better human resource policy and provide evidence in favor of managerial interventions aimed at enhancing employee well-being and, consequently, minimizing the negative effects of an actual turnover in Hotel employee. Implication of this study could help management understand how to reduce the turnover intentions based on the employee and organization needs.

KEYWORDS: Hotel employee, turnover intentions, psychological well-being, cross sectional, human resource, Indonesia.

1 INTRODUCTION

The tourism industry today is probably the most important industry in the world. This industry has been making significant contributions to the economic development of many countries around the world [1]. The survey in 2010 showed that 7,002,944 tourists coming to Indonesia and increasing 10.74% than previous year [2]. The tremendous increase of tourist coming to Indonesia has resulted in the development of hotels in Indonesia.

Globalization and competitive business environment have encouraged service-oriented organizations such as hotel to be proactive in their human resources implementation in order to deliver high quality service to their customer [1]. Hotels fall into the service industry which represents a huge and growing percentage of the world economy; yet customer perception of service are not good. Given the economic growth in services, their profit potential, and overall decline in customer satisfaction with services, it is apparent that the potential and opportunities for establishments who can excel in services delivery have never been greater [3]. Such establishments can only be successful if there are service standards in place and a stable labor [4]. The hotel industry is offering customer service. It mainly adopts a shift-work system; the service offered to the customer emphasizes the harmony and cooperation between each department to achieve the service target [5]. Siao [6] also mentioned that whether a hotel is five-star level hotel is not the most important condition for tourism success. Instead, the important factor is whether the hotel can offer service with good quality. When customers feel good about service quality, the hotel will keep customers and bring further benefits to the enterprise [7].

Corresponding Author: Zulkarnain Amin (zulkarnain3@usu.ac.id)

Investment in employee success leads to employee satisfaction and reduced employee turnover, which in turn allows for the implementation of service standards which are essential for quality customer service. Quality customer service is essential to building customer relationship. Many hotels unfortunately give little priority to stabilizing their labor force, maintaining it preferably as a non-essential variable cost. Such an attitude naturally lays a low emphasis on labor incentives [8].

Numerous problems in the hotel industry have caused high employee turnover rate, making it hard for the enterprise to maintain talent. These problems include: (1) the salary of first-line employees are generally low and their work day is long; (2) they typically work on national holidays, Saturdays and Mondays; (3) the hotel system and welfare are not complete; (4) the ability to be promoted is minimal; and (4) hotels attach importance to service experience [5]. Moreover, employee turnover increases the possibility of losing excellent employees [5].

The increasing awareness of employees' turnover in the hotel industry has become a global phenomenon across the western world. Empirical studies have consistently showed that an excessive employee turnover rate is acknowledged to be one of the most problematic issues and particularly detrimental to the global hotel industry, which ranged about 60 per cent annually [9]-[11]. The traditional constraints such as long working hours, antisocial working hours, low pay, unstable, seasonal employment, low job status make employment within the hotel industry unattractive [1].

Factors affecting employee turnover today have become increasingly complex. The variables affecting job satisfaction are numerous and complex in relation to each other. Nevertheless, among all the possible factors affecting job satisfaction, most studies classify factors such as work environment and coworker relationships as the most influential factors affecting job satisfaction [12]. For the hotel employee, high employee turnover is a serious problem for human resource management at all times [13]. The tourist industry, however, it needs expend a great deal of human resources to offer service [5]. Further, the high employee turnover rate influences the service quality, which causes the cost of personnel, recruitment and training to increase, as well as causes great loss [14]. Therefore, a high turnover rate will induce absolute influence on the hotel's service quality [5].

Employee turnover has adverse consequences for effective organizational functioning. The time and energy devoted to find suitable new employees and the time required for new employees to reach maximum level of productivity may sometimes result in difficulties in achieving organizational objectives [15]. Employee turnover has been costly to organizations [16]. Recruiting, selecting, and training expenses are obvious costs [17]. Employee turnover represents a critical problem to an organization in terms of loss of talent, additional employment and training costs [18].

According to Dess and Shaw [19], indirect costs of turnover include reduction in morale among remaining staff, work overload, and loss of social capital. The impact of employee turnover intentions on organizational effectiveness and employee morale has remained the focus of organizational researchers in recent times [20]-[23]. In the review of antecedents to turnover, Barak, Nissly, & Levin [24] argued that many studies have used turnover intentions rather than actual turnover as the work outcomes because: 1) Employees typically make conscious decision before actually quitting their jobs and 2) It is more practical to inquire from employees their intention to quit in a cross sectional study rather than actually tracking them down via a longitudinal study to see if they have left their organization. For these reasons and consistent with previous researchers [1],[24],[25], turnover intentions is the criterion variable in this study since it has been recommended as a proxy in measuring actual turnover [26].

Cropanzano & Greenberg [27] explained the reasons that led to turnover intentions related to justice. This is the ways in determining whether their workers have been treated fairly in carrying out work activities. Furthermore, the employee intentions to leaving from organization because they feel organization can not provide well-being in the workplace [28]. The well-being of employees is in the best interest of communities and organizations [29]. Employees' well-being is the best interests of employers who spend substantial resources hiring employees and trying to generate products, profits, and maintain loyal customers. To succeed in hiring, employers must provide tangible benefits. However, employees want more than a stable job with pension and benefits. The average adult spends much of his or her life working, as much as a quarter or perhaps a third of his waking life in work. As much as a fifth to a quarter of the variation in adult life satisfaction can be accounted for by satisfaction with work. Employees who have a high level of well-being more cooperative, have a low level of absenteeism, timely and efficient, and can work longer at a company [28]. According to Ryff [30], when individuals have a good psychological wellbeing, they are able to function properly. Psychological well-being has been defined as ''engagement with existential challenges of life" [31].

Ryff and Keyes [32] proposed six dimensions of psychological well-being culled from an extensive literature review, including autonomy, environmental mastery, positive relations with others, purpose in life, self-acceptance and personal growth. Furthermore, Ryff [30] describes autonomy as the individuals' capabilty to regulate and evaluate their behavior.

Mastery of the environment is individual's ability to develop and control of complex environments. A positive relationship with others is positive interpersonal relationships with others, mutual trust, have a deep friendship and ability to identify the others as well. The purpose of life is assumed to have the belief that individuals can give meaning and direction to their life. Self-acceptance means the individuals hold a positive attitude toward their self. Personal growth implies individuals have a desire to continue to develop their potential, grow as individuals and to fully functioning.

In the workplace, employee will face the six dimensions of psychological well-being. This was confirmed by Sirgy, Reilly, Wu & Efraty [33] which agued that the workplace has become a social gathering, exchange ideas, meet and exchange experiences with colleagues. Its clearly that employee could not be separated from interpersonal relationships and it could affect their performance at work. Furthermore Cropanzano and Wright [34] also explained that a good of psychological well-being, a higher of work performance.

Psychological well-being is very important in the workplace to reduce the number of turnover intention or the intention to leave, especially in the hotel industry. In this study, we focused on the psychological well-being dimensions that can contribute to turnover intentions. It is little research has investigated the employees' psychological well-being and its influence on their behavioral intentions in the Indonesian hotel industry. Consistent with notion, a better understanding of the dimensions of psychological well-being which lead to hotel employee intentions to quit becomes increasingly relevant? Several studies have concluded that the psychological well-being can reduce job dissatisfaction, intentions to leave and increasing the role of the individual in the organization [16],[27],[35]. Furthermore, turnover also adversely affects the administration of the organization, and staff turnover impact on the financial and social conditions of the working environment [36]. The focus of this study was to analyze how psychological well-being dimensions can influence turnover intentions in hotel employee.

2 LITERATURE REVIEW

2.1 PSYCHOLOGICAL WELL-BEING AND TURNOVER INTENTION

Turnover intention is an early sign of the emergence of turnover behavior, and usually as last options for employee if workplace not relevant to what the expected. For employee, turnover is way out to get better, but for companies it can be a distinct disadvantage [37]. Study conducted by Pasewark and Strawser [38] showed that intention to leave refers to the intention of employee to seek other employment alternatives and has not manifested in actual behavior. The desire to move is closely related to employee job satisfaction and organizational commitment. Furthermore, Arnold and Feldman [39] concluded that the higher of job satisfaction correlated with the higher level of intention to stay in the company. According to Beach, Brereton and Cliff [16] turnover is not only detrimental to the company in terms of recruitment costs, training a replacement employee productivity and disruption of the company, but turnover also negatively impact social workers.

Furthermore, Vandenberg and Nelson [40] explained organizational commitment, job satisfaction, and well-being at work as a predictor of the emergence of turnover intentions. Employees who have fulfilled their well-being in the workplace are more productive, contributing to the organization's goals, and low of intentions to leave [28]. Samad [25] also stated the process of identifying factors influence of turnover intentions becomes important to be considered in turn lower of turnover rate in the company. Robertson and Cooper [41] suggested employee well-being contributed towards increase of productivity and organizational performance. It will increase service to customers, profitability, and reduced employee turnover and absenteeism.

Based on the description above, the hypothesis

H1: psychological well-being contributed in reducing turnover intentions

2.2 AUTONOMY AND TURNOVER INTENTION

The concept is often understood as self-determination, independence, and the self-regulation behavior. Resist social pressures to think and act incertain ways; evaluate self by personal standards [42]. Study by Ahuja, Chudoba, George, Kacmar & McKnight [43] found autonomy was negatively correlated to fatigue and fatigue was positively correlated with turnover intentions. This condition indicates that if workers lack the autonomy, their job performance tends to decline and feel like they unskills to work. Furthermore, Ahuja et al [43] explained autonomy influence the individual's perception of the workplace and it impacted to their behavior at work. Study by Ohly and Fritz [44] showed autonomy was positively correlated with proactive behavior. Workers were given the freedom to work, looking at the work as a challenge that motivated to complete their work. Gagne [45] describes a work condition can supports autonomy provided by low turnover rate

workplace. In addition, according to Price [26] autonomy and social support can reduce turnover through positive impact of job satisfaction.

Based on the above, the hypothesis

H2: Autonomy contributed in reducing turnover intentions

2.3 Environmental Mastery and Turnover Intention

Individual's ability to choose, create, and manage the environment to fit with psychological condition in order to develop themselves is part of the individual's ability to control the environment. Feel competent and create personally suitable contexts [30]. Individuals are also able to control the activity of using the opportunity of the environment effectively [31]. Horn, Taris, Schaufeli, and Schreurs [46] explained mastery of workplace can help workers to improve commitment. Workers who understand the work environment will help them on the job. By understanding the work environment, it also means workers must follow the policies and regulations of the organization, and show good performance although they have to work in less than ideal environments. The work environment has a positive impact on the performance of workers [47]. Work environment that does not conform to employee expectations is a good reason for employee to leave [48].
Based on the above, the hypothesis
H3: Environmental mastery contributed in reducing turnover intentions

2.4 Positive Relationships with Others and Turnover Intention

Individuals are described as mature individuals and able to foster interpersonal relationships built on mutual trust. Have warm, satisfying, trusting relationships; are concerned about others' welfare; capable of strong empathy, affection, and intimacy; understand give-and-take of human relationships. Individuals also have a strong feelings of sympathy and compassion towards human beings, has a deep friendship and ability to identify the others as well [30]. Robbins [12] stated the work would be an interaction both of colleagues and superiors. When individual have a good psychological wellbeing, they able to function properly. Thus, it would be optimal to do all the duties and responsibilities as individuals and has positive relationships with others. Interpersonal relationships affect to work performance. Strong group bond can form and maintain interpersonal relationships that sense of belonging in the group [49]. The relationship of this group can provide the commitment and job satisfaction, and could prevent negative predictors of intention to leave [35]. Furthermore, the study by Hasan and Subhani [50] found the relationship between superiors and subordinates correlated with turnover intentions. If employees have a positive relationship with co-workers then the employee will be loyal and committed to the organization [51]. According to Lee, Huang and Zhao [52], Workers relationships, salary level, and organizational commitment are important factors that influence employee turnover intention.

Based on the above, the hypothesis

H4: Positive relationship with others contributed in reducing turnover intentions

2.5 Purpose in Life and Turnover Intention

Individuals are assumed to have the confidence, can give meaning and direction to their life. Have goals and a sense of direction in life; present and past lives are meaningful; hold beliefs that give purpose to life. Individuals need a clear understanding of the purpose and direction of life [30]. Thus, it was concluded that the positive function of this dimension is to have a goal, purpose, and direction, all of which contribute to the feeling that life is meaningful. Satisfaction with one's life implies satisfaction with one's life circumstances or fulfillment of one's desires and needs are consistent with the overall purpose of life [53]. Workers' life satisfaction is also related to work behaviors such as performance, passion and desire to get out of work [54],[11]. If workers feel the work is not consistent with the goals of life, they become excited and tend to look for another job that suits their purpose.

Based on the above, the hypothesis

H5: Purpose in life contributed in reducing turnover intentions

2.6 Self Acceptance and Turnover Intention

Ryff and Singer [42] stated individuals should be able to understand themselves and feel positive. Possess positive attitude toward the self; acknowledge and accept multiple aspects of self; feel positive about past life. Individuals who have a

positive self-acceptance will recognize and accept both strengths and weaknesses. When they feel good about themselves, and realize they also have a weakness, it shows a high level of self-acceptance [55]. Employees who have good psychological well-being will be accepted themselves, low level of absenteeism, and can work longer at the organization [28]. If employees behave according to the values of the organization, it will ultimately bring benefits to the organization. Study by Milliman, Czaplewski, and Ferguson [56] found individuals can understand themselves will be more eager to work and consequently they have no intentions to get out of the organization.

Based on the description, the hypothesis

H6: Self acceptance contributed in reducing turnover intentions

2.7 Personal Growth and Turnover Intention

Ryff [30] describes the personal growth as individual feelings of continued development and potential and open to new experiences, an improvement in behavior, and change self-knowledge effectivelly. Price [26] suggests the opportunity for promotion and support of supervisors to remind potential to reduce worker turnover intention through the positive impact of employee commitment to the organization. Moncarz, Zhao & Kay [57] showed that chance of promotion, training and development of employees directly affected the decrease in employee turnover. Furthermore, Yin and Yang [58] revealed that internal environmental factors such as job stress, leadership style, superior and subordinate relationships, opportunities for advancement and administrative policies are significantly related to organizational turnover intentions. Study by Shields and Ward [59] showed dissatisfaction with promotion and training opportunities, have a strong impact on turnover than workload or salary. So, if the organization fails to provide the opportunity for employees to develop, they will quit [60].

Based on the above, the hypothesis

H7: Personal growth contributed in reducing turnover intentions

3 Material and Method

3.1 Participants

The sample of this study was the full time of hotel employees in Medan, Indonesia. Two hundred and forty seven questionnaires (including scales of psychological well-being and turnover intentions) were distributed among employees. The two hundred and twelve questionnaires of the employees returned the questionnaires. In this regard response rate was 86%.

3.2 Instruments

3.2.1 Psychological Well-Being

The instrument was the psychological well-being scale adopted from Ryff and Keyes [32]. The subjects of this study were requested to respond using five-point scaled response options ranging from strongly disagree (1) to strongly agree (5). This scale has six components, including autonomy, environmental mastery, positive relations with others, purpose in life, self-acceptance and personal growth. Higher scores will signify elevated degree of psychological well-being. This scale consists of 33 items and the Alpha Cronbach coefficient of reliability is 0.940.

3.2.2 Turnover Intentions

The instrument was turnover Intentions scale and consisited some aspects of intentions (i.e, attitude towards the behavior, subjective norms, and perceived behavioral control) suggested by Azjen [61]. The subjects of this study were requested to respond using five-point scaled response options ranging from strongly disagree (1) to strongly agree (5). Higher scores will signify elevated degree of turnover intentions. This scale consists of 24 items and the Alpha Cronbach coefficient of reliability is 0.880.

3.3 Data Analysis

This study examined the relationship between psychological well-being (i.e, autonomy, environmental mastery, positive relations with others, purpose in life, self-acceptance and personal growth) and turnover intentions. Furthermore, to find out the determinants of turnover intention, a stepwise regression method was used.

3.4 RESULT

The statistical analyses were computed using SPSS version 17.0 for Windows to extract the data and results. There were several results from the study. Pearson correlation was used to measure correlation between psychogical well-being (autonomy, environmental mastery, positive relations with others, purpose in life, self-acceptance, and personal growth) and turnover intentions. It showed each of the dimensions of psychological well-being correlated with turnover intentions. The results seen in table 1. To find out the determinants of turnover intentions, a stepwise regression method was used. Based on the stepwise method used, the two predictor variables were found to be of significance in explaining turnover intentions. The results seen in Table.2. The two predictor variables were autonomy and positive relations with others. The R-squared was 0.384; it implied that the two predictor variables explain about 38.4 of the variance in turnover intentions. This is quite a respectable result. As depicted in the coefficients table (Table 2), the estimates of the model coefficients for $\beta 0$ is 117.034, β_1 is -1.899 and β_2 is -.860. Therefore, the estimated model is as below:

Y (Turnover intentions) = 117.034 -1.899 (X1) - .860. (X2) + *e*
Where:
X 1 = autonomy
X 2 = positive relations with others

Table 1. Correlations among variables

Variables	Mean	SD	1	2	3	4	5	6	7	8
1. Turnover intention	68.377	12.356								
2. Psychological well-being	117.825	17.814	-.539**							
3. Autonomy	17.193	2.683	-.560**	.773**						
4. Environmental mastery	21.518	5.047	-.324**	.765**	.459**					
5. Positive relations with others	21.080	3.953	-.507**	.865**	.677**	.559**				
6. Purpose in life	18.617	3.403	-.493**	.825**	.622**	.480**	.663**			
7. Self-acceptance	14.839	2.569	-.426**	.811**	.575**	.474**	.682**	.686**		
8. Personal growth	21.033	3.552	-.385**	838**	.575**	.530**	.529**	.537**	.686**	

*Notes: ** p<0.01*

Table 2. Summary of estimates of coefficients for the model

	B (Unstandardized Coefficients)	**Std. Error**	**Beta (Standardized Coefficients)**	**F**	**t**
Constant	117.034	4.665		55.656	25.086**
Autonomy	-1.899	.328	-.412		5.782**
Positive relations with others	-.860	.259	-.237		3.319**

*Notes: ** p<0.01, R = 0.589; R^2 = 0.384*

4 DISCUSSION

The purpose of this study is to analyze the influence of psychological well-being to turnover intentions in hotel employee. The results showed the six dimensions of psychological well-being showed a negative correlation with turnover intentions. It indicates that employees have job satisfaction and met well-being in their workplace, are more productive, contributing to the organization's goals and objectives, and low desire to quit [28]. From this study, it could be said that if the employees feel contented in what they receive compared to those of a referent other, they are more likely to be productive and remain with the organization [62]. Indeed, employees who perceived fair prospects of different work outcomes particularly with regards to pay satisfaction, work schedule, workload, rewards, and job responsibilities may viewed their current organization as favorable and might not see outside alternatives as attractive. Thus, the result implied that hotel employee who perceive a higher level of psychological well-being are less likely to have the intention to leave their current organization. It should be noted that hotel employees put great emphasis on psychological well-being, which in turn significantly affects their satisfaction and emotional bond toward their employing hotel. This indicates that employees who have job satisfaction and

met well-being in their workplace is more productive, contributing to the organization's goals and objectives, and low desire to quit [28]. The result of this study supports the proposition that well-being in the workplace is, in part, a function of helping employees do what is naturally right for them by freeing them up to do through behaviors. Well-being perspective is quite applicable to business and that, as managers and employees focus on satisfying basic human needs in the workplace, clarifying desired outcomes and increasing opportunity for individual fulfillment and growth. The employees may increase the opportunity for the success of their organization [29].

Based on the results of stepwise regression analysis showed the six dimensions of psychological well-being correlated to turnover intentions. In this study, only autonomy and positive relations with other people are predictor of turnover intentions. These results can be explained by Ryff [30] which declared autonomy as maintaining individuality in a social system, individual self-determination, can regulate their behavior, and the ability to withstand social pressures. Autonomy was positively correlated with proactive behavior. Workers were given the freedom to work, looking at the work as a challenge that motivated to complete their work [44]. This finding cosistent with Ahuja et al study's [43]. They explained autonomy influence the individual's perception of the workplace and it impacted to their behavior at work. The empirical results of this study also suggest that positive relations with others significantly negative effects on turnover intentions. Good coworker relationships can help employees address work challenges and enable them to be dedicated to details that affect service quality, enhance service quality stability, and improve job satisfaction [63],[64]. If hotels management can provide a better work environment and facilitate harmonious coworker relationships among employees, job satisfaction can be effectively improved. Job satisfaction indirectly affect the employee turnover rate.

5 CONCLUSIONS

This finding suggests that organizations that provide employees with opportunities to extend their potential and build up their capabilities, which help meet the employees' needs for personal growth, are likely to be perceived as supportive and caring about the employees' well-being. The empirical results of this study suggest that autonomy and coworker significantly positive effects on job satisfaction. If hotels can provide a better work environment and facilitate harmonious coworker relationships among employees, job satisfaction can be effectively improved. Job satisfaction indirectly affect turnover intention through organizational commitment, employees will begin to identify with the company, its goals and business strategies, and their jobs if the company can effectively enhance their organizational commitment. Employees should also be given the opportunity to perform work that requires thought and independent judgment. Employees will regard their work as meaningful when a degree of independence and freedom of choice is allowed in the performance of their tasks. Participation in decision making will enhance the employees' feeling of membership and contribute to the meaningfulness component of sense of coherence. Moreover, the employee should have the freedom to disagree with his/her supervisor, to be able to discuss what to do with his/her supervisor and to act autonomously.

6 IMPLICATION

From the management view, the findings suggest that the relevant hotel authorities concerned with reducing turnover intentions among their employees should be consistent and focus their attention on providing interpersonal treatment that fosters a positive employee-organization relationship. It is hoped that the research findings may offer some insights into hotel organization in Indonesia to retain their valuable employee and reduce turnover intentions. The findings imply that managers can significantly reduce turnover intentions among their competent staff by adopting management styles that increase the feeling that the organization values staff contributions and cares about their well-being. It is worth mentioning that psychological well-being is a vital role in predicting employees' turnover intentions as such autonomy and positive relations with others, in turn, will reduce employees' likehood of quitting the organization. The results confirm that, if workers are to be motivated to grow, and if they are to be affectively tied to the organization and to their organizational roles, the employer must endeavor to establish employee psychological well-being. Understanding the dynamics whereby the psychological well-being arises and is maintained in the worker's mind is an importance for management. If the organization fails to understand and adequately fulfill of the psychological well-being, negative consequences—such as perceptions of injustice or role ambiguity, may ensue for the employment relationship.

7 LIMITATIONS

The present study has the following limitations. Hotel employees are the main subjects in this study. However, it is debatable whether employees of other sectors may face the same problems and difficulties that result in turnover behaviors. Hence, future studies can investigate employees of different service sectors to discuss the differences in the turnover

models. This study used only the cross-section study data as the empirical study basis and did not probe into the interactions of different variables in time; hence, the causal inferences are limited. Therefore, the findings obtained may not be generalized to other samples within the industry. The research was cross-sectional (all the data were collected at the same time), which means that it is not possible to draw conclusions about cause and effect among items based on employee responses. Future longitudinal research is therefore needed to confirm the conclusions drawn by this study. Another limitation is the exclusive use of self-report measures, a strategy often associated with method variance. Via a larger sample from other organizations in the same industry vice versa would improve the consequential of the findings. The second limitation were the availability of data or resources, some of the relevant data to support the research may not be available due to the nature of data or information that desired to be obtained is basically not to offend the companies' confidentiality or policies that remained secretive. In this study, the company refused to expose the data regarding the opportunities for promotion.

REFERENCES

[1] Alan, S.Z.R, Radzi, M.S and Hemdi, M.A, "An empirical assessment of hotel managers turnover intentions: The impact of organizational justice". *Journal of Tourism, Hospitality & Culinary Arts*, pp.1-29, 2009.

[2] Ministry of Tourism and Creative Economy. Strategic plan of directorates of tourism destination development 2012-2014. Jakarta: Directorate General of Tourism Destination Development, 2012.

[3] Zeithaml, V.A., Bitner, M.J and Gremler, D.D, "Services marketing: Integrating customer focus across the firm*".* 4th Edition. New York: McGraw-Hill, 2006.

[4] Kimungu, S.I and Maringa, P.M, "An assessment of the impact of employee turnover on customer service and competitiveness of an establishment". *International Research Symposium in Service Management*, pp. 1-16, 2010

[5] Chang, C.P and Chang, W.C, "Internal marketing practices and employees' turnover intentions in tourism and leisure hotels". *The Journal of Human Resource and Adult Learning* vol. 4, no. 2,pp. 161-172, 2008.

[6] Siao, S.J, "Tourist resort hotels: Super unexpected rival of service industry". *Career*, no.360, pp.112-113, 2006.

[7] Deshpande, R., Farley, J and Webster, F, "Corporate culture, customer orientation, and innovativeness in Japanese firms: A quadrad analysis". *Journal of Marketing*, vol.57, no.1, pp.23-37, 1993.

[8] Torrington, D and Hall, L, "*Personnel management*: *Human resource management in action"*. Europe: Prentice-Hall International, 1995.

[9] Birdir, K, "General manager turnover and root causes". *International Journal of Contemporary Hospitality Management*, vol.14, no.1, pp. 43-47, 2002.

[10] Carbery, R., Garavan, T. N., O' Brien, F and McDonnell, J, "Predicting hotel managers' turnover cognitions". *Journal of Managerial Psychology,* vol.18, no.7, pp. 649-679, 2003.

[11] Ghiselli, R.F., La Lopa, J.M and Bai, B, "Job satisfaction, life satisfaction, and turnover intent among food-service manager". *Cornell Hospitality Quarterly,* vol.*42, no.*2, pp.28-37, 2001*.*

[12] Robbins, S.P, "*Organizational Behavior".* New Jersey: Pearson Education International, Inc, 2003.

[13] Tanke, M. L, "Human resources management for the hospitality industry (2nd ed.)". Delmar: Thomson Learning, 2001.

[14] Huang, P.W, "The relationship between job characteristics and job satisfaction: An example of international tourist hotels". *Journal of Nan-Tai College*, vol.22, pp.143-148, 1996.

[15] Waldman, J.D., Kelly, F., Arora, S and Smith, H.L, "The shocking cost of turnover in health care". *Health Care Management Review,* vol.29, no.1, pp.2–7, 2004.

[16] Beach, R., Brereton, D and Cliff. D, "Workforce turnover in fly-in/fly-out (FIFO) mining operations in Australia: An exploratory study". *Summary Report*. Brisbane: Centre for Social Responsibility in Mining, 2003.

[17] Westlund, S.G and Hannon, J.C, "Retaining talent: Assessing job satisfaction facets most significantly related to software developer turnover intentions". *Journal of Information Technology Management*. vol.19, no. 4, pp. 1-15, 2008.

[18] Loi, R., Hang-Yue, N and Foley, S. "Linking employees' justice perceptions to organizational commitment and intention to leave: the mediating role of perceived organizational support". *J. Occup. Organ. Psychol.,* no.*79*, pp.101-120, 2006.

[19] Dess, G.G and Shaw, J.D, "Voluntary turnover, social capital, and organizational performance". *Acad. Manage. Rev*., no.26, pp.466-456, 2001.

[20] Chen, G., Polyhart, R. E., Thomas, H. C., Anderson, N and Bliese, P. D, "The power of momentum: A new model of dynamic relationships between job satisfaction change and turnover intentions". *Academy of Management Journal*, vol.54, no.1, pp.159-181, 2011.

[21] Kacmar, K.M., Andrews, M.C., Roy, V.D.L., Steilbergand, R.C and Cerrone, S, "Sure everyone one can be replaced but at what cost? Turnover as a predictor of unit level performance". *Academy of Management Journal*, 49, pp. 133-142, 2006.

[22] Pitts, D., Marvel, J and Fernandez, S. "So hard to say goodbye? Turnover intention among US federal employees". *Administration Review*, vol.71, no.5, pp.751-760, 2011.

[23] Sean, V., Godkin, L., Fleischman, G. M and Kidwell, R, "Corporate ethical values, group creativity, job satisfaction and turnover intention: The impact of work context on work response". *Journal of Business Ethics*,no.98, pp.353–372, 2010.

[24] Barak, M. E., Nissly, J. A and Levin, A, "Antecedents to retention and turnover among child welfare, social work, and other human service employees: what can we learn from past research? A review and meta analysis". *The Social Service Review,* vol.75, no.4, pp. 625-661, 2001.

[25] Samad, S, "Predicting turnover intentions: The case of Malaysian government doctors". *The Journal of American Academy of Business*, vol.*8,* no.2, pp.113-119, 2006.

[26] Price, J, "Reflections on the determinants of voluntary turnover". *International Journal of Manpower,* no.*22,* pp.600-624, 2001.

[27] Cropanzano, R and Greenberg, J, "Progress in organizational justice: Tunneling through the maze". *International Review of Industrial and Organizational Psychology*, vol.*12,* no.11, pp.317-372, 1997.

[28] Harter, J. K., Schmidt, F. L and Hayes, T. L, "Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis". *Journal of Applied Psychology,* vol.*87,* no.2, pp.268-279, 2002.

[29] Zulkarnain, "The mediating effect of quality of work life on the relationship between career development and psychological well-being". *International Journal of Research Studies in Psychology*. Advance online publication. DOI: 10.5861/ijrsp.2013.259,2013.

[30] Ryff, C, "Happines is everyting, or is it? Explorations on the meaning of psychological well-being". *Journal of Personality and Social Psychology*, vol.*57,* no.60, pp.1069-1081, 1989.

[31] Keyes, C. L. M., Shmotkin, D and Ryff, C. D, "Optimising well-being: The empirical encounter of two traditions". *Journal of Personality and Social Psychology*, vol.*82,* no.6, pp.1007 – 1022, 2002.

[32] Ryff, C and Keyes C.L.M, "The structure of psychological well-being revisited". *Journal of Personality and Social Psychology,* vol.*69,* no.4, pp.719-727, 1995.

[33] Sirgy, M.J., Reilly, N.P., Wu, J and Efraty, D, "A work-life identity model of well-being: Towards a research agenda linking quality-of-work-life approach to psychological well-being". *Journal of Happiness Studies*, no.*9*, pp.13-39, 2008.

[34] Cropanzano, R and Wright, T, "Psychological well-being and job satisfaction as predictors of job performance". *Journal of Occupational Health Psychology,* vol.*5, no.*1, pp.84-94, 2000.

[35] Harris, G.E and Cameron, J.E, "Multiple dimensions of organizational identification and commitment as predictors of turnover intentions and psychological well-being". *Canadian Journal of Behavioural Science*, vol.*37, no.*3, pp.159-169, 2005.

[36] Butali, N.D., Wesang'ula, P, M and Mamuli. L. C,"Effect of staff turnover on performance of work in Masinde Muliro University of Science and Technology". *International Journal of Human Resource Studies*, vol.*3,* no.1, pp.1-8, 2013.

[37] Nasution, W, "Influence of job satisfaction toward turnover intentions on Telkomsel *call center* employees". *Jurnal Mandiri*, no.*4*, pp.1-11,2009.

[38] Pasewark, W. R and Strawser, J. R, "The determinants and outcomes associated with job insecurity in a professional accounting environment". *Behavioral Research in Accounting*, no.*8*, pp.91-113, 1996.

[39] Arnold, H.J and Feldman, D.C, "*Organizational behavior"*. New York: McGraw Hill Book, 1996.

[40] Vandenberg, R.J and Nelson, J.B, "Disaggregating the motives underlying turnover intentions: When do intentions predict turnover behavior?" *Human Relations*,no.52, pp.1313–1336, 1999.

[41] Robertson, I and Cooper, C, "*Well-being: Productivity and happiness at work".* London: Palgrave MacMillan, 2011.

[42] Ryff, C and Singer, B, "Know thyself and become what you are: An eudaimonic approach to psychological well-being". *Journal of Happiness Studies*, no.*9,* pp.13-39, 2008.

[43] Ahuja, M., Chudoba, K.M., George, J.F, Kacmar, C and McKnight, H, "Overworked and isolated? Predicting the effect of work-family conflict, autonomy, and workload on organizational commitment and turnover of virtual workers". *Proceedings of the 35th Hawaii International Conference on System Sciences*. Big Island, Hawaii, 7 - 10 January. California: IEEE Computer Society. pp. 3586-3593, 2002.

[44] Ohly, S and Fritz, C, "Work characteristics, challenge appraisal, creativity, and proactive behavior: A multi-level study". *Journal of Organizational Behavior*, no.*31*, pp.543 -565, 2010.

[45] Gagné, M, "The role of autonomy support and autonomy orientation in prosocial behavior engagement". *Motivation and Emotion*, vol.*27,* no.3, 199-223, 2003.

[46] Horn, J.E.V., Taris, T.W., Schaufeli, W.B and Schreurs, P.J.G, "The structure of occupational wellbeing: A study among Dutch teachers". *Journal of Occupational and Organizational Psychology, no.77*, pp.365-375, 2004.

[47] Briner, R. B, "Relationship between work environment, psychological environments and psychological well-being". *Occupational Medicine*, vol.*50, no.*5, pp.299-303, 2000.

[48] Milman, A, "Hourly employee retention in the attraction industry: Research from small and medium-size facilities in Orlando, Florida". *Journal of Leisure Property*, no.*2*, pp.40-51, 2002.

[49] Baumeister, R.F and Leary, M.R, The need to belong: Desire for interpersonal attachments as a fundamental human motivation*. Psychological Bulletin*, no.*3,* pp. 497-529, 1995.

[50] Hasan, S.A and Subhani, M.I, "Managerial social wisdom: A major facet for employee turnover intentions, work commitments and manager-subordinate relationships". *European Journal of Economics, Finance and Administrative Sciences, no.40*, pp.132-141, 2011.

[51] Hellriegel, D. Slocum, Jr. J. W and Woodman, R.W, "*Organizational behavior".* Sydney: Thomson Learners, 2001.

[52] Lee, C.C., Huang, S.H and Zhao, C. Y, "A study on factors affecting turnover intention of hotel empolyees". *Asian Economic and Financial Review,* vol.2, no.7, pp.866-875, 2012.

[53] Jan, M and Masood, T, "Factors affecting life satisfaction among women". *Journal of Psychological Researches*, no.*51*, pp.63-73, 2007.

[54] James, J.B and Spiro, A. III, "The impact of work on psychological health and well-being of older Americans". In. J. James, P.Wink, & K.W. Schaie (Eds.), *The crown of life: The dynamics of the early post retirement period.* New York: Springer Publishing. pp. 49 – 64, 2006.

[55] Sheldon, K. M and Kasser, T, "Goals, congruence, and positive well-being: New empirical support for humanistic theories". *Journal of Humanistic Psychology*, no.*41,* pp.30 -50, 2001.

[56] Milliman, J., Czaplewski A.J and Ferguson J, "Workplace spirituality and employee work attitudes: An exploratory empirical assessment". *Journal of Organizational Change Management*, vol.*16, no.*4, pp.426-447, 2003.

[57] Moncarz, E., Zhao, J and Kay, C, "An exploratory study of US lodging properties organizational practices on employee turnover and retention". *International Journal of Contemporary Hospitality Management,* vol.*21,no.*4, pp.437-458, 2009.

[58] Yin, J.T and Yang, K.A, "Nursing turnover in Taiwan: A meta-analysis of related factors". *International Journal of Nursing Studies,* vol.*39, no.*6, pp.573–581, 2002.

[59] Shields, M.A and Ward, M, "Improving nurse retention in the national health service in England: The impact of jobsatisfaction on intentions to quit". *Journal of Health Economics,* vol.*20, no.*5, pp.677–701, 2001.

[60] Rhoades, L and Eisenberger, R, "Perceived organizational support: A review of the literature". *Journal of Applied Psychology*, no.*87*, pp.698–714, 2002.

[61] Ajzen, I, "*Attitude, personality and behavior".* New York: Open University Press, 2005.

[62] Aryee, S and Chay, Y.W, "Workplace justice, citizenship behavior, and turnover intentions in a union context: Examining the mediating role of perceived union support and union instrumentality". *Journal of Applied Psychology*, vol.86, no.1, pp.154 – 160, 2001.

[63] Aryee, S and Zhen, X.C, "Leader-member exchange in a Chinese context: Antecedents, the mediating role of psychological empowerment and outcomes". *Journal of Business Research,* vol.59, no.7, pp.793-801, 2006.

[64] Erdogan, B and Enders, J, "Support from the top: Supervisors' perceived organizational Support as a Moderator of leader-member exchange to satisfaction and performance relationships". *Journal of Applied Psychology*, vol.92, no.2, pp. 321–330, 2007.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 672-679

http://www.issr-journals.org/ijias/

Metallurgical Characterisation of Recovered Aluminium Alloys in Cameroon

T. Tchotang[1], B. D. Soh Fotsing[2], B. Kenmeugne[1], G.F. Anago[3], M. Fogué[2], and E. Nguena[4]

[1]Département des Génies Industriel et Mécanique,
Université de Yaoundé 1/Ecole Nationale Supérieure Polytechnique de Yaoundé/Laboratoire des Matériaux et Mécaniques,
Yaoundé, Cameroun

[2]Département de Mécanique et Productique,
Université de Dschang/Institut Universitaire de Technologie Fotso-Victor de Bandjoun/Laboratoire d'Ingénierie des Systèmes, Industriels et del'Environnement,
Bandjoun, Cameroun

[3]Département de Génie Mécanique,
Université Nationale de Bénin/Ecole Polytechnique d'Abomey-Calavi/Laboratoire d'Etudes et de Recherches des Techniques Industrielles, Cotonou, Bénin

[4]Alucam-Socatral a branch of Alcan Rio Tinto,
Edéa, Cameroun

ABSTRACT: This article is a comparative study of metallurgical characteristics of the different aluminium alloys gotten through recycling of recovered aluminium in Cameroon. A simple experimental device for the foundry of secondary aluminium blend, of very good quality built around a movable charcoal furnace is presented. It enables better energy efficiency, a better distribution of the heat around the crucible and indirectly assures good quality of the products obtained, while respecting the economic constraints and users' safety. Six refining methods are proposed by the addition of polyvinyl chloride (method A), coke rich in carbon CH_S (method C), ammonium chloride NH_4Cl (method E), manganese dioxide MnO (method T), acrylic nitrite $(C_2H_3Cl)_n$ (method P) and sodium chloride NaCl (method S). A critical analysis of the different recycling techniques is presented as well as a proposed process of melting and refining that enables the obtaining products with high degrees of purity. The results are then compared to the results obtained from the industrial methods of aluminium refining such as fractional crystallization (FC), granular filtration (GF) and dissolution in a metal solvent (DS). The later (DS) gives the rate of 6.540% of accumulated alloy elements and enables the best purification (93.460%), while the NaCl gives the lowest global rate of additive elements (9.478%), with the best purity index (90.522%) amount the proposed methods. Results obtained show that this method of refining improves the metallurgical properties of secondary aluminium alloy blends and guarantees better safety, as well as reducing the risks of environmental pollution.

KEYWORDS: Aluminium, Foundry, Recycling, Recovery, Refining.

1 INTRODUCTION

Recycling is the process of treatment which enables the reintroduction of some used objects through recovery, in the production cycle of a product. In Sub Saharan Africa, the methods of recycling used by aluminium artisanal casters to manufacture some kitchens utensils and various tools are misleading. According to M. Fogue and *al*, 1987 [1] impurities that can affect the quality of the product, affect the health of artisanal casters and that of those using the products. It therefore becomes important to find a method of refining that can limit these intrusions while improving the quality of products.

Corresponding Author: B. D. Soh Fotsing (sohfotsing@aol.fr)

This paper proposes a method of refining of the recovered aluminium based on the reduction of the intrusion of impurities by the modification of the microstructure of the recovered aluminium alloys with the help of the molten salts and makes a comparative study of the results gotten with those from the traditional methods used by artisanal casters.

On the basis of models of heated metal furnace studied by Jean Perron, 1987 [2] and concerning the problems of efficiency and safety with respect to the rudimentary state of materials for manufacturing traditional smelting furnaces, a prototype of a mobile fusion oven, that is more detailed and better adapted to the Cameroonian context has been designed in 2005 by Theodore Tchotang [3].

The mobile fusion oven designed [3] is composed of two main parts notably the fireplace and the crucible. The fireplace built with refractory bricks is made up of a cylindrical section consisting of a combustion chamber of about 100 litres in capacity and a lid composed of two parts fitting together one in the other. The crucible is made of refractory steel in which the liquid is melted, with a capacity of nearly 32 litres. The oven prototype enables the assurance of a good thermal insulation, because temperatures can reach 1800°C while functioning for 150min on its own. This oven offers efficiency of 65.28%, a value close to that gotten by P. Lailler, 1983 [4], ranging from 60% to 70% and is better in relation to the one used in traditional fusion oven by artisanal casters that is only about 12.21% [3].

The determination of the content of the constituent elements of the products manufactured based on aluminium alloy recovery has been done on the electronic ionisation impact mass spectrometer used in the ultramodern chemical analysis laboratory belonging Alucam-Socatral company in Edea, Cameroon.

2 CLASSICAL PROCESSES OF RECYCLING OF THE RECOVERED ALUMINIUM ALLOYS

2.1 INDUSTRIAL PROCESSES OF RECYCLING OF THE RECOVERED ALUMINIUM ALLOYS

2.1.1 DESCRIPTION OF THE PROCESSES

Industrial recycling of recovered aluminium alloys takes place in general, according to M. Samuel, 2003 [6] in four main stages notably the collection of scrap, preparation of the materials collected by cooling and heating after grinding and sorting, the smelting and the refining of the liquid. Three techniques of refining of recovered aluminium alloys that enable the production of aluminium of high purity exist according to Bao of Sun and *al*, 2004 [7].

Gaston Riverin, 1993 [8] presents the method of purification of alloys of recovered aluminium by fractional crystallization (FC), that has as advantage the separation of the constituents at low temperatures leading to energy saving. The inconvenience of this method is its low rate of production for a relatively long time of treatment. Its functioning implies high investments and a very expensive maintenance for a low rate of production. According to Harald Gorner, 2009 [9], the method of purification by granular filtration of tabular alumina (GF) is efficient when the concentration in inclusions is weaker and it is recommended when the diameter of inclusions is small. It is costly and reserved for productions on a large scale. In fact the process of purification by dissolution in a noble metal solvent (DS) invented by Paul, R. Kruesi, 2007 [10], remains very efficient for the recovery and the refining of the aluminium. Its principal disadvantage is the very high current density used that implies an increased energy cost and the difficulty to dispose of the noble metals.

2.1.2 COMPARATIVE STUDY OF THE METALLURGICAL CHARACTERISTIC OF RECOVERED ALUMINIUM ALLOYS GOTTEN FROM THE DIFFERENT INDUSTRIAL REFINEMENT

By optimal content comparison [3] of additives and the degree purity of the recovered aluminium alloys gotten by the three industrial techniques of refinement notably the purification by fractional crystallization (FC) [8]; by granular filtration of tabular alumina (GF) [9] and by dissolution in a noble metal solvent (DS) [10], the later (DS) gives the weakest rate (6.540%) of accumulated alloy elements and enables the best purification (93.460%) of aluminium alloys from non refined alloys composed of 85.387% aluminium; and thus a better reduction of the rate of additives (-123.385%) and an increase of the degree of relative purity of 8.634%.

2.2 TRADITIONAL PROCESSES OF REFINING OF THE RECOVERED ALUMINIUM ALLOYS

2.2.1 TRADITIONAL METHODS OF REFINING OF RECOVERED ALUMINIUM ALLOYS

According to studies led by M. Fogue and *al*, 1998 [1], processes of the recovery and recycling of the aluminium in units of traditional foundry include the stages of collection, sorting and compacting, smelting and refining, milling and flow and the

polishing. After collection, material to be melted is introduced in the crucible, starting with the light scrap, followed by the heavy scrap. When the smelting liquid has the required consistence of a paste, the polyvinyl chloride (PVC) with chemical formula $(C_2H_3CL)_n$ is added for the refining. In the presence of free oxygen combustion according to equation (1) leads to the formation of carbon monoxide (CO), hydrochloric acid (HCl) and steam (H_2O).

$$(C_2H_3Cl)_n + \frac{3}{2}nO_2 \rightarrow nH_2O + nHCl + 2nCO \quad (1)$$

From the present metallic oxide reaction in the liquid with the hydrochloric acid (HCl), it forms the metallic chlorides that can be cleared at the surface of the liquid before pouring.

Table 1. Chemical compositions and purities of recovered aluminium alloys produced by the traditional process of refining by the old artisanal-casters , 1998 [1]

Elements of alloys (%)		Si	Fe	Cu	Mg	Mn	Sn	Cr	Ti	Ni	Zn	Pb	Ga	Elements of alloy accumulated	Al*(%)
Sources	**Samples**														
Old samples cast by artisanal - casters of Douala and analysed on the high power electronic microscope (MEB) at Insa in Lyons (Art)	**Art1**	06,750	0,770	0,350	0,515	-	0,030	0,007	0,012	0,025	0,205	0,019	0,015	08,697	91,303
	Art2	16,530	0,940	0,450	0,690	-	0,040	0,007	0,011	0,035	0,235	0,023	0,015	18,975	81,025
	Art3	12,150	1,490	0,560	0,505	-	0,030	0,007	0,013	0,020	0,190	0,014	0,020	14,998	85,002
	Art4	18,960	0,740	0,480	0,350	-	0,110	0,019	0,016	0,040	0,260	0,024	0,015	21,013	78,987
	Art5	08,226	0,700	0,280	0,490	-	0,025	0,007	0,014	0,020	0,170	0,012	0,010	09,953	90,047
	Art6	15,430	0,830	0,360	0,325	-	0,110	0,019	0,015	0,035	0,245	0,020	0,010	17,399	82,602

- **(*) Stands for an estimated value**
- **Numbers put in to show that the composition of the alloyed element is in excess with respect to values of the standard NFA50-105 [11]**

Table 2. Chemical compositions and purities of recovered aluminium alloys produced by the traditional process of refining by the new artisanal-casters, 2008, [3]

Elements of alloys (%)			Si	Fe	Cu	Mg	Mn	Sn	Cr	Ti	Ni	Zn	Pb	Ga	Elements of alloy accumulated	Al*(%)
Refining	**Samples and% of refining**															
A	**A000**	**0**	14,861	1,847	0,832	0, 556	0,278	0,200	0,448	0,025	0,255	0,527	0,126	0,028	19,427	80,573
	A050	**0,2**	15,431	0,898	0,356	0,236	0,025	0,024	0,181	0,008	0,156	0,425	0,145	0,002	17,887	82,113
	A100	**0,4**	13,248	0,785	0,359	0,256	0,056	0,001	0,012	0,012	0,004	0,325	0,062	0,002	15,122	84,878
	A150	**0,6**	12,155	1,494	0,565	0,482	0,039	0,002	0,024	0,023	0,052	0,268	0,087	0,003	15,194	84,806
	A200	**0,8**	13,754	0,836	0,485	0,362	0,045	0,001	0,032	0,008	0,002	0,241	0,074	0,001	15,841	84,159
	A250	**1**	16,354	0,987	0,545	0,458	0,016	0,003	0,426	0,007	0,25	0,189	0,121	0,002	19,358	80,642

- **(*) Stands for an estimated value**
- **Numbers put in to show that the composition of the alloyed element is in excess with respect to values of the standard NFA50-105 [11]**

2.2.2 COMPARATIVE STUDY OF THE METALLURGICAL CHARACTERISTIC OF RECOVERED ALUMINIUM ALLOYS GOTTEN FROM TRADITIONAL REFINING PROCESSES

Table 1 shows the only existing metallurgic data on recovered aluminium alloys melted by artisanal casters in the city of Douala in Cameroon and analysed with a high power electronic microscope (PEM) [1]. These metallurgical characteristics being insufficient, it was necessary to repeat the analysis in order to have some more complete information on the traditional process of refining of recovered aluminium alloys. Six (6) samples (Table 2), of recovered aluminium alloys have been cast by artisanal casters in the city of Yaounde. Each specimen has been tested five times and the final measurement considered for each sample is the arithmetic mean of results of all the specimens.

The analysis of results (Table 2) shows that all the six (6) samples of aluminium alloys recovered by artisanal casters has an excessive rate of Fe+Si, Cu, Mg, Zn and Pb (0.062 to 0.145%> 0.05) in accordance with the NFA50-105 standard [11]. Most specimens contain an excess of chromium and nickel. For each of these two metals an excessive content in 3 samples of the 6 considered can be noticed. The content of the Sn is only greater than the recommended value for the non refined sample. On the other hand the Ti and the Ga are in acceptable proportions. By comparison of the chemical compositions and degree of purity, the traditional refining [3], at 0.4% of $(C_2H_3Cl)_n$ gives the lowest overall rate (15.122%) and thus a relative (-28.468%) reduction of additive element in recovered aluminium alloys produced by artisanal casters in the city of Yaoundé and enables better purification to be done (84.878%) for the recovered aluminium alloys gotten from a non refined alloy composed initially of 80.573% aluminium; and thus an increase in the degree of purity of 5.072% in relative terms.

3 PROPOSED PROCESSES OF REFINING OF RECOVERED ALUMINIUM

3.1 DESCRIPTION OF THE METHOD

The beneficial effect of salt-oxides interactions during the recovery of aluminium raised in J.A.S. Tenorio and *al*, 2002 [12], permitted the putting in place of a purification process of recovered aluminium alloys [3], by the use of melted regular salts to remove the metallic oxides contained in molten aluminium alloys. In the operations of this progress, while following similar manufacturing steps as those of the traditional processes of recovering of salvaged aluminium alloys by the addition of polyvinyl chloride (method A) (PVC or PET), coke rich in carbon CH_S (method C), ammonium chloride NH_4Cl (method E), manganese dioxide MnO (method T), acrylic nitrite $(C_2H_3Cl)_n$ (method P), and sodium chloride NaCl (method S). So, the initial C, E, T, P, S and A are according to the source of the samples and to the nature of refining used in the purification of the fusion liquid as specified above. The contents of the refining elements vary from 0.2% to 1% of the molten aluminium mass; that corresponds to codes ranging from 000 to 250, in steps of 50 assigned to every initial.

3.2 UNDERTAKEN CHEMICAL REACTIONS

The compounds CO_2, CO, SO_2, H_2S, NOx, C_nH_m and MO occur according to combustion formulas (2) to (5) and oxidation of CH_S(C) in presence of the oxygen (6) [6], [12]. The combustion of compounds of CH_S of general formula C_xH_yO is shown in equation (2):

$$C_xH_yO + \left(x + \frac{y}{4} - \frac{1}{2}\right)O_2 \rightarrow xCO_2 + \frac{y}{2}H_2O \quad (2)$$

Equation (3), shows combustion of compounds of CH_S of general formula CH_yO_x written as follow:

$$CH_yO_x + (1 + \frac{y - 2x}{4})(O_2 + 3{,}76N_2) \rightarrow CO_2 + \frac{y}{2}H_2O + 3{,}76(1 + \frac{y - 2x}{4})N_2 \quad (3)$$

To obtain NO_x, the combustion equation (4) is used:

$N_2(air) + O_X \rightarrow NO_X + N$, when the temperature exceeds 1400°C (thermal NO) (4)

If the combustion is carried out at higher temperatures (over 1000°C), metallic oxides MO are reduced by the carbon and to form carbon monoxide according to equation (5):

$$MO + C \rightarrow M + CO \quad (5)$$

The following oxidation equation (6) in the presence of additive elements is gotten:

$$2nCH_S + Al_2S_3 + MS + mH_2O \xrightarrow{\text{Oxidation}} 2C_nH_m + 2nH_S + MO + 2SO_2 + 2Al(OH)_3 + 2H_2S \quad (6)$$

CH_S represents the macromolecules of coke rich in carbon;

C_nH_m represents the heavy hydrocarbons (gaseous)

MO and $Al(OH)_3$ represent oxides and metallic hydroxides

Al_2S_3 represents aluminium sulphides, easy combination of sulphur and aluminium.

In the presence of hydrochloric acid (HCl) produced according to equation (7) from ammonia salt NH_4Cl and oxygen (O_2), there is a reaction with the additive elements in the alloy according to the reaction (8) below and impurities can be released at the surface of the bath as chlorides and metallic hydroxides when the molten metal is mixed :

$$2NH_4Cl + \frac{7}{2}O_2 + 3C \rightarrow 2NO_2 + 2HCl + 3H_2 + 3CO \quad (7)$$

$$2NH_4Cl + Al_2S_3 + 2HCl + 6H_2O + 2Mg \rightarrow 2H_2 + 2Al(OH)_3 + 3H_2S + 2NH_3 + 2MgCl_2 \quad (8)$$

The gases given out are namely hydrogen (H_2), hydrochloric acid (HCl), ammonia (NH_3), carbon monoxide (CO), hydrogen sulphide (H_2S) and other gas traces.

The manganese dioxide MnO_2 and the added elements present in the bath of metal favour the chemical reaction (9), with the production of hydrochloric acid according to equation (7). This enables the impurities that are found at the surface of the bath as chlorides and metallic hydroxides to be removed when molten metal is mixed at the end to refine the metallic mixture.

$$MnO_2 + 2HCl + Al_2S_3 + 6H_2O + Mg \rightarrow Mn(OH)_2 + 2Al(OH)_3 + 3H_2S + MgCl_2 \quad (9)$$

The gases emitted are namely hydrogen (H_2), hydrochloric acid (HCl), carbon monoxide (CO), hydrogen sulphide (H_2S) and other gases like nitrogen dioxide NO_2 and traces of some metals.

After proper cleaning and compacting of the aluminium alloy trash, the combustion of the polyvinyl chloride $(C_2H_3Cl)_n$ is resumed in the bath in the presence of oxygen according to the reaction (1) as seen in the case of traditional refining, to produce carbon monoxide (CO), hydrochloric acid (HCl) and steam (H_2O) :

$$(C_2H_3Cl)_n + \frac{3}{2}nO_2 \rightarrow nH_2O + nHCl + 2nCO \quad (1)$$

According to equation (10), additive elements and the metallic oxides present in the metal in fusion react with the hydrochloric acid, to produce chlorides and metallic hydroxides that are formed at the surface of the vessel:

$$2HCl + MgO + Al_2S_3 + 5H_2O \rightarrow MgCl_2 + 2Al(OH)_3 + 3H_2S \quad (10)$$

The gases formed remain as previously mentioned as carbon monoxide (CO), hydrochloric acid (HCl), hydrogen sulphide (H_2S), traces of other gases and volatile organic compounds (COV).

The sodium chloride NaCl added in presence of oxygen is decomposed according to equation (11) and the dichloride formed immediately reacts with the Mg, Mn, Si and Fe metals, and reduces the additive elements content in the vessel of the aluminium alloy according to equation (12) to eliminate sodium peroxides, chlorides and the metallic hydroxides:

$$2NaCl + O_2 \rightarrow Na_2O_2 + Cl_2 \quad (11)$$

$$MgCl_2 + Al_2S_3 + 6H_2O \rightarrow 3H_2 + 2Al(OH)_3 + Cl_2S_3 + Mg \quad (12)$$

Gases given out at the time of the fusion of the recovered aluminium while using NaCl in refining are essentially hydrogen (H_2) and some traces of other gases and heavy metals..

Table 3. Chemical composition and purities indices of recovered aluminium alloys gotten according to the refining method by the addition of regular molten salts in the metal vessel in fusion in the Solid Mechanics Laboratory at ENSP, Yaoundé Cameroon [3]

Alloyed Elements (%)			Si	Fe	Cu	Mg	Mn	Sn	Cr	Ti	Ni	Zn	Pb	Ga	accumulated alloyed elements	Al*(%)
Refining	Samples and % of refining															
C	C000	0	12,799	0,714	0,368	0,045	0,287	0,002	0,055	0,063	0,101	0,113	0,002	0,057	14,606	85,394
	C050	0,2	09,830	0,662	2,590	0,327	0,192	0,028	0,018	0,049	0,019	0,313	0,012	0,021	14,061	85,939
	C100	0,4	09,019	0,724	2,573	0,315	0,206	0,003	0,026	0,061	0,024	0,254	0,013	0,025	13,243	86,757
	C150	0,6	09,499	0,742	2,387	0,245	0,252	0,014	0,007	0,017	0,017	0,195	0,017	0,020	13,412	86,588
	C200	0,8	09,777	0,747	2,39	0,280	0,205	0,018	0,060	0,014	0,020	0,109	0,020	0,020	13,660	86,340
	C250	1	10,681	0,645	2,364	0,184	0,245	0,023	0,061	0,015	0,015	0,189	0,017	0,025	14,464	85,536
E	E000	0	12,817	0,712	0,364	0,047	0,278	0,002	0,055	0,063	0,101	0,113	0,002	0,057	14,611	85,389
	E050	0,2	10,815	0,352	0,009	0,184	0,012	0,001	0,002	0,012	0,003	0,118	0,008	0,011	11,527	88,473
	E100	0,4	09,947	0,353	0,009	0,221	0,011	0,002	0,002	0,011	0,003	0,129	0,007	0,026	10,721	89,279
	E150	0,6	10,260	0,355	0,016	0,172	0,011	0,002	0,002	0,011	0,003	0,130	0,008	0,020	10,990	89,010
	E200	0,8	10,564	0,352	0,054	0,122	0,020	0,003	0,021	0,030	0,007	0,118	0,005	0,024	11,320	88,680
	E250	1	11,322	0,308	0,061	0,075	0,018	0,002	0,018	0,050	0,021	0,164	0,008	0,015	12,062	87,938
T	T000	0	12,808	0,717	0,363	0,044	0,282	0,002	0,055	0,063	0,101	0,113	0,002	0,057	14,607	85,393
	T050	0,2	11,991	0,693	0,409	0,09	0,066	0,023	0,007	0,015	0,017	0,099	0,018	0,002	13,430	86,570
	T100	0,4	11,133	0,702	0,402	0,099	0,062	0,019	0,006	0,014	0,017	0,103	0,019	0,025	12,601	87,399
	T150	0,6	11,343	0,740	0,387	0,097	0,061	0,025	0,006	0,015	0,015	0,099	0,018	0,025	12,831	87,169
	T200	0,8	11,243	0,708	0,284	0,449	0,207	0,003	0,023	0,059	0,023	0,251	0,009	0,002	13,261	86,739
	T250	1	12,207	0,761	0,404	0,089	0,071	0,014	0,007	0,014	0,018	0,099	0,019	0,026	13,729	86,271
P	P000	0	12,811	0,716	0,366	0,040	0,284	0,002	0,055	0,063	0,101	0,113	0,002	0,057	14,610	85,390
	P050	0,2	11,390	0,705	1,786	0,039	0,200	0,004	0,019	0,057	0,023	0,242	0,009	0,019	14,493	85,507
	P100	0,4	09,774	0,671	2,638	0,068	0,189	0,007	0,018	0,060	0,045	0,222	0,008	0,019	13,719	86,281
	P150	0,6	10,824	0,705	2,196	0,063	0,204	0,010	0,021	0,015	0,016	0,200	0,015	0,020	14,289	85,711
	P200	0,8	10,927	0,732	2,123	0,063	0,245	0,017	0,004	0,014	0,007	0,157	0,016	0,002	14,307	85,693
	P250	1	10,501	0,685	2,747	0,073	0,205	0,003	0,017	0,048	0,019	0,244	0,008	0,027	14,577	85,423
S	S000	0	12,822	0,715	0,365	0,043	0,275	0,002	0,055	0,063	0,101	0,113	0,002	0,057	14,613	85,387
	S050	0,2	10,044	0,529	0,031	0,252	0,027	0,001	0,003	0,012	0,004	0,017	0,012	0,013	10,945	89,055
	S100	0,4	08,641	0,358	0,009	0,355	0,014	0,002	0,025	0,011	0,003	0,030	0,008	0,022	09,478	90,522
	S150	0,6	09,469	0,423	0,030	0,292	0,018	0,001	0,030	0,013	0,004	0,028	0,012	0,017	10,337	89,663
	S200	0,8	09,652	0,347	0,009	0,254	0,014	0,002	0,002	0,011	0,003	0,017	0,008	0,023	10,342	89,658
	S250	1	10,630	0,318	0,030	0,389	0,015	0,002	0,020	0,016	0,001	0,019	0,010	0,017	11,467	88,533

3.3 PRESENTATION AND DISCUSSION OF RESULTS OF SPECTROMETRIC ANALYSIS OF RECOVERED ALUMINIUM ALLOYS GOTTEN ACCORDING TO THE METHOD OF REFINING THE MOLTEN SALT ADDED IN THE METAL VESSEL

Table 3 represents contents for which every refining element is given for maximum reduction of impurities and an increase in the degree of purity in a population of 150 specimens of recovered aluminium alloys purified by molten salt addition.

A comparison of spectrometric analysis results done on samples, with excerpts of the French Standard NFA50-105 [11], has been made and slots in gray are those of which contents of allied elements are more than the recommended values. Furthermore, addition of the contents of Fe+Si elements whose values are greater than the recommended ones according to the standard for all tested samples, those of the other elements namely copper (20 out of 30 samples), magnesium (15 out of 30 samples), zinc (22 out of 30 samples) are superior to the Standard. The high content of manganese is observed in samples (16 out of 30 samples). Nevertheless, the behavior of copper in alloys depends on the different methods of refining. Indeed, the refining CHs and $(C_2H_3Cl)_n$ increases the content of copper alloy after refining as compared to the non refined alloy . The MnO_2 enables the copper content of the alloys to remain almost unaltered. And at the end NH_4Cl and NaCl give the aluminium alloys a copper content that decreases greatly after refining and permits them to respect the norm. On the whole, contents in Sn, Cr, Ti, Ni, Pb and Ga are distinctly lower than the values of the norm. The refining with NaCl is favours a big increase of the Mg content, to the detriment of the other elements of which contents after refining with NaCl are lower than

the norm, except Fe+Si. It is also observed that there are less additive elements in excess when NaCl is used in refining. The metallurgic characteristics (table 3) of the refining of recovered aluminium alloys by the molten salt addition shows that refining with NaCl is the most efficient because of its lowest rate of impurities (9.478%) and a reduction in relative value of -54.178%. It enables the best purification (90.522%) of aluminium gotten from a non refined aluminium alloy (85.387%) to be done; thus an increase in purity index; indicated by the relative value (5.673%). Data of the refining of the metal vessel NaCl is as the values of as the reference values for comparison with the other techniques of refining [3].

4 GENERAL COMPARISON OF THE OPTIMAL METALLURGICAL CHARACTERISTIC OF RECOVERED ALUMINIUM ALLOYS GOTTEN FROM DIFFERENT PROCESSES OF REFINING STUDIED

By comparing of results gotten from the different processes of purification of recovered aluminium alloys studied, the performance of the refining method with NaCl has been investigated and lies (Fig. 1) between that of the industrial purification by dissolution in a noble metal solvent (DS) which is more efficient, and the one of the traditional refining with $(C_2H_3Cl)_n$ whose values are not good enough. It would be necessary for a more reliable use of refining of aluminium alloys with NaCl(S100) at the industrial scale, to consider less use of resources in relation to the industrial refinement by DS, however enormous efforts are to be made extensively in relation to alloys of traditional refining as it is limited in efficiency .

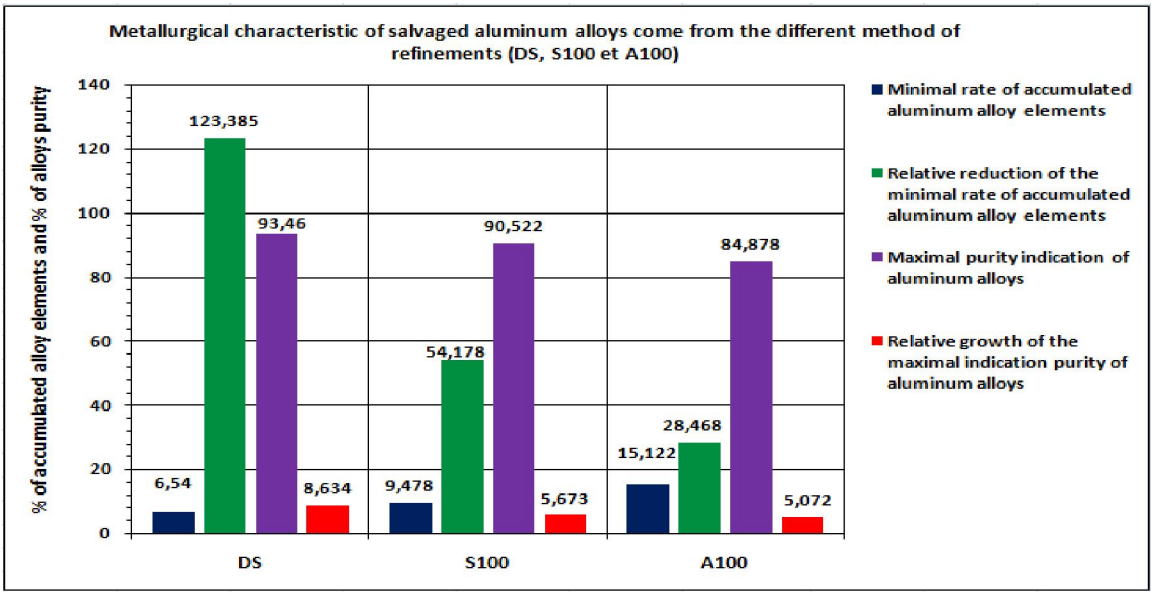


Fig. 1. Metallurgic characteristic according to refining techniques (DS, S100 et A100)

5 CONCLUSIONS AND PERSPECTIVES

Among all the molten salt refining methods proposed, the refining with NaCl gives the lowest global rate of additive elements (9.478%), with the best purity index (90.522%) for aluminium gotten from a non refined alloy initially at 85.387% of aluminium and makes it more suitable to minimize the poisonous effects of alloys tested, for health and to reduce risks of environmental pollution. In addition to the satisfactory result, due to the inconsistency of the gas proportion given out during combustion and fusion, it has proven in addition to be the best alternative in relation to the industrial processes of refining, inaccessible and thus difficult to put place, implying the more energy costs and very high investment cost for a production in small scale.

With the detection of the effective presence of poisonous heavy metals, tests of change of state shall be considered in the future, to deduct in order to foresee the degree of their harmfulness and to contribute to the standardisation of the recovery process and the recycling of aluminium in Cameroon as well as in other countries of sub-Saharan Africa.

ACKNOWLEDGMENT

The authors would like to thank the Alcan Rio Tinto branch in Cameroon, namely Alucam-Socatral, for their experimental support and the provision of the test specimens. The French Cooperation in Cameroon COMETES, is gratefully acknowledged for the funding of the project.

REFERENCES

[1] M. Fogué, G. F. Anago, B. Soh Fotsing et P. Ngnaguepa, "L'aluminium au service de l'alimentaire : fonderie et affinage au Cameroun," *Déchets*, Sciences & Techniques, No. 11, 3ème trimestre, pp. 12-16, 1998.

[2] Jean Perron, "Modélisation mathématique simplifiée d'un four de métal chaud," *Thèse de docteur-ingénieur, Université du Québec à Chicoutimi*, 226 p, Juin 1987.

[3] Théodore Tchotang**,** "Contribution au traitement de recyclage de l'aluminium de récupération au Cameroun," *Thèse de doctorat/Ph.D en sciences de l'Ingénieur, Université de Yaoundé 1*, 297 p, 2013.

[4] P. Lailler, "Performance improvements in aluminum melting furnaces," *Light metal age*, vol. 41, n°. 9, pp. 22-23, October 1994.

[5] Annual Book of ASTM Standards, Section 3, "Metal test methods and analytical procedures," *E466-82*, vol. 3.1, pp. 470-471, 1995.

[6] M. Samuel, "A new technique for recycling aluminum scraps," *Journal of Materials Processing Technology*, n°. 135, pp. 117-124, 2003.

[7] Bao-de Sun, Wen Jiang Ding, Da Shu and Yao-he Zhou, "Purification technology of molten aluminium," *Journal of central south university of technology,* vol. 11, n°. 2, pp. 134-141, 2004.

[8] Gaston Riverin, "Purification de l'aluminium par cristallisation fractionnée," *Mémoire de maîtrise en ressources et systèmes, Université du Québec à Chicoutimi,* 132 p, Juin 1993.

[9] Harald Gorner, "Removal of dissolved elements in aluminium by filtration," *Thesis for the degree of Philosophiae Doctor, Norwegian university of sciences and technology, NTNU*, n°. 225, 228 p, 2009.

[10] Paul R. Kruesi, "Procédés de récupération et de purification d'aluminium secondaire," *OMPI*, WO/2007/062402, 2007.

[11] Norme NFA50-105, "Caractéristiques chimiques du métal dans la fabrication de matériel et d'ustensiles servant à la préparation, à la cuisson des aliments et à la conservation des aliments".

[12] J. A. S. Tenorio, D. C. R. Espinosa, "Effect of salt-oxyde interaction on the process of aluminum recycling," *Journal of Light Metals,* vol. 2, pp. 89-93, 2002.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 680-684

http://www.issr-journals.org/ijias/

Ultra Wideband Slotted Microstrip Patch Antenna for Downlink and Uplink Satellite Application in C band

Tajeswita Gupta and P. K. Singhal

Department of electronics and communication MITS Gwalior
Gwalior, M.P. 474005, India

ABSTRACT: An ultra-wideband slotted microstrip patch antenna has been proposed in this paper for uplink and downlink satellite applications in c band of IEEE 802.11 standards. Various slots have been introduced in the patch to achieve wider bandwidth. Presented work provides a comparative result of the microstrip patch antenna with and without the slots. Fractional bandwidth of the basic antenna is 9% which is increased to 46% after introduction of the slots in the proposed patch antenna.

KEYWORDS: Microstrip, Patch Antenna, Resonating Frequency, Impedance Bandwidth, Return Loss.

1 INTRODUCTION

Microstrip antennas due to their many attractive features have drawn attention of industries for an ultimate solution for wireless communication. A Microstrip patch antenna is a narrowband, wide- beam antenna fabricated by etching the antenna element pattern in metal trace bonded to an insulating substrate. Because such Antennas have a very low profile, are mechanically rugged and can be conformable, they are often mounted on the exterior of aircrafts and spacecrafts, or are incorporated into mobile radio communications devices. This needs very accurate calculation of various design parameters of microstrip patch antennas. Patch dimensions of microstrip antenna is a vital parameter in deciding the utility of a microstrip antennas. Conventional microstrip antennas in general have a conducting patch printed on a grounded microwave substrate [1-2].

Many techniques have been reported to reduce the size of microstrip antennas at a fixed operating frequency. In general, microstrip antennas are half-wavelength structures. The important characteristic of micro strip antennas is their inherent ability to radiate efficiently despite their low profile. The primary source of this radiation is the electric fringing fields between the edges of the conductor element and the ground-plane behind it. Much intensive research has been done in past years to develop enhancement techniques for broadband microstrip antennas. Several bandwidth enhancement techniques are found in literature. Some of these techniques include the use of thick substrates with a low dielectric constant, and stacked or co-planar parasitic patches. The use of thick substrates introduces a large inductance due to the increased length of the probe [3-9].

An antenna, for uplink and downlink application in satellite communication has been presented in this research paper. Antenna is resonating in the C band of IEEE standards. By the addition of various slots in the geometry and defected ground plane the bandwidth of the antenna has been enhanced to provide an ultra-wide band characteristic of the antenna having bandwidth of 1.8 GHz [10-12].

2 ANTENNA GEOMETRY

The geometry of the proposed antenna is shown in Fig.1, was modeled using the classical equations [1] Figure 1 shows the geometry of antenna with and without slots (proposed antenna). A planar antenna includes a rectangular radiator etched

Corresponding Author: Tajeswita Gupta (tajeswitagupta@gmail.com)

with various slots. The antenna has a patch size of 60 × 60 × 1.676 mm^3. The antenna is printed on FR4 substrate (thickness is 1.6 mm relative permittivity 4.4 and loss tangent 0.027) and is fed by 50 ohm microstrip line. On the other side of the substrate ground plane is printed with an area of 60 × 15 mm^2.

Proposed antenna drawn in figure 1 is a novel temple shaped antenna with a microstrip line feed of 50 Ohm. Design specifications of the proposed patch are mentioned in the table 1. The printed patch is of the dimension 30 x 55 mm^2.

Table 1. Design Specifications of the Proposed Antenna

Substrate material	FR4
Relative permittivity	4.4
Thickness of dielectric	1.6 mm
L_1	10 mm
L_2	30 mm
L_3	20 mm
L_4	30 mm
W_1	10 mm
W_2	20 mm
W_3	5 mm
L_4 (length of the feed)	15 mm
Width of the feed	3 mm

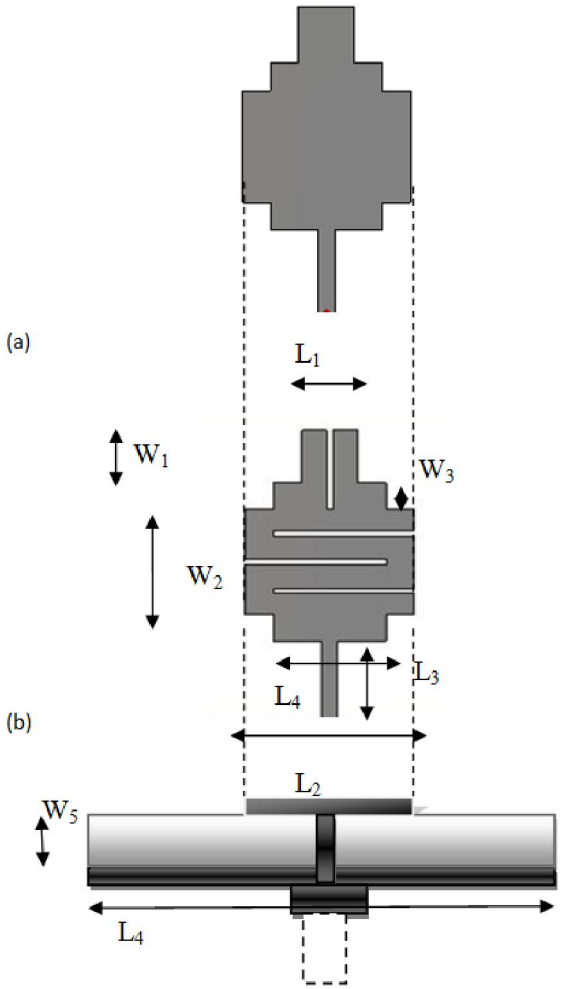


Fig. 1. (a) Basic antenna (b) Modified antenna with slots (proposed)

3 SIMULATED RESULTS

Proposed antenna was experimentally studied and simulated results have been depicted in Figure 2. Simulated result shows the return loss of the patch antenna against the frequency. It is clearly observed that the bandwidth is increased to 1.8 GHz instead of 400 MHz of basic patch antenna. Simulated result shows that the proposed antenna resonates for two band ranging from 3.077-4.9 GHz & 6.98-7.67 GHz.

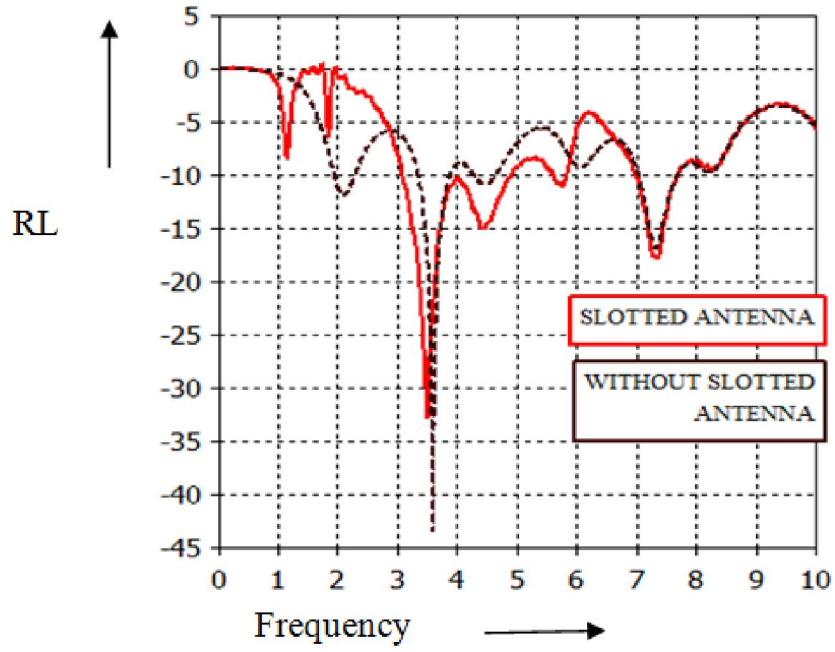


Fig. 2. Return loss against frequency graph

E and H field patterns of the proposed antenna have been depicted in the figure 3 as shown below. Figure 4 shows the directivity of the proposed antenna in both bands. Corresponding directivity is 2.7 dBi and -1.2 dBi with angular width angular 3 dB beam width of 57.7^0 and 36.2^0 and the direction of maximum radiation is 332^0 and 155^0 for uplink and downlink application band respectively.

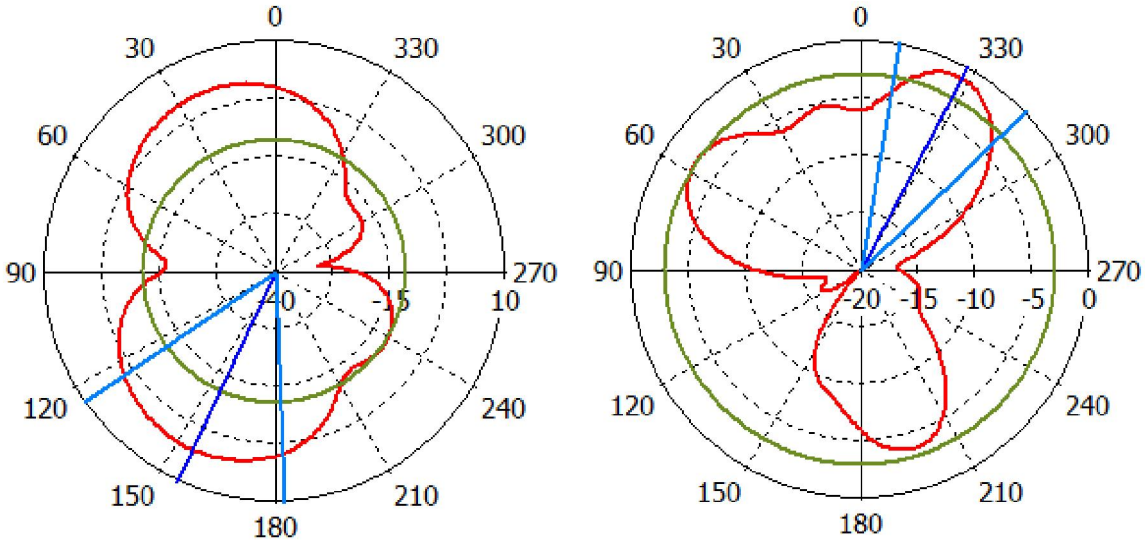


Fig. 3. (a) for downlink application band (b) for uplink application band

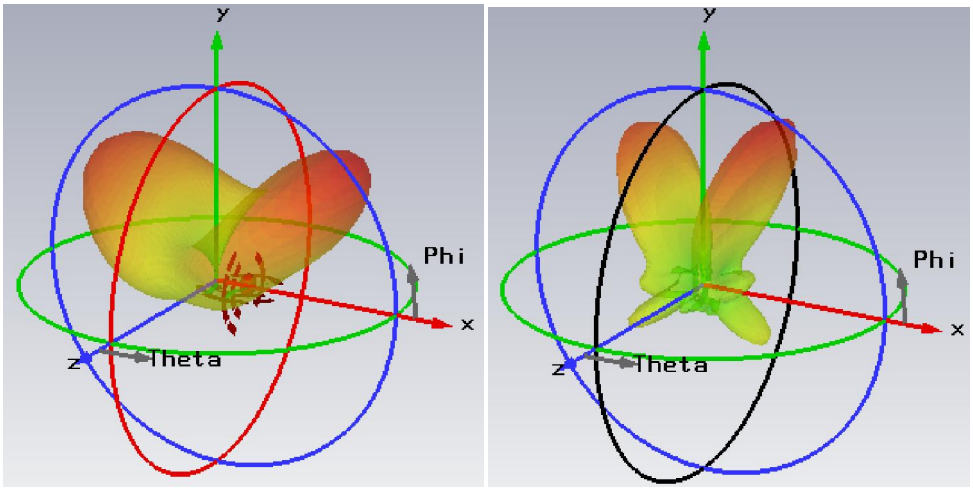


Fig. 4. (a): directivity of antenna for downlink band (b): directivity of antenna for uplink band

The gain of the antenna in both bands has been drawn in the figure 5 for both uplink and downlink applications. The absolute gain is 1.7 and 0.59 in the direction of maximum radiation for downlink and uplink application bands respectively.

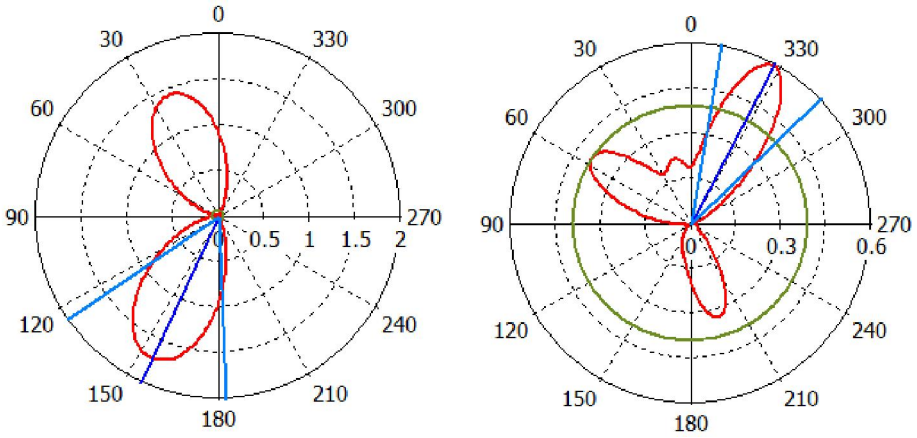


Fig. 5. (a): gain of the proposed antenna for downlink application (b): gain of the proposed antenna for uplink application

4 CONCLUSION AND REMARKS

A novel temple shaped antenna with various slots has been presented in this paper. Proposed antenna achieves the two bands in the IEEE C band for downlink and uplink applications in satellite communication. The achieved bands are 3.077-4.9 GHz & 6.98-7.67 GHz. The bandwidth for lower band is 1.8 GHz which provides it an ultra-wideband application for the achieved band. The proposed antenna has shown Low Profile characteristic and better impedance bandwidth and Return Loss performance.

REFERENCES

[1] C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd Edition, John Wiley and Sons, New York, 2005.

[2] Bahl, I. J. and P. Bhartia, *Microstrip Antennas*, Chap. 4, Artech House, Dedham, MA, 1980.

[3] Jegan, G., Juliet, A.V., and Kumar, G.A., "Multiband microstrip patch antenna for satellite communication," *Recent Advances in Space Technology Services and Climate Change (RSTSCC),* pp. 153-156, 2010. ISBN: 978-1-4244-9184-1.

[4] Zheng Qiurong, Lu Wanzheng, and Liu Feng, "A microstrip antenna element for satellite communication," *Antennas, Propagation and EM Theory, 2003. Proceedings. 2003 6th International SYmposium on*, pp. 81-85, 2003. ISBN: 0-7803-7831-8.

[5] Avisankar Roy, Sunandan Bhunia "Compact Broad Band Dual Frequency Slot Loaded Microstrip Patch Antenna with Defecting Ground Plane for Wi-MAX and WLAN", *International Journal of Soft Computing and Engineering (IJSCE)*, Vol. 1, No. 6, January 2012, pp. 154-157.

[6] Chakraborty, U., Chatterjee, S., Chowdhury, S.K., and Sarkar, P.P., "Triangular slot microstrip patch antenna for wireless communication", *India Conference (INDICON), 2010 Annual IEEE*, pp. 1-4.

[7] M. T. Islam & M. N. Shakib and N. Misran, "Multi-Slotted Microstrip Patch Antenna for Wireless Communication", *Progress In Electromagnetics Research Letters*, Vol. 10, 11, pp. 11-18.

[8] Sona O. Kundukulam, Manju Paulson, C. K. Aanandan, and P. Mohanan, "SLOT-LOADED COMPACT MICROSTRIP ANTENNA FOR DUAL-FREQUENCY OPERATION", *Microwave and Optical Technology Letters*, Vol. 31, No. 5, pp. 379-381.

[9] Jia-Yi Sze and Kin-Lu Wong, "Bandwidth Enhancement of a Microstrip-Line-Fed Printed Wide-Slot Antenna", *IEEE Transactions on Antennas and Propagation*, VOL. 49, NO. 7, pp. 1020-1024.

[10] Rabih Rahaoui and Mohammed Essaaidi, "Compact Cylindrical Dielectric Resonator Antenna excited by a Microstrip Feed Line," *International Journal of Innovation and Applied Studies*, vol. 2, no. 1, pp. 1–5, January 2013.

[11] Mohammed Younssi, Achraf Jaoujal, Yacoub Diallo, Ahmed El-Moussaoui, and Noura Aknin, "Study of a Microstrip Antenna with and Without Superstrate for Terahertz Frequency," *International Journal of Innovation and Applied Studies*, vol. 2, no. 4, pp. 369–371, April 2013.

[12] M. I. Hasan and M. A. Motin, "New slotting technique of making compact octagonal patch for four band applications," *International Journal of Innovation and Applied Studies*, vol. 3, no. 1, pp. 221–227, May 2013.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 685-692

http://www.issr-journals.org/ijias/

Online Tracking of Maximum Panel Power Output in Photovoltaic Stand Alone System with Different Insolation

E. Jensi Miriam and S. Ambalavanan

Lead - acid battery Research Group,
CSIR- Central Electrochemical Research Institute, Karaikudi, TamilNadu, India

ABSTRACT: In solar photovoltaic stand - alone system the basic device is the PV module which is used to charge the storage device during daytime and supplies power to the load during dusk to dawn. This paper presents knowledge based system for evaluating power generation system in PV model. The performance of a PV system depends on the environmental factors such as irradiation and cell temperature. It is a non-linear characteristic and this characteristic is varied in different PV technologies. To solve this problem, an intelligent technique called the Artificial Neural Network (ANN) can be talented solution for obtaining the maximum output power in real-time operation. Initially this work focuses on the simulation of characteristics of the panel power output of PV module at different level of radiation. Thus panel power output is evaluated different level of radiation and the simulated characteristics are figure with the 3D nomogram analysis. The database device using microcontroller is designed as per the simulation studies and it is attached in the solar panel to monitor the real time value of PV standalone system. The above mentioned simulated comparison is validated with results of local climatic data and its accuracy of the proposed methods has been measured with the error estimation method. Thus the proposed method will be very useful for determining the real-time optimum operating condition of PV system with estimated maximum power generation.

KEYWORDS: PV Standalone System, PV Module, Panel Power Output, Knowledge Based System, Artificial Neural Network.

1 INTRODUCTION

Photovoltaic is the field of technology and research related to the practical application of photovoltaic cells in producing electricity from light, though it is often used specifically to refer to the generation of electricity from sunlight [1]. Major components of solar photovoltaic system are solar photovoltaic Panel, battery to store energy, controller to prevent battery from over charging and discharging and a cable that connects the panel to battery and with each other to complete the system [2]. Output voltage from a solar panel will vary depending upon the solar panel design, the attached load, and the amount of light striking on its surface. Most panels are rated in watts of power that can theoretically be produced on a high intensity day. For an example, a 40W panel with a 16V output will generate 40W / 16V = 2.5A peak current. In reality, these peak currents are normally not achieved, and true outputs are about 80 to 90% of the ratings [3].It is the nature of a solar panel, that energy will only be available for a limited time frame during the day and also during winter. Further, conditions with heavy overcast skies or panels coated with snow will also limit energy production. To maintain continuous operation of a load, it is therefore necessary to have a storage battery that can be discharged and recharged upon demand [4]. In order to prevent an overcharge of the battery, it is necessary to use a controller between the panel and the battery .Currently, lead-acid battery is used as storage device for photovoltaic applications on account of low cost and high storage capability. Overcharging of a lead-acid battery will result in hydrogen gas generation, and shorten its life. Regulation of the solar panel output is performed by monitoring the lead acid battery voltage level and applying a shunt load across the panel when the battery is fully charged [5].In order to determine the characteristics of the PV module, the Power vs. Voltage (PV) and Current vs. Voltage (IV) curves must be constructed. The current and power outputs of photovoltaic modules are approximately proportional to sunlight intensity. At a given intensity, a module's output current and operating voltage is determined by the

Corresponding Author: E. Jensi Miriam (jensisandhya@gmail.com)

characteristics of the load and it is necessary to operate the PV at its Maximum Power Point [6].Manufacturers of photovoltaic modules, provide only a few experimental data about electrical and thermal characteristics of the panel. Some of the parameters required for adjusting photovoltaic module models such as the light-generated or photovoltaic current, the series and shunt resistances, cannot be found in the manufacturer's data sheets. All photovoltaic array datasheets basically has the following information: the nominal open-circuit voltage V_{oc}, the nominal short-circuit current I_{sc}, the voltage at the maximum power point V_{mp}, the current at the maximum power point I_{mp}, the open-circuit voltage/temperature coefficient K_V, the short-circuit current/temperature coefficient K_I, and the maximum experimental peak output power P_{max}. In this paper three remarkable parameters namely open circuit voltage (V_{oc}), short circuit current (I_{sc}) and maximum power point tracking (V_{mp}, I_{mp}) given by the manufacturer of the PV module is used for the prediction of PV characteristics of solar panel and panel power output for different level of radiation ranges from 100 to 1000W/m^2 . For the simulation of PV module an equivalent circuit model is proposed and in that, solar cell is modeled as a current source, I_{ph} which is proportional to ambient irradiance level and to the temperature of the panel [7]. For losses estimation, a series resistance R_s and a parallel resistance R_p are included in the circuit. The curve fitting factor was considered as an adjusting parameter so that at rated input values of temperature and irradiance, the data sheet values of the output were obtained as model outputs. The same approach was taken to obtain the temperature and irradiation correction coefficients. In another model active compensation is done at every moment is affected through a function for R_s, V_c and I_{ph} as a function of temperature and irradiance [8]. In the present model the equivalent circuit methodology has been attempted with Shunt resistance in parallel using circuit simulator in MATLAB/SIMULINK.The validity of the model with the new equation has been tested through prediction using ANN. For this purpose, irradiation and temperature are utilized as the input information of ANN Network and in the output layer it predicts maximum performance of PV system power generation it is compared with local climatic data. The details are described in this communication.

2 METHOD

2.1 ARTIFICIAL NEURAL NETWORK

ANN method using back propagation algorithm is utilized to estimate the maximum power generation of PV modules. This approach can be done because the open circuit voltage depends on the variations in insolation and cell temperature, while the cell temperature may be changed due the ambient temperature. From the practical point of view, the open-circuit voltage of PV module can be easily measured by interrupting the normal operation of the system temporarily and storing the measured value [9].On the other hand, the cell temperature can be simply measured at the backside of PV modules.

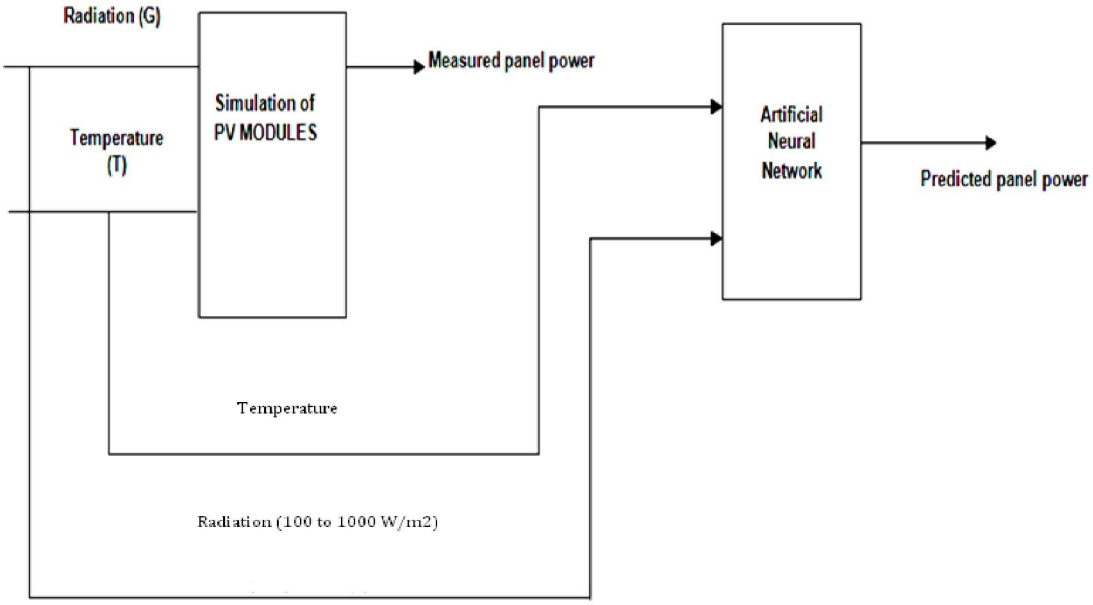


Fig. 1. General block diagram of the proposed system

The proposed system consists of two main blocks simulation of PV modules block and ANN blocks. The main objective of simulation of PV module block is to find the parameters of the nonlinear IV equation by adjusting the curve at three points: open circuit voltage, maximum power, and short circuit current, these three points are provided by all commercial array datasheets, with the parameters of the adjusted IV equation one can build a photovoltaic circuit model with any circuit simulator by using basic math blocks. Computer simulation program of PV module was developed for calculating the IV and PV characteristics at different level of radiation. The Simulation part of PV Module consists of three parts namely User friendly part, input part and internal model part. User Friendly block is designed in such a way that it holds only basic information about the simulation of the particular PV module such as type of model, manufacturer, nominal power and technology of the panel etc. The input block provides the basic information required in the nonlinear IV equation of the equivalent circuit in general form. These parameters are obtainable in manufacturer data sheets. The parameters include are: the reference irradiation (G_{Rref}) and module temperature conditions (T_{Ref}) during measurements, short circuit current (I_{sc}) and open circuit voltage (V_{oc}) at the given conditions. I_{mp} and V_{mp} are the operating point in the region of the maximum power point and current temperature coefficient, by default it can be taken as about 0.05%/°C. After defining the basic parameters, the program still needs some additional parameters to design the internal model parameters of the PV panel, such as shunt resistance and series resistance. In the real-time operation of the PV module some losses exist, they have been added to the model as a resistance in series (R_S) and another in parallel (R_{Sh}) to get a more real behavior [10]. For an ideal cell, R_{Sh} would be infinite and would not provide an alternate path for current to flow, while R_S would be zero. The R_{Sh} value can be obtained on the basis of measured IV characteristics of the module. R_{Sh} is the inverse of the slope around when voltage is zero. It is possible to approximate the series and shunt resistances, R_S and R_{Sh}, from the slopes of the I-V curve at V_{OC} and I_{SC}, respectively, therefore, series resistance and shunt resistance at different level of radiation are predicted. The main aim of internal model block is to develop simple and more realistic models for the photovoltaic peak power and at the same time to use the maximum of information provided by the manufacturer's data sheet. This block design the model of PV model with the information given in the input block and shows the results for any specified operating conditions .This model is very useful for real-time applications where the power needs to be monitored and it is referred to obtain the maximum power at different level of radiation. Thus the simulation block is designed using MATLAB/SIMULINK. Second block is ANN block, in that artificial neural network with specifically feed-forward back propagation algorithm is developed. ANN overcomes the limitations of the conventional approaches by extracting the desired information directly from the experimental data. The fundamental processing elements are neurons. Network is a parallel distributed information processing technique [11]. In this Configuration, networks are arranged in layers, with the first layer taking in inputs and the last layer producing outputs. The middle layers have no connection with the external world, and hence it is called hidden layers. Each layer is connected with Neurons in between them. The sizes of middle (hidden) layers are determined by trial and error methods. A typical ANN operation starts with the training stage, which modifies the connection weights in some orderly fashion using a suitable learning method [12]. To train this network, back propagation algorithm based on experimental result is used. The space of input X of dimension n (Level of the input) is connected to the space of the output Y of dimension m (Level of the output) by the intermediary of a hidden level. This level has a fixed number of neurons, but which varies from a study to another according to the complexity of the problem. In the present study, architecture is considered; multi-layer perception networks with back propagation of the gradient. They are probably one architectures most current and simplest non-linear network. The capacities of modeling of these networks are analyzed. The multi-layer networks are composed of a input layer whose neurons code the information presented at the network, of a variable number of internal layers called "hidden" and of a input layer containing as many neurons as of desired responses. The neurons of the same layer are not connected between them. The training of these networks is supervised. The algorithm used during this training is known under the name of method of Back propagation learning BPL. This method of training is divided into two phases: A phase of propagation, which consists in presenting a configuration of input at the network then to propagate this input gradually layer of input to the output layer while passing by the hidden layers. A phase of back propagation, which consists, after the process of propagation, to minimize the error made on the whole of the examples presented, error considered as a function of the synaptic weights. This error represents the sum of differences squared between the calculated responses and those desired for all the examples contained as a whole of training generally, the stages of construction and validation of the neural networks are divided into phases: the input of the networks, the output of the networks and the tests of the networks with real time data. In our applications, the MATLAB software, "neural networks Toolbox" is used in order to carry out these stages. For our application, irradiation and temperature are utilized as the input information of ANN Network and in the output layer it predicts maximum performance of PV system power generation and the simulated characteristics are explained using 3D nomogarm. Irradiation and temperature are some of the factors which affect panel power output .They are Output rating of the solar panel, intensity of solar radiation and hours of available light. The average current I_{avg} generated by the solar panel is proportional to the ratio of the actual solar radiation to 1000 W/m^2 multiplied by I_{sc} (short circuit current). Therefore, the average current would be

$$I_{avg} = (G/1000)\ W/m^2 * \ I_{sc} \quad (1)$$

G = Level of radiation (100 to 1000 W/m²)

In this proposed model, the radiance level is considered to vary from 100 to 1000 W/m².To show the effect of irradiance on the performance of a module the temperature is kept fixed at 25º C and the values of irradiance are changed to different values. It is quite clear that irradiance has a major effect on the short circuit current and indeed the relationship between irradiance and the short circuit current is a linear one [13]. Power was also calculated at different levels of irradiation. It are seen from the current and power characteristics, the nonlinear nature of the PV array is apparent. The proposed model can be applied to any brand of PV modules by setting the parameters properly. Thus the power produced by the cell in Watts can be easily calculated along the I-V sweep by the equation. [14]

3 RESULTS AND DISCUSSIONS

3.1 REPRESENTATION OF PANEL POWER OUTPUT IN 3D NOMOGRAM

Using this proposed model, we can create a database of the panel power output for various rated panels. This type of data analysis is more suitable in making initial determination of panel power output of the particular rated panel. The database is represented in a 3D nomogram. These types of nomogram are very useful to the quality control engineer for analyzing the power output of group of panels. The 3D Nomogram of PV Module with the Panel Power output of 40, 80 120 and 160 watts are simulated by considering its short circuit current at different level of radiation. The simulated panel power output is compared with the field data.

