

**THE INFLUENCE OF URBAN RESIDENTIAL
SETTLEMENTS ON THE MANAGEMENT OF SELECTED
DRY LAND NATURAL RESOURCES; A CASE OF NGONG
DIVISION, KAJIADO DISTRICT**

By

MAKOKHA, WILLIAM EUGENE

**WORK SUBMITTED IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR MASTERS OF ARTS IN
ENVIRONMENTAL PLANNING AND MANAGEMENT.
FACULTY OF ARTS, DEPARTMENT OF GEOGRAPHY;
UNIVERSITY OF NAIROBI**

2006

**UNIVERSITY OF NAIROBI
EAST AFRICANA COLLECTION**

UNIVERSITY OF NAIROBI LIBRARY



0271466 5

DECLARATION

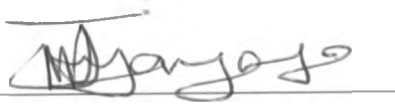
I certify that this is my original work and has not been submitted in any other university for examination.

Signed 

Makokha William Eugene

Date 23/10/2005

This research work has been submitted for examination with my approval as University Supervisor

Signed 

Mr. Nyangaga M. John

Date 23/10/2006

DEDICATION

To my mother Inviolata Natecho and my father Evans Makokha who saw the value of education, stood by me and committed each and every penny to see me through school

ABSTRACT

Urban growth and associated urban functions such as residential settlements are a continuous challenge onto the environment. Residential areas are home to urban areas most important resource; human resource, and are a reserve for urban areas labour, often they act as interlinks and offloads points for human labour arriving from rural areas. They provide markets for ready goods and services within urban areas, and as such, not only consume but drive the urban areas productive processes.

Urban areas demand for food, water, energy (fossil and human), and raw material for production processes, as well as management of waste products from its vital processes. Having this knowledge, this study set out to examine natural resource use; soil, water and vegetation (Products and service) for the urban area occurring on the margins of the ASALs given limited availability of resources in these lands. Particularly attention was paid small urban areas that have grown as an outward influence of growth of the city of Nairobi. The role of increased residential settlements vis a vis natural resource use.

A survey was conducted and data collected on residential units using structured questionnaires and interview schedules. Data was collected in two distinct land use areas for comparison purposes. In addition, short discussion sessions were held with actors and resource specialists (government staff) as key informants on the ground. Photography and observation were used to support arguments and document findings. Data was analysed using basic descriptive statistics and cross tabulations. Hypothesis testing using chi square was used to arrive at conclusions of hypothesized relationships between variables. Qualitative data was based on information from key informants was reported indirectly and verbatim.

Data is presented in tables, charts, photography and actual description of data, Research findings were made on several fronts. Hypothesis testing for land use practices indicated there is no significant difference in land use practices between the different designations of land use areas in the study area. Generally, there is increased settlement in the study area as evident in years of residence, occupation, and mode of tenure. On resources, there is increasing shortage of water in the study area; main source of water is borehole. Water

is a serious occasioned by mechanical breakdown and power failures at borehole installations and sometimes dry seasons. Competition for water has led to increased depth in sinking of boreholes to reach the receding water table.

There is also increased vegetation clearance as a result of residential settlement, while incentive and initiative to green residential areas remain low. Soils are left bare, concretized or covered with stone chips to avoid mud. The study argues, there is increased generation of sediments, surface run-off, and storm waters from residential areas as a result of construction and, soil compaction. Soils removed during construction and carried out of construction sites, are at best used to fill old quarries or at worst, dumped on roadsides, river valleys and open fields destroying the scenic value of the area.

In conclusion, the study argues, urban settlements on the margins of ASALs maybe detrimental to the environment especially where they clear vegetation and interfere with catchments areas and water recharge points. These may not only produce physical and topographical change but may also alter climate of these regions in the long-run.

The study advocates for use of law in environmental management i.e. use of planning instruments such as Environmental Impact Assessments (EIAs), Environmental Audits (EA), Policy for settlement in arid areas, Community Education, advocacy and sensitization in settled areas among residential groups. Community mobilization for resource management should be done through neighbourhood associations, conservation groups, establishment of (Pilot) ecological settlements and policing of settlements to meet certain stipulated requirements.

ACKNOWLEDGEMENT

The preparation and eventual write-up of this work would not have been realized without consultation as well as devotion of certain personalities. To these people I am indebted and humbly oblige to express my gratitude

To my parents, Evans Makokha and Inviolata Natecho, it has been a long journey that they sincerely hoped to see the fruits, for their continuous guidance, encouragement and committing of resources; I am indeed grateful.

To my supervisor Mr. John M. Nyangaga the mentor of my work, he concretized my ideas to merit putting onto paper. For the time, patience and devotion as well as direction to the preparation of the final document; I give my sincere appreciation

To the Ministry of Education science and technology, for granting me the permit to conduct research, Thanks to Mr. B.O. Adewa for facilitating quick processing of my permit

To the provincial administration; Office of the president Mr. D.N Ondabu (OP), Mr. F.K Ndambuki of the District Commissioner's office Kajiado, Mr. Simon G. Gatobu the District officer Ngong Division, respective chiefs of Nkai-Murunya and Ongata Rongai locations, their concerns on my security as well as facilitation of movement I am grateful

To the government staff based at Ngong division; Mr. Joseph Kamau and Ms Eunice Parsimei (Divisional Soil Conservation Officers), Mr. Martin Mburu (Divisional Crop Officer), Ms. Jane Agova (Divisional Agricultural Extension Officer), all of the Ministry of Agriculture. Mr. Thomas Chege (Divisional Livestock Officer) of the Ministry of Livestock and Fisheries, Mr. Mark Miano (Divisional Forestry Officer) of the Environment and Natural Resources and lastly to Mr. Muasya Simon (Deputy Divisional Water Officer) who facilitated access to government records, gave their expert on the ground view. And added a new dimension to resource consideration in the study area for my work; I give my gratitude.

To the Department of Geography; the Chairman and members of staff for having given me an opportunity to pursue this course, special thanks to my teachers who have mentored me through class work to this final stage.

To my classmates and friends: Mr. Kamithi Ng'ang'a, Mr. Tonui Michael, Mrs Librata Muhati, for continuous support and motivation. Special thanks to Mr Omoke Kennedy and Tonui Michael for allowing me use their Computers and facilities during the editing of my work. Again to Mr. Kirema Joseph of the Department of Geography for preparing maps for the work.

Lastly, to my aunts Mrs. Victoria Wataka and Gertrude Mulongo, Cousin Leonora Wataka and my sister Annelyne Makokha for providing stationary and material support and lastly, Sister Jane-Francis (IBVM), for financial support towards my fees, material as well as moral support for the completion of this work; receive my sincere gratitude's

I however absolve all from blame and acknowledge that all errors of commission and omission are mine.

OPERATIONAL DEFINITIONS

Urbanization: Refers to a process reminiscent to urban growth and expansion. In essence urbanization in most cases takes the form of immigration and continuous influx of population into urban areas.

ASALs: Refer to Arid and Semi Arid lands, these are lands or regions recognized by their distinct climate and ecology; these areas are primarily hot, with limited rainfall and less abundant in flora and fauna.

Tenure: refers to the mode of ownership for resources.

Compaction: refers to all possible ways of removing vegetation cover, making the ground firm and sealing the ground surface thereby impairing infiltration of water into lower ground surfaces.

Soil Erosion: Refers to the removal and carrying away of top soil by human, animals, water, and wind.

Sediments: Refers to all materials not miscible in water and less than 2 cm that cannot be transported by water for long distances, and which quickly settle down in any water body.

Soil Resource: Refers to the loose superficial layer found on the top most part of the earth's crust. It ranges from a few centimeters on the ground surface to several metres into the earth's crust.

Vegetative resource: Refers to all vegetation in their diverse forms, for the purpose of the study area, it refers to grass, woody shrubs, thorny thickets and exotic plant species.

Water Resource: Refers to water for drinking, domestic, agriculture and livestock uses found in the study area. It takes the form of both surface and sub-surface waters.

Commercial residential units: This refers to residential units found on the land use area designated as commercial setting.

Agricultural land Residential units: Refers to the residential units found on the land use area designated as agricultural land.

Ecological Settlement: refers to environmental friendly settlements that factor important ecological factors such as environmental conservation, that provide for green areas, necessary for environmental functions such as water soaking, wind breaks and water catchments.

Natural Resource: these refer to resources deriving from the earth and biosphere and exist independent of human use e.g. land, soil, vegetation, air, water minerals etc.

Resource: Is anything that has value or utility. Resource(s) exist regardless of whether humans choose to use or not

Household: People who live in the same house and share a common eating and cooking place.

LIST OF ACRONYMS

- ALRMP** - Arid Lands Resource Management Programme
- APHRC** – African Population and Health Research Centre
- ASALS** – Arid and Semi Arid Lands
- CBS** – Central Bureau of Statistics
- CDA** – Coast Development Authority
- DDP** - District Development Plan
- DFRD** - District Focus for Rural Development
- EMCA** – Environmental Management and Coordination Act
- GDP** – Gross Domestic Product
- GEO** - Global Environmental Outlook
- GIS** – Geographical Information Systems
- GOK** – Government of Kenya
- ICRISAT** – International Crop Research Institute for the Semi-Arid Tropics
- IMCE** – Inter Ministerial Commission on the Environment
- KVDA** – Kerio Valley Development Authority
- LBDA** - Lake Basin Development Authority
- NDP** – National Development Plan
- NEAP** – National Environmental Action Plan
- NEMA** – National Environmental Management Authority
- NES** – National Environmental Secretariat
- OECD** – Organisation for Economic Cooperation and Development
- OLCODO** – Oloolaiser Community Development Organisation
- PPCSCA** – Permanent Presidential Commission on Soil Conservation and Afforestation
- SNV** – Netherlands Development Organisation
- TARDA** - Tana and Athi River Development Authority
- UNCHS (HABITAT)** – United Nations Centre for Human Settlements
- UNEP/ROA**- United Nations Environment Programme/ Regional Office for Africa
- UNESCO** – United Nations Education Scientific and Cultural Organisation

TABLE OF CONTENT

DECLARATION..... I

DEDICATION II

ACKNOWLEDGEMENT V

OPERATIONAL DEFINITIONS VII

LIST OF ACRONYMS.....IX

TABLE OF CONTENT X

LIST OF FIGURES 3

LIST OF TABLES 4

CHAPTER ONE 6

1.0 INTRODUCTION 6

1.1 URBANISATION; GROWTH AND RESOURCE USE 6

1.2 THE DYNAMICS OF ASALS IN KENYA 8

1.4 OBJECTIVES 13

1.6 SCOPE AND LIMITATION 13

1.7 CONSTRAINTS AND LIMITATIONS 14

1.8 ASSUMPTIONS 15

2.0 STUDY AREA 25

2.1 DESCRIPTION OF STUDY SITE 25

2.2 RESOURCES: 32

CHAPTER THREE 36

3.0 METHODOLOGY 36

3.1 RESEARCH DESIGN..... 36

3.2 SAMPLING PROCEDURE..... 36

3.3 DATA COLLECTION 38

3.4 TYPE OF DATA COLLECTED 38

3.5 NATURE OF DATA COLLECTED..... 38

3.6 DATA PROCESSING ANALYSIS 39

3.7 DATA PRESENTATION 39

CHAPTER FOUR..... 40

4.0 RESULTS AND DISCUSSIONS 40

4.1 RESEARCH FINDINGS 40

4.2 DISCUSSIONS 66

4.3 RESOURCES 66

CHAPTER FIVE..... 80

5.0 SUMMARY OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS 80

5.1 RECOMMENDATIONS 81

5.2 BIBLIOGRAPHY 86

5.3 APPENDICES..... 89

LIST OF FIGURES

- Figure 1 - Map of Kenya showing location of study area
- Figure 2 – Map of Kajiado District showing administrative areas by divisions
- Figure 3 – Map of Kajiado District showing Eco- agro- climatic zones
- Figure 4 – Map of Kenya showing drainage and catchments areas
- Figure 5 – Annual rainfall based on six year average
- Figure 6 – Rainfall fluctuation over years (1997 – 2004)
- Figure 7 – Cross tabulation of Age of household head against family size
- Figure 8 – Cross tabulation of Age of household head against occupation
- Figure 9 – Summary of professional qualification by occupation
- Figure 10 – Cross tabulation of household size against water demand
- Figure 11 – Photograph of firewood transit to urban areas
- Figure 12 – Photograph showing open field and river valley dumping of soil
- Figure 13 – Roadside and site dumping of soil
- Figure 14 – Early stages of compaction of sediment generation
- Figure 15 – colonization of new land for urban settlement
- Figure 16 – Vegetation clearing and opening up of new lands
- Figure 17 – Rainfall fluctuation for the last 10 (ten) years

LIST OF TABLES

Table 2.1: Population Distribution by Area and Density

Table 2.2 Administrative Division, Location and Sub-locations

Table 2.3 Rainfall Distribution by Months

Table 4.1: Age of HH head

Table 4.2: Mode of Tenure

Table 4.3: Years of residence in the study area

Table 4.4: Mode of Land Tenure

Table 4.5: Cross tabulation of Age of HH head against Years of residence category

Table 4.9: Proximity of houses/structures

Table4.10: Cross tabulation of Proximity of structures Against Type of residential unit

Table 4.11: Crammer’s V

Table 4.12: Kind of soil cover present

Table 4.13: Estimate extent of Concrete cover

Table 4.14: Cross tabulation of Estimate extent of Concrete cover Against Type of residential unit

Table 4.15: Estimate extent of bare soil cover

Table 4.16: How soils removed during construction

Table 4.17: What happens to soil after removal?

Table 4.18: How rampant is the practice of soil removal from site

Table 4.19: Cross tabulation of what happens to soil to removal * Type of residential unit

Table 4.20: Main source of water

Table 4.21: Alternative sources of water

Table 4.22: Distance to the main Source of water

Table 4.23: Distance to alternative source of water

Table 4.24: Problems with major source (water)

Table 4.25: Problems with alternative source (water)

Table 4.26: Quantity of water in 20lit cans a household’s get in a day

Table 4.27: Extent of Vegetative cover in residential areas

Table 4.28: Cross tabulation extent of Vegetative cover) against Type of residential unit

Table 4.29: Compound layouts and features present

Table 4.30: State of vegetation cover in residential areas

Table 4.31: Kind of fencing material used

Table 4.32 Sources of fuel for residential units

Table 4.33: how does fuel use lead to vegetation loss

Table 4.34: Building material for Homes

Table 4.35: Sources of stones for construction

Table 4.36: Sources of Sand for construction

Table 4.37: Conservation measures on vegetation

Table 4.38: Conservation measures on soil

Table 4.39: Conservation measures on water

CHAPTER ONE

1.0 INTRODUCTION

1.1 *Urbanisation; Growth and Resource Use*

Urban settlement is a global phenomenon; at least every point of the global sphere has evidence of urbanization. Urbanisation is also one of the few phenomena that are revolutionalising the globe. For instance, whereas it is argued that urban areas occupy only 3% of the global surface, these areas command about $\frac{3}{4}$ of the global resources (Girardet 1992)

Growth of urban areas is a function of urbanisation although the process differs distinctly between the first world (otherwise called developed nations) and the 3rd world (developing nations). The growth of urban areas in the developing nations unlike the developed is primarily driven by social factors basically provision of social infrastructure. This contrasts sharply with the developed nations where economic factors; competitive marketing and economies of scale account for the process. The foregoing probably explains why the process is so rapid in the developing nations. In Africa for instance, by early 1950s, the continent was predominantly rural about 95%. During the 1960s urban population accounted for 18.8%. It is reported that urbanisation has grown by 4% in the last 30 years (UNEP; ROA: 2001)

Urban settlement in Kenya like the rest of Africa is a function of rural urban migration. Growth is rapid in major towns. In Nairobi for instance, population has increased six folds from 300,000 in 1962 to over 2 million today (APHRC; 2002). Settlements have grown by about the same margin. Although one would argue that due to rapid population growth amidst deteriorating health and economic standards, the informal settlements have grown immensely. Reports indicate that there are more than 100 informal settlements in Nairobi alone. These cater for people from the lower economic class.

Urban settlement is the greatest invention of the human mind (Girardet 1992) Most of the largest global centers of modernization have grown without manuals. What begun as simple or aggregate residential areas gradually grow into large conglomerates of unknown magnitude. Residential settlements and urbanisation seem to go hand in hand, in most cases residential areas precede the expansion of urban areas, and hence serve as colonizing instruments in urban expansion. Today, because of technology, transport networks are quickly replacing residential settlements in this role. Many people for instance are able to live far away from their places of work and still get to work with ease daily.

Urbanisation poses a great challenge to humanity today, some scholars have equated urban area problems to global problems; Global warming, Biodiversity loss, Transboundary movement of pollutants and general desertification spread. Of late however there is a bid to appraise urban centers as engines of economic growth. This is at a cost and expense of the environment.

Global land degradation trends reveal that approximately some 1900 million hectares worldwide suffer degradation, about 500 million hectare are found in Africa, some 65% of degradation occur in high agricultural areas and which means, some 45% of these degradation is occurring in the Arid and Semi Arid lands (ASALs) (GEO; 2000)

In Kenya, general resource degradation has taken place on several fronts; land in the humid and sub-humid tropics, the Arid and semi Arid land (ASALs), water, forestry. Prime contributors to degradation have been listed as industrialization, urbanization and human settlement. This is rather worrying for a country whose statistics indicate that 80% of its land is considered arid and semi arid. Only 17-20% is arable yet 35% of the arable land lies fallow (UNEP1991)

Degradation on land in high potential areas results from intensification of agriculture and soil loss through soil erosion. Crude statistics indicate that in 1986, approximately 2million and 2.5 million metric tones of soil were washed down the Tana and Sabaki rivers respectively into the Indian Ocean from the hinterland. The scenario is extrapolated

to other highland in the country. Degradation on land has been blamed largely on a collection of factors; Agriculture, Industrialization, and urbanisation.

In dry lands, the situation is no better. Campbell 1977 reports, degradation is a result of competition between different land use practices; settlement, urbanization, agriculture, horticulture, forestry soil and water conservation. These regions are said to occupy over 80% of total land, some 571,416 sq Km of total 649,049 sq Km of land (UNEP: 1999). The importance of these lands lie in the fact that at least 25% of human population, 50% of country's cattle population, 65% of sheep and goats and nearly 100% of camels is found in these areas (NEAP1994). Apart from these, ASALs are regions of high biological diversity. Most of the country's wildlife is found here too and is of immense significance to the country's tourism which account for 12.4% of country's Gross Domestic Product (GDP). ASALs contribute 3% of annual agricultural output and 7% of commercial production (NEAP 1994, NDP 1999-2008)

1.2 The Dynamics of ASALs in Kenya

There are about 22 original ASAL districts in Kenya, classified as 9 Arid and 13 Semi Arid. Communities occupying these areas are primarily pastoralists and agro-pastoralists. Most of these are also frontier districts i.e. found on international borders with communities freely moving between countries at different times of the year. They are prone to insecurity due to failure by different state governments to enact restrictions on citizenship. This has encouraged dual citizenship between states with no loyalty to any particular government.

Similarly, these areas have been neglected by successive governments in providing firm policy direction and infrastructure development, because of their perceived insignificant contribution to national economies (SNV: 2001). Post colonial governments failed to see infrastructure as a way of opening the regions to opportunities and self dependence, instead effected infrastructure only as a declaration of state control and administration of national boundaries (SNV UNPUBLISHED SOURCES: 2001). Internal insecurity has

been perpetuated tribal sense of statehood, and competition for resources; land for grazing and agriculture read inter tribal and cross border conflicts; Kisii-Maasai in Trans-Mara, Maasai-Kipsigis in Bomet, and Turkana-Pokot-Marakwet in North-rift are just but a few examples.

Failure in firm policy direction is exemplified by reactive as opposed to pro-active actions by government and lack of implementation as well as multiplication of organs to oversee overall environmental protection. For example, there was the formation of National Environmental Secretariat (NES) -1972, District Focus for Rural Development (DFRD)-1983, The Permanent Presidential Commission On Soil Conservation And Agro Forestry (PPCSCA)-1983, the formation of various regional development bodies, e.g. TARDA, KVDA, LBDA (1984) again, Inter Ministerial Committee on the Environment (IMCE) mid 1980s, National Environmental Action Plan (NEAP) 1994 and The Environmental Management and Coordination Act (EMCA) in 1999. Today we have the Arid Land Resources Management Programme (ALRMP) based in the office of the president. Most of these organs had overlapping and sometimes contradictory mandate in adjudication of their duties.

Arid lands are hostile climate. Herlocker (1999) in describing this climate notes that these areas are primarily hot, although in Kenya the temperatures seldom exceed 37°C, evapo-transpiration rate range between 2000-2500mm per annum, and rainfall of about 500mm/yr. There are slight variations in rainfall from place to place. Most notable however is the discrepancy in different ASALs classification, normally a range of 250mm/yr in very arid areas, and about 750mm in semi arid. Along coastline and shores of water bodies rainfall is about 1000mm/yr. Rainfall is normally calculated as an average of many years, although in actual sense the annual rainfall falls far below the long-term annual averages, for this reason rainfall is said to be highly erratic and may occur as many years of no rainfall and a few very wet years (Beaumont: 1989, Herlocker: 1999)

Soils and vegetation are also an important aspect of these ecologies. Soil is a function of parent material, climate, time and living organisms. Some scholars have attempted soil

and vegetation classification in East Africa. Morgan (1973) was able to delineate broad soil classification for East Africa using an index of soil units. Soils are depicted as wide area mosaics; these soil mosaics negate minor discrepancies from one area to another and therefore lack small-scale details. Again, characteristic vegetative mosaics depict areas with occasional slight variations showing differences in soils. Pratt and Gwynne (1975) have successfully provided an eco-climatic description of ASALs in East Africa using vegetation. Using modifications of the same from Herlocker (1999) we can delineate Kenya ASALs as falling under the following broad vegetation species areas; *Chrysopogon* mid-grass area, *Cenchrus/Chloris* Sub-region, *Themeda* mid-grass area and *Aristida* annual short grass (Herlocker: 1999)

The research infers impacts of urbanisation in the ASALs by attempted simulations and extrapolation of effects observed in other regions, it also attempts to see how best these explain the local circumstances. The research expects significant variation in coping mechanism between high potential areas and the Arid and Semi Arid areas. While high potential areas are highly dynamic and can recover quickly, arid lands are prone to degradation and their recovery mechanisms are slow.

1.3 PROBLEM STATEMENT

Urban residential settlements are an integral part of the urban settings and are central to social organization for urban areas' most important capital; human resource. Urban functions; commercial, industrial and residential occur as sets of concentric circles each portraying different function of an urban centre. Residential setting will be found on the outermost circle bordering rural neighborhoods, as such, they may be said to be the colonizing tool of urban areas i.e. they convert rural lands for urban use.

Urban residential areas are basic consumption units, most production processes the target these settings and may in turn be said to drive most production processes, and generate values that determine how resources are utilized. As regards natural resources, residential units open up new land for occupation writes Piyush (2001), they clear vegetation for occupation, harness water, food and raw materials (fuel energy). In addition, seal surfaces interfering with hydrological cycle. Thus, residential areas if not well managed present unsustainable means of managing urban ecology.

This paper is interested in natural resources namely; water, soil and vegetation in Arid and Semi Arid lands (ASALs), which are fragile ecologies, their use in residential areas, and how these uses may affect their availability in the long run. While both high potential areas and low potential (ASALs) maybe subject to the same kind of degradation from urban settings, this paper argues, ASALs are more fragile. They have poor coping mechanisms, most coping mechanisms are slow, subject to quick stress, and may take so long to recover or may never recover at all.

In view of these relationships, the study presumes. Residential settlements as an urban function have high demand for water in the ASALs amidst limited water resource availability, most rivers are seasonal and perennial rivers may have low water volumes to adequately carter for various needs. Water is mostly supplemented by underground sources with or without recharge areas, again the recharge of such a source is a slow process sometimes displaced in position. Where settlement occur on upper catchments

areas or resource potential areas as is the case of most ASALs, may significantly affect recharge levels and surface flow.

Secondly, urban settings increase compaction and surface sealing of land through vegetative clearing, constructions, road networks, and concrete covers. Compaction interferes with infiltration of water into the ground; it allows quick water saturation levels of soils as opposed to slow percolation into the ground. This encourages flash action where water is drained as quickly as it is received, generates surface flows and run-offs in large volumes as storm waters which cause massive erosion on the margins of urban areas and delivering large volumes of sediments. These in retrospect interferes with underground recharge levels and ability, increase in rivers thereby reducing their reticulation period.

Lastly, residential settings clear vegetation for occupation, demand construction material wood poles and timber, and bio-energy (wood fuel and charcoal). Construction removes much of the vegetation cover that may or may not be replaced after construction instead surfaces are compacted. Natural resource used in construction; poles, timber, wood and energy use; fuel wood are sourced from nature as is wood fuel and charcoal. These are first sourced from within before expanding to other areas as has previously been reported of Ngong and Oloolua forest in the study area

Vegetation clearing may not only produce visible topographical, hydrological and climatic changes in the long run as observed by Stolgen (1998) but may also severely affect livelihoods.

The study argues there is increased residential settlement in the study area, In pursuit of this knowledge; the study seeks to answer the following Research Questions:

- 1a. what is the evidence pointing to increased residential settlement in the study area?
2. What are the main sources of water for residential areas in the study area, their availability, and sustainability?
3. How and to what extent does a residential settlement contribute to soil compaction and sediment generation?
- 4a) how and to what extent do urban residential areas contribute to vegetative degradation?

1.4 OBJECTIVES

1.4.1 General Objective

The overall objective was to examine natural resource use in residential units within urban areas and to infer effects, which the observed uses are likely to have on general resource use and availability in the area of study

1.4.2 Specific Objectives

- Examine and determine evidence of residential settlement in the study area, Identify key residential settlement population dynamics and their influences on resource use in study area
- Examine the extent to which urban residential settlements contribute to soil compaction and sedimentation in the study area, and compare resource use between different land use areas
- Determine sources of water for use in residential units in the study area, its availability, sustainability and compare use in different land use areas
- Establish extent to which residential units contribute to vegetative reduction and compare use in different land use areas
- Examine extent to which residential areas contribute to sand, stones extraction in the study area
- Suggest and make recommendations for policy direction in the study area

1.6 SCOPE and LIMITATION

1.6.1 Scope of the Study

The research looks at three basic resources; soil, water and vegetation. It examines evidences of soil compaction, sediment generation and their extent. It also looks at sources of water, residents' satisfaction with the available sources and the challenges facing existing dominant sources. It further looks at vegetation and examines residents initiative in revegetation in residential areas (tree and soil cover) further looks at sources of energy in residential areas and examines the place of fuel in environmental conservation measures of the study area

1.6.2 Geographical Scope

The research was done in Ongata Rongai township of Ngong division, one of the three major townships in the division. The findings are however inferred on the other townships of the division; Kiserian and Ngong. The study assumes the townships have undergone similar geo-physical, socio-economic and political development and are bound to be similar in nearly all aspects. Changes in any one town are more likely to be reflected in all others.

1.7 Constraints and Limitations

The study has limited use of technical terms. Use of technical terms has been simplified to facilitate easier comprehension of resource use and their dynamics. The study should serve as an insight and impetus into the research subjects and prompt more study where shortfalls are evident. Resources in question are considered relative to urban demands and source.

Time: The time stipulation of three months was very short to complete everything in time. The study took long and the researcher believes all or most of the important issues have been covered. The study comes later than its expected time of completion hence all the study findings ought to be placed in its rightful context of time i.e. 2004 for the study area.

Finances: The study was primarily financed by the researcher as part requirement for the fulfillment of an academic programme. Finances were a major limitation and the study took longer than the stipulated time, process of data gathering and processing was slow with several intermittent stops. The quality of the work is however assured.

Accuracy of Information: The researcher strived to access most important materials for the work. There may be slight variation in reporting especially dates of various events that stem from various secondary materials used. The research does not therefore take responsibility for inconsistency of dates from different publications. Information

accuracy is backed up by key informant sources who were government resource experts on the ground, and whose views may reflect existing position on the ground

Bureaucracy: The process of obtaining clearance for research was challenging, given that most of the procedures of obtaining clearance are latent without clear procedures and only come out later. For example, while the education ministry gives permit for research, depending on the study area, it is also important to obtain another clearance from the parent ministry hosting the study. Though there seem to be agreement on this and it at times appears to be the discretion of the officers on the ground, which may act as an obstacle.

Suspicion and Hostility: The study was carried out at a time when there was high incidence of insecurity in the study area; hence there was a lot of suspicion and open hostility on the side of residents. Replacements were made for residential units that could not be accessed through dialogue or diplomacy. This however may not possibly affect results in a significant way.

1.8 ASSUMPTIONS

The study makes the following assumptions;

The towns in the study area have undergone similar geo-physical, socio-economic and political development and is bound to be similar in nearly all aspects. Observations in any one town are likely to depict observations in all others.

There is an influx of residential settlement in the study area that may have affected resource use in one way or another, from documentation the study attempt to document these changes tracing to the last 10 (ten) years before study. It documents findings and changes as at time of study assuming ten years before study the area was predominantly rural

The distinctions in land use areas designated as commercial and agricultural, are good pointers of effects of different land use practices. These may depict the growing effects for example of commercial on rural neighbourhoods, again assuming restrictions in land use areas.

1.9 LITERATURE REVIEW

1.9.1 Desertification; Growth and Perceptions

In order to gain insights into this study, literature review was placed within the wider discussion of desertification, its growth and perceptions. Much attempts to draw relationships between urbanisation, human settlement and desertification, globally and within Africa and isolated evidences and influences in Kenya. We also consider policy development, and growing concerns for these regions in Kenya.

Globally the arid and semi arid lands cover some significant portions of the global surface, some 40% of global land is considered arid. Arid like conditions also exist in a number of continents; Africa, Asia, Australia, Europe, North and South America and in some isolated countries reports (Beaumont 1989). GEF (1999) reports that their environmental, social and economic effects are felt in more than 110 countries including more than 80 developing countries and where over one billion people are at risk from effects of serious decline in production and livelihoods.

Although desertification is as old as humanity, a lot of awakening came after the 1979 United Nations Conference to Combat Desertification (UNCCD), augmented by numerous earlier world ecological disasters. Careful serialization and documentation about this time afforded this. For example the Ethiopian famine 1973, later, Sahel drought of the early 1980s. OECD (1988), Timberlake (1985) and Beaumont (1989) have separately talked of the build-up factors and devastation of the Sahel drought in their works. In Kenya the realization was rather slow and most of the development of the ASAL areas was left to Non governmental organizations and other government development partners or donors.

Desertification has a human face; there are links between desertification and human activities agriculture, irrigation, pastoralism, settlements and urbanisation. Desertification can occur anywhere although some places such as the ASALs are more susceptible.

Desertification is an interactive and latent process; no one distinct human activity can be exclusively linked to it. Literature abound talk of slow build up processes that are difficult to tell. Globally, archaeological evidence talked of ancient cities of Mesopotamia

that literary fed on land leaving behind pans of salt in their place (Girardet 1996). More closely in Africa, Timberlake (1985) citing archaeological evidence talked of a once flourishing Sahara ecology that was gradually worked away through unsustainable timber harvesting and export. On the other hand, the Sahel drought was attributed to unsustainable agricultural practices. In Kenya there is no comprehensive evidence of desertification, except for isolated cases of resource degradation. Campbell 1977 talked of destroyed livelihoods of the Maasai, as a result of changed economy; Shanyisa 1991 and Agatsiva 1996 have separately talked of the role of rural and urban energy consumption (wood fuel and charcoal) in vegetative degradation. There is no agreement whatsoever between them on the role of rural wood fuel consumption on vegetative degradation.

Deserts and Dry lands; Deserts are dry lands, dry lands are distinguished for their unique climate; The Unesco; Man and the Biosphere programme- Kenya in trying to distinguish between dry lands developed a classification of these lands based on three indices; Precipitation, Potential evaporation/ transpiration and temperature; calculated as an index and expressed as a ratio i.e. $P / P.E.T$. On the basis of this, they have been able to show different degree of aridity hence susceptibility of different areas and ability to withstand stressful conditions.

Dry lands are basically hostile and fragile ecologies, flora and fauna found in these areas are well adapted, and characteristically these areas have low biomass hence low carrying capacity. Golany (1978) notes that the dynamics of these areas is slow and the period of most coping cycles are long. Because of this, resources are prone to degradation upon subjection to any slight environmental stress. In these lands, vegetation plays a significant role in the maintenance of balance in these ecologies, in most cases desertification of these areas has been linked to loss of vegetative resource (ICRISAT 1995)

Again availability of water is affected by precipitation; normally precipitation is low and the surface evaporation high. Evaporation affect river reticulation most of which originate in the wetter margins of these ecologies, although rainfall is rare, it often comes in large quantities within a short while. Lack of vegetative cover reduces water infiltration into

the soil and also increases the rate at which soils attain saturation levels thereby generating surface flows, run-offs and storm waters.

Human activities not only destroy the environment but equally compromise the system upon which they rely on for their livelihoods. This inevitably is the relationship between human activity and the environment. We attempt to explore relationships and evidences of human activities on the environment. As noted previously, Sudano-sahelian drought was a function of disrupted resilience of the ecology writes OECD 1988, Timberlake 1985. Timberlake talking on agriculture argues that it is erroneous to argue that the Sahara deserts are expanding. Instead, human beings pull deserts outwards by producing localized desertification that eventually merge to form large areas of waste lands.

Rapid population increase, reduced land surface and occupation of wetter margins are some of the significant issues in degradation of ASALs. Today the global agenda is on reducing desertification not only in countries prone to the phenomenon but also desertification in wetter zones. This culminated in the opening of the United Nations Convention to Combat and Control Desertification (UNCCD) in 1992 in Rio de Janeiro. Solutions to desertification are as ambiguous as their causes.

Robin et al (2003) observed that major challenges facing dry lands arise from agriculture, human settlement, desertification, livestock grazing, global warming and mining.

Writing on urbanisation and arid lands, Johnson (1992) observed that modern factors behind the growth and expansion of modern day urban settings are primarily technological. Modern transport and communication has not only encouraged urban dispersal in residential areas but also other urban functions. This technology has of necessity opened up rural lands; humid and dry lands alike encouraging resource exploitation.

The ease of communication has produced a network of cities globally. This way it has afforded exploitation of resources previously unreachable. Girardet (1992) talked of the unsustainable cities of the world. He notes in particular that although cities occupy only

3% of the global surface, they command $\frac{3}{4}$ of global resources. In the 3rd world, urban settlement has been equated to westernization and many societies have embraced the concept fully, many families have tended towards smaller and independent family units thereby requiring large number of houses to house same number of people, consequently this leads to large spreads of residential units.

Urbanization and human settlement alter the functions of dry land ecosystem, which is primarily thought to be rural and agrarian. Urban transport networks not only intersect land but also alter the drainage as well as livestock migratory routes, White and Nickoney (2003) In Kenya, Campbell (1977) has pointed out that dry land degradation is a result of competition between various land uses; agriculture, ranching, wildlife, human settlement, urbanization and conservation processes. Similar observations are cited in NEAP 1994.

Stohlgren working in Colorado USA observed that human activity was capable of influencing both vegetation and climate in a predictable way; modification of natural vegetation was a function of agriculture and urbanization. Alteration of vegetation and climate change speeded degradation processes.

Michael et al (1999) found out that settlement and associated crop production not only reduced vegetative cover but also responsible for replacement of more permanent vegetation with temporal soil covers, these may sometimes seal the soils completely. In dry lands where vegetation is less abundant it may produce an environmental stress and or activate other destructive processes. As already noted, corrective measures in these regions are slow to set in and may quickly be overcome by such processes as wind and water erosion, unpredictable rainfall prompting poor infiltration, surface run-offs etc

Urban areas are seen as engines of growth when they open up rural neighbourhoods to the outside world and incorporate them into the mainstream country's economy. This occurs at the expense of the environment. Piyosh (2001) variously noted; urban problems are as well global problems. Writing about India, he noted modern cities of India like

rampaging demons had encroached on agricultural land, some 2.5 million hectares have been lost in the last 50 years

Another common problem posed by urban areas is the compaction of land as a result of trampling, fouling and concretization or sealing of surfaces with bitumen or concrete. More land surface is covered, impairing infiltration thereby increasing chances of surface run-offs. Large volumes of storm waters are delivered on the edges of urban areas that are not sealed with large velocities that cause soil erosion, land slides, as well as delivering large amounts of mud, sediment, silt rocks, pebbles and cause floods.

Urban areas have also been reported to adversely affect vegetative resources mainly forestry. As urban areas expand and grow, more land is cleared of the natural vegetation mostly for buildings, agriculture, raw materials etc. Where agriculture is possible, land is put under intense use, or more frontiers opened for use. One of the build up factors to Sahelian drought was intense farming, which reduced fallow periods, and cleared forests. The intense activity disregarded traditional practice of sparing shade trees to facilitate soil recovery through shading leaves and protecting soil from the scorching sun as well as wind erosion (OECD 1988; Campbell 1978,)

On the other hand wood fuel energy sourcing has been accused for most of the rural degradation. Joanne and Catherine (1999) and Chidumayo (1983) wrote separately on their encounters. Chidumayo in Zambia noted fuel consumption was responsible for rural deforestation; he decried energy loss through conversion of wood mass to charcoal, (a preference of urban dwellers). Wood was converted to charcoal by traditional earth kiln methods, which converts 14% of wood mass to charcoal. This he noted is an extensive consumer of forests; degradation is a function of urban consumption.

According to Kenya development plan 2002-2008, wood fuel accounts for about 68% of total energy requirements. Wood is a dominant source of energy for rural inhabitants, urban poor and informal sector, wrote Shanyisa (1991). The implications are heavy on the environment as forests either as direct wood or charcoal has to be utilized as a source

of fuel. Shanyisa adds, energy shortage implies that forests particularly the ASALs are over utilized for charcoal to provide for low and middle class urban population. In cases where wood fuel disappears all together, crop residues and animal dung is used instead.

Agatsiva (1996) in comparative study between rural and urban wood energy sourcing in Kenya ASALs (Kilifi district) suggested that rural household fuel consumption had no significant contribution to deforestation. On the other hand, urban demand; Mombasa, Malindi) choice of charcoal over firewood makes the commodity attractive choice for entrepreneurs who derive income from production and distribution. Large demands of charcoal in the two towns have led to large-scale felling of forests in Margarini and Ganze.

There are varied opinions on the role of fuel wood in degradation, sometimes other factors of social and economic preference put in consideration there is no general consensus, thus it can only be inferred contextually.

1.9.2 Policy Development for the ASALs in Kenya

The point of departure is the Stockholm Conference (1972). The United Nations Conference on the Human Environment (UNCHE –1972) in Stockholm Sweden concretized the idea of global environmental protection; many initiatives were taken from that time. Equally, environmental protection was accelerated and a number of conventions envisaged and ratified globally e.g. Convention for Control of International Trade in Endangered Species (CITES –1974), Convention for Migratory Species (CMS-1979).

In Kenya, the country was given the auspice to host the United Nations Environmental Programme (UNEP) offices in Nairobi. Also an international non-governmental organization Environmental Liaison recognized this status for Kenya and equally set its head office in Kenya, this set motion for environmental protection in Kenya (Acts: 1991).