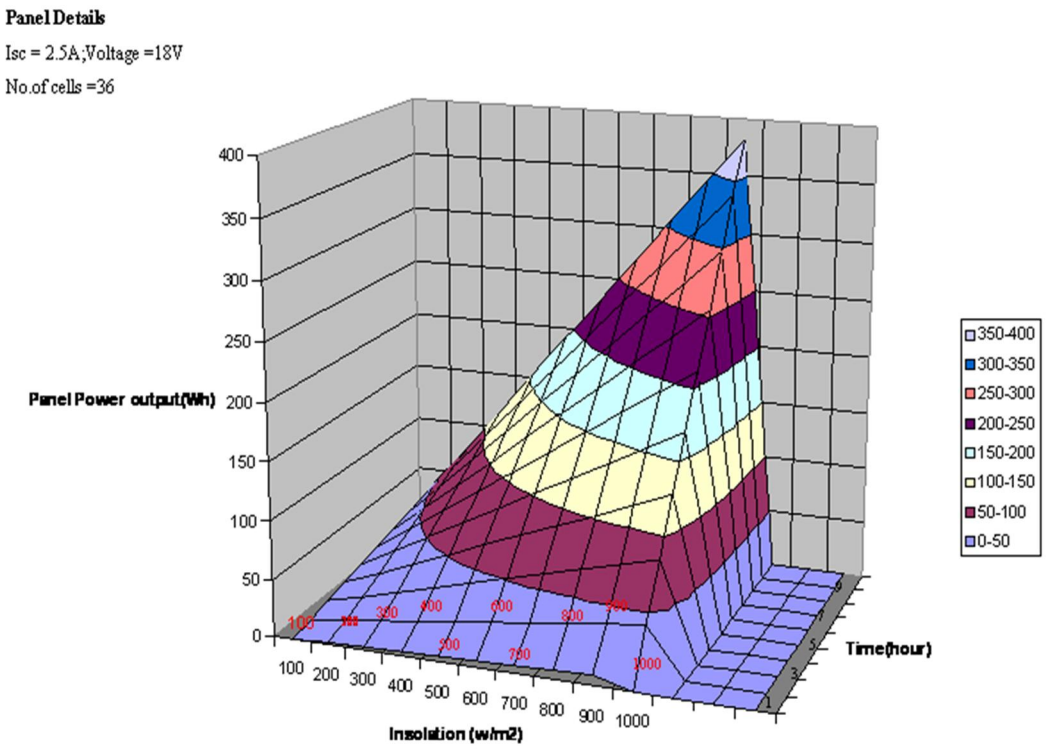


Fig. 2. 3D Nomogram of PV Module with the Standard Panel Power output of 40

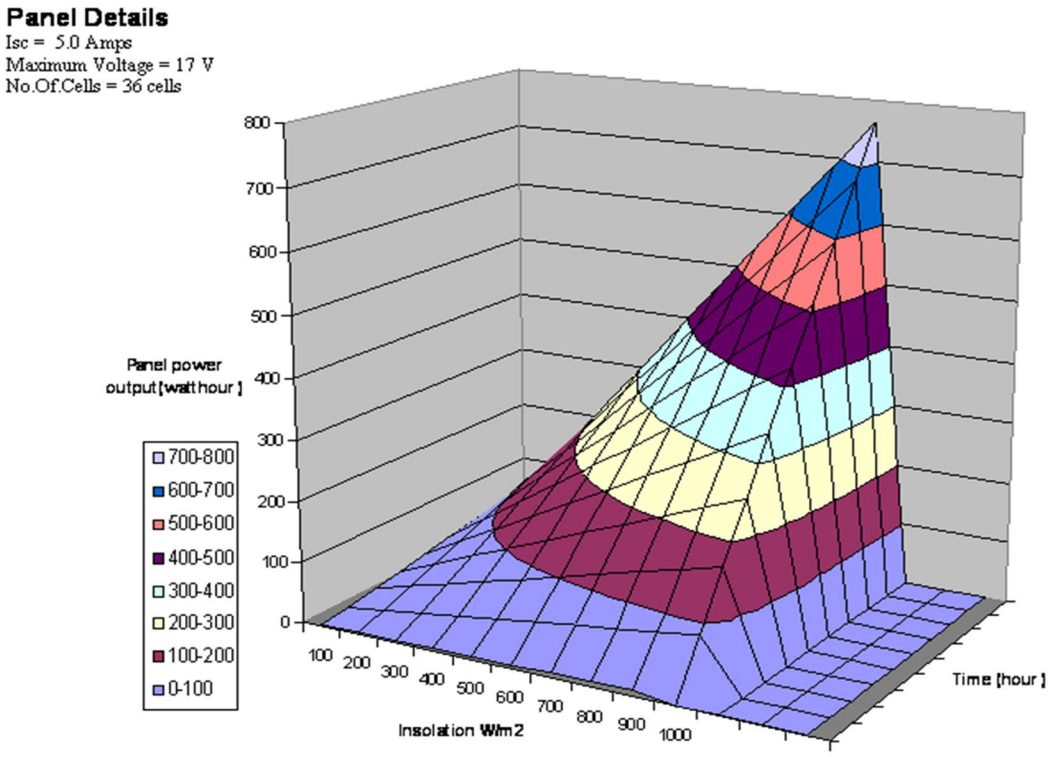


Fig. 3. 3D Nomogram of PV Module with the Standard Panel Power output of 80

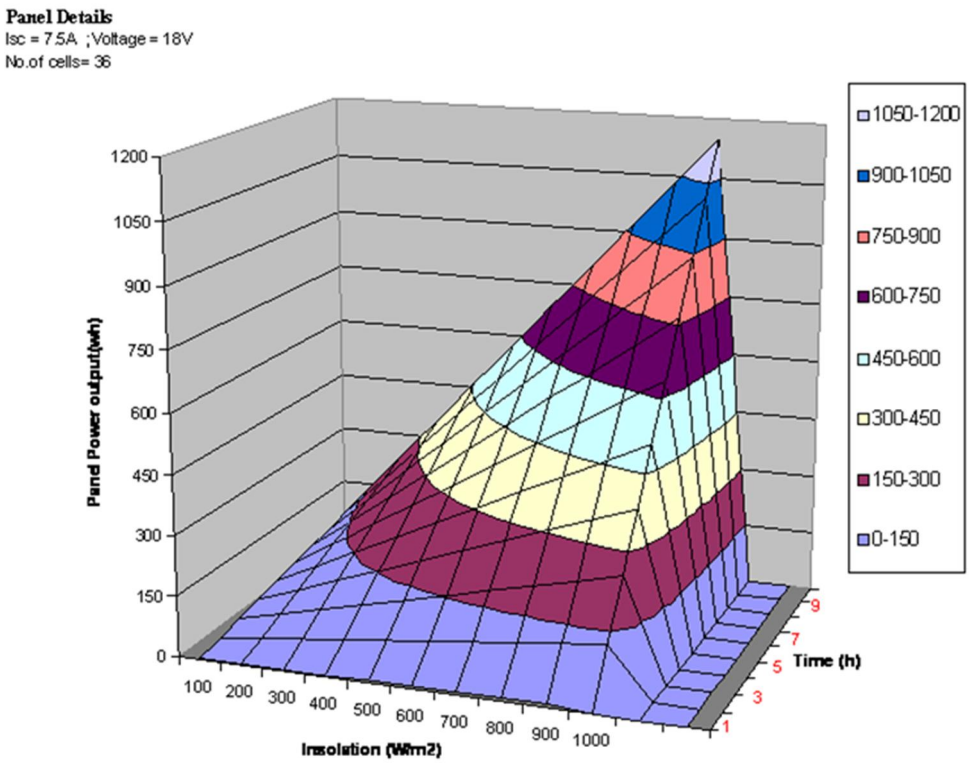


Fig. 4. 3D Nomogram of PV Module with the Standard Panel Power output of 120

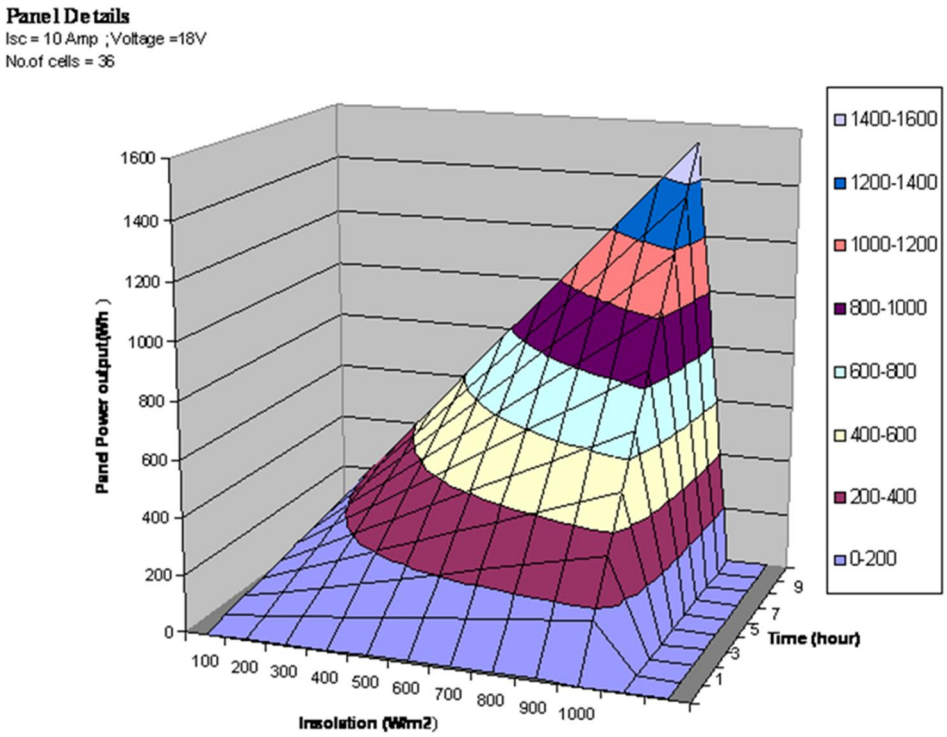


Fig. 5. 3D Nomogram of PV Module with the Standard Panel Power output of 160

3.2 Design of Real-Time Simulator for Power Generation of PV Module

Based on the simulation study using MATLAB; a database device for PV model is being designed to monitor parameters of PV standalone system and to calculate the power generated by the total solar radiation falling on a panel during insolation per day. This database device is designed using microcontroller

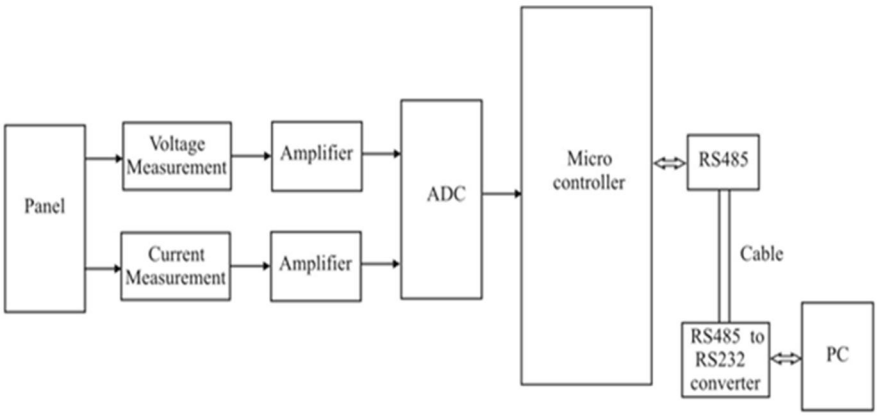


Fig. 6. Block diagram of database device

The data base device consist of sensor for voltage and current measurement, ADC device, Control system using microcontroller and a Monitoring PC connected via RS 232. It contains two LEDs: RED and GREEN to display the status of the watt hour generated per day. Every solar panel is rated by their peak power or Watt-hour. So, for calculating the peak power of the panel, the watt - hour value for the whole day will be compared with the Wh generated per day,

1. Once if it is lower than the required Wh, RED LED will glow otherwise GREEN LED will glow.
2. On the next consequent days, when the Wh lost is achieved in addition to the present day Wh, and then GREEN LED will glow.

In our study we have used two sub-panels each consists of 36 silicon cells. Voltage at maximum power is 17V, short circuit current (I_{sc}) is 2.5A, maximum power is 40+10%, 11 Watt CFL is used as the load which has operating voltage of 12V ,with current consumption of 1.13 A ,dusk to dawn is 12 h/day and discharge output = 1.13 *12 = 13.56 Ah /day. Provision is made to collect battery temperature (both ambient and cabinet), current, voltage, energy input and intensity of solar radiation using data logging system. Average current over particular insolation hours is calculated using the formula (1) and panel power output is calculated for different levels of radiation by knowing the average current and voltage as per designed model. The radiation in terms of lux*1000 is in turn converted to W/m^2. Likewise the Panel Power output is calculated for all sunless days and sunny days. Error estimation method is carried out using sigma plot so as to get relationship between measured and predicted data. For a typical field data collected on cloudy (sun less) days, the maximum lux is $30,000/mm^2$ and sunny days the maximum lux is 90,000/mm. From the result it is seen that our model predicts to a larger extent the output of solar panel within experimental limitations.

Table 1. The results of panel power output and data analysis of PV standalone lighting model for various sunny and sunless days

No. Of Days	Panel voltage (V)		Panel Current (I)		Watt hour Wh	Ampere Hour Ah	Temperature (T)	Battery Voltage (V)
	MAX	MIN	MAX	MIN				
1	14.27	13.03	4.3	0.01	272.23	20.25	35.93	12.92
2	14.1	12.56	4.1	0.01	260.2	20.16	36.13	13.08
3	14.07	12.99	3.72	0.01	251	18.33	37.787	13
4	14.15	12.55	3.89	0.01	311	22.91	35.88	13.09
5	14.21	13.11	3.76	0.01	272.9	19.75	36.09	13.11
6	14.19	12.65	3.754	0.01	253.21	18.47	37.67	13
7	14.18	12.62	3.78	0.01	255.11	18.49	37.71	13.12
8	14.22	12.61	3.7	0.01	259.11	18.52	37.11	13.12
9	14.21	12.69	3.72	0.01	267.78	18.88	38.15	13.12
10	15.3	13.4	3.84	0.01	243.38	17.087	36.62	13.06
11	15.29	13.32	3.97	0.01	254.59	17.99	36.96	13.06
12	14.68	13.15	2.9	0.01	112.7	8.05	37.7	13.06
13	15.35	13.4	3.6	0.01	246.71	18	37.35	13.1
14	15.21	13.28	3.55	0.01	245.52	17.91	37.81	13.08
15	15.05	13.11	3.41	0.01	231.11	17.8	37.88	13.09
16	15.31	13.37	3.55	0.01	229.08	15.94	36.33	13.05
17	15.28	13.28	3.48	0.01	221.11	15.85	36.11	13.09
18	15.31	13.17	3.5	0.01	220	15.81	35.58	13.08
19	14.51	13.37	3.75	0.01	162.25	11.51	36.19	13.05
20	15.32	13.46	3.86	0.01	244.1	17.07	36.62	13.09
21	16.18	13.52	3.11	0.01	173.41	11.86	36.22	13.04
22	13.93	13.66	3.12	0.01	33.66	2.41	34	13.02
23	15.88	13.33	3.15	0.01	168.21	11.21	37.11	13.03
24	14.08	13.36	3.29	0.01	31.09	2.38	34.49	13.03
25	15.49	13.32	3.97	0.01	254.59	17.36	36.96	13.06
26	16	13.51	3.59	0.01	180.2	12.23	35.31	13.04
27	15.3	13.11	3.55	0.01	240.88	17.89	38.72	13.07
28	14.18	12.44	3.65	0.01	258.71	20.01	38.72	13.02
29	14.22	12.34	3.49	0.01	222.11	14.98	36.77	13
30	14.29	12.29	3.29	0.01	223.14	15.09	37.11	13

4 CONCLUSION

In this study, panel current generated at any time are simulated and the behavior of a PV module at different level of radiance is designed. This physical modeling technique does not require the knowledge of internal system parameters, involve less computational effort and offer a compact solution for multivariable problems. The proposed model has the advantage of using the information provided by manufacturer data sheets. Learning from the results of PV module characteristic, a database device is designed and implemented in our solar local climatic data. This type of simulation can save a large amount of time and money. We can often explore a large number of scenarios very quickly. The result will help system engineers to choose the right control strategies for panel and batteries.

ACKNOWLEDGMENT

The authors are thankful to the Ministry of New and Renewable Energy, Government of India and council of scientific and industrial research for the support to carry out this work.

REFERENCES

[1] L. Fahrenbruch., R. H. Bube., *Fundamentals of Solar Cells*, San Francisco, CA: Academic, 1983.
[2] Gow & C. D. Manning, "Development of a model for photovoltaic arrays suitable for use in simulation studies of solar energy conversion systems," In *Proc. 6th International Conference on Power Electronics and Variable Speed Drives*, pp. 69–74, 1996.
[3] Bo Anderson, Catella Generics, Staffan Ulvönäs, Bengt Perers, and Vattenfall Utveckling, *Battery Guide for Small Stand Alone PV Systems,* IEA PVPS Task III, 1997.
[4] Robbins T.P., "Simulation of a stand-alone power system with battery storage," in: *Annual Conference and Annual General Meeting of Anzes, Australian National University*, 27, 1987.
[5] Carl Johan Rydh, "Energy analysis of batteries in photovoltaic systems," Part II: Energy return factors and overall battery efficiencies, *Energy Conversion and Management* 46, (11-12) 1980-2000, 2005.
[6] Singh, V.N., and Singh R.P., "A method for the measurement of solar cell series resistance," *J. Phys. D Appl. Phys.* 16, 1823–1825, 1983.
[7] Altas, I. H., and Sharaf A.M., "A Photovoltaic Array Simulation Model for Matlab-Simulink GUI Environment Clean Electrical Power," *ICCEP '07*. 341 – 345, 2007.
[8] Sera, Dezso, Teodorescu, Remus, and Rodriguez, Pedro, "PV panel model based on datasheet values; Industrial Electronics," *ISIE 2007,* 2007. ISBN: 978-1-4244-0755-2
[9] Krisztina Leban, and Ewen Ritchie, "Selecting the Accurate Solar Panel Simulation Model," *NORPIE/2008, Nordic Workshop on Power and Industrial Electronics*, June 9-11, 2007.
[10] Wook Kim and Woojin Choi, "A novel parameter extraction method for the one-diode solar cell model," *Solar Energy*, Vol. 84, No. 6, pp. 1008-1019, 2010.
[11] Philip D.Wasserman, *Neural Computing: Theory and Practice*, Van Nostrand Reinhold, New York, USA .43-46, 1989.
[12] Robert Hecht Nielsen, *Neuro Computing*, Addison-Wesley, New York, USA .124-126, 1996.
[13] D.R Clark, S.A. Klein, and W.A. Bckman, "A method for estimating the performance PV systems," *Solar Energy*, vol. 33, no. 6, pp. 551-555, 1984.
[14] E. I. Ortiz-Rivera and F. Z. Peng, "Analytical model for a photovoltaic module using the electrical characteristics provided by the manufacturer data sheet," In *Proc. IEEE 36th Power Electronics Specialists Conference, PESC*, pp. 2087–2091, 2005.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 693-700

http://www.issr-journals.org/ijias/

A classification approach using SVM to detect magnetic inrush in power transformers

V. Thamarai Selvi and V. Malathi

Electrical and Electronics Engineering,
Anna University Regional Centre Madurai,
Madurai, Tamil Nadu, India

ABSTRACT: In Order to avoid mal operation of differential relay in transformer it is essential to distinguish between inrush and fault conditions. For accurate discrimination between inrush and fault current SVM technique is proposed. The merit of this method is demonstrated by simulation of different faults and switching conditions using MATLAB/SIMULINK. The inrush current values are obtained by varying the switching angle and the fault currents are obtained by varying the fault resistance. The Proposed method is tested on a 3000MVA, 230 kV Y-Y connected transformer by varying fault resistance, and switching angle. The performance of SVM is compared in terms of classification accuracy. The accuracy obtained using SVM is found to be more than other methods such as neural networks, ANFIS, etc. The results obtained with SVM are far better than other methods earlier used. SVM is preferred here over other methods because it is based on structural risk minimization whereas neural networks and ANFIS are empirical based. Moreover this method seems to be very effective for modern transformers with high harmonic contents and it requires less training. A SVM based protective field programmable gate array relay logic can be implemented further in future which will be verified against the simulation results.

KEYWORDS: Inrush Current, Support Vector Machine, Fault current, Sequential minimization method, Radial basis function.

1 INTRODUCTION

Power transformer is one of the indispensable elements in power systems and thus protection of transformer play an imperative role. When the energization of transformer occurs with no load, inrush current flowing into the transformer will become excessive such that it will be eight to thirty times of full load current. This high current forces the relay to operate falsely. The relay provided has to work only for internal fault and inrush current and not for external fault and normal conditions.

In former protective schemes of transformer, it is confined based on transformer inductance during saturation of current transformer [1] and then algorithms have been developed based on second order percentage differential harmonic restraint concept [2]. Since inrush current usually contains a large second order harmonic component than internal fault, earlier transformer protection systems are planned to restrain during magnetic inrush phenomena by filtering this large second order harmonic component owing to saturation of current transformer or the presence of a dispersed shunt capacitance in a lengthy EHV transmission line to which the transformer may be connected [3]. In undeniable cases, the level of the second order harmonic in an internal fault current can be nearer to or bigger than that of inrush current. In addition, the second order harmonic components in inrush currents likely to be reasonably small in modern large power transformer because of enhancement in the power transformer core substance. In view of the above factors, many researchers accomplish their work to develop new algorithm for transformer protection [4]-[6]. However all these algorithms developed so far are either based on the transformer equivalent circuit model and/ or require some transformer data and thus may vulnerable to parameter variations. In [6], transformer is protected based on the active power flowing in to transformer, which is almost zero in case of energization. New methods have been employed using fuzzy logic and neural network for protection of power transformer [7]-[10]. In [8], neural network based schemes for protection of a single phase transformer has been carried out while applications of the neural network for protection of a three phase power transformer have been shown in [9] and [10].

Corresponding Author: V. Thamarai Selvi (thamsragav123@gmail.com)

However the Artificial Neural Networks used in existing systems are confined to particular power transformers, and would have to be trained again and again for other system which seems to take more time. Several protective schemes of power transformer with wavelet transform are also developed in [11]- [14]. In [15] SVM technique is applied to classify between inrush and fault of a 3 phase 35 MVA, 50 HZ, 132/11 KV Y/Y transformer and 132 KV transmission line. The combined discrete wavelet and SVM technique is applied in detecting and classifying the fault in transmission line [16]. Bus bars are also protected with the help of support vector machines [17].

In this paper, a modern approach is put forward for discrimination of L-G fault and inrush current in power transformer using Support Vector Machine (SVM). The system is tested on a 3000 MVA, 500/230 kV Y-Y connected transformer under variety of fault and inrush conditions. The SVM based approach discriminate between magnetic inrush and L-G fault in power transformer more accurately.

2 SUPPORT VECTOR MACHINES

As with any supervised learning methods training the network is very important. The SVM is therefore first trained and the trained network is used to classify or predict new data. In addition to obtain more accurate results various SVM kernel functions are used and the parameters of kernel functions must be tuned .

The main features of SVM are:

- The upper bound on the generalization error does not depend on the dimension of the space
- The error bound is minimized by maximizing the margin g.

Considering the binary classification task with data point xi(i = 1,2,…,m) having labels yi = ±1 and the decision function be

$$f(x) = sign(w.x + b) \tag{2.1}$$

Where w is the n dimensional vector and b is the scalar. The vector w and scalar b determines the position of the separating hyper plane. If the dataset is separable then the data will be correctly classified where yi (w.x+b)>0"i. Thus canonical hyper plane is such that w.x+b = 1 for closest points on one side and w.x+b = -1 for closest points on other side as in Fig. 2..For separating w.x+b = 0 the normal vector is w and hence, the margin is given by the projection of x1-x2 on to this vector. Since, w.x1+b = 1 and w.x2+b = -1, the margin is g = 1/||w||. To maximize the margin the task is, therefore, subject to the constraints y_i(w.x_i+b) = 0;and the learning task can be reduced to minimization of the primal Lagrangian

$$\min g(w) = \frac{1}{2(||w||)2} \tag{2.2}$$

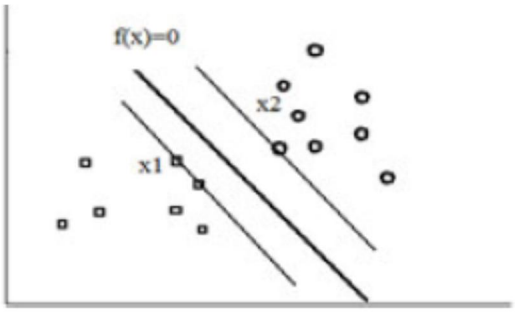


Fig. 1. Sample SVM classifier

A) Separable Data

An SVM classifies data by finding the best hyper plane that separates points of one class from those of the other class. The best hyper plane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyper plane that has no interior data points. The support vectors are the data points that are closest to the separating hyper plane; these points are on the boundary of the slab. The following figure illustrates these definitions, with + indicating data points of type 1, and – indicating data points of type –1.

Mathematical Formulation: The data for training is a set of points (vectors) xi along with their categories yi. For some dimension d, the xi ∈ Rd, and the yi = ±1. The equation of a hyper plane is

$$< w, x > + b = 0 \quad (2.3)$$

where w ∈ Rd, <w,x> is the inner (dot) product of w and x, and b is real.

$$yi(< w, xi > + b) \geq 1 \quad (2.4)$$

The following problem defines the best separating hyper plane. Find w and b that minimize ||w|| such that for all data points (xi,yi),

For mathematical convenience, the problem is usually given as the equivalent problem of minimizing <w,w>/2. This is a quadratic programming problem. The optimal solution w, b enables classification of a vector z as follows:

$$class(z) = sign(< w, z > + b) \quad (2.5)$$

B) Non separable Data

There are two standard formulations of soft margins. Both involve adding slack variables si and a penalty parameter C.

The L1-norm problem is:

$$\min_{w,b,s}\left(\frac{1}{2}(w, w) + C\sum_i s_i\right) \quad (2.6)$$

$$y_i\big((w, x_i) + b\big) \geq 1 - s_i \quad (2.7)$$

$$s_i \geq 0 \quad (2.8)$$

The L^1–norm refers to using s_i as slack variables instead of their squares.

The L^2 -norm problem is

$$\min_{w,b,s}\left(\frac{1}{2}(w, w) + C\sum_i s_i^2\right) \quad (2.9)$$

C) Sequential Minimal Optimization

Platt (1999) proposes to always use the smallest possible working set size, that is, two elements. This choice dramatically simplifies the decomposition method. Each successive quadratic programming sub problem has two variables. The equality constraint makes this a one dimensional optimization problem. A single direction search is sufficient to compute the solution.

3 SYSTEM STUDIED

The single line diagram of the system studied is shown in Fig.2.. Transformer and line configuration details are given in Appendix A. The test system is modeled and simulated using MATLAB/simulink.

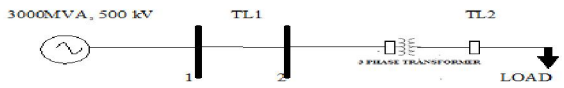


Fig. 2. Test system

3.1 PROPOSED ALGORITHM USING SVM

The schematic order of the fault classification scheme using SVM for power transformer is shown in Fig.2. MATLAB/ Simulink are used to simulate magnetizing inrush conditions due to change in switching angle and L-G fault conditions by varying fault resistance.

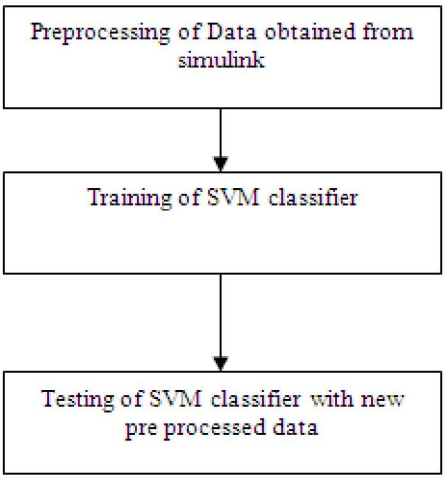


Fig. 3. Block diagram of the proposed method for discrimination between magnetic inrush and L-G fault in power transformer

The output of the SVM is such that the value "1" denotes L-G fault; the value "-1" denotes a magnetizing inrush current.

3.2 FEATURE EXTRACTION

The inrush current signals are obtained by varying the energization angle from 0 to $360^0$and the L-G fault current samples are generated by varying the fault resistance. Matlab/simulink is used to generate the signals.Fig.3-4 shows the magnetizing inrush current and L-G fault current samples.

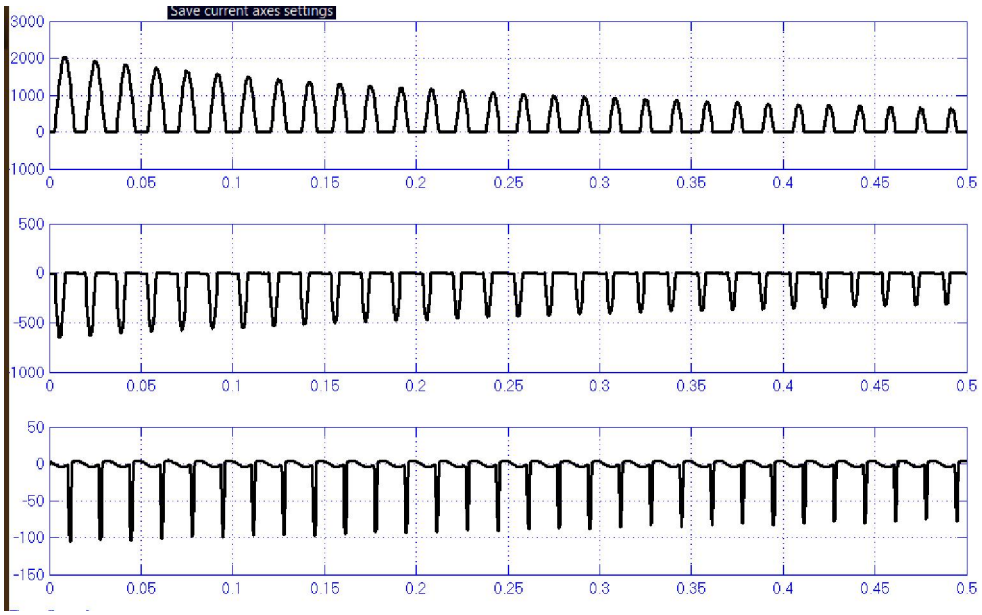


Fig. 4. Simulated three phase current waveforms for inrush current

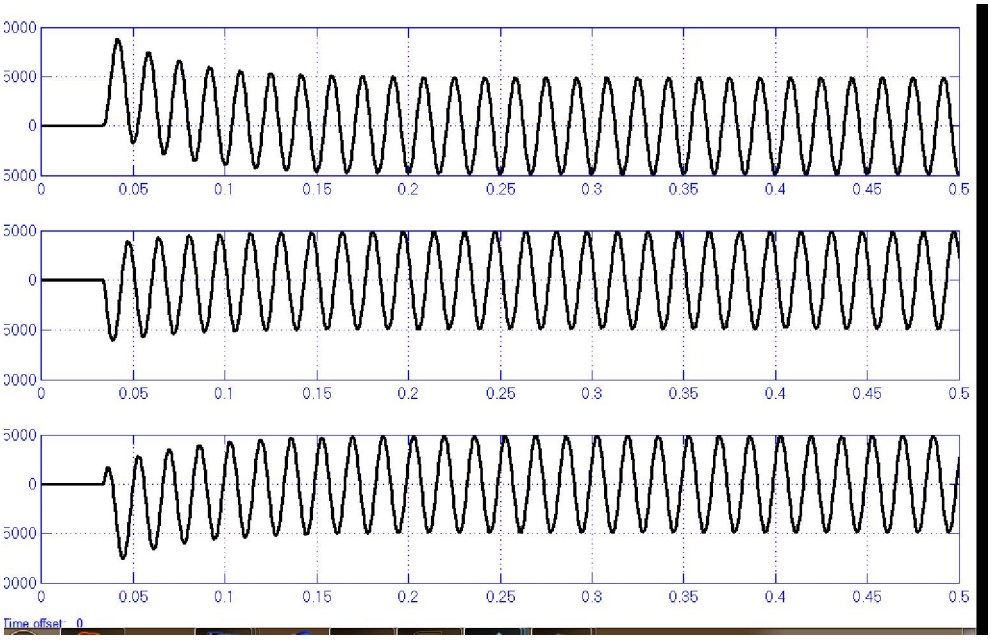


Fig. 5. Simulated three phase current waveforms for L-G fault current

3.3 SVM TRAINING AND TESTING

The SVM is trained for various training patterns of fault and inrush current. 2000 datasets are simulated for different fault and inrush conditions. The fault is simulated by varying fault resistance. The inrush current changes correspond to energization angles of 0^0 and 360^0 from phase-a voltage zero crossing. The 2000 datasets are used for testing the classifier with the SVM structure obtained in training. The same system is trained and tested with LMBPNN to compare the overall performance of SVM.

All the simulations for the SVM and LMBPNN algorithms are carried out in MATLAB 7 platform. In SVM, RBF kernel is used for fault classification of transient events. The simulations for SVM are carried out using SVM toolbox in MATLAB/simulink platform.

4 SIMULATION RESULTS AND DISCUSSIONS

An extensive series of studies are carried out in order to ascertain the overall performance of the two networks. The test sets (which not included as part of the training sets) are composed of over 2000 cases including different fault resistance and magnetizing inrush current. The SVM structure is created by training the SVM classifier with the data obtained from simulation.

Fig.6. shows that the points indicated with red colour denotes the inrush current and with green colour denotes the fault current used for training .The black line is the optimum hyper plane that separates the two data. SMO method is used to optimize the parameters of SVM kernel functions. After training with the simulation data of inrush and fault currents, SVM structure is created The SVM structure includes support vectors, alpha ,bias, kernel functions, support vector indices . The number of support vectors determines the overall performance of SVM classifier. Here the number of support vectors obtained is 15The alpha is the vector weights to support vectors. The sign of weights is positive for the inrush current and negative for fault current. Kernel function used here is radial basis function which maps training data into kernel space. Support Vector indices are the vectors of indices that specify the rows in training, the training data that were selected as support vectors after the data was normalized according to the auto scale argument. Then the SVM structure is used to classify the new sets of fault current and inrush current data. Fig.6. shows both training and testing details of SVM.As already mentioned the new sets of data were discriminated with the help of SVM structure created using the training data.Here the blue colour indicates the no of new set of correctly classified fault data and pink colour indicates the correctlclassified inrush current data. The accuracy is calculated using the formula A= (1-MCR)*100%.

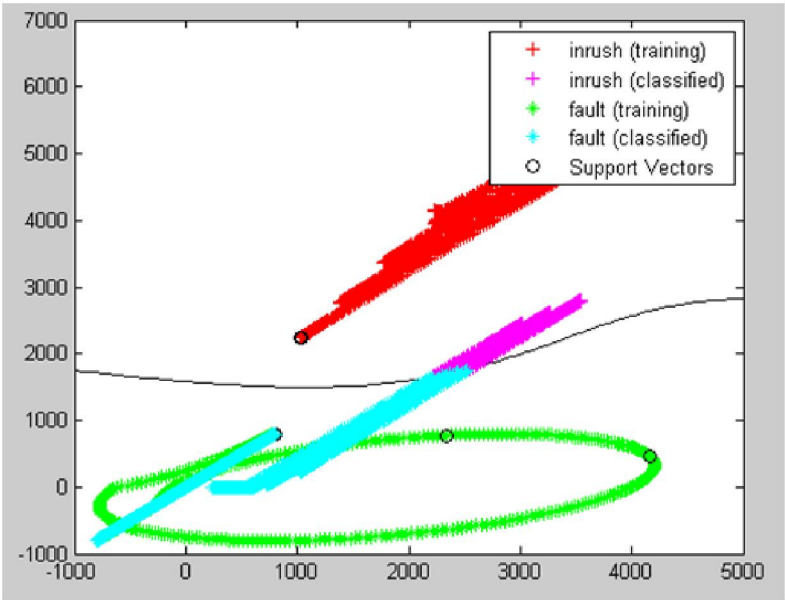


Fig. 6. SVM Classifier result with data

5 CONCLUSION

This study proposes a modern method for discrimination between magnetic inrush and L-G fault in power transformer using SVM. The method is done based on current signature verification. Therefore, for modern transformers with high harmonic components in internal fault current, this method seems to be more effective. The results obtained on simulated data of a three phase two winding Y-Y connected transformer showed that SVM classifier outperforms greatly in discrimination between magnetic inrush and L-G fault currents. In the future the discrimination of magnetic inrush and other types of internal faults using SVM can be done and to be implemented in hardware.

REFERENCES

[1] K. Ingaki, Higaki, M., Matsui, Y., et al., "Digital protection method for power transformer based on an equivalent circuit composed of inverse inductance", *IEEE Trans., on Power Delivery,* vol. 3(4), 1988, 1501-1510.

[2] M.A. Rahman, B. Jeyasurya, "A state-of-the-art review of transformer protection algorithms", *IEEE Trans., on Power Delivery,* vol. 3(2), 1988, 534-544.

[3] Pei Liu, O.P. Malik, Deshu Chen et al., "Improved operation of differential protection of power transformers for internal faults," *IEEE Trans., on Power Delivery,* vol. 7(1*),* 1992,1912-1919.

[4] T.S. Sidhu, M.S. Sachdev, H.C. Wood et all, "Design, Implementation and testing of a microprocessor-based high-speed relay for detecting transformer winding faults," *IEEE Trans., on Power Delivery*, 7(1), 1992, 108-117.

[5] T.S. Sidhu, M.S. Sachdev, "Online Identification of magnetizing inrush and internal faults in three-phase transformers," *IEEE Trans., on Power Delivery,* 7(4), 1992, 1885-1891.

[6] Yabe, K., " Power differential method for discrimination between fault and magnetizing inrush current", *IEEE Trans., on Power Delivery, vol. 12(3), 1997,*1109-1118.

[7] Wiszniewski, A., kasztenny, B., "A Multi-criteria differential transformer relay based on fuzzy logic", *IEEE Trans., on Power Delivery,* 10(4), *1995,* 1786-1792.

[8] Perez, L.G, Floechsig, A.J., Meador, J.L., and Obradovic, Z., "Training an artificial neural network to discriminate between magnetizing inrush and internal faults", *IEEE Trans., on Power Delivery*, 9(1), *1994,* 434-441.

[9] Pihler, J., Grcar, B., and Dolinar, D., "Improved operation of power transformer protection using artificial neural network", *IEEE Trans., on Power Delivery*, 12(3), 1997, 1128-1136.

[10] M.R. Zaman and M.A. Rahman, "Experimental testing of the artificaial neural network based protection of power transformers," *IEEE Trans.on Power Delivery,* 13(2), 1998, 510-517.

[11] Moises Gomez-Morante, Danies, W.Nicoletti, "A wavelet based Differential Transformer protection," *IEEE Trans.on Power Delivery,* 14(4), *1999,* 1351-1358.

[12] P.L. Mao and R.L. Aggarval, "A Novel approach to the classification of the transient phenomena in Power Transformer using combined wavelet transform and Neural Network", *IEEE Trans. on Power Delivery,* 16, 2001,654-660.

[13] Omar A.S. Youssef, "A wavelet based technique for discrimination between faults and magnetizing inrush currents in Transformers", *IEEE Trans. on Power Delivery,* 18*,* 2003, 170-175.

[14] Javad Faiz and S. Lotfi-Fard, "A Novel Wavelet based algorithm for discrimination of internal faults from magnetizing inrush currents in Power Transformer", *IEEE Trans. on Power Delivery*, 21, 2006, 1989-1996.

APPENDIX A

1) Transformer parameters:

Three phase 450 MVA 50 Hz 500/230 kV Y-Y connected windings with earth neutrals.

2) Transmission line parameters:

500 kV 100 km transmission lines.

BIBLIOGRAPHY

V Thamarai Selvi received her Bachelor of Engineering in Electrical and Electronics from E.G.S. Pillay Engineering college, Nagapattinam, Tamil Nadu, India. She is currently pursuing Master of Engineering in Power Management from Anna University Regional Centre Madurai, Tamilnadu, India. Her area of research is intelligent techniques application to power system protection.

V Malathi received her Bachelor of Engineering in Electrical and Electronics from College of Engineering Guindy, Anna University, Chennai, and Master of Engineering in Power Systems from Thiagarajar College of Engineering, Madurai, Tamilnadu, India. She received her Ph.D degree from Ann a University, Chennai. She published over 10 papers in the field of Intelligent techniques applications to power systems.

She is a member of Institution of Engineers (India) and Life Member of the Indian Society for Technical Education.

She is working as Professor in Department of Electrical and Electronics Engineering, Anna University, Regional Centre Madurai, Tamilnadu, India. Her area of research is intelligent techniques application to power system protection.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 701-713

http://www.issr-journals.org/ijias/

Effect of Near-orthogonality on Random Indexing Based Extractive Text Summarization

Niladri Chatterjee[1] and Pramod K. Sahoo[2]

[1]Department of Mathematics,
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi 110016, India

[2]Institute for Systems Studies and Analyses,
Defence Research and Development Organisation,
Metcalfe House Complex, Delhi 110054, India

ABSTRACT: Application of Random Indexing (RI) to extractive text summarization has already been proposed in literature. RI is an approximating technique to deal with high-dimensionality problem of Word Space Models (WSMs). However, the distinguishing feature of RI from other WSMs (e.g. Latent Semantic Analysis (LSA)) is the near-orthogonality of the word vectors (index vectors). The near-orthogonality property of the index vectors helps in reducing the dimension of the underlying Word Space. The present work focuses on studying in detail the near-orthogonality property of random index vectors, and its effect on extractive text summarization. A probabilistic definition of near-orthogonality of RI-based Word Space is presented, and a thorough discussion on the subject is conducted in this paper. Our experiments on DUC 2002 data show that while quality of summaries produced by RI with Euclidean distance measure is almost invariant to near-orthogonality of the underlying Word Space; the quality of summaries produced by RI with cosine dissimilarity measure is strongly affected by near-orthogonality. Also, it is found that RI with Euclidean distance measure performs much better than many LSA-based summarization techniques. This improved performance of RI-based summarizer over LSA-based summarizer is significant because RI is computationally inexpensive as compared to LSA which uses Singular Value Decomposition (SVD) - a computationally complex algebraic technique for dimension reduction of the underlying Word Space.

KEYWORDS: Word Space, Random Indexing, index vector, context vector, near-orthogonal, PageRank.

1 INTRODUCTION

Random Indexing (RI) along with PageRank based algorithm for extractive text summarization has already been proposed in literature by Chatterjee and Mohan [1]. The overall scheme proposed therein is based on the following steps:

- Representation of words using fixed dimension (d, say) ternary index vectors, containing 0, +1 and -1;
- Aggregation of the word vectors first into context vectors, and then from there to sentence vectors;
- Construction of a document graph from the sentence vectors by establishing links between nodes representing sentences;
- Application of PageRank algorithm to rank the sentences according to their importance;
- Finally, selection of the desired number of sentences to form the extractive summary of the document.

Random Indexing [2], [3] is an approximating technique to deal with high-dimensionality problem of Word Space Models (WSMs) [3], [4]. The advantage of this approach is that neither it requires any language-specific resources (e.g. on-line dictionary, thesaurus), nor does it depend on any informed heuristics of the underlying language. A major difference between RI with the most commonly used WSM based approach viz. Latent Semantic Analysis (LSA) [5], [6] is that LSA uses orthogonal unary vectors to represent words of a document, and builds a semantic space over it. On the other hand, in an RI-

Corresponding Author: Niladri Chatterjee (niladri.iitd@gmail.com)

based Word Space the index vectors are not orthogonal. As discussed in section 4, there is some probability of two distinct index vectors being non-orthogonal to each other. This property is called *near-orthogonality* of the index vectors. This near-orthogonality property of index vectors is the key to the RI-methodology as it helps in reducing the dimension of the Word Space before building the context information of a text passage. A study of near-orthogonality is therefore considered primary before advocating RI as a viable alternative to other WSM based approaches for extractive summarization.

A preliminary study in [7] shows that near-orthogonality of index vectors has an effect on the quality of summary produced by RI-based summarization schemes. The primary focus of the present work is to conduct an in-depth study of the effect of near-orthogonality of index vectors on the quality of summaries produced by RI with PageRank based scheme which we term as Random Indexing based Summarization (RISUM). In this study we experimented RISUM with both angular and linear proximity measures, used a larger dataset than [7], and utilized ROUGE [8] metrics to measure the quality of the summaries.

The rest of the paper is organized as follows. Section 2 gives a brief description of LSA-based summarization techniques proposed in literature. In section 3 Random Indexing and RISUM approach for extractive summarization has been presented. Section 4 elucidates the concept of near-orthogonality in an RI-based Word Space. Section 5 deals with experimental details, findings on effect of near-orthogonality on the performance of RISUM and comparative performance evaluation of LSA-based summarizers with RISUM. Section 6 presents concluding remarks.

2 LSA-BASED TEXT SUMMARIZATION

LSA is a Word Space implementation technique built upon the distributional information of words over a text passage. It uses Singular Value Decomposition (SVD); an algebraic dimension reduction technique, to extract the contextual information between the text units (words, sentences, or paragraphs) of the passage. In this technique each distinct word in a document is represented by a distinct unary vector. The dimension of each word vector is equal to the number of distinct words in the document. Each sentence in the document is represented by a sentence vector, which is constructed by taking the weighted algebraic sum of the word vectors. Suppose a document consists of m distinct words $w_1, w_2, \cdots, w_m$; whose word vectors are m-dimensional unary vectors $\vec{w}_1 = [1\ 0\ 0\ \cdots\ 0]^T$, $\vec{w}_2 = [0\ 1\ 0\ \cdots\ 0]^T$,..., $\vec{w}_m = [0\ 0\ 0\ \cdots\ 1]^T$ respectively. If the document consists of n number of sentences $s_1, s_2, \cdots, s_n$ and the weight assigned to the word w_i in sentence s_j is f_{ij}, then the sentence vector of sentence s_j is given by $\vec{s}_j = \sum_{i=1}^{m} f_{ij}\vec{w}_i$. In this way the document can be represented as a words-by-sentences matrix $S = [\vec{s}_1\ \vec{s}_2\ \cdots\ \vec{s}_n]$ of size $m \times n$. Then SVD is performed on matrix S, so that it is factored into a product of three matrices: $S = U\Sigma V^T$; where columns of U are the left singular vectors of S, Σ is a diagonal matrix consisting of singular values of S, and columns of V are the right singular vectors of S. The latent semantic structure of document represented by matrix S can be derived from its factors U, Σ and V [9].

A number of weighting schemes to construct sentence vectors from unary word vectors have been proposed in literature. Ozsoy *et al.* [10] gives a brief description of some of these weighting schemes, which are: (i) word frequency, (ii) binary, (iii) *Tf-Idf* (term (word) frequency-inverse document frequency), (iv) log entropy, (v) root type, and (vi) modified *Tf-Idf*.

Important sentences of a document can be extracted by exploiting the factors U, Σ and V of S. Some of the significant contributions in this regard are Gong and Liu [11], Steinberger and Ježek [12], Murray *et al.* [13] and Ozsoy *et al.* [10], [14]. A comparative study on performance evaluation of these methods were conducted in Ozsoy *et al.* [10], [14] by considering different weighting schemes as mentioned above. Sahlgren [2], [3] criticized LSA for being both computationally expensive and requiring the formation of a full co-occurrence matrix and its decomposition before any similarity computation can be performed. On the other hand, RI allows the incremental building of the semantic space by creating a short index vector for each unique context, and producing the context vector for each word by summing index vectors for each context as one scan through the text.

3 RANDOM INDEXING AND RISUM

This section gives a brief introduction of RI and the summarization scheme RISUM.

3.1 RANDOM INDEXING

RI initially assigns each distinct word in the document a unique and randomly generated vector called the *index vector* of the respective word. These index vectors are sparse, high-dimensional, and ternary. Each index vector of dimension d consists of a large number of '0's and a small number (ε) of '+1's and '-1's with the following probabilities:

$$\begin{cases} +1 \text{ with probability } \dfrac{\varepsilon}{2d} \\ 0 \text{ with probability } \dfrac{d-\varepsilon}{d} \\ -1 \text{ with probability } \dfrac{\varepsilon}{2d} \end{cases} \quad (1)$$

The major advantage of using these index vectors is that they can handle large Word Spaces very efficiently. For example, in LSA vectors of dimension $d = 1000$ can cover a vocabulary of 1000 words only, whereas the same dimension of 1000 with one '+1' and one '-1' can construct 999000 (1000×999) index vectors, and hence can cover a vocabulary of 999000 (~ 10^6) words.

In order to represent the semantics of a document it is assumed that the intended semantics of a word in a document can be found from the context words of the document. This is done by computing the *context vector* for each distinct word. Initially the context vector of each distinct word in the document is initialized to the d-dimensional null vector. Every time a word w_k occurs in the document, the index vectors of the words in its context window are added to its context vector. For example, if a $2 + 2$ sized context window (i.e. a context window spread over two words on either side of the focus word) represented by $[(w_{k-2}w_{k-1})w_k(w_{k+1}w_{k+2})]$ is considered, then the context vector of w_k would be updated as:

$$\mathcal{C}(w_k) := \mathcal{I}(w_{k-2}) + \mathcal{I}(w_{k-1}) + \mathcal{C}(w_k) + \mathcal{I}(w_{k+1}) + \mathcal{I}(w_{k+2}) \quad (2)$$

where $\mathcal{I}(w_k)$ and $\mathcal{C}(w_k)$ respectively are the index vector and context vector of the word w_k. Thus the context vectors of all words in the document are built incrementally by scanning their occurrences in the document.

3.2 RANDOM INDEXING BASED SUMMARIZATION

The major steps of RISUM are as follows:

- Mapping of words into the RI-based Word Space;
- Mapping of sentences into the RI-based Word Space;
- Representing the document as a proximity graph;
- Summary generation.

Subsections 3.2.1 to 3.2.4 discuss these steps in detail. Before being subjected to summarization, a document is pre-processed to identify sentence end-markers and *content words*[1]. The content words of the document are identified by using a list of stop words. We used 'smart_common_words.txt'[2] file which consists of 598 stop words. Then Porter stemming algorithm [15] is used to remove the common morphological and inflectional endings from the words.

3.2.1 MAPPING OF WORDS INTO THE RI-BASED WORD SPACE

Initially each distinct word in the document is assigned a unique index vector of dimension d with '+1's being placed randomly anywhere in the upper half, and the '-1's being placed randomly anywhere in the lower half positions. The context vector for each content word is then constructed by considering a $2 + 2$ sized context window. First, the context vector of each distinct content word in the document is initialized to the d-dimensional null vector. The context of a content word is restricted within the underlying sentence in which the word had occurred. Every time a content word w_k occurs in the document, its context vector $\mathcal{C}(w_k)$ is updated using equation (2). However, to give more weightage to the words nearer to the focus word we have used a weight vector $[0.5 \quad 1 \quad 1 \quad 0.5]$ which assigns weight 1 to the words adjacent to the focus word and weight 0.5 to the words which are at a distance 2.

[1] *Content words are the key words of a sentence. They are the important words that carry the meaning or sense.*

[2] *ROUGE, the evaluation metric used by DUC also uses the same file for listing stop words.*

3.2.2 MAPPING OF SENTENCES INTO THE RI-BASED WORD SPACE

The sentences of the document are mapped into the Word Space by constructing the sentence vectors using equation (3) below:

$$\mathcal{S}(s_j) = \frac{1}{m_j}\sum_{i=1}^{m_j}\left(\mathcal{C}(w_{ij}) - \mathcal{O}\right) \tag{3}$$

where,

- $\mathcal{S}(s_j)$ is the sentence vector of the j-th sentence;
- $\mathcal{C}(w_{ij})$ is the context vector of the i-th content word of j-th sentence;
- m_j is the number of content words in the j-th sentence;
- $\mathcal{O}$ is the central theme of the document computed as arithmetic mean of context vectors of content words of the document.

Subtraction of the mean vector from the context vector in equation (3) reduces the magnitude of the context vectors close in the direction to the central theme of the document, and increases the magnitude of context vectors which are almost in the opposite direction from the central theme. This reduces the influence of the commonly occurring words, such as the auxiliary verbs, articles, on the sentence vector [16].

3.2.3 REPRESENTING THE DOCUMENT AS A PROXIMITY GRAPH

The whole document is represented as a proximity graph, where the nodes of the graph represent the sentences of the document, and weighted edges represent the proximity between sentences. The proximity between any two sentences can be calculated in many different ways. In our implementation we have used two generic schemes: the *angular distance*, and the *linear distance*. In particular, we used (i) Cosine dissimilarity measure as angular distance, and (ii) Euclidean distance to measure the linear distance. Equations (4) and (5) give mathematical formula for the two proximity measures, respectively:

- Cosine dissimilarity measure: $c_{ij} = 1 - \frac{\sum_{k=1}^{d}(s_{ik}\cdot s_{jk})}{\sqrt{\sum_{k=1}^{d} s_{ik}^2}\cdot\sqrt{\sum_{k=1}^{d} s_{jk}^2}}$ (4)

- Euclidean distance measure: $e_{ij} = \sqrt{\sum_{k=1}^{d}\left(s_{ik} - s_{jk}\right)^2}$ (5)

Where $\mathcal{S}(s_q) = [s_{q1} \quad s_{q2} \quad \cdots \quad s_{qd}]$ is the q-th sentence vector.

3.2.4 SUMMARY GENERATION

The weighted PageRank algorithm [17] has been used to get rid of the redundant information in the text by removing the sentences of less importance. If $G = (V, E)$ be an undirected graph with the set of nodes V and set of edges E, then the weighted PageRank of a node v_i denoted by $PR^W(v_i)$ is defined as:

$$PR^W(v_i) = (1-\tau) + \tau \sum_{v_j \in \{v_i : \{v_i, v_j\} \in E\}} \frac{\omega_{ij} PR^W(v_j)}{\sum_{v_k \in \{v_j : \{v_j, v_k\} \in E\}} \omega_{jk}} \tag{6}$$

where ω_{ij} is the weight associated with the undirected edge $\{v_i, v_j\}$ and $\omega_{ij} = \omega_{ji}$ for all i, j. τ is a parameter chosen between 0 and 1 and set to 0.85 as per the recommendation of Brin and Page [18].

Iterative application of the weighted PageRank algorithm on the proximity graph makes the node weights converge. In case of cosine dissimilarity measure, heavily weighted nodes are considered for summary generation; while in case of Euclidean distance measure, light-weight nodes are picked up for the summary.

In our implementation of the summarizer we made two deviations from the one proposed by Chatterjee and Mohan [1].

1. We put restrictions on the placement of '+1's and '-1's in an index vector (see subsection 3.2.1). We found that random positioning of '+1's and '-1's may result in creating two index vectors having '+1' and '-1' occurring at the same coordinate position. This in turn creates erroneous context vectors as '+1' and '-1' cancel each other. To avoid this situation we restrict the placement of '+1's in the upper half positions and '-1's in the lower half

positions of an index vector. This restriction ensures that addition of index vectors while forming a context vector do not end up in producing a null vector; hence the semantics is not misrepresented.

2. Instead of cosine similarity we have used dissimilarity measure. The metric used by them is:

$$C_{ij} = \frac{\sum_{k=1}^{d}(s_{ik} \cdot s_{jk})}{\sqrt{\sum_{k=1}^{d} s_{ik}^2} \cdot \sqrt{\sum_{k=1}^{d} s_{jk}^2}} \quad (7)$$

However, we found that this similarity metric maps the edge weight between any two nodes into the interval [-1, 1]. It is therefore possible that with a mixture of positive and negative edge weights, PageRank value of every node will diverge to either ∞ or $-\infty$. For example, if the weighted PageRank algorithm with $\tau = 0.85$ is applied iteratively on a 3-node undirected graph, where:

- Edge weights are: $\omega_{12} = \omega_{21} = -0.5,\ \omega_{13} = \omega_{31} = 0.9$, and $\omega_{23} = \omega_{32} = 0.7$;
- Initial PageRank of all the three nodes are equal to 0.1;

Then $PR^W(v_1) \to -\infty$, $PR^W(v_2) \to \infty$, and $PR^W(v_2) \to \infty$. This had been an inherent drawback of the earlier scheme. We have not computed theoretically (or estimated experimentally) the probability of occurrence for this divergence. However, in our experiments with the DUC 2002 [19] datasets this divergence had occurred very frequently. This problem has now been rectified by taking the simple linear transformation of the similarity measure: $c_{ij} = 1 - C_{ij}$ (see equations (4) and (7)), which now confines edge weights in the interval [0, 2], and thereby annihilates completely any chance of eventual divergence of the weighted PageRank algorithm.

4 NEAR-ORTHOGONALITY IN A RI-BASED WSM

Let $\mathbb{W}_{RI}$ be a RI-based Word Space consisting of a set of m d-dimensional index vectors $\mathbb{I} = \{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_m\}$. Any two index vectors $\mathcal{I}_j$ and $\mathcal{I}_k$ are said to be orthogonal if their dot product is zero, i.e., $\mathcal{I}_j \cdot \mathcal{I}_k = 0$. This happens when positions of '+1' and '-1's of both the index vectors are different. That means if for any $p \in \{1,2,\cdots,d\}$, the p-th coordinate of $\mathcal{I}_j$ contains either '+1' or '-1', then the p-th coordinate of $\mathcal{I}_k$ must contain a zero, and vice-versa. Since an index vector consists of large number of zeros, there is a high probability that any two index vectors chosen randomly are still orthogonal. However, there is a small probability that the two randomly chosen index vectors are not orthogonal. This probability depends upon the parameters d (the length of the index vector), and ε (the number of '+1's and '-1's in an Index vector) (see equation (1)). Hence we term an RI-based Word Space as near-orthogonal. Near-orthogonality of an RI-based Word Space can be expressed in a probabilistic way in terms of a parameter β; as β-orthogonal, where β is an angle between 0 to $\pi/2$ radian. The following subsection provides a mathematical definition for β-orthogonality with respect to random index vectors.

4.1 β-ORTHOGONALITY IN A RI-BASED WORD SPACE

For $\mathbb{W}_{RI}$ consisting of a set of m index vectors $\mathbb{I} = \{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_m\}$, consider $\mathbb{J} \subseteq \mathbb{I}$. Let $\mathcal{I}_k \in \mathbb{I}/\mathbb{J}$ and $\mathbb{J}_\beta = \left\{\mathcal{I}_j \in \mathbb{J}: \left|Ang(\mathcal{I}_j, \mathcal{I}_k) - \frac{\pi}{2}\right| \leq \left(\frac{\pi}{2} - \beta\right)\right\}$, where β is a predetermined angle between 0 to $\pi/2$ radian, and $Ang(\mathcal{I}_j, \mathcal{I}_k)$ is the angle between the index vectors $\mathcal{I}_j$ and $\mathcal{I}_k$ measured in radian. The subset $\mathbb{J}$ is said be *β-orthogonal* to the index vector $\mathcal{I}_k$ with probability p, where $p = |\mathbb{J}_\beta|/|\mathbb{J}|$.

It may be noted that, the subset $\mathbb{J}$ contains $p \cdot |\mathbb{J}|$ number of index vectors which deviate at most by an angle of $\left(\frac{\pi}{2} - \beta\right)$ radian from being orthogonal to the index vector $\mathcal{I}_k$. These index vectors constitute the subset $\mathbb{J}_\beta$ and lie in the grey colored region as shown in a representative two dimensional plot (see Figure 1). One can also represent the subset $\mathbb{J}_\beta$ as $\mathbb{J}_\beta = \{\mathcal{I}_j \in \mathbb{J}: \beta \leq Ang(\mathcal{I}_j, \mathcal{I}_k) \leq \pi - \beta\}$.

If $\tilde{\mathbb{J}} = \{(\mathcal{I}_j, \mathcal{I}_k): (\mathcal{I}_j, \mathcal{I}_k) \in \mathbb{I} \times \mathbb{I} \text{ and } j < k\}$ and $\tilde{\mathbb{J}}_\beta = \left\{(\mathcal{I}_j, \mathcal{I}_k) \in \tilde{\mathbb{J}}: \left|Ang(\mathcal{I}_j, \mathcal{I}_k) - \frac{\pi}{2}\right| \leq \left(\frac{\pi}{2} - \beta\right)\right\}$, then the RI-based Word Space $\mathbb{W}_{RI}$ is said to be *β-orthogonal* with probability p, where $p = |\tilde{\mathbb{J}}_\beta|/|\tilde{\mathbb{J}}|$.

The Word Space $\mathbb{W}_{RI}$ is close to orthogonality if for large β $(0 < \beta \leq \pi/2)$ the value of p is large. It can be noted that if for $\beta = \pi/2$ the value of p is 1, then the Word Space $\mathbb{W}_{RI}$ is orthogonal. In our discussion, the terms 'orthogonal' and '$\pi/2$-orthogonal' have different meanings. If $\mathbb{W}_{RI}$ is $\pi/2$-orthogonal with $p = 1$, then we call it as orthogonal; otherwise for $p < 1$, we call it as $\pi/2$-orthogonal with probability p.

If in a RI-based Word Space $\mathbb{W}_{RI}$, angle between any two index vectors is one of the angles $\beta_1, \beta_2, \cdots, \beta_{r-1}$ or β_r such that $\beta_1 > \beta_2 > \cdots > \beta_{r-1} > \beta_r$, then $\left|\tilde{\mathbb{J}}_{\beta_t}\right| = \left|\tilde{\mathbb{J}}_{\beta_{t-1}}\right| + N_{\beta_t}$, where $t = 2,3,\cdots,r$ and $N_{\beta_t} = \left|\left\{\left(\mathcal{J}_j, \mathcal{J}_k\right) \in \tilde{\mathbb{J}} : Ang\left(\mathcal{J}_j, \mathcal{J}_k\right) = \beta_t\right\}\right|$.

In order to study the effect of near-orthogonality in a RI-based Word Space we consider following three cases as per equation (1):

Case 1: $\varepsilon = 2$, i.e. index vectors consisting of one '+1' and one '-1';

Case 2: $\varepsilon = 4$, i.e. index vectors consisting of two '+1's and two '-1's;

Case 3: $\varepsilon = 6$, i.e. index vectors consisting of three '+1's and three '-1's.

Subsections 4.1.1 to 4.1.3 discuss these three cases. In this discussion $\mathrm{C}(m, n)$ denotes the combination function: the number of ways of selecting n objects out of m objects. The dimension of index vectors are taken as even numbers, say $d = 2k$. The occurrence of '+1's is restricted in the upper half (positions 1 to k) and occurrence of the '-1's in the lower half (positions $k+1$ to $2k$) of an index vector. Calculation of different probabilities is based on the above positional restrictions that we have followed in our implementation.

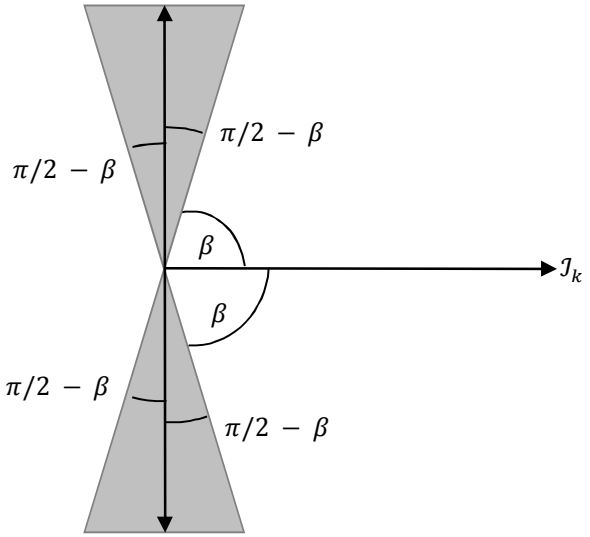


Fig. 1. Near-orthogonality w.r.t. an index vector in two-dimensional plane

4.1.1 CASE-1: INDEX VECTORS CONSISTING OF ONE '+1' AND ONE '-1'

In this case $|\mathbb{I}| = k^2$ and $\left|\tilde{\mathbb{J}}\right| = \mathrm{C}(k^2, 2)$. Any two index vectors are either orthogonal to each other or they make an angle of $\pi/3$ radian between them. Here,

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{2}}\right| = \frac{1}{2}k^2(k-1)^2 \tag{8}$$

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{3}}\right| = \left|\tilde{\mathbb{J}}_{\frac{\pi}{2}}\right| + k^2(k-1) \tag{9}$$

Hence, the RI-based Word Space is $\pi/2$-orthogonal with probability $\left|\tilde{\mathbb{J}}_{\frac{\pi}{2}}\right| / \left|\tilde{\mathbb{J}}\right|$ and $\pi/3$-orthogonal with probability 1.

4.1.2 CASE-2: INDEX VECTORS CONSISTING OF TWO '+1'S AND TWO '-1'S

In this case $|\mathbb{I}| = \left(\mathrm{C}(k,2)\right)^2$ and $\left|\tilde{\mathbb{J}}\right| = \mathrm{C}\left(\left(\mathrm{C}(k,2)\right)^2, 2\right)$. Any two index vectors are either orthogonal to each other or they make an angle of $\frac{\pi}{2.383}, \frac{\pi}{3}$ or $\frac{\pi}{4.347}$ radian between them. Here,

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{2}}\right| = \frac{1}{2}\left(\mathrm{C}(k-2,2)\right)^2\left(\mathrm{C}(k,2)\right)^2 \tag{10}$$

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{2.383}}\right| = \left|\tilde{\mathbb{J}}_{\frac{\pi}{2}}\right| + 2(k-2)\mathrm{C}(k-2,2)\big(\mathrm{C}(k,2)\big)^2 \tag{11}$$

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{3}}\right| = \left|\tilde{\mathbb{J}}_{\frac{\pi}{2.383}}\right| + (\mathrm{C}(k-2,2) + 2(k-2)^2)\big(\mathrm{C}(k,2)\big)^2 \tag{12}$$

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{4.347}}\right| = \left|\tilde{\mathbb{J}}_{\frac{\pi}{3}}\right| + 2(k-2)\big(\mathrm{C}(k,2)\big)^2 \tag{13}$$

Hence the RI-based Word Space is $\frac{\pi}{2}$-orthogonal with probability $\left|\tilde{\mathbb{J}}_{\frac{\pi}{2}}\right|/|\tilde{\mathbb{J}}|$, $\frac{\pi}{2.383}$-orthogonal with probability $\left|\tilde{\mathbb{J}}_{\frac{\pi}{2.383}}\right|/|\tilde{\mathbb{J}}|$, $\frac{\pi}{3}$-orthogonal with probability $\left|\tilde{\mathbb{J}}_{\frac{\pi}{3}}\right|/|\tilde{\mathbb{J}}|$, and $\frac{\pi}{4.347}$-orthogonal with probability 1.

4.1.3 CASE-3: INDEX VECTORS CONSISTING OF THREE '+1'S AND THREE '-1'S

In this case $|\mathbb{I}| = \big(\mathrm{C}(k,3)\big)^2$ and $|\tilde{\mathbb{J}}| = \mathrm{C}\left(\big(\mathrm{C}(k,3)\big)^2, 2\right)$. Any two index vectors are either orthogonal to each other or they make an angle of $\frac{\pi}{2.239}, \frac{\pi}{2.552}, \frac{\pi}{3}, \frac{\pi}{3.735}$ or $\frac{\pi}{5.364}$ radian between them. Here,

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{2}}\right| = \frac{1}{2}\big(\mathrm{C}(k-3,3)\big)^2\big(\mathrm{C}(k,3)\big)^2 \tag{14}$$

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{2.239}}\right| = \left|\tilde{\mathbb{J}}_{\frac{\pi}{2}}\right| + 3\mathrm{C}(k-3,2)\mathrm{C}(k-3,3)\big(\mathrm{C}(k,3)\big)^2 \tag{15}$$

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{2.552}}\right| = \left|\tilde{\mathbb{J}}_{\frac{\pi}{2.239}}\right| + \frac{1}{2}\Big(6(k-3)\mathrm{C}(k-3,3) + 9\big(\mathrm{C}(k-3,2)\big)^2\Big)\big(\mathrm{C}(k,3)\big)^2 \tag{16}$$

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{3}}\right| = \left|\tilde{\mathbb{J}}_{\frac{\pi}{2.552}}\right| + \big(\mathrm{C}(k-3,3) + 9(k-3)\mathrm{C}(k-3,2)\big)\big(\mathrm{C}(k,3)\big)^2 \tag{17}$$

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{3.735}}\right| = \left|\tilde{\mathbb{J}}_{\frac{\pi}{3}}\right| + \frac{1}{2}(6\mathrm{C}(k-3,2) + 9(k-3)^2)\big(\mathrm{C}(k,3)\big)^2 \tag{18}$$

$$\left|\tilde{\mathbb{J}}_{\frac{\pi}{5.364}}\right| = \left|\tilde{\mathbb{J}}_{\frac{\pi}{3.735}}\right| + 3(k-3)\big(\mathrm{C}(k,3)\big)^2 \tag{19}$$

Hence the RI-based Word Space is $\frac{\pi}{2}$-orthogonal with probability $\left|\tilde{\mathbb{J}}_{\frac{\pi}{2}}\right|/|\tilde{\mathbb{J}}|$, $\frac{\pi}{2.239}$-orthogonal with probability $\left|\tilde{\mathbb{J}}_{\frac{\pi}{2.239}}\right|/|\tilde{\mathbb{J}}|$, $\frac{\pi}{2.552}$-orthogonal with probability $\left|\tilde{\mathbb{J}}_{\frac{\pi}{2.552}}\right|/|\tilde{\mathbb{J}}|$, $\frac{\pi}{3}$-orthogonal with probability $\left|\tilde{\mathbb{J}}_{\frac{\pi}{3}}\right|/|\tilde{\mathbb{J}}|$, $\frac{\pi}{3.735}$-orthogonal with probability $\left|\tilde{\mathbb{J}}_{\frac{\pi}{3.735}}\right|/|\tilde{\mathbb{J}}|$, and $\frac{\pi}{5.364}$-orthogonal with probability 1.

Table 1. Near-orthogonality of RI-based Word Spaces

$\beta-$ orthogonal	Case-1 ($\varepsilon=2$)			Case-2 ($\varepsilon=4$)			Case-3 ($\varepsilon=6$)		
	$d=20$	$d=30$	$d=40$	$d=20$	$d=30$	$d=40$	$d=20$	$d=30$	$d=40$
$\pi/5.364$	-	-	-	-	-	-	1	1	1
$\pi/4.347$	-	-	-	1	1	1	-	-	-
$\pi/3.735$	-	-	-	-	-	-	0.997	1	1
$\pi/3$	1	1	1	0.984	0.995	0.998	0.958	0.992	0.997
$\pi/2.552$	-	-	-	-	-	-	0.769	0.921	0.964
$\pi/2.383$	-	-	-	0.830	0.920	0.954	-	-	-
$\pi/2.239$	-	-	-	-	-	-	0.391	0.654	0.783
$\pi/2$	0.818	0.875	0.905	0.387	0.552	0.649	0.085	0.234	0.356
Distinct Words	100	225	400	2025	11025	36100	14400	207025	1299600

Table 1 illustrates the effect of dimension of index vectors on β-orthogonality of RI-based Word Spaces for the three cases discussed above. Irrespective of the dimension of index vector, the Word Spaces of Case-1, Case-2 and Case-3 are $\frac{\pi}{3}$-orthogonal, $\frac{\pi}{4.347}$-orthogonal and $\frac{\pi}{5.364}$-orthogonal respectively with probability 1. For a fixed dimension d and any $\beta \in (0, \pi/2]$, β-orthogonality is achieved with highest probability for $\varepsilon = 2$ (Case-1); and probability of β-orthogonality decreases as ε increases. But for a smaller value of ε, one has to select a very large d to accommodate all the words of a large document. If one prefers a small d to accommodate all the words of the same document, a higher value of ε is to be chosen. Hence there is a trade off between parameters ε and d of index vectors and β-orthogonality of Word Space. The question therefore arises how this trade off actually affects the quality of summaries of RI-based extractive summarizer. Section 5 deals with this issue in detail.

5 EFFECT OF NEAR-ORTHOGONALITY ON THE PERFORMANCE OF RISUM

To study the effect of near-orthogonality on the performance of RISUM, we have used DUC 2002 [19] corpus consisting of newswire documents on various topics as experimental dataset. DUC 2002 is the last version of DUC that included single-document summarization evaluation of informative summaries. Later DUC editions contained a single-document summarization task as well; however, only very short summaries (e.g. headline summaries) were analyzed**.** Since our work is not focused on producing headline summaries, we considered DUC 2002 corpus as our experimental dataset. Each document of DUC 2002 corpus is accompanied by two different abstracts manually created by professionals. These summaries serve as the 'gold' summaries of the corresponding document. The output summary produced by RISUM for each document is limited to 10% of the length of the original document. These summaries are termed as 'system' summaries of the document and were evaluated using ROUGE (Recall Oriented Understudy for Gisting Evaluation) [8] toolbox. ROUGE measures summary quality by counting the overlapping units, such as, *n*-gram, word sequences, word pairs between the system summary and gold summaries. According to the recommendations of Lin [8] for single document summarization, we have used ROUGE-2, ROUGE-L, ROUGE-W-1.2, and ROUGE-S* for summary evaluation. We have considered 95% confidence interval for the evaluation.

5.1 CASE-1: RI WITH ONE '+1' AND ONE '-1'

In this case the dimension of index vector d_{RI} is determined using equation (20):

$$d_{RI} = 2\lceil\sqrt{m}\rceil \tag{20}$$

where m is the number of distinct words in the document and $\lceil\ \rceil$ stands for the *ceiling* function. In fact d_{RI} is the minimum dimension that is required to accommodate all the distinct words of the document. Table 2 shows the ROUGE F-measure scores for RISUM_Cosine[3] and RISUM_Euclidean[4] with dimension of index vectors d_{RI} and $2d_{RI}$. There is almost no change in scores of RISUM_Cosine when the dimension of index vector is doubled, but there is a very little improvement in the scores of RISUM_Euclidean. Also, it can be observed RISUM_Euclidean performs better than RISUM_Cosine irrespective of dimensions of index vectors.

Table 2. ROUGE F-measure scores for Case-1 (RI with one '+1' and one '-1')

Evaluation Measure	RISUM_Cosine		RISUM_Euclidean	
	d_{RI}	$2d_{RI}$	d_{RI}	$2d_{RI}$
ROUGE-2	0.132	0.129	0.155	0.159
ROUGE-L	0.222	0.221	0.265	0.268
ROUGE-W-1.2	0.145	0.145	0.169	0.171
ROUGE-S*	0.108	0.108	0.132	0.137

[3] *RISUM_Cosine: RISUM with cosine dissimilarity measure*

[4] *RISUM_Euclidean: RISUM with Euclidean distance measure*

5.2 CASE-2: RI WITH TWO '+1'S AND TWO '-1'S

In this case d_{RI}, the minimum dimension of index vector required to accommodate all the distinct words of a document is determined using equation (21):

$$d_{RI} = 2\left\lceil 0.5\left(1+\sqrt{1+8\sqrt{m}}\right)\right\rceil \quad (21)$$

Table 3 shows the ROUGE F-measure scores for RISUM_Cosine and RISUM_Euclidean with dimension of index vectors d_{RI} and $2d_{RI}$. When dimension of index vectors taken as d_{RI}, RISUM_Euclidean performs much better than RISUM_Cosine. When dimension of index vectors is increased by two times, performance of RISUM_Cosine improved significantly. However, as in Case-1 above, only a negligible improvement can be noticed in the performance of RISUM_Euclidean. Also, it can be observed with dimension $2d_{RI}$ ROUGE scores of RISUM_Cosine is still inferior to corresponding ROUGE scores of RISUM_Euclidean.

Table 3. ROUGE F-measure scores for Case-2 (RI with two '+1's and two '-1's)

Evaluation Measure	RISUM_Cosine		RISUM_Euclidean	
	d_{RI}	$2d_{RI}$	d_{RI}	$2d_{RI}$
ROUGE-2	0.102	0.127	0.157	0.161
ROUGE-L	0.195	0.221	0.268	0.275
ROUGE-W-1.2	0.126	0.143	0.170	0.175
ROUGE-S*	0.084	0.102	0.136	0.142

5.3 CASE-3: RI WITH THREE '+1'S AND THREE '-1'S

In Case-3, for each document the minimum dimension of index vector d_{RI} is twice of the real root of the equation (22):

$$x^3 - 3x^2 + 2x - 6\sqrt{m} = 0 \quad (22)$$

Table 4 shows the ROUGE F-measure scores for RISUM_Cosine and RISUM_Euclidean with dimension of index vectors d_{RI} and $2d_{RI}$. When dimension of index vectors is taken as d_{RI}, RISUM_Euclidean performs much better than RISUM_Cosine. When dimension of index vectors increased twofold, performance of RISUM_Cosine improved significantly; but surprisingly there is a slight degradation in the performance of RISUM_Euclidean. Also, it can be observed with dimension $2d_{RI}$ ROUGE scores of RISUM_Cosine is still inferior to corresponding ROUGE scores of RISUM_Euclidean.

Table 4. ROUGE F-measure scores for Case-3 (RI with three '+1's and three '-1's)

Evaluation Measure	RISUM_Cosine		RISUM_Euclidean	
	d_{RI}	$2d_{RI}$	d_{RI}	$2d_{RI}$
ROUGE-2	0.064	0.106	0.150	0.147
ROUGE-L	0.149	0.199	0.268	0.260
ROUGE-W-1.2	0.096	0.129	0.169	0.165
ROUGE-S*	0.049	0.087	0.133	0.128

5.4 COMPARISON OF ROUGE SCORES FOR THE THREE CASES

Tables 2, 3 and 4 help us in comparing the performance of the RISUM_Cosine and RISUM_Euclidean algorithms. The findings may be summarized as follows:

1. When dimension of index vector is fixed at d_{RI} for respective cases, the performance of RISUM_Cosine degrades gradually as we move from Case-1 to Case-3; but performance of RISUM_Euclidean remains almost the same for the three cases.

2. Increasing the dimension to $2d_{RI}$improves the performance of RISUM_Cosine in all the three cases. Moreover, the scores in Case-2 with respect to the four ROUGE measures become almost at par with the respective scores of Case-1; while the scores in Case-3 still remains inferior.
3. Increasing the dimension to $2d_{RI}$ does not have a very significant effect on the performance of RISUM_Euclidean unlike in the case of RISUM_Cosine. For Case-1 and Case-2 marginal improvements can be noticed in all the four ROUGE measures, but in Case-3 there is a marginal fall in the scores.

Table 5, Figures 2 and 3 provide more insight into the effect of near-orthogonality on the performance RISUM_Cosine scheme. Table 5 shows the probabilities of $\pi/3$-orthogonality and $\pi/2$-orthogonality of the respective RI-based Word Spaces against the average values of d_{RI} and $2d_{RI}$ of index vector for DUC 2002 experimental dataset. Figures 2 and 3 respectively show the effect of $\pi/3$-orthogonality and $\pi/2$-orthogonality on the performance of RISUM_Cosine.

1. As the probability of $\pi/3$-orthogonality of the RI-based Word Space increases, the performance of RISUM_Cosine also improves. The improvement in ROUGE scores can be noticed till the probability of $\pi/3$-orthogonality reaches 0.99.
2. As the probability of $\pi/2$-orthogonality of the RI-based Word Space increases from 0.003 to 0.146, steady improvement can be observed in the performance of RISUM_Cosine. The performance RISUM_Cosine falls marginally as the probability of $\pi/2$-orthogonality increases from 0.146 to 0.227. Further, increase in probability of $\pi/2$-orthogonality from 0.227 to 0.526 shows improved performance of RISUM_Cosine. Increase in probability of $\pi/2$-orthogonality beyond 0.526 does not show any effect on the performance of RISUM_Cosine.

Table 5. Near-orthogonality of RI-based Word Space over DUC 2002 experimental dataset

β-orthogonal	Case-1		Case-2		Case-3	
	$d_{RI} = 32$	$2d_{RI} = 64$	$d_{RI} = 14$	$2d_{RI} = 28$	$d_{RI} = 12$	$2d_{RI} = 24$
$\pi/3$	1	1	0.955	0.994	0.707	0.979
$\pi/2$	0.882	0.939	0.227	0.526	0.003	0.146

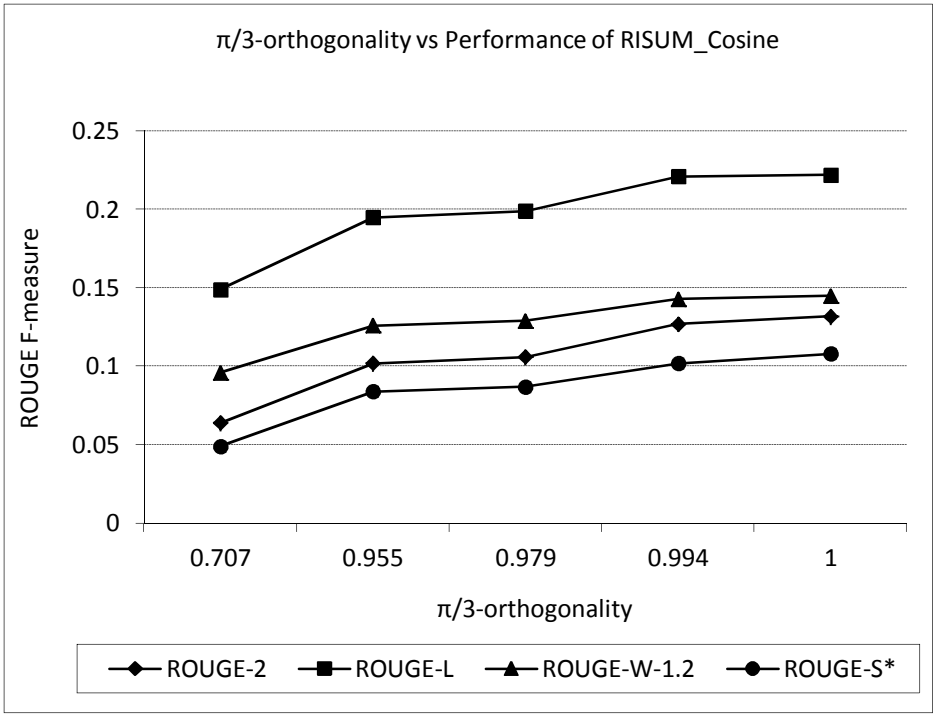


Fig. 2. Effect of π/3-orthogonality on the performance of RISUM_Cosine

From the above discussions it can be concluded that, the increase in performance of RISUM_Cosine is directly proportional to the increase in the probability of $\pi/3$-orthogonality of the RI-based Word Space. But, the probability of $\pi/2$-orthogonality has no relational effect in the performance of RISUM_Cosine. From the study it has been inferred that if one can ensure $\pi/3$-orthogonality with probability more than 0.99 in an RI-based Word Space, then very high-quality performance can be achieved for RISUM_Cosine. Further, it has been observed that the performance of RISUM_Euclidean is mostly invariant of near-orthogonality of the Word Space. This is because in RISUM_Euclidean the proximity measure is the linear distance unlike the angular distance in RISUM_Cosine. Since orthogonality and near-orthogonality are angular properties it does not affect the performance of RISUM_Euclidean.

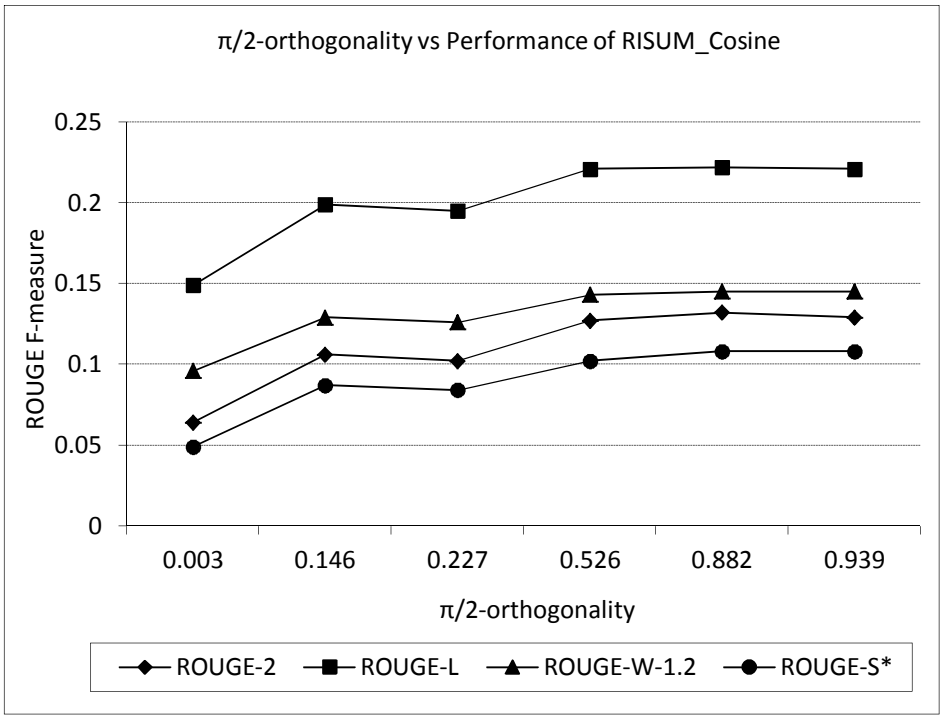


Fig. 3. Effect of $\pi/2$-orthogonality on the performance of RISUM_Cosine

5.5 COMPARISON OF RISUM WITH LSA-BASED SUMMARIZERS

In this section we compare the performance of RISUM with different LSA-based summarizers proposed in literatures. For comparison the ROUGE-L F-measure scores of LSA-based summarizers are taken from Ozsoy *et al.* [14], where 10% summaries of DUC 2002 dataset were produced. Each LSA-based summarizer was evaluated against different weighting schemes. For comparison purpose we have taken the best performance score of the each of the summarizers. Also, the best performance scores of RISUM_Cosine and RISUM_Euclidean were considered for comparison. ROUGE-L F-measure scores of the summarizers are given in Table 6. It has been observed that RISUM_Cosine performs better than cross method of Ozsoy *et al.* [10], [14], and at par with the approach of Steinberger and Ježek [12]. Also, it has been observed that RISUM_Euclidean performs much better than all LSA-based summarizers. The lowest ROUGE-L F-measure score of RISUM_Euclidean is 0.260 (Case-3, $2d_{RI}$), which is much higher than the best scores of all LSA-based summarizers.

The consistently better performance of RISUM_Euclidean over all LSA-based summarization techniques is significant because of two reasons. Firstly, the size of index vector in RI is much lower than the size of the word vector in LSA. Secondly, RISUM avoids complex dimension reduction technique, like SVD, used by LSA-based approaches – which make them computationally expensive.

Table 6. ROUGE-L F-measure scores of summarizers

Summarizer	ROUGE-L F-measure score
Gong and Liu (binary)	0.234
Steinberger and Ježek (binary)	0.224
Murray *et al.* (binary)	0.230
Ozsoy *et al.* (cross method) (binary)	0.196
Ozsoy *et al.* (topic method) (binary)	0.230
RISUM_Cosine (Case-1, d_{RI})	0.222
RISUM_Euclidean (Case-2, $2d_{RI}$)	0.275

6 CONCLUSION

This paper explores near-orthogonality – a significant aspect of the RI-based Word Space. We formulated a definition of near-orthogonality for RI-based Word Spaces. This definition should be equally applicable to other types of Word Spaces as well. We compared the near-orthogonality in RI-based Word Spaces by varying the number of '+1's and '-1's, and dimension of the index vectors. The effect of near-orthogonality on RISUM, an extractive summarization scheme based on RI, has been evaluated for these cases. Two variations of the above scheme have been studied: RISUM_Cosine and RISUM_Euclidean. Our experiments reveal that the RISUM_Cosine performs best when random index vector with one '+1' and one '-1' is considered. Construction of Word Space with index vectors having more numbers of '+1's and '-1's, requires choosing a dimension for the index vectors such that the Word Space should be $\pi/3$-orthogonal with probability more than 0.99. This will ensure a high-quality performance of the RISUM_Cosine. However, the performance of RISUM_Euclidean is almost invariant of near-orthogonality of the Word Space. Also, RISUM_Euclidean performs much better than the LSA-based summarization techniques proposed in literature.

The focus of the present work has been on extraction based single document summarization using RI. In future we plan to focus on multi-document summarization using RI, and study the effect of near-orthogonality in this context.

REFERENCES

[1] N. Chatterjee and S. Mohan, "Extraction-based Single-Document Summarization using Random Indexing," In: *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, vol. 2, pp. 448-455, 2007.

[2] M. Sahlgren, "An Introduction to Random Indexing," In: *H. Witschel (ed.): Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering (TKE 2005)*, 2005.

[3] M. Sahlgren, "*The Word-Space Model: Using Distributional Analysis to Represent Syntagmatic and Paradigmatic Relations between Words in High-dimensional Vector Spaces*," PhD Dissertation, Stockholm University, Sweden, 2006.

[4] H. Schütze, "Word Space," In: *Proceedings of the Conference on Advances in Neural Information Processing Systems (NIPS 1993)*, pp. 895-902, 1993.

[5] T. K. Landauer and S. T. Dumais, "A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge," *Psychological Review*, vol. 104, no. 2, pp. 211-240, 1997.

[6] T. K. Landauer, P. W. Foltz, and D. Laham, "Introduction to Latent Semantic Analysis," *Discourse Processes*, vol. 25, pp. 259-284, 1998.

[7] N. Chatterjee and P. K. Sahoo, "Near-orthogonality of Random Index Vectors and its Effect on Extractive Text Summarization," *Proceedings of the Indian National Science Academy (On the Occasion of the Golden Jubilee of IIT Delhi, India)*, vol. 77, no. 2, pp. 207-218, 2011.

[8] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries," In: *Proceedings of the ACL-04 Workshop on Text Summarization Branches Out*, pp. 74-81, 2004.

[9] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391-407, 1990.

[10] M. G. Ozsoy, I. Cicekli, and F. N. Alpaslan, "Text Summarization of Turkish Texts using Latent Semantic Analysis," In: *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, pp. 869-876, 2010.

[11] Y. Gong and X. Liu, "Generic Text Summarization using Relevance Measure and Latent Semantic Analysis," In: *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001)*, pp. 19-25, 2001.

[12] J. Steinberger and K. Ježek, "Using Latent Semantic Analysis in Text Summarization and Summary Evaluation," In: *Proceedings of the 5th International Conference on Information Systems Implementation and Modelling (ISIM 2004)*, pp. 93-100, 2004.

[13] G. Murray, S. Renals, and J. Carletta, "Extractive Summarization of Meeting Recordings," In: *Proceedings of the 9th European Conference on Speech Communication and Technology*, 2005.

[14] M. G. Ozsoy, I. Cicekli, and F. N. Alpaslan, "Text Summarization using Latent Semantic Analysis," *Journal of Information Science*, vol. 37, no. 4, pp. 405-417, 2011.

[15] M. F. Porter, "An Algorithm for Suffix Stripping," *Program*, vol. 14, no. 3, pp. 130-137, 1980.

[16] D. Higgins and J. Burstein, "Sentence Similarity Measures for Essay Coherence," In: *Proceedings of the 7th International Workshop on Computational Semantics (IWCS 2007)*, 2007.

[17] R. Mihalcea and P. Tarau, "TextRank: Bringing Order into Texts," In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, pp. 404-411, 2004.

[18] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, pp. 107-117, 1998.

[19] Document Understanding Conference 2002. [Online] Available: http://www-nlpir.nist.gov/projects/duc/data.html (accessed August 2012).