The Kenyan government on its part established the National Environmental Secretariat (NES) in 1973, and a number of agencies to deal with environmental issues, a good number of these were in the ASALs, for example, the Soil and Water Conservation

Programme, the Watershed and catchments conservation Programme, regional bodies; the Lake Basin Development Authority (LBDA), Kerio – Valley Development Authority (KVDA), Tana and Athi River Development Authority (TARDA), and Coast Development Authority (CDA). These bodies were to identify and initiate projects in their respective regions to improve and foster sustainable development.

The National Development Plan 1974-79 indicates that by this time, most environmental protection policies in Kenya though highly fragmented, paid attention to developing arid and semi arid areas, marginal groups, urban poor and the rural landless. They advocated for Mainstreaming of these groups so as to improve social equity.

National Development Plan (1979-83) concentrated on land tenure in particular, establishment of a land commission to collaborate with the National Land Use Committee (to be established) in developing issues of land and Soil Conservation Act. The land commission was to identify land for various uses.

Approaches to environmental management has been highly fragmented (Koech 1991) numerous environmental protection organs have sprung up. Harmonisation of their roles was a great challenge, ambiguity prevailed and most had overlapping roles e.g. Ministry of Land Reclamation and Development of Arid and Semi Arid Lands, Ministry of Environment and Natural Resources; with the constituent National Environmental Secretariat (NES), forestry department, mines and geology, tourism, and wildlife among others. Whereas the drafting of policy as a requirement could have been an efficient task, there existed big lapse between formulation and implementation. Some scholars argue that policy development has sounded repetitive and routine in subsequent development plans with little or no modification. Others have never been implemented at all or done partially (ACTS-UNEP: 1991). Still others argue environmental policies have expanded in scope with successive plans.

Attention given to environmental issues vary greatly, more attention has been paid to issues of immediate implication either on agriculture and economy. Such sectors as

wildlife and protected areas have received more attention; forestry on the other hand has been quite controversial due to the competitive nature of agriculture and settlement.

Environmental legislation have also faired poorly, there was absence of a comprehensive environmental law. For instance, before 1999 there were 77 Acts of parliament dealing with different issues of the environment and often contradictory in their adjudication (Acts-UNEP 1991)

The National Environmental Action Plan (NEAP) was the first of its kind, recognizing different ecologies and their challenges. For example, it highlighted challenges, potentials and strategies for the ASALs. Its implementation was however overshadowed by a rush of events. First, it had been formulated as a response and a show of commitment to the Stockholm Conference 1972. Second, there were impending reports of progress for (UNCED- Earth Summit) resolutions dubbed Rio + 5 to be held in Rome, Italy. Lastly in pursuit of the Earth Summit resolutions, the government formulated the Sessional Paper Number 6 of 1999 on Environment and Development.

The paper emphasized the importance of Environment on the economy; it highlighted three issues of development, population growth and the environment. The paper visualized adoption of development strategies that ensured that natural resources are used sustain ably. However in the face of the then coming World Summit on Sustainable Development (2002) in Johannesburg, South Africa, the government overlooked its implementation.

Under the Auspice of the United Nations Environment Programme (UNEP), the government worked to formulate a comprehensive environmental law, which culminated into the draft Environmental Management and Coordination Act (EMCA) in 1999, this was adopted by parliament in December 1999 and given presidential assent on 6th January 2000. It was meant to come into effect a week later (ACTS-UNEP: 1999). However, it was not until November 2003 that necessary parliamentary bills were enacted to bring it into full force.

The National Environmental Management Authority (NEMA) is the body entrusted with the overall supervision and coordination of environmental matters. Today, within the ASALs, we have the Arid Land Resource Management Programme (ALRMP) under the office of the president; its overall mandate is to coordinate efforts by various stakeholders and players within the ASALs

1.10 Research Gaps and Justification

Our literature review reveals the following gaps,

There is significant absence of relevant literature on urbanization in Arid lands, What exists are isolated cases of resource degradation such as studies done by Shanyisa and Agatsiva on rural wood fuel energy and where there is no conclusive agreement

In the face of a new body of environmental law and a national coordinating body, the need for more sets of laws to control environmental relations and policy direction take precedence. This should be preceded by informed decision making on various relation. This work should contribute towards this end

Also, as a new field in environmental planning, this work should trigger more work in the same line to fill the information gap in urban and environmental studies.

CHAPTER TWO

2.0 STUDY AREA

2.1 *Description of Study Site*

The discussion in this section gives a detailed description of the study area i.e. geo-physical location, population characteristics, administrative boundaries, and climate and resource availability.

2.1.1 Location

The study took place in Ngong division of Kajiado district, (figure 1 shows the location of the study area on the map of Kenya). Ngong division is one of the seven divisions of Kajiado district; others are Loitokitok, Magadi, Central, Mashuru, Namanga and Isinya (see figure 2). Ngong division is the most northerly point of Kajiado district. It borders Kiambu district to the North, Nairobi to the Northeast and Narok District to the West. It borders two other divisions of Kajiado district which are; Central to the south and Magadi to the Southwest. Kajiado lies between longitudes 36 5' and 37 5' E and latitudes 1 0' & 3 0' South (District Development Plan 1997-2001, Kajiado)

The division is primarily arid and semi arid with only 20% of the land area being viable for agriculture. Pastoralism and sedentary agriculture are the major forms of livelihoods in the region.



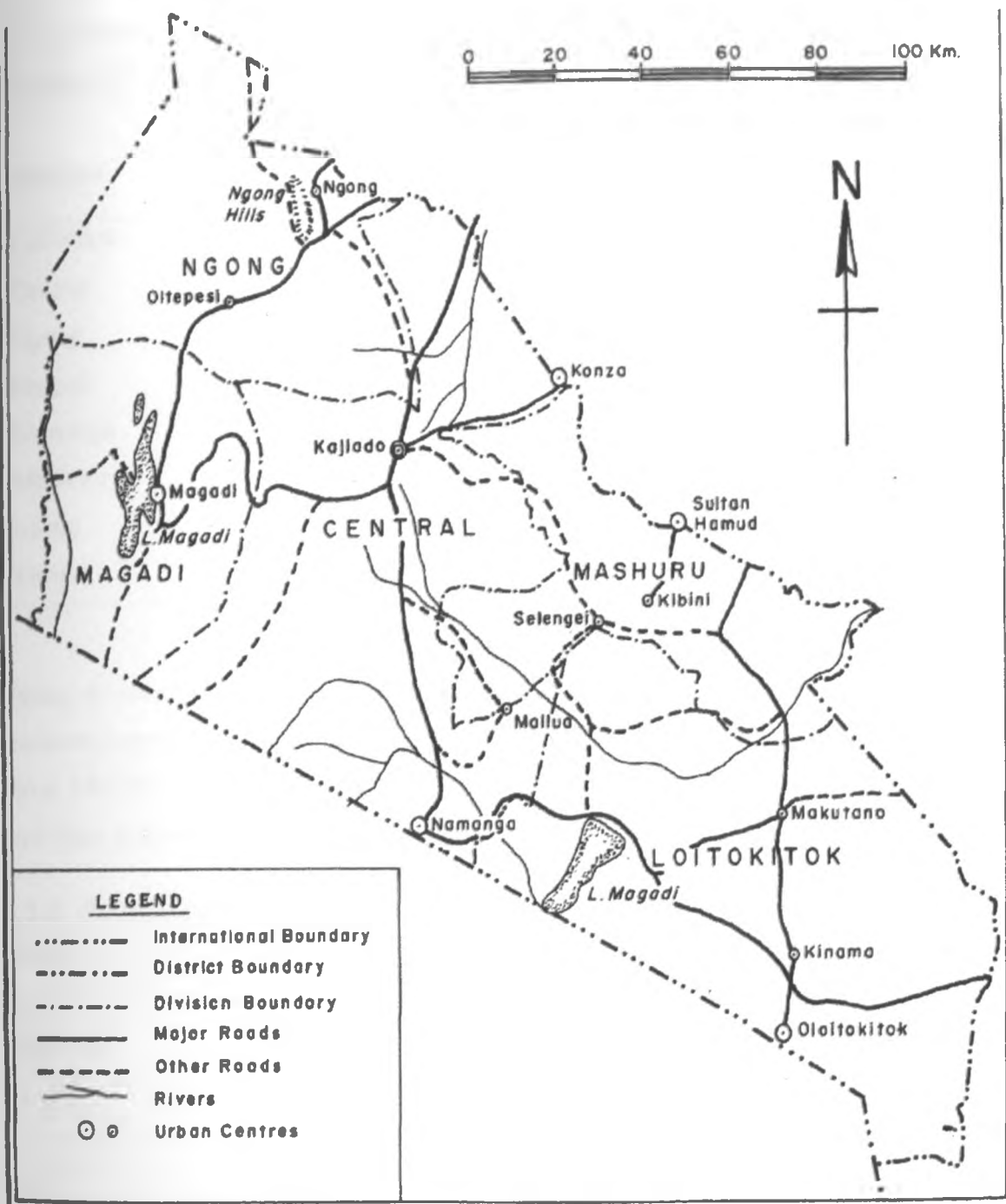


Figure 2 Map of Kajiado District showing Administrative Areas by Divisions and major roads,

Source: Modified from Kajiado District Development Plan, (1997 – 2001)

2.1.2 Population Characteristics

(Table 2.1) Population Distribution by Area and Density

Division	Area in Sq Km	Population	Density
Loitoktok	6,090	95,430	15
Central	3,825	69,402	17
Ngong	3,698	149,771	41
Magadi	2,749	20,112	8
Namanga	2,499	35,673	16
Mashuru	2,550	35,666	12
Isenya	-	-	-
Total	21,107	406,054	18

Source: Adopted from Population and Housing Census, GOK 2001

Ngong division is the third largest in Kajiado district behind Loitoktok and Central yet, the most populous. It occupies an estimated area of 3,698sq Km and with a population of about 149,771. This translates into a population density of about 41 people per sq Km according to the census of 1999 (CBS 2001)

2.1.3 Administrative Boundaries

Ngong division is one of the seven divisions that make up Kajiado district. Division has a total of ten location and 29 sub locations. Most of these have been created recently (See Table 2.2)

(Table 2.2)

Administrative Division, Location and Sub-locations

Locations	No. Of Sub- Locations
Ngong	2
Oloolua	3
Kiserian	2
Nkai- Murunya	3
Ongata Rongai	2
South Keekonyokie	3
Central Keekonyokie	5
Entashart	3
Mosiro	2
North Keekonyokie	4
Total:	29

Source: Participatory land use plan, Olcodo 2002

The total land area is estimated at 3698.1Km² with about 20% being of higher potential for intensive agricultural production leaving 80% to mainly ASALs and rangeland. It is also within this 20% that most of settling is constantly increasing hence reducing further the availability of land for cultivation

2.1.4 Topography and Climate

Figure 3 gives a summary of geo-climatic description using agro-climatic defined zones

The district has four main topographical areas. The Rift valley, consists of the low depression, the Athi-Kapiti which are open rolling lands of altitude of 2460m above sea level and include the Ngong hills, Central broken grounds; a 20-70 km stretch running from NE - SW and forms a permanent draining area, the Amboseli plains; these form undulating plains with deep reddish brown clay loams and flat sedimentary plains. Ngong division is basically part of the Athi-Kapiti Plains.

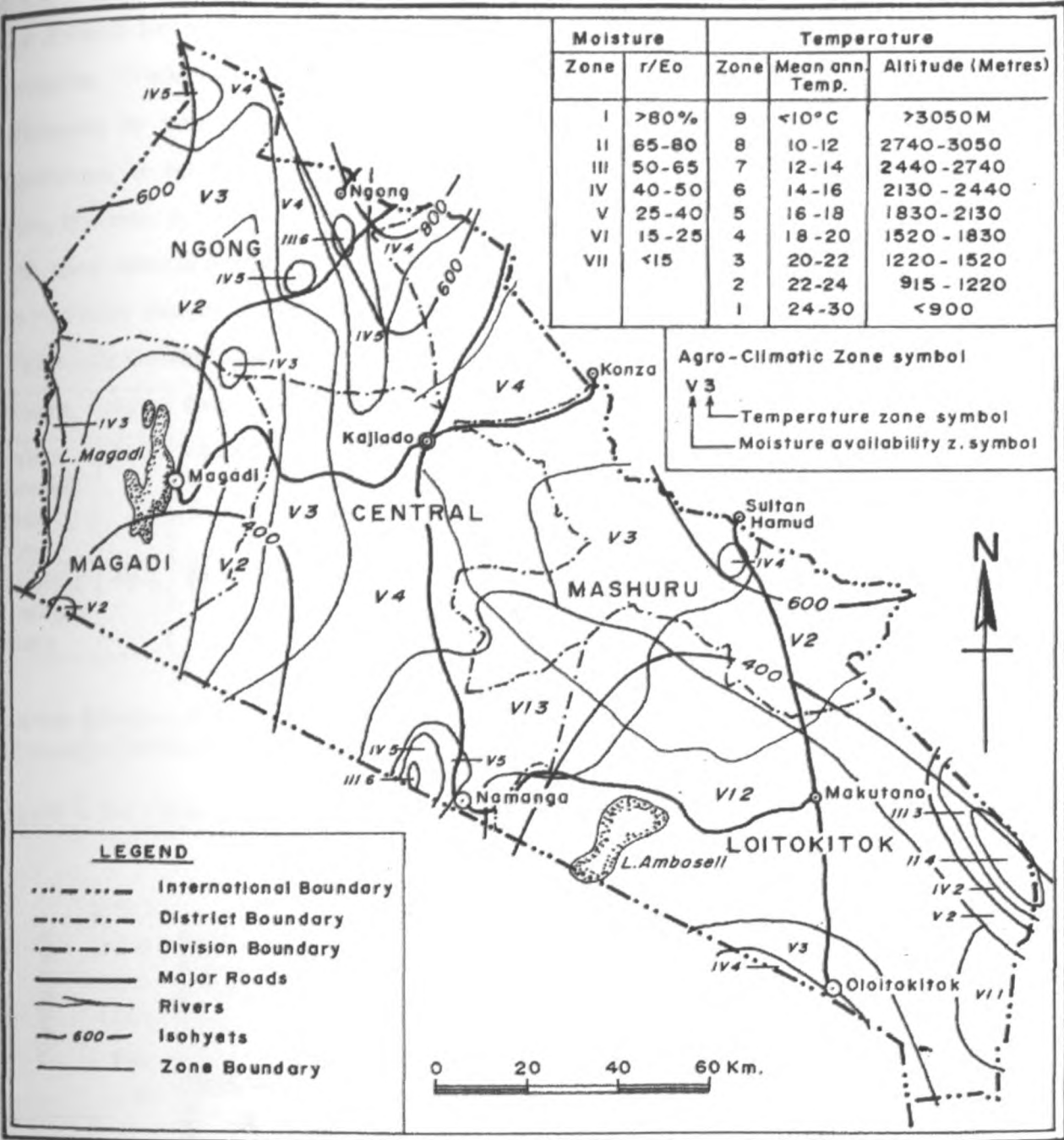


Figure 3: Geo – Physical description of the study area, using delineations of Agro – climatic Zones

Source: Modified from Kajiado District Development Plan, (1997 – 2001)

2.1.5 Rainfall

The division has a bimodal pattern of rainfall. Short rains come in September/October-December. While long rains are expected around March to May. Rainfall is strongly influenced by altitude such that areas of high altitude like the Ngong, Loitoktok, Ngurumani, in the district have more or less even distribution while the low altitude areas, however, in the last 6 yrs the pattern has been quite erratic and quite unpredictable with some seasons overlapping and others almost eliminated and an almost reversing of the two rainy seasons for some years. (See table 2.3 and figure 5 below)

(Table 2.3) Rainfall Distribution by Months

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
5 year Monthly totals (mm)	202	116.8	332	1070.5	551	79.2	66.8	54.9	125.5	262.5	701.9	341.2
Monthly averages (mm)	40.4	23.36	66.4	214	110.2	15.8	13.36	10.98	25.1	52.5	140.38	68.24

Source: Ministry of Agriculture; Ngong Division. Rainfall calculated as a 5 year average by months (2000-4)

Figure 5: Bar Chart Representation of Annual Rainfall based on six year Average

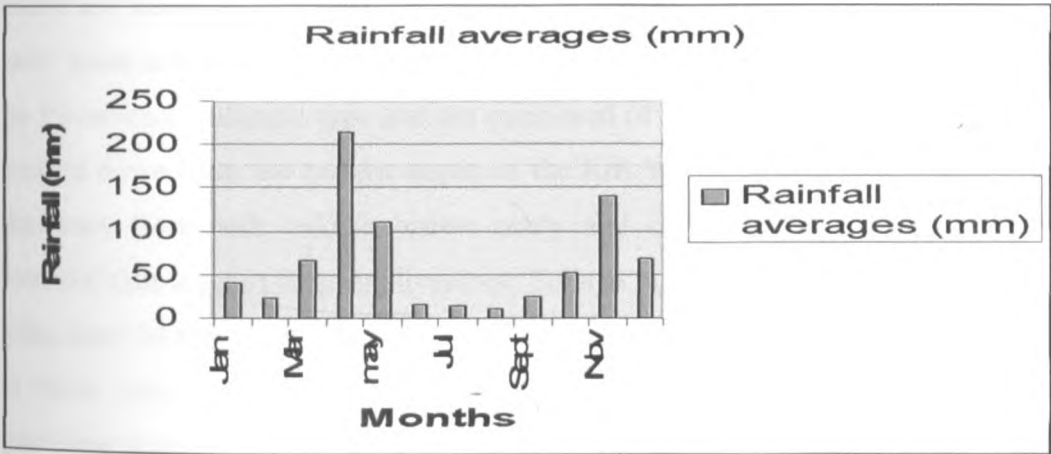
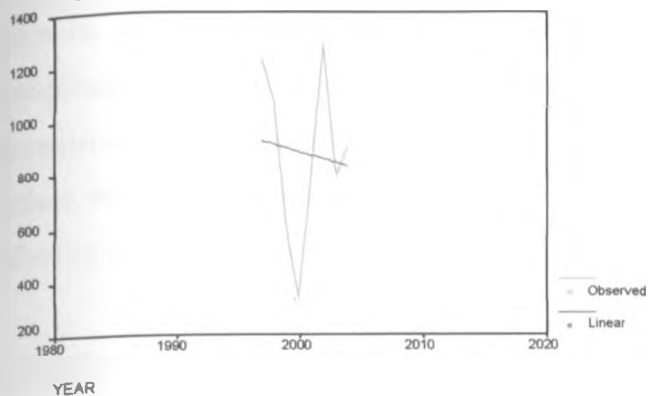


Figure 6: Rainfall fluctuation (1997 - 2004)



Source: Field data 2004

2.1.6 Temperature

Temperature is given as a distinction of the highest and the lowest months in Ngong division, monthly averages of temperature distribution are significantly absent. Minimum temperatures range between 10-12°C, coldest usually in June and July. Maximum temperatures range between 24-28°C hottest usually in February and early September.

2.2 Resources:

2.2.1 Soils and Geology

There are three broad categories of soils in Kajiado District; i.e. Quaternary volcanic soils, basement rock soils and Pleistocene. The soils in Ngong division are basically of the Pleistocene sediment type and are composed of sediment from tertiary volcanic rocks washed down from the eastern slopes of the Rift Valley escarpments (Neap 1994) Soils also vary from dark reddish brown sandy and clay. Normally, they are of medium potential thus support crop and livestock. Soils in Ngong division range from clay in low-lying areas to red loams and some regions have very light sandy and light textured soils. In these areas the soils are quite stony and rather shallow hence the suitability of quarrying activity carried out at various spots of the division like Oloolua, Kiserian and Ewuaso area of North Keekonyokie location.

2.2.2 Water Resources

The Ngong hills in Ngong division are primary catchments for the entire district. The catchments cover about 2080sq km, and forms part of the larger Athi catchments (see

figure 4). The division has two main streams, Kiserian and Mbagathi alongside other perennial and seasonal streams, a large number of streams are seasonal due to low precipitation, high evaporation and severe runoffs drought is frequent (ibid), an alternative source of water for domestic, agriculture livestock and urban use is sub-surface waters, boreholes and springs. Climate, topography and the underlying rocks influence the availability of ground water.

Generally, water tables are low owing to the porous nature of rocks that allow water seepage to low ground levels. Boreholes are drilled to a depth of about 175-250m in Ngong division, the quality of water from the boreholes is good (fluoride and dissolved compounds level is low) hence suitable for consumption. In other regions of the district, water is excessively saline. The average yield is about 27-70 cubic metres in Ngong area. In urban areas, boreholes serve large populations at relatively high costs, also a number of them were sunk without proper hydrological surveys hence have/are drying up.

2.2.3 Communication and Infrastructure

Part of the division is well served by all weather roads. These are mainly parts of Oloolua, Ngong, Nkai-murunya and Ongata Rongai, locations that are also well served with telephone and electricity. There are two other tarmac roads; pipeline and Magadi, which traverse south Keekonyokie to Magadi. These assist in transportation to the south of the division. All other roads are murramed and seasonal and usually not accessible during certain times of the year.

2.2.4 Vegetation Cover

All forests are found mainly on water catchments area; some 2080sq Km gazetted forest land is considered catchments. Reliable details on encroachments and destruction are however lacking, most forests are a combination of indigenous and exotic trees. The dominant

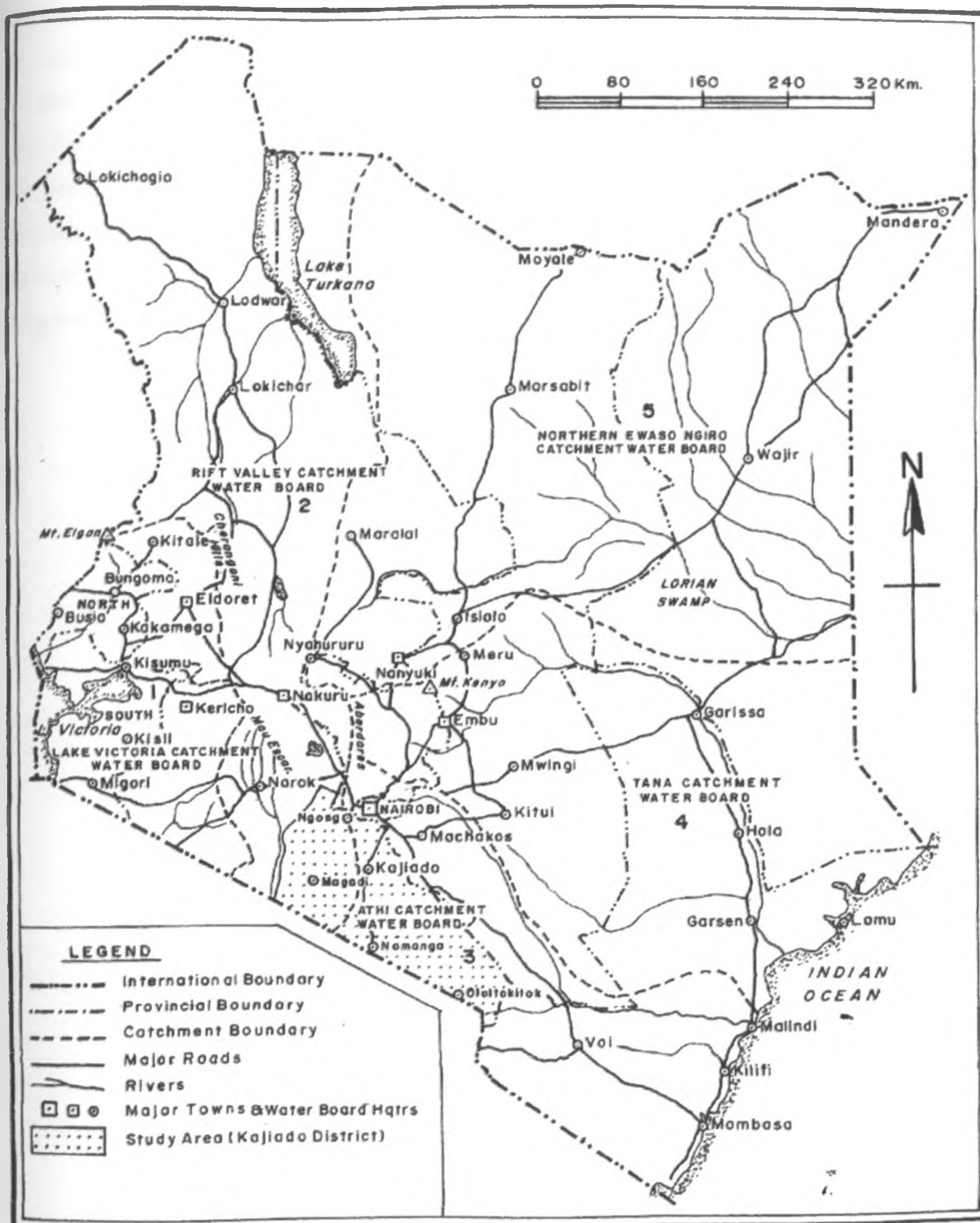


Figure 4: Map of Kenya showing Drainage and Catchment Areas.

Source: Adopted from Kenya National Atlas, Survey of Kenya.

form of vegetation is the Acacia trees and thorny thickets undergrowth and savanna long grass; acacia trees grow to a height of over 10m. The heights however vary with soil depth and aridity; the same is true of dispersion. Grass form an extensive cover of the Ngong forests, characterized by savanna grassland and growing to a height of above 1m. In the past, there was plenty of vegetation in Ngong but due to the increased human settlement and land fragmentation especially in Kandisi and Olkeri respectively; vegetation cover has been severely reduced. Overstocking and excessive charcoal burning has reduced the vegetation in Esonorua and Kimuka Sub- locations.

UNIVERSITY OF NAIROBI
EAST AFRICANA COLLECTION

CHAPTER THREE

3.0 METHODOLOGY

Discussion in this section dwells on the research design and entails the description of procedures and techniques employed, sampling procedure, mode of data collection, type of data collected, data analysis, research findings and presentation

3.1 *Research design*

A field Survey was conducted in Ongata Rongai Township. A distinction was made between existing land use designated areas i.e. commercial and agricultural. Representative samples were drawn from these two areas. Transect lines (hypothetical line of study), were drawn within respective residential areas along which samples were drawn using systematic sampling with a random start. Questionnaire tool were then administered using interview schedules, observation made of residential areas, photographs taken and discussions held with government specialists and resource persons on the ground mainly as key informants.

3.2 *Sampling Procedure*

Before sampling was carried out some considerable time was spent, on reconnaissance, as well as wide consultation with actors and government officials at the Olkejuado county council offices in Ongata Rongai.

Universe According to information at the local county council offices there were about 750 registered residential settings (plots). There were another 500 registered commercial settlements (plots), totaling to about 1,250 units, and a further, 250 registered business premises. At least 10% (Ten) of each category was to be interviewed i.e. 75 units from agricultural residential side, 50 from the commercial side and further 25 of the registered business premises. These sample sizes were divided equally along each transect based on required sample size from each respective residential area. As such, each transect generated its interval of sampling units.

Procedure:

Multi-stage sampling was adopted for the study.

Step 1

One town (Ongata Rongai) was selected from two others; Kiserian and Ngong.

Step 2

Secondly, Equi-distant line transects were generated in respective residential areas within the township, some 500m apart on either side of the road dividing the township into two.

See appendix 1 on generation of line transects in Ongata Rongai Township

Step 3

Ten (10), line transects were generated in the township, five on either side, in respective land use areas; 50units and 75units for commercial and agricultural respectively were then divided equally along the five transects on either side, as such, each transect on the commercial residential side had 10 units and each transect on the agricultural side had 15 units.

Step 4

Houses were listed along individual transect lines on a listing sheet (See appendix 2 for sample listing sheet), residential unit was serialized on a form, described physically e.g. street of location, landmarks and other outstanding features given. This was done to ease identification of unit when coming back to conduct interviews.

Step 5

A separate sample calculation sheet (See Appendix 3 sample calculation sheet)

Provided procedure sampled units (selected units). The calculation sheet provides for determination of the sampling interval, selection of random start, and other units subsequently selected. Once the exercise is complete, the selected units are cross checked on the listing sheet for their location, (Appendix 2). The selected units were marked (using a highlighting pen) and the unit visited for interview.

Interviews were conducted on individual residential unit as where and when a responsible member of the household was available at home. Where no body was present, any other house there about was eligible for interview; this was so as each household had an equal opportunity of being sampled. A total of 126 questionnaires were administered to the households. At least 10% of the residential and the commercial units were successfully interviewed, these translated to 75 and 51 for the residential and commercial households respectively.

Another 25 questionnaires were administered to commercial traders representing 10% of traders. Study in commercial traders targeted dealers in respective resources; timber and timber products charcoal and firewood, water and some as well as sand dealers

3.3 Data collection

Fieldwork was carried out during the month of November 2004, a period when short rains are expected i.e. between October and December, on the contrary the area was considerably dry during this time.

3.4 Type of Data Collected

Primary data: A field survey was conducted, primary data was collected using interview schedules and aided by field research assistants. Short interview schedules were also organized with key informants; government officials from the Ministries of Water, Environment and Natural Resources and Agriculture.

Observation was also used in noting significant occurrences in residential and commercial units; photography and field notes were taken to supplement analysis of data.

Secondary data was used to supplement primary data, this was obtained from published books and government documents as well as unpublished sources from non-governmental organisations operating in the study area.

3.5 Nature of Data Collected

Objective 1

The type of data collected for this objective included: - gender, age, marital status, family size, literacy and education, highest education attained, occupation and mode of tenure. This provided qualitative data analysed alongside mode of tenure. It provided information on the socio-economic characteristic of the dwellers in a bid to determine original occupants.

Objective 2

Type of data collected:- Presence and extent of concrete surfaces, open bare grounds, stone covered surfaces, animal stalls, kind of soil cover present, estimated vegetative cover, estimated cover of bare soils, estimated cover of gravel, proximity of homes, road networks, construction materials, mode of building and use of soils in construction sites.

This enabled the researcher to infer compaction levels, sediment generation and generation of surface run-offs and accounts erosion levels on the margin of urban areas.

Objective 3

Type of data collected for this objective included: major sources of water, alternative sources of water, available water, perceived water needs to estimate demand, state of water resource (quality), difficulties in water sourcing, perceived future demands and supply.

Objective 4

Type of data: layout for residential units, presence of grass lawns, kind of soil cover, extent of bare soils, proximity of homes, road networks, attitude and peoples behaviour affecting vegetative cover. These sought to understand and explain people's behaviour as they affect vegetative cover, and the changes observed

Objective 5

Type of data: Construction materials and their sources

3.6 Data Processing Analysis

Preliminary data treatment entailed processing of data, cleaning, data reduction; i.e. data was captured using computer-aided programme Statistical Package for Social Sciences (SPSS)

Quantitative were used to analyse the data. Frequency was the most basic approach to data, further analysis entailed use of contingency table, cross tabulation and calculation of Pearson's correlation coefficient r : Hypothesis testing using chi square was applied on grouped data measured on interval scale.

Qualitative data was done though summarizing of interview information from key informants, field reports and sorting and editing of photographs after quantitative data analysis and report writing.

3.7 Data Presentation

Data is presented in form of summary contingency tables, summary tables, graphs and interpretation notes or where necessary data description

CHAPTER FOUR

4.0 RESULTS AND DISCUSSIONS

Discussion in this section entails reporting on testing of the hypothesis, research findings and discussions on the research findings. Hypotheses were tested for variables for different

4.1 Research Findings

Objective 1

Examine and determine evidence of increased residential settlement in the study area, Identify key residential settlement population dynamics and their influences on resource use in study area

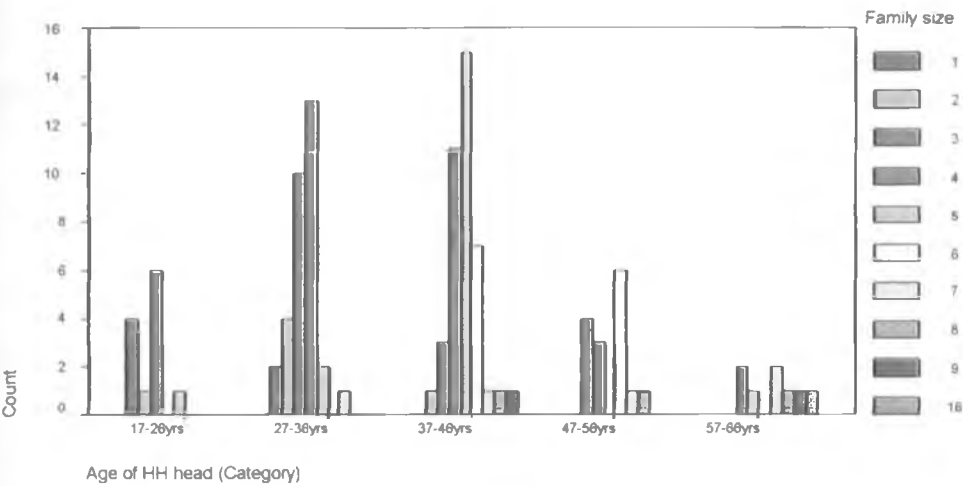
To achieve this objective, evidence was drawn from the socio-economic characteristics of the population such as age of household head, family size and mode of land tenure, occupation and professional qualification, household source of income and years of residence. There is evidence of increased residential settlement in the study area. Table 4.1 shows; age of Household heads lay between 17 -66years, 93.7% of the household heads were between ages 17 -56 with the dominant age group being between 37-46years followed by age group 27-36years representing 32.5% and 29.4%. These are people in their prime ages, both settling in marriage, and trying to accumulate wealth through savings from employment, mortgages or business hence possibility of increased settlement.

Table 4.1: Age of HH head

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	17-26yrs	25	19.8	19.8	19.8
	27-36yrs	37	29.4	29.4	49.2
	37-46yrs	41	32.5	32.5	81.7
	47-56yrs	15	11.9	11.9	93.7
	57-66yrs	8	6.3	6.3	100.0
	Total	126	100.0	100.0	

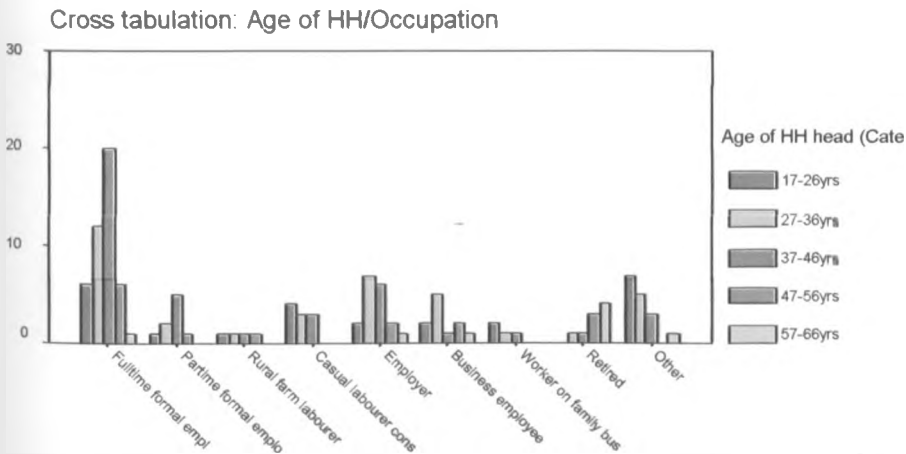
Source: Field data 2004

Across tabulation of age of household head against family size indicates that people between ages 27-36 had between three to four children (see blue and pink colours) while ages 37-46 had between four and five children (see pink and turquoise colours). This still points towards the possibility of these parents looking for stability or somewhere to settle their families. (See figure 7)



Source: Field data 2004

A cross tabulation of age of household head and occupation indicated that most household heads between age category 36-56 i.e. 26 or (57.7%) were in formal employment, on the contrary majority of the age group below 36 years i.e. 7 or (70%) were casual labourers in construction. Casual labour in construction is easily available within the township, whereas formal employment is sought out of township, (See Figure 8 A cross tabulation of Age of household head against occupation).



Source: Field data 2004

This may point to fact that in essence most residents have sought accommodation as a result of formal employment. Table 4.2 shows the mode of tenure, rental was dominantly 65 (51.6%) while individual ownership was only 51 (40.5%). other forms of tenure were insignificant.

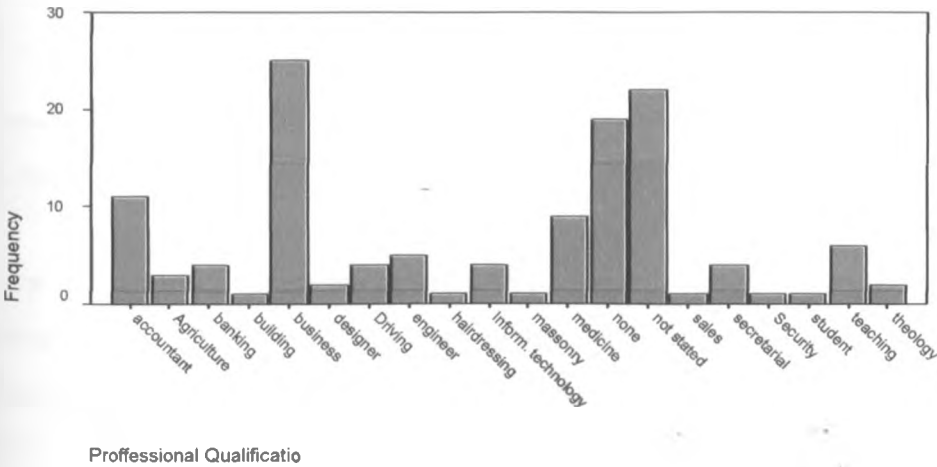
Table 4.2: Mode of Tenure

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Owner	51	40.5	40.5	40.5
	Rental	65	51.6	51.6	92.1
	Family property	6	4.8	4.8	96.8
	Relative's property	3	2.4	2.4	99.2
	Other	1	.8	.8	100.0
	Total	126	100.0	100.0	

Source: Field data 2004

This is an increasing evidence of the fact that the township acts as a lay off for excessive population from Nairobi city. Looking at occupation, most of the residents are workers in formal employment (35.7%), Business employers (14.3%), Business employee (8.7%), and casual labour construction (7.9%), others remain significantly low. Again looking at professional qualification, most residents are qualified in high professions, e.g. banking, accounting, and engineering among others who have sought accommodation in this area (See figure 9 below). Residents in these professions have been encouraged to settle due to availability of land, or have bought houses through mortgage schemes.

Figure 9: Professional Qualification



Source: Field data 2004

A good indicator of evidence of increased residential settlement is the number of years of residence; we can easily tell that most residents are migrants who settled in their present area in the last 12 years. Table 4.3 shows that at least 58.7% of the residents have settled in the study area in the last 6 years, and another 17.5 indicated that they settled in the area in the last 7-12 years.

Table 4.3: Years of residence in the study area

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-6yrs	74	58.7	60.7	60.7
	7-12yrs	22	17.5	18.0	78.7
	13-18yrs	13	10.3	10.7	89.3
	19-24yrs	8	6.3	6.6	95.9
	25-30yrs	5	4.0	4.1	100.0
	Total	122	96.8	100.0	
Missing	System	4	3.2		
Total		126	100.0		

Source: Field data 2004

The dominant form of land tenure was residential 65 (51.6%), against owner occupier 51 (40.5%) other forms of land tenure remain significantly low (See table 4.4).