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 714-718

http://www.issr-journals.org/ijias/

A Minimized Triangular – Meander Line PIFA Antenna for DCS1800/WIMAX Applications

Sonali Kushwah, P.K. Singhal, Manali Dongre, and Tajeswita Gupta

Department of Electronics,
Madhav Institute of Technology and Science,
Gwalior-474001, M.P., India

ABSTRACT: A novel multiband Triangular – Meander Line Planar Inverted - F Antenna (PIFA) has been introduced. The proposed PIFA antenna consists of a triangle shaped meandered – line radiating patch. The proposed antenna operates in the DCS1800 (1.710-1.880 GHz), Mobile WIMAX (2.5-2.69 GHz) and Fixed WIMAX (3.4-3.8 GHz) bands efficiently. Classic triangular PIFA shows very low bandwidth and also operates in a single frequency band. Therefore, the combination of Meander–Line and PIFA antenna improves the proposed antenna characteristics. It also increases the current path. The proposed antenna is covering the size of 50 × 36 × 9.276 mm and showing the overall gain of 3.028 dB.

KEYWORDS: Meander – Line, Planar Inverted – F Antenna (PIFA), Return Loss.

1 INTRODUCTION

In recent years, the wireless communication industry has demanded the equipments of low profile, multifunctional, high efficiency and better performance [1], [2]. Wireless networks such as WIMAX and mobile networks such as GSM need the equipments, which can be handled easily and show better performance as well. Therefore, PIFA and Meander – Line structures have come into interest.

PIFA antenna has the advantage of low profile structure, better efficiency and greater bandwidth [3], but also has the disadvantage of singe band operation. By combining the structures of PIFA and Meander – Line, one can get the better performance in the respect of gain, bandwidth, efficiency and VSWR. Therefore the disadvantage of Meander–Line structure of having low bandwidth and low radiation efficiency has also been overcome [4]-[8].

In this paper, for obtaining better characteristics, Meander – Line approach has been combined with a Triangular PIFA antenna. Also, a comparison among the characteristics of the simple triangular PIFA antenna and Meandered – Line Triangular PIFA antenna has also been included in this research. For each one, the antenna design, return loss graph, 3D pattern, E-field and H-field pattern has also been presented.

2 ANTENNA DESIGN

In this section, the design of the proposed antenna has been discussed along with a general triangular PIFA antenna.

2.1 TRIANGULAR PIFA ANTENNA

The characteristics of Triangular PIFA antenna have first been analyzed. For which, a PIFA antenna with a triangular radiating patch has been designed and simulated. The antenna design has been pictured in Fig. 1. The side of the triangular patch has been kept 32.37 mm. The size of the ground plate is 50 × 36 mm. The radiating patch has been connected to the

Corresponding Author: Sonali Kushwah (sonalikushwah.mits@gmail.com)

ground with the help of a shorting strip of thickness 5 mm. This shorting strip is making the ground defective. The substrate used is Epoxy Glass (FR4) having relative permittivity 4.4, loss tangent 0.02 and height of 1.6 mm.

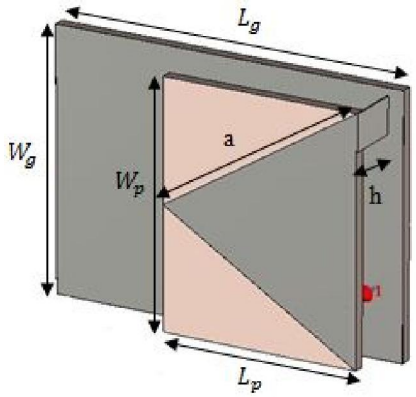


Fig. 1. Triangular PIFA Antenna

2.2 TRIANGULAR MEANDER – LINE PIFA ANTENNA

To improve the characteristics of PIFA antenna and to overcome the disadvantage of PIFA of single band operation, Meander – Line structure has been introduced in the triangular radiating patch. The proposed antenna has been pictured in Fig. 2. The Meander – Line patch consists of 3 mm thick Meandered strip, covering the area of previous triangular patch and this is now connected to the ground plate with the help of a shorting strip. The total size occupied by this whole structure is 50 × 36 × 9.276 mm. The coaxial feed is positioned at 4 mm from the right end. The patch is kept 7.6 mm above the ground.

The physical dimensions of the proposed antenna have been given in table 1.

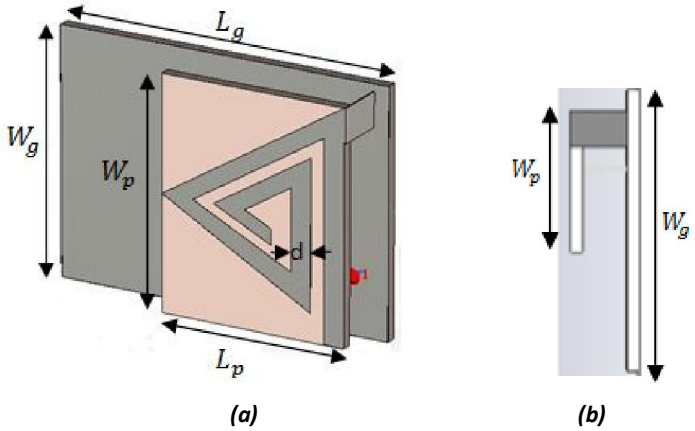


Fig. 2. Proposed triangular Meander – Line Antenna (a) Perspective view, (b) Side view

3 COMPUTATIONAL RESULTS

The PIFA has been simulated in the CST-MWS environment. With the insertion of meandered structure in PIFA antenna, multiband results can be achieved easily. The comparison results of the return loss of both the antennas have been shown in figure 3. The graph shows that the triangular PIFA antenna is operating in a single band centered at 1.9 GHz, while triangular meander-line PIFA antenna is showing triple band operation. The bands achieved, are DCS1800 (1.710-1.880 GHz), Mobile WIMAX (2.5-2.69 GHz) and Fixed WIMAX (3.4-3.8 GHz). The gain (IEEE) of 3.028 dB has been achieved. The bandwidth for the triple bands obtained, are 9.47%, 7.32% and 11.11% respectively.

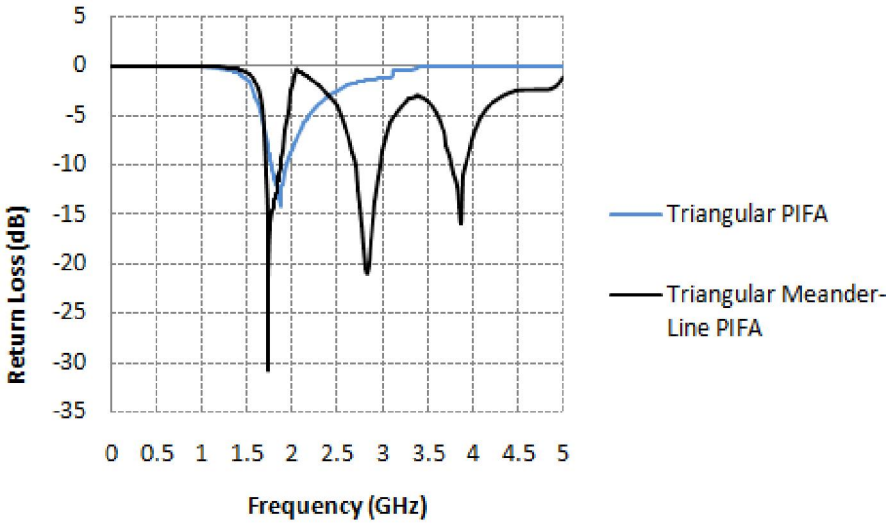
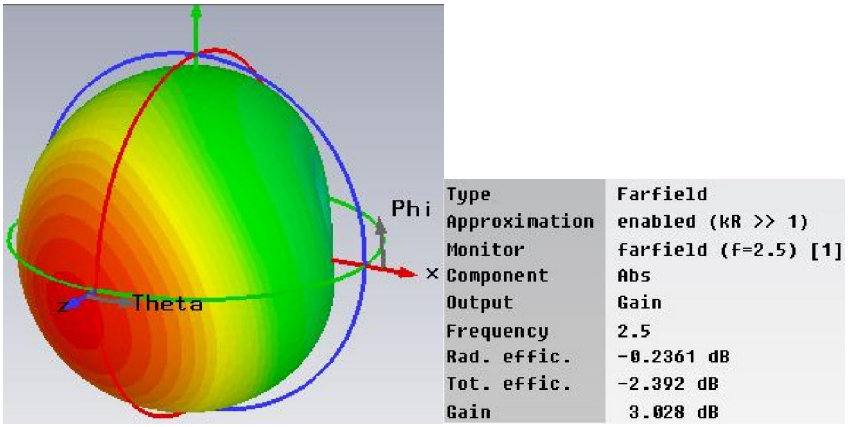


Fig. 3. Comparison between S-Parameter pattern of Triangular PIFA and Triangular Meander-Line PIFA

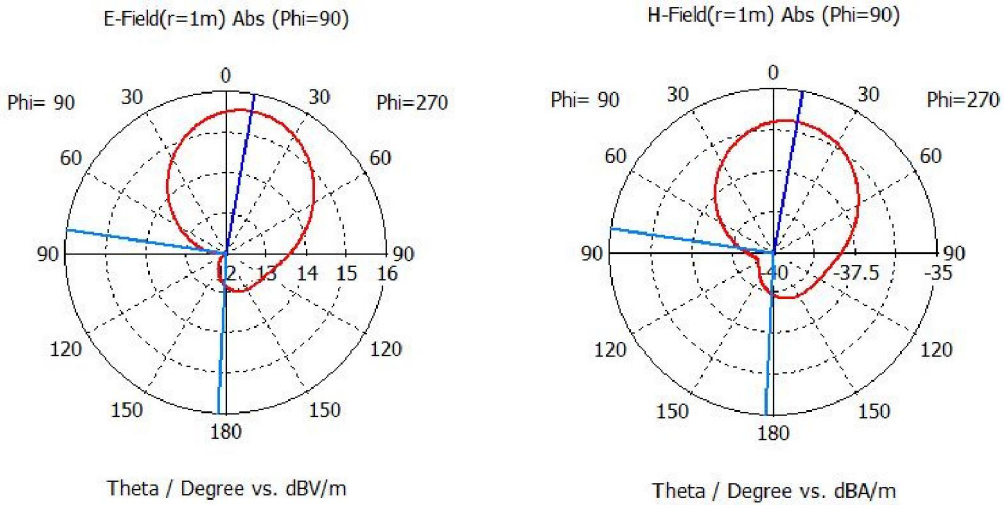
Table 1. Physical Dimensions of the proposed antenna (All in mm)

Elements	Dimensions
Ground Size	50 × 36
Side of the Triangular Patch	32.37
Air Gap Size	7.6
Width of the Shorting Plate	5
Shorting Plate Thickness	0.2
Height of Shorting Plate	7.638

The 3-Dimensional radiation pattern of the proposed antenna has been shown in figure 4(a), with the directivity of 3.268 dBi. The total efficiency of the proposed antenna is 57.65%, while triangular PIFA antenna is showing the efficiency of 43.66%. The E-field and H-field patterns have been shown in Fig. 4(b), 4(c) respectively.



(a)



(b) *(c)*

Fig. 4. (a) 3-Dimensional Radiation Pattern, (b) E-field Pattern, (c) H-field Pattern

The peak gain has also been analyzed which is shown in Fig. 5. This graph is showing that the peak gain is greater than 5.04 dB in the operating bandwidth.

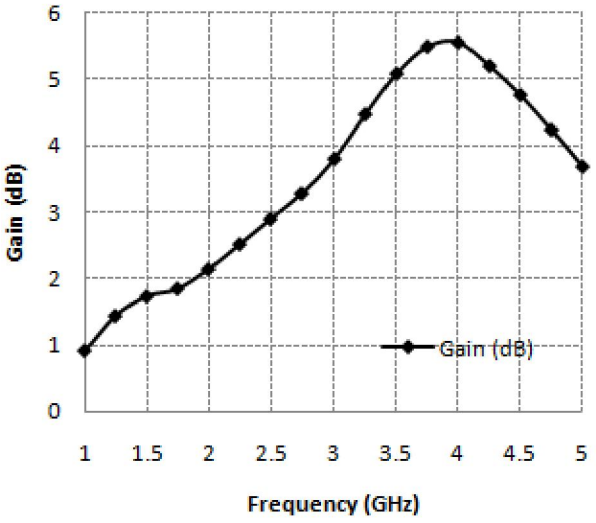


Fig. 5. Simulated Peak Gain of the Triangular Meander-Line PIFA antenna

4 CONCLUSION

A Triangular Meander-Line PIFA antenna has been designed and simulated, which is operating in triple bands and showing the application in mobile and wireless communication. A comparison has been shown among the characteristics of proposed antenna and triangular PIFA antenna. It has shown that the introduction of meander-line structure into the PIFA antenna made the antenna better with respect to bandwidth, gain, efficiency and multiband operation.

REFERENCES

[1] Lin, C.C., S.W. Kuo, and H.R. Chuang, "A 2.4-GHz printed meander-line antenna for USB WLAN with notebook-PC housing," *IEEE Microwave and Wireless Components*, Vol. 15, No. 9, September 2005.

[2] Galehdar, A., D. V. Thiel, S. G. O'Keefe, and S. P. Kingsley, "Efficiency variations in electrically small, meander line RFID antennas," *IEEE Antennas and Propagation Society International Symposium, 2007*.

[3] P. S. Hall, E. Lee, C. T. P. Song, *Printed Antennas for wireless Communications, In: Planar Inverted-F Antennas.* R. Waterhouse, John Wiley & Sons, Ltd, 2001.

[4] A. Jahanbakhshi, Gh. Moradi, R. Sarraf Shirazi, "Design and Simulation of Different Types of Meander Line Antennas with Improved Effciency," *PIERS Proceedings*, Moscow, Russia, August 19-23, 2012.

[5] Balanis, C. A., *Antenna Theory Analysis and Design*, 3rd Edition, John Wiley & Sons, Inc., 2005.

[6] Rabih Rahaoui and Mohammed Essaaidi, "Compact Cylindrical Dielectric Resonator Antenna excited by a Microstrip Feed Line," *International Journal of Innovation and Applied Studies*, vol. 2, no. 1, pp. 1–5, January 2013.

[7] Mohammed Younssi, Achraf Jaoujal, Yacoub Diallo, Ahmed El-Moussaoui, and Noura Aknin, "Study of a Microstrip Antenna with and Without Superstrate for Terahertz Frequency," *International Journal of Innovation and Applied Studies*, vol. 2, no. 4, pp. 369–371, April 2013.

[8] M. I. Hasan and M. A. Motin, "New slotting technique of making compact octagonal patch for four band applications," *International Journal of Innovation and Applied Studies*, vol. 3, no. 1, pp. 221–227, May 2013.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 719-726

http://www.issr-journals.org/ijias/

Détermination des périodes optimales de semis du riz pluvial au Centre-ouest de la Côte d'Ivoire

[Determination of pluvial rice optimum seedling periods in Middle-west of Côte d'Ivoire]

Kouakou Koffi Eugene[1], Kouassi Amani[2], Kouassi Francis William[2], Goula Bi Tie Albert[2], and Savane Issiaka[2]

[1]Département des Sciences de la Terre et des Ressources Minières (STeRMi),
Laboratoire du Génie Civil, des Géosciences et des Sciences Géographiques,
Institut National Polytechnique Félix Houphouët-Boigny (INP-HB),
BP 1093 Yamoussoukro, Côte d'Ivoire

[2]Unité de Formation en Sciences et Gestion de l'Environnement,
Laboratoire Géosciences et Environnement, Université Nangui Abroguoua
02 BP 800 Abidjan, Côte d'Ivoire

ABSTRACT: This study aims to identify the optimal time of planting upland rice in the Center-west of Côte d'Ivoire which is located between longitudes 5º and 7º30 and latitudes 5º30 and 7 º. The method used is that of Eldin that tracks the plant water status by region and decade after decade to determine the vegetative periods in the year. The study focuses on rice 90 and 105 days. The data used are the values of the decadal rainfall on the normal 1971-2000 and information's about upland rice of 90 and 105 days and decadal crop coefficients.
The results indicate that the first growing season (first cycle) is generally between early April and early July and lasts 3 months. The second growing season (one cycle) is between August 11 and November 10 and lasts up to 2 months 10 days. In first crop cycle, the optimum planting dates of upland rice of 90 days must start between 11 and 30 April and the rice of 105 days between March 1 and April 20 depending on the location. In single crop cycle, the upland rice of 90 days should be planted between July 21 and August 20 and rice of 105 between July 1 and August 21 depending on the location of the study area.

KEYWORDS: Pluvial rice, seedling, optimum periods, Middle-West, Côte d'Ivoire.

RESUME: Cette étude qui a pour objectif d'identification de la période optimale de semis du riz pluvial se déroule dans la partie Centre-ouest de la Côte d'Ivoire située entre les longitudes 5º et 7º30 et les latitudes 5º30 et 7º. La méthode utilisée est celle d'Eldin qui permet de retracer l'état hydrique des plantes d'une région décade après décade et déterminer les périodes végétatives dans l'année. L'étude concerne le riz de 90 et 105 jours. Les données utilisées sont les valeurs décadaires de la pluviométrie sur la normale 1971-2000 et les informations sur les cycles du riz pluvial de 90 et 105 jours ainsi que les coefficients culturaux décadaires.
Les résultats indiquent que la première saison végétative (premier cycle) se situe globalement entre début avril et début juillet et dure 3 mois maximum. La deuxième période végétative (cycle unique) se situe entre le 11 août et le 10 novembre et dure au maximum 2 mois 10 jours. En premier cycle cultural, les dates optimales de semis du riz pluvial de 90 jours doivent démarrer entre le 11 et 30 avril et celle du riz de 105 jours entre le 1er mars et 20 avril selon les localités. En cycle cultural unique, le riz pluvial de 90 jours doit être semé entre le 21 juillet et le 20 août et le riz de 105 entre le 1er juillet et le 21 août également selon les localités de la zone d'étude.

MOTS-CLEFS: Riz pluvial, période optimale, semis, Centre-ouest, Côte d'Ivoire.

Corresponding Author: Kouakou Koffi Eugene (kkoffieugene@yahoo.fr)

1 INTRODUCTION

Le riz est une céréale très consommée dans le monde. En Côte d'Ivoire, la production nationale en riz demeure insuffisante à la consommation. En effet, le pays produit annuellement 700 000 tonnes de riz blanchi pour un besoin national de 1 600 000 tonnes ; le déficit de 900 000 tonnes est comblé par des importations[1]. Ces importations correspondent à une sortie de devises de plus de 103 milliards de francs CFA. La riziculture locale est faite essentiellement en condition pluviale. Mais, la répartition aléatoire des pluies pendant la saison culturale limite les rendements à un niveau faible de l'ordre de 800 kg de paddy à l'hectare.

Il existe plusieurs moyens pour améliorer ces rendements. Parmi ces moyens, le calage des cycles de culture du riz pluvial permet d'optimiser la production. Des études de calage des périodes de semis du riz pluvial avaient été effectuées en Côte d'Ivoire [1, 2] à partir de séries pluviométriques des années antérieures à 1970, période à partir de laquelle il est constaté un changement notable tant dans la quantité que dans la répartition des pluies [3, 4, 5]. Ainsi les périodes de semis du riz pluvial liée au zonage climatique [6, 7] et recommandées antérieurement ne sont plus adaptées aux conditions climatiques actuelles.

Face à cette problématique, la présente étude entreprend de déterminer à nouveau les périodes optimales de semis du riz pluvial dans la région Centre-Ouest de la Côte d'Ivoire qui est la plus importante zone de production et de consommation de riz. L'approche consiste à déterminer préalablement les périodes végétatives au cours desquelles la vie végétale est possible à la faveur des pluies et à caler sur celles-ci les dates optimales de semis de deux cycles de riz pluvial (90 jours et 105 jours).

2 PRESENTATION DE LA ZONE D'ETUDE

Le Centre Ouest de la Côte d'Ivoire appartient au climat équatorial de transition caractérisé par une alternance de 2 saisons pluvieuses avec 2 saisons sèches. La première saison des pluies part globalement de mars à mi-juillet et la deuxième d'août à mi-novembre. La première saison sèche encore appelée grande saison sèche se situe entre mi-novembre et fin février. La seconde saison sèche, plus courte, s'apparente parfois à un simple ralentissement entre mi-juillet et mi-août dans certaines localités [8]. La pluviométrie annuelle est comprise généralement entre 1100mm et 1400mm. La température moyenne oscille entre 19°C et 33°C. Les sols y sont ferralitiques, profonds et fertiles [9]. Ils sont généralement aptes à l'igname, au manioc, au riz pluvial, à la banane plantain, au cacao, au caféier, etc.

[1] *La production nationale et les importations annuelles de riz blanc en 2007 d'après le Programme National Riz (PNR) (2008).*

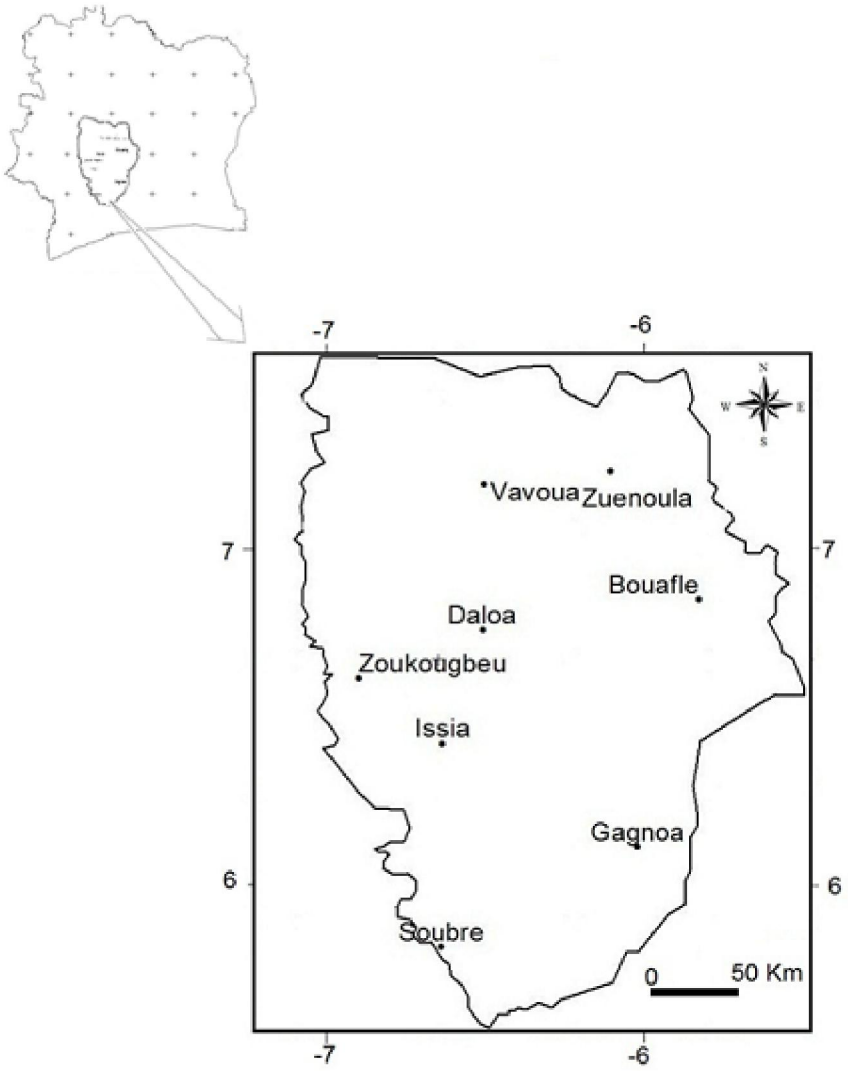


Fig. 1. Zone Centre-ouest de la Côte d'Ivoire

3 DONNEES ET METHODES

3.1 DONNÉES

Deux catégories de données ont été utilisées, des données climatiques et des données agronomiques. Les données climatiques, issues de la base de données de la SODEXAM, sont constituées de précipitations et d'évapotranspirations potentielles décadaires (ETP) des localités. Elles partent de 1971 à 2000. Les données agronomiques sont les durées de phases phénologiques et les coefficients culturaux décadaires des cycles de riz pluvial de 90 et de 105 jours. Elles ont été obtenues auprès du Centre National de Recherche Agronomique (CNRA).

3.2 DÉTERMINATION DES PÉRIODES VÉGÉTATIVES

Parmi les méthodes agro-climatiques de détermination des saisons végétatives, la méthode d'Eldin [10, 11, 12] permet de retracer l'état hydrique des plantes d'une région décade après décade et d'une année à une autre. Elle a servi à la détermination des périodes végétatives. Elle permet de comparer sur chaque décade de la série, la pluie (P) et l'évapotranspiration potentielle (ETP). On sait que cette dernière représente un majorant de la consommation hydrique de la plupart des plantes. Ainsi, obtenir pour une décade donnée, des pluies (l'apport d'eau) supérieures à l'ETP, c'est avoir l'assurance que les cultures ne manqueront pas d'eau. De la même manière, le rapport ETP/2 représente un besoin hydrique moyen qui permet à un couvert végétal de démarrer son cycle de végétation et de le terminer dans les conditions

convenables. Sur cette base, les probabilités que les pluies décadaires atteignent ou dépassent les valeurs décadaires de l'ETP et ETP/2 sont calculées comme suit :

$$\boldsymbol{Probabilit\acute{e}\ (P \geq ETP)} = \frac{\sum \boldsymbol{Ni(P \geq ETP)}}{\boldsymbol{N}}$$

$$\boldsymbol{Probabilit\acute{e}\ (P \geq ETP\ /\ 2)} = \frac{\sum \boldsymbol{Ni(P \leq ETP)}}{\boldsymbol{N}}$$

Probabilité (P≥ ETP) : Probabilité d'avoir une pluie supérieure ou égale à ETP ;

Probabilité (P≥ ETP/2) : Probabilité d'avoir une pluie supérieure ou égale à ETP/2 ;

$\sum N_i$ (P≥ ETP) : nombre de fois la pluie décadaire est supérieure ou égale à ETP décadaire ;

$\sum N_i$ (P≥ETP/2) : nombre de fois la pluie décadaire est supérieure ou égale à ETP/2 décadaire ;

N : nombre d'année totale.

Ces valeurs de probabilités représentées sur le même graphe (Fig. 2) permettent d'obtenir les périodes de cultures sans irrigation.

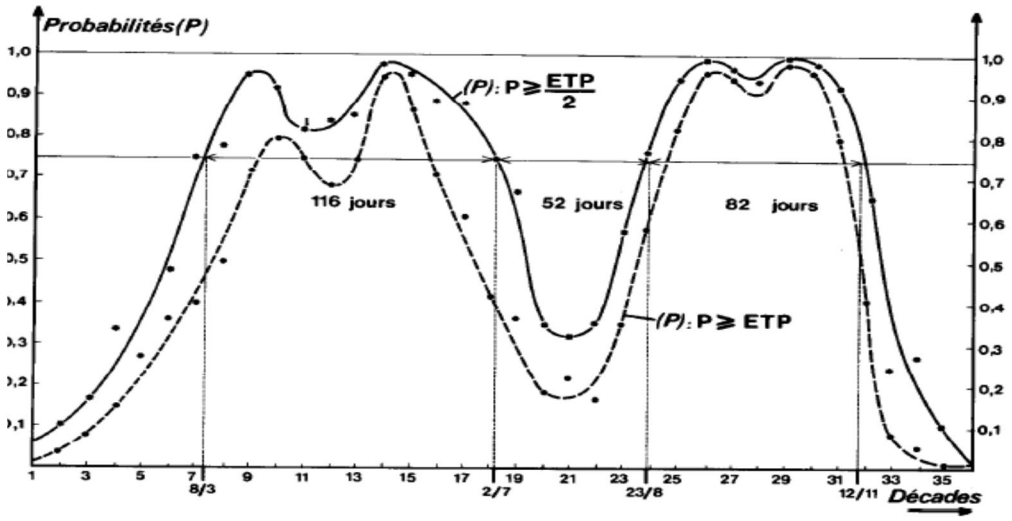


Fig. 2. Détermination des périodes végétatives en régime bimodal

Le choix d'une probabilité (q) traduisant le risque de perte partielle ou totale de la récolte (du fait du déficit hydrique) acceptable par l'agriculteur, permet de définir la période de culture non irriguée correspondante (Figure 2). Elle est lue sur la courbe de P≥ ETP/2. Dans le cadre de cette étude, cette probabilité est égale à 0,25 c'est-à-dire que le cultivateur accepte une perte partielle ou totale de la récolte une année sur quatre. La probabilité complémentaire P=1- q = 0,75 (ou trois années sur quatre) exprime les chances de ne pas perdre la récolte. La droite représentant cette probabilité P = 0,75 sur le graphique de la figure 2 coupe la courbe donnant l'évolution des probabilités d'obtenir des pluies (P) supérieures à ETP/2 en des points qui déterminent le début et la fin de chaque période végétative.

3.3 CALAGE DES DATES OPTIMALES DE SEMIS DES RIZ PLUVIAUX DE 90 ET DE 105 JOURS

En riziculture pluviale, la satisfaction des besoins en eau des plantes est exclusivement assurée par les précipitations. Or le développement du riz est caractérisé par une phase critique (phase épiaison-floraison) durant laquelle le manque d'eau compromet systématiquement la récolte. La détermination(ou calage) de la date optimale de semis consiste donc à trouver la date à laquelle le semis permet d'avoir le meilleur ajustement possible entre les exigences hydriques du riz pluvial et la répartition des pluies. Pour se faire, il faut faire coïncider prioritairement la phase épiaison-floraison avec la période de plus grande probabilité de P supérieures à ETP (lue sur le graphe issue de la méthode d'Eldin). Une projection en arrière de la durée de la phase végétative donne la période de semis.

4 RESULTATS ET DISCUSSION

4.1 RESULTATS

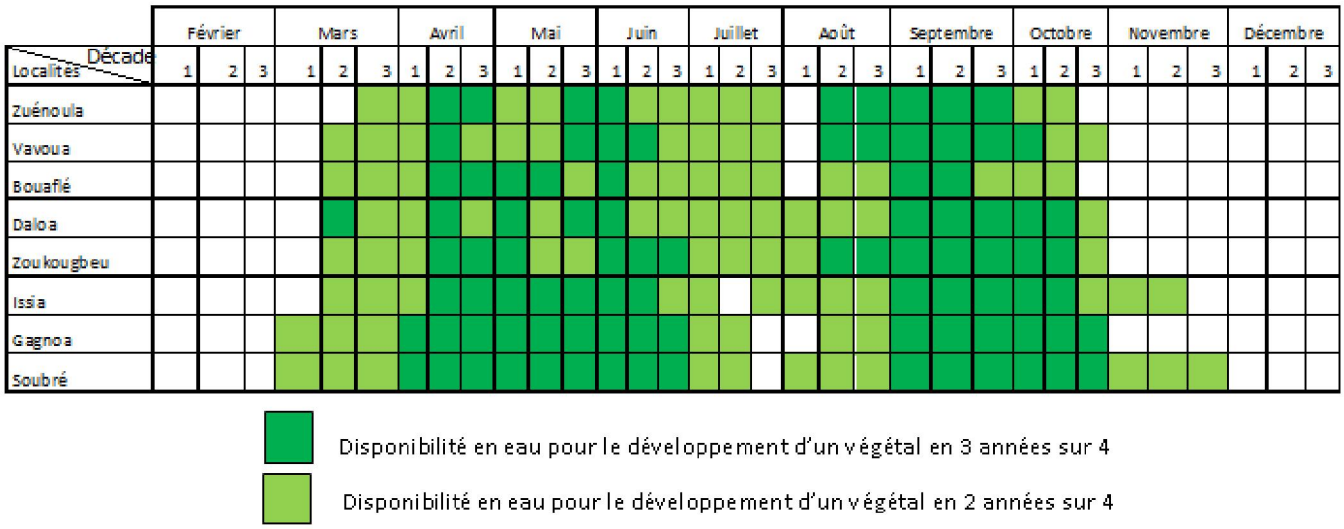
4.1.1 PERIODES VEGETATIVES EN 3 ANNEES SUR 4

Le régime pluviométrique bimodal du Centre-ouest de la Côte d'Ivoire offre deux périodes végétatives (Tableau 1).. Selon les localités, ces périodes végétatives sont plus ou moins marquées en 3 années sur 4.

Les résultats indiquent que la première période végétative s'étale sur le trimestre avril-mai-juin. A Gagnoa et à Soubré, elle démarre dans la première décade d'avril, une décade plus tôt que dans les autres localités. Elle dure au minimum 2 mois et au maximum 3 mois. Lorsqu'on remonte les latitudes en allant de Soubré à Zuénoula, cette première période végétative est de plus en plus irrégulière. Au-dessus de la latitude 6°30', certaines décades entre le 21avril et le 21 mai ne sont pas suffisamment pluvieuses.

Contrairement à la première période végétative, la deuxième période végétative est régulière. Elle est située entre le 11 août et le 10 novembre. Elle est relativement plus courte. Ces dates de démarrage s'obtiennent à partir du 10 août dans les localités de Vavoua, Zouénoula et Zoukougbeu et dans la première décade de septembre dans les autres localités. La durée en 3 années sur 4 va de 20 jours à 2 mois 10 jours

Tableau 1. Périodes végétatives au Centre-ouest de la Côte d'Ivoire



4.1.2 DATES OPTIMALES DE SEMIS

Deux périodes de culture ont été obtenues pour chaque cycle de riz pluvial, une culture en premier cycle (début avril à début juillet) et une autre en cycle unique (mi-juillet à fin octobre). Les dates de semis en premier cycle cultural se situent entre le 1er et le 20 avril excepté à Zoukougbeu où la répartition des pluies exige la pose des semences dès le mois de mars (Fig. 3). En cycle cultural unique par ailleurs, les dates de semis sont comprises entre le 11 juillet et le 11 août (Fig. 4).

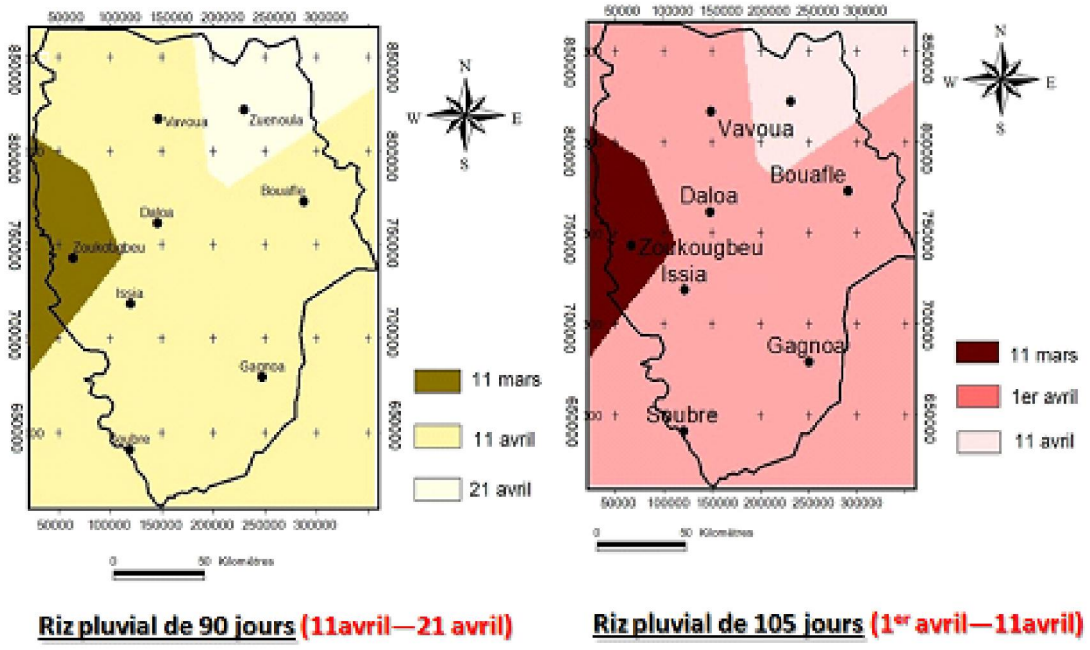


Fig. 3. Date optimale de semis en premier cycle cultural

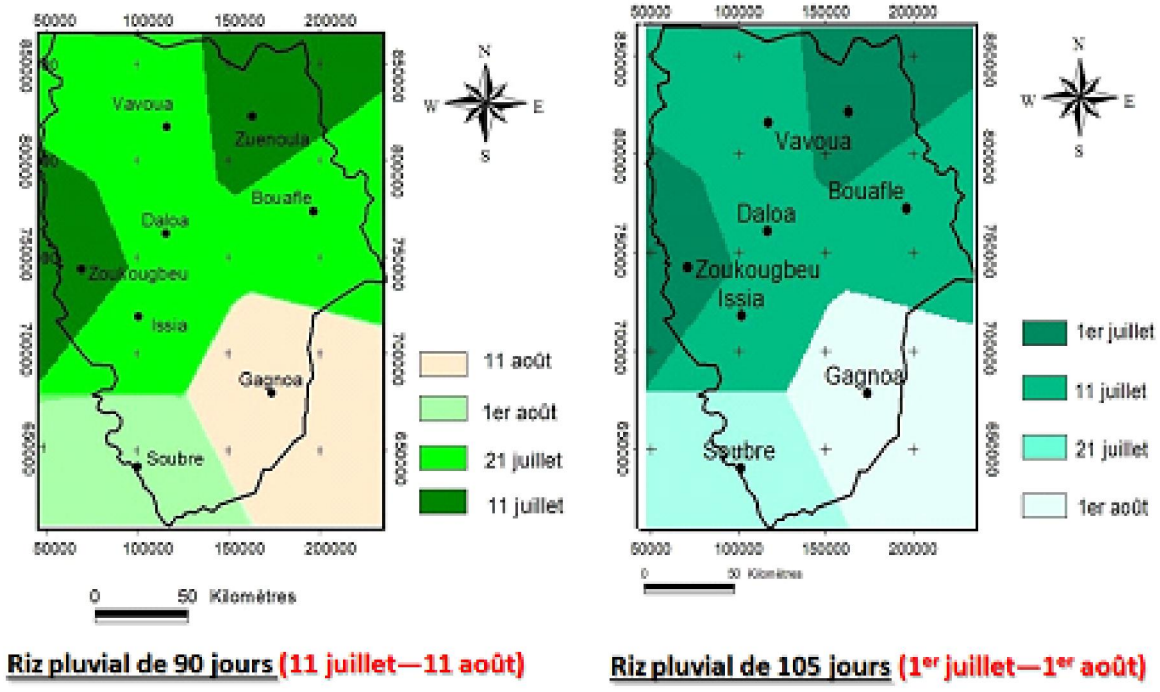


Fig. 4. Date optimale de semis en cycle cultural unique

4.2 DISCUSSION

Avec sa durée moyenne de 2 mois 20 jours en 3 années sur 4, la première période végétative donnerait plus de garantie à la réussite des cycles de riz. Mais son choix ne peut se faire de façon systématique sans tenir compte de sa régularité intra saisonnière. Car il est constaté des périodes plus ou moins sèches qui interviennent après le départ. Ceci ne favoriserait pas la réussite du riz pluvial surtout s'il y a coïncidence avec la phase d'épiaison-floraison. C'est donc pour cette raison que le calage de cycle doit être soigneusement fait pour éviter la compromission de la récolte. Avec les dates de semis obtenues qui sont comprises entre le 1er et 20 avril conformément aux résultats de Koné [2], la phase critique est mise à l'abri de cette période défavorable.

La seconde période végétative, quant à elle n'enregistre pas de sécheresse intra saisonnière. Mais elle est de courte durée (au maximum 2 mois 10 jours). Avec des cycles de riz pluvial qui durent 3 mois minimum, elle ne peut visiblement pas garantir la couverture totale des besoins en eau d'un cycle cultural. Alors, l'option d'une culture basée uniquement sur cette deuxième période végétative n'est pas envisageable. Néanmoins, le caractère moyennement pluvieux des décades entre les deux saisons végétatives permet de réussir des cultures en cycle unique dans certaines localités. Il s'agit de baser le cycle cultural sur les deux périodes végétatives en semant à la fin de la première. Pour cela, il faudra des sols aux mêmes caractéristiques que les sols argilo limoneux [13] qui ont de bonnes capacités de rétention afin de garantir la réserve hydrique nécessaire à la phase végétative. Ce faisant, le riziculteur donnera une plus grande chance à la floraison d'être couverte par les pluies abondantes du mois de septembre.

5 CONCLUSION

Le Centre Ouest est la plus importante zone rizicole de la Côte d'Ivoire. Depuis 1970, la culture du riz pluvial y est perturbée par la variabilité accrue de la pluviométrie. L'objectif de cette étude était de caler les périodes de semis de deux cycles de riz pluvial (90 et 105 jours).

L'approche a consisté à identifier premièrement les dates de début et de fin des périodes végétatives. Il ressort que ces dates qui sont utiles à la planification sont assez instables. Néanmoins les dates d'apparition 3 années sur 4 ont été retenues. Le début de la première période végétative s'obtient entre le 1er et 10 avril et celui de la seconde entre le 10 août et le 1er septembre. Les périodes de semis pour tirer parti de ces périodes végétatives sont également doubles pour chaque cycle de riz pluvial. Elles varient selon les localités. Les semis entre le 1er et 20 avril sont recommandé pour une exploitation en premier cycle cultural et les semis entre le 1er juillet et le 11 août pour une exploitation en cycle cultural unique.

REFERENCES

[1] A. Coulibaly, "Calage des cycles de végétation, Annale du service météorologique", *ASECNA Abidjan*, pp. 17- 19, 1979.

[2] D. Kone, "Calage des cycles du maïs et du riz pluvial dans la zone forestière de Côte d'Ivoire", *Note technique n°12 / 90 / syst. DCV- IDESSA. Bouaké,* 10 p, 1990.

[3] K.E. Kouakou, B.T.A. Goula, I. Savané, "Impacts de la variabilité climatique sur les ressources en eau de surface en zone tropicale humide : Cas du bassin versant transfrontalier de la Comoé (Côte d'Ivoire - Burkina Faso)", *European Journal of Scientific Research*, 16 (1), pp. 31-43, 2007.

[4] B. T. A. Goula, W. F. Kouassi, V. Fatika, K. E. Kouakou, I. Savané, "Impacts du changement climatique et de la variabilité climatique sur les eaux souterraines en zone tropicale humide : cas du bassin versant de la Comoé", *International Association of Hydrological Sciences*, 334, pp. 190-202, 2009.

[5] B. T. A. Goula, I. Savané, B. Konan, V. Fatika, G.B. Kouadio, "Impact de la variabilité climatique sur les ressources hydriques des bassins de N'Zo et N'Zi en Côte d'Ivoire (Afrique tropicale humide)", *Vertigo*, 7 (1), pp. 1-12, 2006.

[6] B.T.A. Goula, B. Srohourou, A.B. Brida, K.A. N'zué, G. Goroza, "Determination and variability of growing seasons in Côte d'Ivoire," *International Journal of Engineering Science and Technology*, Vol. 2(11), pp.5993-6003, 2010.

[7] B.T.A. Goula, B. Srohourou , A.B. Brida, K.A. N'zué, G. Goroza, "Zoning of rainfall in Côte d'Ivoire," *International Journal of Engineering Science and Technology*, Vol. 2(11), 2010, pp. 6004-6015, 2010".

[8] J. M. Avenard, "Aspect de la géomorphologie in : Milieu naturel de Côte d'Ivoire", *Mémoire ORSTOM, Paris, France*, 50, pp. 8-73,1971.

[9] Y. Monnier, *Les sols*, In : *Atlas de la Côte d'Ivoire,* Jeune Afrique, Paris, pp. 20-21, 1978.

[10] M. Eldin, "Risques climatiques, élément des risques encouru pour la production agricole," *Mémoire ORSTOM*, pp. 231-238, 1985.

[11] H.W. Aaiders, "Cartes de pluviométrie minimale par décade à prévoir avec une probabilité de 90%", *Projet d'Agro-Pédologie, 55 cartes, Cotonou - République Populaire du Bénin,* Etude N°188, 5 p, 1978.

[12] P. Franquin "Analyse agro climatique en régions tropicales. Méthode des intersections et période fréquentielle de végétation," *Agrom. Trop*., Vol.28 (6-7), pp. 665-682, 1973.

[13] H. Grimbert, L. Bargeot, "Potentialité agro-climatique des sols, Essai de modélisation du risque de déficit hydrique pour la Bresse," *Mémoire d'Ingénieur d'Agronomie*, France, 23 p, 200.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 727-731

http://www.issr-journals.org/ijias/

Modification in Formula of Resonating Frequency of Equilateral TMPA for Improved Accuracy and Analysis

Tajeswita Gupta, P. K. Singhal, and Vandana Vikas Thakre

Department of electronics and communication MITS Gwalior
Gwalior, M.P. 474005, India

ABSTRACT: In the present work, an equation has been developed to calculate the side length of the equilateral triangular patch for given resonant frequency. With the previously existing equation poor accuracy has been achieved for the performance parameters of the triangular microstrip patch antenna. Modified formula enhances the accuracy of the performance parameters of the triangular microstrip patch antenna. Proposed formulae provide a nonlinear relationship between the resonating frequency and the dimensions of the equilateral triangular patch antenna i.e. the required performance parameters and the design parameters used in the simulation and fabrication process. With the help of calculated dimension, performance of triangular patch antenna has been analyzed for particular resonating frequency. Simulated results of proposed equation have been compared with the existing equation, and better accuracy in the results has been achieved with the improved formula. Proposed equation has been verified by simulating and fabricating various antennas, and measured results are found satisfactory.

KEYWORDS: Triangular microstrip patch antenna, resonant frequency, Curve fitting procedure, Impedance Bandwidth, Return Loss.

1 INTRODUCTION

Literature reveals, a lot of work has already been done on the resonant frequencies of the Equilateral Triangular Microstrip Patch Antenna [1-6], based on cavity model analysis. The whole reported work was unanimous with the various correction factors applied to the existing formula obtained from the cavity model analysis and the same values were compared with the measurements of Dahele & Lee [4]. The resonant frequency of equilateral triangular microstrip patch antenna was grounded on the effective side length and an effective relative permittivity [4]. On the contrary, Garg and Long [5], proposed an alternative expression with effective side length and the actual substrate permittivity was used for the calculation of resonating frequency. However Gang [6] also argued that the effective values of both side length and relative permittivity should be considered for calculations. Triangular shape for printed microstrip patch antenna comes as a good option due to its small surface area as compared to the other printed shapes like rectangular and circular patches. The advantage of such structure is the ability of building compact antennas with low manufacturing cost and high reliability. However, it is in practice difficult to accomplish this while at the same time achieving high bandwidth and efficiency. Nevertheless, improvements in the properties of the dielectric materials and in design techniques have led to enormous growth in the popularity of microstrip patch antennas, and there are now a large number of commercial applications.

To compute the dimensions of radiating triangular patch an improved formula has been developed in this paper. The prime objective of the research is to present a simplified formula with enhanced accuracy. Proposed formula provides a relation between the side length and its corresponding resonant frequency, by considering the effective value of side length and the actual substrate permittivity. Proposed equation yields better accuracy when the calculated design parameter has been simulated as compared to the simulated result of the design parameters provided by Dahele & Lee and Garg & Long [4]-[5]. Authenticity of the equation has been verified by various aspects viz. by simulating a number of antennas in CST

Corresponding Author: Tajeswita Gupta (tajeswitagupta@gmail.com)

simulation software [7-9] using various dielectric materials as substrate and by fabricating few antennas to verify the practicability of the proposed equation.

2 DEVELOPMENT OF EQUATION

For the development of the proposed equation a wide range of data has been generated by simulating a large number of Equilateral Triangular Microstrip Patch Antennas of different side lengths varying from 15-135 mm, in CST microwave studio [7-9], and the resonating frequencies corresponding to them have been noted. With the obtained data a smooth curve has been fitted using curve fitting procedure, among the side length corresponding to their resonating frequencies. In the procedure, a best fitted equation having least root mean square error has been chosen among the various options. In the process of determination of data, the height of the substrate is kept constant at 1.6 mm. Geometry of Equilateral Triangular Patch Antenna has been pictured in figure 1.

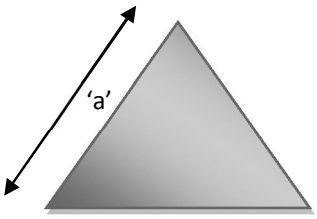


Fig. 1. Geometry of Equilateral Triangular Microstrip Patch Antenna

Let 'a' be the side length of the Equilateral Triangular Microstrip Patch Antenna, and 'f' is the designing resonating frequency. Curve fitting procedure has been used to develop the equation among the data. Curve fitting is the process of constructing a curve, or mathematical function that has the best fit to a series of data points, possibly subject to constraints. Curve fitting can involve either interpolation, where an exact fit to the data is required, or smoothing, in which a "smooth" function is constructed that approximately fits the data. The proposed formulated equation using curve fitting mathematical procedure is given as in equation (1).

$$a = \frac{343.4 * f^{-1.009}}{\sqrt{\in_r}} \quad (1)$$

' $\in_r$' is the relative substrate permittivity ranges between $2.2 \leq \in_r \leq 12$.

3 DETERMINATION OF PATCH DIMENSION USING PROPOSED EQUATION

The side length of equilateral triangular microstrip patch antenna is calculated using proposed formula mentioned in equation (1) for specified parameters viz.

Relative dielectric constant $\in_r$ = 3.75

Thickness of substrate h = 1.58 mm

Designing frequency = 5 GHz

Thus, the side length of the patch calculated theoretically according to proposed equation (1) is given below:

$$a = \frac{343.4 * f^{-1.009}}{\sqrt{\in_r}} = \frac{343.4 * 5^{-1.009}}{\sqrt{3.7\ 5}}$$

= 34.965 mm

When a triangular patch having side of 34.965 mm as stated above, has been simulated in CST, it should resonate at the designing frequency i.e. at 5 GHz or as near as possible to the designing frequency to achieve the accuracy. After simulation using CST microwave studio, the patch resonates at 4.977 GHz. The absolute error between the designing frequency and the resonating frequency of the patch is 23 MHz

4 COMPARISON AND ANALYSIS

A comparative study of the proposed equation and previously existing equation is requisite to analyze the accuracy of the proposed equation. According to the cavity model analysis, the resonating frequency corresponding to various modes is being given by equation (2), [1]-[2]

$$f_T = \frac{2C}{3a\sqrt{\epsilon_r}} \quad (2)$$

Where, C = 3*(10)^8 m/s, is the velocity of light

The side length of an equilateral triangular microstrip patch antenna has been computed using equation (2) for the same parameters stated in section III. When the simulated results of both equations are compared, the design parameter calculated using previously existing equation (2) shows more deviation from the designing frequency as compared to the design parameter computed by new proposed equation, which is drawn in Figure 2. Observations made from Figure 2, that for f = 5 GHz, patch designed using proposed equation resonates at 4.977 GHz whereas when designed using previously existing equation (2), it resonates at 4.4837 GHz. It is clearly observed that simulated results of the patch designed using proposed equation resonates closer at the designing frequency as compared to the previous equation (2).

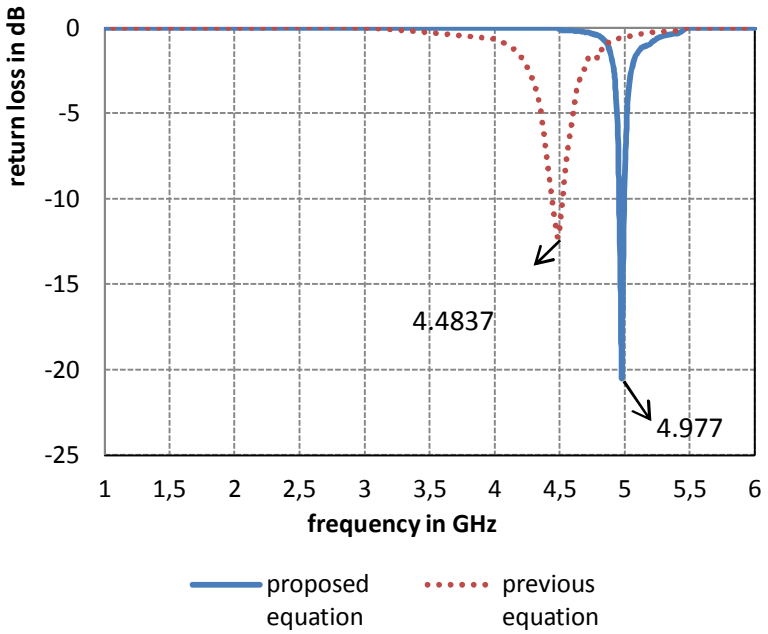


Fig. 2. Comparison of the Antenna Simulated Using Proposed and Existing Equation

To analyze better, few more antennas have been simulated using CST for the design parameter of the patch calculated using proposed equation mentioned in equation (1) and previously existing equation (2) respectively. The dielectric substrate FR4 having dielectric constant 4.4 at substrate height 1.6 mm has been taken as specified designing parameters. Simulated results of patch, calculated using proposed and previous equation have been shown in Table 1.

Table 1 provides a comparative study of both the equations, proposed and previous equation respectively. Simulated results are compared to determine better accuracy. Table 1 reveals the simulated frequencies f1 for the side length calculated using proposed equation (2) are much closer to the corresponding designing frequencies as compared to f1'. It is clearly observed and concluded that when the dimensions of the patch is theoretically calculated, using proposed equation yields better accuracy in the simulated results as compared to the dimensions calculated using previous equation.

Table 1. Comparative Study of Proposed Equation with the Dahele & Lee [4] Measurement Method

Designing Frequency f (GHz)	Simulated Resonating frequency	
	Side calculated using proposed equation f1(GHz)	Side calculated using previous equation (1) f1'(GHz)
1.9	1.901	1.88
3.2	3.25	3.12
4	4.06	3.86
6.5	6.6	6.12
8	8.22	7.44

For the verification of the proposed equation, Table 2 shown below provides simulated resonant frequencies with their different designing parameters. For the various designing frequencies, their corresponding side lengths have been calculated using proposed equation and the patches of the calculated dimension have been simulated in CST microwave studio. Table 2 draws the comparison of results among the calculated values of proposed equation and simulated values. Testing of equation was performed for those values of the designing frequencies which are not included in the development of equation and the results were found satisfactory as shown in Table 2.

It can also be observed from table 2 that the errors among the designing frequencies and simulated frequencies of their corresponding patch are very small. The patches of calculated side length, using proposed equation, when simulated in CST microwave studio, resonate closer to the designing frequency. The mean square error is very low for each value of designing frequency. The root mean square error for the proposed equation is 0.4618.

Table 2. Analysis of Proposed Equation for Various Designing Resonant Frequencies of Equilateral Triangular Microstrip Patch Antenna with Optimum Feed

Designing resonant frequency f (GHz)	Calculated side a(mm) using equation (1)	$\in_r$	simulated frequency f' (GHz)	Absolute error |f-f'|	Mean square error 'e'
0.9	182.07	4.4	0.9	0	0
1.2	192.61	2.2	1.196	0.004	1.6e-05
2	66.92	6.5	2.025	0.025	6.25e-04
3.2	33.75	9.9	3.242	0.042	0.0018
5.1	34.26	3.75	5.075	0.025	6.25e-04
6	22.08	6.5	6.05	0.05	0.0025

In order to affirm the results of the proposed equation, few antennas have been simulated and fabricated as shown in Table 3. Dielectric constant of substrate is 4.4 and the height of the substrate is 1.6 mm has been taken in the account as specified design parameters.

Table 3. Simulated and Measured Results of the Fabricated Antennas

Designing Frequency In GHz	Calculated side In mm	Simulated frequency In GHz	Measured frequency In GHz
2	81.34	2.004	2.1
2.4	67.67	2.412	2.38
2.9	55.913	2.89	2.902

Figure 3(a) shows one of the fabricated antennas, resonating at 2.1 GHz while designed for resonating frequency 2 GHz, as listed in Table 3. Figure 3(b) shows a comparative chart of simulated and measured results of the antenna drawn in Figure 3(a), reveals that measured results are very close to the simulated results. Two more antennas have been fabricated and their measured results have been shown in the Table 3, which reveals that measured results are very close to the designing frequencies. An examination of Table 1 and Table 3 clearly shows that, the results theoretically calculated by the proposed equation are in good agreement with the results of fabricated antenna with good accuracy.

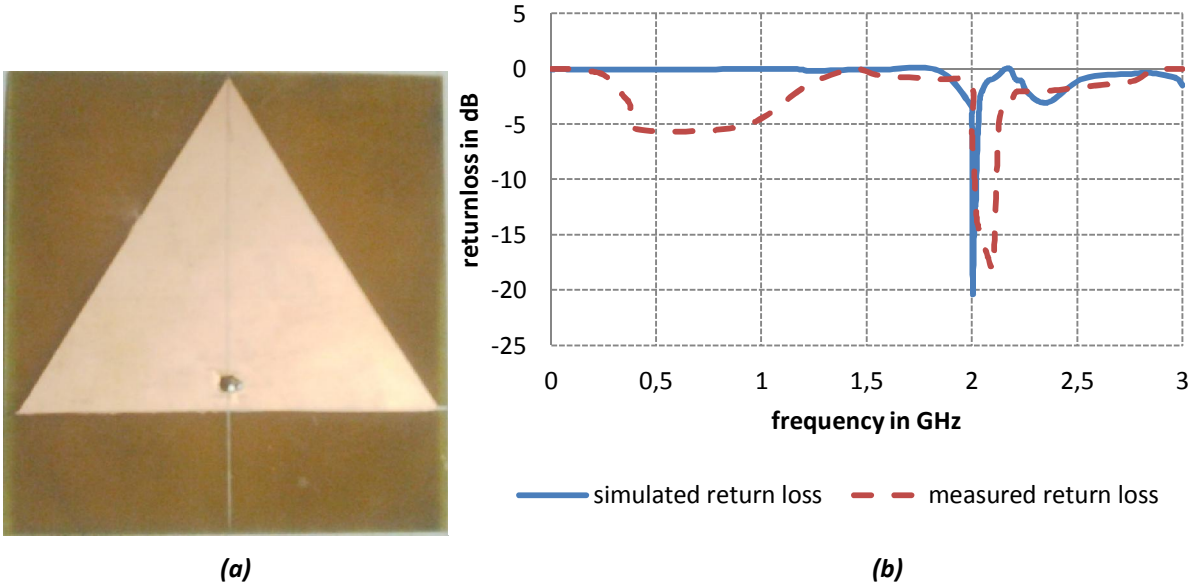


Fig. 3. (a) Fabricated Antenna at 2 GHz (b): Simulated and Measured RL of Fabricated Antenna

5 CONCLUSION AND REMARKS

By analyzing the data very carefully an improved formula relating the side length and the corresponding resonating frequency has been developed in this paper. Resonating frequency of the patch antenna is inversely proportional to its dimensions. Proposed equation relates the dimensions to the patch in nonlinear manner. Calculated side length values using proposed equation for equilateral triangular microstrip patch antenna for various dielectric materials and constant substrate height when simulated using CST for centrally and arbitrarily fed, shows a good measure of accuracy. Few of them have been fabricated in order to verify the relevance of the equation practically. Proposed equation shows a higher measure of accuracy among the theoretically analyzed and simulated values in CST and the fabricated measured values.

REFERENCES

[1] C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd Edition, John Wiley and Sons, New York, 2005.
[2] Bahl, I. J. and P. Bhartia, *Microstrip Antennas*, Chap. 4, Artech House, Dedham, MA, 1980.
[3] J. Helszain, D.S. James, "Planar Triangular Resonators with Magnetic Walls", *IEEE Trans Microwave Theory Tech*, vol. 26, pp. 95-100, 1978.
[4] J. S. Dahele and K. F. Lee, "On the resonant frequencies of the triangular patch antenna," *IEEE Transactions on Antennas and Propagation*, AP-35 100-101, 1987.
[5] R. Garg and S. A. Long, "An Improved Formula for the Resonant Frequencies Of The Triangular Patch Antenna," *IEEE Trans. Antennas Propagation*, Vol. AP-36, 1988, p. 570.
[6] X. Gang, "On the resonant frequencies of microstrip antennas", *IEEE Trans. on Antennas and Propagat.*, Vol. AP-37, pp. 245–247, 1989.
[7] Rabih Rahaoui and Mohammed Essaaidi, "Compact Cylindrical Dielectric Resonator Antenna excited by a Microstrip Feed Line," *International Journal of Innovation and Applied Studies*, vol. 2, no. 1, pp. 1–5, January 2013.
[8] Mohammed Younssi, Achraf Jaoujal, Yacoub Diallo, Ahmed El-Moussaoui, and Noura Aknin, "Study of a Microstrip Antenna with and Without Superstrate for Terahertz Frequency," *International Journal of Innovation and Applied Studies*, vol. 2, no. 4, pp. 369–371, April 2013.
[9] M. I. Hasan and M. A. Motin, "New slotting technique of making compact octagonal patch for four band applications," *International Journal of Innovation and Applied Studies*, vol. 3, no. 1, pp. 221–227, May 2013.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 732-738

http://www.issr-journals.org/ijias/

The BEAC Central Bank and Wealth Creation in Cameroon Economy

Serge MESSOMO ELLE

Department of Economics and Management,
University of Buea,
Cameroon

ABSTRACT: The purpose of this paper was to assess the impact of BEAC's financing on wealth creation in the Cameroon economy. BEAC is a central bank of the six States of the Central Africa. They are Cameroon, Central Africa Republic, Congo-Brazzaville, Chad, Equatorial Guinea and Gabon. The recovery of the Cameroon economy after the economic crisis of years 1980 and 1990 led to question the role played by BEAC central bank in that attainment. The data used in the study were secondary and were collected from Franc currency zone annual reports of 'Banque de France' on Cameroon from the year 1990 to 2011. OLS regression equations, means and variances were used to explain this impact. The descriptive results revealed large variations of standard deviation values from the means. The OLS results in return showed that BEAC's financings affect negatively wealth creation in the Cameroon economy. In other words, wealth creation is significantly negatively affected by BEAC's financings to the banking system for the physical and financial wealth, while negatively affected by BEAC's financings to the government for the financial and human wealth. In order for BEAC to increase the wealth creation in Cameroon economy, it must provide productive loans directly to economic agents through their unions and long-term loans to commercial banks to finance the economy and boost consumption.

KEYWORDS: BEAC's financings, Cameroon government, banking system, Central bank, physical wealth, financial wealth, human wealth.

1 INTRODUCTION

Many scholars and policymakers diverge on the role of central banks in an economy. This is because of different operations carried by these institutions and their genesis. Central banks exercise many operations namely issuance of the country's bank notes, acting as the government bank, acting as the commercial banks' bank, serving as a lender of last resort to the banking and even the financial system as a whole, conducting monetary policy to manage the foreign exchanges and the price level. Other activities have been added to this list: setting of the monetary policy, allocating of credit to promote national goals, setting of banking policy, supervising of secondary banks, publishing of financial information and advising of government on economic policies. There are other roles that need to be associated with central banks. These are the distributive, political and allocative roles of central banks. All these operations undermine activities in the society.

Central banks' policies affect individuals, households, enterprises and the government. This is the distributive role of the central bank. The political role deals with the independence of this financial institution, while the allocative role focuses on the provision of credit to industries ([1]). According to reference [2], the historical role of central banks is to grant resources to government. This role has evolved over time. It goes back to the establishment of the Bank of England in 1964. Today, the operations of central banks can be summarized in two main functions namely the macroeconomic function aiming to maintain price stability and the microeconomic function based on the stability of the banking system ([2]). This stability can be achieved by financing the economy. Central banks finance the economy directly or indirectly. Directly, they provide funds to economic agents without using any financial intermediary. This is the example of different provisions granted by central banks to States' treasuries. Indirectly, central banks use financial intermediaries like commercial banks to finance the economy. These two methods of financing are used by "Banque des Etats de l'Afrique Centrale" (BEAC) to finance Cameroon

Corresponding Author: Serge MESSOMO ELLE (messomo2011@yahoo.fr)

economy. So, the question is: How **do** the provisions of funds by BEAC affect wealth creation in Cameroon economy? There are many types of wealth. But this study limits itself to financial, human and physical wealth to generate these specific questions: How do BEAC's financings affect financial wealth creation in Cameroon economy? How BEAC's funding influences physical wealth creation in Cameroon economy? How BEAC's provision of funds impacts on human wealth creation in Cameroon economy?

The main objective of this study is to examine how financial resources provided by BEAC create wealth in the Cameroon economy. The specific objectives explain how BEAC's funding creates physical, financial and human wealth in the Cameroon economy. This study exploits secondary data from annual reports of Franc zone from "Banque de" France. The data are analyzed using means, variances and Ordinary Least Square (OLS) regressions. The paper is divided in sections. Section 2 looks at the relationship between central banks and wealth creation. Section 3 focuses on the methodology and section 4 gives the results. Section 5 generates concluding remarks.

2 CENTRAL BANKS AND WEALTH CREATION

Many definitions have been given to wealth. But this study uses the one of [3]. He defined wealth as capital. In classical economics, capital referred to durable physical assets such as machinery and buildings that increase the value of production. There are many types of capital: physical, human, financial, social, natural, cultural and intellectual. Each of these capitals is a type of wealth. According to their functions, central banks have two main roles in the process of wealth creation.

The first role is to create a stable macroeconomic environment consisted of a stable price level, low inflation and stable exchange rates. The environment enables the accumulation of net savings for wealth creation and facilitates economic growth. A stable macroeconomic environment facilitates to keep the cost of production low and reduces the speculation in the economy. It avoids the decrease of the value of the assets and the currency. A stable environment permits positive return for investments carried on in the economy. Central banks use general and selective instruments to maintain a stable macroeconomic environment.

The second role is microeconomic. Here, central banks regulate and supervise the banking system. Central banks must regulate and supervise secondary banks in order to allow them to play the functions of the existence of banks defined by [4]. These ones are the provision of credit to the economy, the provision of financial information, the setting and the management of the payment system and the transmission of the monetary policies in the economy. The microeconomic function enables central banks to provide finances to the economy through secondary banks. Central banks by the mechanism of discount rate refinance secondary banks at a rate lower than the market rate. Secondary banks use then the money received from central banks to finance enterprises, governments and households which are agents of wealth creation. Governments from money received create physical wealth like schools, roads, bridges, human wealth like engineers, skilled and high-tech citizens. Enterprises and households increase the volume of stock assets or capital formation, savings and value added in the economy.

Central banks also finance governments directly through the treasury. The treasury is the government bank. It takes care of payment of the government's expenses. The expenses generated allow governments to create physical, financial and human wealth in the economy. All these funds provided by central banks directly or indirectly to the economy increase the Gross Domestic Product (GDP) of countries. A strong economy has a high GDP and a strong currency. It is what BEAC aspires to for Cameroon.

BEAC is the central bank of six states of the central Africa zone. These are Cameroon, Central Africa Republic, Congo-Brazzaville, Chad, Equatorial Guinea and Gabon. BEAC was created on the 22nd November, 1972 ([5]). BEAC uses three instruments to fulfill its macroeconomic and microeconomic roles. The instruments are the year-refinancing objectives, interest rates, reserve requirements. The refinancing objectives are stated yearly for each state of the region and its banking system. BEAC does not go above the amounts of financing stated for each country at the beginning of the budgetary year. The interest rates provide the charges and returns granted by BEAC for various operations of reduction and increase of the quantity of money supply in the economy, buyouts and discounting of marketable securities (commercial papers) held by commercial banks or other monetary creation institutions. The operations of money supply take place on counter A and counter B. Counter A deals with short-term negotiable credits while counter B deals with medium and long-term non-negotiable credit. The reserve requirements permit BEAC to monitor and maintain the liquidity of the Cameroon banking system. The three instruments presented have been contributing to the recovery of Cameroon economy since the devaluation of the Franc CFA on the 11th January 1994 ([6]). From the above literature review, four research hypotheses are derived:

H_1: BEAC's financings to Cameroon government and to Cameroon banking system affect physical wealth in the Cameroon economy;

H_2: BEAC's financings to Cameroon government and to Cameroon banking system affect financial wealth in the Cameroon economy;

H_3: BEAC's financings to Cameroon government and to Cameroon banking system affect the human wealth in the Cameroon economy;

H_4: BEAC's total financings to Cameroon government and banking system affect the total wealth in the Cameroon economy.

The next lines present the methodological framework. It consists of methods and instruments of data collection and analysis.

3 METHODOLOGICAL FRAMEWORK

This study is explanatory and uses time series data. The data were collected from [6] from the year 1990 to 2011 of Cameroon national accounts. Data consisted of dependent and independent variables. The dependent variables were physical wealth (capital formation), financial wealth (net savings), the human wealth and the total wealth of the Cameroon economy. The independent variables were BEAC's financings used by Cameroon government and BEAC's financings used by the banking system. The expenditure approach was used to assess net savings and capital formation. This was done by doing total output produced in the economy minus non-labour cost. The income approach was used to generate human wealth based on income distribution done by Cameroon government for human capital development. In other words, the distribution is done by expenditures made by the State for human development. These expenditures generate wealth to economic agents through salaries, remittances and capital expenditures made by the government. The total wealth is made up of the sum of physical wealth (capital wealth), financial wealth (net savings) and human wealth.

Capital formation was utilized because it portrays in this research the volume of physical wealth produced in the economy. Net savings represented the volume of financial wealth generated in the economy. Income distributions resulting from government expenditures for human development were associated with human wealth, while the total wealth is made up of the sum of all wealth. BEAC financings to Cameroon government and BEAC financings to the banking system were provision of funds by BEAC central bank to the banking system and to the government. The total BEAC's financings were made up of the sum of BEAC's financings to the government and to the banking system. BEAC's financings variables enabled to measure the relationships with the independent variables.

The tools of data used to explain the relationships between dependent and independent variables were mean, variance and the multiple linear regression equations. The descriptive tools that is, the variances and the means were employed to facilitate the understanding of the variation of individual variables vis-à-vis their means. The ordinary multiple linear regression equations permitted to assess whether the relationships studied between the dependent and the independent variables were significant, the direction of the relationships, the significant variables that affected the relationships and the proportion of the variance of the dependent variables explained by the variation of the independent variables using the t-test. The relationships expected between the dependent and the independent variables are negative relationships. This is because less wealth is created when the economy is crisis. Governments and the banking systems make use of central banks' financings mostly in periods of crises; examples of the United States of America and the European Union in the financial crisis of 2008. Central banks' financings affect the economy when they are productive loans and are not used for liquidity financing. In addition, central banks financings to governments and to banking systems in periods of crisis are used primarily for liquidity problems, except when they are long-term debt or make for the recapitalization of banks; case for instance of European banks today. Furthermore, the real actors of wealth creation in the economy are secondary banks (Commercial, development and investment banks) ([7]). In other words, the relationships between BEAC's financings and physical wealth, financial wealth, human wealth and total wealth must be negative, and hence the coefficients of the dependent variables in the equation models must have negative signs. This is because a provision of funds to the economy by BEAC is likely to decrease wealth in the Cameroon economy. The regression models estimated were as follows:

$$W_1 = \beta_0 + \beta_1 Fing + \beta_2 Finb + \varepsilon \quad (1)$$

$$W_2 = \beta_0 + \beta_1 Fing + \beta_2 Finb + \varepsilon \quad (2)$$

$$W_3 = \beta_0 + \beta_1 Fing + \beta_2 Finb + \varepsilon \quad (3)$$

$W_4 = \beta_0 + \beta_1 Fint + \varepsilon$ (4)

W_1 is physical wealth (capital formation); W_2 is financial wealth (Net savings); W_3 is human wealth (social capital expenditures, remittances and government's salaries); W_4 is total wealth (physical + financial + human wealth); Fing is BEAC's financings to the Government; Finb is BEAC's financings to the banking system; Fint is total BEAC's financings. The next lines show the results.

4 RESULTS

This section presents the results. It explores also their interpretation and discussion.

4.1 PRESENTATION AND ANALYSIS OF RESULTS

The presentation of results is done in tables and the analysis describes and explains the results. We start by exploring descriptive results.

Table 1. Descriptive results of BEAC's financings and Wealth creation

Variables	Observations (Years)	Means	Standard deviations
BEAC's financings to the government	22	120267.9	79173.39
BEAC's financings to the banking system	22	45740.68	89921.67
Total BEAC's financings	22	164402.4	114227.9
Physical wealth	22	1667997	983663.3
Financial wealth	22	854953.3	854953.3
Human wealth	22	289111.7	289111.7
Total wealth	22	3845316	1231445

Source: Adapted from "Banque de" France Franc Zone Reports (1990 – 2011)

The descriptive results of the mean and the standard deviation are showing great variation of the standard deviation values from the mean values. This means that the difference between the various observation values is big. This is explained by the difference of financing of BEAC to the government and to the banking system between the years of recession and recovery of Cameroon economy. The data collected to explain the impact of BEAC on wealth creation (see Appendix) showed large amounts of financings of BEAC to the economy during the years of recession and small amount during the years of recovery of the Cameroon economy (see Appendix). In terms of wealth, high variations of the standard deviation values from mean values are explained by the fact that less wealth was produced during the years of recession and more wealth during the years of recovery and positive growth of the Cameroon economy. This gives then great discrepancies between years of recession and recovery and growth of the Cameroon economy (see Appendix). The next analysis focuses on OLS regression results.

Table 2. OLS regression results of BEAC's financings and Wealth creation in Cameroon economy

Independent Variables	Physical wealth	Financial wealth	Human wealth	Total wealth
BEAC's financings to government	- 2.583	- 3.204*	- 2.02***	/
BEAC's financings to the banking system	- 6.536***	- 6.35***	- 0.838	/
BEAC's total financings	5.73**	9.35***	5.03**	- 10.809***
F-test P-value	0.011	0.001	0.017	0.003
R^2	0.38	0.50	0.35	0.35
N	22	22	22	22
Hypotheses	H_1	H_2	H_3	H_4

Source: Adapted from "Banque de" France Franc Zone Reports (1990 – 2011)

The results of OLS regression equation of BEAC's financings and physical wealth (H_1) showed that the equation is significant at 5%. This means that BEAC's financings affect physical wealth creation (capital formation). R^2 showed that the variation of independent variables BEAC's financings to the government and BEAC's financings to the banking system leads to the variation of the dependent variable capital formation of 38%. The t-test showed that BEAC's financings to the banking system are the significant variable affecting the variation of capital formation in the Cameroon economy. The negative coefficient of BEAC's financings to the banking system revealed that they reduce the volume of investment done in Cameroon economy and hence the physical wealth produced by the economy.

The results of OLS regression equation of BEAC's financings and financial wealth (H_2) showed that the equation is significant at 1%. This means that BEAC's financings affect financial wealth creation (net savings). R^2 showed that the variation of independent variables BEAC's financings to the government and BEAC's financings to the banking system leads to the variation of the dependent variable financial wealth of 50%. The t-test showed that BEAC's financings to the banking system and BEAC's financings to the government are significant variables affecting the variation of financial wealth in the Cameroon economy. The negative coefficient of BEAC's financings to the banking system disclosed that they reduce the volume of financial wealth generated in the Cameroon economy.

The results of OLS regression equation of BEAC's financings and human wealth (H_3) explained that the equation is significant at 5%. This means that BEAC's financings influence human wealth creation. R^2 showed that the variation of independent variables BEAC's financings to the government and BEAC's financings to the banking system leads to the variation of the dependent variable human wealth of 35%. The t-test revealed that BEAC's financings to the government are the significant variable affecting the variation of human wealth in the Cameroon economy. The negative coefficient of BEAC's financings to the government divulged that they reduce the volume of human wealth created in the Cameroon economy.

The results of OLS regression equation of BEAC's financings and total wealth (H_4) showed that the equation is significant at 1%. This means that BEAC's financings affect total wealth creation in the Cameroon economy. R^2 showed that the variation of independent variables BEAC's financings to the government and BEAC's financings to the banking system leads to the variation of the dependent variable total wealth of 35%. The t-test showed that total BEAC's financings to the economy are the significant variable affecting the variation of total wealth produced in the Cameroon economy. The negative coefficient of total BEAC's financings disclosed that they reduce the volume of total wealth produced in the Cameroon economy. These results give the following OLS regression equations:

$$W_1 = 2277562 - 2.583Fing - 6.536*** Finb \quad (1)$$

$$W_2 = 2255346 - 3.204*Fing - 6.35***Finb \quad (2)$$

$$W_3 = 878902.8 - 2.02***Fing - 0.838Finb \quad (3)$$

$$W_4 = 5622368 - 10.801***Fint \quad (4)$$

The next lines interpret and discuss the results. The interpretation and discussion center on the results provided by the different OLS equation models, specifically the significant variables provided by the t-test and focus on the impact of the signs of coefficients of independent variables on wealth creation.

4.2 Interpretation and Discussion of Results

The results of H_1, H_2, H_3 and H_4 showed that BEAC's financings affect the variation of physical wealth, financial wealth, human wealth and total wealth. All the four hypotheses studied were significant (see Table 2). BEAC's financings to Cameroon government and to the banking system decrease physical, financial, human and total wealth creation in the Cameroon economy. The R^2 explains that BEAC's financings and wealth creation have an inverse relationship that is, when BEAC's financings decrease, wealth creation increases; vice-versa. The t-test revealed that BEAC's financings to the banking system are significant for the wealth creation for H_1 and H_2. The coefficients of these variables show that an increase of BEAC's financings to the banking system reduces physical and financial wealth creation, while their reduction increases this wealth in the economy. This result is explained by the fact that Cameroon banking system is over-liquid and use rarely BEAC's facilities to finance the economy. Most banks after the restructuration of the year 1990s and the setting of mechanisms of supervision by monetary authorities meet the solvency requirements and are over-liquid ([6]). The financing of the economy by the banking system from their own resources enable banks to avoid the strengthening of regulations on them by the monetary authorities. In other words, the use of the central bank's financings shows that the banking system has liquidity problems and cannot finance the economy, and hence less wealth is created. Thus a negative sign or relation between BEAC's financings to the banking system and physical and financial wealth is explained by the recession of the Cameroon economy. During that period, the banking system borrowed more from BEAC to meet its liquidity problems and could not

finance the economy, hence less wealth was created. In other words, wealth is created when secondary banks use their own resources to finance the economy. This can be justified by the years of recovery and growth of Cameroon economy. In this period of growth of Cameroon economy, BEAC provides smaller amounts of funds to the banking system comparing with amounts provided in the recession time (see the Appendix).These explanations also elucidate the results of the variation of financial and human wealth by the BEAC's financings to the government (H_2 and H_3). The data of the sample studied showed that Cameroon government borrowed large amounts of money from BEAC during the years of recession and smaller amounts and even zero amounts in the era of recovery and growth (see Appendix). During the years of recession, wealth creation could not be carried by the government because the government could generate itself finances. This then justifies the negative sign or relationship of BEAC's financings to the government and financial and human wealth (see Table 2). The OLS regression showed a negative relationship or sign between BEAC's total financings and total wealth (H_4) because BEAC's financings are not an instrument of wealth creation. The wealth creation is mainly affected positively by the secondary banks because they are the providers of liquidity to the economy and transmitters of the monetary policies ([4]).

5 CONCLUSION

The aim of this study was to assess the impact of BEAC's financings on wealth creation in the Cameroon economy. OLS regression equations, means and variances were used to explain this impact. The descriptive results showed large variations of standard deviations from mean values. This is explained by big discrepancies between the years of recession and recovery of the Cameroon economy.

The OLS results in return showed that BEAC's financings affect wealth creation (physical, financial, human and total) in the Cameroon economy. Wealth creation is significantly negatively affected by BEAC's financings to the banking system and to Cameroon government. In order for BEAC to increase the wealth creation in Cameroon economy, it must provide productive loans to economic agents directly through business syndicates and long-term loans to commercial banks to finance the economy.

REFERENCES

[1] G. Epstein, *Federal Reserve Politics and Monetary Instability*, in Alan Stone and Edward J. Harpham, eds., *The Political Economy of Public Policy*, Beverly Hills: Sage Publications, 1982.

[2] F. Capie, C. Goodhart, S. Fischer and N. Schnadt, "The Development of Central Banking," *The Future of Central Banking; The Tercentenary Symposium of the Bank of England,* Cambridge: Cambridge University Press, pp. 1-232, 1994.

[3] A. Marshall, *Principles of Economics*, London, Macmillan and Co, Ltd, 1890.

[4] E .G. Corrigan, *Are Banks Special?* Annual Report, Federal Reserve Bank of Minneapolis, pp. 2-4, 1982.

[5] BEAC*, Présentation de la BEAC,* Banque des Etats de l'Afrique Centrale, 2012. [Online] Available : www.beac.int (October 5, 2012).

[6] Banque de France (1990-2011), *Rapports Annuels de la Zone Franc*, Euro système, France, (1990-2011).

[7] E. S. Messomo, "A prospective Approach on the Role and Challenges of Traditional Banks in African Recovering Economies: A case study of Cameroon", *GERER & DECIDER*, Vol. 2 no. 2, pp. 35-53, 2008.

[8] National Institute of Statistics [NIS], *Literate population in Cameroon*, NSI, Yaoundé, 2010.

APPENDIX

Data of the study (in million Francs CFA) [1]

Years	BEAC's government financings	BEAC's banking system financings	Total BEAC's financings	Physical wealth (capital formation)	Financial wealth (Net savings)	Human wealth (Social expenditures)	Total wealth
1990	100471	273364	373835	416733	313335	501377	1231445
1991	101471	266274	367745	448100	324700	520100	1292900
1992	111692	254231	365923	492500	519400	429200	1441100
1993	112692	52053	164745	515000	526600	391500	1433100
1994	93254	27432	120686	533000	808600	291000	1632600
1995	107692	21716	129408	597500	869200	297500	1764200
1996	107692	4900	112592	789900	831300	269100	1890300
1997	100459	13671	114130	1028400	1126800	339900	2495100
1998	113709	25508	139217	1122200	1120400	358000	2600600
1999	144043	17174	161217	1217600	1433000	522100	3172700
2000	149321	4249	153570	1489300	1584300	542600	3616200
2001	220826	2623	223449	2054400	1700200	544700	4299300
2002	250054	595	250649	2286200	1883800	516400	4686400
2003	256732	11000	267732	2155200	1773500	574900	4503600
2004	252953	5170	258123	2167300	1886300	611600	4665200
2005	176348	484	176832	2297800	2074400	565600	4937800
2006	166054	685	166739	2385900	2306300	621200	5313400
2007	80431	705	81136	2620100	2452300	729900	5802300
2008	0	724	724	2678800	2842000	907200	6428000
2009	0	16295	16295	2908600	2586800	1229800	6725200
2010	0	1327	1327	3371500	2837000	1120800	7329300
2011	0	6115	6115	3119900	2950700	1265600	7336200

Source: "Banque de" France Reports Franc Zone (1990 – 2011)

[1] 655.957 F CFA = 1 euro

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 739-748

http://www.issr-journals.org/ijias/

Caractérisation mécanique et physique d'un béton léger à base de sable calcaire et des granulats de liège

[Mechanical and physical characterization of lightweight concrete based on limestone sand and cork granules]

Brahim LAOUD

Department of civil engineering,
University of Amar Telidji,
Laghouat, Algeria

ABSTRACT: Faced with the growing needs of resource materials and the requirements of environmental protection in a vision of sustainable development, it became necessary to study all the possibilities of reuse and recycling of industrial wastes and by-product especially in the field of civil engineering. There is an increasing interest in limestone fines from limestone quarries in concrete construction to overcome inherent deficiencies in river sand in particular regions of North Africa. Limestone sand is a by-product of the quarry process and typically does not have a significant demand due to its high content of small particles whose diameters are less than 80 µm. This work is part of the promotion of local products such as waste from the quarry crushing (limestone sand) in the region of Laghouat (Algeria) and waste from the manufacture of agglomerated expanded cork, using them in the making of cement materials for construction primarily for thermal insulation. To study the influence of the size of the aggregates on the physico-mechanical and thermal properties of concrete, two classes of granular aggregates are used 3 / 8 and 8 / 15. The results obtained showed that the reduction weight of concrete containing crushed sand with expanded cork significantly reduces its thermal conductivity which improves their thermal insulation, but causes a drop in the strength of concrete produced. Lightweight concrete with expanded cork granules 3/8 are the lightest and most isolated, and they have low mechanical strength compared to concrete with lightweight aggregates 8/15.

KEYWORDS: Sustainable development, valorization of wastes, byproducts, limestone sand, expanded cork, insulation.

RESUME: Devant les besoins croissant des ressources en matériaux, aussi que les exigences de préservation de l'environnement dans une vision de développement durable, il est devenu nécessaire d'étudier toutes les possibilités de réutilisation et de valorisation des déchets et sous-produits industriels notamment dans le domaine de génie civil. Il ya un intérêt croissant pour les fines calcaire à partir de carrières de concassage, dans la construction en béton pour surmonter les déficiences inhérentes à sable de rivière dans certaines régions d'Afrique du Nord. Le sable calcaire est un sous-produit du processus de carrière et typiquement ne pas avoir une demande importante en raison de sa teneur élevée en petites particules dont le diamètre est inférieur à 80 µm. Ce travail s'inscrit dans le cadre de la valorisation des produits locaux tels que les déchets de la carrière de concassage (sable calcaire) de la région de Laghouat(Algérie) et les déchets issues de la fabrication des agglomérées de liège expansé, en les utilisant dans la confection des matériaux cimentaires de construction destinés en premier lieu à l'isolation thermique. Pour étudier l'influence de la grosseur des granulats sur les caractéristiques physico-mécaniques et thermiques des bétons, deux classes granulaires sont employées granulats 3/8 et 8/15. Les résultats obtenus ont montré que l'allégement du béton à base de sable de concassage avec le liège expansé diminue considérablement sa conductivité thermique ce qui améliore leur isolation thermique, mais entraine une chute de la résistance mécanique des bétons élaborés. Les bétons allégés avec les granulats de liège expansé 3/8 sont les plus légers et les plus isolants et possèdent des résistances mécaniques faibles par rapport aux bétons allégés avec les granulats 8/15.

MOTS-CLEFS: Développement durable, valorisation des déchets, sous-produits, sable calcaire, liège expansé, isolation.

Corresponding Author: Brahim LAOUD (brahimdgc@yahoo.fr)

1 INTRODUCTION

Les déchets sont de plus en plus variés et leurs quantités ne cessent d'augmenter, ce qui influe négativement sur l'environnement. La valorisation et le traitement de ces déchets favoriseront la protection et la préservation de l'environnement, ce qui nous permettra de s'inscrire dans une politique de développement durable.

Dans le secteur de la construction, Les matières végétales, renouvelables et respectueuses de l'environnement, sont appréciées pour leur légèreté, qui leur confère des propriétés d'isolation thermique importantes.

Plusieurs travaux ont été faits sur l'incorporation des granulats de liège au ciment et au plâtre pour obtenir un matériau composite thermiquement isolant [1], [2], [3], [4], [5].

Le sable concassé est très abondant dans les carrières de gravier calcaire et son utilisation pourrait réduire considérablement les coûts des matériaux de béton, minimiser le coût d'élimination de la poussière, réduit les pollutions environnementales et dilapidation des ressources naturelles [6], [7].

Les études sur l'utilisation du sable de concassage dans le béton ont montré qu'il est possible d'utiliser le sable de concassage contenant jusqu'à 15% de fines pour la fabrication du béton sans affecter ses performances mécaniques[8], [9].

Ce travail s'inscrit dans le cadre de la valorisation des produits locaux tel que les déchets de la carrière de concassage, et les déchets issues de la fabrication des agglomérées de liège expansé en deux fraction 3/8 et 8/15, en les utilisant dans la confection des matériaux cimentaires de construction destinés en premier lieu à l'isolation thermique, et assurer des caractéristiques physiques et mécaniques convenables.

2 MATÉRIAUX

2.1 SABLE

Le sable calcaire est de granulométrie 0/3 provenant de la carrière de Laghouat (Algérie). Sa répartition granulaire, établie selon la norme NF P18-560 [10], et ses différentes propriétés physiques sont regroupées dans le tableau 1.

Table 1. Propriétés physiques du sable calcaire

Propriété	MF	ρ_{app} (kg/m^3)	ρ_{abs} (kg/m^3)	Porosité *n* (%)	Absorption (%)	Es visuel (%)	Es piston (%)
Valeur	2.66	1530	2670	0.43	7.78	75	70.5

Le sable utilisé est un sable propre, de bonne qualité pour le béton.

2.2 CIMENT

Le ciment utilisé est un ciment CEM II/A 42.5 R répondant à une nouvelle normalisation algérienne NA 442 [11].

2.3 GRANULATS

Deux classes granulaires légères de liège expansé sont employées, granulats 3/8 et 8/15.

Les propriétés physiques des granulats sont regroupées dans le tableau 2.

Table 2. Caractéristiques physiques des granulats de liège expansé

Classe granulaire	ρ_{app} (kg/m^3)	ρ_{abs} (kg/m^3)	Porosité n(%)	Absorption (g/g)
Granulat 8/15	65	145	55	1.4
Granulat 3/8	71	145	51	2.3

D'après les résultats du tableau 2. On remarque que les granulats 3/8 en raison de leurs compacité importants sont moins poreux, plus lourds et possèdent un degré d'absorption d'eau plus élevé par rapport aux granulats 8/15.

2.4 EAU DE GÂCHAGE

L'eau utilisée est une eau potable du robinet. Elle remplit les prescriptions de EN 1008 [12].

2.5 ANALYSE GRANULOMÉTRIQUE

Les courbes granulométriques des constituants du béton sont présentés dans la figure 1.

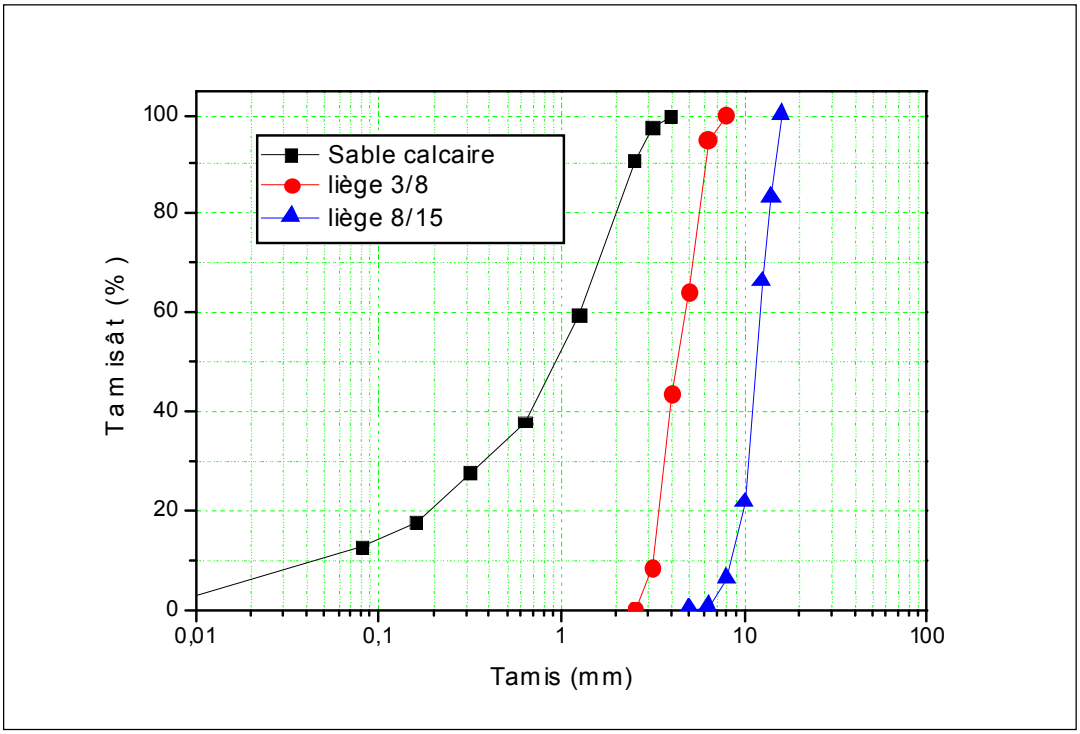


Fig. 1. Courbes granulométriques du sable calcaire et des granulats de liège

La courbe granulométrique du sable calcaire est étalée, elle est bien graduée et sa continuité bien répartie. Le pourcentage de fines du sable utilisé est 12.18 %. Les courbes granulométriques des granulats de liège expansé 3/8 et 8/15 présentent une granulométrie continue et serrée.

3 FORMULATION ET ELABORATION DES BÉTONS CALCAIRES LÉGERS

3.1 FORMULATION DU BÉTON TÉMOIN

D'après les résultats des essais préliminaires, nous avons choisi des rapports C/S=1/3, et E/C = 0.60, pour la formulation des bétons calcaires légers.

3.2 FORMULATION DES BÉTONS ELABORES

Trois pourcentages en volume du liège expansé ont été utilisés 2, 1.5 et 1 en fonction du volume du sable calcaire (Liège/Sable calcaire) c'est-à-dire $V_L/V_S = 2$; $V_L/V_S = 1.5$ et $V_L/V_S = 1$.

Les moules utilisés pour la confection des éprouvettes sont (7x7x28) cm^3. La mise en moule est faite en trois couches damées à 25 coups de chacune, sans vibration pour éviter le problème de ségrégation affectant l'homogénéisation des échantillons [13].

Les éprouvettes ainsi élaborées sont conservées dans deux environnements :

- Les conditions ambiantes du laboratoire (Air libre).
- Une cure par immersion dans l'eau.

4 RÉSULTATS ET DISCUSSIONS

4.1 MASSE VOLUMIQUE APPARENTE

La figure 2 présente la variation de la masse volumique apparente à l'état durci à 28 jours en fonction du dosage en liège expansé.

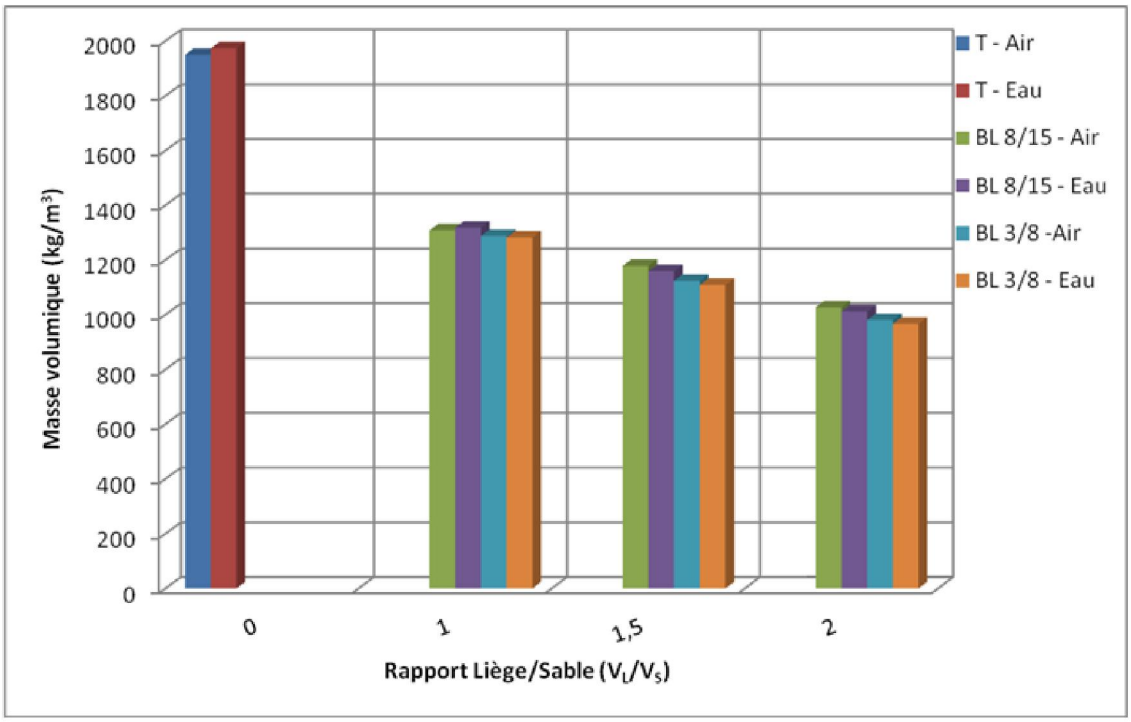


Fig. 2. Evolution de la masse volumique en fonction de dosage en liège

D'après ces résultats, on constate que :

- La masse volumique apparente des compositions diminue lorsque le pourcentage des granulats de liège expansé augmente. On trouve que :

  $1308 > \rho_{app}$ (béton léger de granulats 8/15 (BL 8/15)) > 1027 kg/m^3.

  $1287 > \rho_{app}$ (béton léger de granulats 3/8 (BL3/8)) > 980 kg/m^3.

- Les bétons légers à base de granulats de liège expansé 3/8 ont des masses volumiques moins que ceux de granulats 8/15, dans les deux environnements, Ceci est lié directement par les caractéristiques physiques des granulats de liège, car la porosité des granulats 8/15 est de l'ordre de 55 %. Cette valeur est supérieure à celle des granulats 3/8 qui est de l'ordre 51%. Cela permet d'augmenter le volume de la matrice du ciment qui susceptible d'augmenter la masse du béton.

4.2 RÉSISTANCE À LA FLEXION

La résistance à la flexion à trois points a été mesurée sur les éprouvettes prismatiques (7x7x28) cm^3 pour chaque composition [14].

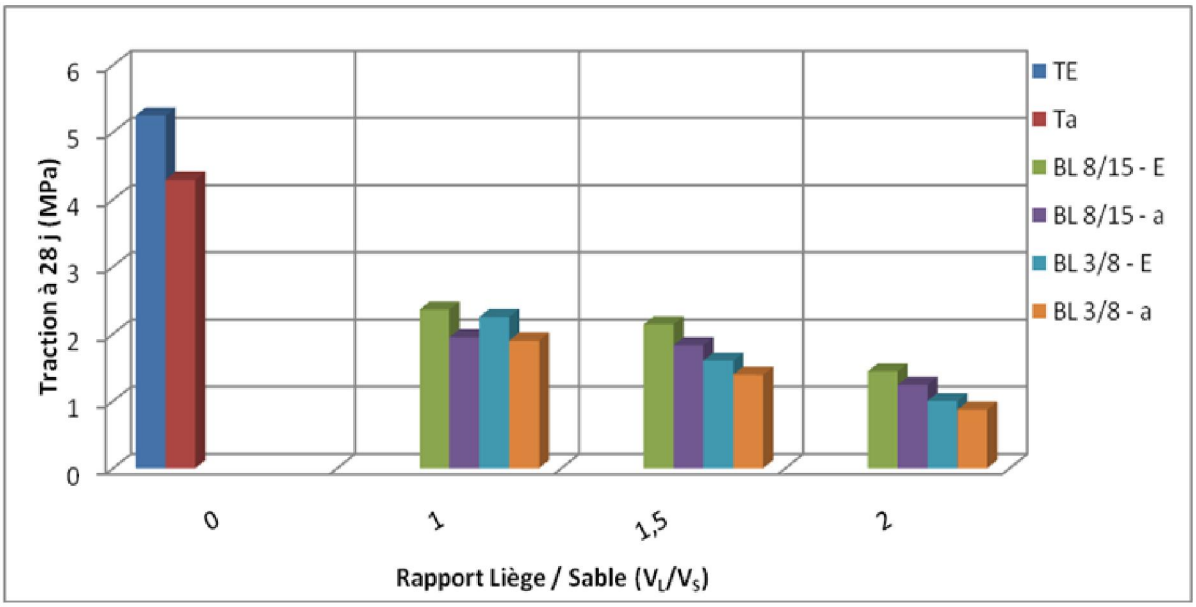


Fig. 3. Evolution de la résistance à la traction à 28 jours en fonction de rapport V_L/V_S

On remarque, avec l'ajout de liège, une chute très rapide de résistance en traction du béton témoin (T) présente 57 % pour le BL 3/8-1 et 55 % pour BL 8/15-1, dans l'environnement de l'eau.

Pour le même rapport (Liége/sable) les valeurs de la résistance en traction des BL 3/8 sont inférieures à celles des BL 8/15.

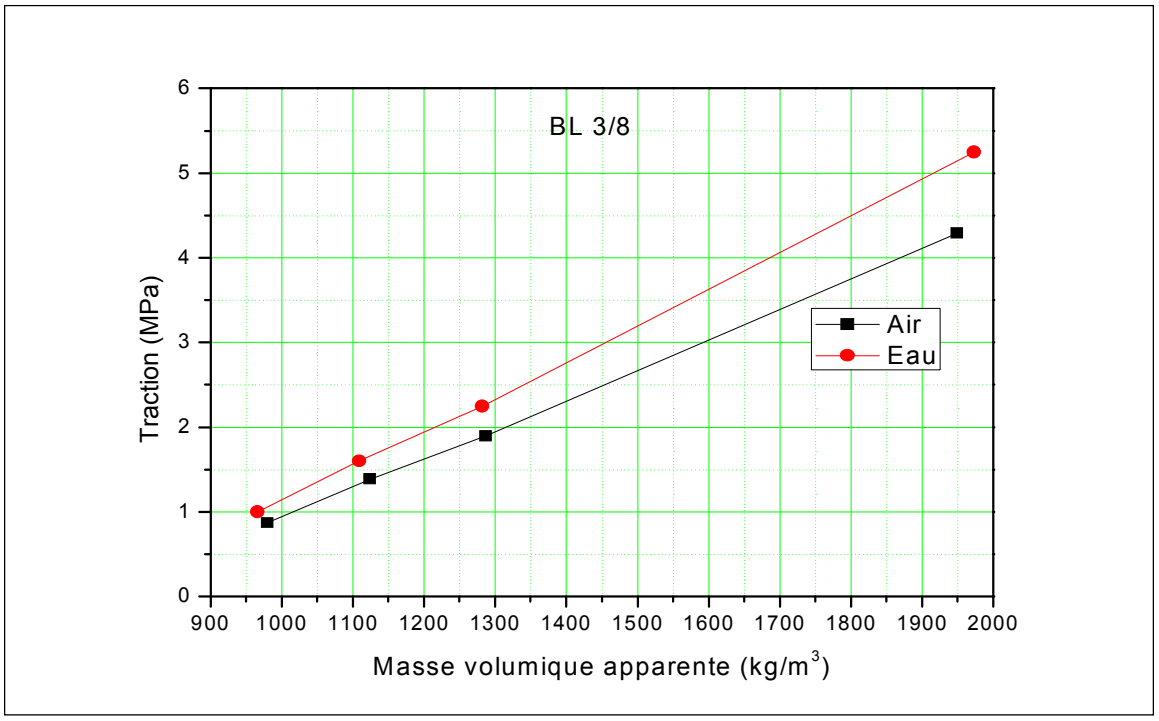


Fig. 4. Variation de la résistance en traction des BL 3/8 et Témoin en fonction de la masse volumique apparente.

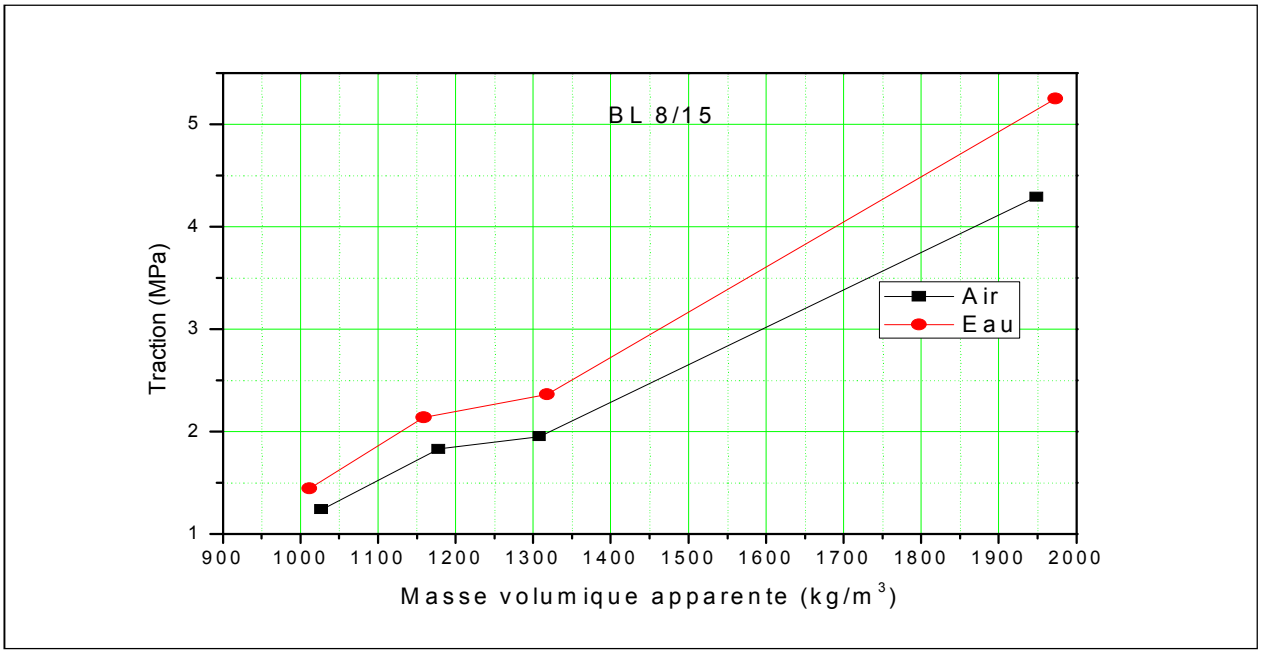


Fig. 5. Variation de la résistance en traction des BL 8/15 et Témoin en fonction de la masse volumique apparente.

On constate que la résistance à la traction diminue lorsque la masse volumique du béton diminue, ceci est tout à fait logique car la résistance à la traction est, comme pour les bétons classiques, une fonction croissante de la masse volumique.

4.3 RÉSISTANCE A LA COMPRESSION

Les figures (6,7 et 8) présentent les résultats des essais de compression du matériau composite effectué sur les demi-éprouvettes prismatiques (7x7x28) cm^3 [15], issues de l'essai de flexion.

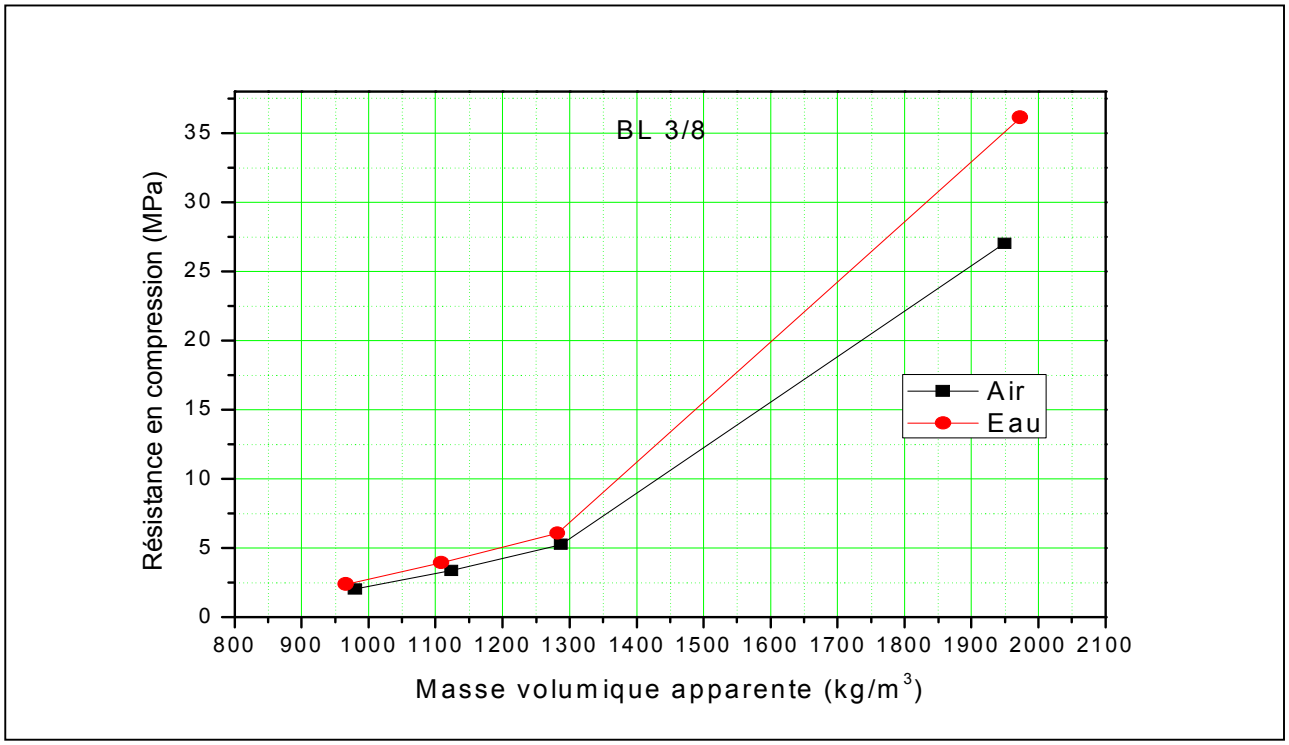


Fig. 6. Variation de la résistance en compression des BL 3/8 et Témoin en fonction de la masse volumique apparente

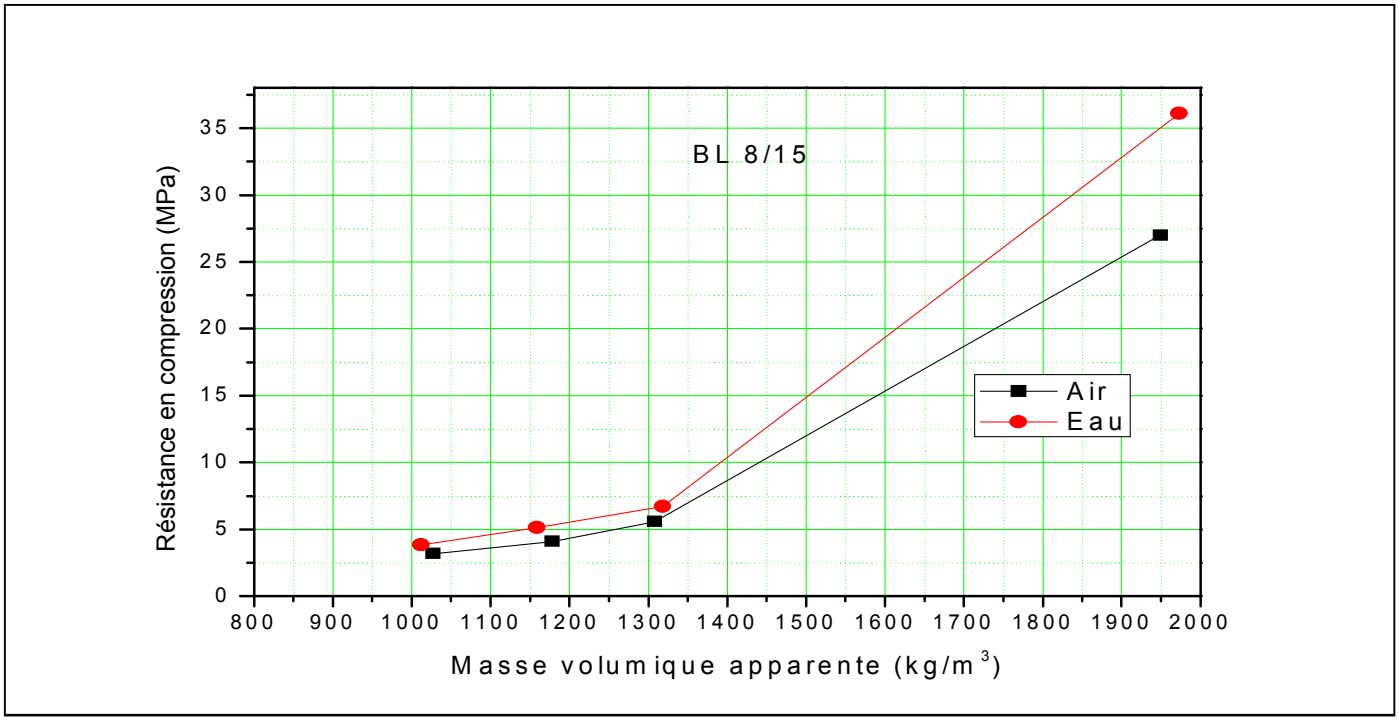


Fig. 7. Variation de la résistance en compression des BL 8/15 et Témoin en fonction de la masse volumique apparente

Les figures 6 et 7 montrent que la résistance à la compression des BL 3/8 et des BL 8/15 diminue lorsque la masse volumique diminue. Une chute très rapide est constatée en passant du béton témoin au béton le moins allégé. Cela est du à l'augmentation des vides et par conséquent la porosité du matériau augmente. Le taux de décroissance est de l'ordre de 79 % et 81 % pour BL 8/15-1 et BL 3/8-1 respectivement dans l'environnement de l'air libre et de 81 % et 83 % dans l'environnement de l'eau. Cette décroissance est de plus en plus faible en augmentant le dosage en granulats. Une amélioration de la résistance est obtenue pour le mode de conservation dans l'eau et ceci pour les deux types de bétons.

La figure 8 résume les résultats obtenus de la résistance à la compression à 28 jours en fonction de rapport liège /sable. On remarque d'après cette figure que l'ajout de liège allant du dosage 1 à 1.5 diminue la résistance des bétons conservés dans l'eau de l'ordre de 24 % pour BL 8/15 et 35 % pour BL 3/8. La décroissance pour BL 8/15-2 est 43 % et de 61 % pour BL 3/8-2.

Donc les granulats de liège expansé 3/8 rendent le béton de sable calcaire moins résistant que les granulats de liège 8/15.

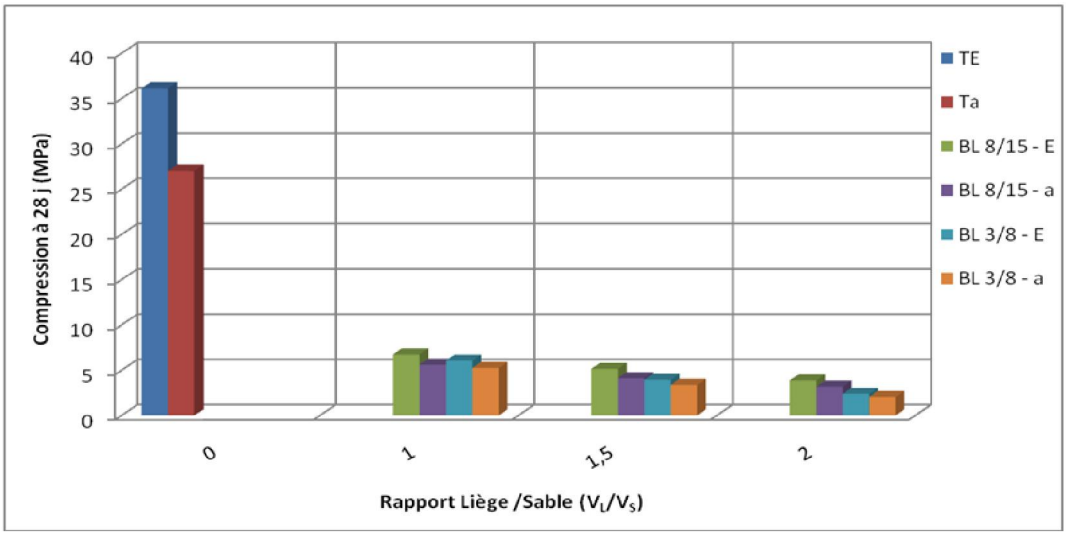


Fig. 8. Evolution de la résistance à la compression à 28 jours en fonction de rapport Liège/sable

4.4 CONDUCTIVITÉ THERMIQUE

L'essai thermique a été réalisé sur des éprouvettes (4x4x10) cm^3 qui sont les résultats de sciage des éprouvettes (10x10x10) cm^3 pour les différents bétons étudiés dans l'état sec et l'état saturé [16], [17].

Les résultats de la mesure de la conductivité thermique (λ) des bétons calcaires légers BL 3/8 et BL 8/15 sont présentés dans les figures (fig. 9 et fig. 10).

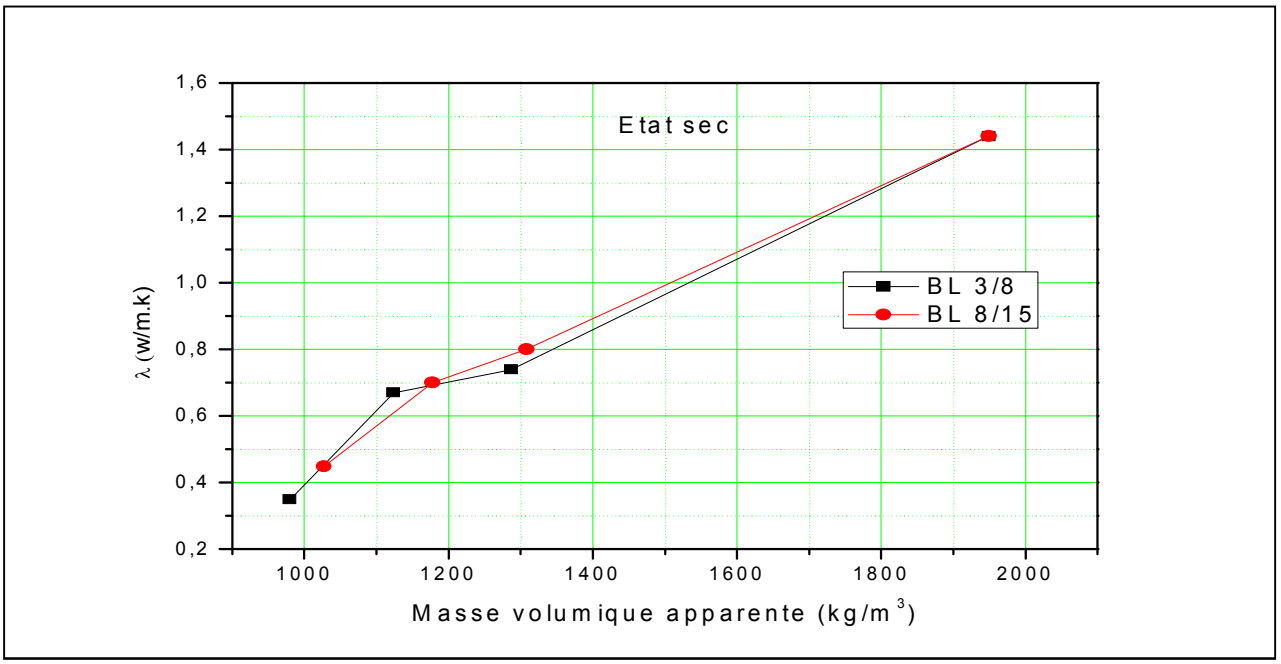


Fig. 9. Variation de la conductivité thermique en fonction de la masse volumique apparente à l'état sec

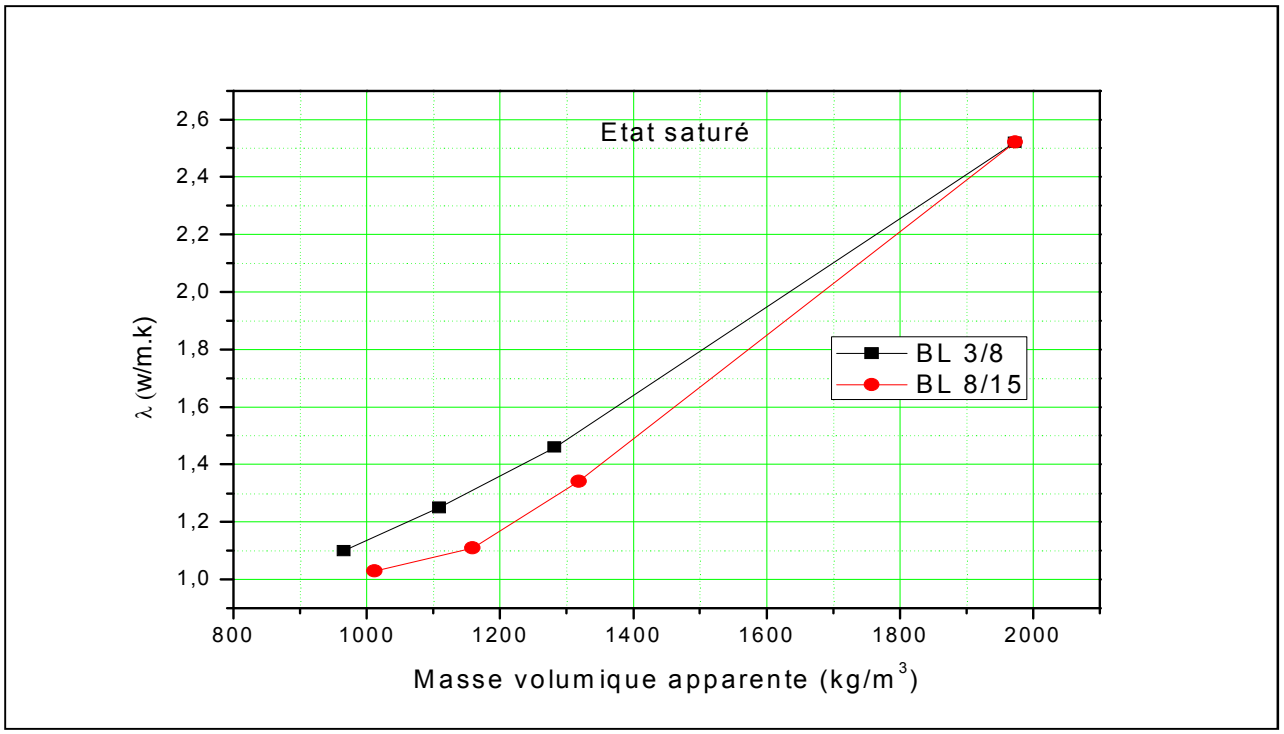


Fig. 10. Variation de la conductivité thermique en fonction de la masse volumique apparente à l'état saturé

D'après les résultats obtenus on remarque que :

- La conductivité thermique du béton élaboré est proportionnelle avec la masse volumique apparente.
- L'humidité a un effet important sur le transfert de la chaleur dans les matériaux étudiés, elle augmente la conductivité thermique.
- Les bétons BL 3/8 ont des conductivités thermiques faible que les bétons BL 8/15 et cela dans l'état sec c'est-à-dire sont les plus isolants. Au contraire, dans l'état saturé, les bétons BL 8/15 sont les plus isolants.
- La conductivité thermique la plus faible est de 0,35 $W.m^{-1}.K^{-1}$ pour le dosage maximum en liège, a été obtenu avec le BL 3/8. Pour le BL 8/15, elle est de 0,45 $W.m^{-1}.K^{-1}$. Celle d'un béton témoin est d'environ 1.44 W/m.k, donc l'allègement du liège permet d'améliorer de 69 % et 75 % le pouvoir isolant par rapport un béton normal.

5 CONCLUSION

La production du béton léger à base de déchets de la carrière des granulats concassés et les déchets issus de la fabrication des panneaux de liège expansé joue un rôle très important dans la protection de l'environnement, et permet aussi de réaliser des gains d'énergie considérables à cause de faible conductivité thermique de liège expansé.

L'objectif de ce travail consiste à étudier le comportement mécanique et physique des bétons légers confectionnés à partir de sable de concassage et allégés avec les deux fractions de liège expansé granulats 3/8 et 8/15.

L'examen des différents résultats obtenus lors de la réalisation de ce travail a permis de tirer les conclusions suivantes :

- La masse volumique apparente des compositions diminue avec l'augmentation de la quantité des granulats de liège expansé.
- Les gammes de masse volumique dans lesquelles varient ces bétons de liège sont : 1027 à 1308kg/m^3 pour le BL 8/15 et de 980 à 1287 kg/m^3 pour le BL 3/8.
- Par sa compacité importante le béton allégé avec les granulats de liège expansé 3/8 est le plus léger.
- L'augmentation du pourcentage de liège est inversement proportionnelle aux résistances mécaniques (compression et traction).
- Le béton léger à base de granulats de liège expansé 8/15 est plus résistant que ce à base de granulats 3/8.
- L'allègement du béton calcaire par l'ajout de liège expansé diminue considérablement la conductivité thermique c'est-à-dire améliore l'isolation thermique
- Pour le béton BL 8/15 et BL 3/8, la conductivité thermique pour le dosage maximum en granulats est 0.45 et 0.35 W/m.k respectivement qui représente le 1/3 et 1/4 de celle du béton témoin. Donc l'allègement du liège permet de diminuer de presque 69 % et 75 % la conductivité thermique par rapport un béton normal.
- La conductivité thermique du béton élaboré est proportionnelle avec la masse volumique apparente.

REFERENCES

[1] M. A. AZIZ, C. K. MURPHY and S. D. RAMASWAMY, "Lightweight concrete using cork granules," *International Journal of Lightweight Concrete*, 1(1), pp. 29-33, 1979.

[2] F. HERNANDEZ-OLIVARES, M. R. BOLLATI, M. DEL RIO, and B. PARGA-LANDA, "Development of cork-gypsum composites for building applications," *Construction & Building Materials*, volume 13, issue 4, pp. 179-186, 1999.

[3] S. R. Karade, F. Caldiera, M. Irle, K. Maher, "Cork granules as lightweight aggregate," *Sustainable concrete construction: proceedings of the international conference*, University of Dundee, Scotland, UK, pp. 253-262, September 2002.

[4] A. EL BAKKOURI, H. EZBAKHE, T. AJZOUL, A. EL BOUARDI, "Etude thermomécanique du béton allégé avec du liège et du béton allégé avec des grignons d'olive," *12èmes Journées Internationales de Thermique*, pp. 307-310, Tanger, Maroc, 15-17 Novembre 2005.

[5] F. G. Branco, A. Tadeu, M. B. C. Reis, "Can cork be used as a concrete aggregate?," *International Journal for Housing Sciences*, Vol. 31, N. 1, pp. 1-11, 2007.

[6] B. Menadi, S.Kenai, J. M. Khatib, A. Ait-Mokhtar, "Strength development and prediction of mortars containing limestone fines," *1st International Conference on Sustainable Built Environement Infrastructures in Developing* Countries, ENSET, Oran (Algeria), October 12-14, 2009.

[7] S. Kenai, A. Attar, B. Menadi, "Cement Industry and Concrete Technology in Algeria: Current Practices and Future Challenges," *International Conference on Durability of Building Materials and Components*, ISTANBUL – TURKEY, May 11-14th, 2008.

[8] B. Li, J. Wang, M. Zhou, "Effect of limestone fines content in manufactured sand on durability of low- and high-strength concretes," *Construction and Building Materials*, Volume 23, Issue 8, pp. 2846–2850, August 2009.
[9] B. Menadi, S. Kenai, J. M. Khatib, A. Ait-Mokhtar, "Strength and durability of concrete incorporating crushed limestone sand," *Construction and Building Materials*, Volume 23, Issue 2, pp. 625–633, February 2009.
[10] Norme française NF P 18-560 : Granulats - Analyse granulométrique par tamisage, AFNOR Paris (1990).
[11] NA 442 : Norme Algérienne, CIMENT, Composition, spécifications et critères de conformité des ciments courants (2000).
[12] EN 1008 : Norme européenne de l'eau de gâchage pour bétons Spécifications d'échantillonnage, d'essais et d'évaluation de l'aptitude à l'emploi (juillet 2003).
[13] Norme française NF P.18-400 : Moules pour éprouvettes cylindriques et prismatiques, Décembre 1981.
[14] NF EN 12390-5, Essai pour béton durci Partie 5: Résistance à la flexion sur éprouvettes, Octobre 2001.
[15] NF EN 12390-3, Essai pour béton durci Partie 3: Résistance à la compression des éprouvettes, Février 2003.
[16] Yves JANNOT, "Métrologie thermique, " *Séminaire PER AUF*, LEA Dakar, pp. 31-34, 12-18/11/08.
[17] Pierre MEUKAM, "Valorisation des briques de terre stabilisées en vue de l'isolation thermique de bâtiments," Thèse de Doctorat, Université de Yaoundé I, pp. 32-33-34, 2004.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 749-757

http://www.issr-journals.org/ijias/

The Survey of the Conservative Relationship between the Intangible Assets and Management Performance Ratio

Roya Darabi[1] and Vahid Vojohi[2]

[1]Accounting Department,
South Tehran Branch, Islamic Azad University,
Tehran, Tehran Province, Iran

[2]Department of Management,
Arak Branch, Islamic Azad University,
Arak, Arak Province, Iran

ABSTRACT: In this study, we investigate the relationship between the management performance and conservatism according to two conservatism scales, namely the profit time- asymmetry scale and the market to book value ratio. The objective of the current study is to survey the conservative relationship between intangible assets and management performance ratios. To achieve this goal, two hypotheses are being posed. To test the study hypothesis, the data from 108 companies, accepted in the Tehran's stock exchange market between the years from 2005 to 2011, was used by taking advantage of targeted systematic sampling method. The company's management performance is related to two factors, intangible assets and conservatism and in fact intangible assets and conservatism are regarded as two independent variables in the present study which have an influence on the management performance. In the present study, the Basu model has been used to measure the conservativeness from the profit and loss perspective and the book to market value ratio has been taken advantage of as well, which is a balance sheet model. The current study methodology is functional from the objective point of view and it is correlation-descriptive from the type perspective. The results obtained are indicative of the direct and significant relationship between the intangible assets and the conservativeness in the intangible assets with management performance.

KEYWORDS: Conservatism, Company Performance, Dupont Ratio, Intangible Assets, Operating Income, Efficiency Analysis.

1 INTRODUCTION

One of the objectives of the financial statements analysis is the company valuation calculation. Many researchers including Feltham and Ohelson [1] believe that the company value is a function of the company expected future growth and its profitability. To analyze the financial statement, analysts usually use the current company growth and profitability as a starting point to forecast the future growth and profitability. Articles and books related to the financial statement analysis deal with the presentation of a simple method to analyze the current state of the company. This simple method and the full capture of the company current state lead to the enhancement of the future profitability forecasting power. One common tool to meet this goal is the use of the financial statements, the basis of which is the use of the figures and the current knowledge extracted from financial statements in order to assess the company's current state. Books and the academic and professional sources and references introduce various, though simple, financial statements to analyze the current profitability, one of the most common of which is assets return. Conservatism is one of the fundamental concepts in accounting which is always paid attention to by compilers in the offering of financial information. In accounting literature, two important features of conservatism have been surveyed.

Corresponding Author: Roya Darabi (royadarabi110@yahoo.com)

First the existence of biases in the less than real representation of stock book value to market value ratio which has been proposed by Feltham and Ohelson [1]. Second, tendency to accelerate the identification of loss and to postpone the profit identification which both have been proposed by Basu. Basu defines conservatism as 'the accountants' tendencies to need a higher degree of verifiability in order to identify good news in profit relative to bad news' (e.g. [2]).

Considering the strategic objectives of a company for choosing a collection of appropriate methods for performance evaluation is a necessity for certain companies. The most important aspect which is in the center of investors' attention is that has their investing value and the management performance results increased?

In other words, is there any value created for them? Some of the economical entities, in the current decades, have used various techniques based on the various approaches to assess the performance but a few numbers of them are satisfied with their organizations processes effectiveness. In practice, there are various approaches to evaluate the performance and in these scales, various indexes and proportions are being used in the calculation of which, the accounting and economical information and a combination of them is used. Accounting conservatism is a disputable issue. Because the costs such as advertising costs (risks resulting from ignoring the future interests stemming from these kinds of costs leads to the accounting information distortion and the study of such investing measures is very difficult due to the intangible values and investments. We concentrate on how the management performance, conservative accounting of intangible assets influence the decisions to allocate managerial resources and the analysis of the conservatism consequences in management performance are dealt with in the current study.

According to competition among public stocks companies and also the relative stagnation occurred, most of the potential investors, managers and other users of financial information, always are searching for the criteria and logical relations and Other indicators of corporations performance. Despite failures in profit measuring, it is possible to be a different between actual earnings and reported profit. If it is assumed that the aim of corporation managers is increasing the wealth of shareholders, therefore their different financial decisions should be taken in accord to fulfillment of this goal Conservation, as one limited accounting principle, are used by accouters for many years and despite of high critiques on it, maintained its status between other accounting principles, such a way that survival of conservatism in contrary with critiques on it in many years, can be an evidence to base principals of this principle. The conservativeness is being measured based on asymmetric timeliness of earning which is criterion of profit and loss, and also it is based on the value of market to the value of share book which is criterion of balance sheet.

2 REVIEW OF LITERATURE AND RESEARCH BACKGROUND

Yoshie Saito [3] in a study with the title of 'management efficiency and intangible conservatism' dealt with the relationship between management efficiency and intangible conservatism in a company. He used a new method in Dupont ratio measurement to measure management efficiency. The results of the study are indicative of the direct relationship between management efficiency and intangible conservatism.

Amir and colleagues [4] found out that operating profit ratio plays a greater role in market response. In the next stage, operating profit ratio was divided into two components which include unspecific profit ratios and other profits ratios. The results showed that non-interactive stability of the gross profit ratio is more than other profits, but no difference was found between interactive stability of these two components. Also, operating assets net flow was divided into two parts: capital in turn flow and fixed assets flow. And they found that interactive and non-interactive stability of the fixed assets results from interactive and non-interactive stability of more flowing of the capital in turn. The other results from this study show that the high (low) operating profit ratio has a positive (negative) relationship with the future surplus return, and this relationship has nothing to do with the flow level of the operating assets net flow. But, in case that the operating profit is lower, the increase in the operating assets net flow does not lead to the surplus return increase and this is indicative of the superiority of the operating profit for surveying the market response.

Soliman [5] in his study, first, dealt with the survey of the efficiency and the reliability of the previous studies and showed that similar to previous researches, the segregation of the net return changes of the operating assets into its element changes plays a role in the analysis and the reaction of the market activists. From among Dupont components, the changes in operating assets net flow lends itself better to future profitability changes forecast. The new aspect of the study was the survey of the reaction and the amount of the use of this ratio and its components to correct (change) short-term and long-term decisions. He showed that in both of the states (short- and long-term) the users pay more attention to the separation of the operating assets net return, and from among them, changes of the operating assets turn play a greater role in the revision (change) of the market activists decision-making. Amir and colleagues [4] dealt with the survey of the financial data users reactions at the time of seasonal announcement of the assets return ratio (not operating assets net return) and its

components. They paid attention to the survey of the components level of the assets return ratio. In the current study it is proved that the market activists react immediately after the announcement of the abnormal assets return, abnormal assets turn and abnormal operating profits ratio (more than expected). Also, abnormal assets return and its components can't afford to account for the abnormal stock return. In the following part, the results lead to the abnormal positive return of the operating profit ratio. Their study shows that the high (low) level and the extant relationship between assets turn level have no relationship. On the other hand, when the operating profit ratio and assets turn are in a low level, increase in the assets turn doesn't lead to the abnormal increase in the stock return.

Wang [6] studied the changes in the performance of the companies accepted in the China's bourse on the preliminary shares offerings and dealt with the effect of the proprietorship on the performance changes at the time of the first shares offerings. The study population was the 747 Chinese companies after the public shares offerings in the time period from 1994 to 1999. Also, the performance was measured by means of assets return, operating income to assets (OI/A) ratio and sales to assets (S/A) ratio. The results of the study of the proprietorship role by different kinds of shareholders and the ownership concentration on the company's performance after the first shares offers, with controlling variables such as company size, financial level and economical activities showed that government proprietorship and ownership concentration have nothing to do with the performance changes but the statutory ownership, non-government ownership have a non-linear relationship with performance changes. That means that the low and high levels of statutory proprietorship (non-governmental ownership concentration) leads to a positive relationship and middle levels of company proprietorship lead to the separation of the control and proprietorship of the company and interest contradictions. Therefore, after the first shares offerings, we will have a low performance relative to the pre-shares offerings.

Shahira shahid [7] in a study, which was performed on 90 companies ,in Egypt in Kario Alexandrio, showed that there is no significant relationship between the ownership type and P / BV, P / E indices but the relationship between the ownership type and companies accounting performances is significant.

He found out that when the managers and the state sector are the major companies' shareholders they influence ROE negatively and, in contrast, holdings companies and private sector influence the ROE positively. Fearfield and Youhen [8] Soliman [9] dealt with the study of the relationship of Dupont ratio and its elements with other financial and accounting variables. Fear field and Youhen [8], at first, dealt with the survey of the next years predictability potential of the profitability changes through segregating operating assets net return ratio to its elements (operating assets net turn and operating profit ratio) and they found out that segregating Dupont ratio to its elements doesn't increase the potential of the future profitability changes predictability. In the next stage, the changes of the operating assets net return ratio and its elements are studied and they found out that the changes in every elements of assets net return ratio is more useful for forecasting the future profitability changes and from among them, operating assets net turn changes have a better predictability potential in comparison with the operating profit ratio changes.

3 RESEARCH HYPOTHESIS

According to the theoretical literature and the study background the following hypotheses were introduced.

H1: There is a significant relationship between intangible assets and management performance.

H2: There is a significant relationship between conservatism in intangible assets and management performance.

4 RESEARCH METHODOLOGY

In the current study, the correlation analysis was utilized to the hypotheses.

The study population includes all the companies in Tehran's securities market during the year from 2005 to 2011 and the total number of them reaches to 468.

Now, with the following assumptions, some of these companies are discarded, perforce:

- The company's fiscal year should end in December.
- The company should not have changed its fiscal year during 2005-2011.
- The company should have been accepted up to the end of 2004 fiscal year in Tehran's securities market.
- The company shouldn't be one of the financial intermediary companies (investing and financial companies).

Sampling was done based on the elimination method by exerting the above conditions and samples of 108 companies were chosen.

4.1 RESEARCH VARIABLES

4.1.1 DEPENDENT VARIABLES

Management performance: Yoshie Saito calculated the Dupont Ratio in 2012 by making use of a new method and introduced it to the financial science field. It was obtained in a very simple way by dividing assets by net profit before tax. But, in the presented study we will act upon the method that Yushi Saito calculated and used Dupont Ratio.

$$ROA_{it} = \frac{INCOME_{it}}{ASSET_{it-1}} = \frac{INCOME_{it}}{SALE_{it}} \frac{SALE_{it}}{ASSET_{it-1}} \quad (1)$$

Dupont ratio has been comprised of assets turn and the operating profit ratio.

ROA_{it} = Dupont ratio

$INCOME_{it}$= net profit before the taxation of the company i in the year t.

$SALE_{it}$= the company i sales in the year t.

$ASSET_{it-1}$= company i assets in the previous year

But, in the new method, Dupont Ratio can be calculated as follows:

$$ROA_{it} = (1 - \frac{EXP_{it}}{SALE_{it}}) \frac{SALE_{it}}{ASSET_{it-1}} \quad (2)$$

In the above proportion which is a combination of the company's income and expenditures:

EXP_{it}= all of the company's expenditure

We can calculate this relation in the form of relation 3 by developing and promoting this proportion:

$$ROA_{it} = \frac{SALE_{it}}{ASSET_{it-1}} - \frac{EXP_{it}}{ASSET_{it-1}} \quad (3)$$

In the current study, Dupont Ratio will be calculated in the form of the relations (2) and (3). And then in the next stage, the Dupont Ratio, which is the management performance, will be measured as relation (4):

$$PERM_{it} = \frac{ROA_{it>0}}{ROA_{it=0}} \quad (4)$$

When the obtained figure is equal to one, the company is in its best management performance and when this figure is equal to zero, the company is in its worst management performance and when this figure is between zero and one the company is in its favorable state.

4.1.2 INDEPENDENT VARIABLES

Conservatism: one of the greatest factors which can affect the management performance proportions is the accounting conservatism and profit management as well, and in the end by making use of the model one we will deal with the conservatism in the intangible assets:

$$BTM_{it} = \alpha_i + \alpha_t + \sum_{j=0}^{j=3} \beta_j R_{t-j.i} + e_{it} \quad \text{Model (1)}$$

BTM is the book to market value of the company's assets.

R is the past three years return lag.

Intangible assets: the amount of the intangible assets for company i in the year t from the company balance sheet which can be measured in the asset part.

4.1.3 CONTROL VARIABLE

The company's total sales: company i sales in the year t which is calculated from the profit and loss account.

Total sales to total expenditures: total sales to total expenditures of company i in the year t which can be calculated from the profit and loss account (in case of having profit, we use the sales and in case of suffering a loss we use the company's total expenditure).

Cost of goods sold: the cost of goods sold for company i in the year t can be calculated from the profit and loss account.

Salary and reward paid to the board of directors: because this amount is not disclosed in Iran and it cannot be calculated, the alternative variable was used, which the maximum amount of money is paid to the board of directors in the business code.

The amount of assets applied in the company: the amount of assets applied in the company i in the year t can be calculated from the right side of the balance sheet. By this amount we mean all of the assets including current assets and non-current assets and intangible assets.

To test the study assets, we use the regression models (2) and (3):

$$ROA_{it} = b_0 + b_1 SALE_{it} + b_2 SAEXP_{it} + b_3 COGS_{it} + b_4 INTG_{it} + b_5 CCOMP_{it} + b_6 EQCOMP_{it} + v_{it} - u_{it} \quad (2)$$

$$ROA_{it} = b_0 + b_1 SALE_{it} + b_2 SAEXP_{it} + b_3 COGS_{it} + b_4 BTM_{it} + b_5 CCOMP_{it} + b_6 EQCOMP_{it} + v_{it} - u_{it} \quad (3)$$

ROA= management performance (dependent variable)

BTM= conservatism (independent variable)

LINTG= intangible assets (independent variable)

LSALE= the companies' total sale

SAEXP= total sales to total expenditures (control variable)

COGS= cost of goods sold (control variable)

CCOMP= the salary and rewards paid to the board of directors (control variable)

LEQCOMP= the amount of the assets applied in the company (control variable)

4.2 Data Collection Method

The data required for the study has been collected via referring to the audited financial statements of the companies accepted in Tehran's securities market (existing in Tehran's securities bourse organization library) and Rahavard Novin company software, as well. The tools used to collect the data include statistical tests observations, statistical tests, information bank, SPSS software and Excel software. The information related to the theoretical and subjective principles has also been collected by taking advantage of the libraries and the books, Persian and English articles.

5 Research Findings

5.1 The Hypotheses Test and Analysis

In the present study to analyze the data and test the hypothesis, the multivariate linear regression has been used and to study the overall model significance F-value and to survey the significance of the independent variables coefficient in every model t-value has been used and the decision to reject or accept the hypotheses has been made over 95% of the confidence level. Also, to survey and determine the experimental data homogeneity with statistical distribution and to survey the errors independence from each other the Kolmogrov-Smearnov test and Durbin-Watson t-value were used, respectively.

5.2 THE RESULTS OF THE HYPOTHESES TEST

To test the study hypotheses, first, we should calculate the conservatism in intangible assets. The calculations related to the conservatism in intangible assets are presented below:

5.2.1 THE CONSERVATISM REGRESSION CALCULATIONS IN INTANGIBLE ASSETS

To test the conservatism in intangible assets, first, we should calculate the conservatism in intangible assets by estimating the coefficients and then we incorporate the conservatism in intangible assets in the regression model in order to be able to evaluate the model and statistically test it. To do so, we use model (1), and the results of the coefficients estimation to calculate the conservatism in intangible assets are as follows:

Table 1 shows the results obtained from the coefficients estimation.

Table 1. Correlation coefficient, determination coefficient and Durbin-Watson test between dependent and independent variables

Durbin-Watson value	The estimation standard error	The offset determination coefficient	determination coefficient	Correlation coefficient	model
1.642	0.68894	0.301	0.315	.523[a]	1

In table 1, correlation coefficient, determination coefficient, and the offsetting determination coefficient and the conservatism estimation model in intangible assets are presented.

Durbin-Watson value is equal to 1.642, which is in the return range of 1.5-2.5. The non-autocorrelation assumption between errors is not rejected and, therefore, the conservatism model regression in intangible assets can be used.

Table 2 includes regression variance analysis to survey the concrete existence of the linear relationship between the dependent and independent variables in the conservative model in intangible assets.

Table 2. Variance analysis

model		Total squares	Degree of freedom	Squares mean	F-value	Significance level
1	Regression	5.437	1	5.437	11.455	.001[a]
	Residuals	355.508	749	.475		
	total	360.946	750			

The results obtained from the conservative model in the intangible assets, which is presented in table 2, is indicative of the significance of the conservative model in intangible assets, and the Fischer distribution statistics and the obtained significance level confirms this matter.

Table 3. Regression model coefficients of the dependent and independent variables

model	abbreviations	No standardized coefficients		Standardized coefficients	t-value	Significance Level	Co linearity statistics	
		B	The column coefficient standard error B	Beta			Tolerance	Variance inflation factor
1	Constant	.675	.027		25.206	.000		
	LAGRjt2	.089	.026	.123	3.385	.001	1.000	1.000

According to table 3, t-value is in the confidence level of 95% for independent variables and p-value amount obtained in the related column is indicative of this claim.

Regression equation is in the following form:

$$BTM_{it} = .675 + .089 * R_{t-j.i} \quad (4)$$

Now, according to the obtained coefficients in the regression equation 4, we can calculate the conservatism in the company's intangible assets. Now, we use the obtained figure to test the model regression related to the study hypotheses.

The results of first hypothesis: First hypothesis: there is a significant relationship between intangible assets and management performance.

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

Table 4. Correlation coefficient, determination coefficient, and Durbin-Watson test between intangible assets and management performance

Durbin-Watson value	The estimation standard error	The offset determination coefficient	determination coefficient	Correlation coefficient	model
1.878	0.15844	0.353	0.369	.412[a]	1

Based on the table 4, Pierson correlation coefficient between the two variables of intangible assets and management performance is 0.412. This table in the 5% error level shows a significant relationship between the two variables of intangible assets and management performance.

According to table 4, the amount of Durbin-Watson value is 1.878 and this figure shows that there is no autocorrelation between errors.

Table 5. Regression variance analysis for the intangible assets and management performance.

model		Total squares	Degree of freedom	Squares mean	F-value	Significance level
1	Regression	3.753	6	0.625	24.916	.000[a]
	Residuals	18.401	733	0.025		
	total	22.154	739			

Table 5 is indicative of the variance analysis between the management performance variable as the dependent variable and intangible assets as the independent variable. Since, significance level is less than 5%, the linearity assumption between the two variables is confirmed.

Table 6. Regression equation coefficients for control and independent variable

Model	abbreviations	Non standardized coefficients		Standardized coefficients	t-value	Significance level	Co linearity statistics	
		B	The column coefficient standard error B	Beta			Tolerance	Variance inflation factor
1	(Constant)	0.143	0.006		22.338	0		
	SALE	1.61E-07	0.0001	0.19	11.928	0	0.003	1.685
	SAEXP	2.57E-05	0.0001	0.014	0.413	0.68	0.987	1.013
	COGS	-1.81E-07	0.0001	-0.808	-11.492	0	0.003	3.754
	INTG	7.434	0.546	0.272	13.615	0	0.355	2.816
	CCOMP	6.65E-07	0.0001	0.038	1.135	0.257	1	1
	EQCOMP	-8.99E-09	0.0001	-0.452	-3.12	0.002	0.054	1.545

In the table 6, the output and in column B, the constant amount and the independent variable coefficient are presented respectively, and this equation takes the following form:

$$ROA_{it} = .143 + 1.61E-07SALE_{it} - 1.81E-07COGS_{it} + 7.434INTG_{it} - 8.99E-09EQCOMP_{it} \quad (5)$$

The second hypothesis results Second hypothesis: there is a significant relationship between conservatism in intangible assets and management performance:

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

Table 7. Correlation coefficient, determination coefficient and Durbin-Watson conservatism in intangible assets and management performance

Durbin-Watson value	The estimation standard error	The offset determination coefficient	determination coefficient	Correlation coefficient	model
1.63	0.13081	0.428	0.433	.658[a]	1

According to table 7, Pierson correlation coefficient between the two conservatism variables in intangible assets and management performance is equal to 0.658. This figure in the error level of 5% shows a significant relationship between the above two. Based on table7, the amount of Durbin-Watson value is 1.63, and this figure shows that the errors are independent from each other.

Table 8. Regression variance analysis for conservatism in the intangible assets and management performance:

Model		Total squares	Degree of freedom	Squares mean	F-value	Significance level
1	Regression	9.686	6	1.614	94.343	.000[a]
	Residuals	12.68	741	0.017		
	total	22.366	747			

Table 8 is suggestive of the variance analysis between management performance and conservatism in intangible assets, because significance level is below 5%, the linearity assumption between the two variables is confirmed.

Table 9. Regression equation coefficient for independent and control variables

Model	Abbreviations	Non-standardized coefficients		Standardized coefficients	t-value	Significance level	Co linearity statistics	
		B	The column coefficient standard error B	Beta			tolerance	Variance inflation factor
1	(Constant)	-0.204	0.019		-10.574	0		
	SALE	1.20E-07	0	5.318	10.658	0	0.003	3.482
	SAEXP	2.15E-05	0	0.012	0.419	0.675	0.988	1.013
	COGS	-1.31E-07	0	-4.924	-10.075	0	0.003	3.232
	CCOMP	3.05E-07	0	0.017	0.631	0.528	0.998	1.002
	EQCOMP	-8.05E-09	0	-0.404	-3.5	0	0.058	1.386
	BTM	0.398	0.021	0.527	18.62	0	0.955	1.048

In table 9 and in column B the constant value and independent variable coefficient are presented in regression model and this equation changes to the following form:

$$ROA_{it} = -.204 + 1.20E-07SALE_{it} - 1.31E-07COGS_{it} + 0.398\,BTM_{it} - 8.05E-09EQCOMP_{it} \quad (6)$$

6 DISCUSSIONS AND CONCLUSIONS

In the first hypothesis test and according to the analyses which have been performed through regression and correlation models, we came to this conclusion that there is a positive correlation coefficient between independent variable (intangible assets) and dependent variable(management performance) in the companies accepted in Iran's capital market and there is a significant relationship between intangible assets and management performance in the companies accepted in Tehran's securities market.

According to the results obtained , there is a significant relationship between intangible assets and management performance in the companies accepted in Tehran's securities market, that means with an increase in intangible assets, the management performance increases and vice versa.

The results of this hypothesis are similar to the results obtained from Yoshie-Saito [3], who expresses that there is a significant relationship between management effectiveness and intangible conservatism and also the results obtained from Soliman study [5] is similar to the other researches results. He dealt with the survey of the companies' efficiency and the operating assets.

In the second hypothesis test, according to the analysis performed by the regression and correlation method we came to this conclusion that there is a positive correlation coefficient between independent variable (conservatism in intangible assets) and dependent variable (management performance) in the companies accepted in Iran's capital market and there is a significant relationship between conservatism in intangible assets and management performance in the companies accepted in Tehran's securities market.

According to the results obtained, there is a direct relationship between conservatism in intangible assets and management performance in the companies accepted in Tehran's securities market, that means with an increase in conservatism in intangible assets, management performance increases and vice versa.

The results obtained from this hypothesis is corresponding to the results of Yoshie-Saito [3] who expresses that there is a significant relationship between management efficiency and intangible conservatism and Soliman's study [5] as well, who dealt with the companies' efficiency and operating assets.

REFERENCES

[1] G. Feltham, J. Ohlson, "Valuation and Clean Surplus Accounting for Operating and Financial Activities," *Contemporary Accounting Research*, pp. 689-731, 1995.

[2] S. Basu, "The conservatism principle and the asymmetric timeliness of earnings," *Journal of Accounting and Economics*, No. 24, pp. 337, 1997.

[3] Yoshie Saito, "Informational and Managerial Inefficiency and Conservatism for Intangible-Intensive Firms," *Department of Accounting, Whitney Bunting School of Business Georgia College & State University*, 2012.

[4] Amir, E., I. Kama and J. Livnat, "Conditional versus unconditional persistence of RNOA components: Implications for valuation," *Review of Accounting Studies,* vol. 16, no. 2, pp. 302-327, 2011.

[5] T. M. Soliman, "The use of dupont analysis by market participants," *The accounting Review*, Vol. 83, No. 3, pp. 823-853, 2008.

[6] Wang, Changyu, "Ownership and operating performance of Chinese IPOs," *Journal of Banking & Finance*, Vol. 29, pp. 1835-1856, 2005**.**

[7] Shahira Abdel Shahid, "Does Ownership Structure Affect Firm Value? Evidence from the Egyptian Stock Market," *Cairo and Alexandria Stock Exchange*, January 2003.

[8] M. P. Fairfield, T. L. Yohn, "Using asset turnover and profit margin to forecast changes in profitability, " *Review of accounting Studies*, Vol. 6, no. 4, pp. 371-385, 2001.

[9] Soliman, M. T., "Using industry-adjusted dupont analysis to predict future profitability," *Stanford Graduate School of Business*, Working Paper, 2004. DOI:10.2139/ssrn.456700

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 758-764

http://www.issr-journals.org/ijias/

Design of CPW-fed Printed Rectangular Monopole Antenna for Wideband Dual-Frequency Applications

Anshul Agarwal, P. K. Singhal, Shailendra Singh Ojha, and Akhilesh Kumar Gupta

Department of Electronics, Madhav Institute of Technology and Science, Gwalior, India

ABSTRACT: A printed rectangular monopole antenna (PRMA) with coplanar waveguide (CPW)-fed is designed for wideband dual frequency application. A proposed monopole antenna which has a size of 31.9 x 33 mm^2, is fabricated to work on a substrate (FR4 lossy) that has the relative permittivity (ε_r) of 4.4 and a thickness of 1.6 mm with CPW-fed in the frequency range of 6 - 16 GHz. Simulation results such as impedance bandwidth are presented and discussed. Simulation results have been verified with good agreement. The parameters which affect the performance of the antenna characteristics are investigated in this paper.

KEYWORDS: Printed rectangular monopole antenna, Wideband antenna, Coplanar waveguide fed, Impedance bandwidth, Return Loss.

1 INTRODUCTION

In the era of modern wireless communication systems, dualband or multiband antennas with omni-directional radiation characteristics play a vital role. Among the printed monopole antennas of various shapes, rectangular monopole antenna is simple in geometry and CPW-fed technique [1] is used for transmission line i.e. transmitting or receiving the electromagnetic (EM) waves.

The design of relatively compact, planar monopole antennas based on the microstrip structures has been reported to meet the bandwidth requirement. Indeed, compared to the classical monopole antennas, the planar monopole antennas are of a significantly wider bandwidth and have similar radiation characteristics. A planar rectangular monopole antenna (PRMA) fed by a coplanar waveguide (CPW) is proposed to promise the aforementioned impedance bandwidth together with satisfactory radiation characteristics through the dual band. The CPW-fed PRMA exploits the CPW configuration to permit easy integration with the monolithic microwave integrated circuits.

A simple structure, dual or multiband operation, and wide bandwidth for use in the modern wireless communication system are the requirement for increasing the information transfer. Several antenna designs such as the planar inverted-F antennas [2], the aperture stacked patch antenna [3], the meander-line chip antennas [4], and the planar monopole antennas [5] have been designed in the last few years to solve this problem. But, such kinds of antennas need a large size of ground plane, which is printed on the different side of the substrate from the radiating plane, and a via-hole connection is always necessary for feeding the signal which increases the manufacture difficulty and cost.

Recently, due to its many attractive features such as wide bandwidth, low cross polarization, uniplanar nature and easy integration with active devices or monolithic microwave integrated circuits, the coplanar waveguide (CPW)-fed antenna has been used a alternative to conventional antennas for wireless communication systems [6-9].

In this paper, a rectangular monopole antenna fed by a CPW transmission line is designed for obtaining wideband dual-frequency operation which is operate in the range of 6.9 – 7.7 GHz and 12.6 - 14.5 GHz frequency bands. The proposed design wide bandwidths cover the C and Ku bands according to IEEE standard frequency spectrum. The Ku frequency band is

Corresponding Author: Anshul Agarwal (agarwal.2712@gmail.com)

wide used for maritime civil and military navigation radars applications like airborne radars for performing the roles of interceptor, fighter, and attack of enemy fighters and of ground targets.

The parameters of the antenna such as return loss, radiation efficiency, directivity and gain are determined using CST (Computer Simulation Technology) Microwave Studio software [10-12]. CST MICROWAVE STUDIO is a fully featured software package for electromagnetic analysis and design in the high frequency range. The software contains transient solver which best fit their particular applications and provide the entire broadband frequency behaviour of the simulated device from only one calculation run. This solver is very efficient for most kinds of high frequency applications such as connectors, transmission lines, filters, antennas and many more.

2 ANTENNA DESIGN

The geometry of proposed wideband dual frequency PRMA with CPW-fed is shown in fig 1. In this design, PRMA with feeding line and ground planes are all printed on the same side of glass epoxy dielectric substrate of thickness (h) 1.6 mm with a relative permittivity (ϵ_r) of 4.4 while the other side of dielectric substrate is without any metallization. A singe rectangular strip line of width Ws and two equal ground planes are used as CPW-fed transmission line. Two equal finite ground planes, each with dimensions of length Lg and width Wg are placed symmetrically on each side of the strip line. The rectangular monopole is connected centrally at the end of the CPW feed line. The space (d) between the rectangular monopole and the edge of the ground plane dominates the resonant mode of the upper band while gap (g) between the strip line and coplanar ground planes vary the impedance bandwidth of upper band. By properly selecting the antenna's geometric parameters, wideband and dual-frequency operation is achieved and even small size of antenna is also obtained.

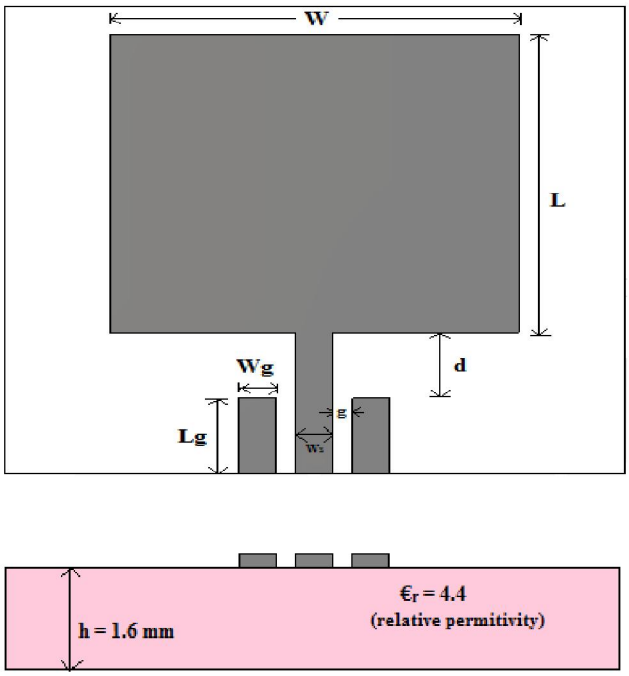


Fig. 1. Geometry of the proposed CPW-fed PRMA for wideband dual frequency operation (All dimensions in mm)

The geometric parameters were adjusted carefully and finally the antenna dimensions were obtained is shown in Table 1.

Table 1. Dimensions of CPW-fed PRMA

Parameter	Dimension	Parameter	Dimension
Lg	8.1 mm	Ws	3 mm
Wg	3 mm	d	7 mm
L	31.9 mm	g	1.6 mm
W	33 mm		

In addition, due to the simulation results show that both the gap distance (g) space (d) have a significant effect on the impedance bandwidth of the proposed antenna.

3 SIMULATED RESULTS

Fig 2 shows the simulated return loss against frequency in the range of 6 – 16 GHz for the proposed wideband dual frequency CPW-fed rectangular monopole antenna. The simulated result is obtained using the CST Microwave Studio software. After analysing the result, it is clearly seen that two resonant modes at about 7.3 and 13.4 GHz are excited with good impedance. The lower resonant band obtain a bandwidth of 0.75 GHz in the range from 6.96 to 7.71 GHz with respect to a resonant frequency of 7.3 GHz and upper resonant band achieves a bandwidth of 1.83 GHz in the range from 12.67 to 14.5 GHz with respect to a resonant frequency of 13.4 GHz.

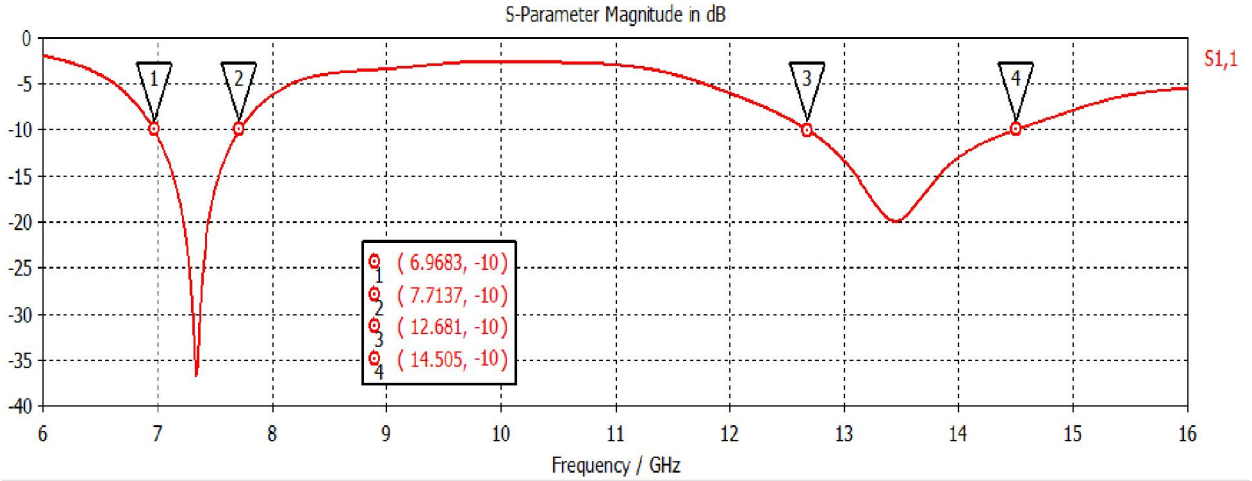
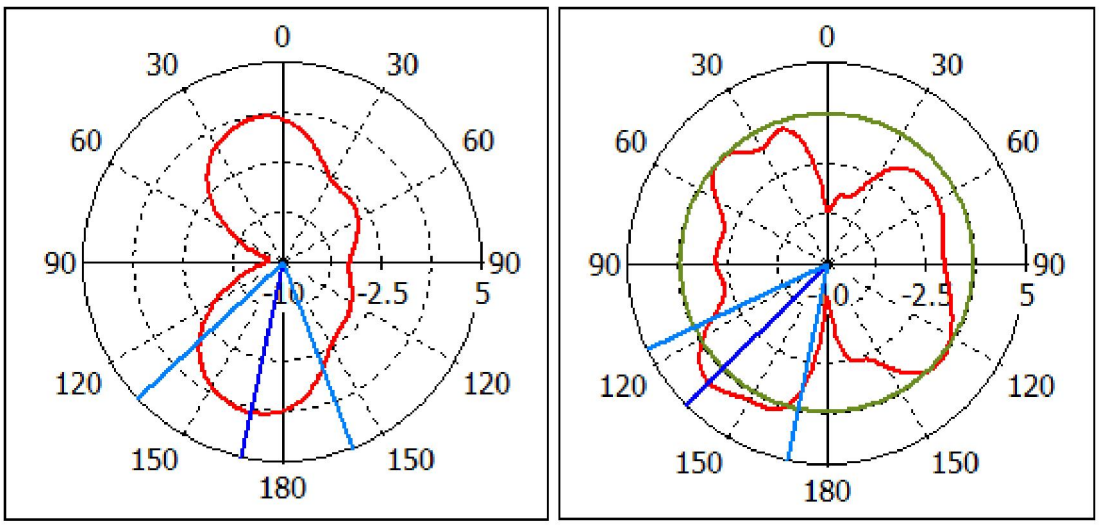


Fig. 2. Simulated return loss of the proposed CPW-fed PRMA.

The simulated radiation patterns of the proposed CPW-fed PRMA at resonant frequencies of 7.3 and 13.4GHz for lower and upper bands are shown in Fig 3. The radiation efficiency, directivity and gain at both lower and upper frequency of 7.3 and 13.4 GHz are 83.33%, 2.908 dBi, 3.693 dB and 84.65%, 4.871 dBi, 4.148 dB respectively.



(a) At 7.3 GHz *(b) At 13.4 GHz.*

Fig. 3. Simulated radiation pattern of the proposed CPW-fed PRMA (a) at 7.3 GHz and (b) at 13.4 GHz

An important feature of the proposed antenna is the capability of impedance matching at both resonant frequencies using a single CPW feed line. For this, the coupling effect between the feed line and the ground planes is investigated. Fig. 4 shows the simulated return loss of the proposed antenna with different gap width (g) and the corresponding data for comparison is shown in Table 2 with all the remaining parameters of the proposed antenna are same as the design shown in fig 1. As seen in Fig. 4, the gap width has a significant effect on the upper band impedance bandwidth of the proposed antenna as compare to lower band bandwidth. The upper band bandwidth is monotonically increases with the decrement in gap width from about 1.35 to 2.23 GHz with respect to simultaneously increment in resonant frequencies.

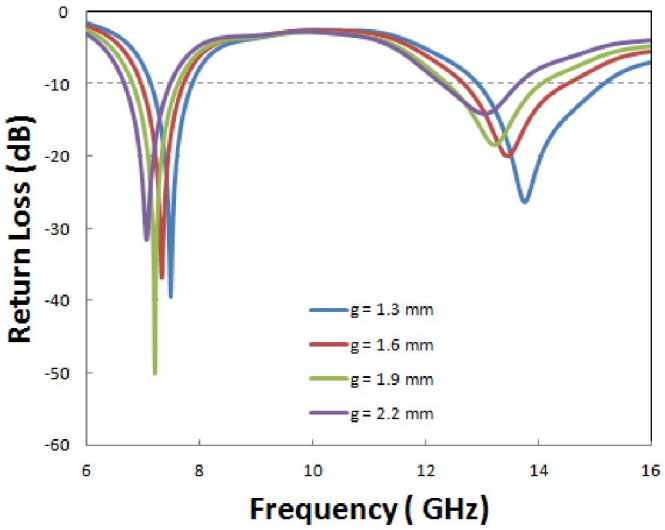


Fig. 4. Simulated return loss of the proposed CPW-fed PRMA with various gap widths.

Table 2. Simulated bandwidths of proposed CPW-fed PRMA as a function of gap width.

Gap width 'g' (mm)	Bandwidth (GHz)	Resonant frequency (GHz)
1.3	0.73, (7.13-7.86) 2.23, (12.92-15.15)	7.5 13.7
1.6	0.75, (6.96-7.71) 1.83, (12.67-14.50)	7.3 13.4
1.9	0.8, (6.81-7.61) 1.66, (12.38-14.04)	7.2 13.2
2.2	0.8, (6.67-7.47) 1.35, (12.29-13.64)	7 13

Also the effect of the space (d) between the rectangular monopole patch and the edge of the ground plane of the proposed CPW-fed PRMA for the impedance bandwidths is studied. Here, the two cases arise for the impedance bandwidth analysis. In the first case, the dimensions of both the rectangular monopole patch and the ground plane are fixed but the space between them is varied and in another case, the space between them is varied according to the change of the ground plane length but dimensions of rectangular monopole patch is fixed. In both the cases, the remaining parameters of the proposed antenna are same as the design shown in fig 1.

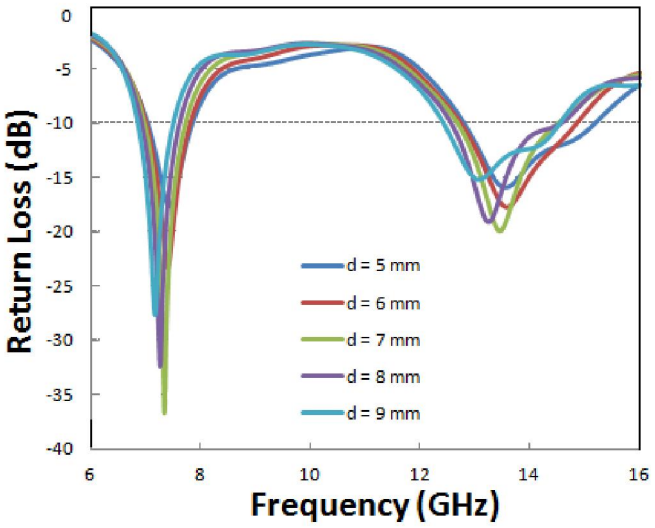


Fig. 5. Simulated return loss of the proposed CPW-fed PRMA as a function of space between monopole antenna and ground.

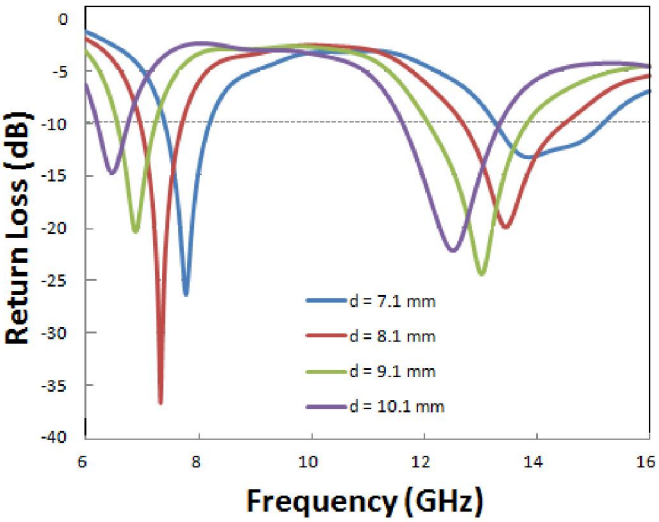


Fig. 6. Simulated return loss of the proposed CPW-fed PRMA as a function of space between monopole antenna and ground plane with respect to the ground plane length.

The simulated return loss of the proposed CPW-fed PRMA with the variation in space between the rectangular monopole and ground plane is shown in fig 5 and their corresponding simulated data are shown in table 3. It is observed from the table 4 that the space (s) has a large impact on the upper band impedance bandwidth, while the lower band almost keeps unchanged.

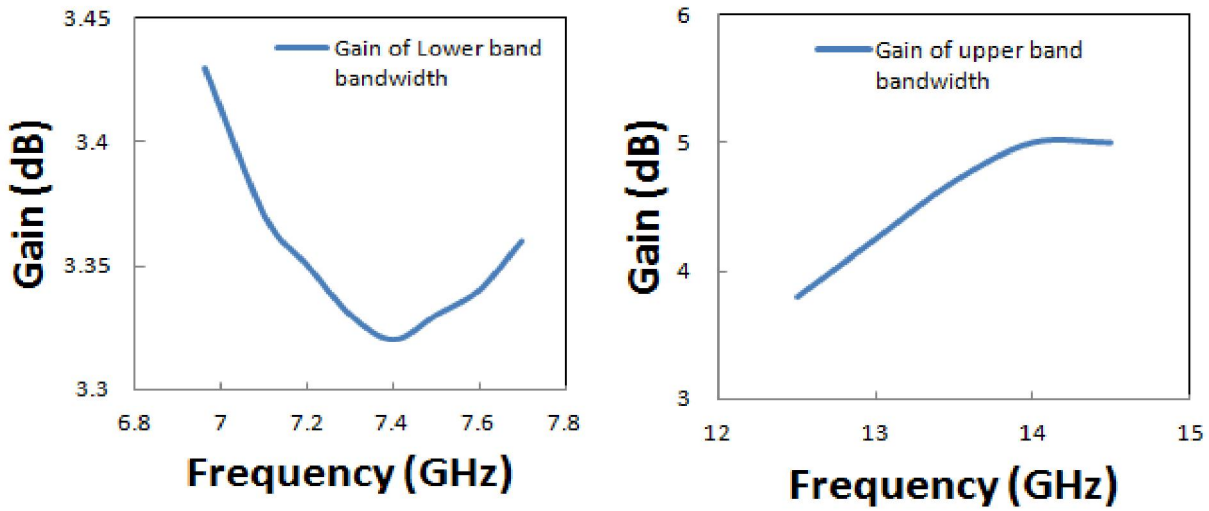
Similarly the simulated return loss of the proposed CPW-fed PRMA with the variation in space between the rectangular monopole and ground plane with respect to change of ground plane length is shown in fig 6 and table 4 lists the corresponding obtained simulated bandwidths for d varying from 7.1 to 11.1 mm for comparison. It is clearly seen in fig 5 that both upper and lower bands of bandwidth and their respective resonant frequencies is varied by the change of the space (s) with respect to the ground plane length. The lower and upper band bandwidths and their respective resonant frequencies are decreases with increase the distance of space (s).

Table 3. Table 3. Simulated bandwidths of proposed CPW-fed PRMA as a function of 'd'

'd' (mm)	Bandwidth (GHz)	Resonant frequency (GHz)
5	0.81, (7.02-7.83) 2.36, (12.80-15.16)	7.4 13.5
6	0.80, (7.00-7.80) 2.10, (12.74-14.84)	7.4 13.6
7	0.75, (6.96-7.71) 1.83, (12.67-14.50)	7.3 13.4
8	0.69, (6.91-7.60) 1.99, (12.56-14.55)	7.2 13.2
9	0.65, (6.84-7.49) 2.12, (12.42-14.54)	7.1 13

Table 4. Simulated bandwidths of proposed CPW-fed PRMA as a function of 'd' with respect to the ground plane length.

'd' (mm)	**Bandwidth (GHz)**	**Resonant frequency (GHz)**
7.1	0.79, (7.40-8.19) 1.91, (13.26-15.17)	7.8 13.9
8.1	0.75, (6.96-7.71) 1.83, (12.67-14.50)	7.3 13.4
9.1	0.67, (6.56-7.23) 1.75, (12.07-13.82)	6.9 13
10.1	0.55, (6.20-6.75) 1.73, (11.61-13.34)	6.5 12.5



(a) *(b)*

Fig. 7. Measured antenna gain for frequencies across (a) the lower band, (b) the upper band of the proposed CPW-fed rectangular monopole antenna.

The antenna gain (dB) of the proposed CPW-fed PRMA across the two bands is shown in Fig 7. For the lower band, the maximum antenna gain of about 3.43 dB is observed and for the upper band, the antenna gain is 5.0 dB. The VSWR (Voltage Standing Wave Ratio) and the efficiency (%) of proposed antenna against the frequency are shown in fig 8.

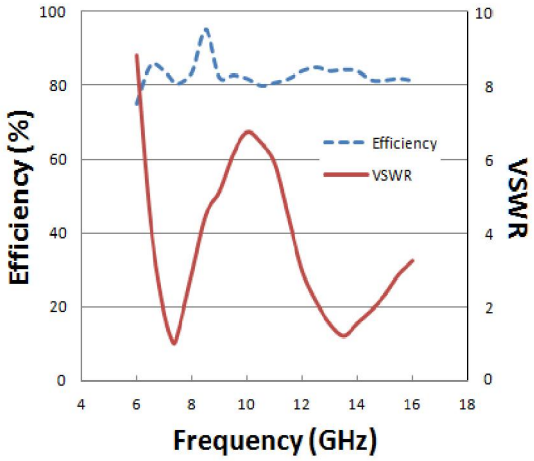


Fig. 8. VSWR and efficiency curve of the proposed CPW-fed PRMA.

4 CONCLUSION

A single-layer rectangular printed monopole antenna based on CPW-fed technology on the FR4 lossy substrate for obtaining two separate wide operating bands has been presented, with simulation results. Good antenna performances such as the wide impedance band-widths, monopole-like radiation patterns, and good antenna gains for the operating frequencies across the two bands have been obtained. The proposed CPW-fed monopole antenna has a very simple structure, which makes the design simpler and fabrication easier, and is very suitable for applications in the access points of wireless communications. It is also investigated that both gap 'g' (distance between the single strip and the coplanar ground plane) and spacing 'd' (distance between the patch and edge of the ground plane) is a frequency dependent parameter which effects the bandwidth of the antenna.

REFERENCES

[1] W.C. Liu, "Broadband dual-frequency meandered CPW-fed monopole antenna," *Electronics Letters*, Vol. 40, no. 21, pp. 1319–1320, 2004.

[2] F.R. Hsiao, K.L. Wong, "Compact planar inverted-F patch antenna for triple-frequency operation," *Microwave Optical Technology Letter*, Vol. 33, no. 6, pp. 459–462, 2002.

[3] S.D. Targonski, R.B. Waterhouse, D.M. Pozar, "Design of wide-band aperture-stacked patch microstrip antennas," *IEEE Transaction Antennas Propagation*, Vol. 46, No. 9, pp. 1245–1251, 1998.

[4] W. Choi, S. Kwon, B. Lee, "Ceramic chip antenna using meander conductor lines," *Electronics Letters*, Vol. 37, no. 15, pp. 933-394, 2001.

[5] L.S. Lee, P.S. Hall, P Gardner, "Compact wideband planar monopole antenna," *Electronics Letters*, Vol. 35, no. 25, pp. 2157–2158, 1999.

[6] N.W. Chen, Y.C. Liang, "An ultra-broadband, coplanar-waveguide fed circular monopole antenna with improved radiation characterisctics," *Progress in Electromagnetic Research C*, Vol. 9, pp. 193-207, 2009.

[7] M.J. Ammann, Z.N. Chen, "Wideband monopole antennas for multi-band wireless systems," *IEEE Antennas Propagation Magazine,* Vol. 45, No. 2, pp. 146-150, 2003.

[8] N.P. Agarwall, G. Kumar, K. Ray, "Wide-Band Planar Monopole," *IEEE Transaction and Antenna Propagation,* Vol. 46, No. 2, pp. 294-295, 1998.

[9] Naffal Herscovici, Christos Christoduoloc, "Wideband monopole antennas for multi-band wireless systems," *IEEE Trans Antennas and Propagation Magazine*, Vol. 45, No. 2, 2003.

[10] Rabih Rahaoui and Mohammed Essaaidi, "Compact Cylindrical Dielectric Resonator Antenna excited by a Microstrip Feed Line," *International Journal of Innovation and Applied Studies*, vol. 2, no. 1, pp. 1–5, January 2013.

[11] Mohammed Younssi, Achraf Jaoujal, Yacoub Diallo, Ahmed El-Moussaoui, and Noura Aknin, "Study of a Microstrip Antenna with and Without Superstrate for Terahertz Frequency," *International Journal of Innovation and Applied Studies*, vol. 2, no. 4, pp. 369–371, April 2013.

[12] M. I. Hasan and M. A. Motin, "New slotting technique of making compact octagonal patch for four band applications," *International Journal of Innovation and Applied Studies*, vol. 3, no. 1, pp. 221–227, May 2013.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 765-773

http://www.issr-journals.org/ijias/

Petrophysical Properties Evaluation for Reservoir Characterisation of SEYI Oil Field (Niger-Delta)

John Olurotimi AMIGUN and Oluwaseyi Ayokunle ODOLE

Department of Applied Geophysics, Federal University of Technology, Akure, Ondo State, Nigeria

ABSTRACT: This paper presents results of a study conducted to determine and evaluate the petrophysical properties of "SEYI" oil field, Niger delta with a view to understand their effects on the reservoirs hydrocarbon prospect and oil productivity of the field. The evaluated properties include porosity, permeability, fluid saturation, net / gross thickness and mobility which are all inferred from geophysical wire-line logs. A suite of wire-line logs comprising of gamma ray, resistivity, spontaneous potential and density logs for four wells from 'SEYI' oil field were analysed for reservoir characterisation of the field. The analyses carried out involves delineation of lithologies, identification of reservoirs and fluid types, wells correlation and determination of petrophysical parameters of identified reservoirs. Seven reservoirs namely: A, B, C, D, E, F and G were delineated with their tops and bases at depth from 2396m to 3429m. Their analysis indicate the presence of hydrocarbon in all the reservoirs. Computed petrophysical parameters across the reservoirs gave porosity as ranging from 0.22 to 0.31; permeability 881.58md to 14425.01md and average hydrocarbon saturation of 41.44%, 20.29%, 30.82%, 37.92%, 51.20%, 91.97% and 85.11% for reservoir A, B, C, D, E, F and G respectively. These results together with the determined movable hydrocarbon index (MHI) values (0.05 to 0.75) of the reservoir units suggest high hydrocarbon potential and a reservoir system whose performance is considered satisfactory for hydrocarbon production.

KEYWORDS: Wells correlation, wire-line logs, hydrocarbon prospect, porosity, Niger Delta.

1 INTRODUCTION

The 'SEYI" oil field is located within the coastal swampy area of Niger delta, Nigeria (Figure 1). The petrophysics of any oil field include reservoir fluid properties and reservoir rock properties which could affect oil recovery and amount of oil production. These properties encompass porosity, permeability, fluid saturation, mobility among others. The successful evaluations of these properties are necessary for determining the hydrocarbon potential of a reservoir system performance and also help to predict the behaviour of complex reservoir situations.

To the reservoir engineering these parameters are important i.e. rock porosity and fluid saturations are the principal factor involved in determining the amount of oil and gas originally in place while permeability is a measure of the ease with which fluid flows through the pore spaces of rock [1]. Table 1 according to [2] provides an effective explanation of porosity and permeability description of reservoirs. The movable hydrocarbon index, MHI value specifies the movability condition of hydrocarbon in a reservoir during invasion.

In this study, the gamma ray (GR), spontaneous potential (SP), resistivity (LLD), and density (PHID) logs were analysed and interpreted in order to define lithologic units of prospective zones, differentiating between hydrocarbon bearing and non-hydrocarbon bearing zone(s), definition of reservoir geometry through well to well correlation and determination of the petrophysical parameters value of zones of interest (reservoirs) in the field such as porosity, permeability, movable hydrocarbon index, water saturation and hydrocarbon saturation.

Corresponding Author: John Olurotimi AMIGUN (johnamigun@yahoo.com)

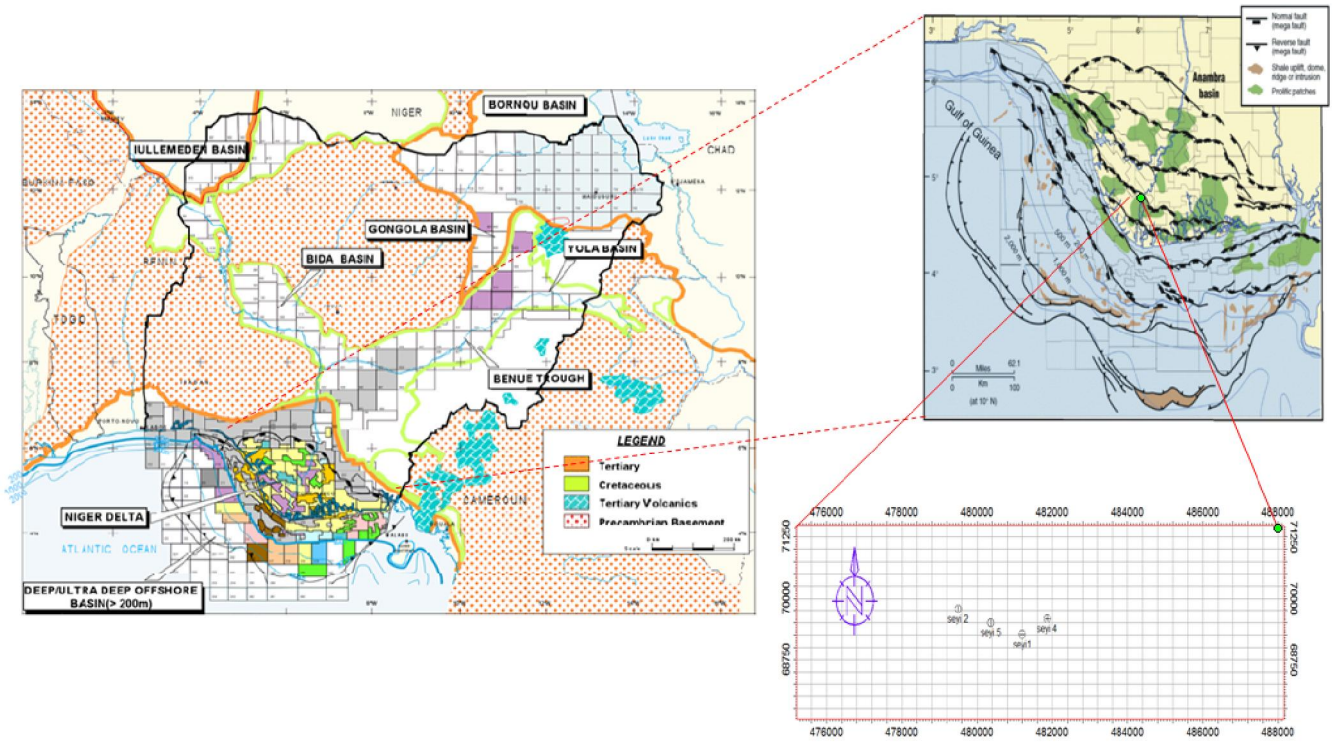


Fig. 1. Map of Nigeria Showing the Location of the Niger Delta and the Base map of study area showing the well locations

Table 1. Porosity and Permeability values for Reservoirs Qualitative Description (Adapted from Rider, 1986)

Qualitative Evaluation of Porosity	
Percentage Porosity (%)	***Qualitative Description***
0 - 5	Negligible
5 - 10	Poor
15 - 20	Good
20 – 30	Very Good
> 30	Excellent
Qualitative Evaluation of Permeability	
Average K_Value (md)	***Qualitative Description***
< 10.5	Poor to fair
15 – 50	Moderate
50 – 250	Good
250 – 1000	Very Good
> 1000	Excellent

2 GEOLOGY OF THE STUDY AREA

The Niger Delta which extend between longitude 3^0 and 9^0E and latitudes $4^0 30'$ and $5^0 20'$ N, covers an area of about 7500 km^2. It is a large acute delta situated in the Gulf of Guinea (Fig.1). From the Eocene to the present, the delta has prograded southwestward, forming depobelts that represent the most active portion of the delta at each stage of its development [3]. These depobelts form one of the largest regressive deltas in the world. Its stratigraphy has been described in detail by [4] and [5] of which they recognized three lithostratigraphic units i.e. the Benin, Agbada and Akata Formations (Fig.2).

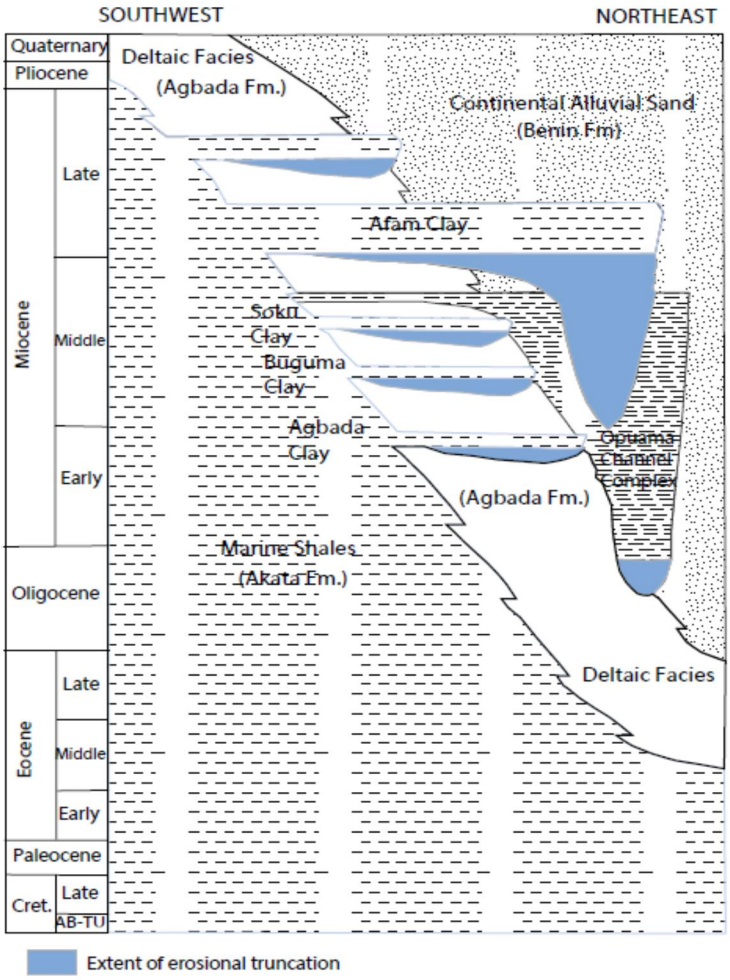


Fig. 2. Stratigraphic Column showing the three Formations of the Niger Delta (Modified after Doust and Omatsola, 1990)

The Benin Formation is a continental Eocene to Recent deposit of alluvial i.e. up to 2000 m thick, the uppermost and youngest rock stratigraphic unit, while the Agbada Formation is a paralic sequence that is characterized by the alternation of sand bodies and shale layers. It is also associated with synsedimentary growth faulting and as well contains the bulk of the known oil accumulation in the Niger Delta. The Akata formation is the lowest unit at the base of the delta. It is of marine origin and is composed of thick shale sequences (potential source rock), turbidite sand (potential reservoirs in deep water), and minor amounts of clay and silt. The Akata formation is undercompacted in much of the delta [6].

The growth faults in the Niger Delta according to [7] are considered to be the major migration conduit and leading factor controlling the hydrocarbon distribution pattern in the Niger Delta. Most known traps in Niger Delta fields are structural although stratigraphic traps are not uncommon.

3 MATERIAL AND METHODS

3.1 DATA ANALYSIS

The analysis of petrophysical logs in this study was aimed at a qualitative and quantitative determination of the properties of delineated reservoirs. The gamma ray (GR) and spontaneous potential (SP) logs were examined for lithologic information. In the reservoir formations vis a vis at shale beds, gamma ray (GR) log which measures natural radioactivity in formations reflects the shale contents while the SP log displays excursion from the shale base line, hence both logs were used for the identification of sand / shale lithology in the study area. The resistivity log in combination with the GR log were used to differentiate between hydrocarbon and non-hydrocarbon bearing zones. In hydrocarbon bearing formation, the resistivity log signatures show high resistivity values than when in water bearing formation. The discrimination of the various

fluid types i.e. oil / gas within reservoirs could not be achieved because of the non availablity of neutron log among materials used in carrying out the study.

Using the gamma ray log, lithologic correlation of equivalent strata across the four wells was performed by matching for similarity the intervals of logs from different wells. At the same time, the potential hydrocarbon reservoirs in the various wells were correlated using the gamma ray and resistivity logs to determine their lateral extent, possible existence of faults, dips and or unconformities. Their outcome is presented as correlation panels shown in Figure 3

Quantitatively, the petrophysical parameters are estimated using empirical formulae as follows;

3.1.1 SHALE VOLUME ESTIMATION:

Shale volume (V_{sh}) was calculated using the [8] formula in equation (1) which uses values from the gamma ray (GR) in equation (2)

$$V_{sh} = 0.083^{(2(3.7 \times IGR))} - 1.0 \quad (1)$$

$$I_{GR} = \frac{GR_{log} - GR_{min}}{GR_{max} - GR_{min}} \quad (2)$$

In equation (2), I_{GR} is the gamma ray index, GR_{log} is the picked log value while GR minimum and GR maximum indicate values picked in the sand and shale base lines respectively.

3.1.2 DETERMINATION OF POROSITY

Porosity, ϕ_D is defined as the percentage of voids to the total volume of rock. According to [8], this parameter is determined by substituting the bulk density readings obtained from the formation density log within each reservoir into equation (3)

$$\phi_D = \frac{\rho_{ma} - \rho_b}{\rho_{ma} - \rho_f} - V_{sh} \frac{\rho_{ma} - \rho_b}{\rho_{ma} - \rho_f} \quad (3)$$

And where ρ_{ma}, ρ_b and ρ_f are matrix density, formation bulk density and fluid density respectively.