Table 4.4: Mode of Land Tenure

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Owner	51	40.5	40.5	40.5
	Rental	65	51.6	51.6	92.1
	Family property	6	4.8	4.8	96.8
	Relative's property	3	2.4	2.4	99.2
	Other	1	.8	.8	100.0
	Total	126	100.0	100.0	

Source: Field data 2004

This may not tell much about the dynamics of resident population in the study area. A cross tabulation of age of household head against years of residence indicated that most residents between ages 17 to 46 had settled in the study area less than six years ago. See table 4.5

Table 4.5: Cross tabulation of Age of HH head against Years of residence category

		Years of residence category					Total count
		1-6yrs	7-12yrs	13-18yrs	19-24yrs	25-30yrs	
Age of HH Head (Category)	17-26yrs	21	2			1	24
	27-36yrs	28	5	2		1	36
	37-46yrs	21	9	6	2	1	39
	47-56yrs	3	5	3	3	1	15
	57-66yrs	1	1	2	3	1	8
Total		74	22	13	8	5	122

Source: Field data 2004

Further, cross tabulation of age of household head against mode of land tenure shows that most of the residents within the age group of 27-36, 26(40%) were tenants and 9 (17.6%) were home owners. For age group 37-46, only 18 (27.7%) were tenants and 21 (41.2%) were home owners. Most of the residents above age 47 were home owners, this means that people tend to settle down and acquire permanent residence as they grow old (See table 4.6)

Table 4.6: Cross tabulation of Age of HH head (Category) against land Tenure

		Tenure					Total count
		Owner	Rental	Family Property	Relatives Property	Other	
Age of HH head (Category)	17-26yrs	2	19	2	1	1	25
	27-36yrs	9	26	1	1		37
	37-46yrs	21	18	2			41
	47-56yrs	12	2		1		15
	57-66yrs	7		1			8
Total		51	65	6	3	1	126

Source: Field data 2004

Lastly findings on the dominant mode of tenure within different land use areas indicate that; Most individual owned residences were found within the agricultural designated land use side 32 (62.7%) as opposed to commercial land use (37.3%). Similarly there were more rental residential units 39 (60%) as opposed to 26(40%) on the commercial designated land use side (see table 4.7)

Table 4.7: Cross tabulation of Mode of Tenure against land use Designation

		Land use designation		Total
		Commercial	Agricultural	
Tenure	Owner	19	32	51
	Rental	26	39	65
	Family property	3	3	6
	Relative's property	2	1	3
	Other		1	1
Total		50	76	126

Source: Field data 2004

Most Home owners were persons living in the Agricultural land use designated areas, on the other hand homeowners on the commercial designated land use area are beneficiaries of trust land bestowed to them by the local county council under trust land ordinance. The increase of residential units within agricultural land use area is a result of the commercial influence where in addition to individual homes, the owners have constructed rental houses on their land and where they sublet as an extra source of income (field notes 2004). According to Olkejuado county council there is an increase in commercialization of the agricultural land use area against the sanctioning of the local Authority.

Objective 2

Examine the Extent to Which Urban Residential Settlements Contribute To Soil Compaction and Sedimentation in the Study Area, and Compare Resource Use between Different Land Use Areas

There were two key elements for this objective i.e. levels of compaction and extent of sediment generation, Extent of compaction looked at extent of surface coverage, levels of laying of concrete, gravel and stones, and bare soils while sedimentation looked at excavation within residential areas, these were supported by observation and photography.

The extent of surface coverage was determined by looking at the first the presence of structures in a residential nit, secondly the proximity of building, which measured on an ordinal scale.

Table 4.8 represents information on presence of surface covers within residential units by counts, percentage of responses and percentage of cases. Residential houses were a common occurrence in all cases 98.4%, and accounted for 62.9% of total responses. Other surface covers were ground covered with stones or gravel (20.6%) of cases, and poultry (10.3%) of the cases. A positive aspect in the residential units were compounds with exotic tree species and grass lawns accounting for 8.7% and 6.3^ of the cases. There was also concrete cover (5.6%) cultivated land area (4%) though very small but may be said to have significant contribution to soil compaction.

Table 4.8: Enumerate presence of the following

Dichotomy label	Count	Percentage of Responses	Pct of Cases
Residential houses	124	62.9	98.4
Animal stalls	3	1.5	2.4
Poultry	13	6.6	10.3
Concrete surfaces	7	3.6	5.6
Ground and stone covered surfaces	26	13.2	20.6
Exotic trees species	11	5.6	8.7
Cultivated land	5	2.5	4.0
Grass lawns	8	4.1	6.3
	-----	-----	-----
Total responses	197	100.0	156.3

0 missing cases; 126 valid cases

Source: Field data 2004

A cross tabulation of presence of surface covers in residential areas against land use designation indicated that there were more animal stalls, poultry and concrete surfaces on the commercial side than the agricultural side. On the other hand, there were more of ground and stone covers, exotic tree cover, grass lawns and cultivated land on the agricultural land use side than the commercial land use side (See appendix 4).

Proximity of buildings was measured on an ordinal scale; buildings were classified as either sparse (ranging over 100m between buildings), moderately sparse (50-100m between buildings, and densely populated (20-30m between buildings. (See table 4.9)

Table 4.9: Proximity of houses/structures

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Scattered houses with open grounds >100m	32	25.4	25.6	25.6
	Moderately sparse with road networks, about 50-75m	44	34.9	35.2	60.8
	Densely populated, with Road networks, 20-30m	49	38.9	39.2	100.0
	Total	125	99.2	100.0	
Missing	System	1	.8		
Total		126	100.0		

Source: Field data 2004

The table reveals that at least 39.9% of the cases were classified as densely populated with road networks, another 34.9% were classified as moderately sparse while 25.4% were scattered houses i.e. in places experiencing new settlement. A cross tabulation of type of residential setting designated by land use type, against proximity of structures indicated that the residential area designated as agricultural had a modal score of each three categories i.e. Sparse (20), moderate (29) and densely populated (27), possibly indication the agricultural side was undergoing rapid differentiation. (See table 4.10). There was also a strong association between type of residential area and proximity of structures or houses. This is shown by a high value of the Crammer's $V = 0.97$ (See table 4.11). Crammer's V is a symmetric measure of association and ranges from 0-1, where 1 indicated a perfect relationship, while 0 indicates independence of variables. It also gives the strength of relationship but does not indicate the direction of influence.

Table4.10: Cross tabulation of Proximity of structures Against Type of residential unit

		Type of residential unit type		Total
		Commercial	Agricultural	
What is the proximity? of houses/structures	Scattered houses with open grounds >100m	12	20	32
	Moderately sparse with road networks, about 50-75m	15	29	44
	Densely populated, with road networks,20-30m	22	27	49
Total		49	76	125

Table 4.11: Crammer's V

		Value	Approx. Sig.
Nominal by	Phi	.097	.552
Nominal	Cramer's V	.097	.552
N of Valid Cases		125	

a. Not assuming the null hypothesis.

Source: Field data 2004

Another measure of compaction was the kind of soil cover, in general, grass was significantly present in residential units as seen from table 4.12, Grass in its diverse form; scanty, grass only or grass and selected tree species accounted for a cumulative Percentage of 63.3%. Bare soils and open grounds accounted for 34.9%

Table 4.12: Kind of soil cover present

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Grass only	21	16.7	17.5	17.5
	Grass and selected tree species	42	33.3	35.0	52.5
	Scanty grass	13	10.3	10.8	63.3
	Bare Soil /open ground	44	34.9	36.7	100.0
	Total	120	95.2	100.0	
Missing	System	6	4.8		
Total		126	100.0		

Source: Field data 2004

Also laying of surfaces with concrete was significantly low; Table 4.13. 50% of residential areas had concrete laying of less than 25% while only 11.1% had over 75%,

and 23% had none at all. On the other hand, a cross tabulation of residential type against extent of concrete surface cover shows that concrete laying of less than 25% was a common occurrence on the agricultural side, although in general the commercial side appears to have more concrete surfaces in all categories. (See table 4.14)

Table 4.13: Estimate extent of Concrete cover

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<25%	60	47.6	50.0	50.0
	25-50%	11	8.7	9.2	59.2
	50-75%	6	4.8	5.0	64.2
	>75%	14	11.1	11.7	75.8
	None	29	23.0	24.2	100.0
	Total	120	95.2	100.0	
Missing	System	6	4.8		
Total		126	100.0		

Source: Field data 2004

Table 4.14: Cross tabulation of Estimate extent of Concrete c ver Against Type of residential unit

		Type of residential unit type		Total
		Commercial	Agricultural	
Estimate extent of Concrete cover	<25%	25	35	60
	25-50%	3	8	11
	50-75%	5	1	6
	>75%	7	7	14
	None	9	20	29
	Total	49	71	120

Source: Field data 2004

Table 4.15 shows extent of bare soil cover, there were no bare grounds in 12.2% of the cases, while soils less than 25% was present in 44.7% of the cases. The greatest percent of bare soils (over75%) was present in 24.4% of the cases. Generally bare soils between 25-75% accounted for a cumulative percentage of 43.1%. Bare solids are a measure of the extent and speed with which grounds reach saturation levels and generate surface flows and mud. Good soil covers reduces the rate at which grounds reach saturation levels, reducing erosion and mud, and enhance infiltration into the-ground and which recharge underground water storage.

Table 4.15: Estimate extent of bare soil cover

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<25%	55	43.7	44.7	44.7
	25-50%	8	6.3	6.5	51.2
	50-75%	15	11.9	12.2	63.4
	>75%	30	23.8	24.4	87.8
	None	15	11.9	12.2	100.0
	Total	123	97.6	100.0	
Missing	System	3	2.4		
Total		126	100.0		

Source: Field data 2004

To measure sediment generation, residents were asked to state what the most common mode of construction process entailed, as it emerged, human labour was common 90.3% of the cases while machine use was not popular and accounted for 7.3% only. (See table 4.16)

Table 4.16: How soils removed during construction

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Use of Machine	9	7.1	7.3	7.3
	Use of Human labour	112	88.9	90.3	97.6
	Others (Specify)	3	2.4	2.4	100.0
	Total	124	98.4	100.0	
Missing	System	2	1.6		
Total		126	100.0		

Source: Field data 2004

Upon removal, soils were disposed away from site such as on road sides, or open fields 52.9% of the cases and used for landscaping and leveling grounds 40.3% of the cases. Soil was left at construction site only 5% of the cases (see table 4.17). The study was also interested in how common these occurrences were, table 4.18 gives a summary of the findings. A further cross tabulation of residential type against soil removal indicated that residential settings within the agricultural land use area had a modal score in all the three categories i.e. soil was used for landscaping and leveling compounds, deposited away from site and washed away gradually by rain, representing in 30, 39 and 2 cases respectively or 62.5%, 61.9% and 100% of the cases respectively. (See table 4.19)

Table 4.17: What happens to soil after removal?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Used for ground leveling and landscaping	48	38.1	40.3	40.3
	Deposited off away from site e.g. roadside, open ground	63	50.0	52.9	93.3
	Left at construction site	6	4.8	5.0	98.3
	Washed away gradually by rain	2	1.6	1.7	100.0
	Total	119	94.4	100.0	
Missing	System	7	5.6		
Total		126	100.0		

Source: Field data 2004

Table 4.18: How rampant is the practice of soil removal from site

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very Rampant almost every construction site	11	8.7	9.4	9.4
	Common	63	50.0	53.8	63.2
	Just a few cases	32	25.4	27.4	90.6
	We rarely see these features	11	8.7	9.4	100.0
	Total	117	92.9	100.0	
Missing	System	9	7.1		
Total		126	100.0		

Source: Field data 2004

Table 4.19: Cross tabulation of what happens to soil to removal * Type c residential unit

		Type of residential unit type		Total
		Commercial	Agricultural	
What happens to soil after removal	Used for ground leveling and landscaping	18	30	48
	Deposited off away from site e.g. roadside, open ground	24	39	63
	Left at construction site	5	1	6
	Washed away gradually by rain		2	2
Total		47	72	119

Source: Field data 2004 (* denotes against)

Objective 3

To Determine sources of water for domestic use in residential units in the study area, its availability, sustainability and compare use in different land use areas

To determine this objective, several questions were posed to residents, they were required to state; their dominant sources of water, alternative source of water, distance to water sources, water supply and demand dynamics and whether there were any challenges facing water sources.

The study presents the following findings;

Dominant source of water: Table 4.20 shows different sources of water and their importance to residents, as indicated, boreholes ranked highest among the sources of water 42.3%, this was followed by piped water with 31.7% of responses, while water vendors came third with 15.5% of the responses.

Table 4.20: Main source of water

Dichotomy label	Count	Pct.of. Responses	Pct.of Cases
Well	1	.7	.8
River	6	4.2	4.8
Bore hole	60	42.3	47.6
Rain Water	8	5.6	6.3
Vendors	22	15.5	17.5
Piped Water	45	31.7	35.7

Total responses	142	100.0	112.7

0 missing cases; 126 valid cases

Source: Field data 2004

Appendix 4 presents a cross tabulation of main source of water against residential type. Four values are given in every square which represent; the actual count, row percentage, column percentage and percentage of the totals. Residentials within the agricultural land use designation had modal score in most categories. There were more people fetching water from rivers, again the potential of tapping rain water was high within the agricultural residential areas than the commercial residential areas.

Alternative source of water: Residents seemed to turn to a number of alternative sources in the face of water difficulties (See table 4.21)

Table 4.21: Alternative sources of water

Dichotomy label	Count	Pct of Responses	Pct of Cases
River	5	3.6	4.0
Borehole	25	18.0	19.8
Rain water	48	34.5	38.1
Vendors	40	28.8	31.7
Piped Water	14	10.1	11.1
N/A	7	5.0	5.6
Total responses	139	100.0	110.3
0 missing cases; 126 valid cases			
Source: Field data 2004			

The main alternative source of water was rain water (34.5% of responses, other alternative sources were vendors (28.8%) who either delivered water at door step or could be found at specific localities, bore holes (18%) and piped water (10%). Difference in preference depends on a person's main source of water. For example those who had piped schemes as main source of water turned to boreholes, vendors and rain water as an alternative source.

The study notes that although there are two major defined sources of water, in real sense the source is one; Borehole. The confusion arises sometimes because whatever considered piped water is a creation of some business entrepreneurs who have tapped borehole water for distribution in the locality and who generate income on the piped borehole water (Field notes) whereas, other borehole owners' retail from the point of sinking the borehole.

Distance to sources of water: Residents reported; main water sources were within door step 60%, however, for most of the responses main water source was less than 500m away as indicated by the cumulative percentage of 92.9%. The same applied for alternative sources of water (94.3%) Table 4.22 and 4.23 present these findings.

Table 4.22: Distance to the main Source of water

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Door step(0-20m)	60	47.6	47.6	47.6
	<200m	25	19.8	19.8	67.5
	200-500m	32	25.4	25.4	92.9
	>1km	6	4.8	4.8	97.6
	>2km	3	2.4	2.4	100.0
	Total	126	100.0	100.0	

Source: Field data 2004

Residents were equally asked to judge the quality of the water they receive along a scale of good, fair or poor, as indicated in table 4.24, most residents felt that the quality of water they received was fair (51.6%) as indicated by valid percentage, 37.1% indicated the water was good

Table 4.23: Distance to alternative source of water

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Door step(0-20m)	66	52.4	52.4	52.4
	<200m	20	15.9	15.9	68.3
	200-500m	14	11.1	11.1	79.4
	>1km	2	1.6	1.6	81.0
	>2km	4	3.2	3.2	84.1
	N/A	20	15.9	15.9	100.0
	Total	126	100.0	100.0	

Source: Field data 2004

Sources of water faced a few challenges which may be classified in issues of reliability (marked with single asterisk in tables), quality (double asterisk) and cost. (See tables 4.24 and 4.25). Cumulatively, cost ranked lowest, while reliability ranked highest for major sources of water. For alternative sources, quality ranked lowest while reliability ranked high consumer costs were placed second, i.e. Water is expensive/extra expenses (28%) and lack of water storage facilities (7%). Costs incurred was related to water vendors, who inflated the price of water in the event of shortfalls emanating from mechanical breakdowns or power failure at borehole installations (field notes 2004)

Table 4.24: Problems with major source (water)

Dichotomy label	Count	Pct of Responses	Pct of Cases
Water borne diseases**	6	4.1	4.8
Dirty water**	14	9.6	11.2
Hard water**	21	14.4	16.8
Mechanical Breakdown*	18	12.3	14.4
Water dry up/low waters*	23	15.8	18.4
Not reliable*	38	26.0	30.4
No Problem	15	10.3	12.0
Water is expensive/extra expenses	9	6.2	7.2
No Response	2	1.4	1.6
	-----	-----	-----
Total responses	146	100.0	116.8

1 missing cases; 125 valid cases

Source: Field data 2004**Table 4.25: Problems with alternative source (water)**

Dichotomy label	Count	Pct of Responses	Pct of Cases
Water borne diseases**	2	1.5	1.6
Dirty water**	11	8.3	8.7
Hard water**	7	5.3	5.6
Mechanical Breakdown*	4	3.0	3.2
Not reliable*	33	24.8	26.2
Water is expensive/extra expenses	28	21.1	22.2
No Problem	41	30.8	32.5
No Storage facility	7	5.3	5.6
	-----	-----	-----
Total responses	133	100.0	105.6

0 missing cases; 126 valid cases

Source: Field data 2004

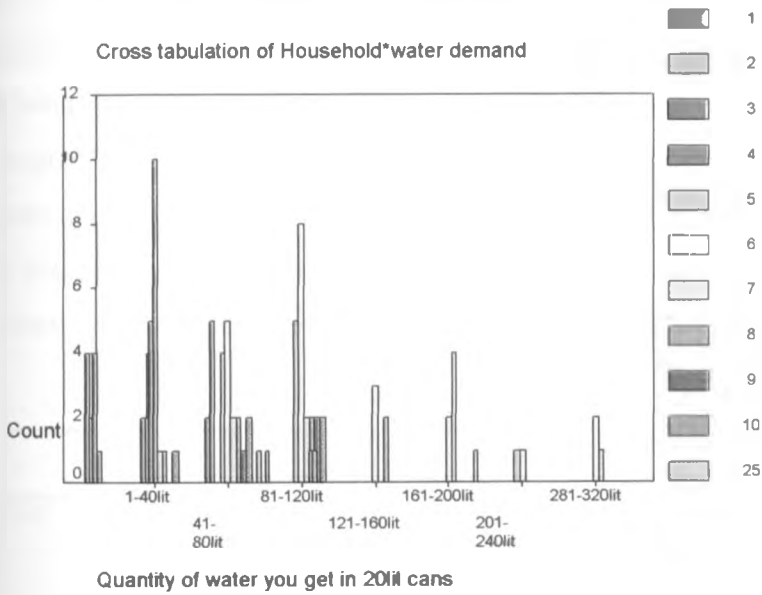
Demand: Demand is an important aspect of resource use as it affects the immediate availability and continuous availability of the resource. Demand is affected by among other factors family size and size of household. A good number of residents (81%) required less than 120 litres to meet their daily needs, with 40-80 litres being the modal category see table 4.26. There was strong relationship between household size and water requirements, in terms of additional water needs, it was easy to determine and predict households' water requirement with some degree of accuracy, hence easy to rule out exaggerated values. Figure 10 present a cross tabulation of water demands per day

Table 4.26: Quantity of water in 20lit cans a household's get in a day

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-40lit	25	19.8	21.6	21.6
	41-80lit	38	30.2	32.8	54.3
	81-120lit	31	24.6	26.7	81.0
	121-160lit	8	6.3	6.9	87.9
	161-200lit	9	7.1	7.8	95.7
	201-240lit	2	1.6	1.7	97.4
	281-320lit	3	2.4	2.6	100.0
	Total	116	92.1	100.0	
Missing	System	10	7.9		
Total		126	100.0		

Source: Field data 2004

Figure 10: A Cross Tabulation of Household size Against Water demand



Source: Field data 2004

Figure above indicates modal water demands per household size, Households sizes are given in coloured boxes to the extreme right y-axis, while x-axis represent amount of water in litres.

Affordability was another issue of demand and which acted to rationalize use. While most people were content that they needed more water than what they get presently, financial implications act as a restraint.

Sustainability

Sustainability is a critical issue and has to do with the continuous availability of the resource in terms of quality and quantity. A good number indicated that their current use was constrained by water availability. Most had demands running higher than current use; availability was limited by shortages, lack of regularity, and changing prices for the commodity. From the study, availability of water is limited by mechanical breakdown and power failure at borehole installations. It remains to be answered as to whether boreholes can adequately supply water if these problems are addressed.

Objective 4

Establish Extent to Which Residential Units Contribute To Vegetative Reduction and Compare Use in Different Land Use Areas

This objective looked at the different ways residential units are beneficiaries of vegetation services and related by products. While it was acknowledged that vegetation had been cleared to pave way for construction, the study looked at how well residents had attempted vegetation conservation measures. Residential units were assessed for the percentages of vegetative cover and its diversity, type of soil coverage, and sources of fuel energy. Questions were also posed to commercial dealers (in Fuel, timber and wood products) to comparatively rank their main clients. The results are discussed below

Table 4.27: Extent of Vegetative cover in residential areas

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<25%	38	30.2	30.6	30.6
	25-50%	23	18.3	18.5	49.2
	50-75%	25	19.8	20.2	69.4
	>75%	21	16.7	16.9	86.3
	None	17	13.5	13.7	100.0
	Total	124	98.4	100.0	
Missing	System	2	1.6		
Total		126	100.0		

Source: Field data 2004

Vegetation cover: Vegetation in whatever form was at least present in most residential units. Table 4.27 shows there was no vegetation in 13.7% of the responses, Scanty less

than 25% was present in 30.6% of the cases. Percentages above 25% were a good indicator of vegetation status i.e. 25-50%, 50-75%, above 75% and accounted for 18.5, 20.2, and 16.9 percentages respectively. Cross tabulations on vegetation cover against type of residential units indicated; Vegetation cover is well conserved within agricultural residential units as opposed to residential units on the commercial side (see table 4.28) Other evidences were derived from residential units layout or presence of surface structures, for instance table 4.29 indicated the importance vegetation is given in residential areas. Grass lawns and exotic tree species accounted for only 4.1 and 5.6 % respectively.

**Table 4.28: Cross tabulation extent of Vegetative cover (%)
Against Type of residential unit**

Count		Type of residential unit		Total
		Commercial	Agricultural	
Estimate	<25%	23	15	38
extent of	25-50%	8	15	23
Vegetative	50-75%	6	19	25
cover	>75%	5	16	21
	None	7	10	17
Total		49	75	124

Source: Field data 2004

Table 4.29: Compound layouts and features present

	Count	Pct of Responses	Pct of Cases
Dichotomy label			
Residential houses	124	62.9	98.4
Animal stalls	3	1.5	2.4
Poultry	13	6.6	10.3
Concrete surfaces	7	3.6	5.6
Ground and stone covered surfaces	26	13.2	20.6
Exotic trees species	11	5.6	8.7
Cultivated land	5	2.5	4.0
Grass lawns	8	4.1	6.3
Total responses		197	156.3

0 missing cases; 126 valid cases

Source: Field data 2004

Table 4.30: State of vegetation cover in residential areas

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very poor (scanty grass/shrubs with open ground)	26	20.6	21.5	21.5
	poor (dry scanty less vegetative grass)	45	35.7	37.2	58.7
	Fair (dry/and vegetative grassland)	37	29.4	30.6	89.3
	Good (moist and expansive grassland/shrub cover)	12	9.5	9.9	99.2
	Very bushy	1	.8	.8	100.0
	Total	121	96.0	100.0	
Missing	System	5	4.0		
Total		126	100.0		

Source: Field data 2004

State of vegetation in residential areas as seen in table 4.30, indicated that the state of vegetation was poor composed of scanty grass with open grounds, and dry scanty less vegetative grasses which cumulatively accounted for 58.7%. Dry and vegetative grassland, and moist and expansive grasses, were present in 30.6% and 9.9% respectively.

Products and services: Households were also asked on the kind of fencing materials they used and cooking energy. The following were the findings. Most homes 39.5% were fenced using live fencing (e.g. Euphorbia), poles were required in fencing; Barbed wire and wooden poles (28.4%), mesh wire and wooden poles (4.8%), and so were rafter (10.5%). This implies that services of vegetation derived products are in high demand, as seen from other alternative to vegetative derived products used; stone and concrete wall accounted for (15.3%). (See table 4.31)

Table 4.31: Kind of fencing material used

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Live/living fence e.g. Cypress	49	38.9	39.5	39.5
	Barbed wire and poles	35	27.8	28.2	67.7
	Mesh wire and poles	6	4.8	4.8	72.6
	Mesh wire and Plastic/metallic poles	1	.8	.8	73.4
	Stone walls	14	11.1	11.3	84.7
	Concrete walls	5	4.0	4.0	88.7
	Wooden rafters	13	10.3	10.5	99.2
	Others (Specify)	1	.8	.8	100.0
	Total	124	98.4	100.0	
Missing	System	2	1.6		
Total		126	100.0		

Source: Field data 2004

Discussions held by dealers in timber and timber products confirmed that reported residential units were their major clients for timber and wood related products, and contributed significantly to the growth of business. However, the place of such a conclusion on vegetative resource extraction may be narrow as this is negated by the influence of urban footprints where growth and influence of township trigger sourcing of materials from distant places.

As found out, most of the consumers of the products were residential units, resource mainly timber products were obtained from outside the district, mainly in Central province; Murang'a and Meru, Rift valley; Kericho, Nakuru and Uasin Gishu and Nyanza; Kisii, Nyamira. The three provinces Central, Rift Valley, and Nyanza accounted for 4%, 16% and 8% respectively.

Sources of cooking energy: Consumption of fuel indicated that only 4.7% of firewood (wood Fuel) was used by respondents. Another 24.4% accounted for charcoal only, Kerosene 35.8% only, cooking gas 29.5% only, and electricity 5.7% only. There was a rich diversification in energy fuel; Residents had a multiple sources (See table 4.32). A cross tabulation of sources of cooking energy by residential unit type indicated that residential units in agricultural land use area, dominated in all fuel categories, (See appendix 5). This possibly stems from the fact that it has a diverse population of people

in formal employment and elitist group; it is also within these areas that new settlements are rapidly emerging with land available for sale.

Table 4.32 Sources of fuel for residential units

Category label	Code	Count	Pct of Responses	Pct of Cases
Firewood	1	9	4.7	7.1
Charcoal	2	47	24.4	37.3
Kerosene	3	69	35.8	54.8
Gas	4	57	29.5	45.2
Electricity	5	11	5.7	8.7
		-----	-----	-----
Total responses		193	100.0	153.2

0 missing cases; 126 valid cases

Source: Field data 2004

The actual intensity of charcoal and wood usage may not be easily tied to vegetative degradation in the study area or division, an interview with dealers in charcoal and wood also indicated that charcoal and firewood were not from the division but the wider district of Kajiado. They are quick to add that only dead wood is collected from interior dry land areas with no significant economic activities. People living in these dry areas earned income out of this collection of dead wood, and waited in anticipation for would be buyers coming from in urban areas. Lorries are used to transport firewood and charcoal from these places, and is a thriving business. See figure 11



Figure 11: shows lorry on brief stop while carrying firewood on transit from interior dry lands destined to urban areas

To the dealers, collection of firewood is a natural way in which the vegetative resource shed off dead wood and rejuvenated younger wood. In areas where vegetation was actually cleared for wood and charcoal the rationale was to paved way for grass and fodder for animals.

Lastly, residents were tested on knowledge and they were required to state whether there was any relationship between their form of energy source and the environment. 34.4% chose not to respond to the question, 31% acknowledged their source of fuel encouraged tree felling, respondents using charcoal well understood its impact to the environment but said they had little options not to exploit the environment, they argued it was the cheapest option available in the markets. They urged the government to ensure that there were better and cheaper alternatives in the markets. Another 7.8% said it contributed to pollution i.e. use of kerosene contributed not only pollution through air but also plastic paper pollution used as packaging. (See table 4.33)

Table 4.33: how does fuel use lead to vegetation loss

Category label	Code	Count	Pct of Responses	Pct of Cases
None	1	12	9.3	9.5
No Opinion	2	45	34.9	35.7
Don't Know	3	18	14.0	14.3
Pollution	4	10	7.8	7.9
Tree felling	5	40	31.0	31.7
Conserve trees	6	3	2.3	2.4
Prevent soil erosion	7	1	.8	.8
Total responses		129	100.0	102.4

0 missing cases; 126 valid cases

Source: Field data 2004

Objective 5

Establish Ways and Extent to Which Residential Units Contribute To Land Resource Exploitation

The objective looked at construction material in residential units and their sources with an aim of deducing how this is likely to have effect on the environment. For homes, common construction materials for walls were stonewall and iron sheets accounting for 48.1%, block wall and iron sheet roof (25.2%, and iron sheet for wall and roof (20%). (See table 4.34)

Table 4.34: Building material for Homes

Dichotomy label	Count	Pct of Responses	Pct of Cases
Iron sheet wall, iron sheet roof	27	20.0	21.4
Rafter wall, iron sheet roof	9	6.7	7.1
Block wall, iron sheet roof	34	25.2	27.0
stonewall, tile/iron sheet roof	65	48.1	51.6
Total responses		135	100.0
0 missing cases; 126 valid cases			107.1

Source: Field data 2004

Sand and stone are therefore important housing materials either as plaster, joinery or reinforced concrete. (Field notes) Stones for construction were obtained from quarries (57.1%) one would either obtain them direct from the quarries by paying transporters to deliver without necessarily visiting the quarries, Table 4.35 gives a summary.

Table 4.35: Sources of stones for construction

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Quarries	72	57.1	57.1	57.1
	Transporters	28	22.2	22.2	79.4
	Buy from dealers	16	12.7	12.7	92.1
	N/A	10	7.9	7.9	100.0
	Total	126	100.0	100.0	

Source: Field data 2004

On the other hand, sand was obtained through transporters (50.8%. There were also dealers who delivered sand in bulk and later sold on retail scale according to individual needs (field notes). See also table 4.36

Table 4.36: Sources of Sand for construction

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Quarries	17	13.5	13.5	13.5
	Transporters	64	50.8	50.8	64.3
	Buy from dealers	35	27.8	27.8	92.1
	N/A	10	7.9	7.9	100.0
	Total	126	100.0	100.0	

Source: Field data 2004

It is important to add at this point also that quarries were run as enterprises and most were licensed. According to the local county council individuals were allowed licenses to mine their own land and not public access land and quarries such as Oloolua in Ngong. Owners therefore paid land rates to the county council. On the other hand, sand was obtained outside the division i.e. in Isenya Division of Kajiado; the county council therefore generated income through collection of gate rates at the point of sand collection (Field notes).

Research finding show; the residents are a group of people who are conscious of their environment and who own reasonable solutions to their problems. Some advocate for research and cultivation of more indigenous environmentally suitable tree species as opposed to exotic varieties that are prone to drought and theft. There is however need for a forum and bodies that will help realize and implement or guide communities to realize these solutions

Residents had suggestions on some conservation measures that ought to be put in place, for instance; revegetation 43.1%, prevent deforestation 15.3% and use alternative sources of energy 9.5%. (See table 4.37)

Table 4.37: Conservation measures on vegetation

Dichotomy label	Count	Pct of Responses	Pct of Cases
Revegetation of areas/awareness	59	43.1	46.8
Use alternative fuel	13	9.5	10.3
Nothing can be done	14	10.2	11.1
Stop burning fuel	7	5.1	5.6
Provide water for irrigation	6	4.4	4.8
Prevent deforestation	21	15.3	16.7
Don't Know	10	7.3	7.9
Encourage people to keep animals	3	2.2	2.4
Policy guidelines for construction	4	2.9	3.2
	-----	-----	-----
Total responses	137	100.0	108.7

0 missing cases; 126 valid cases

Source: Field data 2004

Immediate solutions on soil and land resources needed to address; reduction of soil erosion 9%, build gabions 8.3%. 46.9% wanted more vegetation grown to guard against soil loss, while 4.1% wanted drainage in the area improved to guard under ground water pollution. Some other 5.5% felt there should be regulations in place to regulate resource

use, excavation of quarries and corrective punitive measures imposed on environmental damages such as fines and costs of returning the environment to its initial state; this presupposes some Environmental Impact assessment and return to normal measures. (See table 4.38)

Table 4.38: Conservation measures on soil

	Count	Pct of Responses	Pct of Cases
Dichotomy label			
Control flooding	1	.7	.8
Control erosion	13	9.0	10.3
No Opinion	15	10.3	11.9
Don't Know	11	7.6	8.7
Plant hardy tress species and grass/soil	68	46.9	54.0
Build gabions	12	8.3	9.5
Build water reservoirs	2	1.4	1.6
Make good use of soil from construction	7	4.8	5.6
Improve drainage	6	4.1	4.8
Its is difficult due to occupation	2	1.4	1.6
Government to police land excavation	8	5.5	6.3
	-----	-----	-----
Total responses	145	100.0	115.1

0 missing cases; 126 valid cases

Source: Field data 2004

On water resource, table 4.39 below, gives a summary of residents suggestions,

Table 4.39: Conservation measures on water

	Count	Pct of Responses	Pct of Cases
Dichotomy label			
Drill more boreholes	12	8.4	9.5
No Opinion	15	10.5	11.9
Don't Know	7	4.9	5.6
Plant trees	16	11.2	12.7
Construct water reservoir and storages	26	18.2	20.6
Encourage proper use of water	26	18.2	20.6
Clean dirty water/treat	12	8.4	9.5
Improve drainage and supply	10	7.0	7.9
Policing and law by government	6	4.2	4.8
Get alternative source of water e.g. rain	13	9.1	10.3
	-----	-----	-----
Total responses	143	100.0	113.5

0 missing cases; 126 valid cases

Source: Field data 2004

4.2 DISCUSSIONS

Discussions in this section focus on resource availability in the study area, their potential, expert opinion, as well as views and deliberation of findings by researcher in relation to resource use and management practices

4.3 RESOURCES

Ongata Rongai Township runs parallel sandwiched between two locations (Ongata Rongai and Nkai-Morunya) opposite one another and stretching 4 Kilometers on either side of the Magadi road. The town ship has two distinct land use areas i.e. commercial area; having both business and residential areas, and agricultural land as delineated by the physical plan of Olkejuado County Council. However, due to increased settlement; and accompanied high demand for land particularly for commerce residential settlement, subdivision as well as let-outs and sale has taken place in the adjacent agricultural land. Occasionally, land has been converted to commercial use without prior consent of the local county council. This may not be reflected in the current physical development plans as these processes are so rapid.

4.3.1 Soils:

Soils in Ongata Rongai, Ngong, and Kiserian have a high potential for Agricultural production, although the proximity to Nairobi has put excessive pressure on these areas hence have been diverted to residential or commercial purposes.

Soil is characterized by black heavy clay soils in Ongata Rongai basically termed vertisols and greysols. According to the ministry of agriculture, these soils have approximately 40 – 60 % agricultural potential, potential can be further enhanced through use of fertilizer and irrigation. Most of this area is more suitable for open grazing and ranching especially around Ongata Rongai and bordering the Nairobi National Park. In the past, this area served as wildlife dispersal area and is part of the wider Kitengela conservation site. Today, the area has been converted to settlement due to population pressure.

However these soils are susceptible quick degradation; fertility loss through agriculture, this is a serious limitations. According to agricultural officers in the study area, soils can yield about 8 bags an acre of beans upon first harvest of conversion from

forest soil this quickly declines to 6 bags/acre in second season and may stabilize at 2 – 3 bags an acre with subsequent harvests (notes the divisional crop officer). Agricultural experts on the ground however note, these soils can do well under horticulture e.g. tomatoes, capsicum, onions French beans etc. Again drought resistant cultivars of maize crops taking 3 months e.g. hybrid 513, 511 and sunflower do well here.

The soils are shallow with poor drainage, owing to the presence of sub-surface rock close to the surface; soils are deep in valleys and low lying areas. Again, soils form as a result of weathering of the underlying rock, erosion and sedimentation. Many years of sedimentation has produced sedimentary rock otherwise termed false rocks that are mined in some areas and used in construction. Due to poor drainage soils are poorly formed with poor soil structure. These soils have poor compact ability, have quick water saturation and can flow with ease; sometimes the plasticity is so much that they easily collapse under weight. For these reasons, they are unsuitable for construction and holding buildings or surface structures. Soils have limited Economic importance especially where settlement is concerned. First, there is little if any farming utilities within the settlement as was revealed in our studies, significant source of family income come from some form of formal employment, and informal sector, which include construction business, transport, quarrying garages etc.

Secondly, soils within settlement of little economic and aesthetics importance, for example, because of their poor soil structure they easily form mud, cannot hold water for long to support grass and crack when dry, for these reasons these soils are dug out and disposed off, Where grass lawns are well kept, red loam soil are imported from Kiambu district and used to level compounds or grass lawns

Thirdly, by extension we could say this has created a new phenomenon where top-soils in settlements are seen as a burden. Normally soils are dug out at foundation levels and disposed off. There is however a limitation of disposal sites, hence most find their way into old quarries at best, or at worst, in open grounds, unoccupied plots, roadside, and river valleys, or are used to fill roads. Figure 12 and 13 represents different modes of

surface dumping of soils. And where this highly interferes with land, physical infrastructure development; roads, drainage and sewerage systems, etc



Figure 12: showing Open field and valley dumping



Figure 13: showing Roadside and site dumping

The study observes that these soils are often shallow owing to the presence of an underlying rock. These areas have also been said to be part of the continuous level low

lying areas adjacent to the high potential areas. The geography of this area is such that the soils have formed as a result of erosion in the high areas. These together with the poor drainage owing to the underlying rock are primarily responsible for creation of flood regimes within the settlements.

Residents cite poor drainage hence floods. This is compounded by poor and non-functioning sewer system, heavy reliance on underground water, which poses a threat of pollution to underground water sources.

The study argues, human activities in the study area has adverse impacts on the environment especially on soil and greatly jeopardizes the ability of the system to adequately provide environmental services. First, human activities involve massive earth excavation and movement not only in displaced services such as raw material needs like sand and stone but also disposal of topsoil. Although this occurs at micro-individual level, it has a cumulative effect of generating massive sediments and erosion that are easily worked on by rain especially when soil is discarded in open grounds. It further contributes to soil fouling and trampling. This is so as soil is disposed off.

Massive sediments generated have the ability to block and clog waterways and sewerages and drastically reduce water reservoirs holding capacity. The study cites a case where the current water supply system by the national water corporation project in Mbagathi River has not only been superseded by population growth but increasingly experience reduced holding capacity due to erosion from the upper side of Mbagathi river. In the long run, the corporation is forced to get water from Nairobi. Although the services of other water vendors has come in handy, sediment generation may also clog infrastructure where they find their way into water and sewage lines as has been reported by residents of the study area, especially where such infrastructure cross roads.

Settlements require proper planning especially drainage. It further notes that frequent settlement and activities within the homesteads and outside encourage large sediment

generation at its early stage while compaction especially landscaping and concretization at advanced stages. Outside the compounds, roads infrastructure compacts the ground See figure 14 showing early stages of soil compaction and sediment generation



Figure 14: showing early stages of compaction and sediment generation through laying of roads

through infrastructure road development in residential area. This often results in generation of large volume of storm waters, surface run –offs responsible flood regimes in sections of Ongata Rongai

4.3.2 WATER

Water is an important and most sought after resource in the study area. The study notes that the single vital input of water is rainfall, The study area has a bi- annual pattern of rainfall; the long rains come between March and June while short rains come between October and November. Dry seasons occur between July – October and again December – February. However in the last 6 years the pattern has been highly erratic (see figure 17). Annual rainfall averages range between 700-800mm, distributed between the long and the short rain periods i.e. 400-500 and 300mm respectively. Monthly average range is between 75mm- 125mm. The vast Ngong Hills form the Athi catchments, major catchments for the Athi River that flows into the Indian Ocean. There are two main

streams i.e. Mbagathi river and which originates from the Embakasi forest springs in Ngong escarpments and the Kiserian river which originates in the western escapements of Ngong catchments i.e. the University of Nairobi forest springs. There are numerous small streams such as Kandisi Embakasi, all of which feed the Athi River down in the plains.

Besides river sources, rainfall roof catchments are poorly harnessed. Statistics on this is however varied and largely unreliable. Another major source of water for study area is basically sub-surface (ground) water, mainly boreholes held by individuals (private owners), corporations and government.

Water provision in the study area basically falls into three major regimes i.e. Private owners, governments and its agencies and community. These are spread all over the division. The three water providers have harnessed spring waters, rivers or boreholes. For example, we have the Kiserian and Kataka piped water schemes run by the national Water Corporation; Oloosurutia, Kerarapon-Kibiko run by communities, and a number of borehole piped water supply by individuals in different neighborhoods in the three townships, as well as individual water vendors using either carts or mechanized

It is vital to note that despite these numerous services water is still a problem, this is partly attributed to population pressure that has outgrown the supply, irregular power supply, and low water level during dry spells. Also, poor regulation like where people drill boreholes without proper hydrological survey that has led to increased depth in sinking of existing wells, this could be attributed to lack of knowledge on the dynamics of ground water supply. In this latter case, according to sources at the local county council office situated in Ongata Rongai township, a number of people have dug private boreholes in their compounds in order to ensure independence in water use, this has sometimes disregarded the recommended distance between boreholes i.e. 200m between boreholes. Consequently, water tables have fallen lower over years while people still compete to sink deeper holes than the neighbours. There are only 40 registered boreholes in the township, still many others are yet unregistered as noted by an officer in-charge of water at the local county council office. While it has been indicated in publications that

water is obtained at a depth of 60-75m, (Olcoodo 2001) over the years this has changed and currently the trend is to drill boreholes to a depth of about 180-250m.

The quality of water is saline, there slight variation in degree of salinity basically determined by mineral composition of the basement rock, hence water quality vary from place to place. In some places, surface waters have reported high levels of salinity beyond consumption standards. Some streams such as Olololua are considered too salty and unsuitable for irrigation and animal use. Nevertheless, the mineral content such as iron and fluoride ions in most places is within acceptable standards. Residents complain of water colour, taste, odour and general hardness.

Sometimes water from the boreholes is said not to be enough. This has been taken advantage of by other dealers and transporters of water using Lorries and trailers. The problem is that under such circumstance it is difficult to authenticate the source of water a factor probably explaining constant complain of water borne diseases by residents.

The demand for water in Ongata Rongai township is best illustrated by the number of boreholes in the township. Water is a lucrative business in the township; dealers in water have down played the role played by the government and its agencies.

Other serious questions raised on water and settlements touch on planning. Settlements have numerous demands such as water, sewerage, and solid waste disposal. Since there is a general lack of a central and comprehensive sewerage system, open canals are sometimes used to drain waste water from urban and residential areas into streams. Again the absence of a central sewerage system means most home owners have own disposal system in the form of septic tank and a soak pit. These drain off water and liquids within the compound by directing this deep into the ground therefore affording ease of draining sludge using exhaust services. The prevalence of septic tanks and soak pits pose a threat of contamination to underground water sources through percolation into ground storages. Contamination is prone to occur where such installation and within 75 m of the borehole or fresh water source.

The demand for water in Ongata Rongai in relation to its availability inadvertently suggests that water is indeed a serious problem like in all other arid and semi-arid regions. We could say the area experiences water scarcity and unless measures are taken in future, it has already surpassed its supply hence not sustainable in the long-run. It is common scenario to see people move around with empty water jars or queuing about water points, situation residents' complain consumes a lot of their time.

Worse still we note that some of the already settled places form the high potential regions of the general semi arid region of Kajiado district and which are an extension of the continuous Athi catchments. Resource specialists have decried wanton destruction of vegetation, which act as rain water soak and gradually releasing waters into rivers thereby giving them long reticulation periods. Urban settlements have significantly contributed to this, while agriculture may provide temporal vegetation covers, urban setting on the other hand if not well managed totally replaces vegetation with concrete and tar surfaces. Figure 15 shows colonization of new lands by urban settlements



Figure 15: shows colonization of new lands by urban settlements

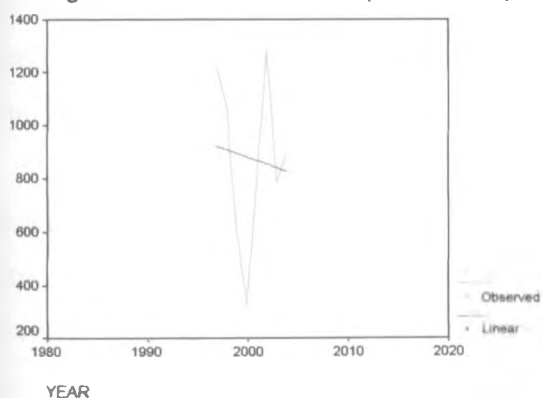


Figure 16: showing vegetation clearing and opening up of new lands for settlements

It is no coincidence that rainfall has been erratic in the last 6-10 years, yet same period coincides with period of increased outside settlement in the region since mid 1990s.

Figure 17 shows fluctuation of rainfall for the past 8 years (1997 -2004)

Figure17: Rainfall fluctuation (1997 - 2004)



Source: Field data

Today most of the existing dry river beds only points to evidence of a once existing rich vegetation that regulated river flow/reticulation and one that might no longer be there. Rivers experience flash action during rainfall with all the water running down the streams. River Kiserian for example is over harnessed as a result of agricultural practices upstream. On the other hand rivers Kandisi only experience seasonal flow while

Naromoru never discharges any water even when rainy. It instead provides an ideal valley used by Kiserian Township as a solid waste Disposal or dumpsite.

Recently the Minister for Environment and Natural Resources (MENR) Mr. Kalonzo Musyoka on Jan 24th 2005 decried wanton destruction of forests especially the Cherangani and Mau forest being the catchments for the rivers Turkwell & Nzoia as well as Sondu and Yala among others. The same might be happening or might have happened in the study area although currently the government has outlawed logging, charcoal burning and other activities in the Ngong forest. There may not be much left in the Ngong forest since most of the catchments is composed of acacia tree species and Savanna grass land as the primary vegetation of catchments. There is a further threat from grazing and burning.

Dr. Dominic Walubengo, the director of forest action network while writing for a local news paper once observed, the corrective measures for these may not entail moving people out of the settlements, which is more political but rather rehabilitating forests and wetlands. People must as of necessity be stopped from busting swamps (wetlands) for example, ploughing within river valley (swamps) must be stopped on the other hand settlers may be required to plant at least 10% of their time with trees (Daily Nation Sunday (January 20, 2005).

In another observation made from a trader in timber and wood products who obtained his supply from homesteads in Nyamira district-Nyanza province, he noted, homestead owners could be encouraged to harvest mature trees, as a means of income generation while planting new seedlings for continuity. This practice maybe harnessed positively if the farmers and settlers are made to understand they could still exercise controlled cropping of trees with replacement and earn income while at the same time ensuring that the environment and its natural cycles are not interfered with (field notes)

Sustainability and sustainable use of resource presupposes inter-generational and intra generational equity. For instance, within a generation (intra generation equity)

presupposes sufficient utilization of resource without compromising the ability of the next person /community to fully utilize the same or derive livelihood from the same resource. Thanks to the new awareness and the new sets of environmental policies under the stewardships of the National Environmental Management Authority (NEMA), the sessional paper no. 6 of 1999 on environment and natural resources and the Water Act 2002 under ministry of water resources are a realization of this.

The Water Act 2002 in particular recognizes intra-generational equity. It vests all waters of the land under the authority of the government. The government therefore regulates the harnessing of river waters upstream for beneficial use by use of permits. It reserves the right to issue and cancel such permits of abstraction as deemed necessary. Hence criminalises any abstraction of water source without permission.

4.3.3 Vegetation

All vegetation in the division serves as water catchments. Forests are a combination of acacia trees and thorny thickets undergrowth. Acacia trees grow to a height of 10m especially in river valleys although the height varies greatly from place to place. Vegetation varies with altitude and soils. Apart from the Acacia, the area is also characterized by the savanna grassland, which is well spread on the Ngong hills catchments and the lower eastern side of the division bordering the Nairobi National Park. In the past this formed part of the Kitengela conservation site, which served as wildlife dispersal area, increased settlements have resulted in further shrinking of these areas. Again, grasslands in the eastern side grass is of sturdy nature according to the agricultural office, it is more suitable for wildlife use with a wide variety of grazers (multi-stage grazers) and can be consumed by domestic animals at secondary level.

One of the greatest challenges as found out in the study is settlement. Settlements require land space and therefore perpetuate vegetation clearing. On the other hand, the incentive to replace vegetation is either lacking or is determined by many other social and economic factors. Even where finances may not be a handicap, concretisation is given preference to vegetation. This is the scenario in residential settlements and most residents

have either bare soil, grounds laid with stone chips or concrete and elaborate waterways especially in populous areas such as storied rental houses.

Homes recorded high percentages of bare soils, stone cover and small percentages in extent of grass cover. The implication is, there is a lot of storm water generated within dwelling units which would otherwise have infiltrated into the ground. Large volumes of storm water are directed out of the compounds into roadsides drainage systems. This eventually find their way into open fields, roads and water ways carrying with it large volume of sediment stones and boulders. Even where roads are covered with stone chips, the costs of maintenance are high since fresh soil through erosion is deposited onto the surfaces. Residents constantly complain of poorly maintained roads especially during rainy periods. The problem does not end here, as these waters are also responsible for erosion or the water margins of these settlements and urban areas.

The ramification maybe more adverse than meets the eye, for instance, one may attribute the situation to reduced periods of river flows within the study area, increased seasonality and water shortage problem to the above. One would argue there is insufficient infiltration of water into underground storage systems. Again, rivers experience flash actions and therefore drain waters as fast as they get them. A view of the rivers indicate dry riverbed but with adversely eroded riverbanks. One is left with the impression of once big rivers that are dead.

The study argues, an aggressive afforestation and greening campaign that would require every settler to plant at least 10% of their compound with trees and grass, this would increase rainfall infiltration into the ground and underground storage thereby reducing surface run offs, gradual release into rivers and therefore long life of river reticulation, to reduce aridity and sustaining livelihoods downstream.

Today the strengths of Kenya's forest policy especially in the arid and semi arid areas lies in monitoring vegetation cover changes in addition to establishing forest estates using improved germ-plasm (GOK 1999) . It notes that a number of people rely in forest for

wood (fuel) charcoal and building. Current indigenous wood demand by sawmills by far outstrips the supply or available stock, threats to forest stem from poor management, subsidized prices of products and services, poor accounting and auditing.

The study notes that indeed, the demand for charcoal and fuel wood from ASALs is high. The uses of these resources exemplify both sustainable and unsustainable use relative to other existing land uses. An interview with commercial charcoal dealers in the study area revealed that use of charcoal and fuel wood is high, charcoal and fuel wood is collected from the most interior regions and out of the district. The division has run out of its supply except in protected areas and relies on other divisions such as Magadi and Isenya.

At the most elementary level of charcoal and wood fuel dealership are rural fuel wood (gatherers/ foragers). This group makes a living out of this activity and therefore pride in seeing interested transporters from urban areas. Most of these people argue their mode of collecting firewood is sustainable as only dead wood is collected. Others still argue that by reducing forest thickets pave way for palatable grass species for domestic animals. The exploitation of these resources however seem to have a strong economic drive, what is more worrying is the fact that such activity takes place in the most remote and arid region , where rejuvenation of new tree crop (natural) is difficult to envisage.

New initiative for arid and semi arid region lie in the strength of the government policy, the sessional paper No 6 of 1999, and in powerful natural resource planning strategies. The policy paper advocate for a reserve of forest and promote tree planting for the private forestry, commercial purpose, public amenity and wildlife. Already initiatives in private forestry for commercial purposes have taken root in other division of Kajiado such as Isenya.

This is a vital exercise in a bid of regulating climate change. We note that the advantages of the vegetation lie in holding soils through root action; wind breaking and reducing sand storms as well as cushioning against drastic climatic changes. It is also important to ensure the catchments within the Arid and semiarid areas, which basically lie on the

margins, are well protected in terms of their vegetation, soil and water so as adequately cater for the changing populations.

It is possible to achieve sustainability levels to some levels and avoid incidences of resource competition between communities such as those witnessed recently in Trans Mara, Narok - Bomet borders, and West Pokot /Trans Nzoia borders and Mai-Mahiu between January-March 2005. Sustainability levels could be realized through adequate forest rehabilitation and management, reforestation, agro forestry and encouraging "Eco forestry settlements". In Ecological settlements settlers ought to be encouraged or educated to adopt green areas in their respective residence especially in the margin of ASALS. Green areas can take forms of trees and vegetation in individual homes, tree lines and grass along roads and pathways to soak excess waters, parks etc.

Greening associations and neighbourhood associations could also be initiated to manage these. Such greening exercise should have aesthetic as well as economic incentive like controlled tree cropping for economic gains and overall, residential areas appraisal, beauty among others. Positive aspects of settlement seem too real in Ruai area of Nairobi where settlements are greening the areas and have increased the value of land in the area. This is a positive change. It goes a long way to improve the vegetative cover on the catchments and watersheds and work towards long period of river reticulation to populations downstream.

CHAPTER FIVE

5.0 SUMMARY OF FINDINGS, CONCLUSIONS and RECOMMENDATIONS

Settlements can be harnessed positively in the marginal areas alongside other livelihoods in the region. Such coexistence however needs proper planning. Proper planning and environmental management here should take the form of good adequate monitoring mechanism for activities in the ASALS. This entails planning for settlement where feasible but above all the ability to anticipate and take prompt action. For instance, although human behaviour is spontaneous, sometimes a priori planning may take the form of setting aside specific lands for specific needs, tailoring concerns and continuous appraisal and direction of vision and objectives set. Critical consideration ought to be borrowed from the interrelationship between environmental degradation, population and development.

Large population mass degrade the environment especially when they lack in vision or direction. When people cannot achieve the idealized environmental goals yet there are no stringent measures or motivation to follow set down guidelines, they generate their own environment where they can be able to work and perform fairly well although not as good as the ideal. Such alternatives represent short/and quick actualization of needs that have immediate gratification but with long-term environmental degradation effects. For example, where settlement set aside is out of reach for many, they settle for a lesser alternative such as settling on open spaces (designated as public or common resource) where they can easily shape and manipulate its dynamic to suit them, hence mushrooming of informal settlement or slums. In most cases, such selected sites are where impacts are least expected and often hazardous or fragile environments such as roadsides, river valleys and dumpsites etc and subject to easy environmental degradation.

Change of policy should take the form of ability to anticipate and direct change. This way policy will be responsive to needs, and will act to prevent environmental degradation

before it actually occurs. A good example is drawn from the beneficiaries of the 167,000 acres of forestland annexed in Mau forest February 2001, and where eviction of such people at this time has moral, legal and political ramifications.

The government should also attempt systematic integrated resource management strategies for resources especially in ASALs. Attention be given to region according to vulnerability and needs.

5.1 RECOMMENDATIONS

We make the following observations and recommendations on water, based on this study. Water is an essential resource in the ASALs both for rural as well as urban use. It is fundamental in determining the holding capacity of these lands and the direction of development.

- a) There is general limited supply of this vital resource in the division, although the available permanent surface flow in rivers is partially underutilized.
- b) A large percentage of water available as fossil water in Private and public boreholes as well as corporate boreholes, springs are significantly limited.
- c) There are too many boreholes sunk in close proximity especially in the urban areas a factor that renders high competition and frequent low water levels.
- d) Regulation levels for water resource remain significantly low. This is true in cases where people first drill boreholes before seeking clearance from the concerned. Hence prompting proximity of distances between boreholes.

Policy direction should therefore

- a) Work to increase the capacity of surface water harnessing especially for rivers and available natural springs in Ongata Rongai (Mbagathi River) Kiserian and Kerarapon- Kibiko spring.
- b) It is also important that the management of community water projects as well as county council water in pastured rural areas be strengthened through capacity building as well as in distribution. In addition,

- c) Promote awareness on the limitation of the resource hence enlightenment and educational awareness prompting conservation as well as friendly water resource consumption patterns and behaviour.
- d) Thorough and stringent regulation of water access, promote distribution, sharing and use of economies tools such as levies for private and entrepreneurship in resources to regulate its harassing.
- e) Consider use of market economy in regulating access and use of resource i.e. levies for heavy users (consumption)
- f) Stringent enforcement of hydrological surveys prior to drilling of boreholes

We also note and make the following recommendations on soil

Soil is a vital resource although a renewable resource its process of renewed is considerably slow.

- a) Soils are basically vertisols often forming large cracks in dry periods. Again these soils have poor drainage and lose fertility easily deterioration to low /excessive infertility levels). Such soils are of medium potential, willingly, their potential can be improved with increased use of fertilization and under irrigation
- b) On the other hand the soils are also subject to quick saturation by water. Because of their poor structure are subject to easy collapse, especially under weight e.g. building and produce cracks on buildings if not laid on a firm rock base. For this reason, soils are therefore unsuitable for settlement when permanent and tertiary buildings are constructed prompting massive excavation, transfer and disposal haphazardly. These create muddy roads, unsightly scenes and overburdens, although may be useful when used to fill old quarries.

Policy direction should take on diverse approaches: -

- a) Improve soil in agriculture by use of, manuring and addition fertilizer, irrigation and ensure vegetation cover. For example, horticulture and use of short-term and early maturity cultivores of crop such as maize and sunflower to ensure continuous soil cover under horticulture and irrigation. This may be distinct soil

cover of different time span. Agriculture generally replaces permanent vegetation cover such as grass/ shrub with more temporary cover; food crops.

- b) Encourage animal husbandry in certain areas especially in lower areas with low agricultural potential. For example deep inside the ASALs and around conservation sites and wildlife dispersal areas to eliminate settlements
- c) Encourage zoning i.e. agriculture settlement, a forestation etc. This may discourage settlement especially in areas covered by vertisols
- d) Impose levies for environmental soil excavation, dispersal charges and encourage soil conservation and rehabilitation of settled areas adversely affected.
- e) Encourage community awareness on conservation and options available to them in terms of soil conservation and encourage them to join such ventures as community group and membership in residence greening groups.
- f) Discourage cultivation on slopes as well as disposal of our soils in river valleys open fields and roadside.

We observe and make the following recommendations on vegetation

- a) Vegetation is vital in soil cover, reduces mud, holds soil and increases water infiltration thereby reducing surface run-offs.
- b) Vegetation is adversely affected by settlement where natural vegetation is removed for residence and other structures
- c) Serious vegetation tramping is present in settled areas (residence) in most bare soils are left behind hence very muddy in rainy seasons.
- d) At a wider scale, vegetation degradation takes the form of quest for fuel wood. A number of dealers note sarcastically that there has never been any fuel wood in Ngong division as it was exhausted several decades ago. Today most of

vegetative resource degradation occurs in neighbouring division in the wider Kajiado district.

- e) On the other hand a right measure of the actual quantity of fuel wood demand in terms of charcoal lacks in details as most urbanites have diversified fuel sources i.e. electricity, cooking gas, paraffin and charcoal.
- f) Most residential units derive construction material from dealers. However, most trader in timber and related products, timber and related products are derived from other districts of the wider republic such as Nyanza Province; (Kisii, Nyamira) Rift Valley; (Uasin-Gishu, Nakuru, Kericho, Bomet) Central province; (Murang'a) and Eastern province; (Meru).

Policy direction should therefore

- a) Address vegetation degradation at both micro level (within residential) and macro, large scale catchments. The study notes, these areas/region catchment's areas (bordering high potential areas of arid and semi arid Settlements have therefore put pressure on catchments interfering and reducing river reticulation.
- b) Policy should address aggressive reforestation measures conservation and protection of catchments among others.
- c) Vegetative conservation ought to recognize not only trees but preservation of grass species natural as well as artificial in residential (dwelling) units' i.e Ngong Hills are actual grass (rank grass) and whistling thorn, and river valleys with lines of isolated tree areas. Basement system rock produces much thicker growth of thorn trees and thickets of evergreen forest that occur in the valleys.
- d) Policy should also encourage /promote awareness and education on environment conservation. Again, policy and aided research ought to encourage vegetative productive, germplasm of tree species that will encourage both commercial-economic return and aesthetics.

- e) Sometimes, encouraging residents through resident membership organization to green their areas, in a combined economic and a improving the economic value or market value but also aesthetic, blending economy and leisure as a behavioral incentive. This maybe synonymous to Dr Dominic Walubengo's thesis of encouraging residents /settlers to maintain at least 10% tree cover of acquired residential and settled areas in a bid to rehabilitate and increase the forest covers.

It should however be noted that search for solution in the ASALs is more than a direct or simple cause-effect relationship and therefore calls for concerted efforts from diverse sectors and disciplines.

5.2 BIBLIOGRAPHY

Articles

Agatsiva J. Lumasia *Contribution of fuel wood harvest to land degradation in Kenya in*

Desertification Control Bulletin: A Bulletin of World Event in Control of Desertification, Restoration of Degraded lands and Reforestation, UNEP: No. 29 1996

Campbell J D & E. S, Mbugua (Unknown) Survey of land use Problems in Kajiado District; A Review of methodology. IDS working paper No.334. IDS: UON

Campbell D. J (1977) Land use competition at the Margins of Rangelands; A proposal for research in Kajiado District, IDS working paper No. 299, IDS: UON

Campbell J. David & Jean Palutikof (1978) Allocation of land Resources in Semi Arid Areas: A Simulation based on the East African Experience IDS working paper no. 262, IDS: UON

Campbell D.J (1978) Development or Decline; Resource, land use, and population growth in Kajiado District IDS working paper No. 352

Chidumayo. E. N (1983) *Urbanisation and Deforestation in Zambia*, in DESERTIFICATION CONTROL NO. 9 December 1983

Daniel Stiles *Linkages between Dry land Degradation and Migration. A Methodology in* **Desertification Control Bulletin**: A Bulletin of World Event in Control of Desertification, Restoration of Degraded lands and Reforestation, UNEP: No. 30 1997

Darkoh, M B K. *Land Degradation and Soil Conservation; A Research Agenda in* **Desertification Control Bulletin**: A Bulletin of World Event in Control of Desertification, Restoration of Degraded lands and Reforestation, UNEP: No. 22 1993

Gideon C. Mutiso *Managing Arid and Semi Arid Areas in Kenya in* **GAINING GROUNDS**: An institutional innovation in land use management in Kenya Amos Kiriro & Calestous Juma (Eds) 1991, Acts Press: Nairobi

Hop craft David *Endangered Ecosystem: Wildlife, Ranching and Research in* **Desertification Control Bulletin**: A Bulletin of World Event in Control of Desertification, Restoration of Degraded lands and Reforestation, UNEP: No. 31 1997

Michael Korir Koech *History of Environmental Management in Kenya in* **GAINING GROUNDS**: An institutional innovation in land use management in Kenya Amos Kiriro & Calestous Juma (Eds) 1991, Acts Press: Nairobi

Ratan Lal *Agriculture Sustainability and Soil Resilience in* **Desertification Control Bulletin**: A Bulletin of World Event in Control of Desertification, Restoration of Degraded lands and Reforestation, UNEP: No. 23 1993

Sian Sullivan; *The Impact of Livestock on Topographically Diverse Open Wood & Shrub land in Arid North Namibia*, source Global Ecology and Biogeography, Vol. 8 Issue 3-4 page 257-77, May 1999

SNV (2002)- NETHERLAND DEVELOPMENT ORGANIZATION *Kenya Draft Policy & strategy paper on community based natural resource management (CBNRM)* Jan 2002 (unpublished)

Thomas J. Stohlgren, Thomas N. Chase, Roger A., Pielke, Sr., Timothy G.F Kitel & Jill S. Baron. *Evidence That Local Land Use Practices Influence Regional Climate, Vegetation And Stream Flow Pattern In Adjacent Natural Areas*, source Global Change Biology Vol. 4 Issue 5 pg 495-June 1998

Publications

Central Bureau of Statistics (2001) 1999 POPULATION AND HOUSING CENSUS: Population Distribution by Administrative Areas & urban Centres. Vol. i GOK: Nairobi

Central Bureau of Statistics (2001) 1999 POPULATION AND HOUSING CENSUS: Socio-economic Profile of the Population Vol. ii GOK: Nairobi

Pratt David I & MD Gwynne (1977) RANGELAND MANAGEMENT AND ECOLOGY IN EAST AFRICA. New York; Robert E Kriegen Pub Co.

Publications

Dennis Herlocker Ed. (1999) RANGLELAND RESOURCE IN EASTERN AFRICA ; *Their Ecology & Development* .Nairobi : GTZ

D.N Parkes, I, H Barnley & S R Walker (1985) ARID ZONES SETTLEMENT IN AUSTRALIA: A Focus on Alice Springs; United Nation University; Tokyo

Gideon Golany (1978) PLANNING URBAN SITES IN ARID REGIONS: *The basic Consideration in Urban Planning for Arid Zones, Americas Experience & Direction*, Gideon Golany (1978) (cds) New York; John Wiley & Sons

Girardet Herbert (1992) THE GAIA ATLAS OF CITIES: *New Direction for Sustainable Urbanisation Living*. UK: Gaia Books Limited

James Johnson (1972) URBAN GEOGRAPHY: *An Introductory Analysis*, Canada: Pergamon Publishers

Jim Barry, Stanley Mcgreal, Bill Deddis, (1993) URBAN REGENERATION: *Property Investment and Development*. London: E & FN SPON

Joanne I. O Abbot and Katherine Homewood: A History of Change: *Causes of Miombo Woodland Decline in a Protected Area in Malawi*, source Journal of Applied Ecology Vol. 36 Issue 3 page 422-June 1999

Joran Fries (1988) THE FIGHT AGAINST DESERTIFICATION IN THE ASALS. Sweden. International Rural Development Agency

J.T. Al Bakri, J. C Taylor & T. R Brewer: *Monitoring Land Use Changes in the Budia Transition Zone in Jordan Using Aerial Photography & Satellite Imagery*, source Geographical Journal Vol. 167 issue 3 pg 248-Sept. 2001

Lands berg J. C.D. James, J. Machonochie, A.O. Nichols, J Stols and R. Tynon: *Scale Related Effects Of Grazing On Native Plant Communities In Arid Rangeland Region Of Australia*, source Journal of Applied Ecology, Vol. 39 Issue 3 pg 427- June 2002

_____ DISTRICT DEVELOPMENT PLAN-KAJIADO (1997-2001) GOK: Nairobi

_____ (1994) NATIONAL ENVIRONMENTAL ACTION PLAN (NEAP) REPORT. GOK: Nairobi

_____ (1999) Sessional Paper No 6 of 1999 on Environment and Development. Ministry of Environment and Natural Resources: Nairobi

Mbithi Phillip & Barnes Carolyn (1975) SPONTANEOUS SETTLEMENT PROBLEM IN KENYA. East African Literature bureau: Nairobi

Michael John Stephen Peel, marc Stallman: The Systematic Reconnaissance flight (SRF) as a tool in Assessing Ecological impacts of Rural Development programme in an Extensive Area of the Lowveld of South Africa Source, African Journal of Ecology Vol.37, issue 4 pg 449-56 Dec 1999

OECD (1988) THE SAHEL FACING THE FUTURE: Increasing Dependence or Structural Transformation

Paul Knox (1987) URBAN SOCIAL GEOGRAPHY New York: John Wiley & Sons Inc.

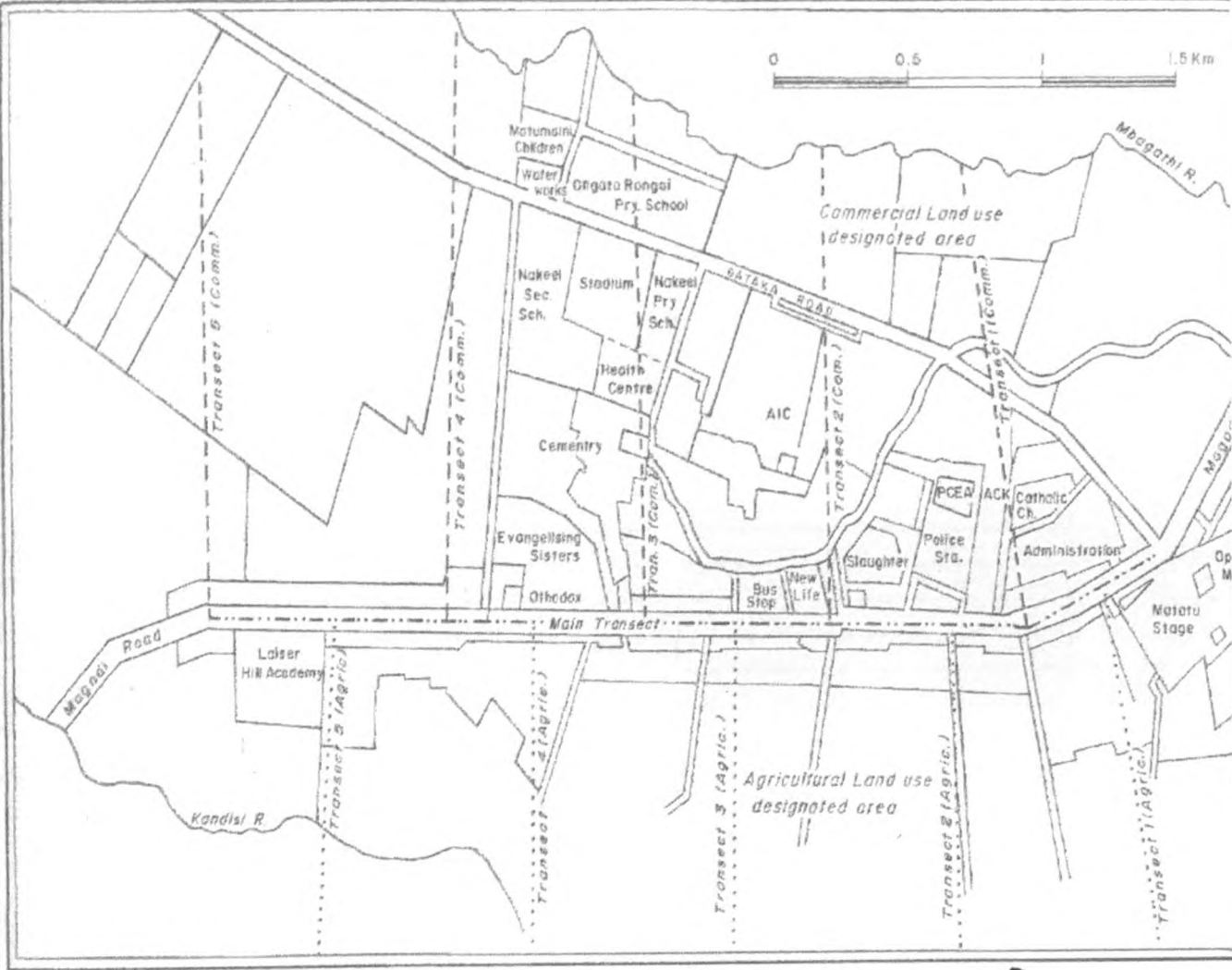
Piyush Tiwari (2001) URBAN ENVIRONMENT AND CITY GOVERNANCE IN INDIA IN INTERNATIONAL REVIEW FOR ENVIRONMENTAL STRATEGIES: Towards the Earth Summit, Vol. 2 No. 2 of 2002, Akoi Morishume et al (eds). Japan: IECS

Robin, P. White and Janet Nickoney (2003) DRY LANDS, PEOPLE & ECOSYSTEM GOODS & SERVICES: A Web Based Geospatial Analysis.

Thomas D.B Eds. (1997) SOIL AND WATER CONSERVATION: MANUAL FOR KENYA Ministry of Agriculture, Livestock Development and Marketing. Nairobi

5.3 APPENDICES

APPENDIX 1
Generation of Line Transects Within Ongata Rongai Township



Ongata Rongai Township: Development Plan

APPENDIX 2

sample Listing Sheet for residential Areas

The Influence of Urban Residential units in the Management of Selected dry Land Natural Resources: (Sample Listing Sheet for Residential Units)

Residential Unit Number	Main Road located on (Identify main road)	Street (Identify and name other streets to aid identification)	Landmark e.g. Junction of.... Opposite, adjacent to.... (Give exact position or relative to a feature e.g. Building name)
1. Example 1.	Kataka road	St Mary's catholic church close	Near over-head water tank, opposite storey house with black gate, simu ya jami at entrance.
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			
11.			
12.			
13.			
14.			
15.			
16.			
17.			
18.			
19.			
20.			
21.			
22.			
23.			
24.			
25.			
26.			
27.			
28.			
29.			
30.			
31.			
32.			
33.			
34.			
35.			

APPENDIX 3**The Influence of Urban Residential units in the Management of Selected dry Land Natural Resources: (Sample Calculation Sheet for Residential Units)**

Residential Totals	Sample Size (10 or 15)	Sample Interval	1st Randomly selected unit between (1-9)	Selected units

APPENDIX 4

Cross tabulation of Main source of water against Type of residential unit

	Count	Commerci Agricult		Row Total
	Row pct	al	ural	
	Col pct			
	Tab pct	1	2	
\$QBA1				
Well	1	0	1	1
		.0	100.0	.8
		.0	1.3	
		.0	.8	
River	3	1	5	6
		16.7	83.3	4.8
		2.0	6.6	
		.8	4.0	
Borehole	4	26	34	60
		43.3	56.7	47.6
		52.0	44.7	
		20.6	27.0	
Rain water	5	1	7	8
		12.5	87.5	6.3
		2.0	9.2	
		.8	5.6	
Vendors	6	6	16	22
		27.3	72.7	17.5
		12.0	21.1	
		4.8	12.7	
Piped water	7	19	26	45
		42.2	57.8	35.7
		38.0	34.2	
		15.1	20.6	
Column		50	76	126
Total		39.7	60.3	100.0

Percents and totals based on respondents

126 valid cases; 0 missing cases

APPENDIX 5

CROSS TABULATION: SOURCES OF COOKING ENERGY AGAINST TYPE OF RESIDENTIAL UNIT

	Count	Commerc. Agric.		Row Total
	Row pct			
	Col pct			
	Tab pct	1	2	
Firewood		4	5	9
		44.4	55.6	7.1
		8.0	6.6	
		3.2	4.0	
Charcoal		17	30	47
		36.2	63.8	37.3
		34.0	39.5	
		13.5	23.8	
Kerosene		31	38	69
		44.9	55.1	54.8
		62.0	50.0	
		24.6	30.2	
Gas		20	37	57
		35.1	64.9	45.2
		40.0	48.7	
		15.9	29.4	
Electricity		4	7	11
		36.4	63.6	8.7
		8.0	9.2	
		3.2	5.6	
Column		50	76	126
Total		39.7	60.3	100.0

Percents and totals based on respondents
126 valid cases; 0 missing cases

APPENDIX 6: QUESTIONNAIRE FOR RESIDENTIAL UNITS

THE INFLUENCE OF URBAN RESIDENTIAL SETTLEMENT ON THE MANAGEMENT OF SELECTED DRY – LAND NATURAL RESOURCES; A CASE OF NGONG DIVISON KAJIADO DISTRICT

Questionnaire for residential Units

Section A

Q001. District.....

Q002.Division.....

Q003 Township

Q004. Residential Unit Number

Q005. Name of Interviewer

Q006. Date of Interview

1.0 Socio-economic characteristics

1.1 Sex of respondent	Male	(01)
	Female	(02)

1.2 Age of Household head

1.3 Marital status	single	(01)
	Married	(02)
	Divorced	(03)
	Separated	(04)
	Widow	(05)
	Widowed	(06)

1.4. Family size (.....)

1.5. No. of Household members (...)

1.6. Literacy and Education	Attended school	(01)
	Never attended school	(02)

1.7. Highest of Education attained	Primary	(01)
	Secondary	(02)
	College	(03)
	University	(04)

1.8. Professional Qualification

- 1.9. Occupation
- Full-time formal employee (01)
 - Part-time formal employee (02)
 - Rural farm labourer (03)
 - Casual labourer (construction) (04)
 - Employer (Businessman) (05)
 - Business employee (06)
 - Worker on family business (07)
 - Retired (worker) (08)
 - Others (Specify) (09)
- 1.10. Religion
- Catholic (01)
 - Muslim (02)
 - Anglican (03)
 - Pentecostals (04)
 - Evangelical Movements (05)
 - Other (06)
 - None (99)
- 1.11. What is the mode of ownership/ tenure?
- Owner (01)
 - Rental (02)
 - Family farm/property (03)
 - Relatives (04)
 - Others (Specify) (05)
- 1.12. How long have you stayed in this place (No. of full years)
- 1.13a. what is the household's main source of income?
- Formal employment (01)
 - Informal Employment (02)
 - Family farm (03)
 - Farm animals (04)
 - Poultry (05)
 - Family enterprise (06)
 - Business (07)
 - Other (Specify) (08)
- 1.13b. what is household's alternative source of income?
- Formal employment
- Family farm (01)
 - Farm animals (02)
 - Poultry (03)
 - Family enterprise (04)
 - Business (05)
 - Other (Specify) (06)
- 1.14 What is the household's average monthly income in Kshs...?

SECTION B: Resource related

B1 Water

1.1 Where do you get your water from? (Main source)

Well	(01)
Spring	(02)
River	(03)
Borehole	(04)
Rain water	(05)
Buy from vendors	(06)
Piped water	(07)
Others (Specify)	(08)

1.2 What is your alternative source of income?

Spring	(01)
River	(02)
Borehole	(03)
Rain water	(04)
Buy from vendors	(05)
Piped water	(06)
Others (Specify)	(07)

B1.3. How far is your main source of water?

Door step (<20m)	(01)
<200m	(02)
200-500m	(03)
>1km	(04)
>2km	(05)
>5km	(06)

B1.4 How far is your alternative source of water?

Door step (<20m)	(01)
<200m	(02)
200-500m	(03)
>1km	(04)
>2km	(05)
>5km	(06)

B1.5. How suitable is the water you get for your domestic uses?

Good	(01)
Fair	(02)
Poor	(03)
Don't Know	(04)

B1.6. what problems do you experience from your:

- a) Main source of water?
- b) Alternative source of water?

B1.7 what quantity of water do you get for your needs (in 20litre jars)

B1.8. How much (many litres) do you need to sufficiently meet your needs?

B1.9 Do you have a borehole?

Yes	(01)
No	(02)

B1.10 If yes, How old is the borehole?