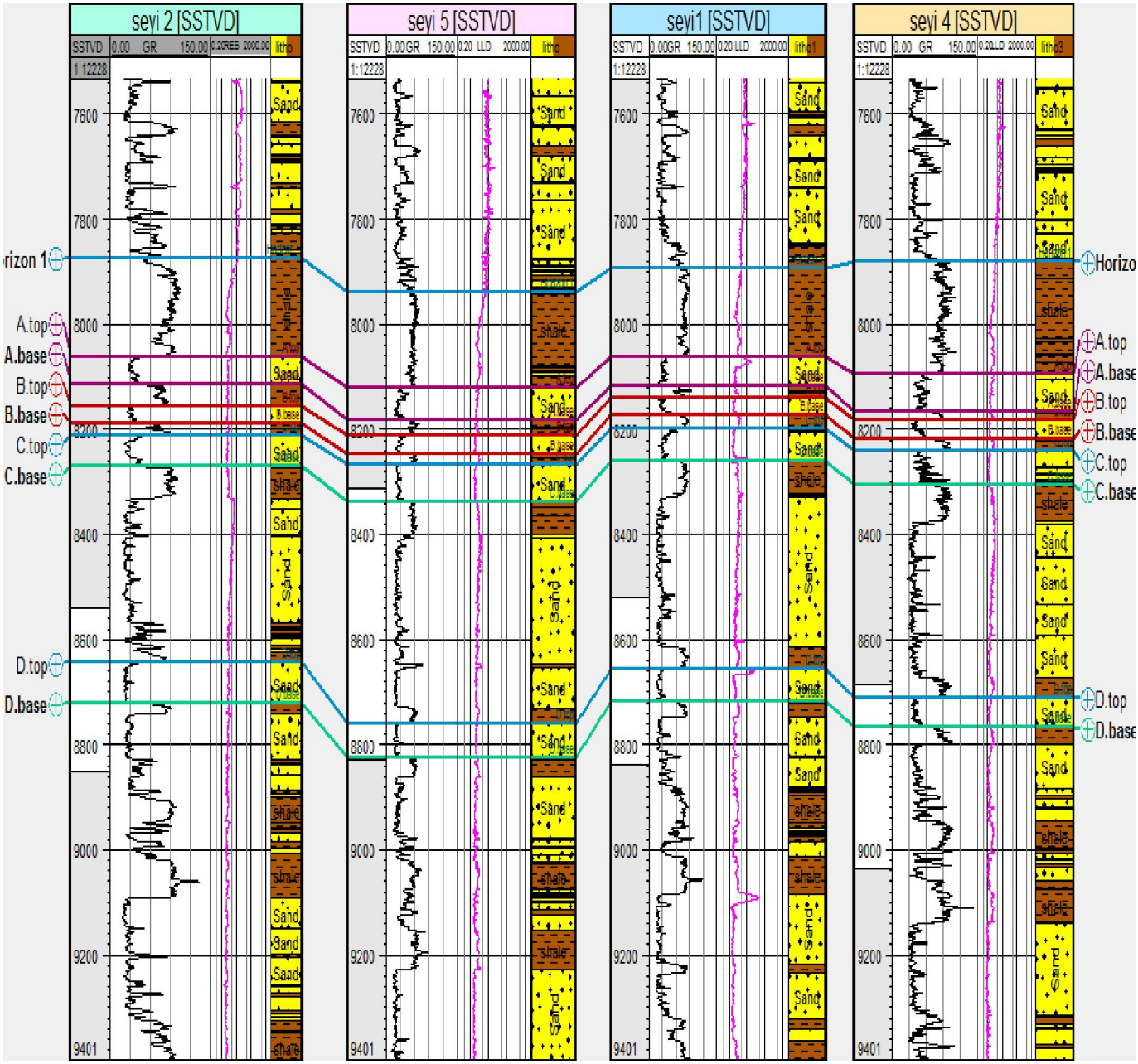


Fig. 3. Lithologic Correlation Panel showing Delineated Reservoirs at Depths Interval (2396m - 2896m) across SEYI field

3.1.3 CALCULATION OF WATER SATURATION

To calculate water saturation, S_w of uninvaded zone, the method used requires a water resistivity R_w value at formation temperature calculated from the porosity and resistivity logs within clean water zone, using the R_o method given by the following equation:

$$R_w = \frac{\phi^m R_o}{a} \quad (4)$$

R_w is the water resistivity at formation temperature, ϕ and R_o are the total porosity and deep resistivity values in the water zone respectively. Tortuosity factor is represented as "a" and m is the cementation exponent, usually 2 for sands [9]. In the water zone, saturation should be equal to 1, as water resistivity R_w at formation temperature is equal to R_{wa}

Water saturation, S_w can then be calculated using Archie's method, given by:

$$S_w = \left(\frac{R_w}{R_{wa}}\right)^{1/n} \quad (5)$$

where n is the saturation exponent and R_{wa} is water resistivity in the zone of interest, calculated in the same manner as R_W at formation temperature [10].

3.1.4 DETERMINATION OF HYDROCARBON SATURATION

Hydrocarbon Saturation, S_h is the percentage of pore volume in a formation occupied by hydrocarbon. It can be determined by subtracting the value obtained for water saturation from 100% i.e.

$$S_h = (100 - S_W)\ \% \quad (6)$$

3.1.5 CALCULATION OF PERMEABILITY

Permeability, K is the property of a rock to transmit fluids. For each identified reservoir permeability, K is calculated using equation (7)

$$K = \sqrt{\frac{250\ x\ \phi^2}{S_{wir}}} \quad (7)$$

where S_{wir} is the irreducible water saturation [11]

3.1.6 ESTIMATION OF MOVABLE HYDROCARBON INDEX (MHI)

The movable hydrocarbon index (MHI) was derived using:

$$MHI = S_w / S_{xo} \quad (8)$$

Where MHI > 1 implies that hydrocarbon was not moved during invasion and MHI < 0.7 implies that hydrocarbon was moved during invasion. The parameters S_w and S_{xo} are water saturation of the uninvaded zone and the flushed zone respectively.

The productivity of each delineated reservoir rock at the zone of interest are estimated by evaluating results of their calculated petrophysical parameters using equation (1) – (8) .

4 RESULTS AND DISCUSSION

4.1 QUALITATIVE INTERPRETATION

For the log interpretation in Figure 3, its litho-stratigraphic correlation furnishes knowledge of the general stratigraphy of the study field. Seven sand bodies marked reservoir A, B, C, D, E, F and G were correlated across the field. This analysis show that each of the sand units extends through the field, varies in thickness and some units occuring at greater depth than their adjacent unit i.e. possibly an evidence of faulting. The shale layers were observed to increase with depth along with a corresponding decrease in sand layers. This pattern in the Niger delta indicates a transition from Benin to Agbada Formation. From the analysis, particularly the resistivity logs, all the seven delineated reservoirs were identified as hydrocarbon bearing reservoir across the four wells i.e. SEYI-2, SEYI-5, SEYI-1 and SEYI-4.

4.2 QUANTITATIVE INTERPRETATION

The methodology as earlier reported was chosen for the quantitative interpretation of the delineated reserviors in each well. Table 2 represents the results of some computed petrophysical parameters for well SEYI-4 in reservoir A - G. In the case of reservoir C in Figure 3, this reservoir was penetrated at depths (2503 - 2521) m in well SEYI-2, (2520 - 2542) m in SEYI-5, (2498 - 2518) m in SEYI - 1 and (2512 - 2530) m in well SEYI-4. The reservoir has an average gross thickness of 19.51m and an average net/gross ratio of 96.11%. The average effective porosity, permeability, water saturation and hydrocarbon saturation of the reservoir are 0.26, 3589 md, 69.18% and 30.82% respectively. The movable hydrocarbon index (MHI) ranges from 0.41 to 0.92 with bulk volume of movable oil ranging between 1.99% and 6.26%. Shown in Figure 4 are charts depicting the relationship in percentage of some of the calculated petrophysical parameters for reservoirs E and G across the wells.

Table 2. Some Computed Petrophysical Parameters for well SEYI-4 Field

PARAMETERS	RESERVOIRS						
	A	B	C	D	E	F	G
Net /Gross (%)	100	100	84.45	100	68.53	77.99	27.41
V_{sh}	0.176	0.121	0.161	0.159	0.376	0.224	0.242
ϕ_D	0.38	0.32	0.32	0.31	0.28	0.2	0.17
ϕ_e	0.31	0.28	0.27	0.26	0.3	0.21	0.24
R_o	0.63	0.63	0.63	0.63	0.63	0.63	0.63
R_t	2.07	1.34	1.29	1.42	58.26	115.29	1097.91
Rw	0.08	0.07	0.06	0.06	0.08	0.04	0.05
F	7.69	9.57	10.35	11.23	8.25	17.77	13.33
Swirr (%)	6.2	6.92	7.19	7.49	6.42	9.43	8.16
Sw (%)	55.17	68.57	69.88	66.61	10.4	7.39	2.4
Sh (%)	44.83	31.43	30.12	33.39	89.6	92.61	97.6
Sxo (%)	88.78	92.73	93.08	92.19	63.59	59.4	47.41
K (md)	14425.01	6292.98	4678.78	3439.94	11042.22	603.4	1791.6
BVW (%)	17.1	19.2	18.87	17.32	3.12	1.55	0.57
I	3.29	2.13	2.05	2.25	92.48	183	1742.71
MHI	0.62	0.74	0.75	0.72	0.16	0.12	0.05

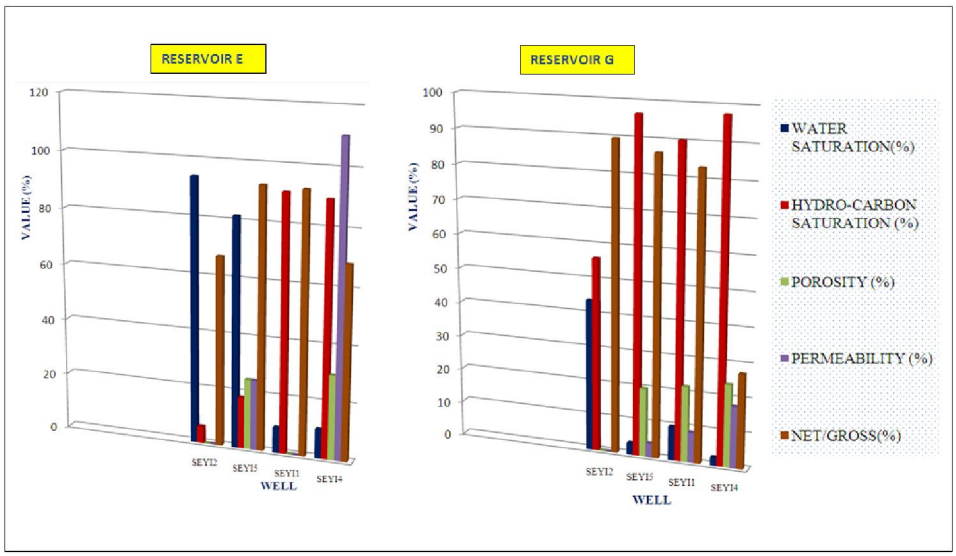


Fig. 4. Chart showing relationship between percentage effective porosity, permeability, water saturation, hydrocarbon saturation and net/gross of RESERVOIR E and G

In Table 3 are summary of results of the important petrophysics parameters utilize as variables that determine reservoir quality. These parameters are subjected to statistical analysis by considering their values across all the delineated reservoirs in the four wells of the study area (Fig. 5). The reservoirs porosity was estimated from density log (RHOB) using porosity formula with shale correction by [8] and these values ranges from 0.22 to 0.31 indicating an outstanding reservoir quality and reflecting probably well sorted coarse grained sandstone reservoirs with minimal cementation. The reservoir units' permeability range was from 881.58md to 14425.01md with reservoir unit A having the highest value. In the appraisal of a well's productivity, the permeability of reservoir unit is an important parameter vis a vis productivity is related to the mathematical product of permeability, k, and thickness, h of that unit. The hydrocarbon bearing reservoir units in the study field have hydrocarbon saturation, S_h ranging from 20.29% to 91.97% indicating that the proportion of void spaces occupied by water is low consequently high hydrocarbon saturation and high hydrocarbon production. The movability of hydrocarbon

in each reservoir was determined (see Table 2) and considered satisfactory for the production of hydrocarbon. According to [12], if OMI is less than 0.7 for sandstone, then movable hydrocarbon is indicated.

Table 3. Summary of the Average Petrophysical Parameters for all Reservoirs in SEYI field

RESERVOIR	WELL NAME	ϕ (%)	K (%)	S_W (%)	S_H (%)	NET/GROSS (%)
A	SEYI-2	-	-	75.48	24.52	100
	SEYI-5	22	8.82	76.91	23.09	100
	SEYI-1	-	-	26.7	73.3	100
	SEYI-4	31	144.25	55.17	44.83	100
	Average	26.5	76.53	58.57	41.44	100
B	SEYI-2	-	-	85.77	14.23	100
	SEYI-5	25	24.99	89.92	10.08	100
	SEYI-1	-	-	74.57	25.43	100
	SEYI-4	28	62.93	68.57	31.43	100
	Average	26.5	43.96	79.71	20.29	100
C	SEYI-2	-	-	84.48	15.52	100
	SEYI-5	25	24.99	89.92	10.08	100
	SEYI-1	-	-	32.44	67.56	100
	SEYI-4	27	46.79	69.88	30.12	84.45
	Average	26	35.89	69.18	30.82	96.11
D	SEYI-2	-	-	74.07	25.93	100
	SEYI-5	24	17.91	85.69	14.31	100
	SEYI-1	-	-	21.97	78.03	100
	SEYI-4	26	34.4	66.61	33.39	100
	Average	25	26.16	62.09	37.92	100
E	SEYI-2	-	-	94.06	5.94	67.37
	SEYI-5	25	24.99	81.65	18.35	92.67
	SEYI-1	-	-	9.1	90.9	92.3
	SEYI-4	30	110.42	10.4	89.6	68.53
	Average	27.5	67.71	48.8	51.2	80.22
F	SEYI-2	-	-	-	-	-
	SEYI-5	-	-	-	-	-
	SEYI-1	28	62.93	8.67	91.33	56.76
	SEYI-4	21	6.03	7.39	92.61	78
	Average	24.5	34.48	8.03	91.97	67.38
G	SEYI-2	-	-	43.91	56.09	89.65
	SEYI-5	20	4.05	3.44	96.56	86.49
	SEYI-1	22	8.82	9.81	90.19	83.25
	SEYI-4	24	17.92	2.4	97.6	27.41
	Average	22	10.26	14.89	85.11	71.7

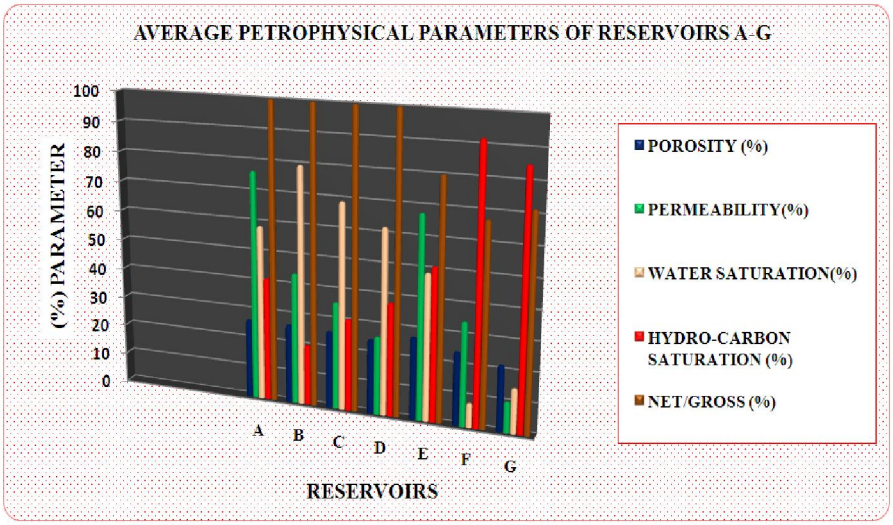


Fig. 5. Relationship between percentage average effective porosity, permeability, water saturation, hydrocarbon saturation and net/gross of RESERVOIR A - G

5 CONCLUSION

The petrophysical properties evaluation of 'SEYI' field for its reservoirs characterisation was made possible by careful analysis and interpretation of its well logs. The results shows the feild's delineated reservoir units having porosity ranging from 0.22 to 0.31 indicating a suitable reservoir quality, permeability values from 881.58md to 14425.01md attributed to the well sorted nature of the sands and hydrocarbon saturation range from 20.29% to 91.97% implying high hydrocarbon production. These results in addition with the reservoir units oil movability index (MHI) values suggest high hydrocarbon potential and a reservoir system whose performance is considered satisfactory for hydrocarbon production.

REFERENCES

[1] E. Abdolla, "Evaluation of Petrophysical Properties of an Oil Field and their effects on production after gas Injection," *1st International Applied Geological Congress*, Department of Geology, Islamic Azad University – Mashad Branch, Iran, pp. 26-28, 2010.

[2] M. Rider, "The Geological Interpretation of Well Logs," *Blackie*, *Glasgow*, pp. 151-165, 1986.

[3] H. Doust and O. Omatsola, *Niger delta*. In: Edwards, J., Santagrossi, P. (Eds.), *Divergent Passive Margins*. American Association of Petroleum Geologists, Memoir 48, pp. 191-248, 1990.

[4] K.C. Short and A.J. Stauble, "Outline of geology of Niger Delta," *American Association of Petroleum Geologists Bulletin*, vol. 51, pp. 761-779, 1967.

[5] E.J. Frankl and E.A. Cordy, "The Niger Delta Oil Province: Recent developments onshore and offshore," *Proceedings of Seventh World Petroleum Congress,* Mexico City, pp. 195-209, 1967.

[6] A. A. Avbovbo, "Tertiary lithostratigraphy of Niger Delta," *American Association of Petroleum Geologists Bulletin*, vol. 62, pp. 295-300, 1978.

[7] K.J. Weber, "Hydrocarbon Distribution Pattern in Nigerian growth fault structures controlled by structural style and stratigraphy," *Journal of Petroleum Science and Engineering*, 1, pp. 1-12, 1987.

[8] A. Dresser, "Log interpretation charts," Houston Dresser Industries, Inc. pp. 107, 1979.

[9] G. Asquith and D. Krygowski, "Basic Well Log Analysis," *AAPG Methods in Exploration* Series, No 16, 2004.

[10] G.E. Archie, "The Electrical Resistivity as an Aid in Determining Some Reservoir Characteristics," *Journal of Petroleum Technology*, vol. 5, pp. 54-62, 1942.

[11] M.P. Tixier," Evaluation of Permeability from Electric Log Resistivity Gradients," *Oil and Gas Journal*, vol. 8, pp. 75-90, 1949.

[12] Schlumberger, "Principle and application of log interpretation," *Schlumberger Education services*, Houston, 1989.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 774-780

http://www.issr-journals.org/ijias/

Supply and Demand of Jeneberang River Aggregate Using Multiple Regression Model

Aryanti Virtanti Anas[1], D.A. Suriamihardja[2], Saleh Pallu[3], and Ulva Ria Irfan[4]

[1]Department of Civil Engineering/Mining Engineering,
Hasanuddin University,
Makassar, South Sulawesi, Indonesia

[2]Department of Geophysics,
Hasanuddin University,
Makassar, South Sulawesi, Indonesia

[3]Department of Civil Engineering,
Hasanuddin University,
Makassar, South Sulawesi, Indonesia

[4]Department of Geology Engineering,
Hasanuddin University,
Makassar, South Sulawesi, Indonesia

ABSTRACT: Aggregate plays an important role in developing infrastructure because it is the major raw materials used in construction such as roads, hospitals, schools, factories, homes and other buildings. Sand and gravel are essential sources of aggregate and exploited often from the active channels of river systems. Jeneberang River is one of the main rivers in South Sulawesi Province which is located at Gowa Regency and mined in order to fulfill the aggregate demand of Gowa Regency and Makassar City. Supply and demand are economic occurrences that affected by several factors, so this research aims to (1) determine influencing factors to aggregate supply and demand, (2) develop supply and demand model. Data was obtained from Central Bureau Statistics of Gowa Regency and Makassar City, and Department of Mines and Energy, Gowa Regency for eleven years (2001 – 2011). In this research, aggregate supply and demand were modeled using multiple regression method. First, relationship among supply and influencing factors were established, followed by demand and its factors. Second, supply and demand model was established using SPSS. The result of this research showed that the model can be used to estimate accurately supply and demand of aggregate using the established relationship among the influencing factors. Supply of aggregate was affected by several factors including price, number of trucks, number of mining companies and mining permit area meanwhile the price, GDP, income per capita, length of road, number of buildings and economic growth had high influence on demand rate.

KEYWORDS: Jeneberang River, infrastructure, aggregate mining, sand and gravel, supply and demand, influencing factors.

1 INTRODUCTION

1.1 JENEBERANG RIVER

Jeneberang River is one of the main rivers in South Sulawesi Province. It is located at Gowa Regency and flows east to west across the province to Makassar City. The river is 85.5 km long with a catchment area of 762.01 km^2, originating from Mt. Bawakaraeng (2,833 m) which is located 90 km from Makassar. In the midstream area of the Jeneberang River, the

Corresponding Author: Aryanti Virtanti Anas (virtanti@gmail.com)

multipurpose Bili-bili Dam provides some benefits to the community such as flood control, water supply, irrigation and power plant [1].

On 26 March 2004, gigantic caldera wall collapsed at the east ridge of caldera of Mt. Bawakaraeng. The volume of the collapsed mass was estimated at about 235 million m^3 (originally) and based on more detailed survey the collapse was estimated to be 231 million m^3. Small to middle scale collapsed deposits were estimated at 1.6 million m^3. Surface erosion (2004-2009) was estimated at 11.7 million m^3. In 2009, the remains of the collapsed deposit in the caldera were estimated at 82.7 million m^3 and sediment volume flowing into the Jeneberang River was estimated at 162.2 million m^3 [1]. After the collapse, the Jeneberang River was supplied with a large amount of sediment that are mined conveniently and economically to supply aggregates demand of Gowa Regency and Makassar City. Mining activities become a part of sediment control plan to mitigate the potential debris flow resulted from the Mt. Bawakaraeng collapse.

1.2 AGGREGATE MINING

The most common natural aggregates of mineral origin are sand, gravel, and crushed rock. Aggregate is basic raw material required for all construction activities. It is impossible to construct a city without using natural aggregate. Aggregate is used in many different applications such as road-building, rail ballast, mass concrete for foundation or major structures, concrete blocks, steel reinforced beams, flooring and walls, mortar, plaster and filter media for sewage and other water treatment [2]. The aggregate produced from natural sources extracted from quarries and gravel pits and used either in their natural state or after crushing, washing, and sizing [3]-[4].

Throughout the history, the Jeneberang River mining started during the development of the Gowa Kingdom Forts under the reign of Karaeng Tumapa'risi Kallonna in 1525. The Forts were built from stones that are presumed to have been taken from the river [5]. Currently, there are eighteen active mining sites in midstream of the Jeneberang River which mining permit area varies between 5 – 60 hectares. A permit for the extraction of sand and gravel is given by Department of Mines and Energy, Gowa Regency. The permit is typically effective for 1-5 years, at which time the applicant has the option to apply for a renewal permit.

1.3 SUPPLY AND DEMAND

Supply in economic term essentially measures quantity of a product or service to be offered at varying price points, meanwhile demand measures quantity of a product or service to be purchased at varying price points [6]. Aggregate demand modeled generally fall under three main methods [7]: (1) historical trend: using such simple assumptions as recent per capita consumption or recent average annual levels; (2) regression models: using either macroeconomic indicators (such as GDP, population, unemployment rate, etc.), and (3) construction input factors: these may be either space based (i.e. tons per sq. ft. of different types of construction).

It is important in aggregate resources to forecast demand because the forecast facilitates an evaluation of the sufficiency of the resource supply to meet the expected demands. In order to forecast aggregate supply and demand we developed an economic model using multiple regression method. Regression method is a tool that widely used in many study areas, because it can be easily modeled using simple assumptions. This method measures the degree of influence of the independent variables on a dependent variable. A regression with two or more explanatory variables is called a multiple regression. Multiple regression is used to test the effects of n independent (predictor) variables on a single dependent (criterion) variable (Supranto, 2005). The general model of multiple regression is given:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon$$

Where y is the predicted trend, x_1, x_2, …, x_n are the influence factors of predicted trend, β_0, β_1 …, β_n are the regression coefficients, and ε is the residual variable.

1.4 AIMS OF RESEARCH

In general, changes in supply and demand level for aggregate occur because of previous or present changes in the characteristics of the overall economy. Thus, it would be expected that a relationship exists between trends in the level of aggregate supply and demand, and various indicators of the general economic activity. Consequently, the aims of this research are (1) to determine influencing factors to aggregate supply and demand, (2) to develop supply and demand model.

2 METHODOLOGY

In this research aggregate is considered as total sand and gravel. Data were collected primarily from Department of Mines and Energy, Gowa Regency and was also obtained from Central Bureau Statistics of Gowa Regency and Makassar City from 2001 to 2011.

The data obtained were subjected to multiple linear regression method using Statistical Product and Service Solutions (SPSS) Version 18.0 from which multiple regression model parameters were estimated and all graphs in this paper were drawn using Microsoft Excel 2010.

3 RESULTS AND DISCUSSIONS

Historical annual aggregate production data were obtained from Department of Mines and Energy, Gowa Regency but the demand for aggregate was determined using the rates of production. This is because there is no accurate source of information to compile the quantities of aggregate actually delivered to consumers. The assumption made, in order to substitute production rates for demand rates. Because only legal mining report their figures to the Department of Mines and Energy, so the figures do not account for the contribution of illegal operators in meeting the demands of the area. Using the Department of Mines and Energy's figures as a basis for projecting supply and demand will, therefore, provide a supply and demand estimation which is lower than the actual.

According to the Department of Mines and Energy, the greatest volume is produced by nine mining sites and the other ones have relatively small operation. There are some sand and gravel illegal mining that take place frequently in several locations but the number and quantity of production is unknown [8].

In this research, dependent variable for supply model was total supply meanwhile potential independent variables considered for supply model here included price, number of trucks, number of mining companies and mining permit area (Table 1).

Table 1. Total supply of aggregate and influencing factors of supply model

Year	Price (Rp/m^3)	Number of trucks (unit)	Number of mining companies (unit)	Mining permit area (ha)	Total supply (m^3)
2001	16,667.00	582	9	38.00	508,528.00
2002	20,833.00	633	9	38.00	602,709.00
2003	29,583.00	430	9	38.00	459,807.00
2004	35,000.00	501	9	38.00	289,078.56
2005	35,416.67	528	9	38.00	363,244.78
2006	44,412.00	522	9	38.00	377,610.57
2007	63,750.00	607	19	105.30	498,403.86
2008	70,062.50	790	25	138.26	555,243.00
2009	76,167.00	891	25	138.26	1,847,896.53
2010	79,375.00	919	12	93.42	1,162,553.79
2011	80,000.00	1164	15	144.42	503,453.92

The supply of aggregate was modeled as:

$$Q_s = -2{,}088{,}760.888 + 15.052\ x_1 + 3667.876\ x_2 + 171{,}417.766\ x_3 - 37{,}302.661\ x_4$$

Where:

Q_s = quantity of aggregate supplied (in metric cubic), annually

X_1 = price of aggregate

X_2 = numbers of trucks

X_3 = number of mining companies

X_4 = mining permit area

The dependent variable for demand model was total demand meanwhile the independent variables were price, GDP, income per capita, length of road, number of buildings and economic growth as shown in Table 2, Table 3 and Table 4.

Table 2. Total demand of aggregate, price and length of road

Year	Price (Rp/m^3)	Length of road (km)	Total demand (m^3)
2001	16,667.00	3041.46	508,528.00
2002	20,833.00	3101.46	602,709.00
2003	29,583.00	3229.27	459,807.00
2004	35,000.00	3287.39	289,078.56
2005	35,416.67	3369.80	363,244.78
2006	44,412.00	3410.81	377,610.57
2007	63,750.00	3256.62	498,403.86
2008	70,062.50	3315.25	555,243.00
2009	76,167.00	3405.81	1,847,896.53
2010	79,375.00	3241.65	1,162,553.79
2011	80,000.00	3159.34	503,453.92

Table 3. GDP, income per capita, number of buildings and economic growth of Makassar City

Year	GDP (in million rupiahs)	Income per capita (Rp)	Number of buildings (unit)	Economic growth (%)
2001	3,621,006.70	13,657,321.80	272,125.00	7.30
2002	3,621,006.70	13,385,901.02	278,596.00	7.14
2003	8,882,254.69	36,262,833.46	253,153.00	8.60
2004	9,785,333.89	39,149,322.02	260,003.00	10.17
2005	10,492,540.67	36,051,885.20	297,398.00	7.16
2006	11,341,848.21	39,493,315.12	298,485.00	8.09
2007	12,261,538.92	42,130,081.50	297,283.00	8.11
2008	13,561,827.18	46,169,494.04	300,544.00	10.52
2009	14,798,187.68	49,930,789.07	303,372.00	9.20
2010	16,252,451.43	52,100,959.69	312,087.00	9.83
2011	17,820,000.00	57,673,075.87	316,030.00	9.65

Table 4. GDP, income per capita, number of buildings and economic growth of Gowa Regency

Year	GDP (in million rupiahs)	Income per capita (Rp)	Number of buildings (unit)	Economic growth (%)
2001	1,134,805.53	1,017,361.00	125,177.00	6.02
2002	1,180,965.05	1,041,973.00	126,797.00	4.60
2003	1,234,698.07	1,038,350.00	128,087.00	4.02
2004	1,294,783.37	2,000,000.00	133,371.00	4.87
2005	1,369,696.51	2,381,679.00	138,191.00	5.74
2006	1,453,592.57	2,480,241.00	142,903.00	6.17
2007	1,543,568.30	2,596,751.00	144,858.00	6.19
2008	1,650,323.75	2,723,864.00	145,757.00	6.92
2009	1,782,158.63	2,889,942.00	152,223.00	7.99
2010	1,890,032.59	2,897,361.00	158,203.00	6.05
2011	2,007,276.99	3,043,579.00	159,757.00	6.20

The demand of aggregate was modeled as:

$$Q_d = 11{,}523{,}838.83 + 19.535\ x_1 + 5981.368\ x_2 + 6.187\ x_3 + 1.348\ x_4 + 0.455\ x_5 - 0.476\ x_6$$
$$- 29.475\ x_7 - 110.561\ x_8 + 34{,}500.193\ x_9 - 250{,}471.053\ x_{10}$$

Where:

Q_d = quantity of material construction demanded (in metric cubic), annually

x_1 = price of aggregate

x_2 = total length of road in Gowa Regency and Makassar City

x_3 = GDP of Gowa Regency

x_4 = GDP of Makassar City

x_5 = income per capita of Gowa Regency

x_6 = income per capita of Makassar City

x_7 = number of buildings of Gowa Regency

x_8 = number of buildings of Makassar City

x_9 = economic growth of Gowa Regency

x_{10} = economic growth of Makassar City

The analysis confirmed that these factors were all correlated with aggregate supply and demand to some degree and had the expected positive or negative signs. The fit of supply model was good with correlation coefficient 0.836 (r) coefficient of determination (R^2) 0.698. The fit of demand model was very good because it had correlation coefficient (r) 1.000, coefficient of determination (R^2) 1.000. The values showed that there was strong relationship between dependent and independent variables. Therefore, the models were good predictors to estimate aggregate supply and demand (Table 5).

Table 5. Regression estimates of aggregate supply and demand

Year	Estimated supply (m^3)	Estimated demand (m^3)
2001	422,065.26	508,528.00
2002	671,831.54	602,709.00
2003	589,534.68	459,807.00
2004	400,906.71	289,078.56
2005	506,210.87	363,244.78
2006	619,596.95	377,610.57
2007	426,141.16	498,403.86
2008	991,386.02	555,243.00
2009	1,453,723.46	1,847,896.53
2010	1,048,929.60	1,162,553.79
2011	568,783.98	503,453.92

The results of the theoretical and estimated data from supply and demand models were plotted and shown in Figures (1) and (2), respectively.

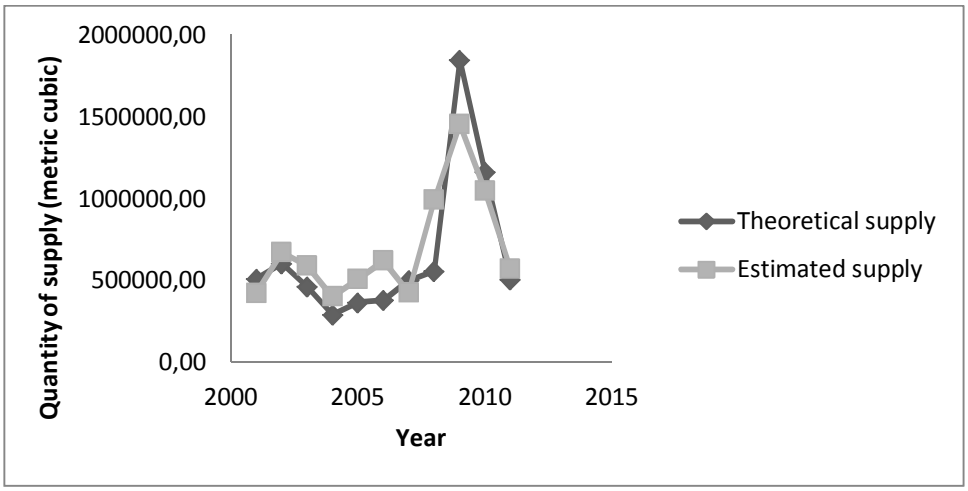


Fig. 1. Relationship between theoretical and estimated quantity of aggregate supply

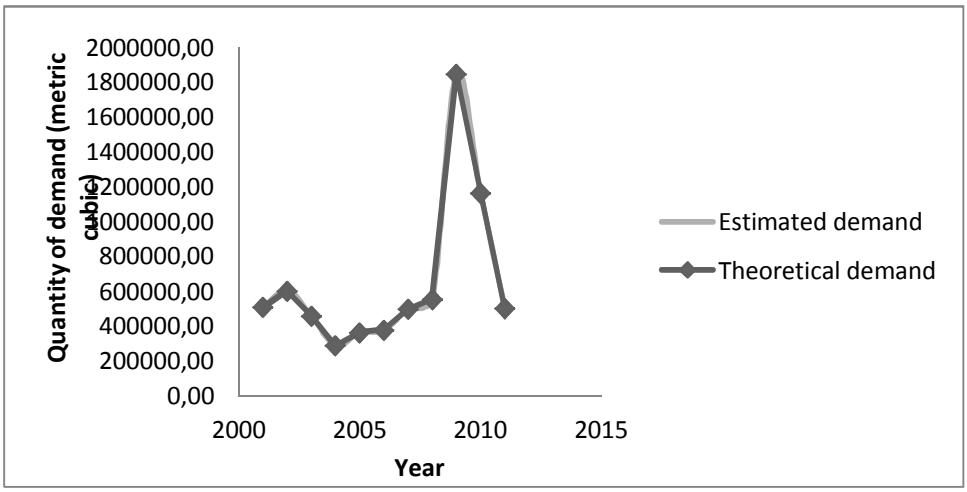


Fig. 2. Relationship between theoretical and estimated quantity of aggregate demand

4 CONCLUSION

It was established that aggregate supply and demand have been affected by number of factors ranging from price, number of trucks, number of mining companies, mining permit area, GDP, income per capita, length of road, number of buildings and economic growth. It was also established that the influencing factors can be effectively modeled using multiple linear regression approach and obtained a good relationship for supply and demand with their influencing factors.

It is recommended that the supply and demand of aggregate be monitored on a periodic basis (such as every year) to see how their trend, as well as to incorporate relevant updates of economic, population, and infrastructure growth.

ACKNOWLEDGMENT

The authors are sincerely thankful and grateful to Department of Mines and Energy, Gowa Regency and Central Bureau Statistic of Gowa Regency and Makassar City for the supports of this research.

REFERENCES

[1] Main Office of Pompengan Jeneberang River Basin, *Water Resources Management in Jeneberang River Basin*, Makassar, South Sulawesi, 2010.

[2] Scott, P.W., Eyre, J.M., Harrison, D.J. and Steadman, E.J, "Aggregate Production and Supply in Developing Countries with Particular Reference to Jamaica," *British Geological Survey*, Keyworth, Nottingham, 2003.

[3] Union Européenne des Producteurs de Granulats (UEPG), "A Sustainable Industry for a Sustainable Europe Annual Review 2011-2012," *European Aggregates Association*, Brussel – Belgium, pp. 1-38, 2012.

[4] Langer, W.H., Drew, L.J. and Sachs, J.S, "Aggregate and the Environment," *American Geological Institute, U.S Geological Survey*, pp. 1-64, 2004. ISBN: 0-922152-71-3

[5] Dariati, T, "Landscape Restoration: Aggregate Mining in the Jeneberang River Basin, From Sky to Sea: Environment and Development in Sulawesi," *Department of Geography Publication Series, University of Waterloo*, pp. 505-525, 2005. ISBN 0-921083-72-6

[6] Rahardja, P., and Manurung, M, "Pengantar Ilmu Ekonomi (Mikroekonomi & Makroekonomi)," *Lembaga Penerbit Fakultas Ekonomi Universitas,* Indonesia, Jakarta, 2004.

[7] State of the Aggregate Resource in Ontario Study (SAROS), *Aggregate Consumption and Demand*, Queen's Printer, Ontario, pp. 1-220, 2009. ISBN 978-1-4435-379-9

[8] Anas, A.V., Suriamihardja, D.A., Pallu, S. and Irfan U.R, "Sustainability Performance Indicators of Sand and Gravel Mining in Jeneberang River South Sulawesi," *Proceedings of 2nd Asia Africa Mineral Resources Conference*, Bandung, Indonesia, 14 – 17 September, 2012.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 781-793

http://www.issr-journals.org/ijias/

Hybrid Distance Based Measures for Geospatial Domain

K. Saruladha and E. Thirumagal

Pondicherry Engineering College, Department of Computer Science and Engineering, Puducherry, India

ABSTRACT: Current researchers of search engines focus more on semantic based information retrieval as syntax based retrieval yield less precision. Retrieving relevant information from diverse heterogeneous web resources remains as a challenge. Distance based measures play a major role in the information retrieval systems. This work focuses on retrieving relevant concepts using geospatial datasources to aid geospatial applications in predicting floods, locating underground pipes and cables and testing the quality of water. Geospatial data characterizes geographical features of the real world using spatial extent and location. This paper proposes HDSM algorithm for geospatial information retrieval which adapts the existing distance based measures viz., Manhattan distance, Euclidean distance, Vector cosine and bray Curtis for the geospatial domain to identify related concepts to the geo-spatial query concept. All these four proposed hybrid distance based measures combine the advantages of geometric and network semantic similarity models. The meaning of the geospatial concepts are captured from the expressive knowledge of the geospatial concept properties and geospatial relations. These proposed four Hybrid distance based measures have been tested using Ordnance Survey Master Map data source and ordnance survey ontology for varying semantic similarity thresholds. The experimental results are reported in this paper. The Hybrid Manhattan distance based measure has yielded the precision of 95%.

KEYWORDS: Hybrid model, Geo-spatial relations, Properties, Semantic similarity, Conceptualisation, Distance Measures.

1 INTRODUCTION

The semantic information retrieval system's [1] goal is to predict the relevant geo-spatial concepts that are best suited for the geo-spatial query concept given by the user. The semantic similarity measure compares the geo-spatial concepts for determining the degree of relevance between them. Based on the relevance of the geo-spatial concepts, the semantic information retrieval task compares the synonyms of the geo-spatial query concepts and search for the geo-spatial information which are relevant to the geo-spatial query concept given by the user. The well thought-out representation of semantics is essential for the competent information retrieval. The semantic description must include the properties [2] i.e. the geo-spatial concepts' features and also the spatial relations [3] i.e. the relationship between the features of the geo-spatial concepts for representing the geo-spatial concepts.

Schwering [4] proposed the Hybrid semantic similarity computational model which uses the geo-spatial properties and the geo-spatial relations for computing the semantic similarity among the geo-spatial concepts. The other semantic similarity computation models except the Hybrid semantic similarity computational model either use geo-spatial properties or the geo-spatial spatial relations for representing the geo-spatial concepts. The Hybrid semantic similarity computational model uses the conceptual hulls for describing the concepts. The properties of the geo-spatial concepts are represented as the vector points in the conceptual hull [5]. For measuring the semantic similarity between the geo-spatial concepts, the Hybrid model compares the vector points between the geo-spatial concepts. The Euclidean distance method is applied between the matching vector points between the geo-spatial concepts. The Hybrid semantic similarity computational model uses the Euclidean semantic distance measure for computing the semantic distance between the geo-spatial concepts which also does not yield better precision and recall. If the semantic distance between the vector points are more, it indicates that the semantic similarity is less. If the semantic distance between the vector points are less, it indicates that the semantic similarity is more. Hence the various hybrid distance measures are used to test the information retrieval.

Corresponding Author: E. Thirumagal (tmagaleng@gmail.com)

The main objective of this paper is to propose various hybrid distance based measures such as Hybrid Manhattan distance measure, Hybrid Vector Cosine distance measure and Hybrid Bray Curtis distance measures for the Hybrid model of the geo-spatial domain and to predict the distance measure which yields good relevant information retrieval. The hybrid distance measures are evaluated using the performance metrics such as precision, recall and f-measure. The Hybrid distance measure which yields high precision is said to be best semantic similarity distance measure for the geo-spatial domain. The hybrid distance measures which are said above are not used for geo-spatial information retrieval, hence this paper attempts to use those distance measures for computing semantic similarity and to retrieve the relevant geo-spatial information.

The geo-spatial datasource used for experiments in this paper is Ordnance Survey Master Map [5] which has four layers such as Topography Layer, OSMasterMap Address Layer, Imagery Layer, and Integrated Transport Network (ITN) Layer provides the information about the topography, address of the location, aerial images and the vehicle movement tracking. Further this paper proposes the HDSM (Hybrid Distance Similarity Measure) algorithm which depicts how the above said hybrid distance measures are used in the geo-spatial domain. In section 2, the related works are depicted. The HDSM (Hybrid Distance Similarity Measure) algorithm is described in section 3. In section 4, the semantic similarity computation using various distance measures are described. In section 5, the experimental results and analysis are discussed. In section 6, the conclusion and the future enhancements are depicted.

2 RELATED WORKS

This section discusses the related works. Peter Gardenfors [6] introduced the design of conceptual space which has the geographical features of the geo-spatial concepts. The semantic similarity model that uses the conceptual space as the knowledge representation is called as the Geometric model. Based on the matching geographical features, the semantic similarity is computed called as the Feature model proposed by Rodriguez [7] and Egenhofer. Rada et al [8] proposed the semantic network which is used in Network model [9] considers only the spatial relations. Goldstone [10] introduced the Alignment model which considers the alignable features for computing the semantic similarity. The Transformational model uses the edit distance method which was introduced by Levenshtein [11] transforms one geo-spatial concept to another. Based on the number of transformations, the semantic similarity is computed. The above said semantic similarity models uses either geographical features or the spatial relations for representing the geo-spatial concepts and does not yield better information retrieval. Hence Schwering [4] introduced the Hybrid model whose semantic description has the geographical features and the spatial relations. The Hybrid model proposed by Schwering uses the Euclidean distance measure for computing the semantic distance which does not yield better precision, recall, f-measure and relevant information retrieval for most of the geo-spatial query concept. Hence various hybrid distance measure are tried for the Hybrid model.

The Euclidean distance measure [5], [12], Manhattan distance measure [12], [13] and the vector cosine method [12] are used for the Image Retrieval Application. The Image Retrieval Application is developed using five histograms for which the distance measures are applied for retrieving the relevant images. The five histograms used are RGB histogram [14], an HSV-based histogram [15], Jain and Vailaya's histogram [16], an HSV-based histogram with soft decision [17] and the histogram used in the QBIC system [18]. For this Image Retrieval Application the Manhattan distance measure yields good precision for all the five histograms. The Bray Cutis distance [19] measure is used in the clustering applications which provides better results. The next section discuss in detail about the HDSM algorithm.

3 HDSM ALGORITHM

The HDSM (Hybrid Distance Similarity Measure) algorithm is given below depicts how various semantic distance measures such as Hybrid Euclidean distance measure, Hybrid Manhattan distance measure, Hybrid Vector Cosine distance measure and Hybrid Bray Curtis distance measure are applied to the Hybrid model of the geo-spatial domain. The geo-spatial query concept is pre-processed. Then the dimensional and the relational semantic distance computation are done using various distance measures. The semantic distance computation is done using the dimensional and the relational semantic distance as the distance measures are applied to the Hybrid semantic similarity computation model. The semantically similar concepts are retrieved for which the documents are retrieved and the performance metrics such are precision, recall and f-measure are calculated.

Here α and β are the ratios which are used for the non-matching and matching dimensions and relations respectively. The α and β values ranges from 0 to 1. The various values of α and β are tested for computing semantic similarity and relevant information retrieval. When α value is 0.8 and β value is 0.2 it yields good results and it also yields good information retrieval. The prediction of the best semantic distance measure for the geo-spatial domain is made with the evaluation of the performance metrics such as precision, recall and f-measure. Precision is the ability of the geo-spatial information retrieval

system to retrieve more number of the relevant concepts from the retrieved concepts of the geo-spatial query concept. Recall is the ability of the geo-spatial information system to retrieve the concepts from the datasource. F-measure is the accuracy of the precision and the recall obtained for the geo-spatial query concept.

Algorithm 1: HDSM Algorithm

HDSM algorithm (QC, DS, SV, GO)
// DS is the Datasource, SV is the Shared Vocabulary and GO is Geo-spatial Ontology.
Input: Query Concept QC
Output: Retrieval of the related geo-spatial concepts.

Step 1: Pre-processing

Pre-processing of the geo-spatial Query concept QC.
Based on the dimensions available in GO, the related geo-spatial concepts from DS are retrieved.
The dimensional and relational quantitative data are obtained from SV.

Step 2: Dimensional semantic distance computation using various distance metrics.

For all the extracted relevant geo-spatial concepts from DS
For all the QC dimensions and the DS concept's dimensions
If the dimensions are matching

Compute $\text{Euclideandistance}_{dim} = \sqrt{(qd_1 - dd_1)^2 + ... + (qd_n - dd_n)^2}$ // qd is the quantitative dimensional value of the Query concept. dd is the quantitative dimensional value of the datasource concepts.

Compute $\text{Manhattandistance}_{dim} = |qd_1 - dd_1| + ... + |qd_n - dd_n|$

Compute $\text{VectorCosinedistance}_{dim} = \frac{(qd_1 \times dd_1) + ... + (qd_n \times dd_n)}{\sqrt{(qd_1)^2 + ... + (qd_n)^2} \times \sqrt{(dd_1)^2 + ... + (dd_n)^2}}$

Compute $\text{BrayCurtisdistance}_{dim} = \frac{|qd_1 - dd_1| + ... + |qd_n - dd_n|}{|qd_1 + dd_1| + ... + |qd_n + dd_n|}$

Else compute standardised distance// for the non-matching dimensions, the concept that have dimensions are aggregated.

End if

$\text{Semdist}_{Euclid dim} = \alpha * \text{standarddistance} + \beta * \text{Euclideandistance}_{dim}$

$\text{Semdist}_{Man dim} = \alpha * \text{standarddistance} + \beta * \text{Manhattandistance}_{dim}$

$\text{Semdist}_{Vector dim} = \alpha * \text{standarddistance} + \beta * \text{VectorCosinedistance}_{dim}$

$\text{Semdist}_{Bray dim} = \alpha * \text{standarddistance} + \beta * \text{BrayCurtisdistance}_{dim}$

End for (all dimensions).

Step 3: Relational semantic distance computation using various distance metrics.

For all the extracted relevant geo-spatial concepts from DS
For all the QC relations and the DS concept's relations
If the relations are matching

Compute $\text{Euclideandistance}_{rel} = \sqrt{(qr_1 - dr_1)^2 + ... + (qr_n - dr_n)^2}$ // qr is the quantitative relational value of the Query concept. dr is the quantitative relational value of the datasource concepts.

Compute $\text{Manhattandistance}_{rel} |qr_1 - dr_1| + ... + |qr_n - dr_n|$

Compute $\text{VectorCosinedistance}_{rel} \frac{(qr_1 \times dr_1) + ... + (qr_n \times dr_n)}{\sqrt{(qr_1)^2 + ... + (qr_n)^2} \times \sqrt{(dr_1)^2 + ... + (dr_n)^2}}$

Compute $\text{BrayCurtis distance}_{rel} = \frac{|qr_1 - dr_1| + ... + |qr_n - dr_n|}{|qr_1 + dr_1| + ... + |qr_n + dr_n|}$

Else compute standardised distance.

End if

$\text{Semdist}_{Euclidrel} = \alpha * \text{standarddistance} + \beta * \text{Euclideandistance}_{rel}$

$\text{Semdist}_{Manrel} = \alpha * \text{standarddistance} + \beta * \text{Manhattandistance}_{rel}$

$\text{Semdist}_{Vectorrel} = \alpha * \text{standarddistance} + \beta * \text{Vectordistance}_{rel}$

$\text{Semdist}_{Brayrel} = \alpha * \text{standard distance} + \beta * \text{BrayCurtis distance}_{rel}$

End for (all relations).

End for (all extracted relevant concepts).

Step 4: Semantic distance computation

//DSC is Datasource Concept

$$Semdist_{Euclid}(QC,DSC) = \sqrt{[Semdist_{Euclid\,dim} - Semdist_{Euclidrel}]^2}$$

$$Semdist_{Man}(QC,DSC) = |Semdist_{Mandim} - Semdist_{Manrel}|$$

$$Semdist_{Vector}(QC,DSC) = \frac{Semdist_{Vector\,dim} + Semdist_{Vectorrel}}{\sqrt{[Semdist_{Vector\,dim}]^2} \times \sqrt{[Semdist_{Vectorrel}]^2}}$$

$$Semdist_{Bray}(QC,DSC) = \frac{|[Semdist_{Bray\,dim}] - [Semdist_{Brayrel}]|}{|[Semdist_{Bray\,dim}] + [Semdist_{Brayrel}]|}$$

//Less the semantic distance, more the semantic similarity. More the semantic distance, less is the semantic similarity.

The relevant semantically similar concepts are retrieved.

Step 5: Retrieval of documents for the semantically similar concepts.

//Rlc is relevant concepts and Rtc is retrieved concepts.

Compute precision, recall and f-measure where

precision=|Rlc ∩ Rtc|/ Rtc,

recall = |Rlc ∩ Rtc|/Rlc and

f-measure = (2*(precision*recall))/(precision + recall).

This section discussed in detail about the HDSM algorithm and also depicts how the various distance measures such as Hybrid Euclidean distance measure, Hybrid Manhattan distance measure, Hybrid Vector Cosine distance measure and Hybrid Bray Curtis distance measure are applied to the Hybrid Model of the geo-spatial information system. The next section discuss about the computation of each distance measure for the sample geo-spatial hydrological ontology and the retrieval of the relevant geo-spatial concepts for the specified geo-spatial query concept.

4 SEMANTIC SIMILARITY COMPUTATION USING VARIOUS DISTANCE MEASURES FOR THE HYBRID MODEL

This section discuss about the Hybrid semantic similarity computational model and the application of various semantic distance measures such as Hybrid Euclidean distance measure, Hybrid Manhattan distance measure, Hybrid Vector Cosine distance measure and Hybrid Bray Curtis distance measure for the Hybrid model of the geo-spatial domain and also depicts the sample calculation of each distance measure.

4.1 EXISTING HYBRID SEMANTIC SIMILARITY COMPUTATIONAL MODEL

Gardenfors introduced the design of the conceptual space which provides the geo-spatial concepts and its corresponding features. Rada et al proposed the semantic networks which uses the relationship between the geo-spatial concepts. The Hybrid model proposed by Schwering combined the work of Gardenfors and Rada et al and this model uses the Euclidean distance measure for computing the semantic distance among the geo-spatial concepts which does not yield better precision. The conceptual space and the semantic network are used for computing the semantic similarity in the Hybrid model. In this Model the dimensions or the properties are considered as the vector points in the conceptual hull.

The Euclidean semantic distance measure is used for computing the semantic distance between the matching vector points of the geo-spatial query concept and the relevant concepts. In the Fig. 1 the query concept hull is watercourse and the related concept hull is channel. q_1, q_2 and q_3 are the matching dimensional vector points between the geo-spatial query concept and the related geo-spatial concept. d_1, d_2 and d_3 are the semantic distances between the dimensions. The other distance measures are not attempted for the geo-spatial domain for computing the semantic distance. Hence this paper attempts various semantic distance measures for the Hybrid model. The next section discuss about the semantic distance measures.

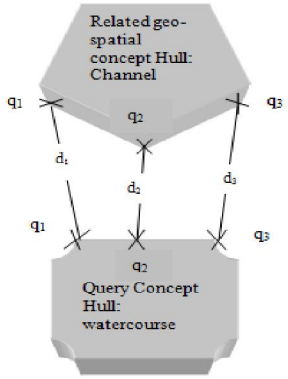


Fig. 1. Hybrid Model Conceptual Hull

4.2 PROPOSED SEMANTIC DISTANCE MEASURES

The proposed semantic distance measures for Hybrid model are

- Hybrid Euclidean distance measure
- Hybrid Manhattan distance measure
- Hybrid Vector Cosine distance measure and
- Hybrid Bray Curtis distance measure

These semantic distance measures are used for computing the semantic distance among the geo-spatial concepts. These semantic distance measures are the distance measures which are used for calculating the distance between two points in the n-dimensional space. The geo-spatial ontology [20] is needed for providing the semantics of the geo-spatial domain. The geo-spatial ontology used for experiments in this paper is Hydrology.owl which is the Ordnance Survey ontology. The datasource used for experiments in this paper is Ordnance Survey Master Map. The Ordnance Survey Master Map has four layers such as Topography Layer, OSMasterMap Address Layer, Imagery Layer, and Integrated Transport Network (ITN) Layer provides the information about the topography, address of the location, aerial images and the vehicle movement tracking. For viewing the Ordnance Survey Master Map, the Geographic Mark-up language viewer is used. The shared vocabulary converts the qualitative dimensional value into quantitative dimensional value. The sample ontology is given by Fig.2, for which the various semantic distance measures are applied and the sample calculation are done and they are tested for the information retrieval and the prediction of the best semantic distance measure for the Hybrid model of the geo-spatial domain is done. Here 'pond' is the geo-spatial query concept considered. Fig.2 shows the related geo-spatial concepts of the query concept pond.

For computing the semantic distance between the geo-spatial query concept and its related concepts, the dimensional values of the geo-spatial concepts are needed which is given by the shared vocabulary. Totally 21 dimensions such as elevation, gradient, gauge level, altitude difference, height relative to the environment, soil moisture, salinity, number of days waterlogged, etc,. Here for the sample calculation three dimensions are considered for which the dimensional values are given by the Table 1.

Table 1. Geo-spatial concepts and its dimensional values.

Geo-spatial Concepts	Dimensions		
	Elevation(meters)	Gradient (%)	Altitude Difference(meters)
River	610	79	80
Pond	600	78	56
Stream	760	84	78
Nant	670	67	89
Sea	800	77	90
watercourse	277	73	87
Lake	345	78	60
Dam	890	75	67

The every semantic distance measure is applied to the sample tree shown in the Fig.2 and for the dimensional values given by the Table 1. The next sub-sections discuss the distance measures and its application in the geo-spatial domain.

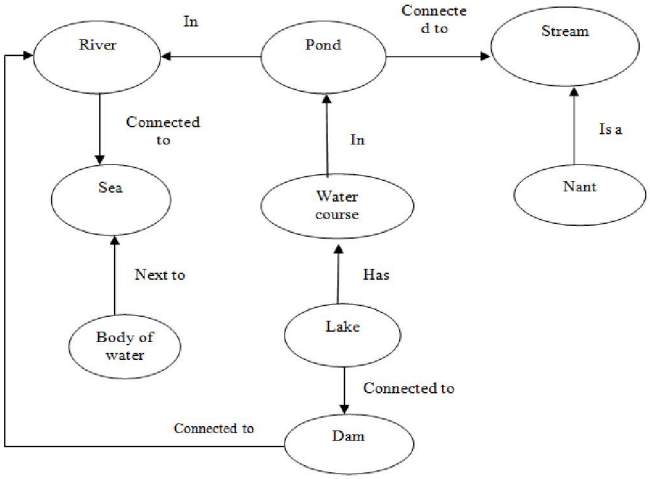


Fig. 2. Sample tree containing the related concepts of the query concept 'Pond'

4.2.1 HYBRID EUCLIDEAN DISTANCE MEASURE

This section discuss about the Hybrid Euclidean distance measure and its application in the geo-spatial domain. The Euclidean distance method is used for computing the distance between two points in n-dimensional space. The general Euclidean distance is given by equation (1).

$$\text{Euclideandistance}(p,q)=\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+....+(p_n-q_n)^2} \quad (1)$$

where $p_1, p_2, ..., p_n$ and $q_1, q_2, ..., q_n$ are the points in the n-dimensional space. This general equation of the Euclidean distance is mapped to the geo-spatial domain which is given by the equations (2), (3) and (4). As said above this distance measure is applied for the Hybrid model which will be considering both the properties or the dimensions and the spatial relations for representing the geo-spatial concepts. For the geo-spatial query concept given, the related geo-spatial concepts are obtained from the datasource based on the semantics in the geo-spatial ontology. The dimensional values of the needed geo-spatial concepts are obtained from the shared vocabulary. For all the matching and the non-matching dimensions of the query concept and the related geo-spatial concept, the dimensional semantic distance is computed which is given by equation (2) and (3).

$$\text{Euclidean distance}_{\text{dim}} = \sqrt{(qd_1-dd_1)^2+.....+(qd_n-dd_n)^2} \quad (2)$$

where qd is the quantitative dimensional value of the Query concept. dd is the quantitative dimensional value of the datasource concepts.

$$\text{Semdist}_{\text{Eucliddim}} = \alpha * \text{standarddistance} + \beta * \text{Euclideandistance}_{\text{dim}} \quad (3)$$

where standardised distance is for the non-matching dimensions, the geo-spatial concepts that have dimensions are aggregated. The α and β are the ratios of the matching and the non matching dimensions. The α value is 0.8 multiplied to the non-matching dimensions and β value is 0.2 multiplied to the matching dimensions. Likewise for the matching and the non-matching relations, the relational semantic distance is computed which is given by the equation (4).

$$\text{Semdist}_{\text{Euclidrel}} = \alpha * \text{standarddistance} + \beta * \text{Euclideandistance}_{\text{rel}} \quad (4)$$

The semantic distance is computed by aggregating the dimensional and the relational semantic distance is given by equation (5).

$$\text{Semdist}_{\text{Euclid}}(QC,DSC)=\sqrt{[\text{Semdist}_{\text{Eucliddim}}-\text{Semdist}_{\text{Euclidrel}}]^2} \quad (5)$$

where QC is the Query Concept and DSC is the datasource concept.

Computation using Hybrid Euclidean distance method

Consider the sample tree shown in the Fig.2 and the dimensional values given in the Table 1 for the geo-spatial concepts. Pond is the query concept considered here. The related concept to the query concept pond is watercourse. So the dimensional semantic distance between pond and watercourse is calculated using equation (2) and (3). The α and β values took are 0.8 and 0.2 respectively. Hence,

$$\text{Semdist}_{\text{Eucliddim}} = 0.8 * 0 + 0.2 * \sqrt{(600-277)^2 + (78-73)^2 + (56-87)^2} = 66.55$$

The geo-spatial concept watercourse has its relation with the lake so dimensions of the watercourse and lake are considered.

$$\text{Semdist}_{\text{Euclidrel}} = 0.8 * 0 + 0.2 * \sqrt{(277-345)^2 + (73-78)^2 + (87-60)^2} = 14.66$$

The semantic distance computation is done by aggregating the dimensional and the relational semantic distance using equation (5).

$$\text{Semdist}_{\text{Euclid}}(\text{pond}, \text{watercourse}) = \sqrt{(66.55-14.66)^2} = 5.189$$

Here all the dimensions are matching and so the standardised distance is 0. The semantic distance between pond and watercourse is 5.189 when three dimensions are considered. For the experiments conducted in this paper 21 dimensions are considered. The query concept pond is connected to stream so the semantic distance between the pond and stream is computed.

$$\text{Semdist}_{\text{Eucliddim}} = 0.8 * 0 + 0.2 * \sqrt{(600-760)^2 + (78-84)^2 + (56-78)^2} = 32.32$$

The geo-spatial concept stream has its relation with the Nant so dimensions of the stream and nant are considered.

$$\text{Semdist}_{\text{Euclidrel}} = 0.8 * 0 + 0.2 * \sqrt{(760-670)^2 + (84-67)^2 + (78-89)^2} = 18.44$$

The semantic distance computation is done by aggregating the dimensional and the relational semantic distance using equation (5).

$$\text{Semdist}_{\text{Euclid}}(\text{pond}, \text{stream}) = \sqrt{(32.32-18.44)^2} = 1.388$$

The semantic distance between pond and stream is 1.388. The semantic distance between pond and watercourse is 5.189. Thus the geo-spatial concept stream is more similar to the query concept pond. Likewise the semantic distance is computed for all the related concepts of the geo-spatial query concept. Then the performance metrics such as precision, recall and f-measure are computed for evaluation. The next sub-section discuss about the Manhattan distance measure.

4.2.2 HYBRID MANHATTAN DISTANCE MEASURE

This section discuss about how the Hybrid Manhattan distance measure is used for computing the semantic similarity between the geo-spatial concepts. The general Manhattan distance is given by the equation (6).

$$\text{Manhattan distance } (p,q) = |p_1 - q_1| + |p_2 - q_2| + \ldots\ldots\ldots + |p_n - q_n| \quad (6)$$

where $p_1, p_2, \ldots p_n$ and $q_1, q_2, \ldots\ldots q_n$ are the points in the n-dimensional space. The mapping of the general Manhattan distance to the geo-spatial domain is given by equations (7), (8), (9) and (10). The matching and the non-matching dimensions are considered. For the matching dimensions the Manhattan distance is used for computation which is given by equation (7).

$$\text{Compute } \text{Manhattandistance}_{\text{dim}} = |qd_1 - dd_1| + \ldots\ldots + |qd_n - dd_n| \quad (7)$$

where qd is the quantitative dimensional value of the Query concept. dd is the quantitative dimensional value of the datasource concepts.

If the dimensions are not matching then the standardised distance is considered. The dimensional semantic distance is given by equation (8).

$$Semdist_{Man\,dim} = \alpha * standard\,distance + \beta * Manhattan\,distance_{dim} \quad (8)$$

Likewise the relational semantic distance is computed and given by the equation (9).

$$Semdist_{Manrel} = \alpha * standard\,distance + \beta * Manhattan\,distance_{rel} \quad (9)$$

The Hybrid computational model aggregates the dimensional and the relational distance which is given by equation (10).

$$Semdist_{Man}(QC, DSC) = |Semdist_{Man\,dim} - Semdist_{Manrel}| \quad (10)$$

where QC is the Query Concept and DSC is the datasource concept.

Computation using Hybrid Manhattan distance method

For the semantic similarity computation, Fig.2 and Table 1 are considered for the sample tree consisting of the geo-spatial concepts and the dimensional values of the geo-spatial concepts. The query concept pond is considered for which the related geo-spatial concept is watercourse and so the semantic distance between pond and watercourse must be computed. The Manhattan dimensional semantic distance is computed using equations (7) and (8).

$Semdist_{Man\,dim}$ = 0.8*0+0.2*[|600 – 277|+|78 – 73|+|56 - 87| = 71.8

The geo-spatial concept watercourse has its relation with the lake so dimensions of the watercourse and lake are considered. The relational semantic distance is computed using equation (9).

$Semdist_{Manrel}$ = 0.8*0+0.2*[|277 - 345|+|73 - 78|+|87 - 60| = 20

The dimensional and the relational semantic distance are aggregated in the Hybrid semantic similarity computational model for computing the semantic distance which is given by the equation (10).

$Semdist_{Man}$ (Pond, watercourse) = |71.8-20| = 51.8/10=5.18

The query concept pond is connected to stream, so the semantic distance between the pond and stream is computed.

$Semdist_{Man\,dim}$ = 0.8*0+0.2*[|600 – 760|+|78 – 84|+|56 - 78| = 37.6

The geo-spatial concept stream has its relation with the nant. So dimensions of the stream and nant are considered.

$Semdist_{Manrel}$ = 0.8*0+0.2*[|760 - 670|+|84 - 67|+|78 - 89| = 23.4

$Semdist_{Man}$ (Pond, stream) = |37.6 – 23.4| = 14.2/10 = 1.42

The semantic distance between pond and stream is 1.42. The semantic distance between pond and watercourse is 5.18. Thus the geo-spatial concept stream is more similar to the query concept pond. Likewise the semantic distance is computed for all the related concepts of the geo-spatial query concept. Then the precision, recall and f-measure are computed. The next sub-section discuss about the Vector cosine distance measure.

4.2.3 HYBRID VECTOR COSINE DISTANCE MEASURE

This section discusses how the Vector Cosine distance can be applied to the geo-spatial domain. The general vector cosine distance is given by equation (11).

$$\text{Vector Cosine distance}(p, q) = \frac{(p_1 \times q_1) + (p_2 \times q_2) + \ldots\ldots + (p_n \times q_n)}{\sqrt{(p_1)^2 + \ldots + (p_n)^2} \times \sqrt{(q_1)^2 + \ldots + (q_n)^2}} \quad (11)$$

where $p_1, p_2, \ldots p_n$ and $q_1, q_2, \ldots\ldots q_n$ are the points in the n-dimensional space.

This general Vector Cosine distance mapped to the geo-spatial information retrieval system which is given by the below equations (12), (13), (14) and (15).

The matching and the non-matching dimensions are considered. For the matching dimensions the Vector Cosine distance is used for computation which is given by equation (12).

$$\text{VectorCosinedistance}_{\text{dim}} = \frac{(qd_1 \times dd_1) + ... + (qd_n \times dd_n)}{\sqrt{(qd_1)^2 + ... + (qd_n)^2} \times \sqrt{(dd_1)^2 + ... + (dd_n)^2}} \quad (12)$$

where qd is the quantitative dimensional value of the Query concept. dd is the quantitative dimensional value of the datasource concepts.

If the dimensions are not matching then the standardised distance is considered. The dimensional semantic distance is given by equation (13).

$$\text{Semdist}_{\text{Vectordim}} = \alpha * \text{standarddistance} + \beta * \text{VectorCosinedistance}_{\text{dim}} \quad (13)$$

The relational semantic distance is computed and given by the equation (14).

$$\text{Semdist}_{\text{Vectorrel}} = \alpha * \text{standarddistance} + \beta * \text{Vectordistance}_{\text{rel}} \quad (14)$$

The Hybrid Model combines the dimension and the relation which is given by equation (15).

$$Semdist_{Vector}(QC, DSC) = \frac{Semdist_{Vector\,\text{dim}} \times Semdist_{Vectorrel}}{\sqrt{[Semdist_{Vector\,\text{dim}}]^2} \times \sqrt{[Semdist_{Vectorrel}]^2}} \quad (15)$$

where QC is the Query Concept and DSC is the datasource concept.

Computation using Hybrid Vector Cosine distance method

For the semantic similarity computation, Fig.2 and Table 1 are considered for the sample tree consisting of the geo-spatial concepts and the dimensional values of the geo-spatial concepts. The query concept pond is considered for which the related geo-spatial concept is watercourse and so the semantic distance between pond and watercourse must be computed. The Vector Cosine dimensional semantic distance is computed using equations (12) and (13).

$$\text{Semdist}_{\text{Vectordim}} = 0.8 * 0 + 0.2 * \left[\frac{(600 \times 277) + (78 \times 73) + (56 \times 87)}{\sqrt{(600)^2 + (78)^2 + (56)^2} \times \sqrt{(277)^2 + (73)^2 + (87)^2}} \right] = 0.194$$

The geo-spatial concept watercourse has its relation with the lake so dimensions of the watercourse and lake are considered. The relational semantic distance is computed using equation (14).

$$\text{Semdist}_{\text{Vectorrel}} = 0.8 * 0 + 0.2 * \left[\frac{(277 \times 345) + (73 \times 78) + (87 \times 60)}{\sqrt{(277)^2 + (73)^2 + (87)^2} \times \sqrt{(345)^2 + (78)^2 + (60)^2}} \right] = 0.99.$$

The semantic distance aggregation is done using the equation (15).

$$\text{Semdist}_{\text{Vector}}(\text{pond}, \text{watercourse}) = \frac{0.194 + 0.99}{\sqrt{(0.194)^2} \times \sqrt{(0.99)^2}} = 6.16$$

The query concept pond is connected to stream, so the semantic distance between the pond and stream is computed.

$$\text{Semdist}_{\text{Vectordim}} = 0.8 * 0 + 0.2 * \left[\frac{(600 \times 760) + (78 \times 84) + (56 \times 78)}{\sqrt{(600)^2 + (78)^2 + (56)^2} \times \sqrt{(760)^2 + (84)^2 + (78)^2}} \right] = 0.199$$

The geo-spatial concept stream has its relation with the nant. So dimensions of the stream and nant are considered.

$$\text{Semdist}_{\text{Vectorrel}} = 0.8 * 0 + 0.2 * \left[\frac{(760 \times 670) + (84 \times 67) + (78 \times 89)}{\sqrt{(760)^2 + (84)^2 + (78)^2} \times \sqrt{(670)^2 + (67)^2 + (89)^2}} \right] = 0.199$$

$$\text{Semdist}_{\text{Vector}}(\text{pond}, \text{stream}) = \frac{0.199 + 0.199}{\sqrt{(0.199)^2} \times \sqrt{(0.199)^2}} = 1.02$$

The semantic distance between pond and watercourse is 6.16 and the semantic distance between pond and stream is 1.02. The stream has the less semantic distance and so stream is more relevant concept to the given geo-spatial query

concept pond. Then the precision, recall and f-measure are computed. The next sub-section discuss about the Bray Curtis distance measure.

4.2.4 HYBRID BRAY CURTIS DISTANCE MEASURE

This section discusses how the Bray Curtis distance can be incorporated into the geo-spatial domain. The general Bray Curtis distance is given by equation (16).

$$\text{Bray Curtis distance}(p,q) = \frac{|p_1 - q_1| + |p_2 - q_2| + \ldots\ldots + |p_n - q_n|}{|p_1 + q_1| + |p_2 + q_2| + \ldots\ldots + |p_n + q_n|} \quad (16)$$

where $p_1,p_2,\ldots.p_n$ and $q_1,q_2,\ldots\ldots q_n$ are the points in the n-dimensional space. The mapping of the general Bray Curtis distance to the geo-spatial domain is given by the equations (17), (18), (19) and (20). The matching and the non-matching dimensions are considered. For the matching dimensions the Bray Curtis distance is used for computation which is given by equation (17).

$$\text{BrayCurtisdistance}_{dim} = \frac{|qd_1 - dd_1| + \ldots\ldots + |qd_n - dd_n|}{|qd_1 + dd_1| + \ldots + |qd_n + dd_n|} \quad (17)$$

where qd is the quantitative dimensional value of the Query concept. dd is the quantitative dimensional value of the datasource concepts.

If the dimensions are not matching then the standardised distance is considered. The dimensional semantic distance is given by equation (18).

$$\text{Semdist}_{Braydim} = \alpha * \text{standarddistance} + \beta * \text{BrayCurtisdistance}_{dim} \quad (18)$$

The relational semantic distance is computed and given by the equation (19).

$$\text{BrayCurtis distance}_{rel} = \frac{|qr_1 - dr_1| + \ldots\ldots + |qr_n - dr_n|}{|qr_1 + dr_1| + \ldots + |qr_n + dr_n|} \quad (19)$$

The Hybrid computational model aggregates the dimensional and the relational distance which is given by equation (20).

$$\text{Semdist}_{Bray}(QC, DSC) = \frac{|[\text{Semdist}_{Bray\,dim}] - [\text{Semdist}_{Brayrel}]|}{|[\text{Semdist}_{Bray\,dim}] + [\text{Semdist}_{Brayrel}]|} \quad (20)$$

where QC is the Query Concept and DSC is the datasource concept.

Computation using Hybrid Bray Curtis distance method

For the semantic similarity computation, Fig.2 and Table 1 are considered for the sample tree consisting of the geo-spatial concepts and the dimensional values of the geo-spatial concepts. The query concept pond is considered for which the related geo-spatial concept is watercourse and so the semantic distance between pond and watercourse must be computed. The Bray Curtis dimensional semantic distance is computed using equations (17) and (18).

$$\text{Semdist}_{Braydim} = 0.8 * 0 + 0.2 * \left[\frac{|600 - 277| + |78 - 73| + |56 - 87|}{|600 + 277| + |78 + 73| + |56 + 87|}\right] = 0.06.$$

The geo-spatial concept watercourse has its relation with the lake so dimensions of the watercourse and lake are considered. The relational semantic distance is computed using equation (19).

$$\text{Semdist}_{Brayrel} = 0.8 * 0 + 0.2 * \left[\frac{|277 - 345| + |73 - 78| + |87 - 60|}{|277 + 345| + |78 + 73| + |60 + 87|}\right] = 0.0217$$

$\text{Semdist}_{Bray}(\text{pond, watercourse}) = \frac{|0.06 - 0.0217|}{|0.06 + 0.0217|} = 0.46*10=4.6$. The query concept pond is connected to stream, so the semantic distance between the pond and stream is computed.

$$\text{Semdist}_{Braydim} = 0.8 * 0 + 0.2 * \left[\frac{|600 - 760| + |78 - 84| + |56 - 78|}{|600 + 760| + |78 + 84| + |56 + 78|}\right] = 0.022.$$

The geo-spatial concept stream has its relation with the nant. So dimensions of the stream and nant are considered.

$$Semdist_{Brayrel} = 0.8 * 0 + 0.2 * \left[\frac{|760 - 670| + |84 - 67| + |78 - 89|}{|760 + 670| + |84 + 67| + |78 + 89|} \right] = 0.0135$$

$$Semdist_{Bray}(pond, stream) = \frac{|0.022 - 0.0135|}{|0.022 + 0.0135|} = 0.23*10=2.3$$

The semantic distance between pond and watercourse is 4.6 and the semantic distance between pond and stream is 2.3. The stream has the less semantic distance and so stream is more relevant concept to the given geo-spatial query concept pond. All the distance measure depicts that stream is more relevant than the watercourse but the semantic distance varies for every semantic distance measure. Based on the semantic distance, the precision, recall and the f-measure values are computed. The next section discusses the experimental results and analysis.

5 EXPERIMENTAL RESULTS AND ANALYSIS

This section discuss in detail about the experimental results and analysis. The experiments are conducted using Netbeans 7.2.1 IDE. The geo-spatial datasource and the ontology are used for capturing the knowledge about the domain. Ordnance Survey Master Map is the datasource used for conducting experiments. The Geographic Mark-up Language (GML) viewer is used for viewing the Ordnance Survey Master Map. For conducting experiments 108 geo-spatial concepts, 20 relations and 21 dimensions or properties are considered. The various distance based measures are applied to the Hybrid model. The performance metrics used are precision, recall and f-measure. The performance metric values for 10 sample relevant geo-spatial concepts for the geo-spatial query concept using various semantic distance measures are shown in Table 2. The precision, recall and the f-measure values are computed for various similarity thresholds. Fig.3 depicts that the Hybrid Manhattan distance measure has yielded the maximum precision of 0.95 at the maximum similarity threshold 5. Then the Hybrid Euclidean distance measure has yielded the precision of 0.8 at the maximum similarity threshold likewise the Vector Cosine distance measure and Bray Curtis similarity measure yields the precision of 0.70 and 0.77 at the maximum similarity threshold. The Fig.4 depicts that except the Hybrid Manhattan distance measure, other hybrid distance measures mostly yield the recall of 1 for varying similarity threshold. Hybrid Vector cosine distance measure yields the maximum recall. The f-measure is computed which is maximum for the Hybrid Manhattan distance measure. Hence from the Figure 3, 4 and 5 it is depicted that the Hybrid Manhattan distance measure yields good precision, recall, f-measure and the relevant information retrieval as the Hybrid Manhattan distance measure retrieves all the relevant neighbour geo-spatial concepts.

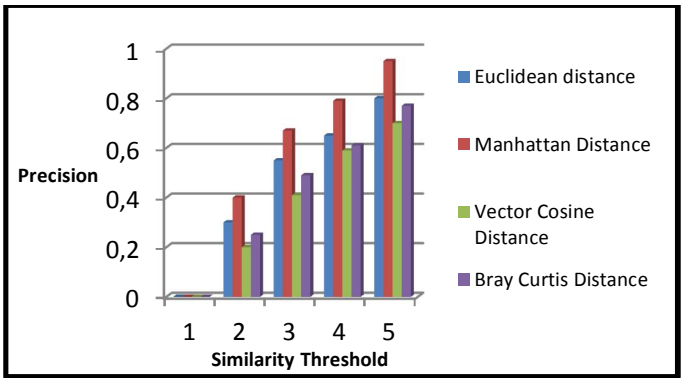


Fig. 3. Average precision yielded by each distance measure

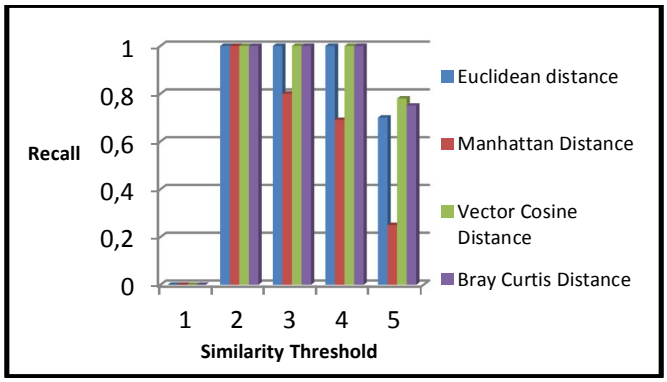


Fig. 4. Average recall yielded by ach distance measure

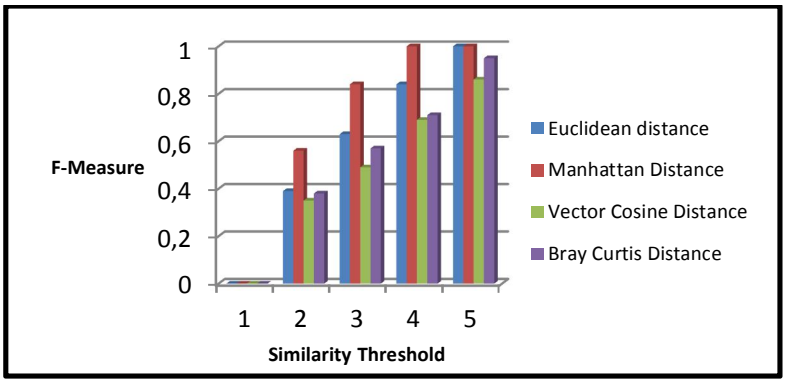


Fig. 5. Average f-measure yielded by each semantic distance measure

Table 2. Semantic distance measures and its performance metrics for the query concept 'Pond'

Semantic Distance Measure	Geo-spatial concepts	Precision				Recall				F-Measure			
		Similarity Thresholds				Similarity Thresholds				Similarity Thresholds			
		2	3	4	5	2	3	4	5	2	3	4	5
Euclidean Measure	Floodplain	0.21	0.39	0.52	0.63	1.00	1.00	1.00	0.77	0.35	0.52	0.69	0.87
	River	0.19	0.46	0.63	0.79	1.00	1.00	1.00	0.69	0.32	0.48	0.64	0.80
	Canal	0.22	0.45	0.55	0.8	1.00	1.00	1.00	0.69	0.36	0.54	0.72	0.90
	Sea	0.26	0.5	0.56	0.81	1.00	1.00	1.00	0.69	0.41	0.62	0.83	1.03
	Lake	0.31	0.48	0.51	0.65	1.00	1.00	1.00	0.76	0.47	0.71	0.95	1.18
	Water course	0.29	0.47	0.49	0.59	1.00	1.00	1.00	0.9	0.45	0.67	0.90	1.12
	Dam	0.23	0.42	0.6	0.75	1.00	1.00	1.00	0.68	0.37	0.56	0.75	0.93
	Channel	0.32	0.47	0.59	0.81	1.00	1.00	1.00	0.69	0.48	0.73	0.97	1.21
	Stream	0.23	0.53	0.62	0.82	1.00	1.00	1.00	0.62	0.37	0.56	0.75	0.93
	Backwater	0.27	0.39	0.51	0.63	1.00	1.00	1.00	0.76	0.43	0.64	0.85	1.06
	Pool	0.22	0.51	0.68	0.83	1.00	1.00	1.00	0.59	0.36	0.54	0.72	0.90
	Reservoir	0.29	0.54	0.65	0.81	1.00	1.00	1.00	0.69	0.45	0.67	0.90	1.12
Manhattan Distance	Floodplain	0.39	0.58	0.75	0.92	1.00	0.81	0.71	0.17	0.56	0.84	1.12	1.40
	River	0.28	0.49	0.6	0.9	1.00	0.79	0.6	0.2	0.44	0.66	0.88	1.09
	Canal	0.3	0.5	0.6	0.91	1.00	0.82	0.72	0.36	0.46	0.69	0.92	1.15
	Sea	0.38	0.55	0.65	0.9	1.00	0.79	0.73	0.2	0.55	0.83	1.10	1.38
	Lake	0.42	0.62	0.8	0.92	1.00	0.8	0.65	0.18	0.59	0.89	1.18	1.48
	Water course	0.41	0.62	0.8	0.92	1.00	0.81	0.73	0.18	0.58	0.87	1.16	1.45
	Dam	0.39	0.6	0.79	0.9	1.00	0.81	0.71	0.2	0.56	0.84	1.12	1.40
	Channel	0.49	0.55	0.65	0.85	1.00	0.79	0.64	0.37	0.66	0.99	1.32	1.64
	Stream	0.39	0.58	0.79	0.9	1.00	0.8	0.73	0.2	0.56	0.84	1.12	1.40
	Backwater	0.43	0.63	0.81	0.91	1.00	0.84	0.71	0.19	0.60	0.90	1.20	1.50
	Pool	0.37	0.53	0.65	0.87	1.00	0.8	0.64	0.36	0.54	0.81	1.08	1.35
	Reservoir	0.39	0.58	0.68	0.88	1.00	0.81	0.75	0.36	0.56	0.84	1.12	1.40
Vector Cosine Distance Measure	Floodplain	0.19	0.3	0.48	0.65	1.00	1.00	1.00	0.8	0.32	0.48	0.64	0.80
	River	0.18	0.4	0.59	0.7	1.00	1.00	1.00	0.71	0.31	0.46	0.61	0.76
	Canal	0.19	0.4	0.5	0.69	1.00	1.00	1.00	0.71	0.32	0.48	0.64	0.80
	Sea	0.2	0.49	0.53	0.7	1.00	1.00	1.00	0.71	0.33	0.50	0.67	0.83
	Lake	0.25	0.45	0.52	0.7	1.00	1.00	1.00	0.8	0.4	0.6	0.8	1.00
	Water course	0.2	0.43	0.41	0.68	1.00	1.00	1.00	1.00	0.33	0.5	0.67	0.83
	Dam	0.19	0.39	0.5	0.72	1.00	1.00	1.00	0.7	0.32	0.48	0.64	0.80
	Channel	0.25	0.44	0.49	0.7	1.00	1.00	1.00	0.71	0.4	0.6	0.8	1
	Stream	0.19	0.5	0.6	0.7	1.00	1.00	1.00	0.8	0.32	0.48	0.64	0.80
	Backwater	0.2	0.35	0.49	0.65	1.00	1.00	1.00	0.82	0.33	0.50	0.67	0.83
	Pool	0.2	0.45	0.6	0.71	1.00	1.00	1.00	0.62	0.33	0.50	0.67	0.83
	Reservoir	0.25	0.5	0.6	0.71	1.00	1.00	1.00	0.75	0.40	0.60	0.80	1.00
Bray Curtis Distance Measure	Floodplain	0.2	0.35	0.5	0.71	1.00	1.00	1.00	0.78	0.33	0.50	0.67	0.83
	River	0.19	0.43	0.6	0.81	1.00	1.00	1.00	0.7	0.32	0.48	0.64	0.80
	Canal	0.2	0.42	0.54	0.7	1.00	1.00	1.00	0.7	0.33	0.50	0.67	0.83
	Sea	0.24	0.49	0.56	0.72	1.00	1.00	1.00	0.7	0.39	0.58	0.77	0.97
	Lake	0.3	0.46	0.52	0.75	1.00	1.00	1.00	0.78	0.46	0.69	0.92	1.15
	Water course	0.25	0.45	0.45	0.7	1.00	1.00	1.00	0.88	0.4	0.6	0.8	1
	Dam	0.22	0.4	0.58	0.75	1.00	1.00	1.00	0.69	0.36	0.54	0.72	0.90
	Channel	0.29	0.46	0.55	0.72	1.00	1.00	1.00	0.71	0.45	0.67	0.90	1.12
	Stream	0.2	0.52	0.61	0.73	1.00	1.00	1.00	0.7	0.33	0.50	0.67	0.83
	Backwater	0.24	0.37	0.5	0.72	1.00	1.00	1.00	0.8	0.39	0.58	0.77	0.97
	Pool	0.21	0.49	0.65	0.79	1.00	1.00	1.00	0.61	0.35	0.52	0.69	0.87
	Reservoir	0.27	0.52	0.64	0.79	1.00	1.00	1.00	0.72	0.43	0.64	0.85	1.06

6 CONCLUSION

Thus this paper has discussed HDSM algorithm which includes various Hybrid semantic distance measures such as Hybrid Euclidean distance measure, Hybrid Manhattan distance measure, Hybrid Vector Cosine distance measure and Hybrid Bray Cutis distance measure for the Hybrid Model (combines conceptual space containing the geographical features of geo-spatial concepts and the semantic network containing the spatial relations between the geo-spatial concepts) of the geo-spatial information system. It is inferred that the Manhattan distance measure for the Hybrid model yields 15% increase of precision, recall and f-measure. It yields better relevant information retrieval as it retrieves all the neighbour geo-spatial concepts. In this paper the semantic similarity measure for computing the relevance of information in done. In future the general approach of semantic similarity computation can be developed for inspecting the semantic similarity between the various application tasks and the similarity judgement of the application tasks can be analysed. The conceptual contexts are not attempted with the semantic distance measures for the geo-spatial domain so the conceptual context can be included with the semantic distance measures.

REFERENCES

[1] Kay Khaing Win, "Measuring Geospatial Semantic Similarity between Geospatial Entity Classes", University of Computer Studies, Yangon, Myanmar, *IEEE,* 2006.

[2] Schwering.A, "Hybrid model for semantic similarity measurement", *4th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE05)*, Agia Napa, Cyprus, Springer, 1449-1465, 2005.

[3] Resnik. P, "Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language", *Journal of Artificial Intelligence Research*, 11: 95-130, 1999.

[4] Schwering, A. and M. Raubal, "Measuring semantic similarity between geospatial conceptual regions", *1st International Conference on GeoSpatial Semantics*, Mexico City, Mexico, Springer, 93-109, 2005.

[5] Schwering, A. and M. Raubal, "Spatial relations for semantic similarity measurement", *2nd International Workshop on Conceptual Modeling for Geographic Information Systems (CoMoGIS2005)*, Klagenfurt, Austria., Springer, 259-269, 2005.

[6] Gärdenfors.P, "How to make the Semantic Web more semantic. Formal Ontology in Information Systems", *Proceedings of the Third International Conference (FOIS2004)*. L. Vieu. Amsterdam, NL, IOS Press. 114: 153-164, 2004.

[7] Rodríguez, A. and M. J. Egenhofer, "Comparing geospatial entity classes: An asymmetric and context-dependent similarity measure", *International Journal of Geographical Information Science*, 18(3): 229-256, 2004.

[8] Rada, R., H. Mili, E. Bicknell and M. Blettner, "Development and application of a metric on semantic nets", *IEEE Transactions on systems, man, and cybernetics*, 19(1):17-30, 1989.

[9] Resnik, P. "Using information content to evaluate semantic similarity in a taxonomy", *14th International Joint Conference on Artificial Intelligence (IJCAI)*, Montreal, Canada, Morgan Kaufmann, 448-453, 1995.

[10] Goldstone, R. L. and J. Son "Similarity - Cambridge Handbook of Thinking and Reasoning" K. Holyoak and R. Morrison. Cambridge, Cambridge University Press: 13-36, 2005.

[11] Levenshtein, I. V, "Binary codes capable of correcting deletions, insertions, and reversals", *Soviet Physics Doklady*, 10(8): 707-710, 1966.

[12] A. Vadivel, A.K. Majumdar, Shamik Sural, "Performance comparison of distance metrics in content-based Image retrieval applications" Computer Science and Engineering, Indian Institute of Technology, Kharagpur, India.

[13] Hosseinabady, M. and Nunez-Yanez, "Run-time stochastic task mapping on a large scale network-on-chip with dynamically reconfigurable tiles", J.L. Computers & Digital Techniques, Vol. 6, pp. 1-11, *IET Journals & Magazines*, 2012.

[14] M.J. Swain and D. H. Ballard, "Color Indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.

[15] S.Sural, G. Qian and S.Pramanik,"Segmentation and Histogram Generation using the HSV Color Space for Content Based Image Retrieval", *IEEE Int. Conf. on Image Processing*, Rochester, NY, 2002.

[16] A. Jain and A. Vailaya, "Image Retrieval using Color and Shape", *Pattern Recognition*, 29(8), pp. 1233–1244, 1996.

[17] S. Sural, "Histogram Generation from the HSV Color Space using Saturation Projection", in Multimedia Systems and Content-based Image Retrieval, S.Deb eds, Idea Group Publishing, Hershey, PA, USA, 2003.

[18] W. Niblack, R. Barber,W. Equitz, M. Flickner, E. Glasman, D. Pektovic, P. Yanker, C. Faloutsos, and G. Taubin,, "The QBIC Project: Querying Images by Content using Color Texture and Shape", *Proc. SPIE Int. Soc. Opt. Eng., in Storage and Retrieval for Image and Video Databases*, vol. 1908, pp. 173-187, 1993.

[19] Michael G. Michie," Use of the Bray-Curtis similarity measure in cluster analysis of foraminiferal data**",** *Journal of the International Association for Mathematical Geology*, Volume 14, Issue 6, pp 661-667,1982.

[20] Ahlqvist, O., "Using Semantic Similarity Metrics to Uncover Category and Land Cover Change", *1st International Conference on GeoSpatial Semantics*, Mexico City, Mexico, Springer, 107-119, 2005.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 794-801

http://www.issr-journals.org/ijias/

The optimization of Spray and Wait routing Protocol by prioritizing the message forwarding order

Qaisar Ayub, Sulma Rashid, M. Soperi Mohd Zahid, and Abdul Hanan Abdullah

Department of computer science and information science,
Universiti Teknologi Malaysia,
Johor bharu, Jhohor, Malaysia

ABSTRACT: In most of the wireless applications, e.g. military networks, vehicular ad-hoc networks, wild life tracking and sensor network, it is impossible to sustain a complete end-to-end data transmission path from source to destination. Thus, the traditional routing strategies (TCP/IP) cannot work with such environments as they need to establish the uninterrupted path before transmission. The disruption-tolerant network (DTNs) has emerged as technology that enables the communication in highly disrupted environments by intermittently connected mobile nodes. The DTN transmission model follows a store carry and forward mechanism wherein the nodes stores the incoming message , carries it while moving and forward when comes in the transmission range of other contacts. The DTN protocol can be divided as single copy and multi copy. In single copy protocol the node is allowed to generate the unique copy of message and forward it on a unique path. The multi copy protocols generate and transmit the multiple copies of each message and forward it along various paths.
In DTN environment, the communication opportunity exists for short duration and a node cannot transmit all copies of its carried message from its forwarding queue. Hence the order at which the messages are forwarded becomes very important. In this paper, we have proposed a message forwarding order for multi copy Spary&wait routing protocol called as smallest message first. Through simulations, we proved that proposed policy out perform as compared to existing FIFO in context of delivery probability, overhead, latency average and buffer time average.

KEYWORDS: DTN, store-carry-forward, Routing, Buffer management, forwarding order, multi copy protocols, FIFO.

1 INTRODUCTION

In conventional routing schemes Ad Hoc [1] it is necessary to launch end-to-end path from source to destination previous to the transmission of data. Hence most of wireless applications such as sensor networks for ecological monitoring [2], sensor networks biological sensor networks and vehicular networks[3] due to highly unstable path which may change or break while being discovered.

Delay tolerant networks [4] permit the transmission of data by means of intermittently connected mobile nodes. DTN has put forward a store-carry-forward paradigm, where every node acts as a hop in the network and store the incoming message in its buffer, carries the message while moving and forward when it encounter with other nodes.

Foundation on the scheme of message forwarding, routing plan for DTN can be separated in to two main categories, single copy[5] and multi copy[6]. Hence, single copy routing schemes only one copy of message exists in the network, which is forwarded along single path. While in multi copy schemes more than one copy of same message are forwarded to multiple paths. Multi copy routing cases according to have high influence on message delivery probability and robustness at the price of extra bandwidth, energy and memory usage such as epidemic, spray & wait. Some recent work has focused on controlling the resource consumption via efficient buffer scheduling [10-11] and message forwarding polices [12-13].

Corresponding Author: Qaisar Ayub (sheikhqaisar@gmail.com)

With extensive delays, restricted bandwidth and limited duration of contact time, a node in DTN multi copy routing cannot exchange all messages from its forwarding queue. Therefore the sequence in which the messages are transmitted becomes very vital [7].

In this paper we studied the impact of packet forwarding order with spray&wait [8] routing protocols and optimize its performance by introducing a new message forwarding scheme (smallest message transmit) SMT as compared to FIFO.

2 EXISTING QUEUING MANAGEMENT POLICIES

2.1 FIRST IN FIRST OUT (FIFO)

In FIFO queue mode all messages are arranged according to arrival time and the message which has oldest arrival time will be transmitted first.

2.2 RANDOM QUEUE MODE (RND)

The message is randomly selected for the transmission.

2.3 GRTR

"Assume A, B are nodes that meet while the destination is D P(X, Y) denote the delivery predictability that a node X has for Destination Y. GRTR forward the message to node only if P (B-D) >P (A-D)" [7].

2.4 GRTRSORT

"GRTRsort looks at difference P (B-D) – P(A-D) values for each message between the nodes and forward the message only if P(BD)> P(A-D)."

2.1 GRTRMax

"Select messages in descending order of P (B-D) forward the message only if P (B-D)> P (A-D)."

3 SPARY&WAIT

Spary&wait routing algorithm work in two phases In phase one, for each message M, N numbers of copies are sprayed/forwarded by source node over network. If the destination is not found in spraying phase, each node carrying the N copy of message and perform direct transmission.

Spary&wait routing inherit the speed of Epidemic routing and simplicity of direct transmission. It initially spread N number of copies in the network just as epidemic routing, when N numbers of copies are forwarded with high probability that at least one of copy will be delivered to destination, it stop and let each node perform direct transmission.

In binary Spary&wait if node A having N>1 copies encounter with node b with no copies, transmit B->N/2 and hold N/2 by itself, when it left with one copy it performs direct transmission.

The existing message forwarding strategy for Spary&wait is First In First Out (FIFO). In the next section we discuss the metrics which will be used to examine the performance of first in first out (FIFO) with proposed Smallest Message Transmit (SMT).

4 APPROACH

Table 1. Symbol and description

Symbol	Description
M	Messages
SM	Size of Message
NM	Number of Copies of Message M
AT	Arrival Time
TTM	Total Transmission Time for NM
DTnT	Delay Tolerance Network Transmission Time
TM	Transmission Time Per Message

Consider a scenario where a source node have to forward M messages such that M ∈ [9]. NM correspond to number of copies of Message M where N={1,2,3,4,5,6,7…20}. SM is the size of message, TTM (Total Transmission Time) represent the net time required to transmit NM copies of message M. When source node starts forwarding then DTnT (DTN Time) is the duration by which two nodes can perform the message exchange. The Time required to transmit the single copy of message M can be calculated from equation [10].

$$\sum_{k=1}^{i} TM_k = \frac{\sum_{k=1}^{i} TTM_k}{\sum_{k=1}^{i} NM_k} \qquad (1)$$

DTnT (Delay Tolerant network Transmission Time) represents the time available for the transmission; we can calculate how many NM can be transmitted.

$$TCT = \frac{DTnT}{\sum_{k=1}^{i} TM_k} \qquad (2)$$

Number of Iterations required to transmit the copies of message with available DTnT can be calculated as

$$Iteration = \frac{\sum_{k=1}^{i} TTM_k}{DTnT} \qquad (3)$$

In the next part we evaluate the performance of existing first in first out FIFO and proposed message Smallest Message Transmit (SMT) forwarding policy.

a. FIFO(FIRST IN FIRST OUT)

In FIFO scheduling mechanisms message which arrives first is the one to be processed first. In table 02, we have three messages M12, M23, M43 with messages size (M), Number of Copies (NM), Arrival time(AT), Total Transmission time (TTM), Delay Tolerant network Time (DTnT).

Table 2. FIFO Transmission order

M	SM	MN	AT	TTM	DTnT
M12	200MB	6	13ms	12s	8s
M23	133MB	6	14ms	8s	
M43	83MB	6	18ms	5s	

The above table describes the message forwarding sequence according to FIFO. Message M12 of size SM (SM=200MB) holding N (N=6) number of copies have TTM (TTM=12s). Time required to transmit single copy of message can be calculated from equation (1). $TM_{12} = TTM_{12}/NM_{12} = 12/6 = 2$. Hence 2s seconds are required to transmit each copy of M12.

The available DTnT (DTnT=8) by putting values in equation (2) we get total number of copies that can be forwarded with in DTnT.TCT =8/2 =>4 copies. The total of 4 copies will be forwarded to network. Total iteration required to forward all copies can be calculated as Iteration = 12s/8s= 1.5 Iterations. Hence, in 0.50 (4s) more Iteration requires to transmit all copies of Message.

b. THE PROPOSED ALGORITHM

(1) Get Messages from Router where Number of Message Copies greater then 1.
For each Message IN RouterMessageList

If Message→Copies>1

ADD (MessageLIST, message)

End if

End loop

(2) Sort the MessageLIST according to size in Ascending order (TSMF)

SortedMessageList=SORT(MessageLIST)

Table 3. Transmission order by SMT

M	SM	MN	AT	TTM	DTnT
M43	83MB	6	18ms	5s	8
M23	133MB	6	14ms	8s	
M12	200MB	6	13ms	12s	

The table 3 describes the message forwarding sequence according to SMT. Message M43 of size SM (SM=83MB) holding N (N=6) number of copies have TTM (TTM=5s). Time required to transmit single copy of message can be calculated from equation (1) $TM_{43} = TTM_{43}/NM_{43} = 5/6 => 0.83s$ are required to transmitted single copy of M43. The available DTnT is 8s by putting values in equation (2) we get total number of copies that can be forwarded with in DTnT.TCT =DTnT/TM=>8/0.83=9.63 copies. Total iteration required to forward all copies can be calculated as Iteration = TTMa/DTnT=5/8= 0.62 Iterations. Hence 0.48 Iteration can be allocated to next message. Simulations results have proved that with SMT policy, Delivery Ratio can be increased while minimize the Overhead, Latency Average, and Buffer time average.

5 PERFORMANCE METRICS

5.1 DELIVERY PROBABILITY

It is the ratio of message received over message send. High probability means that more messages are delivered to the destination. Objective of algorithm is to maximize the delivery probability.

$$\text{DeliveryProbability} = \frac{\text{NumberOfMessageRecieve}}{\text{NumberOfMessageSend}}$$

5.2 LATENCY AVERAGE

It is sum of time when message is generated and when it is received. Mathematically can be represented as

Objective of algorithm is to minimize the value of latency average.

$$LatencyAverage = MessageRecieveTime - MessageGenerationTime$$

5.3 OVERHEAD

It is the negation of number of messages relayed to number of message delivered. Low value of overhead means less processing required delivering the relayed messages. Mathematically it can be represented as

$$Overhead = NumberOfMessageRelayed - NumberOfMessageDelivered$$

Objective of algorithm is to minimize the value of overhead.\

5.4 BUFFER TIME AVERAGE

It is Sum of time spend by a message(s) in buffer divided by message send. Mathematically can be represented as

$$BufferTimeAverage = \frac{\sum_{k=1}^{i} WaitTimeMessage_k}{NumberOfMessageBuffered}$$

6 SIMULATION AND RESULTS

The following sections will provide the performance evaluation of SMT policy as compared to First in First out (FIFO) using DTN ONE simulator. The ONE [6] is a discrete event simulator written in Java. The main objective of simulator is to implement DTN store-carry-forward paradigm where the probability of disconnections and failures is high.

6.1 DELIVERY PROBABILITY W.R.T NO. OF COPIES WITH FIFO AND SMT (SPRAY& WAIT)

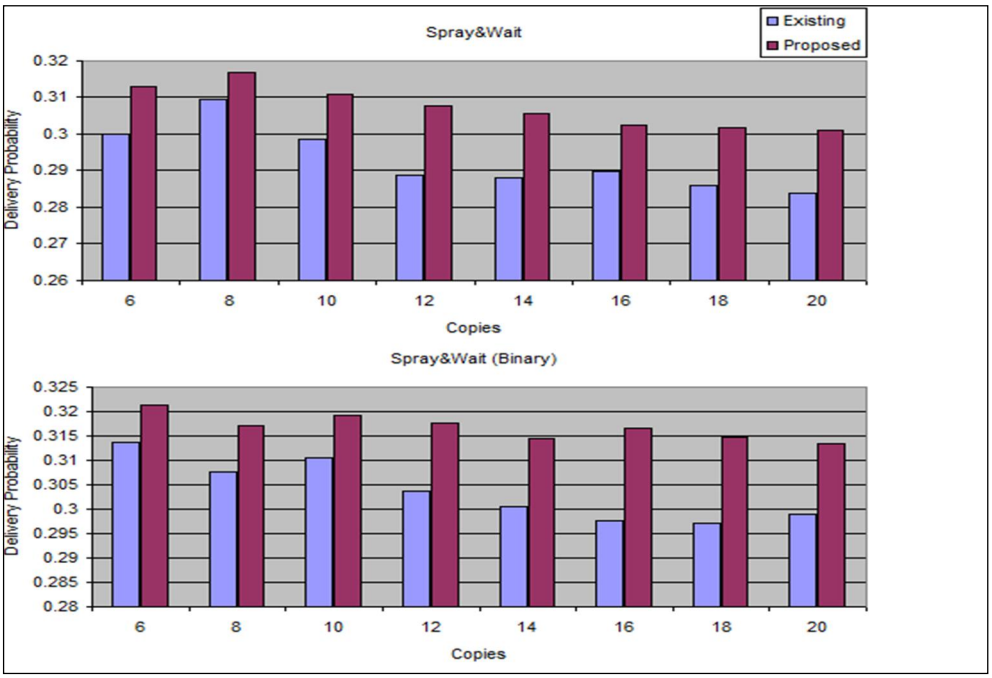


Fig. 1. Delivered Probability w.r.t No. of Copies with FIFO and SMT (Spray& Wait)

As we depict in Fig (1), results has proved that the delivery probability of spray&wait and binary Spray&wait router with proposed scheme (SMT) were performed better as compared to existing FIFO. Moreover by increasing number of copies both routers with SMT scheme performance better but in case of binary spray&wait the delivery probability raises more quickly.

One reason for this is that binary spray&waits forward more copies to relay nodes (N/2), with implementation of SMT small messages gets more opportunity to be transmitted first. In case of spray&wait SMT the average delivery probability is 0.3074 and with existing FIFO 0.2931. In binary spray&wait, average of delivery probability by using SMT is 0.3168 and with FIFO 0.3037. The proposed SMT have optimized the delivery probability.

6.2 LATENCY AVERAGE W.R.T NO. OF COPIES WITH FIFO AND SMT (SPRAY& WAIT)

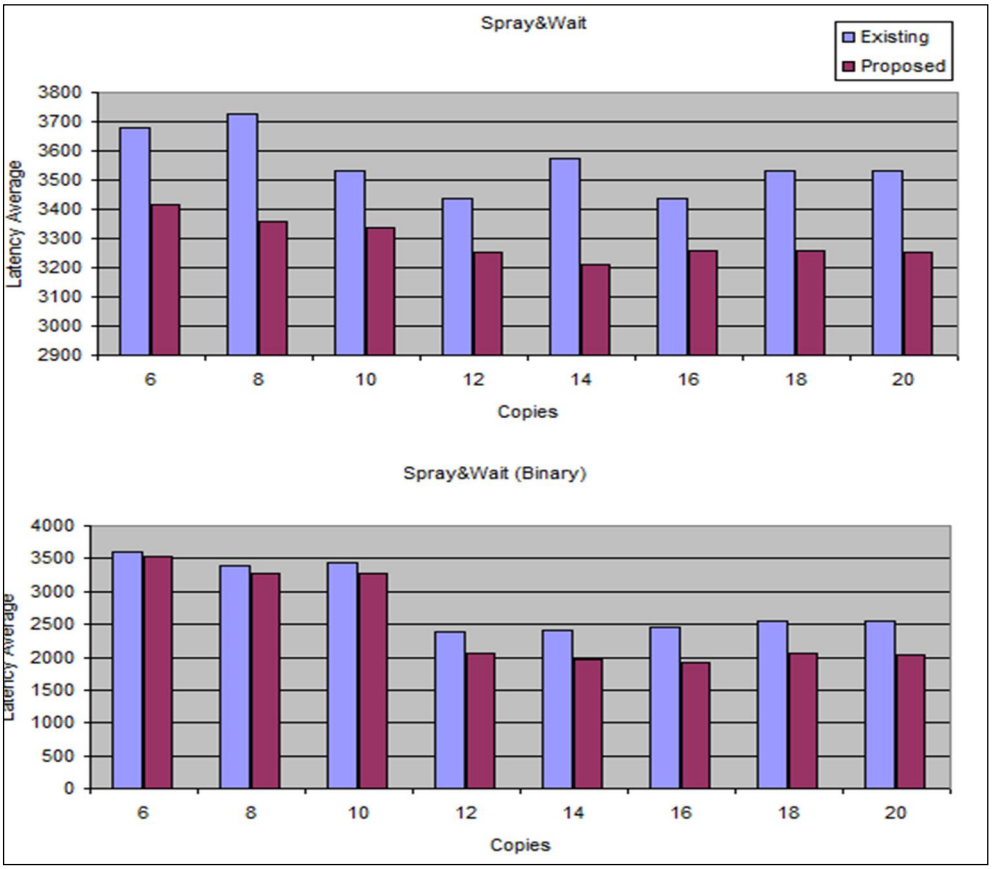


Fig. 2. Latency Average w.r.t No. of Copies with FIFO and SMT (Spray& Wait)

Fig (2) shows the result of latency average by increasing no of copies in spray&wait and binary spray&wait router with existing FIFO and proposed SMT. The objective of algorithm is to minimize the value of latency average. In case of spray&wait with SMT the latency average is 3293.56 and with FIFO 3556.23. In binary spray&wait the SMT the average of latency is 3204.86 and with FIFO 3492.83. The proposed SMT optimize the latency average.

6.3 BUFFER TIME AVERAGE W.R.T NO. OF COPIES WITH FIFO AND SMT (SPRAY& WAIT)

The Fig(3) shows the impact of buffer over Spray & wait and Binary spray & wait with existing (FIFO) and proposed SMT. By increasing number of copies of message the buffer time average drops with a constant ratio. One reason for this is that by increasing number of copies with SMT give high probability of small messages to be forwarded in the network so message spend less time in the buffer. In case of spray&wait SMT the average of buffer time is 2413.69 and with existing FIFO 2429.83. In binary spray&wait the SMT the average buffer time is 1998.60 and with FIFO 1997. The proposed SMT optimize the buffer time average.

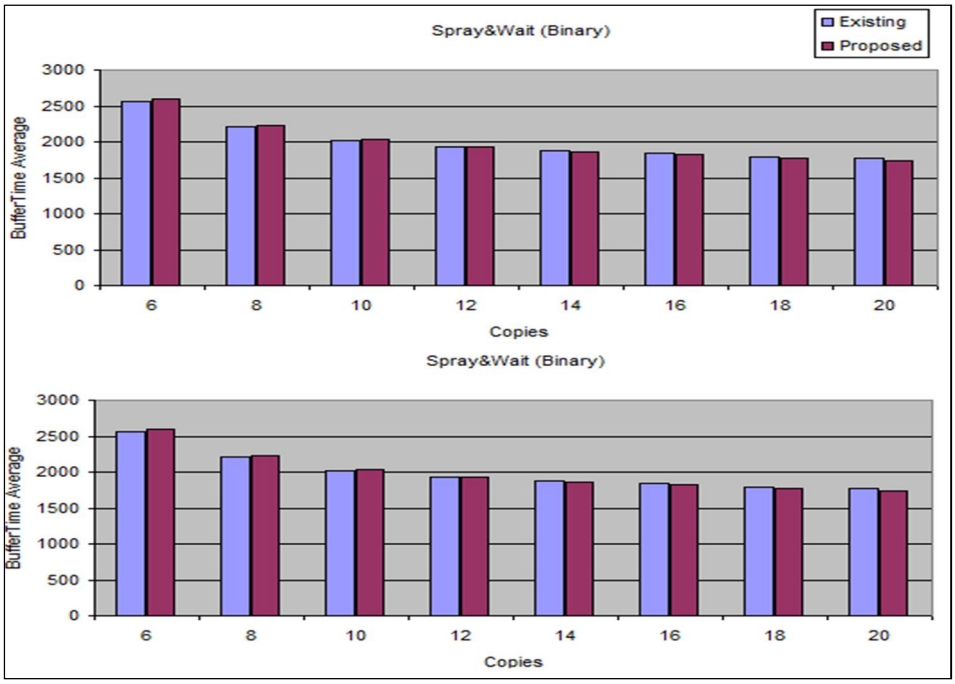


Fig. 3. Buffer Time Average w.r.t No. of Copies with FIFO and SMT (Spray& Wait)

6.4 OVERHEAD RATIO W.R.T NO. OF COPIES WITH FIFO AND SMT (SPRAY&WAIT)

The Fig(4) asserts the of influences of overhead with FIFO and SMT. On average in all cases of no of copies the overhead of proposed (SMT) algorithm is fewer with existing FIFO. In case of spray&wait SMT the average overhead is 16.52 and with existing FIFO 16.62. In binary spray&wait the SMT the average overhead is 20.88 and with FIFO 21.03.

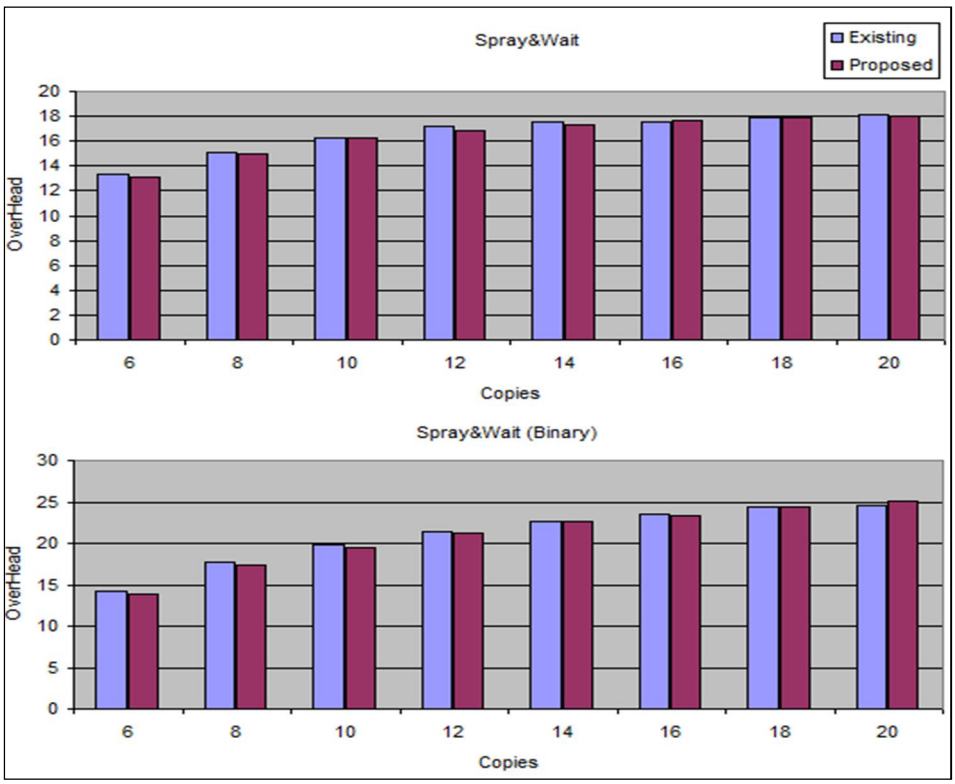


Fig. 4. Overhead ratio w.r.t No. of Copies with FIFO and SMT (Spray& Wait)

7 CONCLUSION

In this paper we have proposed a message forwarding strategy called as Smallest Message First and optimizes the performance of spray&wait routing protocol. The proposed routing method increases the delivery ratio, buffer time average and reduces overhead.

REFERENCES

[1] Bansal, M., R. Rajput, and G. Gupta, Mobile ad hoc networking (MANET): Routing protocol performance issues and evaluation considerations. 1999.

[2] Mainwaring, A., et al. Wireless sensor networks for habitat monitoring. in Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications. 2002: ACM.

[3] Lochert, C., et al. A routing strategy for vehicular ad hoc networks in city environments. in Intelligent Vehicles Symposium, 2003. Proceedings. IEEE. 2003: IEEE.

[4] Fall, K. A delay-tolerant network architecture for challenged internets. in Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications. 2003: ACM.

[5] Spyropoulos, T., K. Psounis, and C.S. Raghavendra. Single-copy routing in intermittently connected mobile networks. in Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on. 2004: IEEE.

[6] Spyropoulos, T., K. Psounis, and C.S. Raghavendra, Efficient routing in intermittently connected mobile networks: the multiplecopy case. Networking, IEEE/ACM Transactions on, 2008. 16(1): p. 77-90.

[7] Lindgren, A. and K.S. Phanse. Evaluation of queueing policies and forwarding strategies for routing in intermittently connecte d networks. in Communication System Software and Middleware, 2006. Comsware 2006. First International Conference on. 2006: IEEE.

[8] Spyropoulos, T., K. Psounis, and C.S. Raghavendra. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. in Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking. 2005: ACM.

[9] Fall, K. and S. Farrell, DTN: an architectural retrospective. Selected Areas in Communications, IEEE Journal on, 2008. 26(5): p. 828-836.

[10] Perkins, C.E. and E.M. Royer. Ad-hoc on-demand distance vector routing. 1999: IEEE. Rashid, Sulma, et al. "Dynamic Prediction based Multi Queue (DPMQ) Drop policy for Probabilistic routing protocols of delay tolerant network." Journal of Network and Computer Applications (2013).

[11] Rashid, Sulma, et al. "Message Drop Control Buffer Management Policy for DTN Routing Protocols." Wireless Personal Communications (2013): 1-17.

[12] Ayub, Qaisar, Abdul Hanan Abdullah, and Sulma Rashid. "Connection Frequency Buffer Aware Routing Protocol for Delay Tolerant Network." Journal of Electrical Engineering & Technology 8.3 (2013): 649-657.

[13] Ayub, Qaisar, et al. "Threshold based locking routing strategy for delay tolerant network." Wireless Networks: 1-12.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 802-812

http://www.issr-journals.org/ijias/

Impact des crimes économiques sur le financement du développement: Cas des pays en développement (PED)

[Impact of economic crimes on the financing of development: Case of developing countries]

Tarik TALII

Laboratoire de Modélisation Mathématique et de Calcul Économique,
Université Hassan I, 26000 Settat, Maroc

ABSTRACT: In the recent years, many studies have showed that weak institutions and the extent of economic crimes in developing countries are one of the deep causes of their poor economic performances. Debt crises which are among other consequences of a low of public levies are accentuated in developing countries, in particular with the development of what is generally known as shadow economy. This research aims to study the potential link existing between economic crimes and the financing of development in developing countries. The economic crimes sector, which is an essential component of the underground economy has a negative effect on the tax mobilization, by cons, the quality of governance is a positive factor in favor of state budgets. This sector is an important fiscal centre which unfortunately for both social and political reasons, sometimes bears fiscal charges beyond its real contributory capacities. This study aims in particularly to investigate the impact of economic crimes on Financing for Development in approximately one hundred developing countries through the channel of public resource mobilization. The main results of empirical analyses using data covering the period 1996-2012 confirm that, it is better for developing countries to fight against economic crimes playing on improving the quality of institutional governance to ensure economic growth sustainability.

KEYWORDS: Economic crimes, underground economy, Shadow Economy, Quality of governance, Development Financing, Public levy, Tax mobilization, Developing Countries.

RESUME: Ces dernières années, de nombreux travaux ont permis de montrer que le retard des pays en développement (PED) en matière de développement économique, serait dû en particulier à l'inefficacité de leurs institutions et à l'ampleur des crimes économiques. Les crises d'endettement conséquences entre autres d'un faible niveau des prélèvements publics se sont accentuées dans les PED notamment avec le développement de plus en plus grand de ce qu'il est convenu d'appeler l'économie de l'ombre. Ce travail de recherche a pour objectif global d'étudier le lien potentiel existant entre les crimes économiques et le financement du développement les PED. Le secteur des crimes économiques, composante essentielle de l'économie souterraine exerce un effet négatif sur la mobilisation fiscale, par contre, la qualité de la gouvernance constitue un facteur positif en faveur des budgets des États. Ce secteur constitue un important foyer fiscal qui, malheureusement pour des raisons aussi bien sociales que politiques supporte souvent des charges fiscales en deçà de ses capacités contributives réelles. Cette étude vise particulièrement à étudier l'impact des crimes économique sur le financement du développement dans environ cent PED via le canal de la mobilisation des ressources publiques. Les principaux résultats issus d'analyses empiriques à base de données couvrant la période 1996-2012, confirment l'idée selon laquelle il est dans l'intérêt des PED de lutter contre les crimes économiques en jouant sur l'amélioration de la qualité des institutions de gouvernance afin d'assurer une soutenabilité de la croissance économique.

MOTS-CLEFS: Crimes économiques, Économie souterraine, Économie de l'ombre, Qualité de la gouvernance, financement du développement, Prélèvement public, Mobilisation fiscale, Pays En Développement (PED).

Corresponding Author: Tarik TALII (rontecktarik@gmail.com)

1 INTRODUCTION

Pour atteindre un niveau de développement satisfaisant et financer les investissements nécessaires à la réalisation d'une croissance forte et durable, l'économie d'un pays doit disposer des ressources internes nécessaires à son financement, d'où l'importance de la mobilisation de l'épargne intérieure. En effet, les PED sont dépendants des financements extérieurs. En plus de son caractère insuffisant, l'épargne intérieure est liquide et de court terme, ce qui la rend volatile et difficilement exploitable pour le financement de l'économie.

En 2002, le Consensus de Monterrey sur le financement du développement, adopté au sein de L'Organisation des Nations unies (ONU), a établi que les ressources financières extérieures ne permettaient pas à elles seules d'atteindre les Objectifs du Millénaire pour le Développement (OMD) et qu'il était nécessaire de mettre au point de nouvelles stratégies en mobilisant les ressources intérieures [1]. Dans les PED, la politique en faveur du développement humain est une priorité inscrite notamment dans les OMD. Une mobilisation optimale de l'Espace Budgétaire est donc nécessaire pour atteindre les objectifs et éviter les gaspillages.

Parmi les causes de la faible mobilisation des ressources publiques nécessaires au financement du développement des PED figurent les crimes économiques assimilés aux dépravations dangereuses qui entravent le développement humain. Ces crimes économiques privent les budgets publics d'un montant considérable de recettes fiscales, faussent la concurrence, violent les droits socio-économiques des citoyens, ralentissent le progrès économique, ponctionnent abusivement les systèmes publics de sécurité sociale et propagent une culture du non-droit.

Pour un pays donné, existe-t-il une relation entre les crimes économiques et le développement ? Nous pouvons répondre par l'affirmative à cette question car, a priori, il y a une corrélation évidente entre ces deux phénomènes, parce que la lutte contre les crimes économiques est indispensable à la bonne marche de l'économie et donc au développement économique et social.

L'influence des crimes économiques sur le développement n'a jamais été véritablement évaluée : il est difficile, voire impossible, de chiffrer cet effet. Néanmoins, les corrélations entre les crimes économiques et le développement des sociétés (entendu au sens large) peuvent être appréhendées à travers de simples régressions économétriques. Sans avoir la possibilité de conclure sur l'existence de causalités entre le crime économique et le type de développement pris en compte, les conclusions confirment toutes l'importance de la lutte contre les crimes économiques pour le développement.

Que ce soit sur la pauvreté et ses différents aspects (monétaire, accès aux biens primaires, santé, éducation), sur la gouvernance ou sur les questions liées à la transparence et à la démocratie, une lutte contre les crimes économiques a toujours une influence positive. Les crimes économiques font l'objet d'une attention croissante, mais la définition du phénomène reste encore insuffisamment élaborée au plan criminologique et demeure absente au plan juridique. D'un point de vue économique, le problème de définition et de mesure des crimes économiques reste encore l'objet d'un fort débat.

L'impact global des crimes économiques sur le développement reste difficile à déterminer, en partie en raison de l'absence d'un concept clair et admis d'un commun accord, du fait que les systèmes de recensement des crimes économiques diffèrent considérablement d'un pays à l'autre et que plusieurs affaires ne sont pas signalées parce que les particuliers, les entreprises ou les établissements financiers choisissent de traiter les incidents en interne. En effet, l'analyse des effets des crimes économiques sur le développement a soulevé une controverse sur le choix et la robustesse des indicateurs de mesure utilisés. La difficulté de produire un indicateur capable de mesurer l'ampleur des crimes économiques semble une tâche pénible, voire impossible.

A partir de ces différents constats, il sera très intéressant d'étudier les crimes économiques comme une forme de l'économie souterraine. En effet, l'économie souterraine s'inscrit directement dans notre recherche, notamment parce que l'objet mesuré est le même, le crime économique est la forme extrême de l'économie souterraine. En pratique, la mesure de l'économie souterraine permet de nous éclaircir sur la nature de l'effet des crimes économiques sur le financement du développement, ainsi que sur les canaux de transmission de cet effet.

2 CRIMES ÉCONOMIQUES : PROBLEME DE DEFINITION ET DE MESURE

Il n'existe pas à proprement parler de définitions de crimes économiques. Ou plutôt, il en existe de nombreuses en fonction des criminologues qui les proposent, des législations différentes entre pays ou en fonction des mœurs et coutumes de certaines cultures et il est difficile d'enfermer le crime économique dans une définition stricte et une mesure universelle.

2.1 DÉFINITION DES CRIMES ÉCONOMIQUES

L'ONU [2] définit la criminalité économique et financière comme étant « toute forme de criminalité non violente qui a pour conséquence une perte financière ».

Le crime économique est défini par le biais d'un ensemble de critères [3], qui tient aux modalités de réalisation de l'infraction et à ses effets : 1. réalisation dans le contexte de la vie économique, 2. méthodes astucieuses ou frauduleuses (et usage marginal de la force ou de l'intimidation) ; 3. savoir-faire propre au monde économique, commercial ou financier ; 4. volonté d'accumulation des profits ; 5. dommages considérables aux circuits économiques.

De son côté, François-Regis Mathieu [4] définit le crime économique comme une atteinte à la bonne conduite économique et surtout financière : Fraude, faillite, délit d'initié, contrefaçon, etc. Ce sont des atteintes à l'économie et à ses institutions, mais jamais les dommages occasionnés à la personne.

Pour Nicolas Queloz [5], le crime économique représente un abus de confiance et de bonne foi dans les relations d'affaires et porte ainsi atteinte à la crédibilité et à la sécurité de certains secteurs d'activités économiques, voire de l'ordre économique et financier dans son ensemble.

Le crime économique est un phénomène global et omniprésent, bien que le concept puisse recevoir différentes définitions dans les législations nationales. Au-delà de la criminalité économique et financière, la criminalité d'affaires et, en termes plus restreints, le crime en col blanc, sont autant de formules que l'on retrouve dans les études empiriques en la matière, rarement avec une identité de significations.

Selon Jean-Luc Bacher [6], le crime économique, qu'il soit ou non imputable aux milieux d'affaires, reste toutefois un phénomène aussi vaste que réel. Dans ses formes les plus bénignes, il est le fait d'un nombre considérable d'individus et d'entreprises. Il suffit de songer à la fraude fiscale ou à la fraude à l'assurance. Dans ses formes les plus aiguës, il porte parfois sur des sommes énormes.

Les pratiques qui sont susceptibles de rentrer dans la catégorie des crimes économiques sont énoncées sur une liste, qui – par sa propre admission - ne se veut pas exhaustive : blanchiment d'argent ; fraudes financières ; fraudes fiscales; fraudes douanières; escroqueries; corruption; criminalité informatique; délits d'initiés ; faillites frauduleuses ; concurrence déloyale ; abus de confiance…

Aujourd'hui, nous parlons des crimes économiques souvent dissimulés derrière des astuces ingénieuses et des moyens sophistiqués impliquant de multiples transactions et un grand nombre d'intervenants, ce qui en complexifie d'autant la détection et la répression. Il importe donc d'être en mesure de rivaliser avec une criminalité complexe, bien souvent investie par des réseaux criminels organisés.

Compte tenu de la gravité de l'impact et des préjudices causés, les formes les plus pénalisantes de l'économie souterraine ont été classées dans la catégorie du crime économique. En effet, une partie des activités souterraines sont illégales; elles relèvent davantage du crime économique ou de « l'économie noire », qui englobe notamment la criminalité des affaires (la «criminalité dans les activités économiques») et le crime organisé (la «criminalité en tant qu'activité économique»).

Les crimes économiques sont des activités économiques illégales menées de façon « informelle » et particulièrement « souterraine » pour échapper à l'impôt et au contrôle de l'État. Faisant partie de l'économie souterraine, le crime économique fait référence aux transactions économiques non déclarées faites entre personnes. En effet, la prise en compte du terme économie souterraine, clandestine ou illicite comme une composante de l'économie informelle, induit trois formes d'activités très différentes : (i) l'économie souterraine générée par le travail non déclaré ou au noir ; (ii) l'économie souterraine générée par les crimes ou délits économiques et (iii) l'économie générée par les activités criminelles ou délictuelles et leur recel.

Cependant, ces activités ont toutes au moins trois points communs : (i) elles échappent aux règles économiques et sociales et à l'intervention de l'État ; (ii) elles ne donnent lieu à aucun prélèvement obligatoire (fiscal ou social) et (iii) elles faussent le jeu de la libre concurrence par rapport aux activités légales, soit du fait de leur propre activité (travail au noir), soit du fait de l'intégration de sommes provenant d'activités délictuelles ou criminelles, cherchant à se réinvestir (blanchiment d'argent).

Pour Schenider et Enste [7], l'économie souterraine couvre un champ beaucoup plus large. Elle comprend toutes les activités légales non déclarées potentiellement assujetties à taxation si elles étaient déclarées et l'ensemble des activités criminelles.

L'économie souterraine, écrit J. Rimbert [8], est communément comprise par les instances internationales (Union européenne, ONU), comme d'une part, l'ensemble des activités productrices licites non déclarées et, d'autre part, l'ensemble des activités illicites, productrices de biens ou de services.

L'économie souterraine est souvent une activité économique légale dissimulée ou dont l'importance est volontairement sous-estimée par les agents économiques (fuite fiscale).

Pour Pierre Lemieux [9], dans le secteur marchand de l'économie monétaire, nous trouvons certes les entreprises publiques ou privées, les coopératives et les mutuelles mais aussi des activités souterraines ou occultes. L'économie souterraine marchande va du travail au noir (marchés noirs productifs) à l'économie du crime en passant par les crimes économiques (fraude fiscale, corruption...). Il s'agit bien d'activités économiques largement marchandes mais qui tentent d'échapper au contrôle de la puissance publique.

Depuis lors, et en dépit de réelles difficultés à le définir, l'expression «crimes économiques» est couramment utilisée à côté de celle de "secteur non enregistré", "secteur informel" et "secteur souterrain". Aujourd'hui encore, la compréhension de ce qu'est exactement ce secteur, ainsi que ses causes et ses conséquences, constitue un sujet de grande préoccupation notamment dans les pays en développement.

2.2 MESURE DES CRIMES ÉCONOMIQUES

Par construction, les crimes économiques impliquent des actes qui s'écartent des règles et qui sont par conséquent déviants au sens propre du terme. Cette vision des crimes économiques comme une sorte de déviance par rapport à l'économie formelle implique l'existence d'une face cachée et frauduleuse de l'économie, celle dont les bénéfices ne sont pas déclarés aux autorités du pays, entraînant ainsi le non-paiement d'impôts.

Les crimes économiques peuvent se définir comme un domaine où les activités ne sont pas régies par une législation économique mais que la notion ne vaut qu'en tension avec l'existence d'une économie formelle. Le secteur informel peut rassembler à la fois les stratégies qui relèvent d'une pure économie de la survie, autant que des activités illégales qui ressortent de la criminalité économique. Les crimes économiques sont liés à la nature humaine. En effet, c'est la quête de « meilleures » conditions de vie qui pousse les hommes à se tourner vers des activités clandestines et souterraines, et ainsi, à contourner les lois, normes, règlements et pratiques officiels.

La mesure de l'économie souterraine et les méthodologies mises en œuvre présentent un double intérêt, méthodologique et quantitatif.

Au plan méthodologique, les mesures de l'économie souterraine sont intéressantes dans la mesure où elles cherchent à cerner, par des biais détournés, l'ampleur d'activités non déclarées, qu'elles soient illégales ou pas.

Au plan quantitatif, les mesures de l'activité souterraine nous fournissent des informations chiffrées qui ne sont pas sans relation avec les différentes activités incluses dans les crimes économiques. En effet, d'une part, dans l'économie souterraine sont parfois incluses les activités illicites productrices de biens et de services ; et d'autre part, les quantifications des activités licites non déclarées peuvent servir à mesurer l'évasion et la fraude fiscale. En outre, l'ampleur de l'économie informelle en soi peut constituer un indicateur de l'existence d'un environnement propice au développement des crimes économiques. Selon Kaufmann et al. [10], la présence d'un secteur informel important dans une économie est un indicateur de faible gouvernance.

Forts de ces raisons qui font que la présentation des diverses mesures de l'économie souterraine est susceptible de nous éclairer sur la mesure des crimes économiques, et avant de présenter les mesures de l'économie souterraine, il convient de signaler que, tout comme les crimes économiques, l'économie souterraine est un objet qui est loin d'être clairement défini et délimité.

De par sa nature même, les crimes économiques sont des phénomènes difficiles à mesurer. Il n'est donc pas étonnant de constater que les estimations de l'ampleur des activités souterraines varient beaucoup selon la méthode retenue. Cependant, depuis une vingtaine d'années, des progrès considérables ont été réalisés dans la mesure du phénomène et plusieurs méthodes sophistiquées ont été proposées.

Il y a d'une part les méthodes directes qui cherchent à quantifier les crimes économiques à partir de données microéconomiques liées aux « micro-observations » et incluent les études fondées sur des enquêtes-ménages ou sur des données de vérification fiscale, et, d'autre part, les méthodes indirectes et modèles fondées sur des hypothèses de type macroéconomique qui relient le phénomène à des variables observables et reposent sur les traces laissées dans l'économie par les activités cachées.

Les mesures directes présentent l'avantage de fournir des informations détaillées sur la structure de l'économie non enregistrée. En revanche, la qualité des résultats obtenus à partir de ces approches dépend de la qualité des questionnaires et de la volonté des agents (opérateurs) soumis aux enquêtes de coopérer. En effet, les personnes interviewées dans ce genre d'enquêtes hésitent à répondre fidèlement aux questions posées ce qui peut biaiser les résultats obtenus. Les mesures directes présentent également l'inconvénient de ne pas pouvoir estimer le développement et la croissance de l'économie non enregistrée sur une longue période.

C'est pour remédier ces inconvénients que des approches indirectes ont été développées.

Contrairement aux approches directes, les approches indirectes sont des approches essentiellement macroéconomiques, utilisant plusieurs indicateurs susceptibles de fournir, dans le temps, des informations sur le secteur des crimes économiques.

Plusieurs indicateurs sont couramment utilisés pour apprécier l'importance des crimes économiques aussi bien dans les PED que dans les pays développés :

- Indicateurs monétaires : le cœur de l'approche monétaire est de supposer que toutes les transactions générées par l'économie souterraine s'effectuent généralement au comptant. La croissance de l'activité des crimes économiques aura probablement pour effet d'accroître la demande de numéraire. Ces indicateurs fournissent en valeur absolue des estimations élevées d'ampleur de l'économie souterraine et peuvent être subdivisées en deux catégories :

La première catégorie dite des agrégats monétaires comprend deux variantes distinctes : (i) la variante de Gutman retient l'hypothèse que les transactions non déclarées sont réglées en espèces et prend en considération la variation du ratio espèces/dépôts au regard d'une année de base comme indicateur de l'économie souterraine ; et (ii) la variante Feige qui s'inspire de la théorie quantitative de Fisher et retient l'hypothèse que les transactions non déclarées sont réglées en espèces comme en chèque.

La deuxième catégorie se réfère à l'approche monétaire de Tanzi qui s'inspire de la fonction de demande de monnaie de Friedman. Cette approche encourt les mêmes critiques que celle des agrégats monétaires concernant les hypothèses relatives au choix de l'année de base à la vitesse au choix de l'année de base à la vitesse de circulation à l'absence d'une théorie des préférences monétaires ; cependant elle incorpore des variables explicatives, taux du pression fiscal, ou taux marginal d'imposition qui ne s'avèrent pas très significatives.

- Indicateurs comptables : l'estimation de la taille des crimes économiques est fondée sur l'écart entre les montants des revenus et dépenses qui ressortent des statistiques de comptabilité nationale ou des données individuelles. Puis à imputer celui-ci à l'économie souterraine celle-ci est réduite à la dimension de la fraude fiscale.

- Indicateurs des facteurs de production : sous l'effet de la croissance des crimes économiques en particulier et de l'économie souterraine en général, les facteurs de production, en particulier le travail, sortent, du moins en partie, de l'économie officielle. Ce déplacement peut entraîner une baisse du taux de croissance officiel de l'économie.

- Indicateurs de marché du travail : comme de plus en plus de personnes travaillent dans le secteur caché, le taux de participation à l'économie officielle risque de baisser. Parallèlement, du fait que les heures ouvrées dans le secteur caché augmentent, celles qui sont travaillées dans l'économie officielle risquent de diminuer.

- Le modèle de la variable latente : la taille des crimes économiques est estimé en faisant de celle-ci une fonction des variables observées qui sont censées influer sur ces crimes économiques — par exemple taux de taxation, indice de réglementation et celui du taux de chômage — et des variables dont l'évolution peut être attribuée aux crimes économiques, comme le numéraire, le temps de travail officiel, la croissance économique, etc. Ce modèle présente des avantages, car il tient compte de plusieurs causes et effets à la fois.

3 IMPACT DES CRIMES ÉCONOMIQUES SUR LE FINANCEMENT DU DEVELOPPEMENT

L'offre de biens publics, notamment de santé, d'éducation, d'infrastructures collectives, de sécurité, de cadre juridique, constitue un facteur crucial de la croissance économique des pays en développement. L'offre des différents biens publics est déterminée par le degré d'élargissement de l'Espace Budgétaire, c'est-à-dire du degré de mobilisation des différentes composantes de l'espace budgétaire [11].

Par ailleurs, l'offre de biens publics dépend du niveau de mobilisation des recettes publiques et des ressources de financement interne et externe y compris les ressources constituées par l'aide ; elle dépend aussi du degré d'efficacité des dépenses publiques.

Pour Chambas et al. [12], la plupart des pays en développement, et notamment les pays africains, ont connu ces dernières années des déséquilibres budgétaires importants; ce qui introduit une forte contrainte sur l'offre des biens publics. Parmi les facteurs de ces déséquilibres figure dans de nombreux pays, Chand et Moene [13] font apparaître l'insuffisance de la mobilisation des recettes publiques. Cela justifie que de nombreux travaux aient pour objectif de mettre en évidence des déterminants des recettes publiques.

Parmi les déterminants des recettes publiques figurent les crimes économiques. Ces fléaux ont pour conséquence de faibles niveaux d'éducation et de santé, des infrastructures collectives insuffisantes. Aussi, la mobilisation des ressources publiques, principal mode de financement des biens publics, constitue-t-elle un enjeu crucial [14]. En général, qu'ils soient de petite ou de grande envergure, les crimes économiques sont susceptibles de remettre en cause l'édifice fiscal à travers les pertes de recettes qu'elles engendrent et les distorsions qu'elles introduisent dans la concurrence. De là, nous pouvons soustraire le canal à partir duquel les crimes économiques affectent le développement.

3.1 CRIMES ÉCONOMIQUES ET MOBILISATION FISCALE

D'un point de vue opérationnel, Tanzi [15] et Schneider [16] soulignent que l'économie souterraine peut aussi être suscitée par le poids de la fiscalité - notamment pour le cas du secteur des crimes économiques (frauduleux). En effet, il semble exister un lien de causalité entre la fiscalité et les crimes économiques. Pour les cas particuliers des pays d'Afrique et ceux en développement en général, il serait difficile d'établir un lien entre le poids de la fiscalité et le développement du secteur non enregistré en raison notamment d'un déficit de références claires qui fassent ressortir une corrélation entre la taille du secteur des crimes économiques et la pression fiscale de manière générale.

Ce secteur peut être envisagé sous un angle positif, puisqu'il crée de l'emploi et des opportunités de revenus à un nombre considérable de personnes qui, autrement, n'auraient pas de moyens de survie ; il peut au contraire être vu sous un angle négatif, en le considérant comme un «secteur hors la loi», pouvant être responsable de graves manques à gagner pour l'Etat, et souvent responsable de distorsions économiques (concurrence déloyale envers les entités économiques légales.).

Schneider [17] définit le secteur non enregistré comme « l'ensemble des activités économiques partiellement ou totalement dérobées au fisc et qui contribuent à la détermination de la production nationale officielle ».

En ce qui concerne les composantes de ce secteur du secteur non enregistré, il nous semble utile de préciser tout d'abord qu'il comporte deux grandes catégories :

La première catégorie sera appelée secteur informel ou secteur «micro informel». Elle regroupe les activités micro informelles exercées par les petits opérateurs, financièrement et économiquement très peu importants. Ceux-ci disposent généralement de capitaux faibles et exercent leurs activités dans les rues de manière ambulante ou fixe.

La deuxième catégorie, la plus importante, sera désignée par l'expression «secteur des crimes économiques». Elle est constituée d'entreprises frauduleuses d'envergure, véritables entités économiques exerçant souvent de manière peu orthodoxe. C'est cette catégorie d'opérateurs qui constitue économiquement la fraction la plus importante du secteur non enregistré d'où l'importance du gisement fiscal qu'elle peut constituer. Par ailleurs, au sein même du secteur des crimes économiques, il importe de distinguer deux types d'activités : (i) les activités de crimes économiques de grande envergure et (ii) les activités frauduleuses de petite envergure. La caractéristique commune à toutes ces activités c'est que contrairement au secteur micro informel, la fraude constitue un facteur décisif de leur développement.

La définition de Schneider que nous retiendrons du secteur non enregistré présente l'intérêt de développer et de mettre en lumière la différence fondamentale entre ses deux composantes à savoir le secteur frauduleux et le secteur micro informel. De cette définition, tirée des travaux de Araujo et Chambas [18], il peut être possible de voir de manière assez claire le poids éventuel de chacune de ces deux composantes en termes de gisement fiscal, de dégager éventuellement leurs impacts sur la mobilisation fiscale ; et par conséquent leurs impacts sur le financement du développement.

3.2 FINANCEMENT DU DEVELOPPEMENT : MESURE ET DETERMINANTS

Le concept de prélèvement public recouvre l'ensemble des recettes fiscales et non fiscales du gouvernement central et des collectivités locales (ressources du domaine de l'État, prélèvements à travers les caisses de stabilisation et dividendes des entreprises publiques) en excluant les aides et dons. Les prélèvements des organismes sociaux sont également inclus.

La mesure des taux de prélèvements publics (recettes publiques) soulève diverses difficultés relatives à l'évaluation des recettes publiques, du produit et enfin des ratios de recettes publiques par rapport au produit. Nous retenons comme

indicateur de ressources publiques le taux de prélèvement public par rapport au PIB, qui mesure les ressources mobilisées par l'État relativement aux ressources produites dans l'économie.

Les travaux précurseurs de Lotz et al. ont souligné le rôle du niveau de développement, du degré de monétisation, de l'ouverture commerciale et de la composition sectorielle du produit dans la détermination du niveau des recettes publiques.

Plus récemment, l'accent a été mis sur le rôle des facteurs institutionnels jusqu'alors négligés. Les comportements de rente, de prédation et de corruption dans l'administration publique et la fraude fiscale affectent fortement les finances publiques et tout particulièrement le niveau des recettes publiques. Diverses analyses ont mis en évidence un impact négatif des crimes économiques sur la mobilisation des recettes publiques mais présentent diverses lacunes. En effet, l'une des principales lacunes réside dans la difficulté à définir les crimes économiques dans un ensemble homogène. De ce fait, la majorité des études de l'impact des crimes économiques sur la mobilisation des recettes publiques procède à l'évaluation par segment.

La littérature sur les déterminants du prélèvement public produit toute une série d'hypothèses testables. De là, il sera possible de ressortir l'impact des crimes économiques (secteur non enregistré) sur la mobilisation fiscale. Nous partons de l'hypothèse qu'il agit négativement sur le niveau des recettes fiscales, et essayons de tester cette hypothèse.

3.3 MODÈLE ET STRATÉGIE ÉCONOMÉTRIQUE

Partant de l'hypothèse selon laquelle les crimes économiques peuvent favoriser le développement du secteur non enregistré (frauduleux notamment), il peut sembler logique de penser que ce dernier puisse avoir les mêmes effets sur les recettes fiscales. Cette hypothèse sera testée sur un ensemble de pays en développement (100 environ).

Suivant les caractéristiques structurelles des économies et l'environnement socio-économique, nous pouvons avoir le modèle de la mobilisation des ressources publiques suivant :

$$TPP = a_0 + \alpha_1 Pibt + \alpha_2 VAA + \alpha_3 VAI + \alpha_4 M2Pib + \alpha_5 Shadow + \alpha_6 QualityINS$$

(+) (-) (+) (+) (-) (?)

TPP = Taux de prélèvement public ; a = constante ; Pibt = Pib par tête ; VAA= valeur ajoutée agricole ; VAI = valeur ajoutée industrielle ; M2Pib = degré de monétarisation de l'économie ; Shadow = taille du secteur non enregistré ; QualityINS = qualité d'institutions politiques.

Toutes les variables utilisées sont extraites du World Development Indicators à l'exception de la variable shadow (considéré comme une proxy du secteur des crimes économiques), pour laquelle une procédure d'interpolation a été utilisée à partir des statistiques de Schneider et al. [19]. Pour procéder à l'interpolation, un taux de variation annuel moyen a été calculé. Nous étions pour cela parti de l'hypothèse d'une faible variation de la taille du secteur des crimes économiques entre deux années consécutives pour dériver les données manquantes.

Dans notre modèle, à la différence des études précédentes, en dehors des variables structurelles qui sont conventionnellement utilisées dans l'explication du niveau du prélèvement public (mesuré par le rapport recettes fiscales et non fiscales sur le PIB), nous ajoutons deux variables concernant les crimes économiques. L'objectif est double, il s'agit d'un côté de mesurer l'effet de la taille des crimes économique sur le niveau du prélèvement public, et dans un autre côté il s'agit tenir compte l'effet de la gouvernance politique et de lutte contre les crimes économique sur la mobilisation des ressources publiques.

La variable Shadow désigne la taille du secteur des crimes économiques (calculée en pourcentage du PIB officiel) conformément à la définition de Schneider de l'économie souterraine saisissant l'ensemble des activités marchandes légales de production de biens et de services que des agents économiques dissimulent délibérément aux pouvoirs publics pour éviter le paiement des impôts sur le revenu et autres taxes ainsi que les cotisations sociales, ou pour contourner certaines normes juridiques relatives au marché du travail ou procédures administratives [20].

La variable QualityINS indique la qualité d'institutions politiques tirée des données de Kaufman et al. dans le programme «Governance matter» qui reflètent l'état de gouvernance et de lutte contre les crimes économiques. Suivant l'étude de Duchene et Zouari [21], nous calculons la qualité de la gouvernance politique comme la moyenne de quatre indicateurs : «Efficacité gouvernementale»; «Qualité de la réglementation»; «État de droit» ; «contrôle de corruption». Nous rééchelonnons, en suite, cet indicateur sur l'intervalle 0-100. Plus cet indicateur est élevé, plus la qualité de la gouvernance est grande. Une autre différence avec des variables économiques est que les variables de la base Kaufman ne sont disponibles qu'à partir de 1996 de façon discontinue.

Il convient de préciser que le choix de la qualité d'institutions politiques comme une variable explicative du taux de prélèvement public est pertinent dans la mesure où elle permet de donner une indication sur l'effet de la gouvernance et d'une éventuelle politique de lutte contre les crimes économiques sur le niveau du prélèvement public; et de tester l'influence de la variable institutionnelle et politique sur l'effort fiscal. En effet, les crimes économiques influencent négativement les quatre indicateurs déjà cités. En effet, les crimes économiques génèrent des gains illégaux, faussent la concurrence sur les marchés légaux et nuisent à la capacité des Etats d'assurer une bonne gouvernance. Ils représentent également une menace directe pour la démocratie, le développement, le respect des droits économiques et l'Etat de droit.

En ce qui concerne la part de la valeur ajoutée agricole, elle est aussi une variable explicative du niveau du prélèvement public dans les pays en développement en général. Cette variable influencerait négativement le prélèvement public en raison du fait que les revenus tirés de leur activité par les paysans sont relativement faibles et constitueraient donc de faibles gisements fiscaux.

L'étude porte sur les PED pour lesquels les données sont disponibles et couvrent la période 1996-2012. La dimension de l'échantillon est donc double, ce qui nous permet de travailler en panel. La dimension transversale l'emporte sur la dimension temporelle; de ce fait, les problèmes spécifiques aux séries temporelles sont négligés.

La méthode d'estimation retenue fait appel à la technique des moindres carrés généralisés. Pour l'échantillon composé de l'ensemble des PED, le test de Hausman révèle une corrélation entre les effets spécifiques et les variables explicatives, ce qui ne permet pas l'utilisation de la méthode classique des effets aléatoires. Aussi, la méthode à effets fixes est-elle utilisée pour cet échantillon.

En effet, l'usage du test d'exogénéïté (d'orthogonalité) de Nakamura Nakamura nous épargne le recours à la technique des variables instrumentales. Pour la variable « shadow », bien que mesurée et donc susceptible de rencontrer un problème d'endogénéité du fait de l'erreur de mesure, ne l'est pas d'après ce test. Comme instrument, nous avons utilisé le taux de scolarisation primaire.

De même, il semble avoir un lien entre le niveau de scolarisation et le choix d'exercer dans le secteur frauduleux et/ou informel. Plus les gens sont éduqués, plus ils sont aptes à accepter l'impôt parce qu'ils seront plus aptes à comprendre que c'est à partir de ses recettes que l'Etat finance les dépenses d'intérêt public telles que les écoles, les centres de soin, les routes,… Toutefois, l'effet des différents niveaux d'éducation varie en fonction du type de crime ; accroître l'éducation primaire n'aura, a priori, pas d'influence sur les crimes demandant un certain niveau de «qualifications» (les crimes en col blanc).

Dans ce modèle l'effet des inégalités d'éducation sur les crimes économiques est écarté.

4 RÉSULTATS ET DISCUSSION

Dans l'ensemble, le pouvoir explicatif de l'équation mesuré à travers le test du Durbin-Watson est satisfaisant (R^2=0,46).

Conformément aux intuitions théoriques, toutes les variables traditionnelles sont positivement et significativement corrélées au taux du prélèvement public à l'exception de la part de la valeur ajoutée agricole qui agit négativement et significativement. Ceci s'explique par la sous-fiscalisation des activités agricoles aussi bien du coté des paysans que des grands commerçants exportateurs bénéficiant de rentes fiscales du fait de système fiscal incitatif de la plupart des pays.

La nature de la corrélation entre ces variables et le taux de prélèvement public confirme les résultats obtenus dans l'étude de Chambas et al..

La taille du secteur des crimes économiques s'avère significativement et négativement corrélée avec le niveau du prélèvement public constitue un frein pour le financement du développement. Par contre, la qualité de la gouvernance politique (QualityINS) constitue un facteur positif pour la mobilisation des recettes publiques. En plus, ses coefficients sont statistiquement tous significatifs. La difficulté de la maîtrise de ce secteur pour certains pays et le caractère invisible et désastreux de cette économie frauduleuse aggravent la mobilisation des ressources.

Le secteur non enregistré constitue un gisement fiscal sur lequel peut être fondée une politique de mobilisation de ressources internes, afin de pallier les difficultés d'accès aux ressources externes.

La présence des crimes économiques et de l'économie de la rente jouent doublement en défaveur de la mobilisation fiscale dans les PED. D'une part, elle constitue un frein à cette mobilisation et, d'autre part, elle incite les opérateurs du secteur privé moderne (qui supporte l'essentiel du fardeau fiscal) à converger vers le secteur frauduleux.

Les résultats trouvés sont quasiment les mêmes qu'il s'agisse de la majorité des PED. Ceci confirme l'impact négatif des crimes économiques sur la mobilisation fiscale dans les pays en développement et par conséquent au financement du développement. En effet, nos analyses empiriques montrent qu'une amélioration de la qualité des institutions démocratiques, des institutions de droits de propriété, et particulièrement des institutions de régulation est favorable à la soutenabilité de la croissance économique dans les PED.

5 TABLEAUX

Tableau 1. Exemple Récapitulatif des variables explicatives et des signes attendus

Variables explicatives	Signes attendus
Shadow	-
QualityINS	+/-
VAA	-
Pibt	+
VAI	+
M2Pib	+

Tableau 2. Impact des crimes économiques sur le prélèvement public dans les PED

Variables explicatives	Constante	Shadow	QualityINS	VAA	Pibt	VAI	M2Pib	R^2
PED	27,01 (15,74)	-0,06 (-5,18)***	0.09 (3. 290)**	-0,22 (-6,18)***	-0,03 (-0,43) **	0,15 (5,41)***	-4,36 (-4,51)***	0,46

*Degré de significativité : *** 1%, ** 5%,*10%. La statistique de Student entre parenthèse.*

N.B. Le secteur agricole ne fait pas partie du secteur des crimes économiques (Shadow)

6 CONCLUSION

Dans les PED, les crimes économiques constituent un manque à gagner important pour le budget de l'État, les crimes économiques créent aussi une somme de distorsions inacceptables pour la société. Ils reportent la charge des impôts et des prélèvements obligatoires sur l'économie officielle, ils perturbent l'équilibre social des secteurs touchés, ils encouragent les migrations clandestines, ils déséquilibrent le jeu de la concurrence, ils suscitent des conditions de travail anormales, ils lèsent les consommateurs qui ne disposent d'aucun moyen de recours ou de garantie, ils accentuent également l'inégalité sociale en favorisant l'enrichissement illicite d'un groupe d'opérateurs, et enfin favorisent l'expansion de la pauvreté en privant l'Etat de ressources indispensables au financement des dépenses de développement.

Le principal enseignement à tirer de cette étude est que les crimes économiques ont un impact négatif sur la mobilisation des ressources internes dans les PED. En effet, et conformément aux intuitions théoriques, les estimations économétriques ont mis à jour une forte corrélation négative entre la taille du secteur des crimes économiques et le niveau du prélèvement public et montrent l'effet positif de la qualité de gouvernance politique.

Plusieurs raisons expliquent la faible exploitation du gisement fiscal du secteur non enregistré. Au rôle défaillant des administrations fiscales vient s'ajouter la collusion entre hommes politiques et hommes d'affaires. Il s'ensuit le développement de la corruption, de l'évasion et de la fraude fiscale très préjudiciables aux finances publiques.

Les réflexions relatives à la psychologie fiscale peuvent aboutir à deux alternatives à la répression classique qui est une simple pénalisation monétaire. Ces alternatives sont connues sous les appellations de stratégies informelles et de stratégies positives :

Les stratégies informelles sont beaucoup plus psychologiques et s'appuient sur le rôle prépondérant donné à la honte. Le sentiment de culpabilité qui anime les mauvais payeurs d'impôt révélé à travers les stratégies informelles peut les conduire à un changement de comportement. La vulgarisation de ces stratégies peut être une alternative crédible dans une communauté où la fraude fiscale est assez répandue. Elles peuvent compléter les sanctions habituelles formelles et conduire à des résultats assez probants.

Les stratégies positives constituent la deuxième grande alternative, ou du moins le deuxième grand complément des sanctions formelles. Contrairement aux sanctions informelles, les stratégies positives n'ont aucun caractère répressif. Elles sont plutôt des appels à la conscience du citoyen, à travers lesquels, l'objectif poursuivi est le changement des mentalités et des comportements des contribuables. Les arguments mis en avant sont les méfaits et les dangers pour la société de leurs actes frauduleux. La communication et la sensibilisation sont les deux principaux piliers de la stratégie positive.

Un élément essentiel différenciant les stratégies informelles et positives est celui du délai au cours duquel les deux approches produisent leurs effets. Alors que les premières peuvent avoir des effets dans le court et moyen termes, les secondes ne seront visiblement productives que dans le long terme.

En effet, les stratégies positives traitent des effets néfastes des crimes économiques et du bienfondé du paiement correct des impôts et taxes dus à l'Etat et aux collectivités locales. Cependant, elles doivent être mises en œuvre avec prudence car l'usage abusif des stratégies positives peuvent être source de motivation à d'autres comportements déviants.

Le problème de la maîtrise des exonérations accordées dans les PED a également été posé. Il est apparu que, dans de nombreux cas, ces mesures «incitatives» sont abusivement utilisées et détournées des objectifs pour lesquels elles ont été accordées. Des réseaux opaques se créent parfois et, en fin de compte, les exonérations profitent à un noyau d'opérateurs au détriment du fisc et des finances publiques en général.

Finalement, nous pouvons voir que toutes les réflexions relatives aux comportements frauduleux ont pour finalité la recherche des stratégies permettant une meilleure mobilisation des recettes fiscales au profit de l'Etat et des collectivités locales. En effet, la lutte contre les crimes économiques et l'amélioration de la qualité des institutions de gouvernance permettent d'assurer une soutenabilité de la croissance économique.

Pour cela, la recherche d'un schéma pouvant conduire les contribuables à plus d'honnêteté dans leur déclaration s'impose.

C'est dans cette optique que la problématique de la fiscalisation du secteur non enregistré est pertinente.

REFERENCES

[1] OCDE, "Perspectives économiques en Afrique 2010," p.82.

[2] Nations Unies, "onzième congrès des Nations Unies pour la prévention du crime et la justice pénale," Bangkok, 2005.

[3] Cartier Bresson J., Josselin C., Manacorda S., "Les délinquances économiques et financières transnationales et globalisation : analyses et mesures du phénomène," *IHESI*, 2001.

[4] David Fongang, "La criminalité économique et ses conséquences sur la PME africaine : Cas de la fraude, la corruption et l'arnaque," *L'Harmattan*, 2011, p. 20.

[5] Nicolas Queloz, "criminalité économique et criminalité organisée," *Altern. économiques* | L'Économie politique, 2002/3 no 15.

[6] Jean-Luc Bacher, "Éditorial : criminalités économiques," *Criminologie*, vol. 30, n° 1, 1997.

[7] Schneider. F, "empirique results for the size of the shadow economy of western European countries over time," working paper 9710 *institut for Linz University*, 2000.

[8] J. Rimbert, "Lutte contre l'économie souterraine," *In Les notes bleues de Bercy*, 2002/229.

[9] Pierre Lemieux, "L'économie souterraine : causes, importance, options," *Les Cahiers de recherche de l'Institut économique de Montréal*, 2007.

[10] Kaufmann D., Kraay A., Mastruzzi M., "Governance Matters VII: Aggregate and Individual Governance Indicators, 1996-2007," *World Bank Policy Research Working Paper* no. 4654, 2008.

[11] Chambas G. & Al., "Afrique au sud du Sahara : Mobiliser des ressources fiscales pour le développement," CERDI, Ministère des Affaires Etrangères. *Edition Economica*, 2005.

[12] Jean-François Brun, Gérard Chambas, Samuel Guerineau, "Aide et mobilisation fiscale dans les pays en développement », Etudes et Documents," *Etudes et Documents,* Centre d'Études et de Recherches sur le Développement International CERDI, 2008.

[13] Chand S. K., Moene K. O., "Controlling Fiscal Corruption," *World Development* 27(7): 1129-1140, 1999.

[14] Gérard CHAMBAS, "Afrique au Sud du Sahara : quelle stratégie de transition fiscale?," *Etudes et Documents*, Centre d'Études et de Recherches sur le Développement International CERDI, 2005.

[15] Tanzi V., "Uses and Abuses of Estimates of the Underground Economy," *Economic Journal*, 109; 456, pp. 338-340, 1999.

[16] Friedrich Schneider, Andreas Buehn, Claudio E. Montenegro, "Shadow Economies All over the World: New Estimates for 162 Countries from 1999 to 2007," The World Bank, Development Research Group, Poverty and Inequality Team & Europe and Central Asia Region, Human Development Economics Unit, July 2010.

[17] Friedrich Schneider, "The size and development of the shadow economy around the world and the relation to the hard to tax," Communication à la conférence "Hard to Tax : An International Perspective," *Andrew Young School of Policy Studies*, *Georgia State University,* Atlanta, Georgia, May 15-16, 2003.
[18] Chambas G., Araujo-B.C., "Taxer l'économie non enregistrée urbaine en Afrique Subsaharienne," Communication à la conférence «Hard to Tax », Atlanta, 2003.
[19] OCDE, "Études économiques de l'OCDE : Mexique," p.169, 2011.
[20] Chambas G., Brun J.F, Combes J.L., "Inégalité, pauvreté et mobilisation fiscale," Communication aux 6ème journées scientifiques du réseau d'Analyse Economique et Développement de l'AUF. Marrakech, 2004.
[21] Duchêne G., Zouari S., "Impact des institutions sur les crises de change."Économies et Sociétés, vol. 5-Série Monnaie, 2006.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 813-819

http://www.issr-journals.org/ijias/

Color Image Interpolation using Optimal Edge Detector

R. Jansi, Asnath Victy Phamila, and R. Amutha

Department of Electronics and Communication Engineering,
SSN College of Engineering,
Chennai, Tamil Nadu, India

ABSTRACT: This paper proposes an efficient image interpolation algorithm using optimal edge detector. The proposed interpolation algorithm is done in two steps. In the first step, the missing pixels with diagonal neighbors are interpolated and in the second step, the missing pixels with axial neighbors are interpolated. In both the steps classification of edge and smooth pixels is performed using canny edge detector. For the pixels classified as edge pixels, the direction of the edge is found using various structuring elements. The edge pixels are then interpolated along the directional orientation of the edges. The smooth pixels are interpolated using proportionate variation based interpolation technique in which more weights are assigned for the direction with minimum variation. The proposed interpolation algorithm is applied in the NTSC color space for color images. Conventional image interpolation algorithms like nearest neighbor and bi-cubic interpolation algorithms produces artifacts like edge blurring, zig- zag effects etc. The proposed interpolation algorithm produces super resolution images with improved image quality and less distortion. Experimental results show that in addition to the significant increase in visual effects, this algorithm also manifests improvements in quantitative analysis. Quantitative analysis is done using metrics like PSNR and correlation coefficient. From this analysis it is evident that the proposed algorithm performs better than conventional interpolation algorithms.

KEYWORDS: Canny, gradient, interpolation, resolution, structuring element, zoom.

1 INTRODUCTION

Interpolation in image processing is a method to increase the number of pixels in a digital image. Interpolation has been widely used in many image processing applications such as facial reconstruction, multiple description coding, and super resolution. Interpolation based super resolution has been used for a long time, and many interpolation techniques have been developed to increase the quality of this task [1]-[6]. There are three well known interpolation techniques namely, nearest neighbor interpolation, bilinear interpolation, and bi-cubic interpolation. Bi-cubic interpolation is more sophisticated than the other two techniques but produces smoother edges than nearest neighbor and bilinear interpolation.

Xin Li et al, proposed a hybrid approach for interpolation using switching between bilinear interpolation local covariance based adaptive interpolation. In this method local covariance was first calculated for the low resolution image. These covariance values were then used for interpolating the low resolution images using the property of geometric duality between the low resolution covariance and the high resolution covariance [7].

Zhou Dengwen et al, proposed a technique for color image interpolation. In this technique the edge pixels were distinguished from the smooth pixels based on a preset threshold value and then the pixels were interpolated using extended bi-cubic convolution algorithm [8].

Jinglun Shi proposed an algorithm for image interpolation using variation based approach. In this algorithm the pixels of low resolution image were projected using various angular orientations and the sum of absolute difference at various orientations were calculated. The pixels were then interpolated along the direction of minimum variation [9].

Corresponding Author: R. Jansi (jansi.rajamani@gmail.com)

In this paper a new approach for image interpolation is proposed. Here the edge pixels are first identified using canny operator. Then using various shapes of structuring elements the direction of edges are identified. Then the edge pixels are interpolated along the direction of minimal variation. The smooth pixels are interpolated using all the neighborhood pixels, with more weightage to the pixels in the direction of minimal difference.

The rest of this paper is organized as follows. Section 2 presents the proposed interpolation algorithm. Section 3 explains the interpolation algorithm for every layer of NTSC color space. Section 4 presents the experimental results and the conclusion is presented in Section 5.

2 PROPOSED COLOR IMAGE INTERPOLATION ALGORITHM

In the proposed method color image interpolation is done using the transformation from RGB to NTSC color space. Interpolation is performed in the NTSC color space and the inverse color transformation returns the interpolated image in the RGB color space. The various test images used for color image interpolation are shown in Fig. 1.

The algorithm for interpolation of every layer of NTSC color space is as follows:

1. Insertion of zeros after every row and column of the image to be interpolated.
2. Apply canny operator to the image after zeros insertion.
3. Classify the pixels as smooth and edge pixels.
4. Interpolate the missing pixels (pixels with the available diagonal neighbors) based on their angular orientation.
5. For the second iteration, repeat steps 1 to 4, to interpolate pixels with axial neighbors.

The same steps are repeated for all the three layers, and inverse transformation of NTSC to RGB conversion yiels the interpolated high resolution image as output.

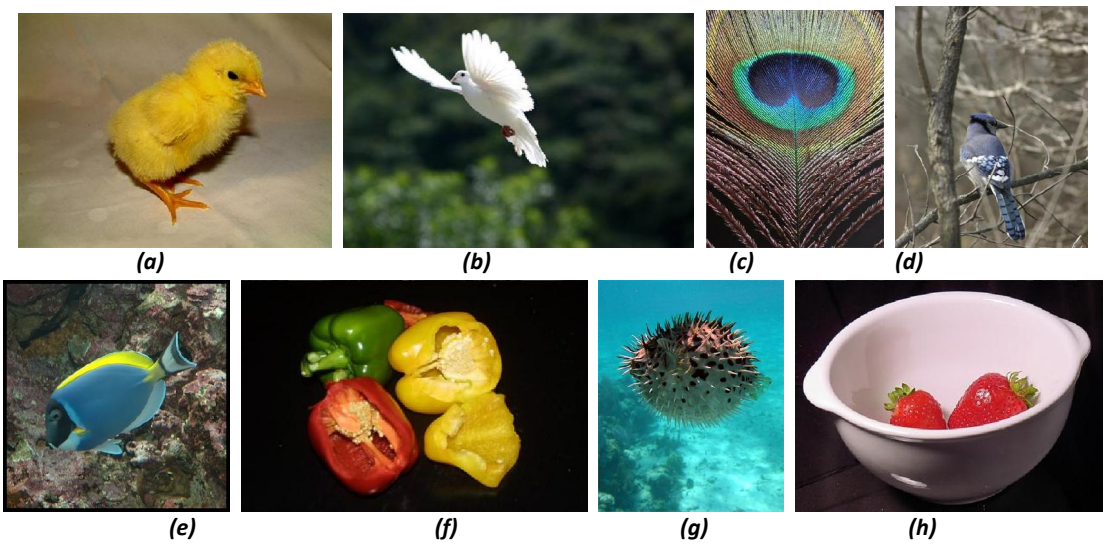


Fig. 1. Set of test images. (a) Chick; (b) Dove; (c) Feather; (d) Bird; (e) Fish; (f) Capsicum; (g) Aqua; (h) Strawberry

3 INTERPOLATION ALGORITHM FOR EVERY LAYER OF NTSC COLOR SPACE

The various steps explained below are performed for the interpolation of every layer of NTSC color space. The three interpolated layers are then converted back to RGB color space to get the interpolated color image. These transformations are performed using Matlab's 'RGB2NTSC' and 'NTSC2RGB' function.

3.1 ZEROS INSERTION

Zeros are inserted after every rows and column of the low resolution image. These zero values are interpolated using the proposed algorithm to yield a good quality high resolution image.

3.2 ITERATION – 1

3.2.1 PIXEL CLASSIFICATION

This section explains steps 2 and 3 of the above mentioned algorithm.

The pixels of gray scale low resolution images are first classified into edge and smooth pixels. Edge pixels are those pixels with abrupt change in the gray scale value. Smooth pixels are those pixels with gradual changes or change in the gray scale value. Two different methods are adopted for interpolating the edge and the smooth pixels.

The proposed algorithm for the classification of edge and smooth pixels is as follows:

- Insert zeros after every row and column of the low resolution image using (1).

$$Q(2i, 2j) = P(i, j) \quad (1)$$

  where *Q* represents the image to be interpolated and *P* represents the low resolution image.
- Apply canny operator (optimal edge detector) and extract the edge of the image *Q.* This gives the gradient image *G*1 as output.
- Apply the following 3×3 structuring element with the centre as the pixel to be interpolated to the gradient image *G*1*.*

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad (2)$$

- There are 3 cases possible

  Case 1. If the output of the structuring element is one of the following, the pixel is classified as an edge pixel oriented along 45^0.

$$\begin{bmatrix} 0 & * & 1 \\ * & * & * \\ 1 & * & 0 \end{bmatrix} \text{ (or) } \begin{bmatrix} 1 & * & 1 \\ * & * & * \\ 1 & * & 0 \end{bmatrix} \text{ (or) } \begin{bmatrix} 0 & * & 1 \\ * & * & * \\ 1 & * & 1 \end{bmatrix} \quad (3)$$

  Case 2. Else if the output of the structuring element is one of the following, the pixel is classified as an edge pixel oriented along 135^0.

$$\begin{bmatrix} 1 & * & 0 \\ * & * & * \\ 0 & * & 1 \end{bmatrix} \text{ (or) } \begin{bmatrix} 1 & * & 1 \\ * & * & * \\ 0 & * & 1 \end{bmatrix} \text{ (or) } \begin{bmatrix} 1 & * & 0 \\ * & * & * \\ 1 & * & 1 \end{bmatrix} \quad (4)$$

  Case 3. The pixels that do not satisfy the above two conditions are classified as smooth or non- edge pixels.

3.2.2 INTERPOLATION OF PIXELS WITH DIAGONAL NEIGHBORS

The edge pixels are interpolated along the direction of the edges and the smooth pixels are interpolated using all the neighborhood pixels. This section explains step 4 of the above mentioned algorithm.

An overlapping window of size 7×7 is moved over the image *Q* with the pixel '*X*' to be interpolated as the center. Fig. 2. represents the 7×7 window with the center pixel 'X' to be interpolated and its neighboring pixels P_1-P_{16}.

The edge pixels oriented along 45^0 are interpolated using (5).

$$X_{45^0} = \frac{-2 * P_{13} + 14 * P_{10} + 14 * p_7 - 2 * P_4}{24} \quad (5)$$

Similarly, the edge pixels oriented along 135^0 are interpolated using (6).

$$X_{135^0} = \frac{-2 * P_1 + 14 * P_6 + 14 * p_{11} - 2 * P_{16}}{24} \quad (6)$$

P_1	0	P_2	0	P_3	0	P_4
0	0	0	0	0	0	0
P_5	0	P_6	0	P_7	0	P_8
0	0	0	**X**	0	0	0
P_9	0	P_{10}	0	P_{11}	0	P_{12}
0	0	0	0	0	0	0
P_{13}	0	P_{14}	0	P_{15}	0	P_{16}

Fig. 2. A 7X7 window showing the center pixel, 'X' to be interpolated and its neighbors

The pixels which were classified as smooth pixels are interpolated using (7)

$$X = (\alpha * X_{45^0}) + (1-\alpha) * X_{145^0} \quad (7)$$

where,

$$\alpha = \frac{S_{135^0}}{S_{45^0} + S_{135^0}} \quad (8)$$

$S_{45}{}^0$ represents the sum of absolute difference of the 4 pixels along 45^0 and $S_{135}{}^0$ represents the sum of absolute difference of the 4 pixels along 135^0.

3.3 ITERATION - 2

In the second iteration, the pixels with the axial neighbors are estimated. This section explains step 5 of the proposed algorithm.

3.3.1 PIXEL CLASSIFICATION

The following algorithm is proposed for the classification of edge and smooth pixels in the second pass:

- The partially interpolated high resolution image, *R* obtained after the first iteration is again passed through the canny operator for edge detection. This gives the gradient image *G*2 as output.
- Apply the following 3×3 structuring element with the centre as the pixel to be interpolated to the gradient image *G*1.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (9)$$

- There are 3 cases possible

Case 1. If the output of the structuring element is one of the following, the pixel is classified as an edge pixel oriented along horizontal direction.

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & * & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{(or)} \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & * & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (10)$$

Case 2. Else if the output of the structuring element is one of the following, the pixel is classified as an edge pixel oriented along vertical direction.

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & * & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (or) \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & * & 1 \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

Case 3. The pixels that do not satisfy the above two conditions are classified as smooth or non- edge pixels.

3.3.2 INTERPOLATION OF PIXELS WITH AXIAL NEIGHBORS

Similar to iteration 1, the edge pixels are interpolated along the direction of the edges, and the smooth pixels are interpolated using all the available neighborhood pixels.

An overlapping window of size 7×7 is moved over the partially interpolated image *R*, with the pixel '*Y*' to be interpolated as the center. Fig. 3. represents the 7×7 window with the center pixel 'Y' to be interpolated and its neighboring pixels P_1-P_{25}.

The edge pixels oriented along horizontal direction or 180^0 are interpolated using (12).

$$X_{180^0} = \frac{-2 * P_{12} + 14 * P_{13} + 14 * p_{14} - 2 * P_{15}}{24} \tag{12}$$

Similarly, the edge pixels oriented along vertical direction or 90^0 are interpolated using (13).

$$X_{90^0} = \frac{-2 * P_2 + 14 * P_{10} + 14 * p_{17} - 2 * P_{24}}{24} \tag{13}$$

0	P_1	0	P_2	0	P_3	0
P_4	0	P_5	0	P_6	0	P_7
0	P_8	0	P_{10}	0	P_{11}	0
P_{12}	0	P_{13}	**Y**	P_{14}	0	P_{15}
0	P_{16}	0	P_{17}	0	P_{18}	0
P_{19}	0	P_{20}	0	P_{21}	0	P_{22}
0	P_{23}	0	P_{24}	0	P_{25}	0

Fig. 3. A 7X7 window showing the center pixel, 'Y' to be interpolated and its neighbors

The pixels which were classified as smooth pixels are interpolated using (14)

$$Y = (\beta * X_{180^0}) + (1 - \beta) * X_{90^0} \tag{14}$$

where,

$$\beta = \frac{S_{90^0}}{S_{90^0} + S_{180^0}} \tag{15}$$

$S_{90}{}^0$ represents the sum of absolute difference of the 4 pixels along 90^0 and $S_{180}{}^0$ represents the sum of absolute difference of the 4 pixels along 180^0.

4 EXPERIMENTAL RESULTS

The efficiency of the proposed interpolation algorithm was quantitatively measured using correlation coefficient and color peak signal to noise (CPSNR) ratio. The color images were decimated by a factor of 2 and the proposed interpolation

algorithm was applied and was the images were interpolated. The images were also interpolated using conventional interpolation algorithms like nearest neighbor and bi-cubic interpolation algorithm. The quantitative measurement metrics was calculated using the ground truth image and the interpolated images and the algorithms were compared. This comparison is shown in Table 1 and Table 2. From these tables it is evident that the proposed algorithm outperforms the conventional interpolation algorithms. Fig. 4. Shows Dove image interpolated by a factor of 2, using various interpolation algorithms.

Table 1. Comparison of Correlation coefficient

Test Images	Correlation Coefficient		
	Nearest Neighbor	Bi-Cubic	Proposed
Chick	0.786	0.882	0.924
Dove	0.794	0.957	0.975
Feather	0.912	0.922	0.963
Fish	0.876	0.893	0.957

Table 2. Comparison of CPSNR

Test Images	CPSNR		
	Nearest Neighbor	Bi-Cubic	Proposed
Chick	29.59	31.67	32.97
Dove	24.97	26.76	28.18
Feather	27.86	29.83	30.64
Fish	28.97	29.86	31.23

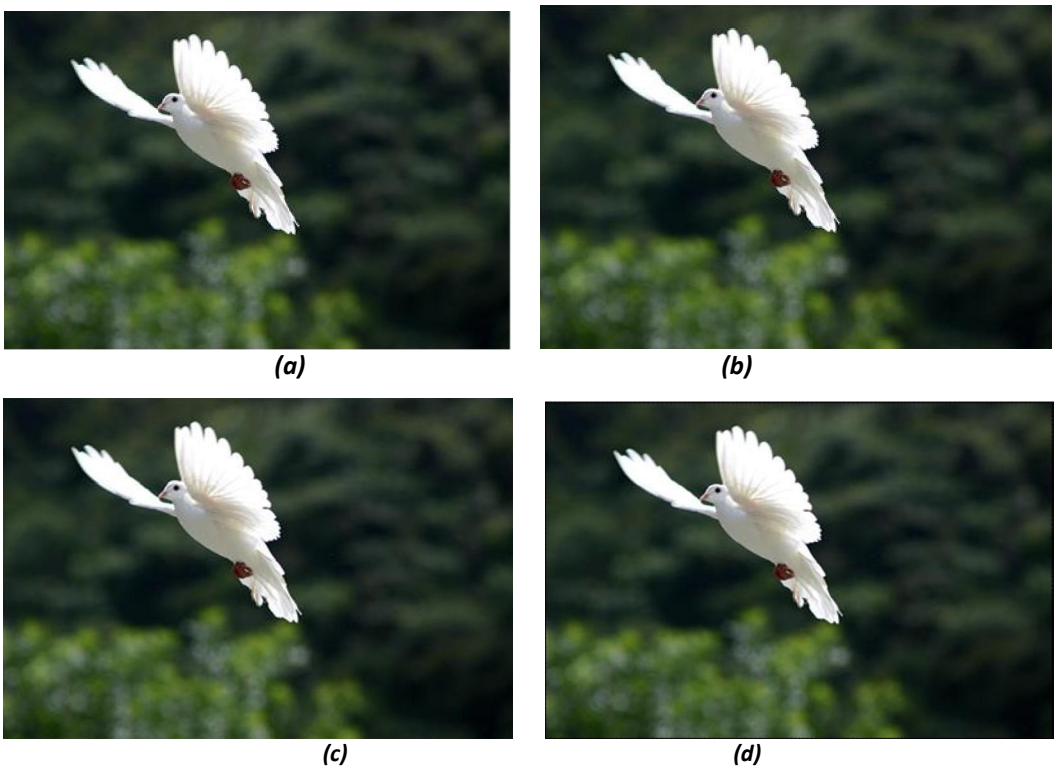


Fig. 4. 2X Zoomed Dove image. (a) Original; (b) Nearest neighbor; (c) Bi-cubic; (d) Proposed

5 CONCLUSION

The proposed color image interpolation algorithm produces good results in both visual and quantitative analysis. The edges of the images are interpolated along the direction of the edge and hence the interpolation artifacts are greatly reduced in the proposed method compared to other conventional interpolation techniques.

ACKNOWLEDGMENT

I profusely thank my supervisor Dr. R. Amutha, Professor, Department of Electronics and Communication, for her innovative ideas, keen interest, suggestions, guidance and co-operation that paved way for the successful completion of the paper.

REFERENCES

[1] Lei Zhang and Xiaolin Wu, "An Edge-Guided Image Interpolation Algorithm via Directional Filtering and Data Fusion," *IEEE Transactions on Image Processing,* vol. 15, no. 8, pp. 2226-2238, August 2006.

[2] Hakran Kim, Youngjoon Cha and Seongjai Kim, "Curvature Interpolation Method for Image Zooming," *IEEE Transactions on Image Processing,* vol. 20, no. 7, pp. 1895-1903, July 2011.

[3] Dong-Ho Lee, "A New Edge-based Intra-field Interpolation Method for Deinterlacing Using Locally Adaptive-thresholded Binary Image," *IEEE Transactions on Consumer Electronics,* vol. 54, no. 1, pp. 1110-115, February 2008.

[4] P. Longère, X. Zhang, P. B. Delahunt, and D. H. Brainard, "Pereceptual assessment of demosaicing algorithm performance," *IEEE proceedings*, vol. 90, no.1, pp. 123–132, January 2002.

[5] R. Ramanath, W. E. Snyder, and G. L. Bilbro, "Demosaicking methods for Bayer color arrays," *Journal of Electronic Imaging*, vol. 11, pp. 306–315, July 2002.

[6] R. Kimmel, "Demosaicing: image reconstruction from color CCD samples," *IEEE Transactions on Image Processing*, vol. 8, no. 9, pp. 1221–1228, Sept. 1999.

[7] Xin Li and Michael T. Orchard, "New Edge-Directed Interpolation," *IEEE Transactions on Image Processing,* vol. 10, no. 1, pp. 1521-1527, October 2001.

[8] Zhou Dengwen and Shen Xiaoliu, "An Effective Color Image Interpolation Algorithm," *International Congress on Image and Signal Processing*, Vol. 2, pp. 984-988, 2011.

[9] Jinglun Shi and Zhilong Shan, "Image Interpolation Using a Variation-Based Approach," *IEEE Proceedings*, 2011.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 820-828

http://www.issr-journals.org/ijias/

Qualitative analysis of land use change pressures, conditions and drivers in rural-urban fringes: A case of Nairobi rural-urban fringe, Kenya

Aggrey Daniel Maina Thuo

Department of Land Resources Planning and Management,
Jomo Kenyatta University of Agriculture and Technology,
P.O. Box 62000, -00200- Nairobi, Kenya

ABSTRACT: This paper is grounded on the truism that the planet's future is urban. However, urban growth process is bringing rapid economic, social and physical changes. These changes are bringing rapid transformations in areas concerned, especially on land uses in rural-urban fringes. While this is happening, the pressures and drivers are not well documented and understood, particularly so for rural-urban fringes in developing countries such as Kenya.

This paper is based on a qualitative research approach and used Nairobi rural-urban fringe as a case study in its attempt to document and analyse pressures and drivers of land use changes. The paper concludes that land use changes are contingent upon many pressures and drivers, primary of which is population increase through natural growth and immigration. The process that produces population growth is however a subset of the structuration processes that produces land use changes in the rural-urban fringe.

In suggesting ways of reconciling the likely to emerge multiple perspectives and differences in managing rural-urban fringes, the paper observes that there are always sufficient points of intersection to support dialogue and collaboration. However, this requires a strategy that looks for intersections among different positioning and rationalities and enters into a dialogue at such situated moments is needed. This entails going beyond scientific or technical forms of knowledge to involve emotional sensitivity and judgment, practical wisdom, ethics and deliberation that touches on values with reference to praxis.

KEYWORDS: Rural-urban fringe, land use, drivers, agriculture, urbanisation.

1 INTRODUCTION

The process of urbanisation is leading to rapid economic, social and physical changes particulary in developing countries [1]. As it is being obeserved in other parts of the developing world, the urban population in Africa (especially those in Sub-Saharan Africa) is expected to double by 2025 [2]-[3]. Urban growth will increase demand for land not only for housing but also for various other urban uses [21]. This land is not available withing cities' boundaries but in rural-urban fringes, Nairobi city is not an exception (Ibid). In many countries, the encroachment of rural-urban fringes is most likely to affect (or is affecting) these areas [1]-[7].

This paper is based on an inductive in-depth qualitative research that was meant to produce contextualized data through a case study based on Town Council of Karuri, (hereafter referred as TCK) (See Figure 1), representing Nairobi rural-urban fringe (henceforth refered as Nairobi fringe). The paper sought a qualitative understanding of a place (rural-urban fringe), and processes and perceptions associated with it. To achieve the study goal, in-depth interviews with diverse individuals and groups were done, documents reviewed and, casual and participants observations were also made.

Insights from an actor-oriented approach [13] were also borrowed to augment the idea of how actors make choices and how those choices lead to the shift of the structures and activities of various institutions over time. It is the actions of the actors that illustrate how they exercise their agency within existing structures to address the circumstances that local and

Corresponding Author: Aggrey Daniel Maina Thuo (athuo@jkuat.ac.ke)

extra-local conditions are creating in their areas. It is with this notion of agency in mind that this study sought to answer the question, 'why is agriculture being squeezed out by urban land uses in the Nairobi rural-urban fringe?'

This paper does not claim to have 'discovered' a full answer as to what are the drivers of urbanisation on Nairobi fringe. However, it does offer a window for partial understanding of urbanisation in the fringe and its drivers using a broad lens of a qualitative enquiry. On the other hand, it was not the intention of the paper to provide 'solutions' to (or a better way out of) urbanisation in the fringe problems as is common with studies done by urban planners' heeding to the John Friedmann's [8] call for planning to "attempt to link scientific knowledge to actions." This paper however, in general, sought to provide information that create an understanding of issues affecting urbanisation in Nairobi rural-urban (hereafter referred to as Nairobi fringe) in a way that can inform the process of policy development rather than proposing activities or normative prescriptions on how to address land use problem (*for more on planning theory see,* [10]-[12]).

The next section of this paper is structured along the broad headings which are solely meant to aid in the discussion on the drivers of urbanisation and not to portray issues addressed as mutually exclusive of each other. The paper appreciates that the drivers are interrelated and recursive. There is no single theory that provided an explanation on how different influences, factors and linkages could be established or identified. An approach which stretches across the twin poles of structure and agency was needed. Therefore to understand the various dimensions of land use/conversions and the links among them, I adopted a conceptual framework that focused on neo-classical, political economy, and structure and agency theoretical perspectives. Neo-classical and political economy theoretical perspectives could not provide insights into localised/individualised aspects of land use such as the agency of the landholders, therefore in developing my argument, I further drew insights from Giddens' theory of structuration [14] to conceptualise how an actors' agency interacts with structural determinants of change to shape the conditions for land use/conversions at the local level.

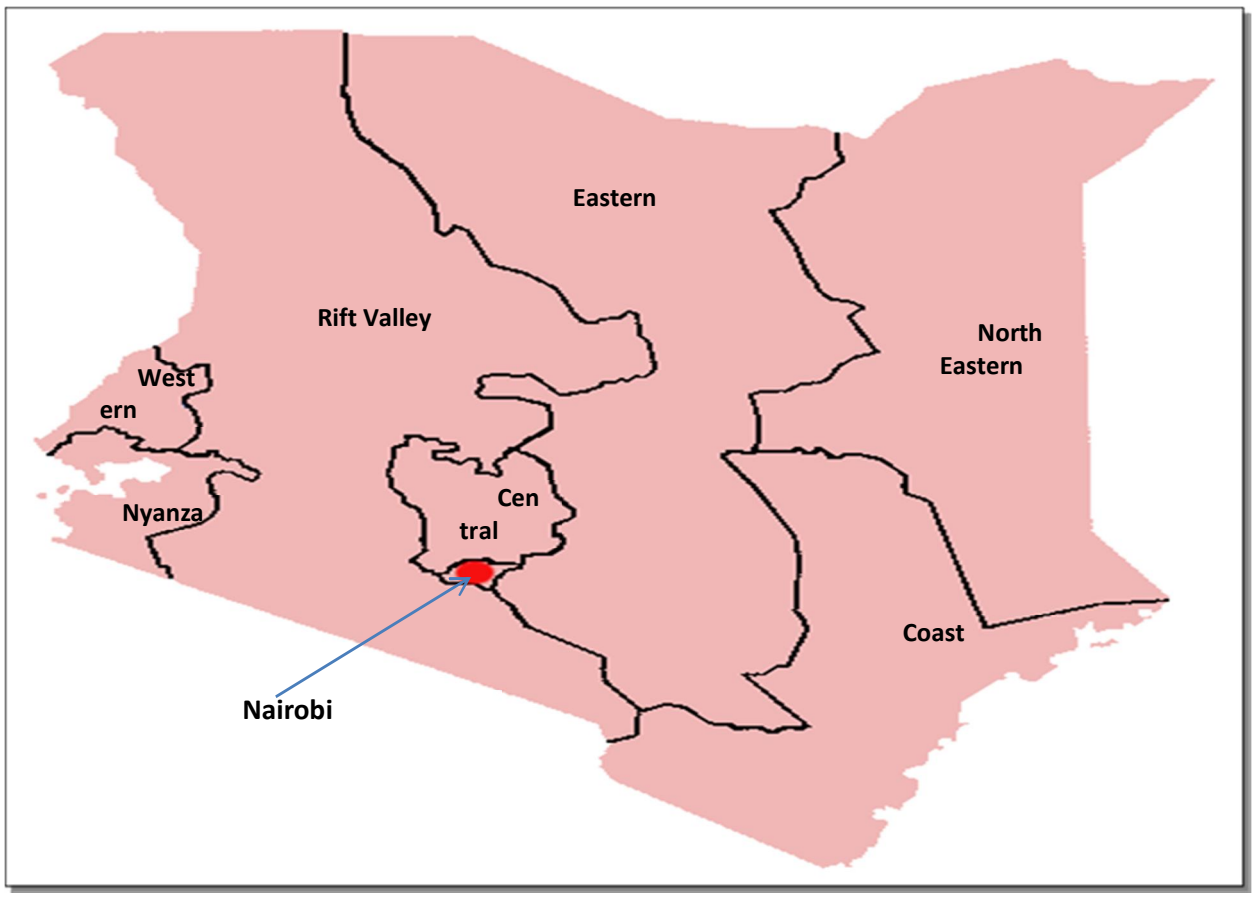


Fig. 1. Location of Nairobi in Kenya

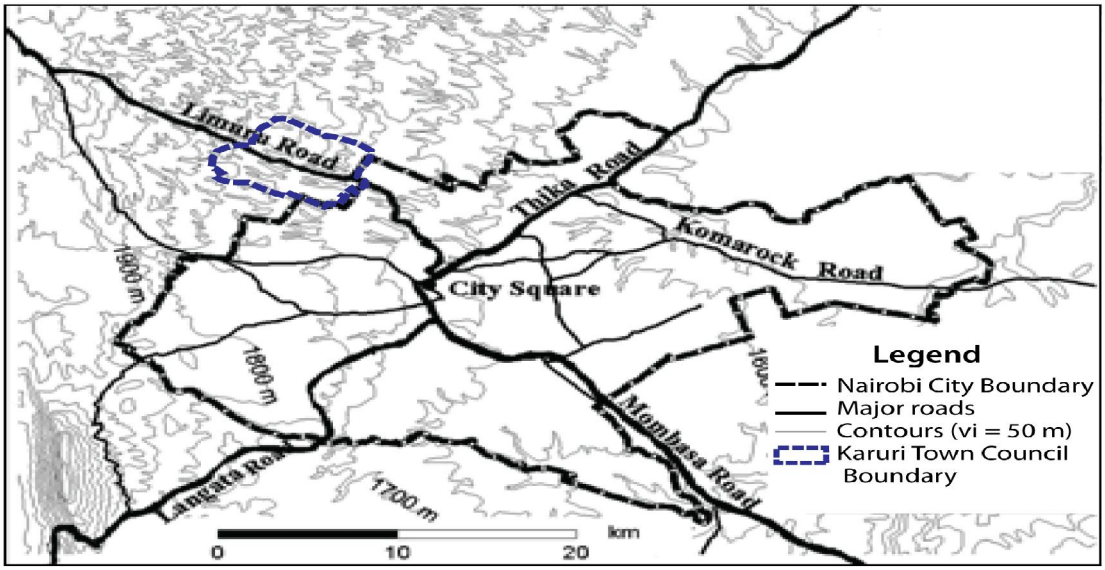


Fig. 2. The Town Council of Karuri in the context of Nairobi city [15]

2 LOCATING PRESSURES, CONDITIONS AND DRIVERS INFLUENCING LAND USE CHANGE IN THE NAIROBI FRINGE

On the surface of the above question, (why is agriculture being squeezed out by urban land uses in the Nairobi rural-urban fringe?), it may seem obvious and straightforward that this paper sets to identify various factors leading to land use changes in the Nairobi fringe. However, as this paper indicates, different aspects of land use interact in a contingent and recursive manner in the conversion of agricultural land to non-agricultural uses. It thus follows that, although this paper focuses on a seemly parochial case study based on a micro-scale, it shows that there are factors at macro scale that influence land use activities in the Nairobi fringe.

2.1 HOUSING AND LAND MARKET FAILURE

Lack of affordable housing due to high costs of land and non-provision of public housing in the Nairobi city is making people to resort to rural-urban fringe for housing. This can partly be blamed on colonial land ownership legacy which has lingered since independence. Also, the city inherited colonial building codes, standards and regulations which were partly meant to perpetuate restrictions to the native Africans from permanently staying in the city. These have thus created additional bureaucratic obstacles in housing. In regard to land there are problems of ascertaining ownership particularly where land lease titles has elapsed and political elites have taken advantage to illegally allocate such land. These have resulted to lack of transparency in land transactions within the city, further scaring potential homeowners away.

As a result of inadequate and unaffordable housing in the city and unclear land market, Nairobi rural-urban fringe, has become attractive site for those in need of land for housing. Here most land is owned through the *Mbari* (clan) system. Under the *Mbari* system, locally specific social norms and systems of trust operate in effecting land sale and transfers. These local practices are based on locally sedimented social, economic and cultural systems which are widely regarded and accepted by the community members and even those outside the community.

In the Nairobi fringe, there are also other aspects of flexibility when it comes to complying with the land use planning regulations. Majority of the land in the area is classified as agricultural and are thus covered under Ministry of Land and Agriculture where the role of the Department of Physical Planning is just advisory. This is in contrast to land use planning compliance requirements in the city which due to its lack of explicit urbanisation and housing policy and vagueness of the land market has given room to corruption and other non-civil behaviours among council officials and land dealers.