- B1.11 Do you experience any problems with your borehole?
- Yes (01)
- No (02)
- B1.12 what would you judge as the distance between one borehole to another in this area?
- <200m (01)
- 200-500m (02)
- >1km (03)
- >2km (4)
- >5km (5)
- B1.13 Do they provide sufficient water?
- Yes (01)
- No (02)
- B1.14a. is there any problem you know of boreholes around this area?
- Yes (01)
- No (02)
- B1.14b If yes, what is/are the problems facing boreholes in this area?
-
- B1.15 In your opinion, what are the major problems facing water sources around?
-

B2 LAND AND SOIL

Compaction and soil cover

- 1.1 Make an observation of the residential unit and enumerate presence of the following
- Residential units (01)
- Animal stalls (02)
- Poultry structure (03)
- Concrete surfaces (04)
- Ground and stone covered surfaces (05)
- Trees (06)
- Cultivated land (07)
- Grass lawns (08)
- 1.2 What kind of soil cover is present?
- Grass only (01)
- Grass and selected tree species (02)
- Scanty grass (03)
- Bare soils and open grounds (04)
- 1.3 Estimate the extent of
- a. Vegetative cover?
- < 25% (01)
- 25-50% (02)
- 50-70% (03)
- >75% (04)
- None (05)

- b. Bare soils?
- < 25% (01)
 - 25-50% (02)
 - 50-70% (03)
 - >75% (04)
 - None (05)
- c. Concrete/gravel surfaces
- < 25% (01)
 - 25-50% (02)
 - 50-70% (03)
 - >75% (04)
 - None (05)
- 1.4 What is the proximity of houses/structures in this residential setting?
- Scattered houses with open ground >100m (01)
 - Moderately sparse with road networks, about 50-75m (02)
 - Densely populated, with road networks, 20-30m (03)
- 1.5 Briefly describe the road network to yours residential area
- Earth road, dusty and occasionally muddy (01)
 - Earth roads with loose gravel surface (02)
 - Earth road with stone cover (03)
 - Earth road with loose stone chips (04)
- 1.6 What evidence of building type is present
- Iron sheet wall, iron sheet roof (01)
 - Rafter wall, iron sheet roof (02)
 - Brick wall, iron sheet roof (03)
 - Stone wall, tilled roof (04)
- 1.7 Where do you get your construction materials from?
- a. Stones
- Quarries (01)
 - Transporters (02)
 - Buy from dealers (03)
- b. Sand
- Quarries (01)
 - Transporters (02)
 - Buy from dealers (03)

c. Bricks	Quarries	(01)
	Transporters	(02)
	Buy from dealers	(03)
1.8	What does the most common mode of construction entail?	
	Heavy earth movement	(01)
	Light earth movement	(02)
	Use of building poles for structures	(03)
	Other (Specify)	(04)
1.9	How are soil removed during construction?	
	Use of Machine	(01)
	Use of human labour	(02)
	Other (Specify)	(03)
1.10a	what happens to soil after it has been removed?	
	Used for ground leveling and landscaping	(01)
	Deposited away from site	(02)
	Left at construction site	(03)
	Washed away gradually by rain	(04)
1.10b	how rampant/ widespread is this practice of disposal of soils from construction sites?	
	Very rampant (almost all sites)	(01)
	Common	(02)
	Just a few cases	(03)
	Rare	(04)
1.10c	Do you think this is a problem?	
.....		
1.10d	if yes, what do you think are some of the problems associated with this practice?	
.....		
1.11	Tick below the commonest fencing material used	
	Living fences e.g. Cypress	(01)
	Barbed wire and poles	(02)
	Mesh wire and poles	(03)
	Mesh wire and plastic/metallic poles	(04)
	Stone wall	(05)
	Wooden rafters	(06)
	Others (Specify)	(07)

B3 VEGETATION

1.1	How do you rate the quantity of vegetation cover?	
	Very poor (scanty grass with open ground)	(01)
	Poor (dry scanty less vegetation)	(02)
	Fair (dry and expansive grassland)	(03)
	Very bushy	(04)

- 1.2 What is the species diversity for vegetative cover within the residential unit?
- Very poor (scanty grass with open/bare ground) (01)
 - Poor (diverse natural grass species with seasonal growth) (02)
 - Fair (less grass and few selected exotic tree species) (03)
 - Good (rich and diverse grass and other tree species) (04)
- 1.3 How has people's behaviour affected vegetative cover within residential units?
- Don't Know (01)
 - Clearing for construction (02)
 - Clearing for farming (03)
 - Clearing for charcoal (04)
 - Planted more exotic tree species (05)
- 1.4 What kind of energy (fuel) do you use for cooking?
- Firewood (01)
 - Charcoal (02)
 - Kerosene (03)
 - Cooking gas (04)
 - Electricity (05)
- 1.5 What quantity of fuel (energy source) do you use?
-
- 1.6 How often do you get your energy supply?
- Weekly (01)
 - Daily (02)
 - Monthly (03)
 - 2-3 times a month (04)
- 1.7 Where do you get your fuel from?
- 1.8 How does your fuel source contribute to:
- a. Vegetative loss.....
 -
 - b. Vegetative conservation?.....
 -
 -
- 1.9 In your opinion, how can we help conserve the following:
- a. Soil?.....
 -
 - b. Water?.....
 -
 - c. Vegetation?.....
 -
 -
- 1.10 Any other comment?
-
-

Thank you for your cooperation

APPENDIX 7: ANALYSIS SUMMARY
Ongata Rongai = Frequency Tables/Cross Tabulation Charts
District

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Kajado	126	100.0	100.0	100.0

Division

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Ngong	126	100.0	100.0	100.0

Township

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Ongata Rongai	126	100.0	100.0	100.0

land use designation

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Commercial	50	39.7	39.7	39.7
Agricultural	76	60.3	60.3	100.0
Total	126	100.0	100.0	

Name of Interviewer

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Christopher	32	25.4	25.4	25.4
Sammy	32	25.4	25.4	50.8
Teresia	32	25.4	25.4	76.2
Eugene	30	23.8	23.8	100.0
Total	126	100.0	100.0	

Date

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Nov 12	51	40.5	40.5	40.5
Nov 13	18	14.3	14.3	54.8
Nov 15	40	31.7	31.7	86.5
Nov 16	17	13.5	13.5	100.0
Total	126	100.0	100.0	

Sex

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Male	78	61.9	61.9	61.9
Female	48	38.1	38.1	100.0
Total	126	100.0	100.0	

Age of HH head (Category)

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 17-26yrs	25	19.8	19.8	19.8
27-36yrs	37	29.4	29.4	49.2
37-46yrs	41	32.5	32.5	81.7
47-56yrs	15	11.9	11.9	93.7
57-66yrs	8	6.3	6.3	100.0
Total	126	100.0	100.0	

Marital Status

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid single	31	24.6	24.6	24.6
Married	92	73.0	73.0	97.6
Widow	2	1.6	1.6	99.2
Widowed	1	.8	.8	100.0
Total	126	100.0	100.0	

Family size

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	6	4.8	5.6	5.6
2	6	4.8	5.6	11.2
3	23	18.3	21.5	32.7
4	29	23.0	27.1	59.8
5	19	15.1	17.8	77.6
6	13	10.3	12.1	89.7
7	5	4.0	4.7	94.4
8	3	2.4	2.8	97.2
9	2	1.6	1.9	99.1
16	1	.8	.9	100.0
Total	107	84.9	100.0	
Missing System	19	15.1		
Total	126	100.0		

No. of HH members

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	7	5.6	6.1	6.1
2	6	4.8	5.3	11.4
3	12	9.5	10.5	21.9
4	19	15.1	16.7	38.6
5	15	11.9	13.2	51.8
6	25	19.8	21.9	73.7
7	13	10.3	11.4	85.1
8	8	6.3	7.0	92.1
9	3	2.4	2.6	94.7
10	5	4.0	4.4	99.1
25	1	.8	.9	100.0
Total	114	90.5	100.0	
Missing System	12	9.5		
Total	126	100.0		

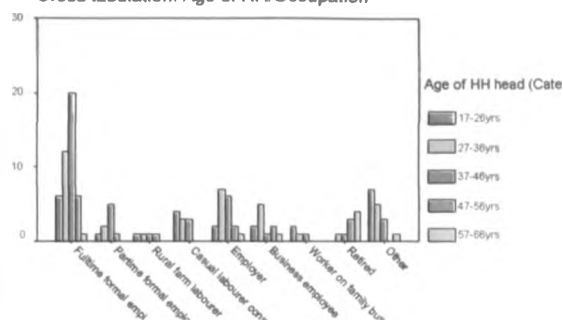
Literacy and Education

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Attended school	126	100.0	100.0	100.0

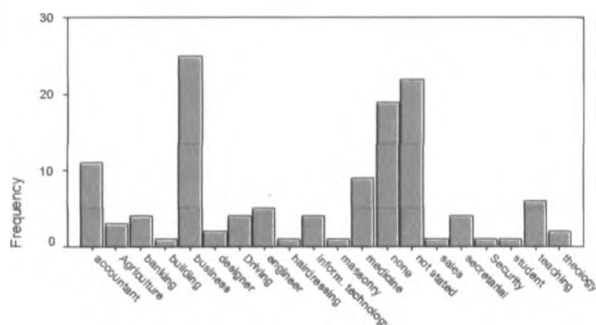
Highest level of Education

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Primary	21	16.7	16.7	16.7
Secondary	36	28.6	28.6	45.2
College	46	36.5	36.5	81.7
University	23	18.3	18.3	100.0
Total	126	100.0	100.0	

Cross tabulation: Age of HH/Occupation



Professional Qualification



Professional Qualification

Religion

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Catholic	38	30.2	30.2	30.2
	Muslim	4	3.2	3.2	33.3
	Anglican	17	13.5	13.5	46.8
	Pentecostals	42	33.3	33.3	80.2
	Evangelical movements	10	7.9	7.9	88.1
	Other	8	6.3	6.3	94.4
	None	7	5.6	5.6	100.0
	Total	126	100.0	100.0	

Occupation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fulltime formal employee	45	35.7	35.7	35.7
	Partime formal employee	9	7.1	7.1	42.9
	Rural farm labourer	4	3.2	3.2	46.0
	Casual labourer construction	10	7.9	7.9	54.0
	Employer	18	14.3	14.3	68.3
	Business employee	11	8.7	8.7	77.0
	Worker on family business	4	3.2	3.2	80.2
	Retired	9	7.1	7.1	87.3
	Other	16	12.7	12.7	100.0
	Total	126	100.0	100.0	

Occupation Others (Specify)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		1	.8	.8	.8
	home keeper	5	4.0	4.0	4.8
	missionary	1	.8	.8	5.6
	N/A	110	87.3	87.3	92.9
	self employed	3	2.4	2.4	95.2
	student	1	.8	.8	96.0

Tenure

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Owner	51	40.5	40.5	40.5
	Rental	65	51.6	51.6	92.1
	Family property	6	4.8	4.8	96.8
	Relatives property	3	2.4	2.4	99.2
	Other	1	.8	.8	100.0
	Total	126	100.0	100.0	

Years of residence category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-6yrs	74	58.7	60.7	60.7
	7-12yrs	22	17.5	18.0	78.7
	13-18yrs	13	10.3	10.7	89.3
	19-24yrs	8	6.3	6.6	95.9
	25-30yrs	5	4.0	4.1	100.0
	Total	122	96.8	100.0	
Missing	System	4	3.2		
Total		126	100.0		

main Source of HH Income

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Formal employment	52	41.3	41.3	41.3
	Informal employment	16	12.7	12.7	54.0
	Family farm	3	2.4	2.4	56.3
	Farm animals	2	1.6	1.6	57.9
	Poultry	2	1.6	1.6	59.5
	Family enterprise	13	10.3	10.3	69.8
	Busines	26	20.6	20.6	90.5
	Other	5	4.0	4.0	94.4
	Not stated	7	5.6	5.6	100.0
	Total	126	100.0	100.0	

HH Alternative source of income

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Family farm	4	3.2	3.2	3.2
	Farm animals	3	2.4	2.4	5.6
	Poultry	5	4.0	4.0	9.5
	Family enterprise	11	8.7	8.7	18.3
	Busines	40	31.7	31.7	50.0
	Other	28	22.2	22.2	72.2
	Not stated	35	27.8	27.8	100.0
	Total	126	100.0	100.0	

Others (Specify)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	N/A	98	77.8	77.8	77.8
	none	28	22.2	22.2	100.0
	Total	126	100.0	100.0	

Average monthly income (category)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0-10000	29	23.0	46.8	46.8
	11-20000	17	13.5	27.4	74.2
	21-30000	3	2.4	4.8	79.0
	31-40000	2	1.6	3.2	82.3
	41-50000	4	3.2	6.5	88.7
	51-60000	2	1.6	3.2	91.9
	71-80000	2	1.6	3.2	95.2
	81-90000	1	.8	1.6	96.8
	91-100000	2	1.6	3.2	100.0
	Total	62	49.2	100.0	
Missing	System	64	50.8		
Total		126	100.0		

Main source of water

Dichotomy label	Count	Pct.of. Responses	Pct.of. Cases
Well	1	.7	.8
River	6	4.2	4.8
Bore hole	60	42.3	47.6
Rain Water	8	5.6	6.3
Vendors	22	15.5	17.5
Piped Water	45	31.7	35.7
Total responses	142	100.0	112.7

0 missing cases; 126 valid cases

QBA2 Alternative source of water

Dichotomy label	Count	Pct of Responses	Pct of Cases
River	5	3.6	4.0
Borehole	25	18.0	19.8
Rain water	48	34.5	38.1
Vendors	40	28.8	31.7
Piped Water	14	10.1	11.1
N/A	7	5.0	5.6
Total responses	139	100.0	110.3

0 missing cases; 126 valid cases

Distance to the main Source of water

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Door step(0-20m)	60	47.6	47.6	47.6
	<200m	25	19.8	19.8	67.5
	200-500m	32	25.4	25.4	92.9
	>1km	6	4.8	4.8	97.6
	>2km	3	2.4	2.4	100.0
	Total	126	100.0	100.0	

Distance to alternative source of water

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Door step(0-20m)	66	52.4	52.4	52.4
	<200m	20	15.9	15.9	68.3
	200-500m	14	11.1	11.1	79.4
	>1km	2	1.6	1.6	81.0
	>2km	4	3.2	3.2	84.1
	N/A	20	15.9	15.9	100.0
	Total	126	100.0	100.0	

How suitable is the water you get for use

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Good	46	36.5	37.1	37.1
	Fair	64	50.8	51.6	88.7
	Poor	13	10.3	10.5	99.2
	Dont know	1	.8	.8	100.0
	Total	124	98.4	100.0	
Missing	System	2	1.6		
Total		126	100.0		

BA6A Problems with major source (water)

Dichotomy label	Count	Pct of Responses	Pct of Cases
Water borne diseases	6	4.1	4.8
Dirty water	14	9.6	11.2
Hard water	21	14.4	16.8
Mechanical Breakdown	18	12.3	14.4
Water dry up/low waters	23	15.8	18.4
Not reliable	38	26.0	30.4
No Problem	15	10.3	12.0
Water is expensive/extra expenses	9	6.2	7.2
No Response	2	1.4	1.6
Total responses	146	100.0	116.8

1 missing cases; 125 valid cases

QBA6B Problems with alternative source (water)

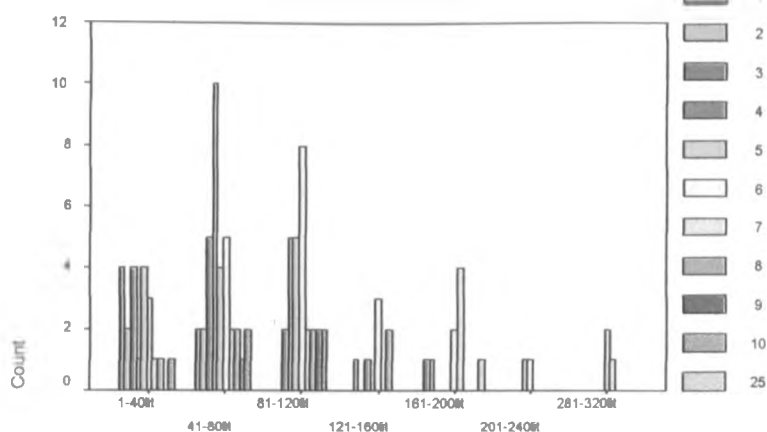
Dichotomy label	Count	Pct of Responses	Pct of Cases
Water borne diseases	2	1.5	1.6
Dirty water and raises issues of safety	11	8.3	8.7
Hard water	7	5.3	5.6
Mechanical Breakdown	4	3.0	3.2
Not reliable	33	24.8	26.2
Water is expensive/extra expenses	28	21.1	22.2
No Problem	41	30.8	32.5
No Storage facility	7	5.3	5.6
Total responses	133	100.0	105.6

0 missing cases; 126 valid cases

Quantity of water you get in 20lit cans

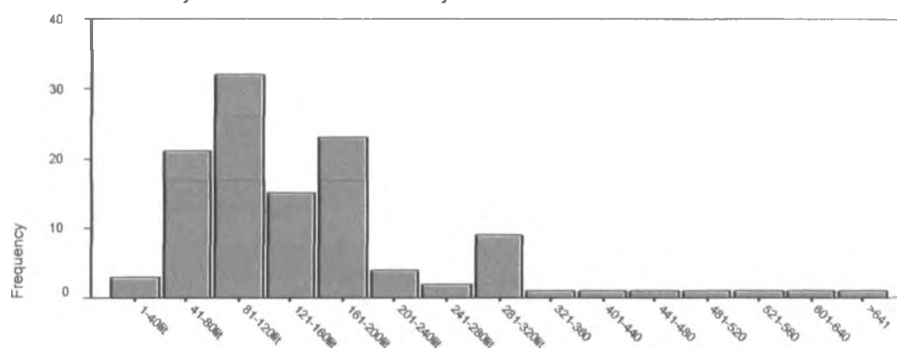
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-40lit	25	19.8	21.6	21.6
	41-80lit	38	30.2	32.8	54.3
	81-120lit	31	24.6	26.7	81.0
	121-160lit	8	6.3	6.9	87.9
	161-200lit	9	7.1	7.8	95.7
	201-240lit	2	1.6	1.7	97.4
	281-320lit	3	2.4	2.6	100.0
	Total	116	92.1	100.0	
Missing	System	10	7.9		
Total		126	100.0		

Cross tabulation of Quantity of water (l) against household size



Quantity of water you get in 20lit cans

How many lit would be sufficient for your use



How many lit would be sufficient for your use

Do you have a borehole

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Yes	5	4.0	4.0	4.0
No	121	96.0	96.0	100.0
Total	126	100.0	100.0	

How old is the borehole

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	1	.8	.8	96.8
2months	1	.8	.8	97.6
2yrs	1	.8	.8	98.4
5yrs	2	1.6	1.6	100.0
Total	126	100.0	100.0	

Do you experience any problems with your borehole

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Yes	2	1.6	1.6	1.6
No	3	2.4	2.4	4.0
N/A	121	96.0	96.0	100.0
Total	126	100.0	100.0	

What would you judge as the distance between boreholes in this area

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<200m	12	9.5	10.7	10.7
	200-500m	46	36.5	41.1	51.8
	>1km	38	30.2	33.9	85.7
	>2km	14	11.1	12.5	98.2
	>5km	2	1.6	1.8	100.0
	Total	112	88.9	100.0	
Missing	System	14	11.1		
Total		126	100.0		

Do they provide sufficient water

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	60	47.6	53.1	53.1
	No	53	42.1	46.9	100.0
	Total	113	89.7	100.0	
Missing	System	13	10.3		
Total		126	100.0		

Is there any problem you know of boreholes in this area

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	72	57.1	63.7	63.7
	No	41	32.5	36.3	100.0
	Total	113	89.7	100.0	
Missing	System	13	10.3		
Total		126	100.0		

QBA14B If yes! What are the Problems facing boreholes in this area

	Count	Pct of Responses	Pct of Cases
Dichotomy label			
Not reliable	53	39.8	42.1
Not Applicable	46	34.6	36.5
Dirty water	13	9.8	10.3
Low water level due to proximity	3	2.3	2.4
Salinity	7	5.3	5.6
Low water tables	8	6.0	6.3
Expensive	3	2.3	2.4

Total responses	133	100.0	105.6

QBA15 Opinion on problem facing water sources around

	Count	Pct of Responses	Pct of Cases
Dichotomy label			
Hard water	11	7.3	8.7
Dirty water	15	9.9	11.9
Mechanical breakdowns	13	8.6	10.3
Insufficient water	8	5.3	6.3
Irregular supply of water	20	13.2	15.9
Water is expensive	16	10.6	12.7
No Opinion	30	19.9	23.8
Drought	23	15.2	18.3
Poor water and sewerage infrastructure	15	9.9	11.9

Total responses	151	100.0	119.8

QBB1 Enumerate presence of the following

	Count	Pct of Responses	Pct of Cases
Dichotomy label			
Residential houses	124	62.9	98.4
Animal stalls	3	1.5	2.4
Poultry	13	6.6	10.3
Concrete surfaces	7	3.6	5.6
Ground and stone covered surfaces	26	13.2	20.6
Exotic trees species	11	5.6	8.7
Cultivated land	5	2.5	4.0
Grass lawns	8	4.1	6.3

Total responses	197	100.0	156.3

0 missing cases; 126 valid cases

Kind of soil cover present

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
Grass only	21	16.7	17.5	17.5
Grass and selected tree species	42	33.3	35.0	52.5
Scanty grass	13	10.3	10.8	63.3
Bare Soil /open ground	44	34.9	36.7	100.0
Total	120	95.2	100.0	
Missing				
System	6	4.8		
Total	126	100.0		

Estimate extent of Vegetative cover

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
<25%	38	30.2	30.6	30.6
25-50%	23	18.3	18.5	49.2
50-75%	25	19.8	20.2	69.4
>75%	21	16.7	16.9	86.3
None	17	13.5	13.7	100.0
Total	124	98.4	100.0	
Missing				
System	2	1.6		
Total	126	100.0		

**Estimate extent of Vegetative cover * Type of residential unit
Crosstabulation**

Count		Type of residential unit		Total
		Commercial	Agricultural	
Estimate	<25%	23	15	38
extent of	25-50%	8	15	23
Vegetative	50-75%	6	19	25
cover	>75%	5	16	21
	None	7	10	17
Total		49	75	124

Estimate extent of Bare soil cover

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<25%	55	43.7	44.7	44.7
	25-50%	8	6.3	6.5	51.2
	50-75%	15	11.9	12.2	63.4
	>75%	30	23.8	24.4	87.8
	None	15	11.9	12.2	100.0
	Total	123	97.6	100.0	
Missing	System	3	2.4		
Total		126	100.0		

Estimate extent of Concrete cover

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<25%	60	47.6	50.0	50.0
	25-50%	11	8.7	9.2	59.2
	50-75%	6	4.8	5.0	64.2
	>75%	14	11.1	11.7	75.8
	None	29	23.0	24.2	100.0
	Total	120	95.2	100.0	
Missing	System	6	4.8		
Total		126	100.0		

What is the proximity of houses/structures

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Scattered houses with open grounds >100m	32	25.4	25.6	25.6
	Moderately sparsed with road networks, about 50-75m	44	34.9	35.2	60.8
	Densely populated, with road networks, 20-30m	49	38.9	39.2	100.0
	Total	125	99.2	100.0	
Missing	System	1	.8		
Total		126	100.0		

Briefly describe road network in this area

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Earth roads, dusty and occassionally muddy	88	69.8	69.8	69.8
	Earth roads with loose gravel surface	12	9.5	9.5	79.4
	Earth roads with stone cover	20	15.9	15.9	95.2
	Earth road with loose stone chips	6	4.8	4.8	100.0
	Total	126	100.0	100.0	

QBB6 Building types present

Dichotomy label	Count	Pct of Responses	Pct of Cases
Iron sheet wall, iron sheet roof	27	20.0	21.4
Rafter wall, iron sheet roof	9	6.7	7.1
Brick wall, iron sheet roof	34	25.2	27.0
Stonewall, tile/iron sheet roof	65	48.1	51.6

Total responses 135 100.0 107.1
 0 missing cases; 126 valid cases

Where do you get your construction material(stones)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Quarries	72	57.1	57.1	57.1
	Transporters	28	22.2	22.2	79.4
	Buy from dealers	16	12.7	12.7	92.1
	N/A	10	7.9	7.9	100.0
	Total	126	100.0	100.0	

Where do you get your construction material(sand)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Quarries	17	13.5	13.5	13.5
	Transporters	64	50.8	50.8	64.3
	Buy from dealers	35	27.8	27.8	92.1
	N/A	10	7.9	7.9	100.0
	Total	126	100.0	100.0	

Where do you get your construction material(bricks)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Quarries	8	6.3	11.9	11.9
	Transporters	25	19.8	37.3	49.3
	Buy from dealers	34	27.0	50.7	100.0
	Total	67	53.2	100.0	
Missing	System	59	46.8		
Total		126	100.0		

What does the most common mode of construction entail?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Heavy earth movement e.g.machine excavation	61	48.4	51.7	51.7
	Light earth movement e.g. trenches	42	33.3	35.6	87.3
	Use of Building poles for structures	15	11.9	12.7	100.0
	Total	118	93.7	100.0	
Missing	System	8	6.3		
Total		126	100.0		

How soils removed during construction

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Use of Machine	9	7.1	7.3	7.3
	Use of Human labour	112	88.9	90.3	97.6
	Others (Specify)	3	2.4	2.4	100.0
	Total	124	98.4	100.0	
Missing	System	2	1.6		
Total		126	100.0		

What happens to soil after removal

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Used for ground levelling and landscaping	48	38.1	40.3	40.3
	Deposited off away from site e.g. roadside, open ground	63	50.0	52.9	93.3
	Left at construction site	6	4.8	5.0	98.3
	Washed away gradually by rain	2	1.6	1.7	100.0
	Total	119	94.4	100.0	
Missing	System	7	5.6		
Total		126	100.0		

How rampant is the practice of soil removal/disposal from site

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very Rampant almost every construction site	11	8.7	9.4	9.4
	Common	63	50.0	53.8	63.2
	Just a few cases	32	25.4	27.4	90.6
	We rarely see these features	11	8.7	9.4	100.0
	Total	117	92.9	100.0	
Missing	System	9	7.1		
Total		126	100.0		

BB10D If yes, is it a problem

Dichotomy label	Count	Pct of Responses	Pct of Cases
Erosion is rampant/ mud	34	24.5	27.0
Flooding	1	.7	.8
Not Applicable	80	57.6	63.5
Destroys beauty through overburdens	4	2.9	3.2
Block infrastructure/ hard to develop	16	11.5	12.7
Clear vegetations	2	1.4	1.6
No Problem	2	1.4	1.6
	-----	-----	-----
Total responses	139	100.0	110.3

0 missing cases; 126 valid cases

Tick bellow common fencing material present

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Live/living fence e.g. cypress	49	38.9	39.5	39.5
	Barbed wire and poles	35	27.8	28.2	67.7
	Mesh wire and poles	6	4.8	4.8	72.6
	Mesh wire and plastic/metallic poles	1	.8	.8	73.4
	Stone walls	14	11.1	11.3	84.7
	Concrete walls	5	4.0	4.0	88.7
	Wooden rafters	13	10.3	10.5	99.2
	Others (Specify)	1	.8	.8	100.0
	Total	124	98.4	100.0	
Missing	System	2	1.6		
Total		126	100.0		

How do you rate quantity of vegetation cover?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very poor (scanty grass/shrubs with open ground)	26	20.6	21.5	21.5
	poor (dry scanty less vegetative grass)	45	35.7	37.2	58.7
	Fair (dry/and exensie grassland shrub cover)	37	29.4	30.6	89.3
	Good (moist and expansive grassland/shrub cover)	12	9.5	9.9	99.2
	Very bushy	1	.8	.8	100.0
	Total	121	96.0	100.0	
Missing	System	5	4.0		
Total		126	100.0		

Rate the species diversity in residential unit

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very poor (scanty grass/shrubs with open ground)	24	19.0	20.0	20.0
	poor (divers grass with seasonal grass)	43	34.1	35.8	55.8
	Fair (few grass with few selected tree speicies)	45	35.7	37.5	93.3
	Good (rich and diverse grass and tree species)	8	6.3	6.7	100.0
	Total	120	95.2	100.0	
Missing	System	6	4.8		
Total		126	100.0		

How has peoples behaviour affected vegetation cover in this area?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Dont know	18	14.3	14.5	14.5
	Clearing for construction	88	69.8	71.0	85.5
	Clearing for farming	8	6.3	6.5	91.9
	Clearing for chacoal	4	3.2	3.2	95.2
	Planted more exotic trees	6	4.8	4.8	100.0
	Total	124	98.4	100.0	
Missing	System	2	1.6		
Total		126	100.0		

QBC4 Energy source for cooking

Dichotomy label	Count	Pct of Responses	Pct of Cases
Firewood	9	4.7	7.1
Charcoal	47	24.4	37.3
Kerosene	69	35.8	54.8
Gas	57	29.5	45.2
Electricity	11	5.7	8.7

Total responses	193	100.0	153.2

0 missing cases; 126 valid cases

How often do you get your energy supply?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Weekly	34	27.0	29.6	29.6
	Daily	32	25.4	27.8	57.4
	Monthly	40	31.7	34.8	92.2
	2-3 times a month	9	7.1	7.8	100.0
	Total	115	91.3	100.0	
Missing	System	11	8.7		
Total		126	100.0		

QBC8A how does your fuel use contribute to vegetation loss

Dichotomy label	Count	Pct of Responses	Pct of Cases
None	12	9.3	9.5
No Opinion	45	34.9	35.7
Don't Know	18	14.0	14.3
Pollution e.g. plastic paper	10	7.8	7.9
Felling trees	40	31.0	31.7
Conserves trees	3	2.3	2.4
Loss of soil leading to soil erosion	1	.8	.8
Total responses	129	100.0	102.4

0 missing cases; 126 valid cases

QBC8B how does your fuel use contribute to vegetative degradation

Dichotomy label	Count	Pct of Responses	Pct of Cases
None	15	11.9	11.9
No Opinion	48	38.1	38.1
Don't Know	18	14.3	14.3
Causes no Pollution	4	3.2	3.2
Reduces deforestation	9	7.1	7.1
Using alternative fuel to charcoal	32	25.4	25.4
Total responses	126	100.0	100.0

0 missing cases; 126 valid cases

QBC9A how can we help conserve soil

Dichotomy label	Count	Pct of Responses	Pct of Cases
Control flooding	1	.7	.8
Control erosion	13	9.0	10.3
No Opinion	15	10.3	11.9
Don't Know	11	7.6	8.7
Plant hardy tress species and grass/soil	68	46.9	54.0
Build gabions	12	8.3	9.5
Build water reservoirs	2	1.4	1.6
Make good use of soil from construction	7	4.8	5.6
Improve drainage	6	4.1	4.8
Its is difficult due to occupation	2	1.4	1.6
Government to police land excavation	8	5.5	6.3
Total responses	145	100.0	115.1

0 missing cases; 126 valid cases

QBC9B how can we help conserve water?

Dichotomy label	Count	Pct of Responses	Pct of Cases
Drill more boreholes	12	8.4	9.5
No Opinion	15	10.5	11.9
Don't Know	7	4.9	5.6
Plant trees	16	11.2	12.7
Construct water reservoir and storages	26	18.2	20.6
Encourage proper use of water	26	18.2	20.6
Clean dirty water/treat	12	8.4	9.5
Improve drainage and supply	10	7.0	7.9

Policing and law by government	6	4.2	4.8
Get alternative source of water e.g. rain	13	9.1	10.3
	-----	-----	-----
Total responses	143	100.0	113.5

0 missing cases; 126 valid cases

QBC9C how can we help conserve vegetation

	Count	Pct of Responses	Pct of Cases
Dichotomy label	59	43.1	46.8
Revegetation of areas/awareness	13	9.5	10.3
Use alternative fuel	14	10.2	11.1
Nothing can be done	7	5.1	5.6
Stop burning fuel	6	4.4	4.8
Provide water for irrigation	21	15.3	16.7
Prevent deforestation	10	7.3	7.9
Don't Know	3	2.2	2.4
Encourage people to keep animals	4	2.9	3.2
Policy guidelines for construction	-----	-----	-----
Total responses	137	100.0	108.7

0 missing cases; 126 valid cases

QBC9D Any other Comments

	Count	Pct of Responses	Pct of Cases
Dichotomy label	61	41.5	48.4
No comment	1	.7	.8
Dont Know	1	.7	.8
Conserve soil	8	5.4	6.3
Provide for cheap alt.sources of energy	4	2.7	3.2
Prevent deforestation	9	6.1	7.1
Plant more trees and cover crop	13	8.8	10.3
Improve security	31	21.1	24.6
Improve infrastructure	1	.7	.8
Stop plastic pollution	18	12.2	14.3
Advocacy, awareness and policy direction	-----	-----	-----
Total responses	147	100.0	116.7

0 missing cases; 126 valid cases

UNIVERSITY OF NAIROBI
EAST AFRICANA COLLECTION



UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

**E-PROCUREMENT MODEL FOR THE PUBLIC
SECTOR OF KENYA**

By:

Malela, Grace Akinyi

P56/P/7751/2005

Supervisor:

Mr. C. A. Moturi

University of NAIROBI Library



0439113 2

**Submitted in partial fulfilment of the requirement of the Master of Science in
Information Systems**

August, 2010

Declaration

I, Malela Grace Akinyi do declare that this research project as presented in this report is my original work and has not been presented for any other University award.

Signed:  Date: 1.9.2010

Name: Malela, Grace Akinyi

Reg. No. P56/P/7751/2005

This research project has been submitted as part fulfilment of the requirements for the Master of Science in Information Systems of the University of Nairobi with my approval as the University Supervisor.

Signed:  Date: 6/9/2010

Mr. C. A. Moturi

School of Computing and Informatics

University of Nairobi

Dedication

To My father, Alphayo Malela Odidi

And

My husband and Children, John, Fabians, Renny and Florah Sophie respectively.

Acknowledgement

First and foremost, I pay a multitude of thanks to the Almighty God who gave the strength and the opportunity to undertake this challenging project.

To my supervisor Mr. C. A. Moturi, whom through his understanding always gave a listening ear even with the frequent visits to his office, Prof Rodrigues and the panellists for their guidance and criticisms. To my dear father who saw me through up to this level. Without his efforts I would not have reached this far. To my family, loving husband John Okwaroh, Sons Fabians and Renny, daughter Florah Sophie who had to do without their mother around them for most of the duration. Lastly, I cannot forget my friends and classmates especially Esther who were always there to advice and give me courage to move on.

I also wish to extend my gratitude to the respondents who took their valuable time with patience and cooperation to respond to my questionnaires and interviews during data collection exercise.

Table of Contents

Declaration	i
Dedication	ii
Table of Contents	iv
List of Tables	vi
List of Figures.....	vii
Acronyms and Abbreviations.....	viii
Abstract	1
CHAPTER 1 - INTRODUCTION	2
1.1 General Introduction.....	2
1.2 The Public Procurement and Disposal Act and its objectives	4
1.3 Statement of the Problem.....	6
1.4 Objectives of the Study	7
1.5 Research Questions.....	7
1.6 Importance of the Study.....	7
1.7 Scope of the study.....	8
CHAPTER 2 - LITERATURE REVIEW.....	10
2.1 Introduction	10
2.2 E-procurement Models.....	16
2.3 Selected Reviews	27
2.3.1 South Korea.....	27
2.3.2 E-Procurement by Brazil's Federal Government - COMPRASNET.....	29
2.3.3 E-Procurement in the Danish Public Sector - PePP	31
2.3.4 ChileCompra of Chile	32
2.3.5 Italian Consip Business Model.....	36
2.4 Challenges of e-procurement	39

2.4.1	Legal Challenges and E-procurement.....	39
2.4.2	Other Obstacles to e-Procurement Adoption	41
CHAPTER 3 - RESEARCH METHODOLOGY.....		43
3.1	Introduction.....	43
3.2	Research Design	43
3.3	Research Population.....	44
3.4	Data Collection Methods	45
3.5	Assumptions.....	45
3.6	Scope and Limitations.....	45
CHAPTER 4 - RESEARCH FINDINGS		46
4.1	Situational Analysis	46
4.2	Level of e-procurement Use/Adoption.....	47
4.3	Number of Suppliers as a factor for slow Adoption	48
4.4	Perceived Benefits of e-procurement	49
4.5	Factors for Slow Adoption	51
CHAPTER 5 – THE PROPOSED MODEL		53
5.1	Stakeholders.....	53
5.2	The Public E-Procurement Model.....	56
5.2	Proposed Network Design.....	62
5.3	Security	63
CHAPTER 6 - CONCLUSIONS AND RECOMMENDATIONS.....		65
6.1	Review of Work Done	65
6.2	Conclusion	65
6.3	Recommendations and Further Research.....	68
REFERENCES.....		69
APPENDICES		71
Appendix A: Data Collection Instruments.....		71

List of Tables

Table 1-1: Listed Cases by Public Procurement Board 3

Table 1-2: Objectives of e-Procurement system in private and public sectors (Source: AMS, 2003)...6

Table 2-1: Comparing the Models in Use 25

Table 2-2: Comparison of the two approaches to Compra e-procurement 36

Table 4-1: Perceived Benefits of e-Procurement.....50

Table 4-2: Factors for Slow Adoption of e-Procurement in the Public Sector of Kenya 52

List of Figures

Figure 2-1: Evolution of E-Procurement (<i>Source: Dagg, 2005</i>).....	13
Figure 2-2: Buyer Centric Model	19
Figure 2-3: Seller Centric Model.....	20
Figure 2-4: Illustration of Seller-Centric.....	21
Figure 2-5: E-Marketplace or Third Party Managed Model.....	22
Figure 2-6: E-marketplace.....	23
Figure 2-7: Open Trading Environment.....	26
Figure 2-8: Efficiency Gains through PePP.....	32
Figure 2-9: 2003 – 2008 Approach.....	35
Figure 2-10: Current Approach.....	35
Figure 2-11: Consip's Fields of Action.....	37
Figure 2-12: Different Systems of Consip (<i>Source: Arthur Andersen, Booz Allen & Hamilton, Consip</i>)	38
Figure 2-13: Potential Legal Issues/Challenges (<i>Source: Ting and Wong, 2003</i>)	41
Figure 4-1: Means of Communicating with Suppliers	46
Figure 4-2: Internet Use.....	47
Figure 4-3: Level of e-Procurement Use/Adoption.....	48
Figure 4-4: Number of Suppliers Doing Business with the Institutions	49
Figure 5-1: E-Procurement System Stakeholders.....	54
Figure 5-2: Proposed e-Procurement Model for Kenyan Public Sector	58
Figure 5-3: The Enhanced e-Procurement Model.....	59

Acronyms and Abbreviations

API	-	Application Programming Interface
CEO	-	Chief Executive Officer
DTE	-	Electronic Tax Documents
EDI	-	Electronic Data Interchange
e-MRO	-	electronic Maintenance, Repairs and Operations
ERP	-	Enterprise Resource Planning
G4C	-	Government for Citizens
ICT	-	Information Communication Technology
IFMIS	-	Integrated Financial Management Information System
JDBC	-	Java Database Connectivity
KEPSA	-	Kenya Private Sector Alliance
MRO	-	Maintenance, Repairs and Operations
OASIS	-	Organization for the Advancement of Structured Information
ODBC	-	Open Database Connectivity
PePP	-	Public e-Procurement Portal
PKI	-	Public Key Infrastructure
PPARB	-	Public Procurement Administrative Review Board
PPOA	-	Public Procurement Oversight Authority
PPOAB	-	Public Procurement Oversight Advisory Board
RDBMS	-	Relational Database Management System
SAP	-	Systems Application and Products
SMEs	-	Small and Medium sized Enterprises
SOAP	-	Simple Object Access Protocol
SQL	-	Structured Query Language
SSL	-	Secure Socket Layer
UDDI	-	Universal Description, Discovery and Integration Standards
WSDL	-	Web Services Description Language
XML	-	Extensible Mark-up Language

Abstract

The public procurement in the Kenyan has been undergoing reforms starting with the Public Procurement and Disposal Act 2005 that saw the creation of Public Procurement Oversight Authority. The next step was the implementation of e-procurement for the public sector. According to e-government strategy paper 2004, e-procurement was one of the medium term objectives which was to be implemented by June 2007, but the process has been very slow. The manual processes are costly, slow, inefficient and data storage and retrieval poor.

This project sought to determine e-procurement practices in public sector, find the factors for slow adoption, existing models in use, and then develop a model for the Kenyan public sector.

The methodology used was questionnaires sent out to respondents, interviews and also through existing literature.

Several models have been tried by different countries to implement e-procurement. These are seller centric buyer centric, e-marketplaces or third party managed. In some countries these models are summarized into three: public, the mixed model and public private partnerships. The findings show that most of the procurement processes are still manual with the internet only being used for e-mails and web browsing. The factors for slow adoption include limited legislation, poor infrastructure, lack of awareness and top management support, integration with internal systems/solutions, lack of technical standards, lack of cooperation on the part of suppliers, costs associated with adapting web-enabled purchasing system

The proposed model for the Kenyan public sector is the public model where the government will own all the investment risks as well as benefits. This model favours the SMEs doing business with public sector, offers easier integration and since its government owned, it will gain a lot of support from to managers.

E-procurement is not as simple as just providing a catalogue on the Internet. E-procurement is not suitable for all products and several systems needs to be put together to achieve results. Sound strategy also needs to be put in place.

CHAPTER 1 - INTRODUCTION

1.1 General Introduction

In the past decades, the public procurement system in Kenya has undergone significant developments. From being a system with no regulations in the 1960s, and a system regulated by Treasury Circulars in the 1970s, 1980s and 1990s, the introduction of the Public Procurement and Disposal Act (PPDA) of 2005 and the Procurement Regulations of 2006 has introduced new standards for public procurement in Kenya.

In line with the country's public procurement reform agenda, Kenya in 2006 committed itself to become one of the 22 countries participating in the pilot testing a new Methodology for Assessment of National Procurement Systems (version 4) developed by the OECD-DAC Joint Venture for Procurement.

The Government of Kenya has been reforming the procurement process countrywide. Indeed after the publication of Public Procurement Regulations 2001, both the Government and multilateral donors embarked on further reforms by reviewing the procurement regulations and draft procurement bill that was eventually accented by the President of Kenya in 2005. Procurement Law was put in place under the name of Public Procurement and Disposal Act, 2005 and Public Procurement and Disposal Regulations 2006. The procurement law has now changed the way procurement process is undertaken by public entities. Similarly, a standardized insurance services tender document 2006 was formulated by Public Procurement Oversight Authority (PPOA), to enhance fair play to all stakeholders.

Prior to the publication of the new procurement law, the Public Procurement Regulations 2001 had put in place a mechanism of protest for the aggrieved bidders who considered awards by public entities not executed according the regulations. In it was established the Public Procurements Review and Appeals Board. In the intervening years, 2001-2006, a number of protests against award of insurance contracts were heard and determined by the Public Procurements Review and Appeals Board. For example: these are some of the few listed cases by Public Procurement and Appeals Board. These are shown in the table below.

	Year/Date	Insurance Broker	Procuring Entity	Case Type	Remarks	Verdict
1.	5 th March, 2003	Kenbright Insurance Broker	Agricultural Development Co.	Insurance cover for Development House	No public tender carried (single Source Procurement)	Board recommended a re-tender
2.	6 th August 2003	Getro Insurance Broker	City Council of Nairobi	Breach of Public Procurement Regulations and Minister of Local government awarding tender to M/s Invesco Insurance Broker	Lack of using Public Procurement Regulations properly	A Re-tender was recommended
3.	13 th January, 2004	Boma Insurance Broker	Kenya Police	Breach of Public Procurement Regulations	Failure to use the procurement Regulations properly	Re-tender was recommended
4.	8 th August 2004	Momentum Insurance Broker	Kenya Police	No grounds for appeal	Appeal	Rejected by Board

Table 1-1: Listed Cases by Public Procurement Board

Source: Schedule 1 of Public Procurement Complaints Review and Appeals Board Minutes

In the recent past, reviews are done by PPOA and reports are published in their website (www.ppoa.go.ke/downloads/reviews/reports). Some of those listed reports are due to none compliance with the recent Public Procurement and Disposal Act 2005.

The drivers for procurement reform include:

- Globalization
- Rapid pace of technological change
- Litigious society
- Increased transparency and public scrutiny
- Government restructuring
- Expanding socio-economic benefits

1.2 The Public Procurement and Disposal Act and its objectives

In 2003 the Government proposed the Public Procurement and Disposal Bill to establish the Public Procurement Oversight Authority. The bill aimed at making procurement more transparent and accountable and would require procurement agencies to carry out an annual update of pre-qualified firms, especially when dealing with restricted tenders such as military tenders.

The purpose of the act is to establish procedures for efficient public procurement and for the disposal of unserviceable, obsolete or surplus stores, assets and equipment by public entities and to provide for other related matters. The Act has therefore addressed one common area of potential abuse by which Government property is significantly undervalued for disposal to private entities. The objectives of the Act include:

- Maximize economy and efficiency
- Promote competition and ensure that competitors are treated fairly
- Promote integrity and fairness of those procedures
- Increase transparency and accountability in those procedures
- Increase public confidence in those procedures
- Facilitate the promotion of local industry and economic development

According to the Act, there are three bodies involved in the regulation of public procurement.

- Public Procurement Oversight Authority (PPOA).
- Public Procurement Oversight Advisory Board (PPOAB).
- Public Procurement Administrative Review Board (PPARB).

The third major reform programme in the public procurement is the implementation of e-procurement for the public sector under the guidance of PPOA.

Typically, e-procurement web sites allow qualified and registered users to look for buyers or sellers of goods and services. Depending on the approach, buyers or sellers may specify costs or invite bids. Transactions can be initiated and completed. Ongoing purchases may qualify customers for volume discounts or special offers. Companies participating expect to be able to control inventories more effectively, reduce purchasing agent overhead, and improve

procurement cycles. E-procurement is expected to be integrated with the trend toward computerized supply chain management.

Within the study the term e-procurement is understood in its broad sense. It consists of two core processes: the contract establishment process that contains notification, tendering, evaluation and contract conclusion; the order and payment process that consists of buyer order, order acceptance, delivery of goods, delivery note, invoice and payment.

E-procurement is done with a software application that includes features for supplier management and complex auctions.

There is a growing consensus that e-procurement is the single most important area of development in procurement. Within the public sector context e-Procurement has been widely embraced by governments seeking the administrative and cost reductions experienced in the private sector. As a result a number of ‘proven’ private sector e-Procurement solutions such as e-marketplaces, desktop purchasing systems, and tendering platforms have been employed by various public sector organizations. However, public e-Procurement differs from the private sector in various aspects mainly because of its economic and social considerations. These differences result in a number of specific regulations and standards that have been developed for public e-Procurement.

The general objectives of e-procurement systems in public and private sectors are shown in the table below.

Private	Public
• Increase bottom line profitability • Allow uniform access to standardized supplier information • Leverage buying power of corporation to provide uniform pricing and contracts among multiple divisions and groups • Increase the speed of the procurement cycle • Increase compliance with established contracts by reducing off-contract (maverick) purchasing	• Increase top-level funding availability • Create open markets in which every provider can compete • Harness aggregated buying power of the government to achieve dynamic pricing of goods and services • Improve efficiency of the procurement cycle while meeting all legislative mandates • Ensure compliance with purchasing policies, both within departments and across the government agencies

Table 1-2: Objectives of e-Procurement system in private and public sectors (Source: AMS, 2003)

1.3 Statement of the Problem

E-procurement is one of the public procurement reform programmes being pursued by the government of Kenya having come up with the Public Procurement and Disposal Act 2005. E-procurement was one of the medium term objectives to be accomplished by the year 2007, but nothing much has been done to date. Given these issues, the study seeks to determine the factors that have affected the adoption and implementation of e-procurement in the public institutions of Kenya. Like other public management reform, adopting advanced IT in the public procurement process faces many problems, theoretically and practically.

Despite the advances in technology and the downward trends in the cost of computer hardware, the Kenyan public sector has been very slow to embrace e-procurement. Over the years many e-procurement management models have been developed. Most of the organizations in Kenya run fully developed procurement departments, and many organizations are thinking about the concept of e-procurement. The existing models are not known or well understood to be applied.

The bulk of corrupt practices in Kenya have occurred in public procurement. The explanation for this state of affairs is to be found in the political dynamics of the state and its role in the economy.

Some of the problems with the current procurement system include: high costs in publicity and advertisements; very long cycle time invitation of bidders to award of contract; poor storage of historical information on previous contracts; inflated costs of goods and services far beyond market prices; lack of transparency and accountability; lack or inadequate information on suppliers. This project therefore seeks to find out the current practice, factors for the slow pace, existing models and design a model that can be adopted by the Kenyan public sector.

1.4 Objectives of the Study

The study was designed to fulfil the following objectives:

- i. Determine the e-procurement practices by public institutions in Kenya
- ii. Identify the factors that have led to the slow adoption of e-procurement in public sector of Kenya
- iii. Study the available e-procurement models, their strengths and challenges
- iv. Develop e-procurement model that can be adopted by the public sector.

1.5 Research Questions

The study seeks to answer the following research questions:

- i. What are the e-procurement practices in the Kenyan public sector?
- ii. Which models are currently available and in use?
- iii. What factors affect the public e-procurement adoption in Kenya?

1.6 Importance of the Study

Section 45 of The Public Procurement and Disposal Act No. 3 of 2005 states that, “A procuring entity shall keep records for each procurement for at least six years after the resulting contract was entered into, or no contract resulted, after the procurement proceedings were terminated”. In Government, information is a key resource. It is basic to the management of resources, execution of functions, measurement of performance, pursuit of democratic processes and service delivery. Owing to the information intensive nature of Government operations, an essential element of the reform process entails changes in the management information systems. The manual nature of managing information systems today, the stand-alone information technology systems and the multiplicity of equipment in the public service have impacted negatively on the delivery of certain essential services. Poor record keeping and cumbersome filing systems have been major impediments to efficient service delivery.

Driven by demand for further control over public spending and more efficient acquisition processes, procurement has become a key public function. Significant technological changes,

international trade and global competition have forced Governments and public institutions to extend policy interests beyond cost minimization, and to pursue new challenges including innovation, competition, transparency and reforms. Further, policy makers have become aware of the role of public procurement for many other socio-economic purposes. Crossing both economics and business, public procurement has raised the interest of practitioners directly involved in the management of purchases and with academic researchers to answer key questions, explore new issues, and to discuss an extensive range of topics.

The completed study will benefit the following institutions/persons:-

1. The policy makers in the Government of Kenya will find these results important as they will be able to know the factors that have impeded the quick adoption and implementation of e-procurement in public institutions. This will further help in formulating policies geared towards addressing this.
2. The successful completion of the study is important to the public institutions in Kenya in that after clearly understanding the factors affecting the adoption and implementation of e-procurement in their institutions they will allocate resource toward addressing this.
3. The scholars and researchers who may wish to study on e-procurement or on the public sector in Kenya will also find this study useful as a basis for their further research or other academic discourses. The recommendations given here will form the basis for further research to be undertaken.

1.7 Scope of the study

This study covered the e-procurement practices of public institutions in Kenya. Further, other organisations other than public institutions will not include the scope of this study. The study covered:

- The Central Government .i.e. government ministries
- Local Authorities and
- State Corporations.

E-procurement is not an easy endeavour and several systems can be implemented and integrated for the procurement of different goods/services. Due to this technicality this study only considered procurement of direct goods i.e. consumables whose procurement process is straight forward.

CHAPTER 2 - LITERATURE REVIEW

2.1 Introduction

This chapter focuses on the review of literature related to e-procurement and logistics in general.

2.1.1 Definitions

Although there are many definitions of e-Procurement and confusion exists in defining the term (Vaidya et al., 2004; World Bank, 2003; Murray, 2001), *e-Procurement is basically a tool that enables procurement activities such as sourcing, ordering, commissioning, receipting and making payment.*

E-procurement is the buying and selling of goods and services (business to business) in a digital environment by means of software applications that are “web” or “internet” based. An e-procurement network provides a secure market place, sales platform and transaction tracking system to buyers and sellers.

E-procurement is either the business-to-business or Business-to-Consumer purchase and sale of supplies and services through the Internet as well as other information and networking systems, such as electronic data interchange (EDI) and Enterprise Resource Planning (ERP). An important part of many B2B sites, e-procurement is also sometimes referred to by other terms, such as supplier exchange.

There are six main types of e-procurement namely web-based ERP, electronic Maintenance, Repair and Operating (e-MRO), e-sourcing, e-tendering, e-reverse auctioning, and e-informing.

E-marketplace - is an arena on the Internet where a trusted intermediary (e-market) offers trading functionality to registering companies (Swedish Trade Council, 2001).

B2B - Business-to-business (B2B) describes commerce transactions between businesses, such as between a manufacturer and a wholesaler, or between a wholesaler and a retailer. Contrasting terms are business-to-consumer (B2C) and business-to-government (B2G).

E-Catalogue - An e-catalogue is an online presentation of information on products and services that are offered and sold by an organisation.

E-Government - The delivery of government information and services using Internet technology.

Portal - also known as a links page, presents information from diverse sources in a unified way.

E-Marketplace - Commonly known as electronic marketplace, an e-marketplace is a virtual market where Buyers and sellers meet just like in a traditional market just that in this case, all interactions are done virtually. e-marketplace includes b2b, b2c, c2c etc. For example B2B e-Marketplace is Business-to-Business e-marketplace.

ERP - Enterprise resource planning is an integrated computer-based system used to manage internal and external resources, including tangible assets, financial resources, materials, and human resources. It is a software architecture whose purpose is to facilitate the flow of information between all business functions inside the boundaries of the organization and manage the connections to outside stakeholders. Built on a centralized database and normally utilizing a common computing platform, ERP systems consolidate all business operations into a uniform and enterprise-wide system environment.

E-Tendering – This is the procurement process simply conducted online, ie supplier registration/expression of interest, contract download, submission of bid document, evaluation of tenders. May or may not involve e-auctions.

PPP - Public-private partnership (PPP) describes a government service or private business venture which is funded and operated through a partnership of government and one or more private sector companies.

Private Sector - In economics, the private sector is that part of the economy which is run by private individuals or groups, usually as a means of enterprise for profit, and is not controlled by the state.

Public Sector - The public sector, sometimes referred to as the state sector is a part of the state that deals with the production, delivery and allocation of goods and services by and for the government or its citizens, whether national, regional or local/municipal.

2.1.2 Brief History of e-Procurement

In the last half of the twentieth century, there was a revolution in the industry as a result of the harnessed power of seemingly ever-increasing capacity, speed and functionality of computers and microprocessors. This trend made a way for management and workers within industries with new capabilities for management, planning and control, design, quality assurance and customer support. Structured information flow became the foundation of industrial companies. New applications, tools and information technology systems emerged and evolved to facilitate companies to integrate the various departments like Design, Procurement, Manufacturing, Sales and Finance within companies, predominantly the larger ones, including international corporations, providing opportunities for them to meet new demands for product time to market, just in time supply of orders, and customer support.

The earliest literature on e-procurement is that relating to electronic data interchange – a technology that has been in use in organizations since the 1960s. Most dialogue about electronic inter-organizational systems in the academic literature up until the mid 1990s involved electronic data interchange. It was not until the mid 1990s that there was a shift towards the discussion of the use of the internet for electronic commerce. In contemporary times it still stands that electronic data interchange continues to be the primary medium of electronic commerce. According to (Neef, 2001), the emergence of e-procurement is not just as an improvement of EDI technologies, but it is a way of conducting purchasing transactions over the Internet.

From Figure 2-1 below, e-procurement effectively began in late 1990s when several start-up software companies, led particularly by Ariba and Commerce One, began to develop a suite of applications that allowed vendors to create electronic catalogues (Neef, 2001). This turn of events dramatically altered the activities of purchasing and transformed the purchasing process from a tactical into a strategic activity and tried to eliminate maverick buying, or purchases that did not involve the purchasing department.

eProcurement history

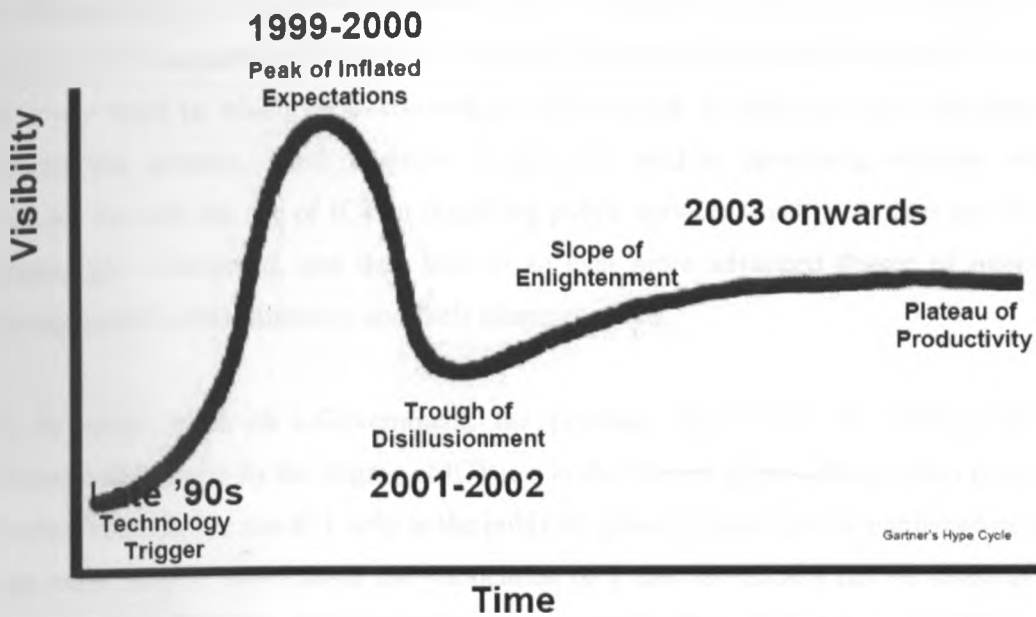


Figure 2-1: Evolution of E-Procurement (Source: Dagg, 2005)

Figure 2-1 above illustrates the fact that initially there was excitement over using electronic means to do business and e-procurement was one of the first business application areas to make use of the Internet in the 1990s. Many companies wanted to be part of this technological trigger. However, the period 1999 to 2000 shows that the expectations were rather too high as a lot of companies needed to put a lot of things in place before this new technology could work well for them. This era was followed by the trough of disillusionment in the period 2001-2002 as many companies' expectations were not met. From 2003 onwards, there has been a steady appreciation of e-procurement as a lot of companies have put things in place and they have seen e-procurement work for them.

2.1.3 Public e-Procurement

Public e-procurement is the use of electronic means for publishing, processing, exchanging and storing all the information related to institutional purchases and provisioning in public organizations. Even if public e-procurement platforms are similar to B2B marketplaces used in the private sector, the underlying processes are complex and require sophisticated technological tools which must comply with legal and organizational constraints.

Public e-Procurement refers to the use of information and communication technology (ICT) on the part of the public administration in order to manage its relationship with suppliers with regards to the acquisition of supplies, services, and labour for the public sector. It is one of the many ways in which e-Government is implemented. In planning and implementing e-Government systems, much attention is typically paid to optimising relations with the citizenry through the use of ICT in supplying public services. Such initiatives are becoming increasingly widespread, and they lead to an ever more advanced degree of interactivity between public administrations and their clients.

As for other types of e-Government, the practical applications of e-Procurement are differentiated mostly by the degree of ICT use in the various phases of any given project. The simplest applications use ICT only in the publicity phase: tenders can be published online, or, even more simply, news about the publication of a call for tenders can be made available online. More complex ICT infrastructures are necessary in order to make it possible for potential suppliers to accept economic offers online. Even more complex systems are needed for the electronic management of purchased supplies, which require shared standards for electronic data exchange in the ordering, invoicing, accounting, and payment phases.

Within the galaxy of e-Government applications, e-Procurement is characterized by an inversion in the relationship described above: the administration is the one requesting the services of the private sector; the positive effects of using new technologies manifest themselves mostly in an improved allocation of resources in the public sector. The use of electronic means makes it possible to cut the transaction costs of public sector purchases through the automation and the centralization of the fulfilment of administrative requirements. This then makes it possible to reallocate the administration's human resources towards organizational units directly responsible for furnishing services to the citizenry.

Financial advantages for the public administration can also arise out of lower unit costs for the goods and services that are required. This can happen both because the centralization of orders makes it possible to purchase goods in bulk, and thanks to the improved functioning of competition and market mechanisms. Indeed, the electronic management of procurement process, if implemented correctly from a technical point of view and supported by an adequate normative regime, can make such processes more transparent, widen the base of potential suppliers (both in geographic terms and in terms of the size of potentially interested

firms), and improve the flow of information; these are essential conditions for competition that the administration can take advantage of.

It is thus evident that the main motives behind the implementation of e-Procurement programmes on the part of public administrations are often financial in nature. Citizens enjoy these benefits indirectly, since they lead to a more efficient use of public funds.

The implementation of e-Procurement programmes also has other objectives, which are more difficult to quantify but probably more important.

Indeed, process automation is an excellent tool to promote transparency in public tenders. For example, an e-auction system can register bids automatically, and does not allow their consultation before the closing of bids. Minimizing human intervention also makes it easier to identify the individuals managing the process, thus making them more responsible. These characteristics make e-Procurement a particularly valid solution in places in which corruption is widespread and negatively impacts the functioning of markets within which public administrations operate.

The introduction of new management modalities is also an opportunity to re-organize work flows, including both those of the public administration and those of private economic operators. Thanks to ICT, procedures can be sped up and optimized, thus pursuing the simplification goals usually associated to e-Government programmes.

The implementation of e-Procurement system can also be seen as a development policy to promote electronic commerce and in more general terms to encourage the computerization of the private sector. Public administrations that automate their supply system encourage the private sector to invest in ICT in order to participate in public tenders; it introduces and spreads communications standards and protocols for the exchange of documents that can then be used for transactions between private firms; in other words, it creates an electronic market in which e-Procurement's benefits in terms of efficiency and transparency are transferred to the private sector.

2.2 E-procurement Models

For today’s industrial age organizations to become tomorrow’s leaders in e-business, their current procurement practices must change (Kalakota, Robinson 2001). As a consequence, for organizations to compete successfully and retain customers who are of primary concern, organizations need to deploy new models of e-procurement applications to improve their procurement processes. There is a range of different models for procurement. The models are evolving fast, so it is difficult to know which to select. These various new models of e-procurement application exploited by different researchers attempt to solve similar business process problems.

Kalakota and Robinson (2001) divide these e-procurement trading models into seven (Table 2-1) which, they categorize as either a public or private exchange. The table below shows some of the generic e-procurement trading models which are in use and also their key differences. These are some of the generic options for implementing e-procurement.

S/N	Trading Model	Characteristics
1	EDI networks	• Handful of trading partners and customers • Simple transactional capabilities • Batch processing • Reactive and costly value-added network (VAN) charges
2	Business-to-employee (B2E) requisitioning applications	• Make buying fast and hassle-free for a company's employees • Automated approval routing and standardization of requisition procedures • Provide supplier management tools for the professional buyer
3	Corporate procurement portals	• Provide better control over the procurement process and let a company's business rules be implemented with more consistency • Custom, negotiated prices posted in a multi-supplier catalog • Spending analysis and multi-supplier catalog management
4	First-generation trading exchange: community, catalog, and storefronts	• Industry content, job postings, and news • Storefronts: new sales channel for distributors and manufacturers • Product content and catalog aggregation services
5	Second-generation trading exchanges: transaction-oriented trading exchanges	• Automated requisition process and purchase order transactions • Supplier, price, and product/service availability discovery • Catalog and credit management
6	Third-generation trading exchanges: collaborative supply chains	• Enable partners to closely synchronize operations and enable real-time fulfillment • Process transparency resulting in restructuring of demand and supply chain • Substitute information for inventory
7	Industry consortiums: Buyer and supplier led	• The next step in the evolution of corporate procurement portals

Table 2-1: Generic Trading Models

From a buying organization's perspective, the key differences between e-procurement models in the market today lie in where the buying application software and electronic catalogue(s) reside. Buying application software allows an organization to carry out purchasing processes. Depending on the sophistication of the software, it allows buyers to perform some or all of the following functions: access supplier information, including catalogues; generate

requisitions; obtain approvals in accordance with the business rules of the organization; send purchase orders to suppliers; receive invoices; and process payment.

Three basic architecture models have evolved over a number of years and can be broadly described as buyer-centric, seller-centric and e-marketplace. These electronic procurement models take a 'hub and spokes' approach (the hub being the organization at the centre of the system, while the buyers and/or suppliers who interact with the system are at the ends of the spokes) These models have particular advantages and disadvantages from the perspective of buyers and sellers, but they generally involve 'closed' systems, limiting the flexibility of those on the periphery. There is therefore an emerging trend towards a more open trading environment, based on open standards, which is intended to overcome the limitations of these closed systems.

The model that is best for an organization depends on its specific requirements. Some of the questions you will need to ask are: How much control do you want over the buyer application software you use? If you want to ensure the software meets your specific needs you may need to buy the software, customizing it, as required. Who should manage the electronic catalogue(s) you buy from? - Suppliers, your organization, or a third party (the e-marketplace provider)?

Wilson (2002) describes and divides the models into three based on application and functionality.

1. **Buyer Centric or Buy- side** procurement is a form of procurement system developed and implemented by large buyer organizations to web-enable their purchases with selected suppliers. These suppliers are also using e-procurement in the entire management processes relating to purchase, product development, transactions, etc. This actually creates a virtually integrated inter-organizational system between the buyers and the seller like EDI systems but with greater scope and capabilities. This type of model is designed predominantly to serve the needs of the buying organization.

In a buyer-centric or buyer-managed purchasing model, the buying organization implements software to support its procurement processes, obtains catalogue data from its contracted suppliers, and aggregates the catalogue data into a single internal catalogue for the use of its purchasing officers. In terms of a hub and spokes model, the buyer is at the hub with suppliers connected at the end of the spokes, as illustrated in Figure: 2-2.

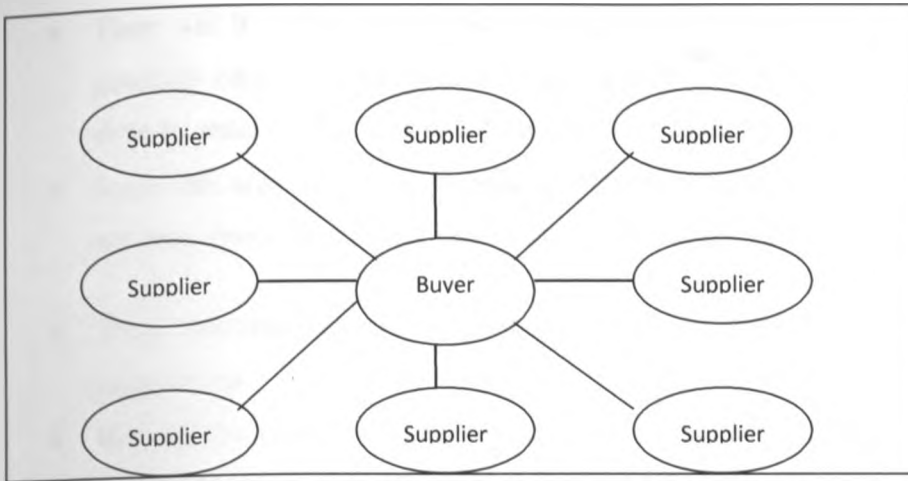


Figure 2-2: Buyer Centric Model

Benefits

- The main benefits to the buying organization are that tight control can be maintained over the procurement software, catalogue (and other procurement) data and processes.
- The solutions can be highly customized to meet the buying organization's needs.
- These systems also make it much easier for procurement solutions to be fully integrated into the buying organization's financial management system so that the whole procurement process can be easily automated from beginning to end.
- They tend to offer the most procurement functionality to buyers.
- They lead to fast and easy product search and selection.
- They give buyers access to multiple suppliers' offerings and can usually do price comparisons.
- For suppliers the costs of participating in these systems are generally low, there are no transaction fees and most require little investment from suppliers to participate.

Disadvantages

- This model generally results in a closed purchasing environment, allowing suppliers to trade only with the central buying organization. Suppliers need to use other arrangements to trade electronically with other buyers.

- There will be higher system maintenance costs in these solutions for the buyer, but generally catalogue maintenance (updating items, prices and availability) can still be done by sellers, with the updated data provided to the buyer.
- Issues can arise as to the currency of data, stock availability etc, when suppliers do not have direct control over items listed in the buyer's catalogue.
- These solutions can also require fairly high up-front investment for the buying organization.
- Because this model will generally not allow suppliers to trade universally, suppliers wishing to move to on-line trading may need to accommodate a number of different systems used by their customers, increasing costs and complicating content management.

2. **Seller Centric or Sell-side** procurement also referred to as e-sales is a form of procurement system by which one supplier sells to large number of buying organizations using e-procurement systems. In a seller-centric procurement model the seller is at the centre of the hub, with buying organizations connected at the spokes, as illustrated in Figure 2-3 and 2-4 below. Buyers use the supplier's system, accessible over the Internet, to browse the supplier's catalogue and place orders. This space in the marketplace has been largely the domain of business-to-consumer selling, although increasingly business-to-business trading is occurring on these sites.

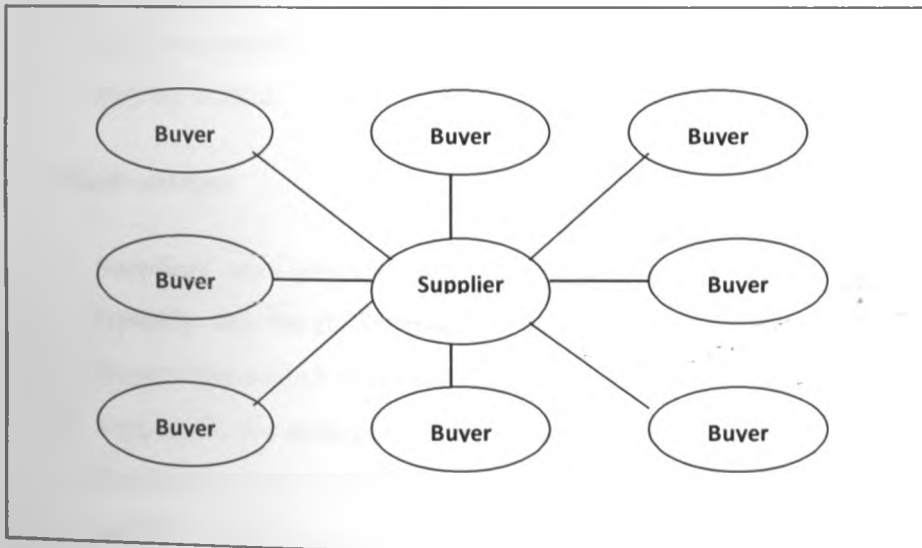


Figure 2-3: Seller Centric Model

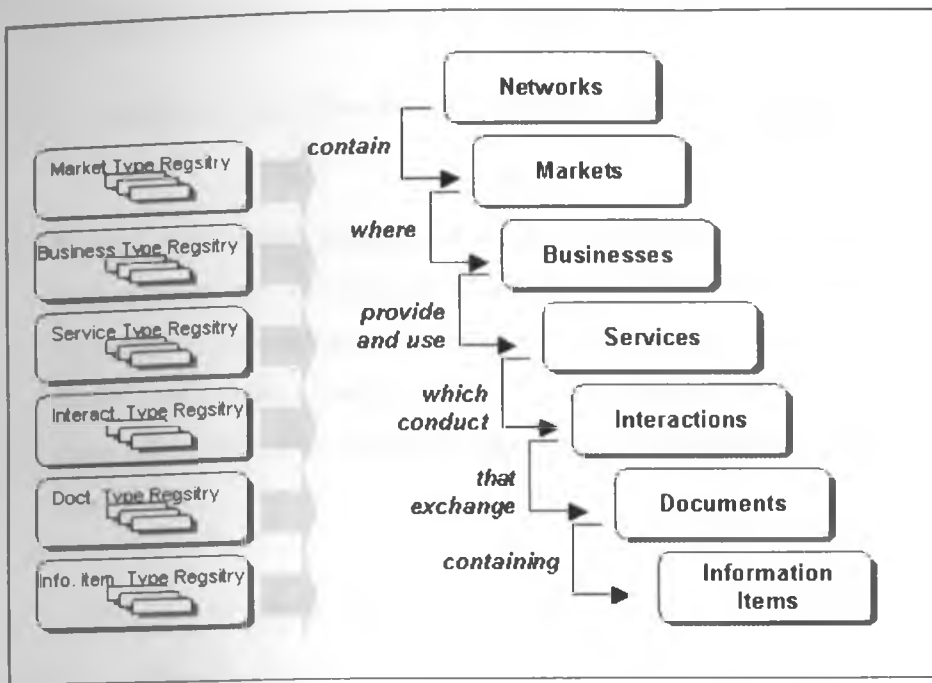


Figure 2-4: Illustration of Seller-Centric

Benefits

- For the buyer these solutions generally offer the lowest investment cost.
- Some sites are being aggregated into virtual e-mails to reduce the need to search many different sites, but buyers still need to know where to find the site.
- For suppliers these solutions may be optimal in terms of control, cost, maintenance and functionality, but their customers need to know where to find them so their use may be limited.

Disadvantages

- Suppliers' sites provide only limited support for the buyer's purchasing process, typically only the ability to browse catalogues and place orders.
- Buyers lose control over catalogue data.
- Support for pre-arranged contracts may be limited.
- Buyers must access each supplying organization's system individually, and only one supplier's catalogue can be accessed at a time.
- Buyers must learn to use multiple supplier systems and manage multiple access right details (e.g. user-name and password).

- Buyers must know where to find the site.
 - Integration into financial management systems at the buyer's end may be very difficult.
3. **E-Marketplace** and trading hub is the combination of industrial consortium and the trading exchanges. This type of model brings together many different buying and selling organizations in one trading community as depicted in Figure 2-5 below. The prominent e-marketplace task is auction where a variety of products is displayed. These create a close relationship and increase collaboration between organizations in a particular trade sector.

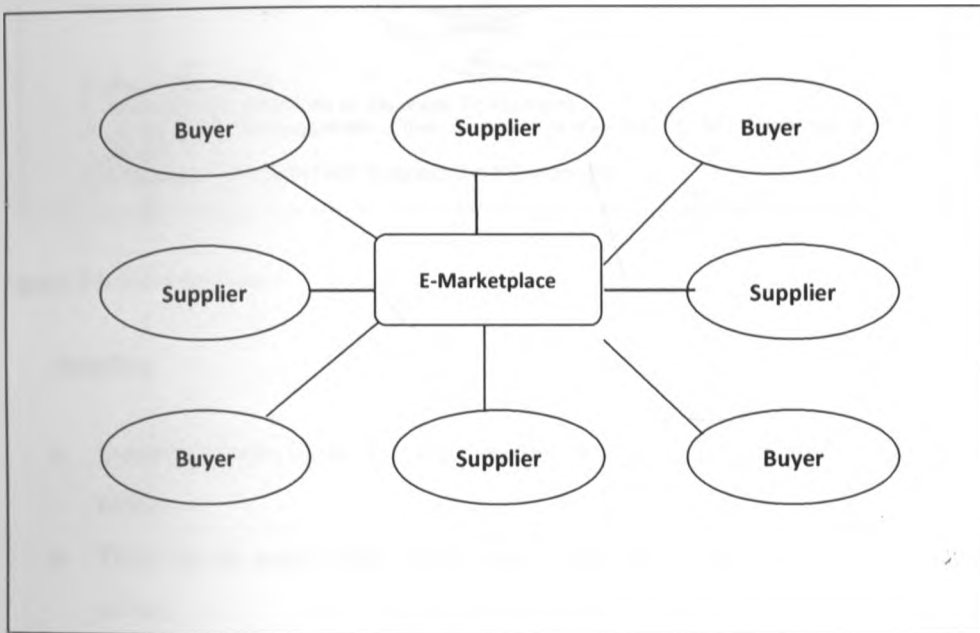


Figure 2-5: E-Marketplace or Third Party Managed Model

A further illustration of e-marketplace is shown in Figure 2-6 below.

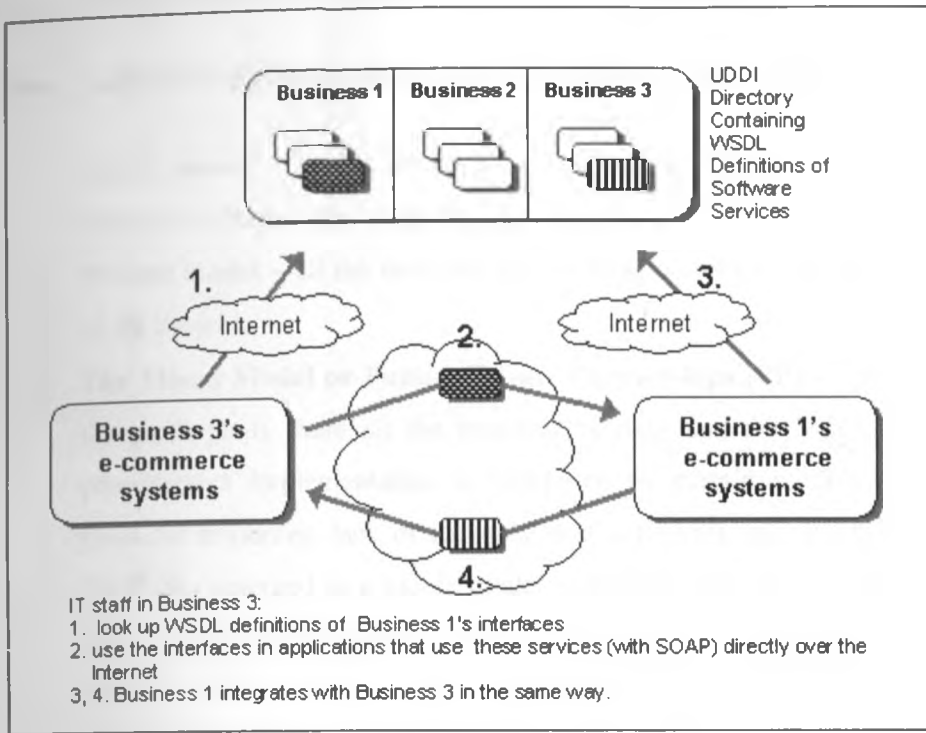


Figure 2-6: E-marketplace

Benefits

- E-marketplaces allow extended trading between many organizations - both buyers and sellers.
- They can be particularly useful for organizations that are large purchasers as well as sellers.
- These solutions are usually least expensive, in terms of overall initial costs, from the combined perspective of buyers and sellers.