2.2 POPULATION INCREASE

At independence, Kenya saw an upsurge of urban population due to relaxed rural-urban migration policies. The migrants to the city came to look for better life and formal employment which were lacking in many rural areas. Due to unavailability

of formal jobs to accommodate the then growing numbers of job seekers, a good number of people joined the informal sector. Majority of these people are hardly eligible for the mortgage due to their non-regular income and also due to their fear of erratic banks interest rates. Owing to land problems within the city, these people find the Nairobi fringe as a preferable site to build their houses.

Increased population through natural growth and immigration into Nairobi fringe is creating new income opportunities for indigenous groups. These income opportunities are in the service sector such as repair, construction sector and ready market for the farm produce. The newcomers are also more exposed to the 'outside world' than the locals thus contact with them is bringing in new awareness of income opportunities outside the areas.

2.3 WEAK AND CONFLICTING INSTITUTIONAL REGULATIONS

Land use conversions in the Nairobi fringe are taking place in the absence of explicit land and urbanisation policy in Kenya. As result there are legal and jurisdictional overlaps among the institutions/departments involved in land use development control. This worsened by the weak local authorities in these areas (which in most cases were initially design to cater for rural interests). These local authorities lack the capacity and capability for managing rapid land developments taking place. They also lack upto date information which to reference decisions on plan approvals. This has created apathy among Planning Officers who have to rely on the goodwill of the developers to 'do good' in the implementation of their proposed land development activities.

Multiple governmental institutions/departments with separate goals and mandate on land, coupled with lack of planning guidelines such as Local Physical Development Plans, has created a jurisdictional ambiguities/vacuum. This vacuum is being misused especially by the Land Board members to grant approvals for the subdivision of agricultural land (though knowingly it is meant for residential purposes) without seeking the advice from the Physical Planning Office. The ambiguities in responsibilities and jurisdiction are also promoting corruption among officers involved in land development approvals. Even where attempts have been to provide planning guidelines, there has been poor involvement of diverse actors and thus denying such guidelines the legitimacy and widespread acceptance by the majority of land users. There is also political interference in land use planning and control which reduces the effectiveness and operations of the Physical Planning Office in enforcing land development control within the peri-urban areas.

There are also constraints to the land use control due to dual legal system whereby customary and formal land ownership systems co-exist. In most Nairobi fringe areas, customary land use practices, such as sub-division of land for inheritance among the family members, are predominant. When subdivision of land is done for family use residential purposes, it is leading to an aspect of *in situ* urbanisation.

2.4 ECONOMIC DRIVERS

The era of Structural Adjustment Programmes (SAPs) saw the liberalisation of export and import markets in Kenya. These created a class of business people within a short span of time. However due to the reduced donor support to the budgetary deficit, there was reduced investment in infrastructures necessary for the stimulation and location of industrial and manufacturing sectors. This meant that the 'new' businesswomen and men did not have many options where to re-invest their trade proceeds. Land and housing became the easy (if not the only) option thus the peri-urban areas became a preferred investment site.

Various arterial roads from the Nairobi city to other areas of the country pass through the peri-urban areas. These have made many of the areas easy to access by the use of public transport. This was particularly enhanced by the liberalisation of public transport sector which went further to reduce the costs and time of travel for those seeking to stay outside the city.

Presence of the Nairobi city has also affected labour and income aspects of the peri-urban areas in that in the city there exists a steady and high source income against that of agricultural sector which is seasonal. There is also a negative attitude, which is reproduced through an education system that castigated agriculture in favour of 'white collar' jobs. These have made people especially young people shuns farming in favour of urban-based employment. This leads to inadequate labour which consequently raises the level of farm work wages especially during the peak seasons, such as harvesting.

The reduced income from agriculture has also been occasioned by the implementation of SAPs that saw the costs of input increasing against the reduced earnings from the farm produce. Furthermore, the reduced earnings from farming and emergence of other land uses willing to pay high prices for the same land leads to the cost of land rising beyond the normal price for land between farmer to farmer. In addition, due to high demand for peri-urban land for residential purposes has

encouraged ingenuity among landholders who convert portions of their land to residential housing purposes or selling whole or part of their land in order to buy bigger but cheap parcels of land other (rural) areas.

Coffee growers' companies/cooperatives, which have been on the decline, are been subdividing their coffee estate among the members who then sell their plots or build rental housing blocks. There are also land owners who are subdividing their land parcels to cash in or to avoid restrictive zoning regulations, which are likely to be put in place, as a result of the prospective boundary extension of the Nairobi city. This proposal is creating a condition of 'impermanence' for those still practicing large-scale farming in the Nairobi fringe areas.

In addition, availability of the paid farm and non-farm labour in the city and within the peri-urban areas has led to the reduced availability non-paid family farm labour especially for the smallholders' farming households. Lack of non-paid family labour is affecting making smallholding farming system thus making such farmers prefer to subdivide sections of their land for sale or they themselves construct rental houses. There is however those who have stuck with their parcels of land due to cultural attachments and this explain why there is patchy residential development in the rural-urban fringe areas.

2.5 SOCIAL AND CULTURAL CHANGES

The entry of newcomers who are individual-oriented rather than communal-oriented has complicated the local settings in the rural-urban fringe. They are bringing new challenges which are complex than existing local social and institutional structures can handle such as crimes and prostitution. Non-alignment to the customary norms further make the newcomers treat land, which is cultural and symbolic good to the indigenous group, like any other commodity that can be traded on the market. The 'newcomers' values' are slowly being transmitted to the residents who see them as representing a more progressive system than their own. Once this happens especially to young people, they lose foothold in their families which make them vulnerable to vices such as crime and prostitution. Furthermore, other than loosing family connections, young people have no attachment to ancestral land as source of employment through agriculture and thus increased incidences of heirs selling their inherited parcels of land once their parents die.

2.6 ENVIRONMENTAL POLLUTION/CONFLICTS

Increase in the paved surfaces such as houses and roads have reduced storm water ground infiltration. The increased volume of storm water have been washing soils from building sites into water bodies where it is causing siltation of dams and river channels, thus flooding along the river valley. This is affecting small-scale horticultural farming along the valleys. Excessive pressure on land as result of land shortage is causing erosion of top soil that is suitable for crops cultivation. Mining of top-soils for sale to those growing flowers and building stone quarries are also reducing the amount of land available for farming.

Poor solid wastes disposal such as polythene papers is causing problems on livestock once ingested, and further clogging water drainage systems thus causing flooding. Dumping sites are also becoming breeding grounds for the pests such as rats and mice which are destroying crops in the farms or in stores.

Poor management of human waste is also posing a great threat to continued farming and public health. Due to the small sizes of the land, digging new pit latrines is becoming impossible. Once pit latrines are emptied, the waste collectors dump the waste into open grounds especially in the isolated farming areas or into the river channels.

In areas where farming and residential land uses are co-existing, conflicts are now becoming manifests. Such conflicts are due to complaints of the foul smell and flies from livestock stables. There are also complaints from the residents on the continued use of pesticides and acaricides by the farmers. On the side of non-farming residents, conflicts are also resulting from free range livestock and poultry destroying their flower gardens and orchards. Some of the residents are resorting to unorthodox means of spraying their orchards and flower gardens with chemicals to ward-off the livestock and poultry. This is further making it difficult for free range livestock and poultry keeping while caging them require additional financial resources to feed them thus making such enterprises/activities expensive and economically unviable in the long run.

2.7 LAND HOLDERS' RESPONSES AS A DRIVER

Land conversions are producing intended and unintended consequences in the Nairobi fringe. These consequences are leading to changes in these areas. Landholders are however not passively accepting their fate of being victims of these changes but instead they have been routinely monitoring local and extra-local circumstances affecting their surroundings.

They have with time evolved varieties of local/human-level responses to enable them live in a rapidly changing environment. The actions resulting from their responses are unintentionally creating an enabling conditions for further land conversions either through making the hitherto unfavourable areas for settlement becoming favourable for residential settlement or creating more obstacles for the continuation viable agricultural activities. It then follows that, land use conversions in the Nairobi fringe is as a result of both the actions of the landholders and outcomes of such actions.

3 DISCUSSION

This paper aimed at identifying, describing and examining the conditions and driving forces that influence land use in the Nairobi fringe using TCK as a case study. The paper noted that different processes influencing land use are interwoven (see Figure 3) in a way that the singular treatment of each influence can only hold for the purpose of clarity in discussion and not in actual terms. For example, some of the influences of the land use changes at the micro-level are not necessarily local but are a product of wider social, cultural, political and economic conditions that directly or indirectly affect the way land use operates in the area. Attempts were made to link the findings to the wider processes that play a role in influencing land use activities in the Nairobi fringe. Landholders' actions and perceptions were not taken as given but attempts were made to frame them within the wider structural forces at different levels.

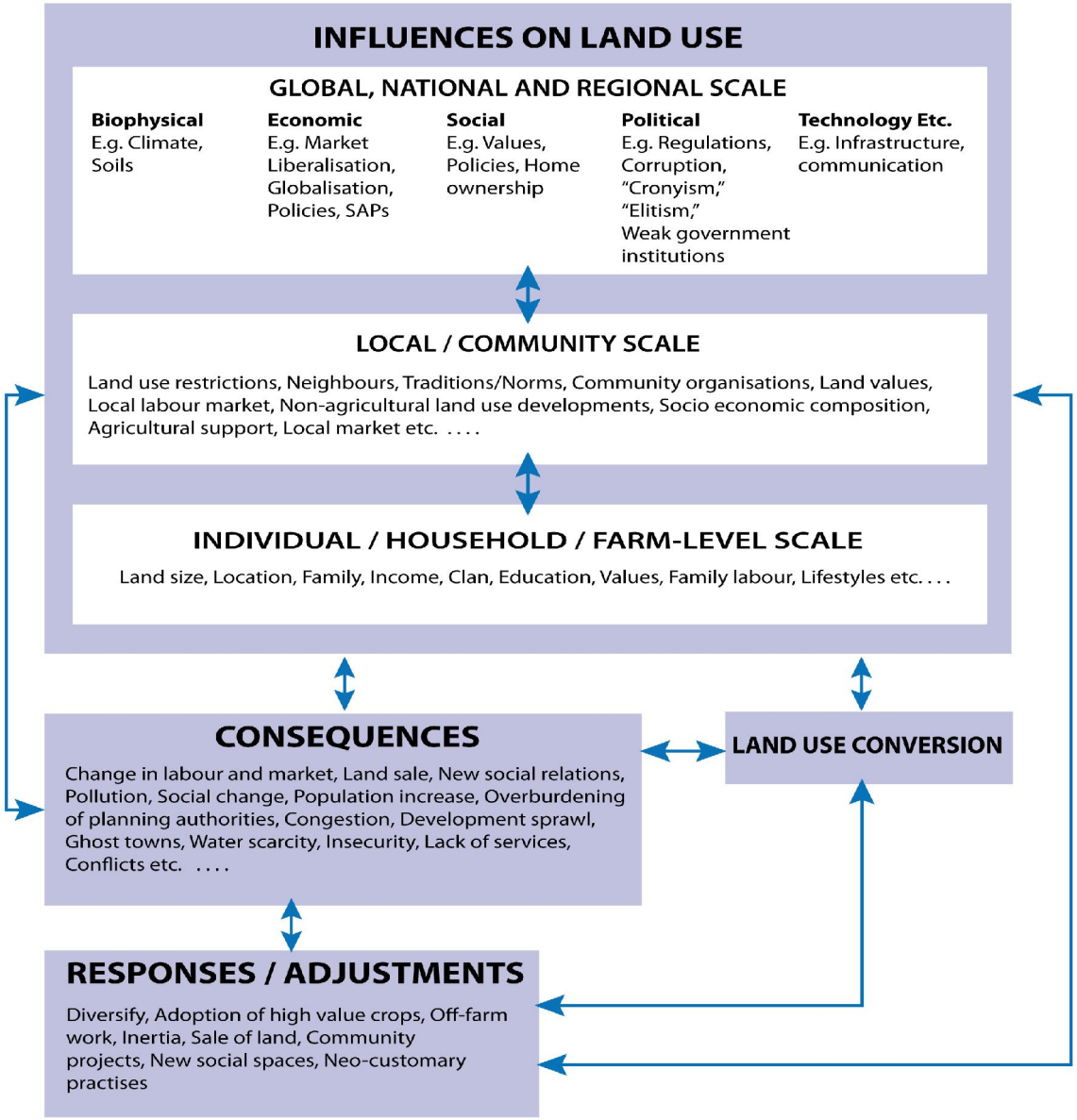


Fig. 3. Schematic diagram of the complex links in land conversion in the Nairobi fringe (Adapted from [16])

The genesis of land and housing problems within the City can be traced to pre-colonial and colonial period. After independence, the government adopted the colonial land legal system which was based on the English Common Law with some attempts to incorporate customary laws. This created a dual system of land legislation. Having different sets of laws dealing with the same land has created confusion and overlaps in its management. In addition, the institutions vested with the authority for land use planning have inadequate capacity to enforce the land laws. Developers, given the obstacles in obtaining land in the city, are exploiting the legal confusion and institutional weakness to access land for residential development in the Nairobi fringe. The confusion has also created room for corruption and other non-civil behaviour amongst government officers dealing with land.

At independence, Kenya saw an upsurge of urban population due to relaxed rural-urban migration policies. The migrants came to look for jobs in an already formal job-scarce environment. This created a need for more houses to accommodate people whose economic situation was not stable. The Nairobi City Council could not cope and therefore there was inadequate public housing. This was coupled with the inefficiency of the land and housing market which have resulted in lack of transparency in land transactions within the city and a consequent focus on developing land in the rural-urban fringe such as at TCK.

In general, as is identified in this paper, Kenya's economic situation has gone through turbulent moments which have in part affected how people (the majority whom rely on land for agriculture) have structured their livelihood. Changing economic circumstances have also affected performance of the government in terms of governance, services and infrastructure delivery. Both individual and governmental situations have created conditions that greatly influence land use in the Nairobi fringe. There are also influences coming from the implementation of SAPs and other neoliberal policies. These policies saw the liberalisation of export and import markets and reduced government investment in physical and social infrastructure. This affected farming due to increased costs of inputs against the reduction in income from agricultural produce. There was also reduced investment in housing within the city by the government.

Much of the Nairobi fringe is accessible by roads that traverse to the country-side. This coupled with cheap bus fares has made the fringe a preferred place for actors seeking self-built or rental houses. Proximity to the City has also affected labour and income aspects of the fringe. Steady City jobs against seasonal farm work have made many actors change their income focus from farming to urban jobs. There is also a negative attitude towards farming as a source of income and livelihood especially by young actors. The negative attitude to farming gets an additional boost from breakdown in traditional familial and societal ties that held the farming households in the fringe. Other than shortages of labour as result of attitude and better income available to young actors in the City, shortage of farm labour has also been worsened by the high incidences of HIV/AIDS, where able bodied actors are not able to optimally contribute their labour in the agricultural sector (which is still labour intensive) due to illness.

As the Nairobi fringe continues to be converted from agricultural to residential purposes, continued farming for the remaining farmers is being jeopardized. This is partially a result of pollution from household wastes which include solid and liquid wastes. There is also a problem of flooding and soil erosion due to reduced ground infiltration of surface water and also from siltation of rivers and dams from wastes from building sites. Soil erosion together with excess pressure on land as a result of intensive farming due to land shortage, is affecting farming in the area by reducing land productivity.

It is worth repeating again that it is not easy to separate the forces that influence land use in the Nairobi fringe. The attempt made to classify various factors and conditions into different sub-sections of this paper does not in any way show that they are working/can work independently of each other. They however interact in a contingent and recursive manner in conversion in influencing various decisions on land use.

4 CONCLUSION, LESSONS AND SUGGESTIONS

Following from the above discussions, it is therefore reasonable for this paper to conclude in regards to the central research question that the reason why agricultural land use is being edged out by non-agricultural uses in the Nairobi fringe is contingent on many factors/conditions, primary of which is population increase through natural growth and in-migration. Population growth thus is a necessary condition for land conversions from agricultural to residential use in the Nairobi fringe. The process that produces population growth is, however, a part of the structuration processes that produces land conversions in the Nairobi fringe.

I should however point out that population growth by itself does not cause land conversion in the Nairobi fringe. This argument is informed by an understanding that urban areas can have (or have had) population growth within controlled and designated zones. However, population growth comes with an increase in number and diversity of activities/actors and

extensive linkages. Increased number of actors/activities and linkages are, on the other hand, necessary in increasing the need for land for various purposes/uses.

Thinking about land use planning and management in areas such as Nairobi fringe, questions arise about how to reconcile multiple perspectives and differences that are likely to emerge from actors with diverse interests. Harrison [17] observed that within many settings actors with different traditions and values have historically intersected and continue to do so, and that there are always sufficient points of intersection to support dialogue and collaboration. The problem is not an "…inability to see across multiple positions but it is rather unwillingness to do so…" [17] by researchers and planners. To address the problem, "epistemic humility" a strategy that looks for intersections among different positioning and rationalities and enters into a dialogue at such situated moments is needed [17]. This entails going beyond scientific or technical forms of knowledge to involve emotional sensitivity and judgement, practical wisdom, ethics and deliberation that touches on values with reference to praxis [17- 20].

ACKNOWLEDGMENTS

I am grateful to Professor Lex Chalmers and Dr. Mairi Jay, both of The University of Waikato, New Zealand, for their comments on the project from which this paper is based.

REFERENCES

[1] C. Rakodi (Ed.), *The urban challenge in Africa: Growth and management of its large cities*, Tokyo; New York: United Nations University Press, 1997.
[2] UN- Habitat, *The state of the world's cities*, Nairobi, Kenya: UN- Habitat, 2001.
[3] Hall, P. and Pfeiffer, U., *Urban Future 21: A Global Agenda for Twenty-First Century Cities,* E&FN Spon, London, 2000.
[4] Aguilar, A.G., "Peri-urbanization, illegal settlements and environmental impact in Mexico City," *Cities*, vol. 25, no. 3, pp. 133-145, 2008.
[5] Aguilar, A.G., P. Ward, "Globalization, regional development and mega-city expansion in Latin America: Analysing Mexico city's peri-urban Hinterland," *Cities,* vol. 20, no. 1, 2003.
[6] C. Tacoli, "Understanding the opportunities and constraints for low-income groups in the peri-urban interface: The contribution of livelihood frameworks," *Paper produced for the research project on Strategic Environmental Planning and Management for the Peri-Urban Interface Research Project*, 1999.
[7] C. Tacoli, "Rural-urban interactions: A guide to the literature," *Environment and Urbanization*, vol. 10, pp. 147-166, 1998.
[8] J. Friedmann, "Planning in the public domain: From knowledge to action," *Princeton University Press*, 1987.
[9] D. Satterthwaite, *Urbanisation and third world cities*, In: Clark, D., *The Elgar companion to development studies*, Cheltenham: Edward Elgar Publishing, 2006.
[10] A. D. M. Thuo, "Tradition, culture and development in Africa: Historical lessons for modern development planning," *Book Review*, *Urban Policy and Research*, vol. 26, no. 1, pp. 125-139, 2008.
[11] M. Jay, "Symbolic order and material agency: A cultural ecology of native forest remnants on Waikato dairy farms," Unpublished PhD Thesis, *Department of Geography, Tourism and Environmental Planning, University of Waikato*, 2004.
[12] A. Faludi, *Planning theory*, Oxford: Pergamon press, 1973.
[13] N. Long, A. Long, "Battlefields of knowledge: The interlocking of theory and practice in social research and development," London: *Routledge*, 1992.
[14] A. Giddens, *The constitution of society: Outline of the theory of structuration*, Cambridge: Polity Press, 1984.
[15] C. Mundia, M. Aniya, "Dynamics of land use/cover changes and degradation of Nairobi City, Kenya," *Land Degradation & Development*, vol. 17, no. 1, pp. 97-108, 2006.
[16] J. Smithers, P. Johnson, *Development trajectories*, In: *The dynamics of family farming in North Huron County, Ontario*, The Canadian Geographer, vol. 48, no. 2, 191-208, 2004.
[17] P. Harrison, "On the edge of reason: Planning and urban futures in Africa," *Urban Studies*, vol. 43, no. 2, pp. 319-335, 2006.

[18] B. Flyvbjerg, *Making social science matter: Why social inquiry fails and how it can succeed again*, Cambridge University Press, 2001.

[19] L. Sandercock, "Commentary: Indigenous planning and the burden of colonialism," *Planning Theory & Practice*, vol. 5, no. 1, pp. 118 – 124, 2004.

[20] S. Thompson, *Diversity, difference and the multi-layered city*. In: Freestone, R., editor, *Urban Planning in a Changing World: The Twentieth Century Experience*, pp. 230, 2000.

[21] A.D.M. Thuo, "Community and social responses to land use transformations in the Nairobi rural- urban fringe, Kenya," *Journal of Field Actions Science* (*Field Actions Science Report)*, Special Issue 1, 2010.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 829-846

http://www.issr-journals.org/ijias/

A Survey of Vehicular Ad hoc Networks Routing Protocols

Marwa Altayeb[1] and Imad Mahgoub[2]

[1]Department of Computer Information & Networks,
College of Computer Science & Information Technology,
Sudan University of Science & Technology,
Khartoum, Sudan

[2]Department of Computer & Electrical Engineering and Computer Science,
College of Engineering & Computer Science,
Florida Atlantic University,
Florida, USA

ABSTRACT: In recent years, the aspect of vehicular ad hoc network (VANET) is becoming an interesting research area; VANET is a mobile ad hoc network considered as a special case of mobile ad hoc network (MANET). Similar to MANET, VANET is characterized as autonomous and self-configured wireless network. However, VANET has very dynamic topology, large and variable network size, and constrained mobility; these characteristics led to the need for efficient routing and resource saving VANET protocols, to fit with different VANET environments. These differences render traditional MANET's protocols unsuitable for VANET. The aim of this work is to give a survey of the VANETs routing mechanisms, this paper gives an overview of Vehicular ad hoc networks (VANETs) and the existing VANET routing protocols; mainly it focused on vehicle to vehicle (V2V) communication and protocols. The paper also represents the general outlines and goals of VANETs, investigates different routing schemes that have been developed for VANETs, as well as providing classifications of VANET routing protocols (focusing on two classification forms), and gives summarized comparisons between different classes in the context of their methodologies used, strengths, and limitations of each class scheme compared to other classes. Finally, it extracts the current trends and the challenges for efficient routing mechanisms in VANETs.

KEYWORDS: VANET, Route, Routing Protocols, Topology-based, Position-based, V2V.

1 INTRODUCTION

Vehicular Ad hoc networks (VANETs) are a special type of mobile ad hoc networks; where vehicles are simulated as mobile nodes. VANET contains two entities: access points and vehicles, the access points are fixed and usually connected to the internet, and they could participate as a distribution point for vehicles [1]. VANET addresses the wireless communication between vehicles (V2V), and between vehicles and infrastructure access point (V2I). Vehicle to vehicle communication (V2V) has two types of communication: one hop communication (direct vehicle to vehicle communication), and multi hop communication (vehicle relies on other vehicles to retransmit). VANET also has special characteristics that distinguish it from other mobile ad hoc networks; the most important characteristics are: high mobility, self-organization, distributed communication, road pattern restrictions, and no restrictions of network size [2]-[4], all these characteristics made VANETs environment a challenging for developing efficient routing protocols.

VANETs applications types are classified into safety and efficiency application [1], [5], [6]. There are many difficulties facing VANETs systems design and implementation, including: security, privacy, routing, connectivity, and quality of services. This paper will focus on routing problem in vehicle to vehicle communication (V2V); discusses some proposed routing solutions, routing protocols classifications, and illustrates some challenges and open issues in VANET routing.

Corresponding Author: Marwa Altayeb (mamdrw@gmail.com)

The main goal for routing protocol is to provide optimal paths between network nodes via minimum overhead. Many routing protocols have been developed for VANETs environment, which can be classified in many ways, according to different aspects; such as: protocols characteristics, techniques used, routing information, quality of services, network structures, routing algorithms, and so on. Some research papers classified VANETs routing protocols into five classes: topology-based, position-based, geocast-based, broadcast, and cluster-based routing protocols, this classification is based on the routing protocols characteristics and techniques used [2], [5], [7]. As well, other papers classified VANETs routing protocols according to the network structures, into three classes: hierarchical routing, flat routing, and position-base routing. Moreover, they can be categorized into two classes according to routing strategies: proactive and reactive [8]. On the other hand other papers classified them into two categories: geographic-based and topology-based, according to the routing information used in packet forwarding [4]. Also based on quality of services classification, there are three types of protocols that dealing with network topology (hierarchical, flat, and position aware), concerning with route discovery (reactive, proactive, hybrid and predictive), or based on the MAC layer interaction [9]. However all previous classifications did not concern by transmission strategies classification (such as unicast, broadcast, and multicast).

This paper will address two types of classifications as shown in Fig. 1; the first one is the routing information which used in packet forwarding, it mainly focuses on topology-based and graphic-based routing. And the other class is the transmission strategies, which is we thought it has a significant impact in protocol design and network performance (in case of network overhead, delay, and packet loss).

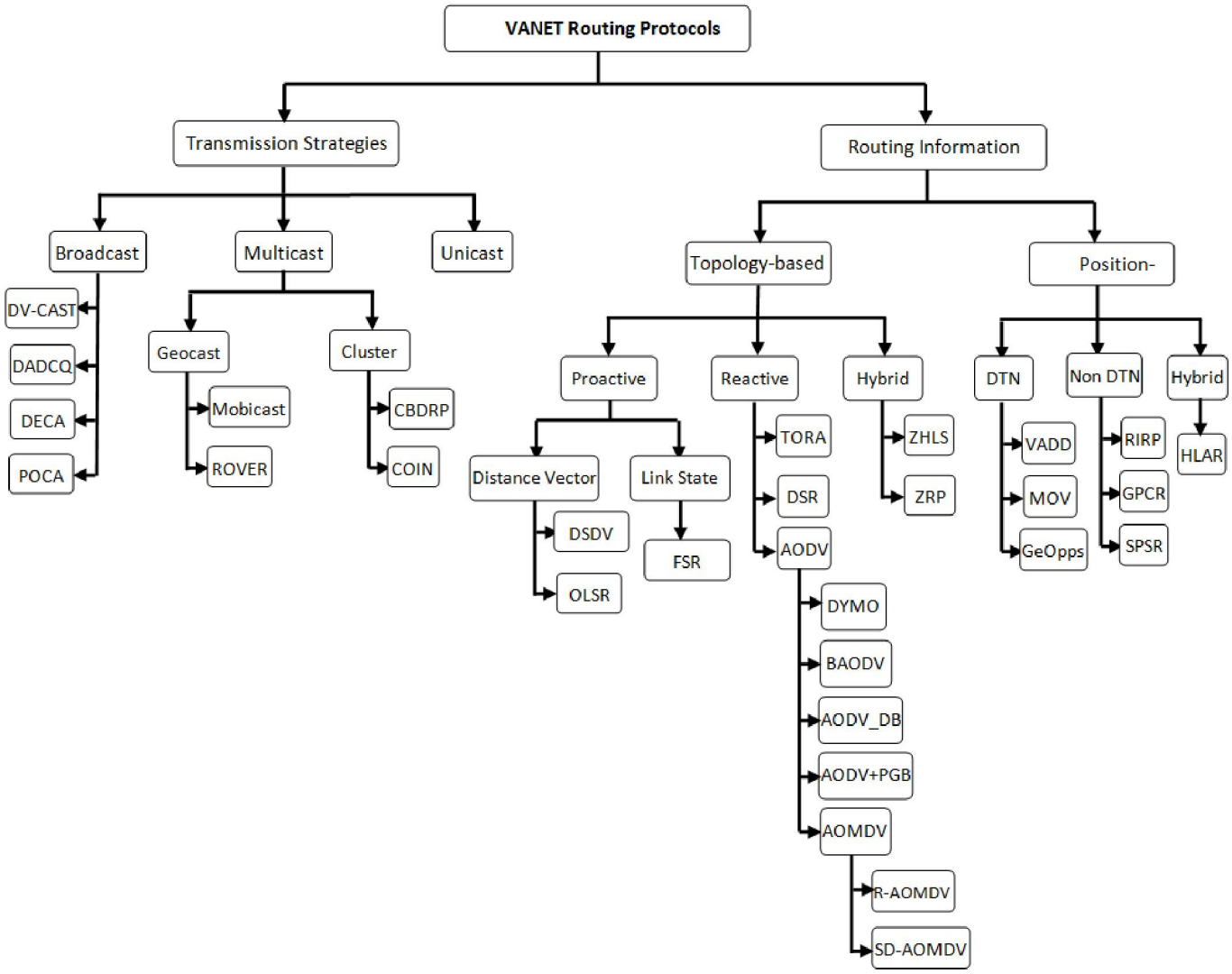


Fig. 1. Classification of VANET routing protocols

The rest of the paper is organized as follows: section 2 presents class one: the routing information used in packet forwarding, discusses topology-based and position-based routing protocol, and illustrate some related protocols with a brief showing to their strengths and limitations, and comparisons between the different class types. Section 3 just like section 2, represents the second class types; transmission strategies, shows various categories of routing protocols, and discusses some related protocols along with their strengths and limitations, also gives brief comparisons between the different categories. Section 3 discusses some research area and open issues in VANETs. And final section summarizes and concludes this paper.

2 ROUTING INFORMATION USED IN PACKET FORWARDING

This class is divided into two subclasses: topology-based and position-based routing protocols. In topology-based routing, each node should be aware of the network layout, also should able to forward packets using information about available nodes and links in the network. In contrast, position-based routing should be aware of the nodes locations in the packet forwarding.

2.1 TOPOLOGY-BASED ROUTING PROTOCOL

Topology-based routing protocol usually a traditional MANET routing protocol, it uses link's information which stored in the routing table as a basis to forward packets from source node to destination; it commonly categorized into three categories (base on underlying architecture) [3],[10]: Proactive (periodic), Reactive (on-demand) and Hybrid

2.1.1 PROACTIVE ROUTING PROTOCOLS

Proactive protocols allow a network node to use the routing table to store routes information for all other nodes, each entry in the table contains the next hop node used in the path to the destination, regardless of whether the route is actually needed or not. The table must be updated frequently to reflect the network topology changes, and should be broadcast periodically to the neighbors. This scheme may cause more overhead especially in the high mobility network. However, routes to destinations will always be available when needed [4]. Proactive protocols usually depend on shortest path algorithms to determine which route will be chosen; they generally use two routing strategies: Link state strategy and distance vector strategy.

2.1.1.1 DESTINATION SEQUENCE DISTANCE VECTOR ROUTING (DSDV)

DSDV protocol it is an earliest ad hoc routing protocol, it implements the distance vector strategy and uses a shortest path algorithm to implement only one route to destination which stored in the routing table, each routing table contains information about all accessible network nodes, as well as the total number of hops needed to reach these nodes, and each entry in the routing table is labeled with a sequence number initiated by the destination node. To maintain routes reliability, each node must periodically broadcast its routing table to its neighbors. DSDV protocol guarantees the loop free routs, excludes extra traffic caused by frequent updates, as well as reduces control message overhead, it also keeps only the optimal path to every node, rather than keeping multi paths which will help to reduce the total size of routing table [8]. However, DSDV increases the overhead in the large network; because of unnecessary updating broadcast even if there is no change in the network topology. Besides that, DSDV don't provide multi routes to destination node [8] and has no control over the network congestion which decreases the routing efficiency [11]. As the result of these limitations, Randomized DSDV protocol (R-DSDV) is proposed to support congestion control over DSDV; by maintaining nodes randomized decision which allows each node to make a decision whether to forward or discard a packet. However the R-DSDV produces more overhead compared to the DSDV protocol.

2.1.1.2 OPTIMIZED LINK STATE ROUTING PROTOCOL (OLSR)

OLSR protocol implement the link state strategy; it keeps a routing table contains information about all possible routes to network nodes. Once the network topology is changed each node must send its updated information to some selective nodes, which retransmit this information to its other selective nodes. The nodes which are not in the selected list can just read and process the packet [10].

Some researchers thought that OLSR has easy procedure which allows it to built-in different operating systems, besides it works well in the dynamic topology, also it is generally suitable for applications that required low latency in the data transmission (like warning applications) [11]. However, OLSR may cause network congestion; because of frequent control

packets which sent to handle topology changes, moreover OLSR ignore the high resources capabilities of nodes (like transmission range, bandwidth, directional antenna and so on) [12]. Therefore, some researchers propose Hierarchical Optimized Link State Routing (HOLSR) protocol as enhancement of the OLSR protocol, which decreases routing control overhead in the large size networks, also maximizes the routing performance; by the defining network hierarchy architecture with multiple networks [13]. Also some researchers propose QOLSR as a solution of providing a path such that the available bandwidth at each node on the path is not less than the required bandwidth. QOLSR considers delay as a second for path selection [12]. These protocols usually provide average enhancement for the QoS of packets. However, they cause more complexity, increasing packet overhead, and only suitable for some limited applications [9].

2.1.1.3 FISHEYE STATE ROUTING (FSR)

In FSR, the node periodically updates its table based on the latest information received from neighboring nodes. The updating of the routing table entries that concern a certain destination must be broadcast by different frequencies for neighbors. Table entries that are further in the distance are broadcast with lower frequency than entries that are nearer, this scheme doesn't guarantee decreasing broadcast overhead in large distances routing process. However, it could be accurate, if the packets come closer to the destination [4], [14]. The problem with the FSR is that, the growing network sizes will also increase the routing tables, also if the topology changes increased, the route to a remote destination becomes inaccurate. Moreover if the destination moves out of scope of source node then it can not discover the route [4], [15].

The advantage of proactive routing protocols can be abbreviated to there is no need to route discovery process; because the route to the destination is kept in the background, moreover proactive protocols periodically update the routing information which lets these protocols to perform well in low mobility networks. However, they have degraded performance in highly mobility and density network that when compare them with the reactive routing protocols, moreover unused routes consume the available bandwidth and increase the network overhead [2].

Recent studies show that proactive routing protocols (such as OLSR) generally outperform the reactive protocols in terms of network throughput and end to end delay [16]. However; there is no much research in the proactive routing protocols for VANET compared with existing VANET reactive protocols researches.

2.1.2 REACTIVE ROUTING PROTOCOLS

Reactive routing protocols (also called on-demand) reduce the network overhead; by maintaining routes only when needed, that the source node starts a route discovery process, if it needs a non existing route to a destination, it does this process by flooding the network by a route request message. After the message reaches the destination node (or to the node which has a route to the destination), this node will send a route reply message back to the source node using unicast communication [17]. Reactive routing protocols are applicable to the large size of the mobile ad hoc networks which are highly mobility and frequent topology changes [18]. Many reactive routing protocols have been developed, the following sections will illustrate characteristic of some reactive protocols, as well as illustrates the existing enhancement protocols.

2.1.2.1 AD HOC ON-DEMAND DISTANCE VECTOR (AODV)

AODV routing protocol is proposed for mobile ad hoc network, it has been evaluated in several researches and shows good results compared to related routing protocols; so it has a good documentation [19]. AODV offers low network overhead by reducing messages flooding in the network; that when compared to proactive routing protocols, besides reducing the requirement of memory size; by minimizing the routing tables which keep only entries for recent active routes, also keeps next hop for a route rather than the whole route. It also provides dynamically updates for adapting the route conditions and eliminates looping in routes; by using destination sequence numbers. So AODV is flexible to highly dynamic network topology and large-scale network [20]. However, it causes large delays in a route discovery, also route failure may require a new route discovery which produces additional delays that decrease the data transmission rate and increase the network overhead [17]. Moreover, the redundant broadcasts without control will consume extra bandwidth (broadcast storm problem), this problem grows as the number of network nodes increases, that besides collisions which lead to packet lost problem [19]. There are several protocols have been proposed to enhance AODV protocol; by decreasing its problems.

2.1.2.1.1 AD-HOC ON-DEMAND MULTIPATH DISTANCE VECTOR ROUTING (AOMDV)

AOMDV protocol is a multi path on-demand protocols comes as an extension of the AODV protocol, it discovers many paths from source to destination in a single route discovery process. Multi path On-demand protocols perform better than

single path protocols especially in reducing route discovery retransmission. It stores multi paths to destination using a single route discovery process; therefore no need to discover a new route if only a single path is failing; it's easy use any one of the existing redundancy paths. In multi path protocols a new route discovery is required if and only if all replicated paths to the destination are failing. In contrast, different thing happens in the single path protocols which establish a new route discovery every time a single path from the source to the destination is failing. This made multi path protocols have a better performance in term of uninterrupted communications for the packet transmission, and provide lower overhead; due to decreasing frequent route discovery transmission. AOMDV protocol is an enhancement of AODV protocol, it uses same control messages used in AODV, it just adds extra fields for AODV routing control messages; that to reduce the overhead occurring by discovering multiple paths. Moreover discovering multiple paths doesn't increase the delay of the route discovery process; because the latency of a route discovery is defined by the total source waiting time, before the source received the first path. AOMDV keeps all available paths in the routing table, then the source chose one of the stored paths; the preferred path will be the first established one [21].

2.1.2.1.1.1 ENHANCING AOMDV ROUTING PROTOCOL FOR V2V COMMUNICATION (SD-AOMDV)

SD-AOMDV is proposed as an enhancement of the AOMDV protocol; to deal with VANET characteristics. SD-AOMDV appends new factors (speed and direction) to the hop count field to determine the next hop during the rout discovery process. The next hop node is an intermediate node, selected based on two factors: intermediate nodes that can move in the same direction of the source and the destination, or the same direction of the source, or the same direction of the destination, and its speed is equal or near to the average speed of the source and destination. The protocol merges these two factors with a hop count field to choose a route [22].

2.1.2.1.1.2 A CROSS-LAYER AOMDV ROUTING PROTOCOL FOR V2V COMMUNICATION IN URBAN VANET (R-AOMDV)

R-AOMDV is a protocol built on AOMDV; it uses a method which merges transmission count and hop counts at the MAC layer, taking into account minimizing delay and performance of intermediate links. In the route discovery process, R-AOMDV is similar to AOMDV; it depends on route request and route reply control packets. This protocol adds two fields to route replay message fields; to compute quality of the whole path, one of these fields is the maximum retransmission count (MRC) which is computed by the MAC layer, and the other is the total hop count which is computed by the network layer. In R-AOMDV, a source node sends a route discovery packet when it hasn't a path to the required destination, stored in its route table. When a route replay packet received to the source, an intermediate node updates its retransmission count value; in case if it was greater than the current MRC. So, when the route replay packet arrives to the source, the source can identify which path has maximum MRC. R-AOMDV protocol inherits all good properties of multi path routing protocols, such as reducing rebroadcast route discovery. This protocol shows better performance than AOMDV in both rural and city vehicular networks [22]-[23], as well as enhances the routing operations by getting information about route's quality based on the neighbors IP addresses. However, the technique based on IP addresses is not convenient for VANET; because it sends the packets to nodes IPs, even though they change their locations; in this case the source node must search for a new intermediate forwarding node; and as a result this may lead to increase the packet delays and packet loss. This problem has largely appeared in city vehicular networks which have multiple paths, several intersections, different node density, and high congestion [23].

2.1.2.1.2 THE DYMO ROUTING PROTOCOL IN VANET SCENARIOS

Dynamic MANET On-demand (DYMO) protocol is a reactive multi hop routing protocol. Like AODV protocol, it uses the sequence numbers to provide loop free routes; it also has two essential processes: Route Discovery and Route Maintenance. DYMO is different from AODV protocol in other characteristic; that in DYMO a new route request process has to maintain information about all intermediate nodes, however in AODV, it just collects information about the destination node and the next hop, moreover in DYMO, every node which participates in a recent route discovery process should collect information about a requested node and all other intermediate nodes in the new path. Particularly at higher node densities, which commonly occurred in VANETs, routing protocols and transport protocols may increase the network overhead. When establishing a new route is required in the congested network, the use of the simple retry mechanism will only cause furthered congestion [24].

2.1.2.1.3 AODV_DB

It will be a challenge in VANET if a route was failing; according to this, AODV protocol spent a long time in discovering a new route; by sending a route request and route replay messages. However, VANET topology is highly dynamic change, which requires an establishment of a new route rediscovery. The route caching scheme that used by the on-demand protocol has a bad performance in the high mobility environment. In this case the using of flooding mechanism is the more suitable for routes maintenance; however it produces many redundant messages and network overhead. AODV_BD discovers and maintains a route by broadcasting the data packet instead of the route request packet. The data packet header contains the reverse path, so intermediate nodes will store the reverse path and rebroadcast the data packet, finally the destination will receive the data packet and at the same time send the route replay to the source. This approach will reduce the time of route setup packet transmission [25]. However it leads to heavy network overhead by flooding data packet broadcast. The problem arises if there is no route to destination.

2.1.2.1.4 AD HOC ON-DEMAND DISTANCE VECTOR PREFERRED GROUP BROADCASTING (AODV+PGB)

This protocol enhances the AODV protocol by Preferred Group Broadcasting (PGB) algorithm, this algorithm aims to reduce control message overhead in addition to offer routes availability which is an important feature in VANET environment, as the reducing routing overhead is a significant issue in ad hoc networks, also the routes consistency is a desirable issue in fast moving environment. There are many issues that critically decreasing ad hoc network performance can be abbreviated in [19]:

- The problem of hidden terminal which arises if the signal from the source to the destination is weaker; this makes easy to interrupt the communication between two nodes by a hidden terminal.
- No particular scheme is used to select intermediate hops. A large number of hops involve the short distance selection; however the link can simply be fail if one of the intermediate nodes goes out of the range, otherwise the weak signal may be changed.
- The larger numbers of errors may reduce the quality of links; which lead to decreasing network throughput.
- Also if the data transmission rate is adapted according to the network congestion, it could be affected by the large data error rate decreases the data transmission rate and may cause a bottleneck in the current node.

PGB tried to deal with all these issues via permits some particular nodes to re-broadcast a route request packet. However, if the node that allowed rebroadcasting the route request is not the nearest node to the destination, then the route discovery could be longer than it should. Also broadcast can be halted if there is no specific node which had a rebroadcast permit (case in light networks). Moreover packet duplication may occur if any two nodes rebroadcast the same packet at the same time.

2.1.2.1.5 THE BUS AD HOC ON-DEMAND DISTANCE VECTOR (BAODV)

BAODV focused on extending the AODV protocol to make it more suitable for VANETs, it has been designed to select a route with the minimum number of vehicles (buses), and it minimized end-to-end packet delay in the network. The protocol also allows drivers to avoid congested roads which contain a large number of buses. The AODV protocol is not an efficient VANET protocol; it has a limitation on when it works for VANET; which mainly shows in the frequent routing table updates. The main idea in the designing BAODV is to overcome AODV protocol limitations that made BAODV works on sensing of the vehicle characteristics and behaviors. In the AODV protocol, the update occurs according to two important factors: the newer sequence number and the lower hop count. But BAODV considers the lower hop count factor alone (mean lower number of vehicles) is insufficient for many VANET environments, so it uses additional factors like vehicle type and behavior to achieve better routing performance. The BAODV protocol modified ADOV messages: RREP and RREQ; by adding new fields defines the number of vehicles (bus, car, trucks, Trailers and so on) in the route by adding new fields in order to calculate the number of buses on the route. This information should be stored in the routing table; thus it required modifying the routing table also. The simulation results show the BAODV protocol performed better than the traditional AODV in terms of end-to-end packet delay in city areas which is a significant factor in VANETs, particularly for warning messages. The protocol also enables drivers to choose preferable routes that help them in avoiding congestion [26]. However, there is a shortcoming of BAODV documentation, it didn't illiterate carefully the modification method used in the routing table and what is the algorithm used to select a suitable route. Also the protocol focused on solving end-to-end delay, but neglected the frequent changeable topology, so the selection of routes based on the less hop count may not work properly in long distance between nodes, because if any intermediate node moved out of the range, the route will break, also it may cause the problem of weak signal.

2.1.2.2 DYNAMIC SOURCE ROUTING PROTOCOL (DSR)

DSR protocol aims to provide a highly reactive routing process; by implementing a routing mechanism with an extremely low overhead and fast reaction to the frequent network changes, to guarantee successful data packet delivery regardless of network changes. DSR is a multi hop protocol; it decreases the network overhead by reducing periodic messages. This protocol has two main processes: route discovery and route Maintenance. In the route discovery, when a source node needs an unavailable route, it initially broadcasts a route request message. All intermediate nodes which received this message will rebroadcast it, except if it was the destination node or it has a route to the destination; in this case the node will send a route replay message back to the source, later the received route is cashed in the source routing table for future use. If a route is failing, the source node will be informed by a route error message. In DSR protocol, every data packet contains a complete list of the intermediate nodes; so the source node should delete the failed route from its cache, and if it stores other successful route to that destination in its cache, it will exchange the failed one by the other successful route. But if there is no alternative route, it will initiate a new route discovery process [27]. The benefit of DSR protocol is clearly shown in a network with low mobility; because it can use the alternative route before starts a new process for route discovery. However, the multi routes may lead to additional routing overheads by adding all route information to every data packet, besides, as the network span larger distance and including more nodes, the overhead will frequently increase and as result network performance will be degraded [28].

2.1.2.3 TEMPORALLY ORDERED ROUTING ALGORITHM (TORA)

TORA is a distributed routing protocol using multi hop routes; it is designed to reduce the communication overhead related to adapting frequent network changes. This protocol does not implement a shortest path algorithm; thus the routing structure does not represent a distance. TORA constructs a directed graph which contains the source node as the tree root. Packets should be running from higher nodes to lower nodes in the tree. Once a node broadcasts a packet to a particular destination, its neighbor will broadcast a route replay if it has a downward link to the destination, if not, it just drops the packet. TORA ensures multi path loop free routing; since the packet always flows downward to the destination and don't flow upward back to the sending node [29]. The advantages of TORA are that it offers a route to every node in the network, and reduces the control messages broadcast. However, it causes routing overhead in maintaining routes to all network nodes, especially in highly dynamic VANETs [4], [15].

2.1.3 HYBRID ROUTING PROTOCOLS

Hybrid protocol is a mixture of both proactive and reactive protocols; it aims to minimize the proactive routing protocol control overhead and reduce the delay of the route discovery process within on-demand routing protocols. Usually the hybrid protocol divides the network to many zones to provide more reliability for route discovery and maintenance processes. Each node divides the network into two regions: inside and outside regions; it uses a proactive routing mechanism to maintain routes to inside region nodes and using a route discovery mechanism to reach the outside region nodes [3].

2.1.3.1 ZONE ROUTING PROTOCOL (ZRP)

ZRP is the first protocol developed as a hybrid routing protocol, it allows a network node to divide the network into zones according to many factors; like: power of transmission, signal strength, speed and many other factors. The area inside the zone is the routing range area for the node and vice versa for outside zone. ZRP uses the reactive routing schemes for outside the zone and the proactive routing schemes for inside the zone; with a view to keep the latest route information within the inside zone. In the local inside the zone, the source node uses a proactive cached routing table to initiate a route to a destination, which can be helped in transmitting packets directly without delay. ZRP uses independent protocols inside and outside the zone; it may use any existing proactive and reactive routing protocols. For outside zone, the ZRP reactively discover a route; that the source node transmits a route request packet to the border nodes of its routing zone; the packet includes a unique sequence number, the source address and the destination address. When the border node receives a route request packet, it looks for the destination within its inside zone. If the destination is found, it sends a route reply on reverse path to the source node; else if it doesn't find the destination in its local zone, the border node adds its address to the route request packet and forwards it to its own border nodes. After the source received a reply, it stores the path included in the route reply packet to use it for data transmission to the destination [30]. The weakness of ZRP protocol is that it performs like a pure proactive protocol particularly for large size zones; however for small zones it performs similar to a reactive protocol [17]. Thus ZRP protocol is not applicable for large size VANET with highly dynamic topology and frequently change environment.

2.1.3.2 ZONE-BASED HIERARCHICAL LINK STATE (ZHLS)

ZHLS protocol divides the network into non overlapping zones; every network node has its own ID and a zone ID, which is measured by a GPS. There are two levels for structural topology: zone level topology and node level topology. In ZHLS there is no position administrator or cluster head are used to manage the data communication; that means there is no traffic bottleneck. Besides that the ZHLS reduces the transmission overheads when compared it with the reactive protocols. ZHLS broadcast scheme showed lower overhead compared to the flooding scheme in pure reactive protocols. Also in ZHLS, the routes is flexible to the dynamic topology because it required only the zone ID and the node ID of the destination node for routing; that means there is no need to search for the location, if the destination node does not move to another zone. The shortcoming of ZHLS, it needs a static zone map into each node, and this may not be sufficient for a network with dynamic zone edges. Moreover, it is not appropriate for highly dynamic topologies [17].

Generally, the hybrid routing protocols have a higher scalability than pure proactive and pure reactive protocols; because of reducing the number of rebroadcast messages which achieved by allowing network nodes to work together and the most appropriate nodes are used to setup a route [17]. However, pure proactive and pure reactive routing protocols could be more suitable to some highly dynamic level in a network environment.

OTHER TOPOLOGY-BASED PROTOCOLS

Some researchers [31] assumed it's more efficient to develop a routing protocol based on the topology of roads; that mean road to road transmission rather than the conventional node to node routing scheme; they justify that because of two reasons: the vehicles highly mobility and the data delivery is constrained by road pattern. However, the challenging will be the transmission of packets at intersections. So they proposed Buffer and Switch (BAS) protocol which allow each road to store packets along with many transmitting copies to offer additional chances for a packet to switching at intersections. Different from conventional protocols in VANETs, BAS is a bidirectional duplicate transmission. Also, BAS controlled duplicate transmission by spatiotemporally, which leads to significantly minimum cost compared by other flooding protocols. BAS performance shows better than the traditional protocols, mainly for network with limited resources. However, it may cause transmission delay, and packet loss due to packet expired.

2.2 POSITION-BASED ROUTING PROTOCOL

Position or geographic routing protocol is based on the positional information in routing process; where the source sends a packet to the destination using its geographic position rather than using the network address. This protocol required each node is able to decide its location and the location of its neighbors through the Geographic Position System (GPS) assistance. The node identifies its neighbor as a node that located inside the node's radio range. When the source need to send a packet, it usually stores the position of the destination in the packet header which will help in forwarding the packet to the destination without needs to route discovery, route maintenance, or even awareness of the network topology [3], [4]. Thus the position routing protocols are considered to be more stable and suitable for VANET with a high mobility environment, compared to topology-based routing protocols. Geographic routing protocols commonly classified into three classes: Delay Tolerant Network (DTN) Protocols, Non Delay Tolerant Network (Non DTN) Protocols and hybrid [4].

2.2.1 DELAY TOLERANT NETWORK (DTN) PROTOCOLS

DTN is a wireless network designed to perform efficiently in networks with some characteristics; like frequent disconnection communication, large scale, long unavoidable delays, limited bandwidth, power constraints and high bit fault rates [15]. In this network, all nodes help each other to forward packets (store and forward scheme). These nodes may have a limited transmission range; so packets transmission will take large delays. Commonly, the DTN node is a mobile node, so it establishes routes to other nodes when they reach its transmission range. In DTN protocol, there is no guarantee of unbroken end to end connectivity, so the packets may be cached for a time at intermediate nodes [4], [14], [3]. To design of a routing protocol for DTN network with these characteristics is a significant problem. This section, review many DTN routing protocols that fall under this category.

2.2.1.1 MOTION VECTOR ROUTING ALGORITHM (MOVE)

MOVE algorithm is designed for light networks, especially for road side vehicle communication. This protocol assumes that each node has global locations information, that's beside the knowledge of a mobile router speed and its neighboring

nodes velocity. From this information the node can estimate the nodes which are the closest distance to the destination [14]. In this protocol each node regularly broadcasts a HELLO message; and its neighbor replays by a RESPONSE message; by this replayed message the node will know its neighbors and their locations. Given this information, the node can estimate the shortest distance to destination, in that case the node decides how to forward the message according to the information about nodes which are currently located nearby the destination. MOVE protocol uses less memory size compared with Non DTN position-based routing; it also has a higher data transmission rate in light environments [30]. However, Non DTN position-based routing could have better performance only if the routes are stable and consistent [3].

2.2.1.2 VADD: VEHICLE-ASSISTED DATA DELIVERY IN VEHICULAR AD HOC NETWORKS

VADD protocol designed to handle frequently disconnected vehicular networks and highly mobility problems. It implements the store and forward scheme; while a node is moving it stores the packet, until a new node arrives to its zone range, and then it forwards the stored packet to this new node. This protocol predicts node mobility based on two factors: network traffic and route type; that help a node to discover the next forwarding node. VADD protocols usually deliver the packet to the path with the least transmission delay; following three main principles [4], [14]:

- Continue use the available wireless channel
- Deliver the packet to the higher speed node in the route to carry it
- VANET is a high mobility environment, so it's difficult to estimate packet delivery by a predefined optimal path, which may lead to frequent discover a new optimal path to transmit a packet.

To break the routing loop, each node adds information about its former hop/hops before forwarding the packet, containing its own information as a former hop. Once the packet received to a node, it looks at the previous hops information to avoid forwards the packet to the previous hops and try to find other available hop; so that may avoid the routing loop problem. To forward a packet, VADD implements four different schemes [4], [14]:

- Location First Probe (L-VADD): it used to deliver the packet to the closest node to the destination without consideration of the movement direction. The drawback in this scheme the occurring of the routing loop.
- Direction First Probe (D-VADD): the selection of the next hop is based on the node has the same movement direction as the destination, which helps in avoiding the route loop.
- Multi-Path Direction First is the Probe VADD (MD-VADD): it provides a multi path rather than one path; however, it consumes the bandwidth by redundancy packets.
- Hybrid Probe VADD (H-VADD): it is a hybrid scheme that takes the advantages of L-VADD and D-VADD, to deliver a packet, it initially uses the L-VADD; but if a route loop is identified, it changes to D-VADD. As a result this scheme performs better than pure L-VADD and D-VADD.

2.2.1.3 GEOGRAPHICAL OPPORTUNISTIC ROUTING (GEOPPS)

GeOpps is a forwarding protocol uses the available navigation system in collecting information about geographical position; this information is used to select vehicles that are closest to a certain destination. The protocol uses store and forward technique, it works just like the Move and Non DTN protocols but it uses navigation system to provide efficient packet delivery. In the GeoOpps, to send a packet from the source to the destination, there are three main steps used to select the next hop of the intermediate nodes [3], [5]:

- Each neighboring node at the estimated routes calculates the future closest point to the destination which it will reach soon.
- Each neighbor node then calculates estimated shortest delay time to reach the specified packet's destination.
- Use the estimated shortest time calculated by each neighbor node; that any node estimated to be closer to the destination in lowest delay time, should be selected to become the next hop carrier to transmit the packet faster to the specified destination.

The protocol concerned some cases which affect its efficiency [3], [5]:

- The node ignores the estimated calculated route and follows other different path; in this case the system will forward the holding packet to any neighbor node.
- The node stops its movement (switch off the engine or long pause time); in this case its packets should be forward to another neighboring node.

The benefit of GeOpps does not require all nodes to calculate the routes; and GeOpps transmission rate depends only on the route topology and the mobility of the nodes. However, it has some complexities in calculating delay time depending on a navigation system measurement.

2.2.2 NON DELAY TOLERANT NETWORK (NON DTN) PROTOCOLS

The non-DTN protocols are geographic routing protocols, but it does not consider a dis-connectivity issue; it assumes there are always a number of nodes to achieve the successful communication; so, this protocol is only suitable for high density network. In these protocols, the node forwards its packet to the closest neighbor to the destination, but this approach may be unsuccessful if there is no closest neighbor to the destination rather than the current node itself. Many non-DTN routing protocols handle this failure; by different strategies will be shown in the following sections [1].

2.2.2.1 GREEDY PERIMETER STATELESS ROUTING (GPSR)

GPSR is a famous greedy routing protocol in VANETs. In this protocol, each node forwards packets to other intermediate nodes that are constantly nearer to the packet's destination (greedy forward), until the packet reaches its final destination. If there is no neighboring node close to the destination, it uses perimeter forwarding to decide to which node it will deliver the packet. GPSR is a stateless protocol that keeps information of its first hop neighbor's positions, which could increase protocol scalability more than shortest path ad hoc routing protocols. Another advantage is the dynamic forwarding packet decision [3]. However, GPSR could face a link failure due to the high mobility network and frequent topology changes (it holds old position information). This problem can be handled by perimeter forwarding, but it may cause high packet loss and more latency time due to the large number of hops in perimeter forwarding mode. Moreover, if the destination node moves to a new location, its information which embedded in the packet header will never be updated [2].

2.2.2.2 GREEDY PERIMETER COORDINATOR ROUTING (GPCR)

GPCR protocol is designed to be suitable for the high mobility environments (as in city) based on the greedy forwarding technique; this technique aims to forward the packet to a neighbor node which is closest to the location of the destination. Each node has to be aware of its location gotten by a navigation system, it knows its neighbor by periodic beaconing, and the position of the destination is obtained from the location service. When a node forwards a packet, the packet will be spread over the road until it reaches the next intersection. The maintenance process covers two components: decision making, to decide which intermediate node the packet will be passed on the intersection (a coordinator node selection), and forwarding the packet to the next intersection. The coordinator node decides to which route the packet will be forwarded. But if no coordinator node found in the route, the packet will be forwarded to furthest node [10]. GPCR does not need any global information; however it is based on the connectivity of the destination node and the density of the next roads, it could not connect the destination if the node density is low, which will increase the transmission delay [32].

2.2.2.3 RELIABILITY-IMPROVING POSITION-BASED ROUTING (RIRP)

RIPR is a position-based routing algorithm designed for VANETs, it aims to solve the problems of links failures that found in a position-based routing; which appear due to storing old information about a stale intermediate node. RIPR predicts the vehicle speeds and their moving directions, as well as estimates the characteristics of the city road. In this protocol, the sender selects an intermediate node to forward its packet, based on the mobility estimation for neighboring nodes that done by initially deciding whether a neighbor node exists or not. The sender creates a position record for each neighboring node, this record contains the recent position of the node and its mobility speed; that helps in the selection of the forwarder node which is done based on the route characteristics and the node position record which arranged after the exchange of beacon messages. This record avoids the local problem which prevents a node to select a neighbor node as a forwarder node; that happens because there is no node that is closest to the destination [33]. RIPR protocol is similar to GPSR protocol uses two modes: a greedy mode and perimeter mode, as well as the route characteristics consideration, and the position of the nodes. Therefore, RIPR can solve the link failure problem caused by storing information about a stale intermediate node; so it can reduce the possibility of link failure [32].

2.2.3 HYBRID POSITION-BASED ROUTING

Position routing protocol reduces control routing overhead, it doesn't need to construct or maintain a routing table; because it only uses the location information about the neighbors and destination nodes, these issues made position-based

routing protocols scalable. However, position routing protocols have many limitations that restrict their usage; these limitations can be summarized in the following points [6]:

- The performance of position routing can be significantly decreased according to the location accuracy; because the accurate locations information is an essential factor to get a good performance in position routing.
- Position routing could be failing, if there is no any neighbor node which is closer to the destination (null area).
- Position routing solves the absence of closest neighbor toward the destination, by the backup process. However, it required packets to travel larger distances to reach destinations, also packets could be travel in a close circle, or could be dropped.

So no existing routing protocol performs efficiently in all circumstances. Therefore, many researchers developed hybrid schemes, they merge characteristics of two or more position-based routing protocols (non-DTN and DTN schemes), sometimes they merge one or more topology routing protocols (reactive, proactive and hybrid schemes) with position-based routing. The hybrid position routing protocol is a mixture protocol that takes advantage of more than one protocol schemes. The next section will illustrate HLAR protocol which being an example of is a hybrid position-based routing protocol.

2.2.3.1 HYBRID LOCATION-BASED AD HOC ROUTING PROTOCOL (HLAR)

HLAR is a hybrid position routing protocol designed to efficiently use all the available location information and to minimize the routing control overhead. This protocol is planned to switch to the on-demand routing when sufficient location information is unavailable or limited, it also deals with the problem of no closest neighbor to the destination (void regions), and so it is almost a scalable protocol. HLAR works as a reactive protocol in the route discovery process, however if there is no route to the destination node, the source node adds information about its location and the location of the destination in the route request packet then it searches for a closer node near the destination. If the node finds a neighbor which is close to the destination then it forwards the request packet to it. But if no closer neighbor node is found, it floods the route request packet to all its neighbors. The source node repeats these steps until it reaches the desired destination. The simulation results showed that the HLAR protocol minimizes the routing control overhead compared with the on-demand routing protocols, furthermore it generally provides a fresh large size location information [6]. However, HLAR doesn't guarantee the best reliable route; because the intermediate node doesn't have a reverse link to the source, and could not inform other neighboring nodes if it finds a better route to source [34].

Actually, all categories of routing protocols have the same objectives; that they aim to decrease the network overhead, minimizing the transmission delay and increasing the network throughput. However, in VANETs it is more difficult to find a specific routing protocol that works efficiently in all network environment situations; that some protocol may be suitable for the high mobility environment but suffer from end to end delay, in contrast other protocols could provide fast packet delivery, but unsuitable for the high mobility environment, and so on. So it could be not easy work to precisely compare the existing VANETs routing protocols, or even claims which one is the best in all environment situations; however some research papers analyzed the two classes and compare them using some related protocols; and their results is concluded that the position-based routing performs better than topology-based routing for both urban and rural scenarios [35]. Table 1 illustrates a comparison between topology-based and position-based routing, focusing on strengths, limitations, and methods used in each class.

3 TRANSMISSION STRATEGIES USED IN PACKET FORWARDING

Delivery of information from a source to a destination can be classified into four types: unicast, broadcast, multicast, and geocast, however the multicast and geocast can be merged in one class because geocast usually is a special type of multicast transmission.

3.1 UNICAST ROUTING PROTOCOLS

Unicast routing refers to information delivery from a single source to a single destination using the wireless multi hop scheme; where the intermediate nodes are used to forward data from the source to the destination, or by using the store and forward scheme. It is the most class that widely used in the general ad hoc networks. This scheme required the source vehicle to hold its data for a time and then forward it [36]. There are many unicast routing protocols proposed for VANETs; most of the topology-based routing protocols belong to a unicast class; such as VADD, AODV, DSR and many other, which presented in the previous sections.

3.2 BROADCAST ROUTING PROTOCOLS

Broadcasting routing enables packets to flood into the network to all available nodes inside the broadcast domain. Broadcasting routing is widely in VANETs, it mainly used in the route discovery process, some protocols (like AODV) allow nodes to rebroadcast the received packets. This routing scheme allows packets to deliver via many nodes which may achieve a reliable packet transmission, however it could consume the network bandwidth by sending replicated packets, so each node need to identify which packet is replica (it has received it before) to discard.

Table 1. Comparison of Topology based Routing and Position based Routing

VANET Routing Protocols	Methods Used	Strengths	Limitations	Comments
Topology-based Routing	Link's information stored in the routing table as a basis on forwarding a packet	The shortest route from source to destination Support of messages unicast, multicast and broadcast Less resource consumption Beaconless Save bandwidth	More overhead Routes discover and maintaining delays Fail to discover a complete path (frequent network changes) Unnecessary flooding	These protocols generally are proposed for MANETs Can helpful for small networks (less overhead)
Position-based Routing	Beaconing Vehicles position information Global positioning service	No need to create and maintain global routes More stable in high mobility environment More fitting for network distributed nodes Lowest overhead More scalable	Obstacles in highway scenario Deadlock problem in location server Position services may fail in tunnel or obstacles (missing satellite signal)	More suitable for VANETs; but need more researches for small networks and control congestion

3.2.1 DENSITY-AWARE RELIABLE BROADCASTING PROTOCOL (DECA)

DECA is a density aware protocol; it uses beacon messages to get knowledge about its neighboring nodes and to share information between nodes. It is a reliable broadcast protocol utilizes store and forward transmission scheme. When a node broadcasts a packet, it initially chooses a next hop to rebroadcast the packet; the next hop selection is based on the amount of node information; that means the next hop node will be the node that has the largest density information, after the next hop selection, the node adds the next hop ID, to the packet then broadcast the packet. Other nodes which aren't next hops, should store the packet and startup a waiting timer; if the time is over and no rebroadcast packet received then they rebroadcast the packet by themselves. Any neighboring node which received the broadcasted packet, will add its ID to the regular beacon, to enable other nodes to determine which one of its neighbors that haven't received the broadcasted packet, in order to rebroadcast the packet to it [10]. Mainly, DECA protocol doesn't use any global position information (like GPS) in its processes; that help it to be more flexible and do well in many network environments [10]. However, transmission of periodic beaconing could cause a broadcast storm problem which increases the network overhead and decreases the performance. Also if the waiting time is ended without receiving any broadcasted packet, the network will flood by rebroadcast from neighboring nodes.

3.2.2 POSITION-AWARE RELIABLE BROADCASTING PROTOCOL (POCA)

POCA similar to DECA protocol, it select certain neighbor nodes to rebroadcast a packet, however in this protocol the selection of rebroadcast nodes is based on their position; other unselected nodes stores the packet and startup a waiting timer, if the time is over and no rebroadcast received, they rebroadcast the packet by themselves. POCA also depends on adaptive beaconing which minimizes beacon overhead, to obtain information about neighbors' locations, their speed, and their connectivity status. Thus nodes recognize if their neighbors did no received some packets and rebroadcast to them. POCA provided a good reliability in higher density [10]. However, it will be a problem if the waiting time is over; the network will flood by rebroadcast from neighboring nodes.

3.2.3 DISTRIBUTED VEHICULAR BROADCAST PROTOCOL (DV-CAST)

DV-CAST is a broadcast routing protocol uses multi hop scheme. In this protocol, each node monitors the status of its neighboring connectivity all the time, in order to broadcasts to them. DV-CAST deals with different classes according to many aspects; such as: traffic state, connected state of the neighboring nodes, light traffic, and normal traffic. It uses the periodic beacon messages to get information about the network topology. In a smaller amount of the connected nodes, the node can be rebroadcast along with nodes moving in the same way. In disconnected case of neighboring nodes, the source node should use the store and forward scheme; that it store the broadcasting packet until it find another node move into its broadcast domain, but if there is no node, it will discard the packet after the packet live time is ended. The protocol also enables network nodes to decide if the packet is received before or not; that by using flag parameter [33]. Certainly, DV-CAST protocol minimizes the broadcasting overhead, which made the protocol appropriate for both of light and crowded traffic situations. However, it could cause a highly control overhead and increase the end to end delay in the data transmission [14].

3.2.4 DISTRIBUTION-ADAPTIVE DISTANCE WITH CHANNEL QUALITY (DADCQ)

DADCQ aims to provide a well performed adaptive multi hop broadcast protocol for large networks with high node distribution. It selects forwarding nodes to rebroadcast packets according to positional information. In rebroadcast decision, when a node received a packet, it first checks its distance from the destination; if it was very close there then no need to rebroadcast; because its rebroadcast will not cover a further area. However, if this distance is large, then the node has to rebroadcast the packet. DADCQ has a minimum transmission overhead, because it depends on the node a distributed measurement which is the lower changes than network topology changes [37]. However it may cause a large message overhead.

3.3 MULTICAST-BASED ROUTING PROTOCOL

Multicast is defined by sending packets from a single source to specific group members by multi hop communication [36]. Multicast routing in VANETs can be classified into two categories: geocast and cluster-based routing. The following section illustrates each class in more detail.

3.3.1 GEOCAST-BASED ROUTING PROTOCOL

Geocast routing protocols belongs to a multicast routing protocol which based on sending packets from a source to a particular group of destinations. Some publications remark geocast routing is actually a multicast position-based routing [2], [7]. In VANETs, the geocast routing protocol is a multicast service which enables a single vehicle to transmit a packet to all other vehicles located in the specific geographical area which labeled zone of relevance (ZOR) [7]. Nodes are elements in a one ZOR group, if they located in the same and a specific geographical area. The node membership is changed when the node moves out of the defined geographical area scope, and in this case it drops the packet. A zone of forwarding (ZOF) is defined as the geographic area which vehicles in this area must deliver the packets to other ZOR vehicles. ZOF aims to achieve a reliable packet's delivery in highly dynamic topology. It provides a periodic retransmission, to deal with the network changes. The one drawback of geocast is packet transmission delay that caused by network disconnection. There are a variety of proposed Geo cast routing protocols available.

3.3.1.1 ROBUST VEHICULAR ROUTING (ROVER)

ROVER is a geographical multicast protocol; which permits each vehicle to deliver a packet to all vehicles inside a specific ZOR; using on-demand routing to discover packets inside a ZOR. ROVER similar to AODV protocol; it only floods control

packets in the network, and unicasts the data packets, this scheme rising the consistency and efficiency. This protocol assumes vehicles have identification numbers, digital map, and global locations information. The source node starts discover a route by flooding its ZOR by route request packet, this packet included source ID, its location, its recent ZOR, and a sequence number of the route. When a vehicle received the route request packet, it accepts the packet; if it was nearly close to the source and located inside the ZOF and ZOR. If the vehicle was outside the ZOR, it doesn't send a reply. After a vehicle accepts the route request packet, it sends back a reply packet contains its ID to one-hop neighbors, besides recorded the route request packet information in its routing table. And then retransmit the route request packet [1].

ROVER is different from AODV in that it sends the reply back to the node which transmitted the route request packet not to the source. So each node can construct a tree of multicast routing which has the source node as the tree root. After the tree is constructed, the data packet is then broadcasted in the tree. ROVER has achieved a reliable geographical multicast routing scheme for VANETs. However, the control packet overhead is increased as well as data delivery delay; due to the increased number of retransmission packets [2].

3.3.1.2 MOBILE JUST IN TIME MULTICASTING PROTOCOL (MOBICAST)

Mobicast is a multicast geographical protocol, different to conventional geocast routing protocol, Mobicast routing protocol takes into account the time aspect. It's designed to provide a management for spatiotemporal needs in VANETs; that by transmitting a Mobicast packet to Vehicles inside a ZOR at time t (ZORt). All vehicles belong to the ZOR at a time t should stay connected to preserve the communication of the real-time data among the entire ZOR vehicles. The communication of ZOR is failing if any ZOR vehicle unexpectedly speeds up or slows down its speed. A location provider (GPS) is used to know the location of each vehicle. When the network is temporal fragmented, vehicle within a ZOR may not efficiently receive Mobicast packets. Mobicast protocol designed to solve this problem using ZOF, and efficiently disseminate Mobicast packets. ZOFt is a ZOF that enables to disseminate a Mobicast packet to every vehicle found in the ZORt. The size of the ZOFt may not usually be the optimal one. That if the ZOFt size is bigger than the ZOFt optimal size, in this case many unrelated vehicles can be requested to deliver Mobicast packets. As well if the ZOFt size is smaller than the optimal ZOFt size, it will cause a temporal network fragmentation problem; actually the ZOFt is not easy to accurately determine it, especially in the high mobility environment, which may lead to misuse of the network resources. Mobicast estimates the ZOFt size to a value is close to the optimal, as well as determines the structure, and location of the ZOFt, to dynamically form flexible ZOF [32]. However, Mobicast has a limitation because of it relies on the location provider (GPS) to know a global knowledge about the network density; because it may not perform well in a large network with a dynamic and density environment. Also the ZOF is may be large without the need (due to incorrect configuration estimation), this may increase the network overhead.

3.4 CLUSTER-BASED ROUTING PROTOCOL

This protocol divides the network to clusters, where nodes have the same characteristics, like same direction or same velocity, or so on. Each cluster has a cluster head, its task is to manage communication processes inside, and to outside its cluster. Nodes inside the cluster communicate by direct paths, but their communication with other nodes outside the cluster is achieved by their cluster header, and this creates a virtual infrastructure for networks. This scheme can provide a good scalability for large networks; however it may increase network overhead and delays in highly dynamic network [2], [14].

3.4.1 COIN: CLUSTERING FOR OPEN IVC NETWORK

COIN is a clustering mechanism designed to improve network scalability, it divides the network to clusters; but not like conventional other clustering protocols, COIN selects clusters according to three parameters: mobility of nodes, nodes positions and behavior of nodes. The protocol provides each cluster specific time which is a time to live; in order to decrease control overhead. Inter vehicles communication system (IVC) deals with the unstable distances of inter vehicles. To enable a head of cluster node and the cluster member node stay continue communicate, their mobility should be low and related to the mobility of each, in this case they can reside in radio contact for a longer time [6], [14].

3.4.2 CLUSTER-BASED DIRECTIONAL ROUTING PROTOCOL (CBDRP)

In CBRP the network is divided into several clusters. Each cluster has a cluster head which is responsible for the routing procedure. These cluster heads communicate with each other via gateway nodes which are nodes that have more than one cluster head. When a source node requests a route, it floods the network by request packet. The CBRP clustered structure reduces traffic overhead, because request packet only passes among cluster heads. However, in density networks, the packet

overhead will increase and the transmission delay will increase; because information of each node in the route should be added to the routed packet which increases the size of the packet. One more limitation of the CBRP, it maintains unidirectional links, while most of link layer protocols support only bidirectional links [38].

Generally, cluster-based routing protocols may perform well in network scalability for large networks; however, they may cause more networks overhead due to structure of clusters and cluster heads [34].

4 VANETs Routing Open Issues

Through our literature review in VANETs routing protocols, we found there are still some open issues and challenges in VANETs routing, which it is one of the most active topics in VANETs research area, it has several recent publications. This section represents some open research issues in VANETs routing problem (for example, but not limited to). It required designing a single routing protocol that can:

- Works efficiently for both urban and rural, has the ability to improved networks throughput and packet delivery ratio. Also reduces resource consumption, and guarantee optimal paths.
- Scalable; has the ability to handle dynamic connectivity for broken links, as well as deals with the conditions of a single network, like crowded, congestion, available bandwidth, transmission interference, allowed speed, and so on.
- Adaptive reliable broadcasting/ multicast transmissions, that works efficiently in sending packets for all nodes with minimum overhead, duplications, collisions and congestions.
- Can solve hidden terminal problems; to avoid out range collision.

Intelligent to adapt and deal with unexpected conditions; like driver behavior, signal loss, interference by tunnels, high building, and intersections condition. Briefly, an expected VAENETs routing protocol should be able to providing communication with minimal overhead and delay, highest scalability and adaptable for VANETs environments; by using optimal route selection and powerful reconfiguration algorithm.

Table 2. Comparison of VANET Transmission Strategies

Transmission strategies	Methods Used	Strengths	Limitations	Comments
Unicast	Information delivery from a single source to a single destination	Less network overhead More privacy Minimum packet delay	Links should be frequently configured and maintain Less reliability Packet loss	Need more researches to enhance reliability, packet retransmission, scalability and avoid collision
Broadcast	Packets flooding to all network nodes inside the broadcast domain	More reliable data transmission Less packet loss	Consumes bandwidth Routes loop Network congestion Less network throughput More packet delay Packet collisions	Required reducing bandwidth consumption Could be useful for alert messages Need some packets flooding constrains
Multicast	Geocast sending packets from a source to a group of destinations using geographic addresses Cluster divides the network to clusters, each cluster has a cluster head to manage communication inside and outside the cluster	Efficient routing by sending one copy to multiple nodes Minimum network consumption Minimum packet delivery delay Easy to implement Transparent to changeable addresses (no requirement to receiver's address)	Consumes bandwidth More overhead in dividing network nodes into groups Routes loop	Scalability control for dynamic groups The cluster may not very efficient because frequent changing heads (like Mobile routers in network mobility but without a guarantee the network nodes will travel as one unit)

5 CONCLUSION

This paper has presented an overview of Vehicular s ad hoc networks (VANETs), illustrates their motivation and characteristics, it studied in detail VANETs routing problem, mainly vehicle to vehicle (V2V) communication, providing two classifications of VANETs routing protocols that exist in the last few years, investigated them and showing how do they work and their main advantages and limitations The paper also summarized comparisons between the main classes.

Thorough this study of different VANETs routing protocols, many related open issues and research challenges are found and represented, these issues still required more effort and research to address them. We hope that the instrument presented in this paper to be useful and helpful to students and researchers in the field.

REFERENCES

[1] Ramin Karimi, Norafida Ithnin, Shukor Abd Razak, and Sara Najafzadeh, "Non DTN Geographic Routing Protocols for Vehicular Ad Hoc Networks," *International Journal of Computer Science Issues (IJCSI)*, vol. 8, Issue 5, no. 3, September 2011.

[2] Bijan Paul, Mohammed J. Islam, "Survey over VANET Routing Protocols for Vehicle to Vehicle Communication," *IOSR Journal of Computer Engineering (IOSRJCE)*, ISSN: 2278-0661, ISBN: 2278-8727, vol. 7, Issue 5 (Nov-Dec. 2012), pp. 01-09.

[3] Moath Muayad Al-Doori, "Directional Routing Techniques in VANET," *PhD Thesis, Software Technology Research Laboratory, De Montfort University, Leicester United Kingdom*, November 2011.

[4] Lee, Kevin C., Uichin Lee, and Mario Gerla., "Survey of Routing Protocols in Vehicular Ad Hoc Networks," *Advances in Vehicular Ad-Hoc Networks: Developments and Challenges reference*, *IGI Global*, 2010, pp. 149-170, 25 Mar. 2013.

[5] Salim Allal and Saadi Boudjit, "Geocast Routing Protocols for VANETs: Survey and Geometry-Driven Scheme Proposal," *Journal of Internet Services and Information Security (JISIS)*, vol. 3, no. 1/2, pp. 20-36, February 2013.

[6] Mohammad Al-Rabayah and Robert Malaney, "A New Hybrid Location-based Ad Hoc Routing Protocol," *Global Telecommunications Conference (GLOBECOM 2010) IEEE* , vol. 1, no. 6, pp. 6-10, Dec. 2010.

[7] Rakesh Kumar and Mayank Dave, "A Comparative Study of Various Routing Protocols in VANET," *International Journal of Computer Science Issues (IJCSI)*, vol. 8, Issue 4, no. 1, July 2011.

[8] Vijayalaskhmi M., Avinash Patel, Linganagouda Kulkarni, "QoS Parameter Analysis on AODV and DSDV Protocols in a Wireless Network," *International Journal of Communication Network & Security*, vol. 1, no. 1, 2011.

[9] lajos hanzo and rahim tafazolli, "Asurvey of QoS Routing Solutions for Mobile Ad hoc Networks," *IEEE Communications Surveys & Tutorials, 2nd quarte,* 2007.

[10] James Bernsen and D. Manivannan,"Greedy Routing Protocols for Vehicular Ad Hoc Networks," *Wireless Communications and Mobile Computing Conference* IWCMC 08, vol. 632, no. 637, pp. 6-8, Aug. 2008.

[11] Jamal Toutouh, Enrique Alba, "Optimizing OLSR in VANETS with Differential Evolution: A Comprehensive Study," *ACM DIVANet'11, Miami, Florida, USA.*, 2011.

[12] Sonam Jain and Sandeep Sahu, "Topology vs. Position based Routing Protocols in Mobile Ad hoc Networks: A Survey," *International Journal of Engineering Research & Technology (IJERT),* vol. 1, Issue 3, May 2012.

[13] Y. Lacharite and M. Wang, "Hierarchical OLSR draft-lacharite-manet-holsr-02," *Mobile Ad hoc Networking (MANET), IETF Internet-Draft,* 2009. [Online] Available: http://tools.ietf.org/html/draft-lacharite-manet-holsr-02 (July 13, 2009)

[14] Yatendra Mohan Sharma and Saurabh Mukherjee, "A Contemporary Proportional Exploration of Numerous Routing Protocol in VANET", *International Journal of Computer Applications,* vol. 50, no. 21, pp. 14-21, July 2012.

[15] Sandhaya Kohli, Bandanjot Kaur and Sabina Bindra, "A comparative study of Routing Protocols in VANET", *International Journal of Computer Science, Issues-IJCSI*, *Proceedings of ISCET*, 2010.

[16] Haerri, J., Filali, F., and Bonnet, C., "Performance comparison of AODV and OLSR in VANETs urban environments under realistic mobility patterns", *In Proceedings 5th IFIP Mediterranean Ad-Hoc Networking Workshop*, June 14-17, 2006, Lipari, Italy.

[17] M. Abolhasan, T. Wysocki and E. Dutkiewicz, "A review of routing protocols for mobile ad hoc networks", *Ad Hoc Networks 2, Elsevier*, pp. 1–22, 2004.

[18] Raphael Frank, Eugenio Giordano, Thomas Engel and Mario Gerla "Performance Bound for Routing in Urban Scenarios", *ACM AINTEC'11*, November 9–11, 2011, Bangkok, Thailand.

[19] Naumov, V., Baumann, R., Gross, T. (2006), "An evaluation of Inter-Vehicle Ad Hoc Networks Based on Realistic Vehicular Traces," *Proc. ACM MobiHoc'06 Conf.*, May, 2006.

[20] S. Das, C. Perkins, E. Royer, "Ad hoc on demand distance vector (AODV) routing", *Internet Draft, draft-ietf-manet-aodv-13.txt*, *Mobile Ad Hoc Networking Working Group,* 2003.

[21] M. K. Marina, and S. R. Das, "On-demand multipath distance vector routing in Ad Hoc networks," *Proc. 9th International Conference on Network Protocols, IEEE Press*, pp. 14-23, November. 2001.

[22] Hafez Maowad and Eman Shaaban, "Enhancing AOMDV Routing Protocol for V2V Communication", In *Proceedings of the 6th international conference on Communications and Information Technology, World Scientific and Engineering Academy and Society (WSEAS),* 2012. [Online] Available: http://dl.acm.org/citation.cfm?id=2209535.2209538 (2012)

[23] Boangoat Jarupan and Eylem Ekici, "A survey of cross-layer design for VANETs," *Ad Hoc Networks (Elsevier) Journal Special Issue on Vehicular Networks*, doi:10.1016/j.adhoc.2010.11.007, 2010.

[24] Sommer, C., Dressler, F. "The DYMO Routing Protocol in VANET Scenarios," *Vehicular Technology Conference*, *VTC-2007 fall. IEEE 66th*, pp: 16–20, 2007.

[25] Annu Mor, "A STUDY OF IMPROVED AODV ROUTING PROTOCOL IN VANET," *International Journal of Computer Applications & Information Technology*, vol. II, Issue I, January 2013 (ISSN: 2278-7720).

[26] Sufyan T. Faraj Al-Janabi, Yaseen Saleem Yaseen, and Bob Askwith, "The Bus Ad Hoc On-demand Distance Vector (BAODV) Routing Protocol", *PGNet,* ISBN: 978-1-902560-26-7, 2012.

[27] Johnson, David B., David A. Maltz, and Josh Broch. "DSR: The dynamic source routing protocol for multi-hop wireless ad hoc networks," *Ad hoc networking 5*, pp. 139-172, 2001.

[28] Gurleen Kaur Walia, "A Survey on Reactive Routing Protocols of the Mobile Ad hoc Networks," *International Journal of Computer Applications*, vol. 64, no. 22, pp. 45-51, February 2013.

[29] V. Park, S. Corson, "Temporally-Ordered Routing Algorithm (TORA) Version 1 Functional Specification," *IETF Internet draft, 2001*. [Online] Available: http://tools.ietf.org/pdf/draft-ietf-manet-tora-spec-04.pdf (2001)

[30] Jan Schaumann, "Analysis of the Zone Routing Protocol," 2002.
[Online] Available: http://www.netmeister.org/misc/zrp/zrp.pdf (December 8, 2002)

[31] Chao Song, Ming Liu, Yonggang Wen, Jiannong Cao and Guihai Chen "Buffer and Switch: An Efficient Road-to-road Routing Scheme for VANETs", *Seventh International Conference on Mobile Ad-hoc and Sensor Networks,* 2011.

[32] Si Ho Cha1, Min-Woo Ryu and Kuk-Hyun Cho, "A Survey of Greedy Routing Protocols for Vehicular Ad Hoc Networks," *Smart Computing Review*, vol. 2, no. 2, April 2012.

[33] M. W. Ryu, S. H. Cha, J. G. Koh, S. Kang and K. H. Cho, "Position-based routing algorithm for improving reliability of inter-vehicle communication," *Transactions on Internet and Information Systems (TIIS)*, vol. 5, no. 8, 2011. [Online] Available: http://freepatentsonline.com/article/KSII-Transactions-Internet-Information-Systems/272246267.html (Aug. 2011)

[34] James. Jong Hyuk Park, Victor C.M. Leung, Cho-Li. Wang and Taeshik Shon, "Future Information technology, Applications, and Services", *FutureTech Springer,* vol. 1, pp. 82-98, 2012. [Online] Available: http://books.google.com/books?id=pggzUBASE4IC&printsec=copyright#v=onepage&q&f=false (2012)

[35] Akhtar Husain, Ram Shringar Raw, Brajesh Kumar and Amit Doegar, "Performance Comparison of Topology and Position Based Routing Protocols in Vehicular Network Environments", *International Journal of Wireless & Mobile Networks (IJWMN)*, vol. 3, no. 4, August 2011.

[36] Yun Lin, Yuh Chen and Sing Lee, "Routing Protocols in Vehicular Ad Hoc Networks: A Survey and Future Perspectives", *Journal of Information Science and Engineering*, vol. 26, pp. 913-932, 2010.

[37] Michael Slavik and Imad Mahgoub, "Spatial Distribution and Channel Quality Adaptive Protocol for Multihop Wireless Broadcast Routing in VANET," *IEEE Transactions on Mobile Computing*, vol. 12, no. 4, pp. 722-734, April, 2013.

[38] Tim Daniel Hollerung, "The Cluster-Based Routing Protocol", *project group Mobile Ad-Hoc Networks Based on Wireless LAN*, *University of Paderborn*, 2003/2004.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 847-856

http://www.issr-journals.org/ijias/

The Survey of the Relationship between Management Expected Profits and Disclosure Quality Associated with Market Surprise

Roya Darabi[1] and Saeid Gholami[2]

[1]Accounting Department,
South Tehran Branch, Islamic Azad University,
Tehran, Tehran Province, Iran

[2]Department of Management,
Arak Branch, Islamic Azad University,
Arak, Arak Province, Iran

ABSTRACT: In the present study, we deal with the survey of the relationship between the management forecasted profits and disclosure quality with the market surprise in Tehran's securities market. Since managers, analysts and investors pay a greater attention to the companies' reported profit in a way that they use it to evaluate the company's performance and also because the decision-making for purchasing, maintain, or the sale of the stock shares is of a great importance for the investors, and from among other evidences and information, the capability to forecast the stock return rate has a greater influence on such decision-making, the aim of the present study is the survey of the relationship between the management forecasted profits and the disclosure quality with market surprise in the companies accepted in Tehran's Securities Exchange.
To reach the aforementioned objective three hypotheses are being proposed in which it has been dealt with the profit forecast accuracy, getting surprised with the management announced profit and systematic risk with stock price response. To test the study hypotheses, the data from 116 companies accepted in Tehran's Securities Exchange was selected based on the goal-oriented systematic sampling method and the data from the time span from 2001 to 2011 was used to statistically test the hypotheses in the form of multiple-regression and the data panel was used in two softwares, namely SPSS17 and Eviews7.
The obtained results are suggestive of a significant and reverse relationship between the profit forecast accuracy and the stock price response and there is a direct relationship between getting surprised from the management announced profit and the systematic risk with stock price response.

KEYWORDS: Management Predicted Profits, Disclosure Quality, Market Surprise, Behavioral Finance, Accuracy.

1 INTRODUCTION

From among the information that the companies' financial data users pay attention to in their decision making one can refer to the information related to the companies' stock return. In between, any kind of information which is somehow related to the companies and their function and financial status can be effective on the stock return. For this reason, various studies devoted to the changes in the companies' stock return have been undertaken and we have dealt with the stock return behavior in confrontation with the diffusion of the companies' related information by taking advantage of various methods.

It is expected that the accounting profit is part of the data which is being used by the investors to assess the venture-taking and return. Therefore, it can be predicted that accounting profit and specially the difference between predicted net profit and the realized net profit during various periods has information content. Also, the expectation of the price

Corresponding Author: Roya Darabi (royadarabi110@yahoo.com)

fluctuations and as a consequence information content such as investors response in comparison to the management forecast and measures have been compared with the forecast undertaken in reality which in the current study will measure the investors responses in a different manner.

Prospective Stock return is seeking for the management forecast and the effect of the undertaken disclosure quality from these forecasts which in the end leads to the investors surprise in case of good news announcement or bad news relative to the confronted conditions. The forecast have turned into wonders for the investors and make them respond excessively to the performed forecasts. Management forecast is based on the company business strategy and the degree of the awareness of all of the performed forecasts and such forecast will lead to abnormal returns. This strong return has been estimated based on the conventional assets pricing and the subtraction of the transaction costs. We also predict that these abnormal returns are stemmed from the good or bad news forecast and finally they are based on the amount of the investors being conservative and their awareness of such activities and the company background in comparison to the announced news related to the predicted profits. These responses, in fact, are the investors surprise in view of the announced news in relation to the management forecasts [1].

2 REVIEW OF LITERATURE AND RESEARCH BACKGROUNDS

Ng, Tuna and Verdi [1] in a study named 'management forecasts, disclosure quality and market efficiency' dealt with the survey of the getting surprised from predicted profits by the management and the systematic risk with stock market responses. The study results are indicative of the direct and significant relationship between management forecast with getting surprised. And, there is a significant and direct relationship between the accuracy of the profit forecast as an indication of the disclosure and the stock price response.

Bloomfield, Libby and Nelson [2] dealt with the survey of the investors' reliance amount on the time series components of the previous profits. Their study results showed that the investors rely excessively on the information obtained from the previous profits in order to do the forecasts and this leads the investors to expected errors in the forecast process.

Best and Best [3] dealt with the survey of the information content of the predicted net profit change statement by making use of the predicted error standardized model and they used Toobin's Q to measure the shareholders' responses. To do so, they used the fifteen-day data regarding the net profit change statement. To make their study more meaningful and significant, they divided the net profit change statement in to two groups: the decrease and increase in the predicted net profit statement. They considered many control variables in their studies such as stock price, assets market value, the amount of the difference in the stock profit statement. They came to this conclusion that the market reacts to the net profit statement and the stock profit change statement possesses information content.

Collins and Kothari [4] dealt with the relationship between systematic risk and the profit response coefficient and they found out that the only reductive factor for the profit response coefficient is the systematic risk (β). Besides, they figured out that the growth opportunities rate factor has a positive effect on the ERC. In fact, their study showed that as ERC makes a swift test of the relationships between price and return feasible which is a latent financial valuation, some of the simultaneous ERC differences can be elaborated by incorporating the systematic risk variables.

Xie [5] , [6] seeking to find new models for the profit forecasts and developing the extant payment models, stated that the income forecasts can suffer from entering personal factors in the forecasts dispersion. He referred to the purging of the personal analysis about the profit forecasts reports. He dealt with the development and the presentation of models to standardize the profit forecasts.

Brid [7] surveyed the importance of accounting information for predicting profit change. He gathered his information from the companies based in the U.S.A., England and Australia. His final model financial proportions which are chosen based on the step-wise regression ratios are as follows:

Asset output ratios,

Change in the sales to inventories,

Change in total assets,

Change in inventories,

Change in capital expenditure to total assets,

Operating profit to sales,

Profit before tax on sales,

Cash to debts ratio,

Net profit to sales,

Equity return.

Beaver, clarke and wright [8] studied the amount of market response. They examined a sample of 276 companies for a 10-year period from 1965 to 1974. Comparing unexpected changes in stock abnormal return, these three researchers concluded that the stronger the intensity of the changes in the unexpected net profit, the stronger the response of the securities market. This conclusion is consistent with the application of capital assets pricing pattern and with the method based on the utility of the information from the decision-making point of view. The more changes in the unexpected profit, on average, more investors rethink their attitude towards the positive orientation related to the company's prospective profitability strength.

Management earnings forecasts are increasingly common voluntary public disclosures through which managers can influence price formation in the equity markets. Many papers have examined the short-term returns in response to management forecasts to draw inferences from investors' immediate reaction to the forecasts (e.g., [9], [10], [11], and [12]). Although these papers show that management forecasts are informative, they do not address whether the market under reacts or overreacts to them. We address this question by examining the future long-term abnormal returns following management forecasts. In addition, we study the influence of disclosure quality on the magnitude of these returns.

Brav and Heaton [13], relaxing the assumption those investors has complete information, propose rational structural uncertainty as another explanation for under reaction to news. In their model, rational investors with incomplete information about the structure of the economic environment react in a Bayesian manner. Any under reaction then occurs as a result of mistakes or risk premia that arise from incomplete information. As noted by Brav and Heaton, the similar mathematical properties and empirical predictions of rational structural uncertainty and the behavioral models make it difficult to empirically distinguish which of these models best explain under reactions to news.

Bartov, Radhakrishnan, and Krinsky [14] in a study dealt with the investor sophistication and patterns in stock returns after earnings announcements. Their study shows that the institutional holdings variable is negatively correlated with the observed post-announcement abnormal returns. Their findings also show that traditional proxies for transaction costs (i.e., trading volume, stock price) as well as firm size have little incremental power to explain post-announcement abnormal returns when institutional holdings is an explanatory variable. If institutional ownership is a valid proxy for investor sophistication, these findings suggest that the trading activity of unsophisticated investors underlies the predictability of stock returns after earnings announcements. However, tests evaluating the validity of institutional holdings as a proxy for investor sophistication yield only mixed results.

Rogers [15] studied the disclosure quality and management trading incentive. Using the changes in market liquidity to proxy for disclosure quality, he found that trading incentives are associated with disclosure quality choices. Tests were performed across three disclosure samples: management forecasts, conference calls and press releases. Consistent with a desire to reduce the probability of litigation, he found evidence that managers provide higher quality disclosures before selling shares than they provide in the absence of trading. Consistent with a desire to maintain their information advantage, he found some, albeit weaker, evidence that managers provide lower quality disclosures prior to purchasing shares than they provide in the absence of trading.

Lee [16] in his study 'Market efficiency and accounting research' assumed that the price adjustment process to information is instantaneous and/or trivial. This basic assumption has had an enormous influence on the way we select research topics, design empirical tests, and interpret research findings. In his discussion, he argues that price discovery is a complex process, deserving more attention. He highlights significant problems associated with a naïve view of market efficiency, and advocates a more general model involving noise traders. Finally, he discusses the implications of recent evidence against market efficiency for future capital market research in accounting.

Williams [17] studied the relationship between a prior earnings forecast by management and analyst response to a current management forecast. He investigated whether the accuracy of prior earnings forecast by management serves as an indicator to analysts of the believability of a current management forecast. Regression analysis is used to examine the relationship between the usefulness of a prior forecast by management and analyst response to a current forecast, after controlling for other determinants of believability. The results suggest that management establishes a forecasting "reputation" based on prior earnings forecasts.

3 RESEARCH HYPOTHESIS

According to the theoretical literature and study background the following hypotheses are introduced.

H1: There is a significant relationship between profit forecast accuracy with the stock price response.
H2: There is a significant relationship between getting surprised from the management announced profit and the stock price response.
H3: There is a significant relationship between systematic risk and the stock price response.

4 RESEARCH METHODOLOGY

The current study method is functional from the objective point of view and it is correlation-descriptive from the type point of view. In the current study in order to test the hypothesis, the correlation analysis method was used.

The study population is all of the existing companies in the Tehran's securities market during the period from 2001-2011, but these companies should be qualified according to the following conditions to be chosen as the statistical sample:

- Companies shouldn't change their fiscal year in the study time period.
- The studied companies shouldn't be engaged in the investing and holding affairs.
- Their fiscal year should end in December.
- The number of the total companies participated in the study, based on the above qualifications was 116.

4.1 RESEARCH VARIABLES

4.1.1 DEPENDENT VARIABLES

Stock price response

In the current study abnormal return on the company stock was considered as the stock price response scale, which is calculated according to the relation one:

$AR_{i,t}= R_{it} - (R_{mt})$ (1)

$AR_{i,t}$ is the real stock return on the day t. And it is calculated based on the relation 2:

$$R_{i,t} = \frac{(P_{t+1} - P_t) + D + M + N}{P_t} \quad (2)$$

P_{t+1} = price on the day after t-th day.

P_t = price on the t-th day.