Disadvantages

- E-marketplaces may provide only limited ability for buyers and sellers to determine business rules.
- These solutions have the highest on-going costs (subscriptions, maintenance and/or transaction fees).

- Although they allow extended trading, you must be a subscriber to the e-marketplace to access buyers and suppliers linked to the e-marketplace.

These models have also been categorized by Oliveira and Amorim as:

- **Public model** – in this model all the tasks including the investments and risks of building the portal are run by the government.
- **Private model** – all the tasks are run by private entities that bear the investment risk of the project.
- **The Mixed Model or Public /Private Partnerships (PPP)** – this is a model where the participants share all the investments risks and the benefits of the project. E-procurement implementation is hampered by certain roadblocks such as lack of financial resources, lack of technical and soft skills, etc. Public-private partnership (PPP) has emerged as a viable model to counter these factors, apart from improving the economic sustainability. PPP essentially implies sharing of risks and rewards of a venture. This model combines accountability with efficiency, as the services are governed by strict service level contracts. Advantages are:
 - ☐ No capital cost to the Government
 - ☐ State-of-the-art technology
 - ☐ High quality resources
 - ☐ Improved risk allocation – risk borne by the entity best able to handle that risk
 - ☐ Combining accountability with efficiency
 - ☐ Services governed by Service Level Agreements

The above three models are being used worldwide in many countries especially in South American countries, Europe and Asia. The first two models required fresh Government investment in an area where there is no prior experience.

Comparison of the models is shown in Table 2-2 below

	Model	Advantages/strengths	Disadvantages
1.	Public	• Profound legislation changes not necessary	• Time necessary to manage the process changing within the public sector • Determination of appropriate pricing for participation of suppliers • Prior experience required
2.	Private	• Coexistence between private and public suppliers conceivable	• Choice of an entity to manage the system can raise legal problems in terms of public law • Price of participation could be very high to suppliers especially SMEs thereby locking them out • More profound legislation changes. Some laws preserved for the public entities • Requires strong regulating and supervision authority
3.	PPP	• State of the art expertise in technology • Combines accountability with efficiency because services are governed by strict SLAs • High quality resources • No capital cost to government • Shared risks of hardware, software (improved allocation)	• Expected total cost higher at the beginning • More complex, expensive and time consuming stages of preparing agreements and project development • Also requires regulating and supervision authority and revision of contracting litigation rules • Insufficient stability and reliability of political organization

Table 2-3: Comparing the Models in Use

Developing an open trading environment

As noted, the three models described above have some disadvantages, connected with the fact that they generally involve closed communities. They are also based on one dominant player (the hub), which establishes the trading rules for the system, and this limits other participants' flexibility. However, these limitations can be overcome in an open trading environment which allows buyers and sellers to establish systems that meet their requirements, without imposing these on their trading partners.

Under an Open Trading model, buyers implement software that allows them to establish and maintain local internal catalogues (should they wish to do so), but also to access e-marketplaces and supplier's stand-alone catalogues. Suppliers are free to choose whether to host their own catalogues or use marketplaces. These arrangements benefit buyers by offering them flexibility in accessing supplier catalogues, regardless of where the catalogues are hosted. These arrangements benefit suppliers, who are able to make their catalogues visible to the widest possible market, regardless of the e-procurement arrangements individual buyers have implemented. The figure below illustrates some buyers and suppliers transacting between themselves, but through a marketplace with other buyers and suppliers.

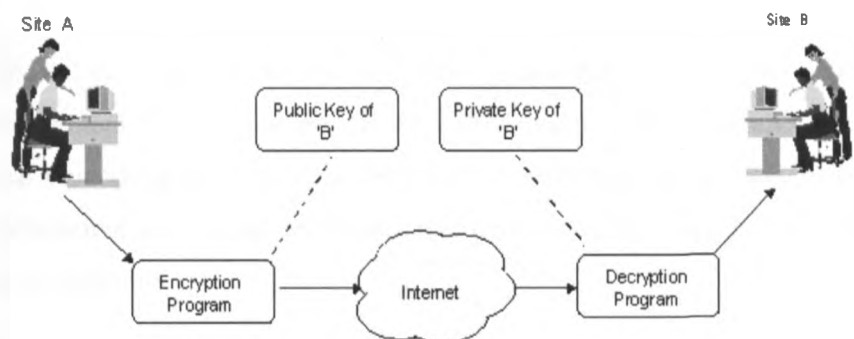


Figure 2-7: Open Trading Environment

Open trading arrangements of this kind, however, require the systems of buyers, suppliers, and e-marketplaces to be able to 'talk' to each other. For this to happen, the systems must be based on common, open standards. An open trading environment therefore requires a commitment to open standards on the part of buyers, suppliers, and e-marketplace providers.

Two of the most prominent of these standards are the OBI (Open Buying on the Internet) Standard and XML (Extensible Markup Language). The OBI architecture is based on the

premise that process owners should be responsible for maintaining information associated with their business processes.

2.3 Selected Reviews

2.3.1 South Korea

According to the World Bank report of July 2004 (www1.worldbank.org/prem), the **S. Korean Government** began reforming its notoriously complicated, non-transparent, corrupt public procurement system, introducing e-procurement to exploit the country's well-developed information and communications infrastructure in 1997. Through extensive business process reengineering and information strategy planning, the Public Procurement Service, the agency formerly responsible for buying Government goods and services has been transformed into a one-stop information centre. E-procurement has generated numerous benefits, including:

1. Enhanced transparency and public trust - by reducing contacts between officials and suppliers and by sharing information between Government agencies and the public.
2. Increased managerial efficiency - by achieving economies of scale in procurement, with an estimated \$2.5 billion a year in savings from the \$26 million investment.

The Public Procurement Service was established in 1949 to run Korea's centralized procurement system. The Procurement Business Law requires all Government agencies and local Governments to procure their supplies through the service. In practice, however, the procurement system has centralized and decentralized components. Some local Governments are exempted.

Before e-procurement was introduced, legal and regulatory requirements resulted in complicated procurement procedures. Companies seeking Government business had to scan newspapers, magazines, and Government gazettes to identify opportunities, and then register separately with each Government agency. Moreover, procuring agencies did not share information. Instead, each maintained its own register of suppliers. This time-consuming, costly application process discouraged suppliers, and prearranged bidding was not uncommon contributing to corruption. The process also limited economies of scale. Although a few ministries had computerized procurement systems when reform began in 1997, sophistication was low. And because these standalone systems were incompatible, they could not handle large volumes of procurement. Meaningful change began only when former

President Dae Jung Kim initiated comprehensive e-Government reform, vowing in his 1998 inaugural speech to have state-of-the-art systems in place by the end of his term in 2002. A Presidential E-Government Committee was created to oversee e-Government reform, which involved 11 major projects: the National Pension System, Internal Tax Service, Integrated Local Administration System, National Education Information System, Financial Management Information System, Human Resource Management System, Electronic Data Interchange, e-Signature System, Government wide Information System, Government for Citizens (known as G4C), and e-Procurement.

Some of the benefits the Korean Government has experienced include *enhanced transparency and competitiveness* and also *Savings on procurement management*. Transparent online information disclosure has been a powerful tool for increasing the accountability of e-procurement decision makers and system operators. Each system has its challenges which hamper its success. Korea's e-procurement system has not been an exception to some of these challenges. Although the system's hardware is in place, its operating system must be stabilized, users continuously trained, content reliability enhanced, and product choices broadened. Computer viruses and worms have become a serious threat to the system's stability. One infection already shut down the entire system resulting in financial losses to the Public Procurement Service and associated suppliers, and raising concerns about the system's reliability. Information security has become an urgent issue. Furthermore, some parts of the population have been left out of e-Government. A Government wide policy is needed to narrow the widening societal gap in information and communications technology capacity. For the full effect of e-procurement reform to be felt, its changes must be internalized, and savings in Government agencies' procurement should be reflected in their expenditure management.

There are several lessons that can be learnt from the Korea's e-procurement system. Every country interested in e-procurement has distinct goals, and environments for public procurement differ considerably across countries. Korea's experience offers four lessons that can help other countries implement such reform and overcome bureaucratic and other obstacles.

- Strong leadership
- Nationwide e-Government reforms, rather than as an isolated change

- Certain level of information and communications infrastructure.
- comprehensive process reengineering

2.3.2 E-Procurement by Brazil's Federal Government - COMPRASNET

This is a government e-procurement system set up by the Secretariat of Logistics and Information Technology in the Brazilian Ministry of Planning, Budget & Management. The system is a Web-based on-line procurement system used by all the more than one thousand Federal Government procurement units. It enables on-line price quoting and reverse auction commodity purchases. It has client/server architecture, resident on secure 32-bit Pentium III Xeon corporate servers. The operating system is Windows 2000 Advanced Server. The front end is supported by three Unisys Aquanta STD Web application servers. The solution software used is the Vesta Business Services Suite.

COMPRASNET is a system where Federal Government organizations register their procurement needs (i.e. goods and services they need to buy). The system automatically informs registered suppliers by e-mail and the supplier may download the bidding documents. The procurement officer uses a Federal Catalogue to specify the description of the good or service required. If the item is classified as a commodity, the whole process may be done through the Internet, using the price quoting system (which is a two- to three-day purchase posting site for non-competitive small purchases). Reverse auction procedure is used for larger procurement of general-purpose goods and services (such as building maintenance services or office supplies and equipment). In the reverse auction the bids (prices the suppliers will charge for that item) are submitted on the Web. Each supplier reduces their bid price competitively with others during the auction and the one offering the lowest price at a pre-agreed end time for the auction will be the one awarded the contract. Auctions and prices are open for inspection by the public, and auction results are posted immediately.

COMPRASNET was introduced to automate the procurement process. The aim of automation was to make the procurement process uniform without centralizing the buying process of the Federal organizations. It was also intended to reduce procurement costs and give more transparency to the process. Other aims were to increase the number of government suppliers; to reduce participation costs for these suppliers; and to increase

competition among suppliers, which should also bring about costs reductions and better quality of goods and services acquired.

Federal Government agencies and organizations, as well as the suppliers of goods and services to the Federal Government (there are over 150,000 registered suppliers) are the main affected parties. Citizens and society are affected in the sense that e-procurement is intended to provide an instrument for social control of public expenditure, through its public transparency.

During COMPRASNET's first three years the Federal Government spent about US\$7m on system development and maintenance. During the first two years of on-line reverse auction use, the Federal Government is estimated to have saved up to US\$1.5m. Besides this positive return on investment, the system enables better and more transparent procurement, as well as reducing the red tape in the process. For example, a normal procurement process takes more than two months. The on-line reverse auction may be completed in less than 15 working days. The use of on-line procurement has also increased the participation of small businesses in government supplies.

There is insufficient data from a broad enough range of stakeholders to describe the system as a total success, but it can certainly be described as largely successful, bringing an estimated average 20% reduction of final price for goods and services acquired through reverse auction and price quoting. Suppliers also see it as successful due to it being linked to the financial payment system, guaranteeing timely payments on supplies sold to the government.

Enablers or Critical Success Factors were: *Political will inside government, External pressures particularly from citizens and citizens' groups for greater transparency and efficiency in government spending and acceptance by suppliers of transparency.*

Constraints/Challenges have been: *technological factor* causing temporary unavailability of the system at times, *legislative delays* caused by the need for new legislation and rules to allow for new forms of procurement. Recommendations are: *Get the technology right* – this system needed a robust platform, scalability, a basis in open systems, with heavy investment in back office sustainability. Secondly, *provide intense training* - this needed to cover users on both government and supplier sides. Lastly, *adopt a phased approach* - this project

worked well by having modules, the first of which could provide a demonstration effect for the system that would develop usage and interest.

2.3.3 E-Procurement in the Danish Public Sector - PePP

The model that the Danish public sector chose for the development and operation of their e-procurement system was based on the public-private partnership model. Public sector representatives specified and developed the PePP system in collaboration with private players. While the public sector paid for using the portal, financial risk was placed with a private entity, the gatetrade.net portal, which was in charge of operation and management of the portal. Founding partners of the gatetrade.net portal included TDC, a large telecommunications company; Danske Bank, a leading Danish Bank; Maersk Data, a subsidiary of Maersk industry group; and Post Denmark, which operates the national mail service.

When gatetrade.net was launched in 2000 it was introduced as a neutral third party marketplace. The four founders agreed upon placing all their procurement on the electronic marketplace in order to secure the necessary critical mass, both with respect to supply and demand. This, in principle, aimed to solve one of the classic problems of whether buyers or suppliers act as the catalyst for marketplace activity. However, one of the initial problems gatetrade.net faced was the limited number of willing suppliers. This is not surprising, given the general weakening in e-business during the launch of the marketplace. The strategy pursued by the consortia behind the PePP was to select a marketplace where there was a potential for business transactions from day one. This objective was to a large extent fulfilled, given that the PePP involved itself in an already ongoing electronic marketplace. Despite initially high expectations, adoption of PePP in public administration has been low.

The benefits derived from e-procurement (be it employed by private business or public institutions) include, a wider choice of suppliers, better quality, improved delivery, and reduced cost. In particular, we argue that the PePP enhances efficiency through two main avenues: (a) transaction cost savings and (b) reduced direct procurement costs.

Procurement carried out in an electronic marketplace helps in consolidating separate procurement functions on the state, county, and municipality level as illustrated in Figure 2-5.

Examples of procurement activities potentially undertaken by the PePP include bidding, ordering, and payment related data processing. This could lead to direct procurement cost savings in terms of human resources expenditures and administrative overhead in all three levels of public administration. In addition, demand bundled across public institutions reduces direct purchasing costs as public bargaining power increases relative to suppliers.

Positive side effects of the PePP include (a) increased transparency and accountability as electronically conducted processes allow better monitoring and benchmarking, and (b) easier, cheaper, and quicker exchange of information about prices and product offerings. This creates a huge potential for transaction cost saving in terms of reduced search, matching, and order control costs.

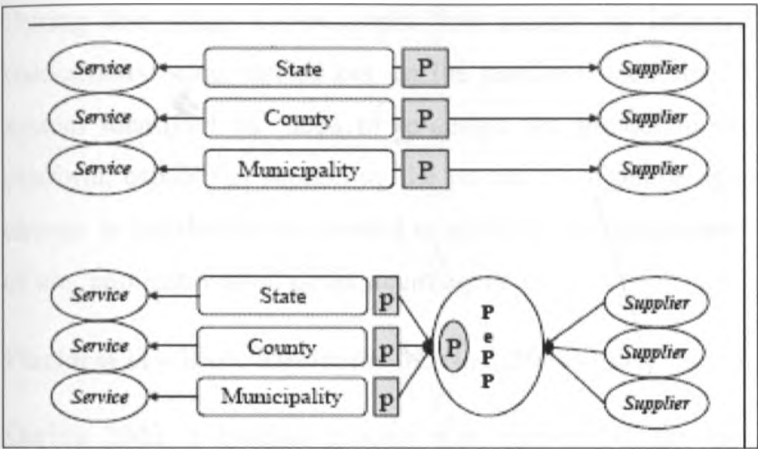


Figure 2-8: Efficiency Gains through PePP

2.3.4 ChileCompra of Chile

ChileCompra is the Chilean Procurement and Contracting System of goods and services for the public sector, a business marketplace managed by the Public Procurement and Contracting Bureau that allows state buyers to contact private suppliers. Its objective is to guarantee high levels of transparency, efficiency and use of technologies in the public procurement market, thus benefiting entrepreneurs, public agencies and the general citizenry.

Using a business model that privileges technological instruments and management excellence, ChileCompra has the following basic and strategic principles: free access to information, universality and non-discrimination

The mission of the Chilean Public Procurement System, ChileCompra, was to develop policies and initiatives that provide the public procurement market with greater transparency and added value by introducing technological instruments and management excellence.

ChileCompra has gone through various stages since 2003 —when it was basically an electronic “bulletin board” for public agencies wishing to publish their requirements— gradually becoming a full-fledged e-Marketplace.

This new electronic marketplace has been implemented step-by-step, and its operation has been largely outsourced to private businesses.

Platform I – Electronic Bulletin Board (2000-2003)

During this stage, ChileCompra was simply an informational bulletin board, with no transactions being carried out on the platform itself. In 2002, an evaluation of the entire system identified the need to re-design the project in order to create an actual trading platform, capable of supporting the purchases of public agencies. It was also decided that a change in legislation was needed to establish mandatory use of this new system and its rules of use, applicable to all government agencies.

Platform II – Basic Electronic Market (2003–2004)

During 2003, a bidding process was carried out for the construction and operation of ChileCompra’s new information system, which began implementation in February 2004. This platform allowed for a totally online purchase process, including: publication of calls for bids and the use of an electronic catalogue (framework agreements), publication and downloading of bidding terms and conditions, the bidding process, electronic emission of internal authorizations for different stages of the procurement process, questions and answers for each bidding process through a specific forum, emission of purchase orders by the buyers, reception of electronic invoices, and electronic payment.

Platform III – Electronic Trading Marketplace (2004–2007)

During this period, the following new applications and platform improvements were developed for www.chilecompra.cl:

- Pilot Electronic Payment (or e-Payment): allows for payment of invoices linked purchasing orders issued by public agencies through ChileCompra’s website.

- Inter-connectivity with the Official Electronic Registry of Government Suppliers, www.chileproveedores.cl.
- Website “ease-of-use”: new usability criteria have allowed for substantial improvement in user-system interaction, reducing processing times and contributing to lower costs in the procurement process.
- Electronic Tax Documents (DTE): new functions have been added to facilitate DTE management within the electronic platform.
- Integration Models: initial integration has begun between the public procurement platform and different management systems of other public agencies, seeking to produce automated, efficient internal processes and avoid redundant information processing.
- ChileCompra Express: a purchase mode featuring an e-Catalogue.

Platform IV – Advanced Trading e-Marketplace (2007–2010)

The new platform, called ChileCompra 2.0, includes a series of improvements to the services offered by the ChileCompra, ChileProveedores and ChileCompra Express websites, along with the creation of different applications that shall allow for substantial quantitative and qualitative advances, maintaining our reputation at the forefront of management excellence and State modernization, and adding value to the Chilean public marketplace. The main new functions include:

- Time zone management; Management for multiple organizational levels; Multi-language; Management of multiple currencies; Evaluation tables for advanced bids; Powerful search engine (with capabilities similar to Google); Budgetary management; Financial accounting system; Inventory management; Consolidation of inter-connectivity with other systems.

In the PPP model below of Chile, initially the software, hardware and operations was done by a single private operator while the government provided policy and regulations. One major problem was that the moment the Service Level Agreement (SLA) the contractor was not willing to improve the software.

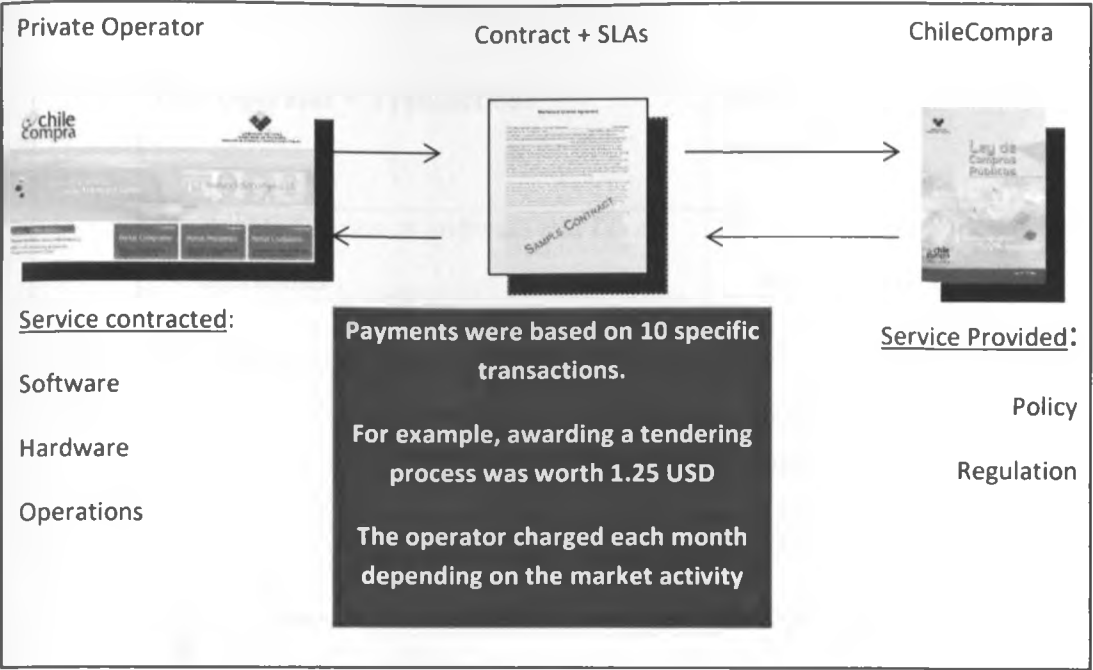


Figure 2-9: 2003 – 2008 Approach

They later changed and engaged more than one contractor/operator to provide for the data centre, contact centre and complementary software.

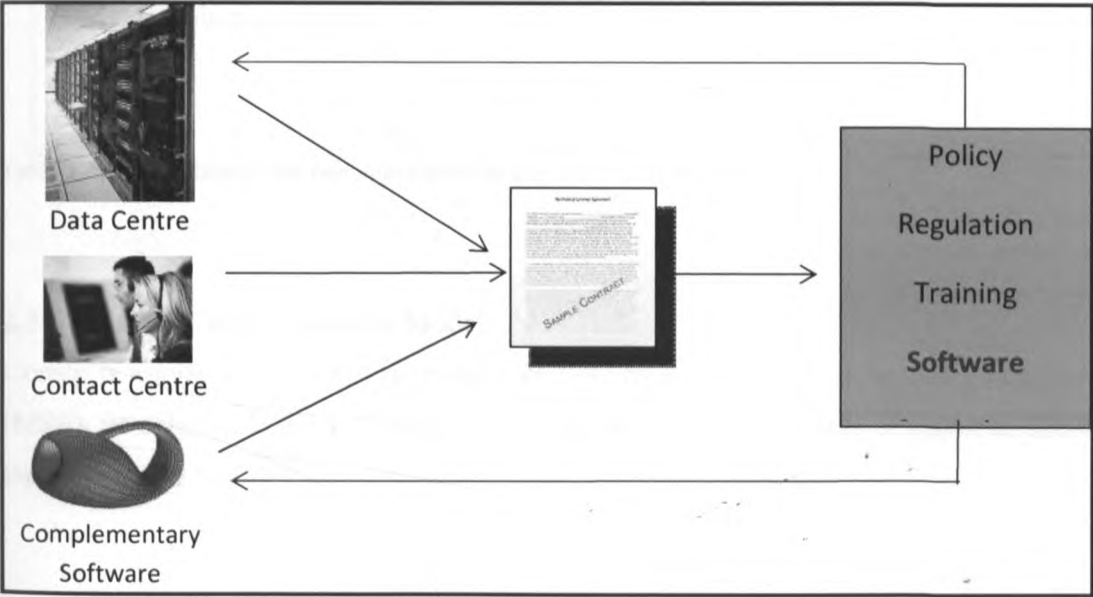


Figure 2-10: Current Approach

Table 2-2 below compares the strengths and weaknesses of the two approaches of the PPP model for the ChileCompra.

	One Operator + Transactions	Multiple Operators + Fixed Monthly
Strengths	☐ Private focuses on software and IT operations ☐ Public sector focuses on policy	☐ Procurement policy is better backed up by operations ☐ Flexibility for permanent service and applications improvement
☐ Both approaches are cost-equivalent		
Weaknesses	☐ Procurement policy and contract design were not always aligned ☐ Contract administration extremely expensive ☐ Lack of incentives for software improvements	☐ Public sector focuses on policy + software development ☐ Internalization requires a larger and more complex public agency

Table 2-4: Comparison of the two approaches to Compra e-procurement

2.3.5 Italian Consip Business Model

Consip is a joint-stock company, created in 1997 by the Ministry of Economy and Finance (MEF), its sole stockholder. Consip is entrusted with two main fields of action as shown in the Figure 2-12 below:

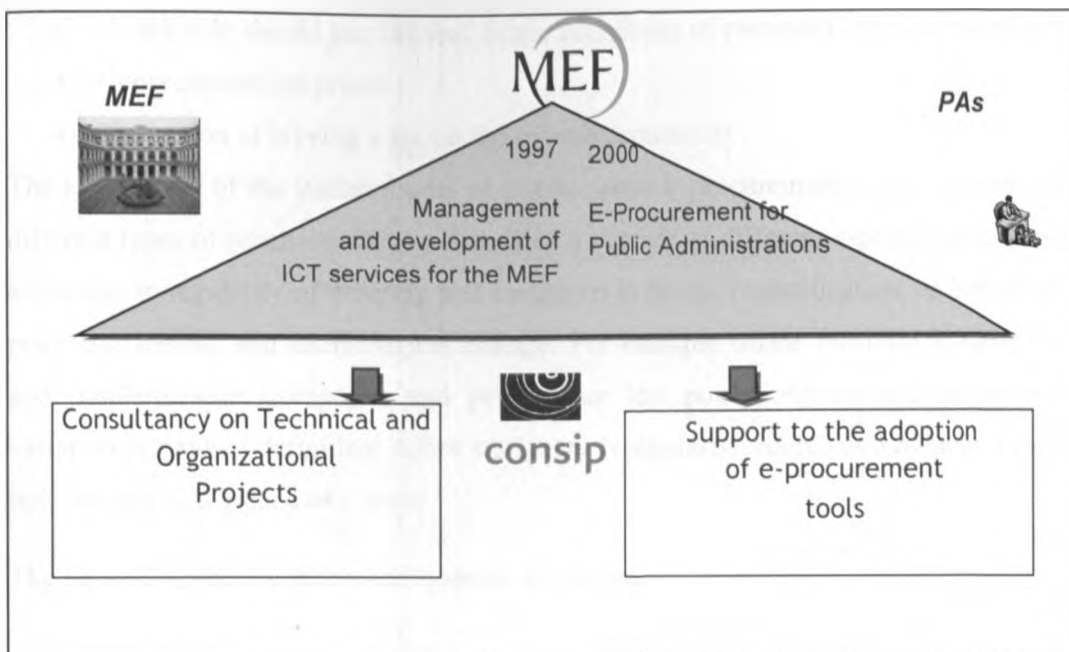


Figure 2-11: Consip's Fields of Action

The aims of Consip are manifold:

- To provide a better service for both buyers and suppliers
- To improve visibility and therefore accountability of public sector contracting
- To reinforce Italian government's commitment to the goals of e-Europe
- To minimize transaction costs through standardization

Why a joint-stock company?

- 1997: Need to hire (quickly) highly skilled professionals at market wages
- Flexibility in job assignments
- More suited for a company deemed to grow rapidly

However,

- Financial resources entirely provided by the Ministry on a cost-reimbursement basis according to a three-year plan of activities

Why a no-fee model for the e-procurement platform?

- On-boarding of both sides of the market of paramount importance (most SME's in Italy are in facto micro firms)
- Main objective was to get both sides of the market acquainted with e-procurement tools
- Large areas of the public sector still reluctant to use "anonymous"-purchasing tools

Any other model should consider:

- Which side should pay the fee? Some economics of two-sided markets would help
- Consequences on prices
- The option of levying a tax on the overall community

The key feature of the Italian model of public sector e-procurement is the differentiation of different types of purchase. Consip identified a number of different e-procurement scenarios, according to regularity of ordering and variations in product specification, as a result of, say, price fluctuations and technological change. For example office furniture is rarely ordered and standardizable; computers and printers are less rarely ordered and subject to great variation in product definition; office stationery is regularly ordered and of limited variation, but comprising a great many items.

The figure 2-13 below shows the systems developed.

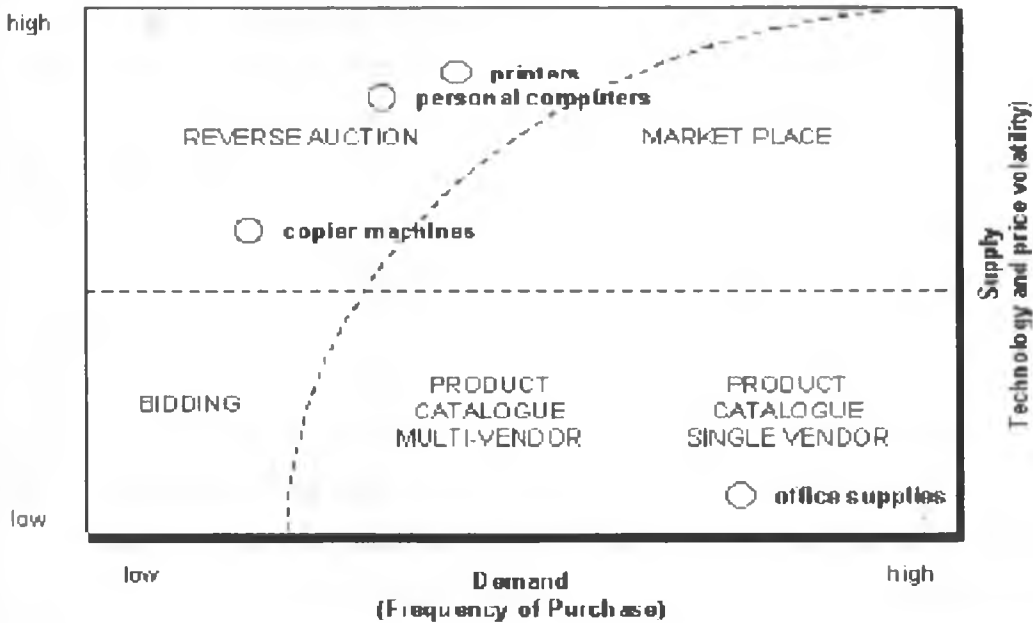


Figure 2-12: Different Systems of Consip (Source: Arthur Andersen, Booz Allen & Hamilton, Consip)

The likes of office furniture would be procured via a simple bidding system, not vastly different to the one previously in force. Computers and printers, with their high variations in price and technological specification, would be the subject of a reverse-auction, with interaction of supply and demand criteria on both price and specification whilst for office stationery approved suppliers will be able to post their catalogues on the net. For low value (less than €200,000), nonstandard items an electronic marketplace will be established for pre-qualified suppliers of high frequency, high specification volatility products. This model is

still being developed. For the time being, frame contracts have been constructed whereby ordering of individual products can be done quickly by individual public sector users.

2.4 Challenges of e-procurement

A number of recent studies have also looked into difficulties faced by firms in launching e-procurement and different authors have identified many challenges and drawbacks of e-procurement as follows:

Innovation & Information Consultants, Inc Concord, MA (2004) has noted the following challenges of e-procurement especially in relation to small firms;

- **Technological barriers**

Technological barriers represent obstacles to the adoption of e-procurement due to technological factors such as lack of high-speed connections and software incompatibility, integrating e-procurement with internal solutions and difficulties encountered in obtaining high speed access and download capabilities. Both cost and availability have a direct impact on a small firm's choice of access mode.

- **Market barriers**

Market barriers include those barriers that are external to the firm, and are driven by market forces (supply and demand) as opposed to other entities such as the government.

- **Regulatory**

Regulatory barriers include barriers created by governmental action or intervention in the market or action directly affecting electronic commerce including procurement.

- **Barriers unique to firm size**

Some barriers relate specifically to the size of the firm. For example, some small firms have concerns that the high cost of investing in e-procurement will prevent them from competing for such business. This "cost" is not necessarily large in absolute terms, but it is relative to any perceived benefits that small firms expect they will receive.

2.4.1 Legal Challenges and E-procurement

Ting and Wong (2003) have identified some of the legal challenges in e-procurement as shown in Figure 2-13. This follows a study they performed in the construction industry. They contend that the legal issues relating to e-procurement could be categorised into global trading, contract enforcement, liability risks, security breaches, and intellectual properties protection. The global issues involve lack of alignment, legacy variations (laws and



Figure 2-13: Potential Legal Issues/Challenges (Source: Ting and Wong, 2003)

2.4.2 Other Obstacles to e-Procurement Adoption

A study carried on the e-procurement adoption in the Local Authorities of the United States by Eric Prier and Clifford P. McCue found out the following obstacles which have also been identified in other countries:-

- High costs involved in implementing e-procurement procedures
- Resistance from different groups within the organization. Resistance may come from the ICT, Finance or even from top management
- Difficulty in integrating the e-procurement system to the existing financial systems within an organization
- Limited resources

- Human resource capacity or skills
- Too few suppliers to engage in the process
- Lack of supplier buy-in (rejection). Most suppliers fear sharing information with their competitors

Another study done by Prof. Hajah Rugayah for the State of Selangor Local Governments, Malaysia points out difficulty in effecting change, lack of funding, complex and expensive initiatives and rigid government bureaucracies as some of the factors for slow adoption of e-procurement.

CHAPTER 3 - RESEARCH METHODOLOGY

3.1 Introduction

This chapter seeks to explain how the research was carried out. This includes the method of selecting the sample from the target population, data collection methods and the limitations of the study. The study seeks to answer the following research questions:

- i. What are the e-procurement practices in the Kenyan public sector?
- ii. Which models are currently available and in use?
- iii. What factors affect the public e-procurement adoption in Kenya?

The major aim is to design a model that can be adopted by the public sector of Kenya to assist implement e-procurement.

3.2 Research Design

There are three main methods of research design. The first one is the exploratory research studies that are also", termed as formulative research studies" (Kothari, 2004). It is conducted when a problem has not been clearly defined or its scope is not clear and it enables the researcher to familiarize himself/herself with the problem at hand and to determine the research design, the method of collecting data and the selection of the subject. The second type is the descriptive design "...which is concerned with describing the characteristics of an individual or group" (Kothari, 2004). The third type is the causal research design that seeks to identify a causal relationship between an independent variable and a dependent variable. The relationship is suggestive but may not be proven by the research.

This study was descriptive in nature and survey method was used. A survey research attempts to collect data from members of a population and describes existing phenomena by asking individuals about their perception, attitudes, behaviour or values. Moreover, it explores the existing status of two or more variables at a given point in time. Primarily data collected from such a population or census is more reliable and up to date. The descriptive research was to enhance a systematic description that is as accurate, valid and reliable as possible regarding the e-procurement management practices by public institutions in Kenya.

This research is an academic project type whose objective was to establish and disseminate findings regarding e-procurement practice in the public institutions of Kenya, factors for slow adoption of e-procurement and the existing models that have been used to implement e-procurement around the world with the aim of coming up with a proposal of a model.

The methodology adopted included:

- Literature review of relevant issues of e-procurement in general; this includes study of online journals and papers
- Study of models in use world wide
- Survey and case studies about existing practice and problems among Kenyan public institutions
- The survey was conducted by use of questionnaires. Follow up interviews were also done for the questionnaire responses that were not clear to the researcher.

3.3 Research Population

The population of interest was the public sector which consists of the government ministries, local authorities as well as state corporations in Kenya. Of particular interest were the heads in both procurement and ICT departments. The sample was selected in a random fashion from the list of public institutions. Only one local authority was considered for the study due to its size and proximity to the researcher. The coverage included one hundred (100) institutions of which only fifty two (52) responded. The spread is as shown below:

Category of institution	No. of Questionnaires sent	Respondents
Central Government	40	35
Local Authorities	2	1
State Corporations	48	16
Total	100	52

Table 3-1: Population Spread

To evaluate the proposed model the project was done very closely with PPOA in order to get the real picture of e-procurement in the public sector of Kenya. The PPOA gave good and elaborate information as to where we are and where we are moving to being the lead agency in the quest to implement e-procurement for the public sector of Kenya.

3.4 Data Collection Methods

The main instruments of data collection used were through semi structured questionnaires targeting the ICT as well as procurement officers in the public institutions. . questionnaires were sent via mail, some hand delivered. This was followed by face to face interviews with a few respondents. The other instrument used was observation and documentary evidence.

3.5 Assumptions

The assumptions of the research were that:

- the sample population represents the targeted population.
- the respondents were truthful in their responses and gave the correct situation as on the ground.
- bulk of the procurement in central government is done at Ministry headquarters and so only respondents in Nairobi i.e. headquarters were considered.

3.6 Scope and Limitations

The research was only intended to be carried out from public institutions i.e. the central government, state corporations and local authorities. Only one local authority was chosen due to its proximity and size.

E-procurement is a wide and very complex area and research shows that it may not be suitable for all types of goods and services. Several systems have been combined in order to achieve the benefits. As such this research only considered consumables. The rest is left for further research.

CHAPTER 4 - RESEARCH FINDINGS

This Chapter presents details of the analysis and findings of the study. The research sought to find out the current practice as concerns e-procurement in the Kenyan public sector and the factors for slow adoption.

4.1 Situational Analysis

Majority of the public institutions have procurement departments and the head reports to the CEO of the organizations and are bound by the Public Procurement and Disposal Act 2005.

- ☐ Procurement is decentralized in that each procuring entity can advertise, evaluate and do all the other processes independently.
- ☐ Tenders are invited through advertisement in Daily Newspapers. Most institutions can now place the advertisements on the websites for information only and the subsequent processes are purely done manually.
- ☐ Websites are also used to get the best specifications available in the market.
- ☐ Major communications done via telephone. The figure 4.1 below shows the means by which the procuring entities communicate with their suppliers.

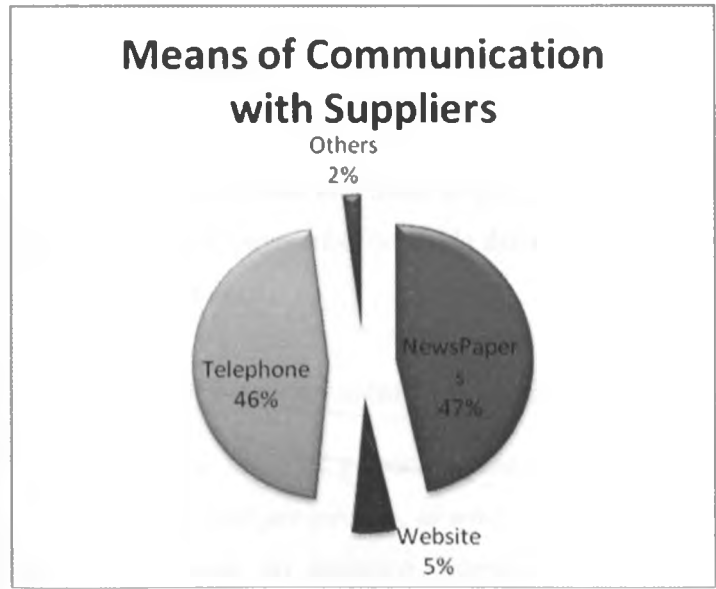


Figure 4-1: Means of Communicating with Suppliers

- ☐ Manual Sealed Bids are invited from Suppliers
- ☐ Manual evaluation process is conducted which is time consuming & prone to human errors
- ☐ Tenders are awarded in an offline manner
- ☐ Tenders are tracked in an Offline Manner. Figure 4.2 below shows how the internet is used. Business transactions rank the lowest. The only use in most entities is limited to search for specifications and price ranges of goods to help give price guidelines.