D= net profit.

M= precedence benefits.

N= sharing profit benefits.

$R_{m,t}$ is the expected market return which is calculated based on the relation 3:

$$Rm,t = \frac{TEDPIX_{t+1} - TEDPIX_t}{TEDPIX_t} \quad (3)$$

Where TEDPIX is the price index and cash return in Tehran's Stock Exchanges.

4.1.2 INDEPENDENDENT VARIABLES

Getting surprised from the management announced profit: when the management profit forecast has been announced and it is found different from the investors and financial analysts' expectations, a phenomenon known as 'market surprise' occurs.

In the current study this phenomenon (surprise) is calculated by making use of Ng and Tuna and Verdi [1] method in the form of relation 4:

$$Surprise = \frac{Management\ Frecast - Analyst\ Forecast}{Price} \quad (4)$$

Surprise= getting surprised.

Management forecast= Management forecast from the company profit.

Analyst forecast: the undertaken forecast average from the profit by the management during the recent years.

Price: the stock price at the end of the fiscal year.

Disclosure: Disclosure means that the increase in the profit forecast accuracy is interpreted as the higher disclosure quality and vice versa.

$$\text{Accuracy} = -1 * |\text{Prior Earnings} - \text{Prior Management Forecast}|/\text{Price} \quad (5)$$

Prior Earnings = every shares real profit (EPS) related to the previous fiscal year.

Prior management forecast= expected by the management related to the previous fiscal year.

Accuracy: profit forecast accuracy.

Systematic risk: beta is a scale of systematic risk which is calculated by relation 6.

$$\beta = \frac{\delta_{im}}{\delta_m^2} = \frac{E[(R_i - \mu_i)(R_m - \mu_m)]}{E(R_i - \mu_i)^2} \quad (6)$$

In Which R_i , R_m are the company return and market return, respectively, and μ_i , μ_m are the companies returns average and market returns average.

4.1.3 CONTROL VARIABLE

Company size: This is obtained from the total assets logarithm.

To test the study hypothesis, we use regression models (1), (2), and (3), in order:

$$\text{Abret} = \beta_0 + \beta_1 \text{ Accuracy} + \varepsilon \quad \text{Model 1}$$

$$\text{Abret} = \beta_0 + \beta_1 Bad\ News + \beta_2 Good\ News + \beta_3 BadNews \times Surprise + \beta_4 GoodNews \times Surprise + \beta_5 Control\ Variable_j + \beta_6 Surprise_{it} + \varepsilon \quad \text{Model 2}$$

$$\text{Abret} = \beta_0 + \beta_1 Risk\ Controls + \beta_2 Surprise + \beta_3 Surprise \times Quality_m + \varepsilon \quad \text{Model 3}$$

Abret= abnormal return

Bad News= it is the news of profit reduction proportional to the last year and it is equal to one if the surprise is positive/one, it is zero, otherwise.

Good news: it is an increase in the profit proportional to the last year. It is equal to one if the surprise is negative and it is zero, otherwise.

Control variable: company size.

Risk controls: systematic risk.

Accuracy: profit forecast accuracy.

Surprise: getting surprised.

4.2 DATA COLLECTION METHOD

The information required for the current study to test the study hypothesis is collected through a reference to the audited financial statements (existing in the Tehran securities organizations library) of the companies accepted in Tehran securities market and Rahavard Novin software (Iranian software) as well.

The utilized tools to collect the data include observation, statistical tests, information banks, SPSS software, and Excel software. The data related to the theoretical and subjective principles are collected mostly from the libraries and by making use of the books, Persian and English articles.

4.3 EMPIRICAL RESULTS

4.3.1 HYPOTHESIS ANALYSIS AND TEST METHOD

In the current study, the multivariate linear regression model has been used to analyze the data and test the hypotheses and F-value was used to test the significance of the overall model, and T-value was used in every model to assess the significance of the independent variables coefficient. To accept or to reject the hypotheses, decisions has been made in the confidence level of 95%. Also Colmogrof - Smearnov test and Durbin-Watson D - value were used to survey and determine the homogeneity of the experimental data with the statistical distributions and to test the error independence from each other, respectively.

4.3.2 VARIABLES NORMALITY ASSUMPTION

Since, variable normality results in the model residuals normality, it is necessary to control its normality before testing the goodness of the fitness. To test the dependent variable normality we use the Colmogrof - Smearnov test.

Table 1. Colmogrof-Smearnov test (K-S)

Abbreviations	Abret
Number of data	1044
Average	-0.1347
Standard deviation	0.75232
Most deviated modulus	0.101
The most positive deviations	0.101
The most negative deviations	-0.057
Z-value	3.254
Significance level	0

According to table 1, because dependent variable significance level is less than 0.05, H_0 hypothesis is rejected and H_1 hypothesis is accepted. That means the data follow a normal distribution. To normalize the variables, the mathematical conversion (log^2) is used.

Table 2 surveys the converted variables normality hypothesis.

Table 2. Colmogrof-Smearnov test (converted amount)

Abbreviations	LNABRT
Number of data	1044
Average	-2.4154
Standard deviation	2.5049
The Most deviated modulus	0.081
The most positive deviations	0.052
The most negative deviations	-0.081
Z-value	2.623
Significance level	0.111

According to table 2, because variables significance level is more than 0.05, H_0 hypothesis is accepted and H_1 hypothesis is rejected. In other words, data have normal distribution. Therefore, variables normality assumption is accepted.

4.3.3 THE FIRST HYPOTHESIS TEST RESULTS

H1: there is a significant relationship between the profit forecast accuracy and stock price response.

Table 3 is indicative correlation coefficient, determination coefficient, and Durbin-Watson test between profit forecast accuracy and profit forecast response.

Table 3. Correlation coefficient

Durbin-Watson value	Estimation standard error	offsetting determination coefficient	determination coefficient	Correlation coefficient	model
2.011	1.50931	0.291	0.312	.340[a]	1

Based on the table 3, Pierson correlation coefficient between the two variables, profit forecast accuracy and stock price response, is equal to 0.340. This figure shows a significant relationship between the two variables, profit forecast accuracy and stock price response, and also the calculated offsetting determination coefficient is 0.291, which is a good number and it offers an appropriate goodness of fitness for stock price response variable by using the profit forecast accuracy and according to table 3 the amount of Durbin-Watson value is 2.011 and this figure shows that the errors are independent from each other and there is no autocorrelation between errors and the regression can be used.

Table 4 shows regression variance analysis for the profit forecast accuracy and the stock price response.

Table 4. Regression variance analysis

Significance level	F-value	Squares mean	Degree of freedom	Total squares	Model	
.000[a]	5.637	18.587	1	10.587	regression	
		3.297	1037	6529.607	residual	1
			1038	6540.195	total	

Table 4 is indicative of the variance analysis between stock price response as the dependent variable and profit forecast accuracy as the independent variable. Considering the significance level which is less than 5%, the linearity assumption of the relationship between the two variables is confirmed.

Table 5. Regression equation coefficients for independent and control variables

Co linearity Statistics		Significance level	T-value	Standardize coefficient	Not Standardize coefficient		abbreviations	model
Variance inflation factor	tolerance			Beta	The column coefficient standard error B	B		
		0	-30.911		0.078	-2.409	Constant	1
1	1	0	-10.597	-0.44	0.43	-4.557	Accuracy	

In table 5 and in the column B the constant amount and the independent variable coefficient in the regression equation have been introduced and this equation is in the form of relation 7:

$$\mathrm{ABRET}_{\mathrm{it}} = -2.409 - 4.557 * \mathrm{Accuracy} \quad (7)$$

4.3.4 THE SECOND HYPOTHESIS TEST RESULTS

H2: There is a significant relationship between the management announced profit and the stock price response.

Table 6 shows correlation coefficient, determination coefficient and the Durbin-Watson test between getting surprised from the management announced profit and stock price response:

Table 6. Correlation coefficient

Durbin-Watson value	Estimation standard error	offsetting determination coefficient	determination coefficient	Correlation coefficient	model
2.006	1.49184	0.305	0.318	.435[a]	1

Based on the table 6, Pierson correlation coefficient between the two surprise variables from the management announced profit and stock price response is equal to 0.435. This figure shows a significance relationship between the two surprise variables from the management announced profit and stock price response in the error level of 5%.

Also, the calculated offsetting determination coefficient is 0.305 which is a good figure and it offers an appropriate goodness of fitness for the variable changes by use of getting surprised from the management announced profit and according to table 6 the amount of the Durbin-Watson value is 2.006, and this figure shows that the errors are independent from each other and there is no autocorrelation between errors and the regression can be taken advantage of.

Table 7 shows regression variance analysis for surprised from the management announced profit and stock price response.

Table 7. Regression variance analysis

Significance level	F-value	Squares mean	Degree of freedom	Total squares	Model	
.001[a]	4.825	29.957	4	119.83	Regression	
		6.209	1034	6420.365	Residual	1
			1038	6540.195	Total	

Table 7 shows the variance analysis between stock price response, as the dependent variable and getting surprised from the management announced profit, as the independent variable and according to the significance level which is below 5%, the linearity assumption between the two variables is confirmed.

Table 8. Regression equation coefficients to the independent and control variable

Co linearity Statistics		Significance level	t-value	Standardize coefficient	Non-Standardize coefficient		abbreviations	model
Variance inflation factor	Tolerance			Beta	The column coefficient standard error B	B		
		0.027	-2.198		0.66	-1.451	**Constant**	1
1.484	0.674	0.022	2.301	0.086	0.193	0.444	**Good news**	
1.436	0.696	0.021	2.306	0.085	2.126	4.903	**Bad news surprise**	
1.024	0.976	0.063	-1.864	-0.058	0.13	-0.241	**Control variable**	
1.49	0.671	0.018	2.361	0.089	0.646	1.525	**surprise**	

In table 8 and in column B, constant amount and independent variable coefficient has been introduced in the regression equation, respectively and this equation is in the form of relation 8:

$$ABRet_{it} = -1.451 + .444 * GoodNews + 4.903 * BadNewsSurprise + 1.525 * Surprise \quad (8)$$

4.3.5 THE THIRD HYPOTHESIS TEST RESULTS

H3: there is a significant relationship between systematic risk and stock price response.

Table 9 shows correlation coefficient, determination coefficient and Durbin-Watson test between systematic risk and stock price response.

Table 9. Correlation coefficient

Durbin-Watson value	Estimated standard error	Offset determination coefficient	Determination coefficient	Correlation coefficient	model
1.986	1.50447	0.316	0.349	.497[a]	1

According to table 9, the Pierson correlation coefficient of the two variables, systematic risk and stock price response, is equal to 0.497. This figure shows a significant relationship between the two variables, systematic risk and stock price response. Also, the calculated offsetting determination coefficient is 0.316, which is a good figure and it offers an appropriate goodness of fitness from the changes of variables for the stock price response by taking advantage of the systematic risk. Based on the table 7, Durbin-Watson value is 1.986 and this figure shows that the errors are independent from each other and there is no autocorrelation between errors and the regression can be used.

Table 10 shows regression variance analysis for the systematic risk and the stock price response.

Table 10. Regression variance analysis

Significance level	F-value	Squares mean	Degree of freedom	Total squares	model	
.001[a]	4.749	20.291	3	60.873	Regression	1
		4.272	1027	6441.746	Residual	
			1030	6502.62	Total	

Table 10 shows the variance analysis between stock price response variable as the dependent variable and systemic risk as the independent variable and because the significance level is less than 5%, the linearity assumption of the relationship between the two variables has been confirmed.

Table 11. Regression equation coefficient for the control and independent variables

Co linearity statistics		Significance level	t-value	Standardized coefficient	Non- Standardized coefficient		Abbreviations	Model
Variance inflation factor	Tolerance			Beta	Column coefficient standard error B	B		
		0	-30.24		0.082	-2.491	Constant	1
1.002	0.998	0	-4.86	-0.32	0.093	-0.452	Risk control	
1.24	0.309	0	-3.578	-0.325	0.958	-3.428	Surprise	
1.24	0.309	0	-3.979	-0.374	1.128	-4.489	Surprise quality	

In table 11 and in column B, the constant value and the independent variable coefficient are introduced, in order, in the regression equation and this equation takes the following form:

$$ABRet_{it} = -2.491 - .452 * RiskControl - 3.428 * Surprise - 4.489 * SurpriseQoulity \quad (9)$$

5 CONCLUSIONS AND DISCUSSIONS

In the first hypothesis test, according to the analysis performed based on the regression and correlation methods we came to this conclusion that there is a positive correlation coefficient between the independent variable(profit forecast accuracy) and dependent variable (stock price response) in the companies accepted in Iran's securities market and there is a

significant relationship between the accuracy of profit forecast and the stock price response in the companies accepted in Tehran's securities market.

According to the results obtained, there is an inverse relationship between the profit forecast accuracy and the stock price response in the companies accepted in the Tehran's securities market. That means that with an increase in the profit forecast accuracy, there would be a decrease in the response of the stock price and vice versa.

In the second hypothesis test and according to the analysis which has been performed by regression and correlation method, we came to this conclusion that there is a positive correlation between independent variable (getting surprised from the management announced profit) and dependent variable (stock price response) in the companies accepted in Iran's capital market and there is a significant relationship between getting surprised from the management announced profit with the stock price response in the companies accepted in Tehran's securities market, which means that with an increase in the surprise from the management announced profit, the stock price response goes up and vice versa.

In the second hypothesis test, according to the analysis which is performed by the regression and correlation method, we came to this conclusion that there is a positive correlation between independent variable (systematic risk) and the dependent variable (stock price response) in the companies accepted in Iran's capital market and there is a significant relationship between the systematic risk and stock price response in the companies accepted in Tehran's securities market.

According to the results obtained , there is a direct relationship between systematic risk and stock price response in the companies accepted in Tehran's securities market which means that with an increase in the systematic risk, the stock price response increases and vice versa.

REFERENCES

[1] J. Ng, I. Tuna, R. Verdi, (2012), "Management Forecasts, Disclosure Quality, and Market Efficiency," The Wharton School University of Pennsylvania 1303 Steinberg Hall-Dietrich Hall.

[2] R. J. Bloomfield, R. Libby, and M. W. Nelson, (2003), "Do Investors Over relies on Old Elements of the Earnings Time Series?"Contemporary Accounting Research 20 (1), (spring), pp. 1-31.

[3] R. J. best, (2002), "Prior information and the market reaction to dividend changes", Review of Quantitative Finance and Accounting, pp. 361–376.

[4] D.W. Collins and S.P. Kothari, (1989), "An Analysis of Inter temporal and Cross-sectional Determinants of Earnings Response Coefficients", Journal of Accounting and Economics 18, 49.

[5] W. Xie, (2008), "A New Measure of Analyst Earnings Forecast Dispersion", Working Paper, University of New Hampshire.

[6] W. Xie , (2009), "Analyst Earning Forecast Accuracy": A Second Look , University of New Hampshire.

[7] R. Brid, (2001), "The prediction Earnings Movements Using Accounting Data": An Update and Extension of Ou And Penman, Journal of Asset Management, 2, (February).

[8] W.H. Beaver, R. Clarke, and W. Wright, (1979), "The Association between Unsystematic Security Returns and the Magnitude of the Earnings Forecast Error," *Journal of Accounting Research* (Autumn 1979), pp. 316-340

[9] B. B. Ajinkya and M. J. Gift, (1984), "Corporate managers' earnings forecasts and symmetrical adjustments of market expectations," Journal of Accounting Research 22 (2), pp. 425-444.

[10] G. Waymire, (1984), "Additional evidence on the information content of management earnings forecast" Journal of Accounting Research 22 (2), pp. 703-718.

[11] R. Jennings, (1987), "Unsystematic security price movements, management earnings forecasts, and revisions in consensus analyst earnings forecasts," Journal of Accounting Research 25 (1), pp. 90-110.

[12] J. L. Rogers and P. C. Stocken, (2005), "Credibility of management forecasts," The Accounting Review 80 (4), pp. 1233-1260.

[13] A. Brav, and J. B. Heaton, (2002), "Competing theories of financial anomalies," The Review of Financial Studies 15 (2), pp. 575- 606.

[14] E. Bartov, S. Radhakrishnan, and I. Krinsky, (2000), "Investor sophistication and patterns in stock returns after earnings announcements," The Accounting Review 75, pp. 43-63.

[15] J. Rogers, (2006), "Disclosure quality and management trading incentives," University of Chicago working paper.

[16] C. Lee, (2001), "Market efficiency and accounting research: a discussion of 'capital market research in accounting' by S.P. Kothari," Journal of Accounting and Economics 31 (1-3), pp. 233-253.

[17] P. A. Williams, (1996), "The relation between a prior earnings forecast by management and analyst response to a current management forecast," The Accounting Review 71 (1), pp. 103-113.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 857-865

http://www.issr-journals.org/ijias/

Literature Review of Automatic Single Document Text Summarization Using NLP

Md. Majharul Haque[1], Suraiya Pervin[1], and Zerina Begum[2]

[1]Department of Computer Science & Engineering,
University of Dhaka,
Dhaka, Bangladesh

[2]Institute of Information Technology,
University of Dhaka,
Dhaka, Bangladesh

ABSTRACT: In the time of overloaded online information, automatic text summarization is especially demanded for salient information retrieval from huge amount electronic text. For the blessing of World Wide Web, the mass of data is now enormous in its volume. Researchers realized this fact from various aspects and tried to generate an automatic abstract of the gigantic body of data from the commencement of the last half century. Numerous ways are there for characterizing different approaches to passage recapitulation: extractive and abstractive from single or compound document, objective of content abridgement, characteristic of text summarization, level of processing from superficial to profound and sort of article's content. A significant précis is very much helpful in our day to day life which can save valuable time. The investigation was at first commenced naively on single document abstraction. In this paper, automatic single document text summarization task is addressed and different methodologies of various researchers are discussed from the very beginning of this research to this modern age. This literature review intends to observe the trends of abstraction procedure using natural language processing. Also some promising approaches are indicated and particular concentration is dedicated for the categorization of diversified methods from raw level to similar like human professionals, so that in future one can get precious direction for further analysis.

KEYWORDS: Information retrieval, World Wide Web, electronic text, automatic abstract, human professionals.

1 INTRODUCTION

Outcome of the information retrieval becomes necessary for user to find out concrete information for the abstraction because of the stridently escalation of data on the web. Internet is widely used by people to come across information using proficient information retrieval (IR) tools, such as Google, Yahoo, AltaVista, etc., where findings are abundant. In most of the cases, users feel bore with the very tedious and time consuming job to reveal the main gist of the outcome of the IR. Academics and researchers are very much benefitted by using automatic text summarization system as a tool to lessen the amount of time spent manually extracting the chief thoughts from large documents. In addition to the above reason, automatic text summarization also provides its users with numerous benefits as well as:

(i) Increase efficiency of other researches to choose documents/information from search engines' output, which usually contain an excess amount of replicated information.

(ii) Solve the limitation of information presentation on small communication devices such as PDA and mobile phone etc., which is able to display abridged version of the full document.

(iii) The running time of machine for translation is significantly reduced if a short version of text is given.

Corresponding Author: Md. Majharul Haque (mazharul_13@yahoo.com)

American research libraries spawned the initial interest in automatic text shortening during sixties [1]. A considerable body of research over the last sixty years has explored different levels of analysis of text, to help determine what information in the text is salient for a given summarization task [1]. Radev et al [2] in 2002 defined a summary as a text that is engendered from one or more texts, which conveys essential information of the original text(s), and that is no longer than half of the original text(s) and usually notably less than that. Simply a summary text is a derivative of a source text condensed by selection and/or generalization on important content [3]. A large document is entered into the computer and a recapitulated content is returned, which is a non redundant extract from the original passage. Automatic text summarization can be classified into single document text summarization and multiple documents text summarization [4] as in figure 1. This paper focused on single document text summarization as single document was the target from the commencement of such research on automatic abstraction.

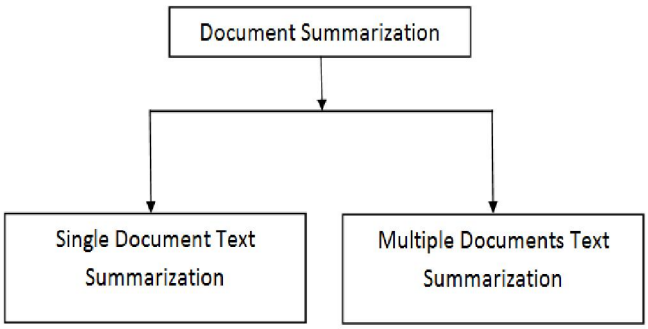


Fig. 1. Classification of document summarization

Various methods that utilize passage categorization, such as neural networks, semantic graphs, decision trees, regression models, fuzzy logic and swarm intelligence, etc. are involved on the study on finding crucial portion of text. The objective of this paper is to present a comprehensive literature review on automatic single document text summarization using natural language processing and investigate the movement of document abridgement.

The rest of the paper is organized as follows. Section 2 briefly describes about Natural Language Processing. Section 3 presents a comprehensive literature review about different procedures of automatic single document text summarization. At a glance comparison among the various techniques is depicted in section 4. Section 5 turns conclusion with a brief about this paper.

2 NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is a field of computer science, artificial intelligence and linguistics as all those specified arena brought it into play. Generally it deals with the interactions between machines and human languages that accomplish task on analyzing, understanding and generating the language, which human use naturally in order to interact with computers in both oral and written contexts using natural human languages instead of computer languages [5]. It is an interdisciplinary area based on versatile arena of study including computer engineering, which provides methods for model illustration, algorithm devise and accomplishment; linguistics, which categorizes linguistic forms and practices; mathematics, which provides formal models and methods; psychology, which studies models and theories of human behavior; statistics, which offers procedures for predicting measures based on sample records; and biology, which travels around the underlying architecture of linguistic processes in the human brain [6].

3 REVIEW ON AUTOMATIC SINGLE DOCUMENT TEXT SUMMARIZATION

In the very beginning of the research in the arena of launching artificial intelligence to generate abridged version of a large document, disclosed the paradigms for extracting salient features.

Automatic text summarization process model can be divided into three steps [7] as in figure 2: (1) in the preprocessing step source text interpretation to source text representation, (2) source representation transform to summary text representation with an algorithm and (3) in the final step, summary text generation from summary representation.

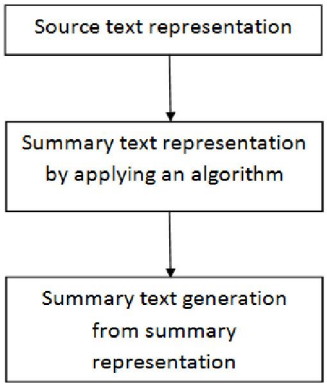


Fig. 2. Process flow of automatic text summarization

Numerous types of research work have been accomplished by various researchers where we can be familiar with multiple types of way of summary generation from single document text. In the following part of this section, methodologies of several researchers that incorporated this topic are depicted in brief, from pioneering works to the era of modern science where the thoughts of similar like human professionals' abstraction techniques are being explored.

3.1 PIONEERING WORKS

The voice over automatic i.e. computerized abstraction initiated around sixty years ago. As accomplishment of automatic passage summarizer was often cited in the oldest publication in 1958 by H. P. Luhn [8]. This method is based on the word frequency and clearly emphasized that the most frequent words represent the most important concepts of the text. In the next step, frequencies are used to score and decide on sentences to be extracted. In this paper, application of machine is emphasized and expressed that because of the absence of the variations of human capabilities and orientation, auto-abstracts have a high degree of reliability, consistency, and stability, as they are the product of statistical analysis of the author's own words. This paper is worthy of being appreciated as it is almost the earliest paper in this arena of automatic text abstraction [9]. Besides this, the proposed method mostly works on scientific article and journal paper.

P. B. Baxendale [10] in 1958 investigated machine techniques for reducing scientific credentials to their important sharp indices. Human scanning patterns were tried to be simulated here for selecting topic sentences special phrases. It was declared in this paper that sentence' position and containing certain cue-words (i.e., words like 'crucial' or 'pertinent' etc.) or the word exist in the heading are special indicator for being in the salient parts of the document.

G. J. Rath et al [11] in 1961 in their research scrutinized about four types of lexical indicators of content to determine which one is the best for detecting relevant document from repository and to answer a set of question. After their experiment, it was claimed that utilizing complete text is better than only abstract for answering question. But for distinguishing relevant document only abstract is enough. They also proclaimed that purely statistical method of producing extracts was suspected of being inadequate, and hence other methods were sought.

H. P. Edmundson [12] in 1969 accomplished a notable progress after ten years of the beginning of the research on text recapitulation. He described three additional methods with the standard keyword method, disregarding the very high frequency common words to determine the sentences' weight. Those are:

(i) Cue Method: The hypothesis of this technique is that the presence or absence of certain cue words will compute the significance of a sentence.

(ii) Title Method: The weight of a sentence is calculated as a sum of all the content words materializing in the title, headings and sub-headings of a text.

(iii) Location Method: Here relevance is assumed on the basis of location, expressed that sentences taking place in initial position of paragraphs have a higher probability of being pertinent.

The result was very fruitful and assumed that by using a combination of these three schemes the best correlation between the automatic and human-made abstracts can be achieved. This paper emphasized on indicative abstracts rather than on the production of informative abstracts. So, if a user doesn't know about the document to be summarized earlier, can't get summary through this proposed methodology.

3.2 ALGEBRAIC METHODS

Julian Kupiec et al [13] in 1995 explored an innovative algebraic method. Using naïve-bayes classifier, this classification procedure sorts out each sentence as worthy of being in the abstract or not. A trained classifier is generated at first, where the authors used a corpus of 188 couples of complete documents/summaries. The distinguishing features used uppercase words, length of sentence, structure of phrase, position in paragraph besides word frequency.

ChinYew Lin et al [14] in 1997 in their paper addressed the difficulty of identifying probable topics of texts by their location in the corpus, and scored sentences by its position in given text. They considered techniques of tailoring the position method towards optimality over a variety and how it can be estimated for efficiency. The idea that texts generally follow a predictable discourse structure, and that the sentences of greater topic centrality tend to occur in certain specifiable locations (e.g. title, abstracts etc), arises the position method for topic recognition. Here also predicted that discourse structure significantly varies over domains, for which position method is a bit tough. Given a text T and the list of topic keywords, each sentence of the text is labeled with the paragraph number and the sentence number. With the number of 2097 documents in this method, the result illustrated that the first and the last 50 positions fully cover majority of the text.

Eduard Hovy et al [15] in 1999 attempted to create a robust automated text summarization system and named it SUMMARIST. Instead of irregular term counting, SUMMARIST combines symbolic world knowledge (embodied in WordNet, dictionaries, and similar resources) with strong NLP processing (using IR and statistical method) to resolute concept relevance. Their procedure based on the following equation:

Summarization = topic identification + interpretation + summary generation

S. P. Yong et al [16] in 2005 introduced an automatic text summarization system that incorporates learning ability by combining a statistical approach, keywords extraction and neural network with unsupervised learning. It was claimed that their proposed technique can extract 83.03% of significant contents. The procedure is made up of three steps, as follows:

(i) Text pre-processing subsystem: two pre-processing methods are applied, one is stop words removal to remove words like “the”, “a”, “can” and “will” etc. and another one is stemming to convert each word to its stem by eliminating suffixes and prefixes.

(ii) Keywords extraction subsystem: from a group of sample heterogeneous text documents, main features of each word are determined by computing term frequency tf(w, s) as the number of times that the word w appears in the sentences s, and the inverse sentence frequency isf(w) is calculated as the number of sentences in which the word w occurs. From the generated tf-isf(w, s) matrix, most frequent terms are listed as keyword in the text to be summarized.

(iii) Summary production subsystem: in the final step the system chooses sentences, which contain the keywords, as part of the summary. It is suggested to run through another round of stop words checking procedure before selecting sentences for being ensured that there is no stop word is working as keyword.

3.3 PROCEDURES BASED ON TEXT CORRELATION

Previously discussed methods used statistical probability for choosing a sentence for generating abridged version of text. But those panoramas are fully dumb about the cohesion of sentences with each other. So, the relations between concepts in a text can’t be captured using extractive methods. If a sentence is extracted which contains link to any previous context then the summary will be difficult to understand [17]. In those circumstances various research work done for exploring the cohesive properties to comprise relations among expressions of text.

M. A. K. Halliday et al [18] in 1976 performed the first research to explore the degree of subjectivity of two aspects of the lexical cohesion perceived by readers of text: the word cluster (lexical chains) that are formed and the lexical semantics relations that are perceived between the words. Building blocks of lexical cohesion, cohesive harmony, and the concept of patterns of lexical affinity are the lexical semantic relations. The linguistic study was emphasized here and tried to forms inter sentence groups of related words, and the original outlook of them was very wide and universal; but there had to be a recognizable relation between two words as the only decisive factor.

Ruqaiya Hasan [19] in 1984 extended the investigation to include the concept of cohesive harmony, which adds lexico-grammatical structure to word groups by first dividing them into two types: (i) identity-of-reference chains that combine reference and lexical cohesion, (ii) similarity chains that used only classical relations, and linking these chains with grammatical intra-sentence relations.

Rhetorical Structure Theory (RST) of text summarization can also be included into the groups of methods based on text correlation as the text coherence is attributed principally to the presence of rhetorical relation. RST was developed during 1980s by researchers in natural language generation, which models the discourse structure of a text by means of a hierarchical tree diagram [20].

William C. Mann et al [21] in 1988 established a new definitional foundation of RST and also scrutinized three claims: the predominance of nucleus/satellite structural patterns, the functional basis of hierarchy, and the communicative role of text structure. Various kinds of consequences of RST have also been reviewed in this paper. A quantity of allegorical relations, which hook up together text units, forms a hierarchical structure called RST tree. The relations tie collectively a nucleus as central to the author's objective and a satellite as subsidiary parts. Texts in terminal nodes of RST tree are supposed to be encoded and non-terminal nodes represent contiguous text spans, whose sibling spans are joined by discourse relations. At last the nucleus and some tightly connected terminal nodes are selected as the principal theme of passage.

Jane Morris et al [22] in 1991 utilized cohesion chains and Regina Barzilay et al [23] in 1997 exploited lexical chains to characterize contents of a document. A perception in the document is symbolized by a sequence of associated words in these representations. The sequence is a list of words that confines a segment of the consistent structure of the document and is generated as independent of the grammatical formation of the text. The procedure usually launch from a set of words in the heading of the document to construct lexical chains, adjoining of words that have similar meaning or are related to the title. WordNet thesaurus has been used for determining cohesive relations between terms (i.e., repetition, synonymy, antonymy, hypernymy, and homonymy). In the next step, scores are calculated for sentences considered to be important by the previous step and those scores will be used to generate the final summary.

Branimir Boguraev et al [24] in 1997 proposed a novel technique for saliency-based content characterization of text document. A procedure was offered, which uses a salience feature as the basis for a "ranking by importance" of an unstructured referent set, and ultimately for topic stamp identification. Co-reference resolution system is used here, which is a process of determining whether two expressions in natural language refer to the identical entity. Then a threshold value is calculated and where the occurrences of frequently mentioned objects overcome the value is included into the summary.

Li Chengcheng [25] in 2010 presented an effective method using RST for successful automatic text summarization, which is based on natural language generation. The rhetoric structure of the text is extracted with a compound that relates the sentences in this theory. Here the summarization is accomplished using the nodes i.e. nucleus, those are given weights based on script based analysis. Then the entire text is divided into individual sentences based on commas, quotes, semicolons and punctuation marks exist in the sentences. This is then done into a graph, deletes the unimportant sentences and then summarizes the entire text. Principal scheme of this procedure is as follows: (i) analyzing the candidate sentence, (ii) discover the rhetoric relations and (iii) forming the essential part of sentence constructive for ultimate recapitulation.

3.4 METHODS CLOSE TO HUMAN ABSTRACTION CONCEPT

Artificial Intelligence is much matured in this modern age, as the application of this is noticeable in various regions such as robotics, machine learning and knowledge based system etc. Automated summary is also the grace of Artificial Intelligence through natural language processing. But, still there is a qualitative dissimilarity between synopsis generated by existing automated summarizers and the abstracts written by human. Extraction is a common technique utilized by most of the summarizers, which may be ambiguous or incoherent to the original abstract. In spite of the existence of some shortcomings, a number of methods have started to emerge lately with either sentence compressing capability or re-producing technique.

Hongyan Jing [26] in 2000 presented a novel sentence lessening method, which eliminate extraneous phrases from the extracted sentences that are chosen for abstraction purpose. A parse tree is generated at first in this procedure and then grammar is checked to decide which terms of a sentence must not be removed to keep its' structure accurate. Then it finds the sentences that are closely related to the main topic by using corpus evidence, consisting of sentences compressed by human and original sources, delete unnecessary sub-tree from parse tree for producing the final outline.

Kevin Knight et al [27] in 2000 proposed an innovative method by incorporating the procedure of regenerating sentence for coherent abstract, rather than just extracting from original source. A training corpus is used here, which is a collection of newspaper articles paired with human written abstracts. Their first goal is to rewrite a compressed version of given input parse tree. The rewriting process starts with an empty Stack and Input List that contains the sequence of words subsumed by the given large parse tree. Four types of operations are incorporated here: SHIFT operations move the first word from the input list into the stack; REDUCE operations pop the top syntactic trees, combine them and again push them to the stack; DROP operations are used to remove from the input list that has already been compressed; ASSIGN TYPE operations change label of trees at the top of the stack. The procedure ends when the input list is vacant and the stack contains only one tree.

After that an in-order traversal of the leaves of this tree generates the compressed form of the sentences those are given as input.

K. McKeown et al [28] in 2000 stated that there is a significant difference between the summaries produced by current automatic system and the human professionals. Because automatic summarizer cannot always recognize key topics of an article and automatic procedure has no robust text generation method. This paper presented a "Cut and Paste" strategy for text summarization that derived from an analysis of human written abstract. "Cut and Paste" method not just extracts sentences but smooth the extracted sentences by editing them which mainly cutting phrases and pasting them together in a novel ways. It was stated that generated summary bears a resemblance to the human summarization process more than extraction does. Six operations were defined to transform chosen sentences from an article into the corresponding summary sentences in its human written abstracts: (i) sentence reduction, (ii) sentence combination, (iii) syntactic transformation, (iv) lexical paraphrasing, (v) generalization and specification, and (vi) reordering. An evaluation was done for this procedure with 50 human-written abstracts, consisting of 305 sentences in total and claimed that 93.8% of sentences were correctly decomposed.

4 COMPARISON AMONG THE TECHNIQUES

At a glance comparison among the discussed techniques of single document text summarization has been shown in table 1:

Table 1. Comparison Among the Techniques of Single Document Text Summarization

#	Researcher(s), Year, Reference	Category	Technique	Basis of procedure
1	H. P. Luhn, 1958, [8]	Pioneering works	Word frequency	Frequent words represent the most important concepts of the text.
2	P. B. Baxendale, 1958, [10]	Pioneering works	Position in text	Sentence position and containing certain cue-words or the word exist in the heading are special indicator for being the salient parts of the document.
3	G. J. Rath, 1961, [11]	Pioneering works	Lexical indicator	Lexical indicators of content could be utilized best by subjects to determine relevant from irrelevant documents.
4	H. P. Edmundson, 1969, [12]	Pioneering works	Cue words and heading	The sentences that contain specific cue words, words those are exist in title and heading or sub-heading, are significant to be selected for summary.
5	Julian Kupiec, 1995, [13]	Algebraic method	Naïve-Bayes classifier	Uppercase words, length of sentence, structure of phrase, position in paragraph with word frequency are the distinguishable features of sentence.
6	ChinYew Lin, 1997, [14]	Algebraic method	Position in text	Texts generally follow a predictable discourse structure, and that the sentences of greater topic centrality tend to occur in certain specifiable locations (e.g. title, abstracts, etc).
7	Eduard Hovy, 1999, [15]	Algebraic method	Symbolic world knowledge	Instead of irregular term counting, this method combines symbolic world knowledge (embodied in WordNet, dictionaries, and similar resources) with strong NLP processing (using IR and statistical method) to resolute concept relevance.
8	S. P. Yong, 2005, [16]	Algebraic method	Neural network	Using neural based system this method generate summary using three sub system such as: Summarization = Text pre-processing sub-system + Keywords extraction sub-system + Summary production sub-system.
9	M. A. K. Halliday, 1976, [18]	Text correlation	Lexical cohesion	Building blocks of lexical cohesion, cohesive harmony, and the concept of patterns of lexical affinity are the lexical semantic relations.
10	Ruqaiya Hasan, 1984, [19]	Text correlation	Lexical cohesion	Identity-of-reference chains and similarity chains, that linking with grammatical intra-sentence relations.
11	William C. Mann, 1988, [21]	Text correlation	Rhetorical Structure Theory(RST)	Texts in terminal nodes of RST tree are supposed to be encoded and non-terminal nodes represent contiguous text spans.
12	Jane Morris, 1991, [22]	Text correlation	Cohesion chains	A perception in the document is symbolized by a sequence of associated words in the representation.
13	Regina Barzilay, 1997, [23]	Text correlation	Lexical chains	The procedure usually launch from a set of words in the heading of the document to construct lexical chains, adjoining of words that have similar meaning or are related to the title.
14	Branimir Boguraev, 1997, [24]	Text correlation	Saliency-based content characterization	This method uses a salience feature as the basis for a "ranking by importance" of an unstructured referent set.
15	Li Chengcheng, 2010, [25]	Text correlation	Rhetorical Structure Theory(RST)	Principal scheme of this procedure is as follows: (i) analyzing the candidate sentence, (ii) discover the rhetoric relations and (iii) forming the essential part of sentence constructive for ultimate recapitulation.
16	Hongyan Jing, 2000, [26]	Human abstraction concept	Lessening method	Find the sentences that are closely related to the main topic and eliminate extraneous phrases from the extracted sentences that are chosen for abstraction purpose.
17	Kevin Knight, 2000, [27]	Human abstraction concept	Regenerating sentence	Rewrite a compressed version of given input parse tree. The rewriting process starts with an empty Stack and Input List that contains the sequence of words subsumed by the given large parse tree.
18	Hongyan Jing, 2000, [28]	Human abstraction concept	Cut and paste	Smooth the extracted sentences by editing them which mainly cutting phrases and pasting them together after reduction, combination, transformation, specification, reordering, etc.

5 CONCLUSION

In this paper the concepts of single document text summarization that categorize different approaches in this arena have been reviewed. This literature review emphasizes on extractive approaches for text summarization. Deriving the classification of the automatic text abstraction procedure has also been attempted. Recent trend in summarization system that comes from novice procedure to resemble with human written summary has been scrutinized here. We believe that the study of document summarization is a productive region for further research, both by linguists performing text analysis and by computational linguists trying to create techniques to produce summaries conforming to one or more of the characteristics listed above. Around 18 papers have been briefly discussed and various key topics from other historical publication relevant with text abstraction have been analyzed here from 1958 to 2010. There exist some other procedures similar with those briefed in this paper, the discussion of which has not been included here as it will be a large corpus. Nowadays, in the age of computer and internet, information is found from various sources about a single topic, so the research of multi-document text abstraction is a burning issue. After all it is expected that any researchers can get help from this literature review for better understanding of different types of summary generation techniques even for multi-document text abridgement. It will also assist for better perception of the diversified sorts of abstraction procedure, which will help to construction of new formula and systems that significantly serve the various principle of summarization in broad-spectrum.

REFERENCES

[1] Martin Hassel, "Evaluation of Automatic Text Summarization - A practical implementation," Licentiate thesis, Department of Numerical Analysis and Computer Science, Royal Institute of Technology, Stockholm, Sweden, 2004. [Online] Available: http://www.csc.kth.se/~xmartin/papers/licthesis_xmartin_notrims.pdf (May 21, 2013)

[2] Dragomir R. Radev, Eduard Hovy and Kathleen McKeown, "Introduction to the special issue on summarization," *Journal of Computational Linguistics*, vol. 28, no. 4, pp. 399-408, December 2002.

[3] Abdelwahab Hamou-Lhadj and Timothy Lethbridge, "Summarizing the Content of Large Traces to Facilitate the Understanding of the Behaviour of a Software System," Proceedings of the 14th IEEE International Conference on Program Comprehension, pp. 181-190, 2006.

[4] S. Suneetha, "Automatic Text Summarization: The Current State of the art," *International Journal of Science and Advanced Technology (ISSN 2221-8386)*, vol. 1, no. 9, pp. 283-293, November 2011.

[5] Eugene Charniak and Drew McDermott, Introduction to artificial intelligence, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 1985.

[6] Bill Z. Manaris, Natural Language Processing: A Human-Computer Interaction Perspective, University of Southwestern Louisiana, Academic Press, New York, vol. 47, pp. 1-66, 1998.

[7] Karen Sparck Jones, Automatic summarizing: factors and directions, In: Inderjeet Mani and Mark T. Maybury (Eds.), Advances in Automatic Text Summarization, Cambridge MA: MIT Press, pp. 1-12, 1999.

[8] Hans P. Luhn, "The Automatic Creation of Literature Abstracts," *IBM Journal of Research and Development*, vol. 2, no. 2, pp. 159-165, 1958.

[9] Dipanjan Das and André F. T. Martins, "A survey on Automatic Text Summarization," Language Technologies Institute, Carnegie Mellon University, 2007.
[Online] Available: http://www.cs.cmu.edu/~afm/Home_files/Das_Martins_survey_summarization.pdf (May 21, 2013)

[10] P. B. Baxendale, "Machine-made Index for Technical Literature -An Experiment," *IBM Journal of Research and Development*, vol. 2, no. 4, pp. 354-361, October 1958.

[11] G. J. Rath, A. Resnick and T. R. Savage, "Comparisons of four types of lexical indicators of content," *Journal of the American Society for Information Science and Technology*, vol. 12, no. 2, pp. 126-130, April 1961.

[12] H. P. Edmundson, "New Methods in Automatic Extracting," *Journal of the Association for Computing Machinery*, vol. 16, no. 2, pp. 264-285, April 1969.

[13] Julian Kupiec, Jan Pedersen and Francine Chen, "A Trainable Document Summarizer," Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 68-73, 1995.

[14] Chin-Yew Lin and Eduard Hovy, "Identifying Topics by Position," In Proceedings of the Fifth conference on Applied natural language processing, San Francisco, pp. 283-290, 1997.

[15] Eduard Hovy and Chin-Yew Lin, Automated Text Summarization in SUMMARIST, In: Inderjeet Mani and Mark T. Maybury (Eds.), Advances in Automatic Text Summarization, MIT Press, chapter 8, pp. 18-24, 1999.

[16] S. P. Yong, A. I. Z. Abidin and Y. Y. Chen, "A Neural Based Text Summarization System," 6th International Conference of Data Mining, pp. 45-50, 2005.

[17] Karel Jezek and Josef Steinberger, "Automatic Text Summarization (The state of the art 2007 and new challenges)," Znalosti, pp. 1-12, 2008.
[18] M. A. K. Halliday and Ruqaiya Hasan, Cohesion in English, Longman, London, 1976.
[19] Ruqaiya Hasan, Coherence and Cohesive Harmony, In: Flood James (Ed.), Understanding Reading Comprehension: Cognition, Language and the Structure of Prose. Newark, Delaware: International Reading Association, pp. 181-219, 1984.
[20] William C. Mann and Sandra A. Thompson, Relational Propositions in Discourse, Defense Technical Information Center, Information Sciences Institute, Marina del Rey, 1983.
[Online] Available: http://www.tandfonline.com/doi/abs/10.1080/01638538609544632?journalCode=hdsp20#preview
[21] William C. Mann, Sandra A. Thompson, "Rhetorical Structure Theory: Toward a functional theory of text organization," Text, vol. 8, pp. 243-281, 1988.
[Online] Available: http://www.cis.upenn.edu/~nenkova/Courses/cis700-2/rst.pdf (May 21, 2013)
[22] Jane Morris and Graeme Hirst, "Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text," *Journal of Computational Linguistics*, vol. 17, no. 1, pp. 21-48, March 1991.
[23] Regina Barzilay and Michael Elhadad, "Using Lexical Chains for Text Summarization," In Proceedings of the ACL/EACL-97 Workshop on Intelligent Scalable Text Summarization, Association of Computational Linguistics, pp. 10-17, 1997.
[24] Branimir Boguraev and Christopher Kennedy, "Salience-based Content Characterization of Text Documents," In Proceedings of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization, 1997.
[25] Li Chengcheng, "Automatic Text Summarization Based On Rhetorical Structure Theory," International Conference on Computer Application and System Modeling (ICCASM), vol. 13, pp. 595-598, October 2010.
[26] Hongyan Jing, "Sentence Reduction for Automatic Text Summarization," In Proceedings of the 6th Applied Natural Language Processing Conference, Seattle, USA, pp. 310-315, 2000.
[27] Kevin Knight and Daniel Marcu, "Statistics-Based Summarization Step One: Sentence Compression," In Proceeding of the 17th National Conference of the American Association for Artificial Intelligence, pp. 703-710, 2000.
[28] Hongyan Jing and Kathleen R. McKeown, "Cut and Paste Based Text Summarization," In Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics, Seattle, USA, pp. 178-185, 2000.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 866-870

http://www.issr-journals.org/ijias/

Tile adhesive production by Inorganic materials

Fasil Alemayehu Hayilu and Omprakash Sahu

Department of Chemical Engineering,
KIOT, Wollo University,
Kombolcha, Ethiopia

ABSTRACT: In modern construction, ceramic tile and mosaic which are used for finishing and decoration are attached to the surface by using tile adhesives. It was a long way for tiling technology to arrive at the current cement based modified adhesive. The development in additives and modifier are the paramount factor to improve workability, higher flexibility, and better adhesion. In this document tile adhesive has been produced for economical and high performance formulation. These products have been produced by considering the effect of aggregate. These two products with different size of aggregate have been compared and tested. The test made was slip, bending, and compression test. Economical formulation consists of components like cement, quartz sand, cellulose ether and tartaric acid. But high performance consists of limestone and cellulose fiber in addition to these components. The modifier added has enhanced the final product resistance to sliding, bending and compression strength. In terms of compression strength test about 17.27% high performance is stronger than economical formulation. And in addition high performance is stronger than economical formulation by about 16.89% in terms of bending strength. The other thing is the effect of grain size, the component that has low grain size have shown great strength and resistant to slide.

KEYWORDS: Aggregate, Compression, Formulation, Modifier, Strength.

1 INTRODUCTION

An adhesive is a material that is applied to the surfaces of articles to join them permanently by an adhesive bonding process. Substance capable of forming bonds to each of the two parts when the final object consists of two sections that are bonded together. [1]. Adhesive or glue is a material, usually in a liquid, semi liquid and solid state that adheres or bonds items together. It is a substance capable of holding material together by surface attachment [5]-[6]. Cement-based mineral adhesives were exclusively produced and applied by so-called job-site mixing technology. Job-site mixing means transportation of the individual raw materials to the job-site and then mixing on site in the appropriate ratio. Thus, cement, the most common mineral binder, is mixed with fillers (sand) before water is added to create the wet adhesive for application. Adhesives are produced in specially designed dry-mix mortar plants in which mineral binder(s) and aggregates (sand) are mixed together in the appropriate way [2]. This factory-based process also allows different additives and admixes to be added to these adhesive to improve significantly their technical performance. Based on this technology individual dry mortars for specific applications can be produced according to formulations developed and pretested in the laboratory. Because of there is a fast-growing demand in the construction industry for new building materials and technologies, it resulted shortage of skilled workmen, the need for shorter construction times together with cost reduction, increasing labor costs, the diversification of building materials suitable for specific applications, new materials, and an increased demand for better quality of construction. Job-site mortar technology was not able to adequately meet all these requirements. [3].

The majority of the adhesive used in the construction industry is concerned with fastening decorative finishing materials to the insides and outside of building. For instance, the attachment of ceramic tiles and mosaics to floors and walls, wooden and flexible floor coverings, ceiling tiles, thermal insulation materials, wall veneers, covings, nosing, and so on, accounts for the usage of large amounts of a variety of adhesive materials. Tile adhesives are used for fixing tiles on various substrates [8].

Corresponding Author: Omprakash Sahu (ops0121@gmail.com)

They are powder based on cement as a binder and require the application of water prior to the application. Pre-packed polymer modified tile adhesives are widely used in the modern building industry, e.g. as adhesives for fixing tiles and the installation of thermal insulation materials, as tile grout mortars, mineral decorative finishes and stuccos, self-leveling screeds and underlayment, waterproofing sealing slurries, repair mortars, jointing compounds, key-coats, masonry adhesives etc [9]. Due to the high diversification of modern building materials, these mortars must meet various technical requirements like good adhesion on all kind of substrates, high deformation ability (flexibility) and excellent durability even under extreme climatic conditions all over the world [4]. The Production of tile adhesive uses local mineral resources and industrial polymers, the raw material composition is usually as follows: cementing material made of Portland cement; the main aggregates are silica sand or quartz sand, limestone; minerals admixture mainly industrial products [7]. This thesis focuses on the production of tile adhesive and investigating the effect of modifiers. And as well it studies the additives added, whether it is selected depending on the climatic condition of the environment. After producing this adhesive, its strength is checked using strength testing machine. The production of tile adhesive has its own quality and performance, so in this thesis the task accomplished is to make quality product for given environment. The aim of the work is to determine the effect of aggregates and additives, preparation of standard tiles adhesive, and determining the best composition of tile adhesive

2 METHODS

2.1 MATERIAL REQUIRED

Jaw crusher, rotor mill, rotor disk mill, mixer, sieve, ceramic tile, heat oven balance, hammer, wood, nails, and universal tensile- compression testing machine. Portland cement, sand (quartz), limestone, fibers (e.g. cellulose fibers), cellulose ether, calcium chloride (calcium formate or lithium carbonate), citric acid or tartaric acid.

2.2 EXPERIMENTAL WORK

This document tries to address the issue of tile adhesive production. And it relies on cement-based tile adhesive because of it has some advantage and availability of chemicals and raw material. The work of this project depends on literature review on tile adhesive and related products. The final goal of this thesis is also dependent on laboratory work on tile adhesive from mineral resource.

2.2.1 PREPARING OF ECONOMICAL FORMULATION TILE ADHESIVE

This is an adhesive prepared for cost minimization and need minimum numbers of raw materials [10].

Table 1. Raw material for economic formulation with composition

No.	Raw material	Grain size	Composition (%)
1	Quartz	0-0.5mm	74.25
2	Portland cement	As prepared by the supplier	25
3	Cellulose ether	As prepared by the supplier	0.5
4	Tartaric acid	As prepared by the supplier	0.25

For the purpose of meeting this, raw materials are ground using grinding machine that exist in the laboratory. After these raw materials are prepared with appropriate ratio and then mixed [6].

2.2.2 PREPARING OF HIGH PERFORMANCE FORMULATION TILE ADHESIVE

This product has been produced using modifier and chemical additives with regard to environmental condition. And it has been prepared with different grain size [10].

Table 2. Raw material for high performance formulation with composition

No.	Raw material	Grain size	Composition (%)
1	Quartz sand	0 – 0.5mm	55
2	Ground limestone	0-20 µm	5
3	Portland cement	As prepared by the supplier	38
4	cellulose fibers	As prepared by the supplier	1
5	Carbomethyl cellulose	As prepared by the supplier	0.5
6	Tartaric acid	As prepared by the supplier	0.25

2.2.3 STRENGTH TEST FOR THE PRODUCTS

Using universal tensile strength test machine the strength of the product has been tested [7].

1. For the purpose of this test, the specimens have been prepared in mold form.
2. After seven days the samples have been tested.
3. After this the samples have been inserted in heater to make it dry around 70^0C.
4. Universal strength test machine has been used for test of compression and bending strength of specimen.

2.2.4 SLIP TEST

A tile adhesive should have a good slip resistance mainly for the following two reasons: for tiling with heavy tiles like marble tiles, and second for countries where the tiling of a wall starts from the top [8].

3 RESULTS AND DISCUSSION

After accomplishing the experiment, the anticipated results that were proposed have been found. The production of high performance and economical formulation was done. The next step is testing the products. From experiments the following result has been found.

3.1 SLIP TEST RESULTS

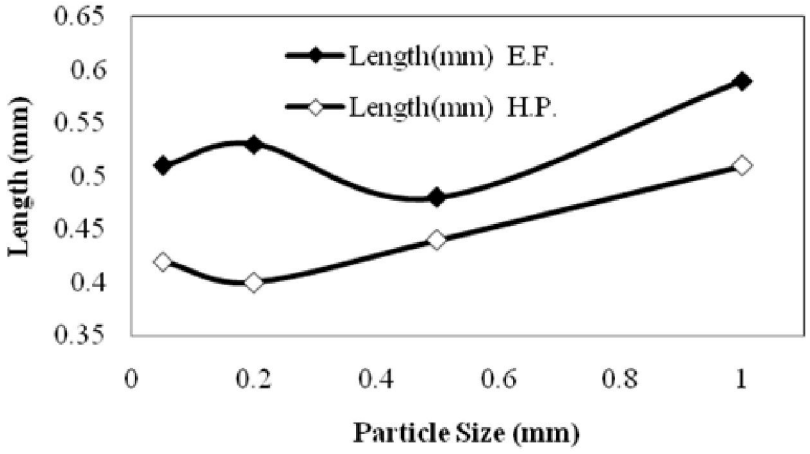


Fig. 1. Graphical representation for economic formulation and high performance

For the two data, the graph has been noticed that when the fineness changes the length travelled by tile is changed. And for the middle product was good relatively as compared to the other. The other thing is high performance have shown better result than the economic formulation.

3.2 COMPRESSION TEST RESULT

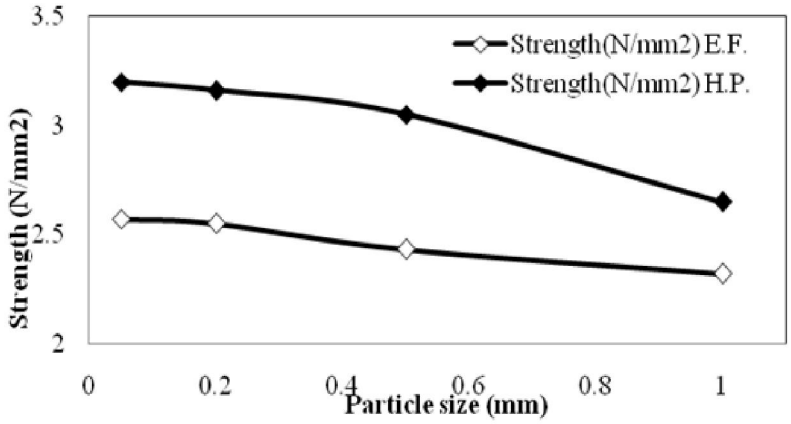


Fig. 2. Graphical representation for economic formulation and high performance

The graphs above shows that fine particle have high compressive strength. And the other thing is, in percentage high performance is stronger than economical formulation by about 17.27%.

3.3 BENDING TEST RESULT

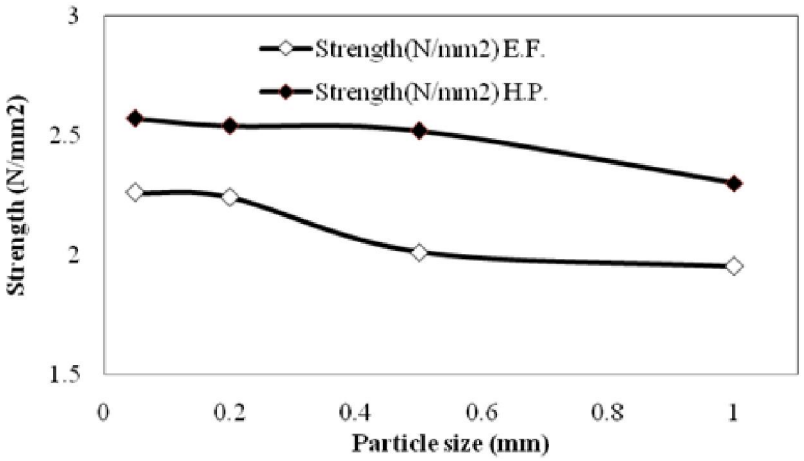


Fig. 3. Graphical representation for economic formulation and high performance

In the test of tile adhesive bending strength we have get the data shown in the graph above. From the graph we noticed that finer particle has good bending strength. And in percentage high performance is stronger than economical formulation by about 16.89%.

4 CONCLUSION

Tile adhesive have been produced with two different formulations. These two formulations are economical and high performance. Economical formulation is a product produced from components like cement, sand, cellulose ether and tartaric acid. While high performance is produced from cellulose fiber, limestone in addition to the components mentioned above. The work done in this thesis is production of tile adhesive as per the standard which is comfortable for our environment. Additives were added to enhance and inhibit the rate of drying and hydration. For acceleration chemicals like Calcium chloride, calcium formate and lithium carbonate were added for wet climate region. For retard and slow down of the rate of hydration chemicals like citric acid and tartaric acid are good for hot climate. Cause of these it was possible to produce tile adhesive suitable for our environment.

Tile adhesive which prepared from different modifiers for high performance and without modifier for economical purpose has difference in terms of strength. In terms of compression strength test about 17.27% high performance is stronger. And in addition high performance is stronger than economical formulation by about 16.89% in terms of bending strength. The two products have difference in terms of composition and content. Because the modified product has cellulose fiber, and limestone it has a characteristic like good workability, long open time, high anti slip resistance, high compression and bending strength. The economical formulation has shown slip resistant which greater than the standard as was indicated in the result except medium size product. For the case of high performance the two products, (fine and medium) have shown less than maximum requirement. Because of these, fine size of economical formulation is an acceptable product. And fine and medium size of high performance formulation have shown good result. The other thing is the bending and compressive strength test results of economical formulation were unable to satisfy the minimum requirement. But high performance formulation, (fine and medium size) have satisfied minimum requirement but that is not sufficient strength. This might have resulted because of raw material quality and absence of redispersible powder which enhance the strength of the product.

ACKNOWLEDGEMENT

Authors gratefully acknowledge Dashen Cement Company for providing the lab facility and guidance.

REFERENCES

[1] Arthur H, Sina E, *Adhesives technology handbook*, Second Edition, Norwich, NY: William Andrew Inc., 2009.
[2] J.K. Felixberger, "Polymer modified thin bed tile adhesive", *Institut De Promocio Ceramica*, Dec 8, 2008.
[3] R. Bayer, H. Lutz, *Encyclopedia of Industrial Chemistry*, Sixth Edition, Wiley-VCH., 2003.
[4] M.P. Edward, *Handbook of adhesive and sealant*, 1st edition, McGraw-Hill Professional, 1999.
[5] G. C. Mays, A.R. Hutchinson, *Adhesive in civil engineering*, Cambridge, 1 edition reprinted, 2005.
[6] G.A. Rowland, "Adhesives and Adhesion", CHEM NZ, no. 71, pp. 17-27, 1998.
[7] W. Brockmann, *Adhesive Bonding Materials Applications and Technology*, 1 edition, Wiley-VCH, 2009.
[8] W. Frank, K. Josef, H. Erwin, H. Sandy, W. Alexander, Z. Roger, "Construction and Building Material: Moisture induced length change tile adhesive mortars and their impact on adhesion", Science direct, pp. 426 – 438, 2012.
[9] A. Wetzel, M. Herwegh, R. Zurbriggen, F. Winnefeld, "Cement and Concrete Research: Influence of shrinkage and water transport on mechanisms on microstructure and crack formation of tile adhesive", Elsevier, pp. 39 – 50, 2012.
[10] CS4 Methocel for cement based, 1999.

International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 3 No. 3 July 2013, pp. 871-878

http://www.issr-journals.org/ijias/

Design and Simulation of STATCOM to Improve Power Quality

Md. Nazrul Islam[1], Md. Arifur Kabir[1], and Yashiro Kazushige[2]

[1]Research & Technical Service,
Bangladesh National Power Grid, Bangladesh

[2]JICA Experts, Japan

ABSTRACT: The performance of power systems decreases with the size, the loading and the complexity of the networks. This is related to problems with load flow, power oscillations and voltage quality. Such problems are even deepened by the changing situations resulting from deregulation of the electrical power markets, where contractual power flows do no more follow the initial design criteria of the existing network configuration. Additional problems can arise in case of large system interconnections, especially when the connecting AC links are weak. FACTS devices, however, provide the necessary features to avoid technical problems in the power systems and they increase the transmission efficiency. This paper presents a study on the design of a shunt connected FACTS device (STATCOM) and investigates the application of this device to control voltage dynamics and to damp out the oscillation in electric power system. STATCOM is one of the key shunt controllers in flexible alternating current transmission system (FACTS) to control the transmission line voltage and can be used to enhance the load ability of transmission line and extend the voltage stability margin. In this paper, the proposed shunt controller based on the voltage source converter topology as it is conventionally realized by VSC that can generate controllable current directly at its output terminal. The performance and behavior of this shunt controller is tested in 3-machine 9-bus system as well as the performance is compared in the test system with and without STATCOM at three cases in MATLAB/Simulink. Simulation results prove that the modeled shunt controller is capable to improve the Power quality significantly.

KEYWORDS: FACTS, STATCOM, VSC, CSC, PLL, PWM, PID.

1 INTRODUCTION

The traditional steady state stability studies and Transient stability studies take into account the active power flow P and power angle ð and generally assume constant receiving and sending end bus voltage. The reactive power flow Q and voltage fall during heavy current flow is neglected. This approach could not explain the several black-outs in USA, Europe, Japan etc. during the last quarter of the twentieth century. The blackouts were due to voltage collapse. During voltage collapse, the bus voltage starts falling and as a result power transfer P through the transmission line starts reducing resulting in ultimate voltage collapse and loss of system stability of entire network. That's why voltage stability studies have received more attention and have acquired a vital place in power system Studies. Voltage collapse phenomena take place where reactive power management is inadequate.

The application of power electronics in the electric power transmission plays an important role to make the system more reliable, controllable and efficient [1]. Due to deregulation, environmental legislations and cost of construction, it is becoming increasingly difficult to build new transmission lines. Thus it is essential to fully utilize the capacities of the existing transmission system. The flexible AC Transmission system (FACTS) has become a popular solution to our large/over extended power transmission & distribution system. FACTS devices are proving to be very effective in using the full transmission capacity while increasing power system stability, transmission efficiency and maintaining power quality and reliability of Power system. These devices are mainly based on either voltage source converter (VSC) or Current Source Converter (CSC) and have fast response time. As an important member of FACTS devices family, STATCOM has been at the centre of attention

Corresponding Author: Md. Nazrul Islam (nazrul.pgcb@yahoo.com)

and the subject of active research for many years. STATCOM is a shunt connected device that is used to provide reactive power compensation to a transmission line. This controller can either absorb or inject reactive power whose capacitive or inductive current can be controlled independent of the AC line voltage. Thus, STATCOM can enhance the transmission line load ability by extending the MW margin and improves the oscillation of voltage transients through efficient regulation of the transmission line voltage at the point of connection [1]-[3].

This paper deals with the modeling of a SPWM based STATCOM with a PID controller implemented on a 3- machine 9-bus test system. The device is connected to a load bus with a converter transformer. The modeling of shunt controller and testing is simulated in the MATLAB/Simulink environment. The controller is represented as block diagram that presents practical electronic model of shunt controller. PID controller is used to control the current injection at the connection point by varying the desired parameters, one is Modulation index (AM) and another is power angle (∂). Mainly there are four loop tuning methods for a PID controller; those are manual tuning, Ziegler-Nichols, software Tools and Cohencoon method. Firstly, Ziegler-Nichols method is chosen for loop tuning and then manual tuning is applied to the PID controller by trial and error method to take its performance at optimum level. In fact, there are four different control strategies for a STATCOM controller, direct control, decoupling control, cross control and matrix control. The direct control method is used here to control the output of shunt connected FACTS device.

2 BASIC CONFIGURATION AND PRINCIPLE OF OPERATION

Basically, shunt connected FACTS device can be realized by either a VSC or a CSC [4]. But the VSC topology is preferred because CSC topology is more complex than VSC in both power and control circuits. In CSC such as GTO (Gate Turn Off Thyristor) is used, a diode has to be placed in series with each of the switches. This almost doubles the conduction losses compared with the case of VSC. The DC link energy storage element in CSC topology is inductor where as that in VSC topology is a capacitor. Thus, the efficiency of a CSC is expected to be lower than that of a VSC [6]-[9]. The modeled STATCOM using VSC topology is being used in the test system to supply reactive power to increase the transmittable power and to make it more compatible with the prevailing load demand. Thus, the shunt connected FACTS device should be able to minimize the line over voltage under light load condition and maintain voltage levels under heavy load condition. Two VSC technologies can be used for the VSC. One of them, VSC is constructed with IGBT/GTO-based SPWM inverters. This type of inverter uses sinusoidal Pulse-Width Modulation (SPWM) technique to synthesize a sinusoidal waveform from a DC voltage source with a typical chopping frequency of a few kilohertz. Harmonic voltages are cancelled by connecting filters at the AC side of the VSC. This type of VSC uses a DC link voltage Vdc. Output voltage is varied by changing the modulation index of the SPWM modulator. Thus modulation index has to be varied for controlling the reactive power injection to the transmission line. In another type VSC is constructed with GTO-based square-wave inverters and special interconnection transformers. Typically four three-level inverters are used to build a 48-step voltage waveform. Special interconnection transformers are used to neutralize harmonics contained in the square waves generated by individual inverters. In this type of VSC, the fundamental component of output voltage is proportional to the voltage Vdc. Therefore Vdc has to be varied for controlling the reactive power.

The shunt controller is like a current source, which draws from or injects current into the system at the point of connection. The shunt controller may be variable impedance, variable source or a combination of these [10]. Variable shunt impedance connected to the line voltage causes a variable current flow and hence represents injection of current into the line. As long as the injected current is in phase quadrature with the line voltage, the shunt controller only supplies or consumes reactive power. When system voltage is low, the STATCOM generates reactive power (STATCOM capacitive). When system voltage is high, it absorbs reactive power (STATCOM inductive). The variation of reactive power is performed by means of a VSC connected on the secondary side of a coupling transformer. The VSC uses forced-commutated power electronic devices (GTOs, IGBTs or IGCTs) to synthesize a voltage V2 from a DC voltage source. Any other phase relationship will involve handling of real power as well [11]. So, the shunt controller is therefore a good way to control the voltage at and around the point of connection through injection of reactive current (leading or lagging) alone or a combination of active and reactive current for a more effective voltage control and damping of voltage dynamics [12]. The real power (P) and reactive power (Q) are given by:

$$P = \frac{E.V}{X} Sin\partial \quad (1)$$

$$Q = \frac{E^2}{X} - \frac{E.V}{X} Cos\partial \quad (2)$$

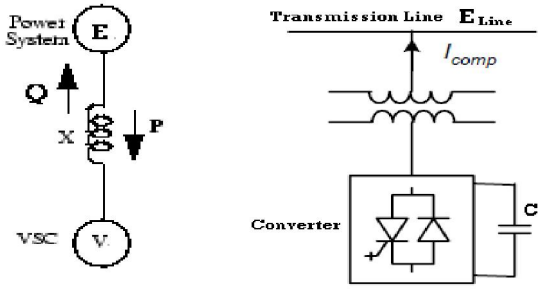


Fig. 1. Static Synchronous Compensator

E is the line voltage of transmission line. V is the generated voltage of VSC. X is the equivalent reactance of interconnection transformer and filters and δ is the phase angle of E with respect to V. In steady state operation, the voltage V generated by the VSC is in phase with E (δ=0), so that only reactive power is flowing (P=0). If V is lower than E, Q is flowing from E to V (STATCOM is absorbing reactive power). On the reverse, if V is higher than E, Q is flowing from V to E (STATCOM is generating reactive power). Since we are using here a VSC based on SPWM inverters hence modulation index is varied for controlling the reactive power injection to the transmission line. A capacitor is connected on the DC side of the VSC acts as a DC voltage source. In steady state the voltage V has to be phase shifted slightly behind E in order to compensate for transformer and VSC losses and to keep the capacitor charged.

3 MODELING OF SHUNT CONVERTER

Figure shows a construction of STATCOM utilizing one interconnection transformer and three GTO/Diode based double arm H-bridges. Each H-bridge is connected to the each phase of the secondary of interconnection transformer. Transformer primary is connected to the transmission line. Conventionally, shunt controllers are constructed of three phase converters or inverters. But it is possible to replace the three phase converter with three single phase converter. The three phase converter constructed with three single phase converter produces less switching ripples than the conventional three phase converter [16]. So, three phase converter constructed with three single phase converters is used. T1, T2 and T3 represent the transformer coils of phase A, B and C respectively that form a three phase transformer secondary connected to shunt converter. A capacitor (C) which acts as a voltage source is used. The original circuit diagram of each GTO/Diode bridge (B1,B2 and B3) is shown in figure . Each H-bridge consists of four GTO and four diodes where the GTO and diode are connected in anti-parallel. So, four different control pulses are required to control each of bridges. Therefore to apply firing pulses to three different bridges properly, total twelve different pulses are required for controlling the shunt converter.

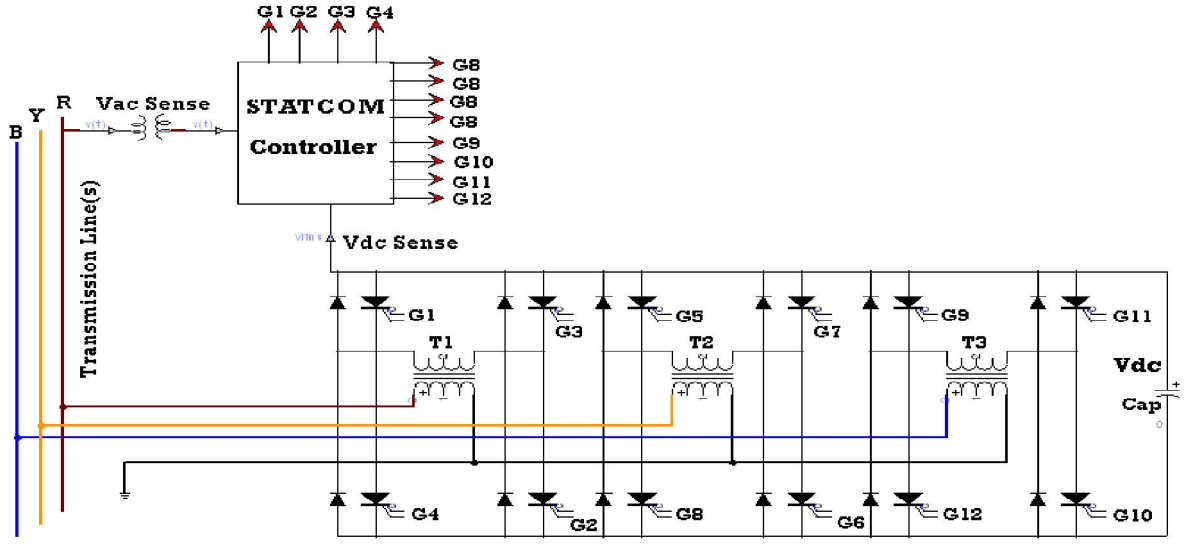


Fig. 2. Modeled Shunt Connected FACTS Device

4 CONTROL STRATEGY

STATCOM can be controlled in voltage control mode and Var control mode. The control used in this simulation is AC voltage control mode. Mainly, the control is divided into two parts. One is for angle order and another is for the order of modulation index. The shunt converter is operated in such a way as to demand this DC terminal power from the line keeping the voltage across the storage capacitor Vdc constant. So, according to equation-1, the angle is ordered in such a way that the net real power absorbed from the line by this shunt FACTS device is equal to the losses of the converters and the transformer only. The remaining capacity of this shunt converter can be used to exchange reactive power with the line so to provide VAR compensation at the connection point. The reactive power according to equation-2 is electronically provided by the shunt converter and the active power is transmitted to the DC terminals. The shunt converter reactive current is automatically regulated to maintain the transmission line voltage at the point of connection to a reference value.

The line voltage and Dc link voltage across capacitor are measured to calculate the amount of reactive power to regulate the line voltage and consequently the modulation index is varied in such a way as to calculate reactive power can be injected at the point of connection and thus the shunt FACTS device acts as a voltage regulator. The SPWM firing pulses to the GTOs are obtained by comparing the PWM carrier signal and the reference sine wave. The amplitude of reference sine wave is 1 Volt and frequency is 50 Hz which is similar to system operating frequency. The carrier frequency is set at 1.5 KHz which is 30 times the system operating frequency. The phase lock loop (PLL) plays an important role in synchronizing the switching to the system voltage and lock to the phase at fundamental frequency. The converter is consisted of 12 GTO with additional components. The controller controls the firing pulses from G1 to G12 which are sinusoidal pulse width modulated signals. The following figure shows the block diagram of control strategy to generate only one pulse width modulated signal and 11 signals can be generated similarly.

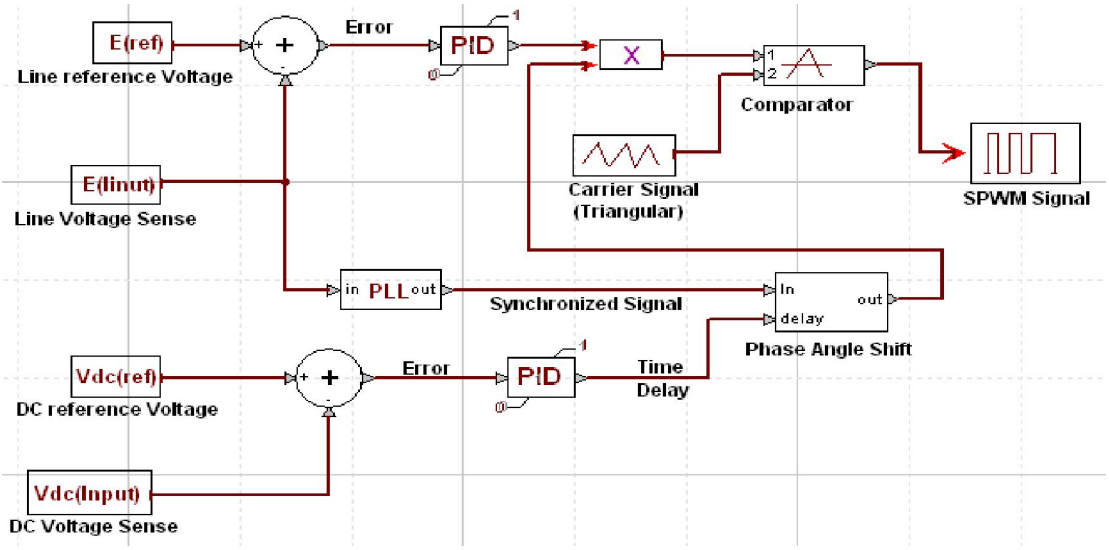


Fig. 3. Block Diagram of STATCOM Controller

5 SIMULATION SETUP AND RESULTS

Figure 4 shows the 3 machine 9 bus test system for simulation. The test system includes machines, transmission lines and loads at different buses. The modelled FACTS device (STATCOM, 300 MVA) is installed at Bus-9. Two types of large loads (Load-1 & Load-2) are also connected at bus-9. Power is flowing to Bus-9 by TL1 and TL2.

Case 1: At starting, Load-1(50 MW & 30 MVAR) and Load-2 (100 MW & 50 MVAR) is not connected to the Bus-9. A fault takes place at t=0.3 second at TL1 and instantaneously Bk-1 and BK -2 is tipped to isolate the transmission line TL1 from the system. At the same time (t=0.3s) Load-1 is connected to the system and Load-2 is connected at t=0.7 second. The start time of simulation at t=0 s and the end time is t=1 second. Waveforms Scales are zoomed so that the voltage oscillations can be seen clearly.

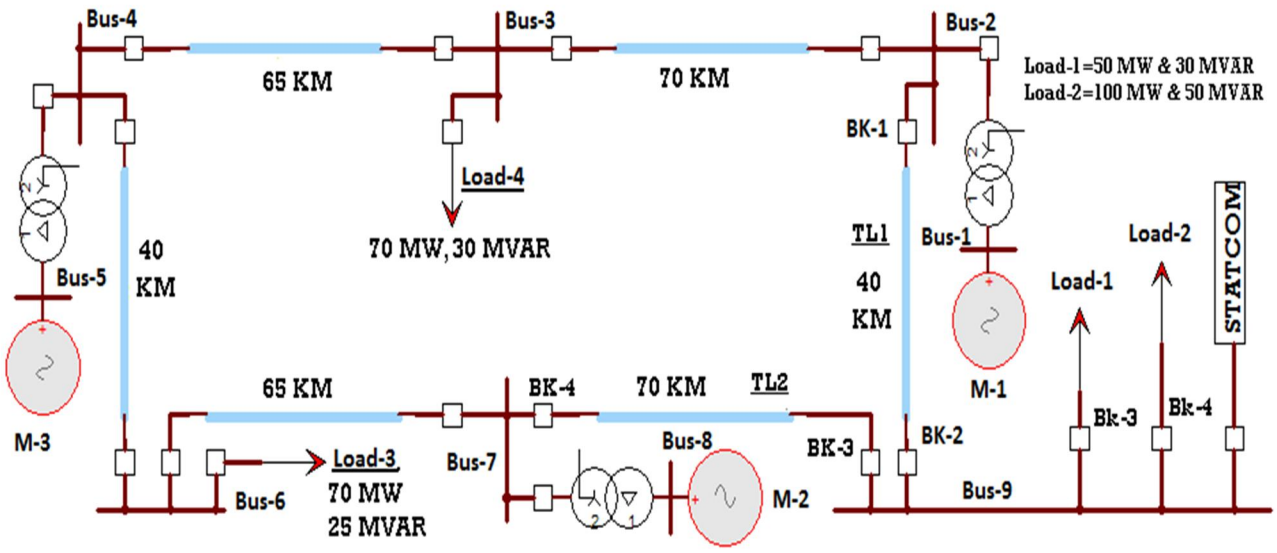


Fig. 4. STATCOM connected to the 3 machine 9 bus test System

Figure 5 shows the voltages of Bus-9 in per unit (p.u) when STATCOM is not in use. The solid line represents the voltage in p.u when modelled STATCOM is connected to the bus. As it is installed at Bus-9 the voltage is controlled at that bus successfully. Figure 6 shows the injected reactive power to the Bus-9 to control the voltage and real power drawn by the STATCOM to keep the DC link voltage (Vdc) across the capacitor constant because the DC voltage tends to change during operation. In steady state the voltage V has to be phase shifted slightly behind E in order to compensate for transformer and VSC losses and to keep the capacitor charged. The DC link voltage (Vdc) across the capacitor is shown in Figure 7. The modulation index is required to change due to generate or absorb the required amount of reactive power. To maintain the transmission line voltage at Bus-9 the modulation index is controlled by PID controller which is shown in Figure 7.

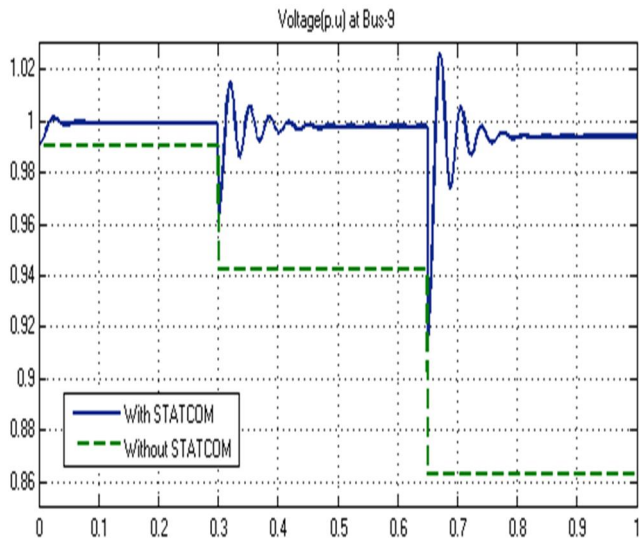


Fig. 5. Transmission line voltage at bus-9

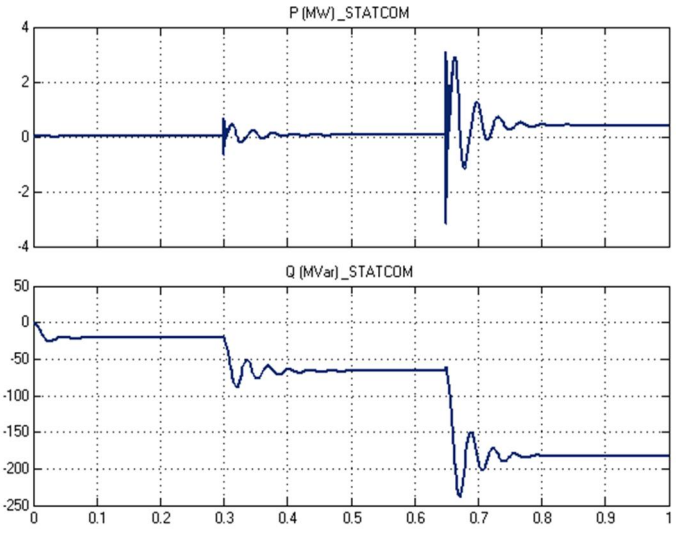


Fig. 6. STATCOM Real Power (Drawn) & Reactive Power(Supplied)

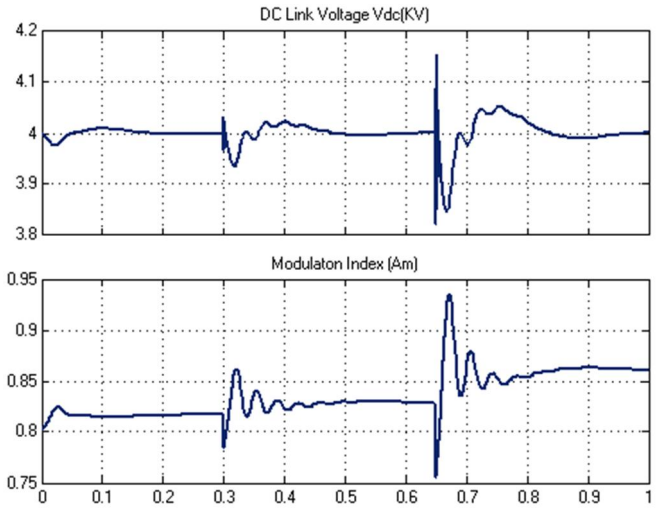


Fig. 7. DC Link Voltage & Control of Modulation Index

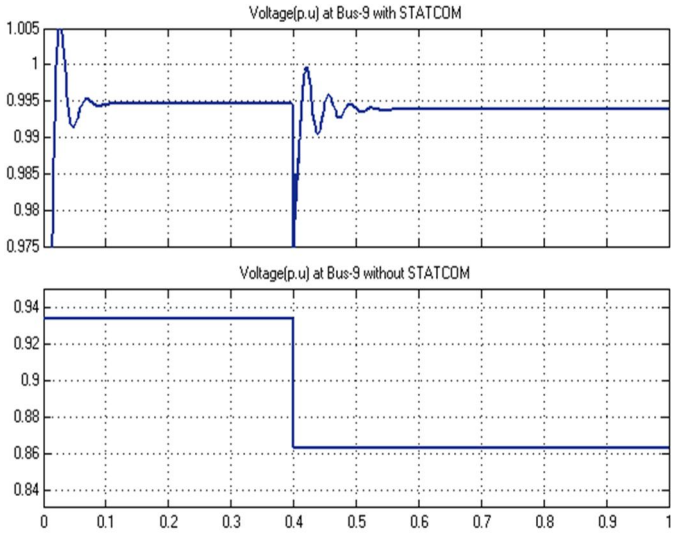


Fig. 8. Transmission line voltage at bus-9

Case 2: At starting (t=0.0 s), Load-1 and Load-2 is connected to the Bus-9 and drawing huge power. But for any reason (either manual or protection) transmission line TL1 is disconnected from the system at t=0.4 second. Simulation starts at t=0 second and ends at t=1 second. Waveforms Scales are zoomed so that the voltage oscillations can be seen clearly. Figure 8 shows the voltages of Bus-9 in per unit (p.u) when STATCOM is not in use. The solid line represents the voltage in p.u when modeled STATCOM is connected to the bus. As it is installed at Bus-9 the voltage is controlled at that bus successfully. Figure 9 shows the injected reactive power to the Bus-9 to control the voltage and real power drawn by the STATCOM to keep the DC link voltage (Vdc) across the capacitor constant because the DC voltage tends to change during operation. In steady state the voltage V has to be phase shifted slightly behind E in order to compensate for transformer and VSC losses and to keep the capacitor charged. The DC link voltage (Vdc) across the capacitor is shown in Figure 10. The modulation index is required to change due to generate or absorb the required amount of reactive power. To maintain the transmission line voltage at Bus-9 the modulation index is controlled by PID controller which is shown in Figure 10.

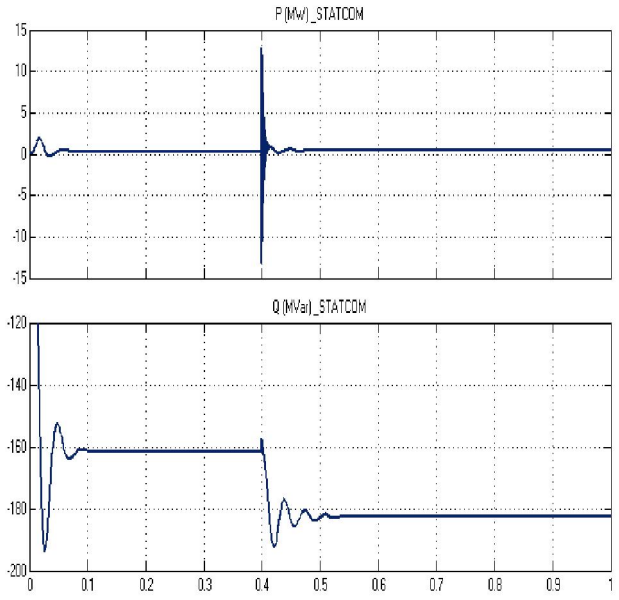


Fig. 9. STATCOM Real Power (Drawn) & Reactive Power (Supplied)

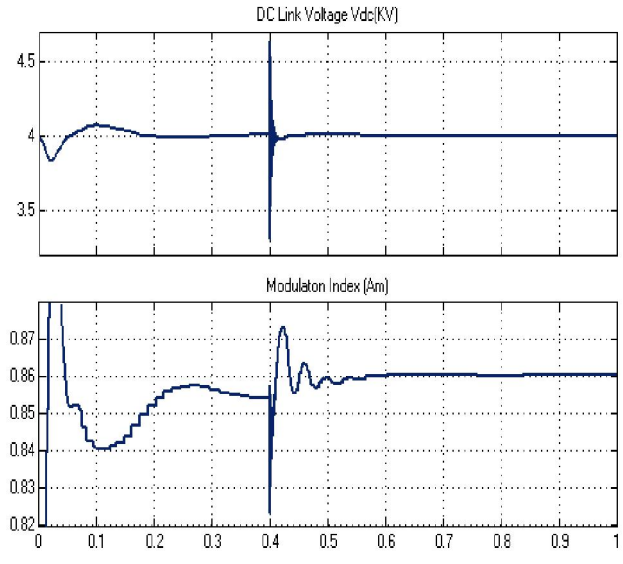


Fig. 10. DC Link Voltage & Control of Modulation Index

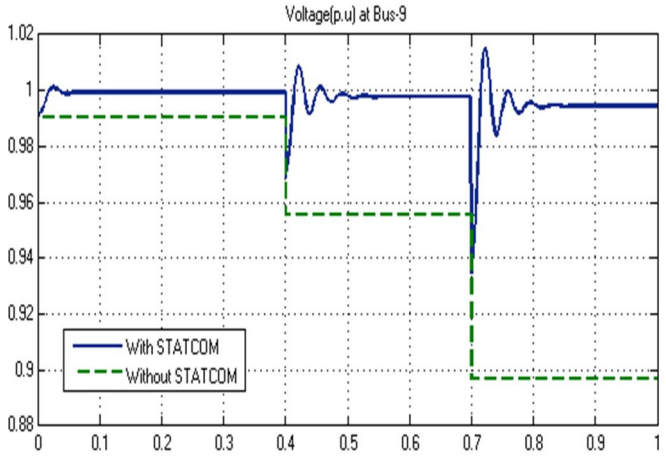


Fig. 11. Transmission line voltage at bus-9

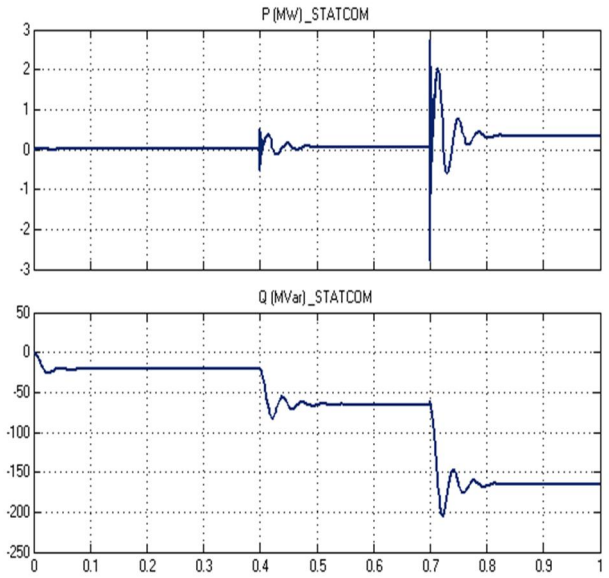


Fig. 12. STATCOM Real Power (Drawn) & Reactive Power (Supplied)

Case 3: At starting, Load-1 and Load-2 is not connected to the Bus-9. A fault takes place at t=0.4 second at TL1 and instantaneously Bk-3 and BK -4 is tipped to isolate the transmission line TL2 from the system. At the same time (t=0.4s) Load-1 is connected to the system and Load-2 is connected at t=0.7 second. Simulation starts at t=0 second and ends at t=1 second. Waveforms Scales are zoomed so that the voltage oscillations can be seen clearly.Figure 11 shows the voltages of Bus-9 in per unit (p.u) when STATCOM is not in use. The solid line represents the voltage in p.u when modelled STATCOM is connected to the bus. As it is installed at Bus-9 the voltage is controlled at that bus successfully. Figure 12 shows the injected reactive power to the Bus-9 to control the voltage and real power drawn by the STATCOM to keep the DC link voltage (Vdc) across the capacitor constant because the DC voltage tends to change during operation. In steady state the voltage V has to be phase shifted slightly behind E in order to compensate for transformer and VSC losses and to keep the capacitor charged. The DC link voltage (Vdc) across the capacitor is shown in Figure 13. The modulation index is required to change due to generate or absorb the required amount of reactive power. To maintain the transmission line voltage at Bus-9 the modulation index is controlled by PID controller which is shown in Figure 13.

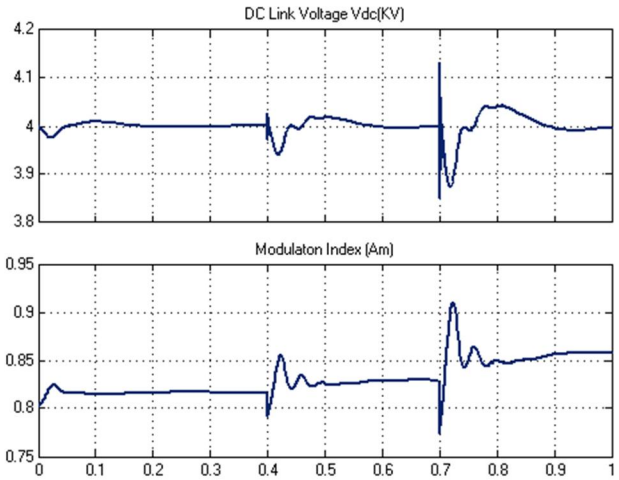


Fig. 13. DC Link Voltage & Control of Modulation Index

6 CONCLUSION

In simulation, worst events are considered to examine the performance of modeled shunt connected FACTS device. The simulation results show that the modeled STATCOM is capable enough to control the transmission line voltage dynamics as well as the same shunt controller can be used in VAR control mode. Vdc is regulated by controlling proper phase shift and transmission voltage is regulated by varying the modulation index. Two single input and single output (SISO) closed loop systems are used. The response of controller is very fast due to apply direct control method. The simulation results also prove that the shunt device with proposed switching scheme functions successfully as the real time voltage controller and it improves the dynamic stability with a wide range of control the reactive power. The magnitude of voltage oscillation in simulation and other figures are zoomed along y axis to observe the oscillation clearly but actually the oscillations are very low. Three single phase converters are used rather than three phase converter to reduce switching ripples. A part of Dhaka ring of Bangladesh National Grid is modeled as transmission line in this paper.

REFERENCES

[1] Ye, Y., Kazerani, M., and Quintana, V.H., 2005. Current-source converter. Based STATCOM : Modeling and control. IEEE Transactions on Power Delivery 20(2): 795-800.

[2] Jower F.A., 2007, Improvement of synchronizing power and damping power by means of SSSC and STATCOM: A comparative study. Electric Power Systems Research 77(8): 1112-1117.

[3] Puleston, P.F., Gonzalez, S.A. and Valenciaga, F., 2007. A STATCOM based variable structure control for power system oscillations damping. International Journal of Electrical Power & Energy Systems 29(3): 241-250.

[4] Hingorani, N.G. and L. Gyugyi, 2000. Understanding FACTS: Concepts and Technology of Flexible AC Transmission Systems. New York: IEEE Press.

[5] Moran, P.D., Ziogas, L.T., Joos, G. and Hingorani, N.G., 1989. Analysis and design of a three-phase current source solid-state var compensator. IEEE Transactions on Industrial Application 25(2): 356-365.

[6] Gyugyi, L., 1994 Dynamic compensation of AC transmission lines by solid-state synchronous voltage sources. IEEE Transactions on power Delivery 9(2): 904-911.

[7] Gyugyi, L., Schauder, C.D., Williams, S.L., Reitman, T.R., Torgerson, D.R., and Edris, A., 1995. The unified power flow controller: A new approach to power transmission control. IEEE transactions on power delivery 10(2): 1085-1097.

[8] Lehn, P.W. and M.R. Iravani, 1998. Experimental evaluation of STATCOM closed loop dynamics. IEEE Transactions on Power Delivery 13(4): 1378-1384.

[9] Schauder, C.D. and H. Mehta, 1993. Vector analysis and control of advanced static VAR compensators. IEEE Proceedings 140(4): 299-306.

[10] Sen, K.K., 199. STATCOM-STA Ticsynchronous COM Pensator: Theory, modeling, and O. Farrok. M.G. Rabbani and M.R. Islam/International Energy Journal 11(2010) 43-50.

[11] Applications, In Proceedings IEEE Power Engineering Society Winter Meeting. 1177-1183. Yankui, Z., Yan, Z., Bei, W, and Jian, Z., 2006. Power injection model of STATCOM with control and operating limit for power flow and voltage stability analysis. Electric Power Systems Research 76(12): 1003-1010.

[12] El-Moursi, M.S. and A.M. Sharaf, 2006. Novel reactive power controllers for the STATCOM and SSSC. Electric Power Systems Research 76(4): 228-241.

[13] Tavakoli Bina, M., Eskandari, M.D. and Panahlou, M., 2005. Design and installation of a # 250 kV Ar D-STATCOM for a distribution substation. Electric Power Systems Research 73(3): 383-391.

[14] Iyer, S., Ghosh, A. and Joshi, A., 2005. Inverter topologies for DSTATCOM applications-a simulation study. Electric Power Systems Research 75(3-2): 161-170.

[15] Ricardo D.M., Juan M.R. and Ruben T.O., 2005. Three-phase multi-pulse converter Stat Com analysis. International Journal of Electrical Power & Energy Systems 27(1): 39-51.

[16] Fujita, H., Akagi, H. and Watanabe, Y., 2006. Dynamic control and performance of a unified power flow controller for stabilizing an AC transmission system. IEEE Transactions on Power Electronics 21(4): 1013-1020.