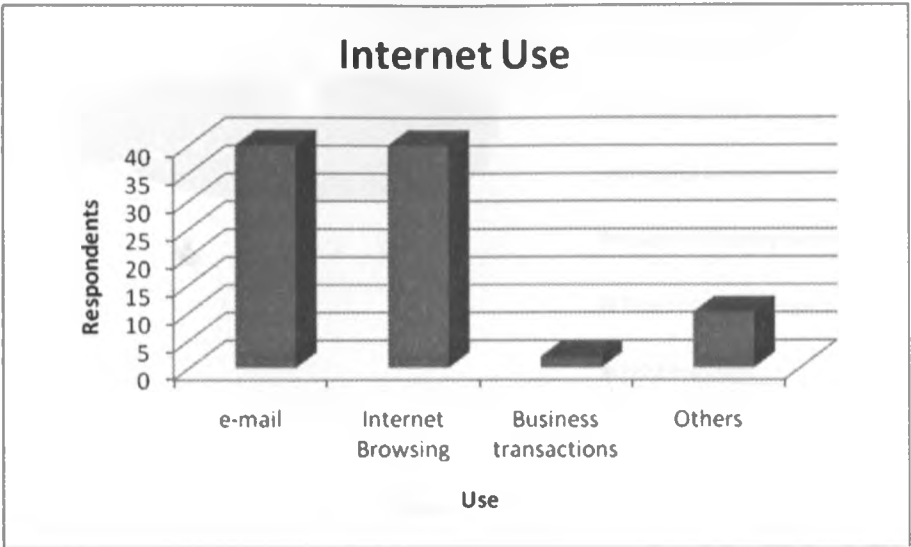


Figure 4-2: Internet Use

Only one institution has an online enquiry form for suppliers to help them know or enquire the status of their payments for goods delivered and of the services rendered. The rest of the processes are manual.

4.2 Level of e-procurement Use/Adoption

To find out the level of e-procurement use/adoption, the question asked was “From an organizational level perspective, to what extent has your organization used e-procurement?” The options were: no intention, intention in the future, pilot use, commitment, limited deployment, general deployment, and rejection after evaluation and no response. From the respondent who answered this question the research shows that many of the institutions are willing to and have the intention to implement e-procurement because of the benefits that can

be achieved from the same. Only a few are not willing to venture into this new way of doing business.

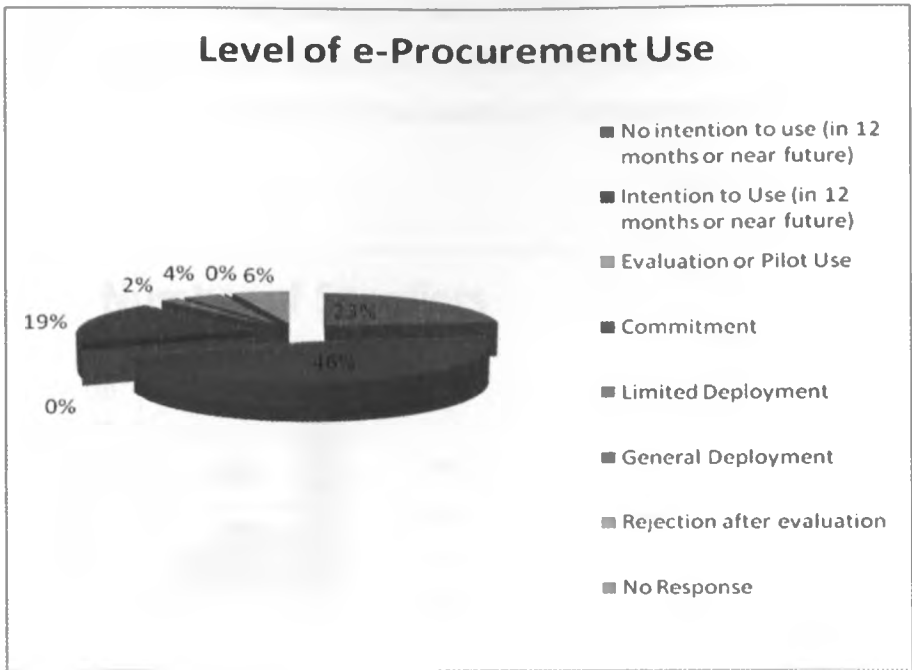


Figure 4-3: Level of e-Procurement Use/Adoption

The figure above (Figure 4-3) shows that 46% of the respondents are willing to implement e-procurement in the near future or in the next 12 months. Only a meagre 2% have no intention of implementation in the near future. 6 % of the respondents did not give an answer as to the position of their organization as concerns implementation of e-procurement.

There was conflicting results from one Local Authority as regards the level of use of e-procurement. ICT department is very optimistic in regards to the implementation of e-procurement. On the other hand, the procurement department feels that it will never be possible to implement e-procurement for the council.

4.3 Number of Suppliers as a factor for slow Adoption

From the literature review it was found out that the number of suppliers an organization deals with also has an impact on the adoption of e-procurement. Those organizations with fewer suppliers did not find it viable to implement e-procurement. The study carried out in the

public sector of Kenya shows that this may not be a big issue because as can be seen from the results of the survey

Few suppliers mean that the venture is not viable. It also emerged that many suppliers are not willing to trade online because of fear of leaking out their secrets to competitors. Figure 4-4 below majority of the institutions are dealing or trading with a good number of suppliers. So this factor is not a hindrance to the adoption of e-procurement in the public sector of Kenya.



Figure 4-4: Number of Suppliers Doing Business with the Institutions

4.4 Perceived Benefits of e-procurement

It was also important to find out if the perceived benefits of e-procurement are known. To find out how the ICT and Procurement Heads felt on the perceived benefits of e-procurement a table having some of the benefits was provided for the respondents to rate some of the benefits with 1 being high rating and 5 the very lowest rating of that particular benefit.

	Very High	High	Average	Low	Very Low	
Benefits	1	2	3	4	5	Total
Increased speed, accuracy and transparency	<u>69.23</u>	23.08	7.69	0.00	0.00	100
Wider competition and economy in procurement	46.15	38.46	15.38	0.00	0.00	100
Simply standardize and speed up process and communication	30.77	<u>61.54</u>	7.69	0.00	0.00	100
Real time generation of detailed and accurate data with seamless sharing of information	<u>76.92</u>	23.08	0.00	0.00	0.00	100
Facilitate online pre/post contracts management including bidding, contracting, ordering and payments with online participation of all stakeholders effectively	46.15	46.15	7.69	0.00	0.00	100
Reduction in cost of Government purchases with reduced cycle time	<u>61.54</u>	30.77	7.69	0.00	0.00	100
Elimination of delays arising out of processing tenders manually	61.54	38.46	0.00	0.00	0.00	100
Reduced paperwork and lower transaction costs	<u>76.92</u>	23.08	0.00	0.00	0.00	100
Problem of mafia and cartel formation arrested.	30.77	30.77	23.08	0.00	15.38	100
Database on goods, services, works and contractors get build up	<u>69.23</u>	23.08	7.69	0.00	0.00	100
Economy of scale achieved by aggregation of requirements	<u>61.54</u>	30.77	7.69	0.00	0.00	100
Quick and easy access to all information and audit trails	<u>61.54</u>	30.77	7.69	0.00	0.00	100
Better monitoring of tenders and contracts	<u>69.23</u>	23.08	7.69	0.00	0.00	100
Reduced inventory costs	<u>53.85</u>	38.46	7.69	0.00	0.00	100
Improved work culture in the department and better vendor relationship	38.46	15.38	38.46	7.69	0.00	100
Winning the trust of people and good governance	30.77	30.77	23.08	15.38	0.00	100
Ability to link directly into existing systems, such as ERP	<u>61.54</u>	30.77	7.69	0.00	0.00	100
Self-invoicing on behalf of clients can add to the bottom line.	46.15	30.77	23.08	0.00	0.00	100
Month-end reconciliation can end the problem of the wrong items being ordered or the wrong price being offered as business processes have been streamlined and all are working off the same catalog	38.46	38.46	15.38	7.69	0.00	100
Maverick spending is reduced	46.15	23.08	15.38	15.38	0.00	100.00
Total Average	53.94	31.73	11.83	2.69	1.25	100.00

Table 4-1: Perceived Benefits of e-Procurement

The above Table 4-1 shows the percentage of the perceived benefits of e-procurement by respondents. From the above figures it can be deduced that the public sector are aware what e-procurement is all about and the benefits they can derive from it. Real time generation of detailed and accurate data with seamless sharing of information leads with a percentage of 76.9%. Most of the respondents have realized what benefits can be derived from the implementation of e-procurement in their institutions.

4.5 Factors for Slow Adoption

To help find out the major factors for slow adoption a lot was borrowed from the literature review to help design the question used to collect this information. Some of the factors for this included:

- Lack/poor infrastructure. Use of IT and access to the internet is a prerequisite for e-procurement. The infrastructure in Kenya is still undergoing a lot of development and has only covered a small percentage of the country and especially in the urban areas. Even the urban centres covered still experience a lot of disparities and very slow connectivity.
- Lack or no known appropriate model or technical standards.
- High costs involved in implementing e-procurement procedures. These include cost of both hardware and software, infrastructure development costs, firm wide training in both technical and other business process areas. There are also costs associated with system migration and integration, data conversion and many other issues such as awareness for all the stakeholders.
- Resistance from different groups within the organization. Resistance may come from the ICT, Finance or even from top management. Lack of support and resistance to change is one of the biggest barriers to the introduction of e-procurement. Cultural change is required prior to adoption. Champions need to be appointed with full senior management support. Organizational culture and upper management support range from complicated procedures and extended relationships, excessive state intervention, bureaucratic dysfunctions, absence of clear national IT policy, large volume of paper, lack of flexible central control.
- Difficulty in integrating the e-procurement system to the existing financial systems within an organization

- Limited resources in terms of skilled personnel.
- Too few suppliers to engage in the process
- Lack of supplier buy-in (rejection). Most suppliers fear sharing information with their competitors

		Very High	High	Average	Low	Very Low	No Response	Total
1	Poor/ Lack of Infrastructure	7.69%	15.38%	<u>30.77%</u>	15.38%	15.38%	15.38%	100.00%
2	No Appropriate Model	23.08%	15.38%	<u>30.77%</u>	7.69%	15.38%	7.69%	100.00%
3	Integration with Back Office Systems	<u>30.77%</u>	15.38%	<u>30.77%</u>	-	15.38%	7.69%	100.00%
4	Security Issues	15.38%	-	15.38%	30.77%	30.77%	7.69%	100.00%
5	Lack of appropriate technology	15.38%	7.69%	7.69%	23.08%	30.77%	15.38%	100.00%
6	Lack of Technical Standards	7.69%	7.69%	15.38%	23.08%	30.77%	15.38%	100.00%
7	Lack of support	<u>46.15%</u>	23.08%	7.69%	7.69%	-	15.38%	100.00%
8	Lack of awareness	<u>46.15%</u>	30.77%	-	15.38%	-	7.69%	100.00%
9	In Planning stage	15.38%	7.69%	23.08%	23.08%	23.08%	7.69%	100.00%
10	Lack of funds	-	-	23.08%	23.08%	23.08%	<u>30.77%</u>	100.00%
11	Regulatory	23.08%	-	38.46%	-	23.08%	15.38%	100.00%
12	Too few suppliers	-	14.29%	-	28.57%	42.86%	14.29%	100.00%
13	Size of Organization	-	7.69%	7.69%	30.77%	38.46%	15.38%	100.00%
14	Organization Culture - Resistance	23.08%	7.69%	30.77%	15.38%	15.38%	7.69%	100.00%
15	Lack of suppliers willing to corporate	-	-	-	38.46%	38.46%	23.08%	100.00%
16	Lack of skills	7.69%	7.69%	7.69%	15.38%	<u>46.15%</u>	15.38%	100.00%
	Total Average	16.35%	10.03%	16.83%	18.61%	24.31%	13.87%	100.00%

Table 4-2: Factors for Slow Adoption of e-Procurement in the Public Sector of Kenya

CHAPTER 5 – THE PROPOSED MODEL

This chapter presents the proposed e-procurement model for the Public sector of Kenya and the stakeholders.

5.1 Stakeholders

Stakeholders for e-procurement can include any person or functional area that “touches” the systems, tools or information being held. This may include people at all levels of the organization and external contacts such as suppliers and outsourcing partners. An agency can have multiple procurement processes. It may be necessary to look across the entire agency for the appropriate stakeholders and to understand the different business rules as they apply to e-procurement.

The key players are:

- ☐ PPOA-
 - ❖ Responsibility as the lead agency.
 - ❖ Enhance its oversight role
 - ❖ Provide information on procurement activities
- ☐ Central Government for policies and regulations, human resource, software, infrastructure
- ☐ State Corporations and Local Authorities as users
- ☐ Suppliers and business communities
 - ❖ Register as suppliers
 - ❖ Access public tenders and Quotations
 - ❖ Offers goods, works and services to government
 - ❖ Access public procurement information
- ☐ General Public
 - ❖ Procurement Policies and procedures
 - ❖ Tender advertisement and awards
 - ❖ Lodge Queries and suggestion on the public procurement process

Figure 5-1 below gives a summary of the stakeholders of the e-procurement system. The financiers include Ministry of Finance, Ministry of Planning and Vision 2030 and the donors. Donors are the major financiers of major projects in the public sector of Kenya.

Users include the procurement entities such as the government ministries, state corporations, local Authorities as well as top management. E-government include the Directorate of E-government, Ministry of Information and Communication which are mainly concerned with providing the required infrastructure , the necessary advice in terms of ICT related issues as well as policy guidelines and standards. The marketplace includes the involved consultants, Kenya Private Sector Alliance (KEPSA) and the suppliers/business community.

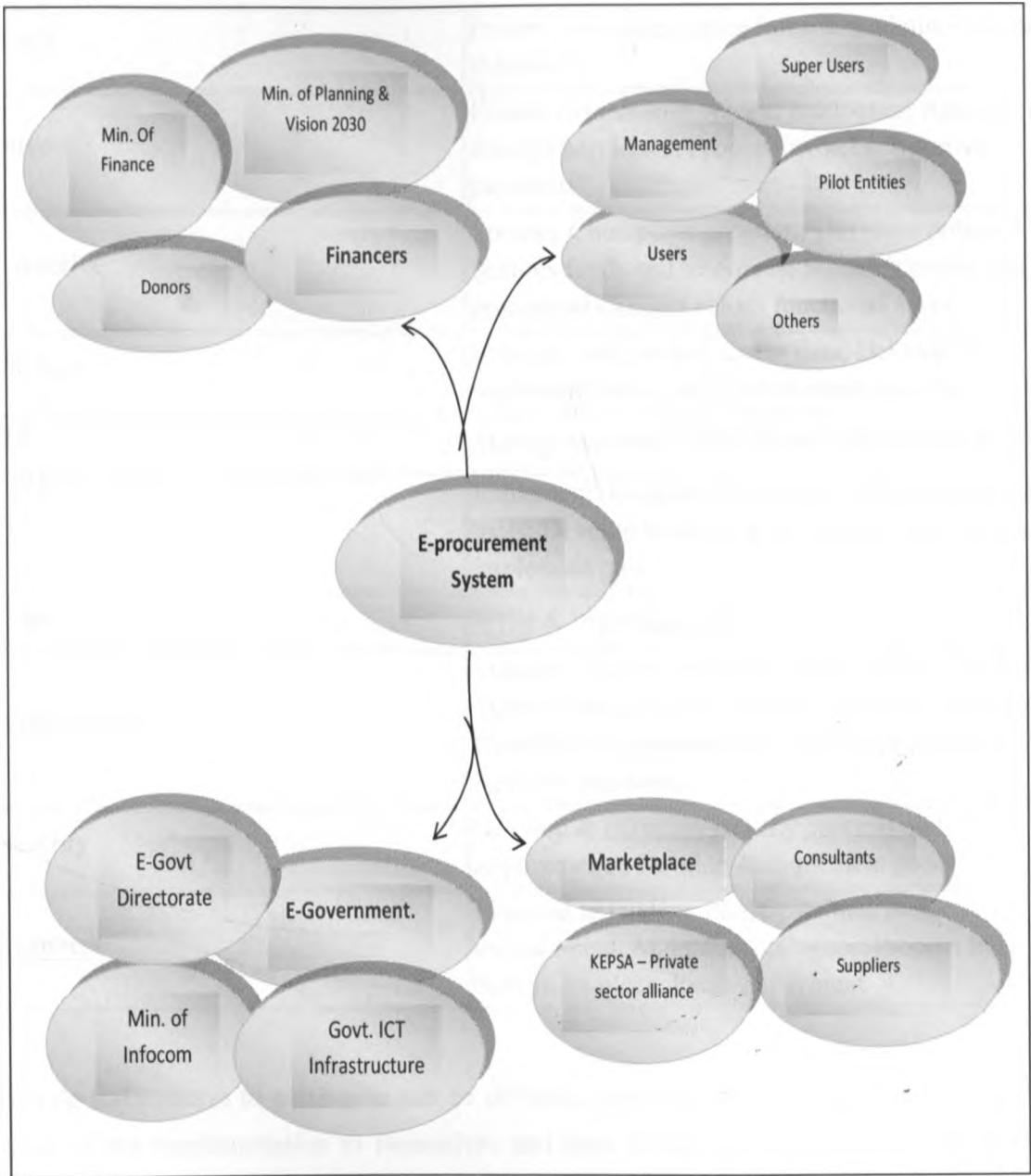


Figure 5-1: E-Procurement System Stakeholders

Functional Procurement Areas and Their Activities

The following list provides a general view of functional procurement areas and their potential e-procurement-related activities:

Area	Activity
Accounts Payable	Maintain catalogues; Process POs; Process invoices
Audit	Ensure laws, rules, procedures & guidelines are being followed
Buyers	Initiate requisitions; Access catalogues; Approve POs; Receipt purchases; Process invoices; Approve payments
Executive	Sponsor Champions program; Develops policy; Secures funds and resources; Supports policy and procedural changes across functional areas
Finance	Manage, analyze and access data; Develop & implement policy and procurement processes
HR	Manage personnel data; Develop & implement policy
IT	Implement, integrate & manage online systems and tools; Develop business & IT strategy; Develop & implement policy
Legal	Write & negotiate contracts
Procurement	Manage, analyze and access data; Source suppliers; Manage categories & contracts; Manage suppliers; Develop & implement policy; Receipt purchases; Approve payments
Security	Develop & maintain security controls and requirements; Develop & implement policy
Suppliers	Respond to tenders; Participate in contract management; Maintain catalogues; Process POs; Submit invoices; Receive payment

Getting stakeholders to participate can be difficult, especially if they do not understand the value of the implementation to themselves and their functional area. A successful model includes a cross-functional team with members from each functional area. The team can support testing, development and training activities along with acting as product champions.

Well-managed stakeholders at different levels in the organization can positively impact an e-procurement implementation in the following ways:

- key stakeholders can spear head the change management activities by reviewing and updating local policy and procedures that will be impacted by the implementation
- the Executive Sponsor can ensure new policies and procedures are put into place to drive uptake of e-procurement
- managers who are committed to e-procurement can encourage its use on a daily basis
- end users prefer a fast and easy procurement process. Their roles and responsibilities often change as a result of e-procurement. Changes should be identified and communicated in a way that allows them to absorb the information and incorporate it into their work processes.

5.2 The Public E-Procurement Model

The proposed e-procurement model for the Kenyan public sector is drawn from the existing models being used by other countries in consideration to those factors that have led to slow adoption discussed in chapter 4. In the Literature Review of Chapter 2, the models in use are:

- The Public model or Government owned. This is where all the e-procurement tasks (notifications, dissemination of notice and tender documents, submission of bids, e-catalogues, reverse auctions e.t.c.) are run by the government which also incurs all the investment risks of the project
- Private model. This is where e-procurement tasks are run by private entities which bear all the investment risks of the project. It requires more profound legislation changes because the entity that manages the e-procurement system can perform and is called upon to do so some acts that are traditionally reserved by the law to the public entities. It can also raise some legal problems in terms of public law and in terms of competition.
- The Public/Private Partnership (PPP) model

The factors that have led to slow adoption of e-procurement in the Kenyan public sector include:

- Poor/lack of infrastructure
- Integration with the existing financial systems
- Management issues such as lack of support
- No known appropriate model
- Lack of awareness among the staff
- Regulatory Issues
- Organizational culture e.g. resistance

The model chosen is the public or simple the buyer centric model. This model where the buyer or in this case the government source for goods and services from several suppliers and all the investments risks and benefits are taken by the government. This model offers several advantages which include:

- Tight control can be maintained over the procurement software, catalogue (and other procurement) data and processes.
- Can be highly customized to meet the buying organization's needs.
- It is much easier for procurement solutions to be fully integrated into the buying organization's financial management system so that the whole procurement process can be easily automated from beginning to end.
- Offers the most procurement functionality to buyers.
- Faster and easier product search and selection.
- Access to multiple suppliers' offerings and can usually do price comparisons.
- For suppliers the costs of participating in these systems are generally low, there are no transaction fees and most require little investment from suppliers to participate.

This type of model is designed predominantly to serve the needs of the buying organization. In a buyer-centric or buyer-managed purchasing model, the buying organization implements software to support its procurement processes, obtains catalogue data from its contracted suppliers, and aggregates the catalogue data into a single internal catalogue for the use of its purchasing officers. One of the major advantages of this model is that profound legislation changes are not necessary

A diagrammatic representation of the model is as shown in Figure 5-2 below.

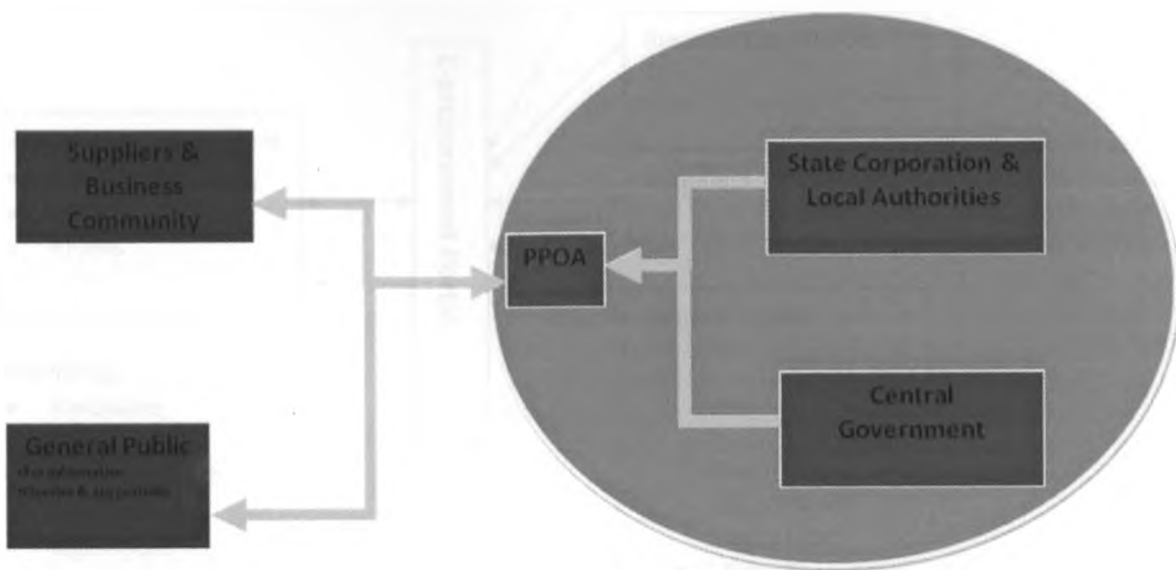


Figure 5-2: Proposed e-Procurement Model for Kenyan Public Sector

In this model all the investments, risks and tasks are taken by the govt through PPOA. The services are provided by PPOA for free which is financed through the national budget and donors. The government is providing the policy and regulations required in public procurement. The software provided by PPOA, infrastructure and part of hardware by central government. PPOA is to provide the required basic and advanced training as well.

Suppliers and Business Community

Business community is included because they need to get access to public procurement information, make queries/complaints, to register as suppliers, offer goods, works and services. Suppliers appreciate e-procurement when it reduces data entry, improves data quality and streamlines their work. By communicating objectives and strategy and understanding suppliers' concerns it is possible to ensure suppliers' continued support.

General Public

The general public is in the model because they are also a major stakeholder in the system. They need information, lodge queries and make suggestion which can improve the procurement process. They also need to access the tender advertisements and awards and get to know the procurement policies and procedures i.e. for information.

Unlike private sector procurement, public sector procurement must work within regulations and policies established to accomplish desirable social as well as economic goals. Public sector procurement emphasizes inclusiveness and broad competition instead of using a small number of suppliers with whom a trusted relationship has been established.

One of the most serious inclusiveness issues facing government procurement offices today is minority business owner's digital divide. As many businesses tap the promise of the Internet, minority entrepreneurs or SMEs [many small business owners] are struggling to harness the power of computers, information technology and e-commerce.

One of the major factors hindering adoption of e-procurement is integration with existing systems within the organizations. There are several ERPs/financial systems that have been implemented across public sector that requires to be integrated to the e-procurement system. Some of these include Oracle Financials for the Central government, SAP implemented by Kenya Power and Lighting Company and several other systems. This model offers the following advantages:

- Tighter controls can be maintained over the procurement software, catalogue (and other procurement) data and processes.

- The solutions can be highly customized to meet the buying organization's needs.
- These systems also make it much easier for procurement solutions to be fully integrated into the buying organization's financial management system so that the whole procurement process can be easily automated from beginning to end.
- They tend to offer the most procurement functionality to buyers.
- They lead to fast and easy product search and selection.
- They give buyers access to multiple suppliers' offerings and can usually do price comparisons.
- For suppliers the costs of participating in these systems are generally low, there are no transaction fees and most require little investment from suppliers to participate.

Most of the suppliers dealing with the public sector are small and medium sized enterprises (SMEs) who have not established themselves well. They need a lot of support from the government in terms of provision of infrastructure. Left on their own they may not be able to develop their own portals to enable them compete effectively with the few large enterprises.

Implementation of any system requires a lot of both technical and basic training. The government offers training at all levels and always invests a lot of funds in training of staff.

5.2 Proposed Network Design

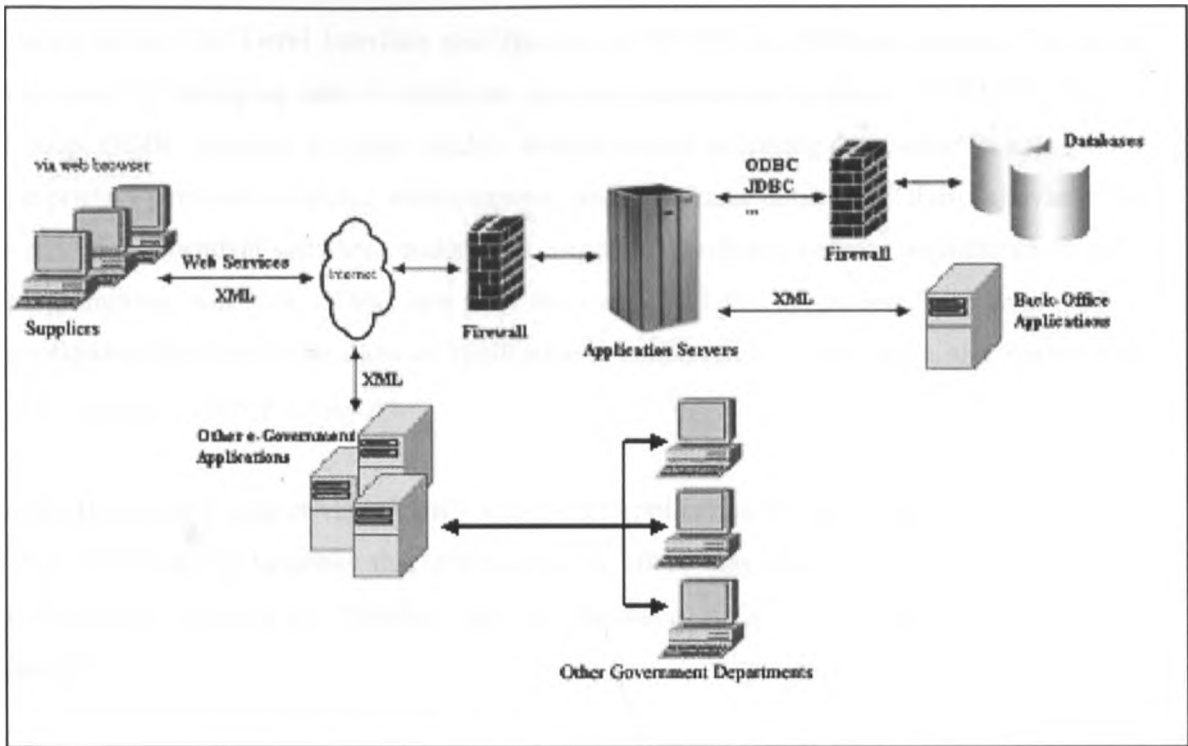


Figure 5-4: Proposed Network Design

Figure 5-4 above is a proposed network design for the e-procurement system. There are currently other database systems running in government such as the Oracle database for the IFMIS system. This database is centralized and accessible to government ministries and is yet to be rolled out to the provinces and districts within the country. The procurement system needs to be integrated to this system for the purposes of sharing information. There is also one state Corporation that is currently running SAP system as an ERP.

Other government applications may be specific to a certain procuring entity or a government wide system. One of such systems includes the National Registration system, Companies Registry system, Lands Registry system. These systems may not have a direct link to the e-procurement system.

Extensible Mark-up Language (XML) is a general-purpose specification for creating custom mark-up languages. It is classified as an extensible language, because it allows the user to define the mark-up elements. XML's purpose is to aid information systems in sharing structured data, especially via the Internet.

Open Database Connectivity (ODBC) is Microsoft's strategic interface for accessing data in a heterogeneous environment of relational and non- relational database management systems. Based on the Call Level Interface specification of the SQL (a database computer language designed for managing data in relational database management systems (RDBMS)). Access Group, ODBC provides an open, vendor- neutral way of accessing data stored in a variety of proprietary personal computer, minicomputer, and mainframe databases. ODBC alleviates the need for independent software vendors and corporate developers to learn multiple application programming interfaces. ODBC now provides a universal data access interface. With ODBC, application developers can allow an application to concurrently access, view, and modify data from multiple, diverse databases.

Java Database Connectivity (JDBC) is an API (Application Programming Interface) for the Java programming language that defines how a client may access a database. It provides methods for querying and updating data in a database. JDBC is oriented towards relational databases

5.3 Security

The other important point in the design of the e-Procurement system is to ensure security because electronic public procurement process will only be reliable when confidential information can be transmitted safely and securely between all involved parties. For the government, e-Procurement security is essential to achieve online tendering, to enable enhancements like comparison among the goods and services and to ensure that the e-Catalogues are valid. Also, the dissemination of e-Procurement application depends on reliable security. Therefore, to provide confidence and privacy, to prevent damages and to ensure trusted distribution of services, some level of security should be provided. In this respect, the government has to make commitment to provide secure communication in e-Procurement application.

Threats to the e-Procurement system can be in the form of resource stealing, tampering with documents and leakage of information. These can be done by the following methods:

- Obtaining data without authority

- Using another users' identity without authority
- Intercepting and altering messages

From the government's point of view, in e-Procurement system, a different and digital identity should be assigned to everyone who wants to do business with the government. Therefore, as a general solution, data are required to be encoded and users are required to be authenticated so that data can only be understood by the authorized users.

In this proposed architecture, Public Key Infrastructure (PKI) can be chosen as a security technology. PKI involves one public and one private key to encrypt and decrypt data. If data were sent to a user using PKI, it would be encrypted using the public key. To decode the data, the recipient uses the private key, which is stored on his/her computer. Also, secure socket layer (SSL) can be preferred to transfer data securely between users and the servers.

Besides PKI and SSL, to separate public and private networks for preventing unauthorized access to databases, firewalls are used both in front of the databases and before accessing to the Internet where private sector is located and Intranet where public sector is located.

CHAPTER 6 - CONCLUSIONS AND RECOMMENDATIONS

6.1 Review of Work Done

The objectives of the project sought to find out the current practice of e-procurement in the public sector of Kenya, identify factors for slow adoption of e-procurement, study the available models and finally design a model that can be adopted.

E-procurement is in the political agenda in Kenya. Most public institutions in Kenya are willing to implement e-procurement and are very much aware of the benefits they can get from adopting e-procurement. Despite the importance of and perceived benefits of e-procurement and the government's effort in pushing the use of ICT, the adoption rate of e-procurement is very low in Kenya. The research found out that that the public sector is very much willing to adopt e-procurement, but are not yet there.

The procurement processes are still manual and some of the factors for slow adoption include inadequate infrastructures, lack of awareness and top management support, inadequate skills and human capacity, lack or unknown appropriate model and technical standards, resistance from different groups, difficulty in integrating e-procurement system with the existing systems. Though implementing e-procurement can be very complex and costly, the research indicates very low response from the respondents. This suggests that the cost implications are not known or not clear to the respondents.

Despite the above factors most institutions are aware of the benefits that can be gained from the implementation of e-procurement

During the research some institutions were very reluctant to give information and several visits had to be made to such institutions before they could open up to fill the questionnaire or give a chance for a short interview. The other issue was that the top management in procurement were reluctant to participate, but referred the researcher to the other staff.

6.2 Conclusion

When launching an e-procurement strategy, there are certain key points that need to be taken into consideration before it can be rolled out. E-procurement is about much more than just buying online – it is also about improving the purchasing process and reducing costs. And this holds true in any line of business, whether one is procuring office supplies, maintenance,

repair and operation products, or IT equipment, software and services. By introducing e-procurement, time consuming and repetitive routines can be reduced and businesses can control 'maverick buying' outside of negotiated contracts or 'preferred supplier' agreements.

E-procurement is still a relatively recent phenomenon and is not as simple as just providing a catalogue on the internet. As with every business decision, a sound strategy has to be in place to ensure needs are met and the best possible processes are in place. Three pre-requisites need to be fulfilled before implementing any e-procurement solution. These are technological maturity; business adoption and consumer/supplier buy in.

Technological Maturity

Traditionally, procurement professionals have too often been focused on processing and administrative tasks rather than 'strategic' procurement activities. With e-procurement in place, organizations have a chance to integrate these systems with other back-office systems to ensure that activities are no longer duplicated and that processes run more smoothly, achieving an overall improved efficiency – meeting the objective of saving time and money. Looking at the advantages of implementing e-procurement systems, one wonders why every organization is not switching over. There is one very easy explanation – technological and business maturity. Many suppliers are not ready to trade electronically, as their IT infrastructure is not yet well developed enough. Although web-based trading is becoming more common, not every business seems to have realized the enormous potential.

Business Adoption

It is important to appreciate that e-procurement is not an end in itself, but should form part of an overall procurement strategy. If an organization makes the decision to introduce e-procurement, it is not enough to just implement a new system. It should be part of an organization-wide change programme. It is an opportunity to focus the whole organization – and procurement professionals – away from purchasing administration and onto adding value and core activities.

Customer/Supplier buy-in

In order to make e-procurement initiatives work, organizations need buy-in from their suppliers. Suppliers need to see a value proposition and, at the moment, too many companies only focus on commodity cost, ignoring quality and services. E-procurement should enable suppliers, where possible, to use their existing e-commerce infrastructure, not requiring them

to invest in a new one solely to support an organization's e-procurement initiative. Key suppliers must be involved from a very early stage of the project. Without their acceptance and involvement, the project will most certainly fail.

Another crucial factor is which business model to choose. There are several e-procurement models available, such as buy-side applications and e-marketplaces. The model chosen must fit the business' procurement strategy and objectives. It is therefore essential to conduct research into the company's individual needs to avoid duplication of efforts or the project ending in failure.

There are a large number of system vendors on the market, so checking their track record and experience is critical. As e-procurement is a relatively new area, businesses are advised to examine vendor's reference sites. Has the application been fully rolled out, or are the projects still in pilot phase?

Content management is key to a successful e-procurement project and encompasses more than just uploading spreadsheet data. Even the best e-procurement platform will be compromised if the catalogue content is not properly cleaned, structured and maintained. Items must be easily found by the end-user and the system must enable the searching of items across suppliers in a consistent way. A further consideration is how the e-procurement application will handle product name synonyms and variations between different suppliers.

However, taking all these factors into account will enable an organization to successfully plan and implement an e-procurement solution to meet its needs. These guidelines are valid for both the private and the public sector, as both sectors need to consider the same pre-requisites when implementing e-procurement solutions and very often face the same problems. There are, though, a few very important differences.

While private businesses are not bound to any policies but their own, public organizations have to comply with the government's guidelines. For example, there may be policies regarding sourcing, where possible, from local or small businesses. There are huge benefits to be reaped from investing in a carefully planned, well-chosen, and wisely implemented e-procurement solution, whatever the organization's business area. However, the e-procurement initiative must form part of a considered e-business strategy, taking into account

the entire company's processes and procedures, as well as supplier and customer relationships.

6.3 Recommendations and Further Research

E-procurement is not suitable for all goods and services to be purchased within an organization. It is recommended that a further study to be on how institutions can deploy e-procurement for all types of goods and services so as to reap the full benefits of the system.

The model chosen is rarely used and reasons needs to be sought as to the reasons why this is so. Most countries studied have implemented e-procurement used the PPP model which they kept on reviewing to suit their needs.

It is also recommended that a further research be done where a single system can be used to implement e-procurement. Many cases cited in the literature shows that several systems are used making integration such a big issue for implementation.

REFERENCES

1. A Green Paper on e-Commerce for South Africa, Department of Communications Republic of South Africa (Nov. 2000)
2. Antonio Davila (2002). Moving e-Procurement Systems to the Internet: The Adoption and Use of e-Procurement Technology Models. www.Olin.wustl.edu
3. Babbar, S., Prasad, S. (1998), "International purchasing, inventory management and logistics research: an assessment and agenda", International Journal of Operations & Production Management, Vol. 18 No.1, pp.6-36.
4. D. Neef (2001), "e-Procurement". Pearson Education
5. E. Cloete (2003), Retail Organizations in the Western Cape: Acceptance and Adoption of e-Procurement, www.stuartwarden.com/
6. Efraim Turban, D King, D. Viehland, Jae Lee (2006). "Electronic Commerce, A Managerial Perspective". Prentice Hall
7. Gunasekaran A., Erick W. T, International Journal of Production Economics. Volume 113, Issue 1, May 2008, pages 159-126
8. Helle Z. H. and Volker M. E-procurement Adoption in the Danish Public Sector: The Influence of Economic and Political Rationality. <http://www.e-sjis.org/journal/volumes/volume17>
9. Jeffrey F. Rayport, Bernard J. Jaworski (2002). "Introduction to E-Commerce", Tata McGraw-Hill
10. Joyce Chan (2002). "Factors Leading to Electronic Procurement Adoption Behaviour" AM CIS 2002 Proceedings, Page 317, <http://aisel.aisnet.org/amcis2002/317>
11. Kalakota R. and Robinson M, 2001 '*e-Business: Roadmap for Success*', Addison – Wesley, Canada.
12. Kothari, C.R, (1990). Research Methodology: Methods and Techniques. 2nd Edition Wishira Prakashan.
13. Laudon K. C. and Laudon J. P. (2006). Management Information Systems: Managing the Digital Firm". Prentice Hall
14. Marra M. (2004), Innovation in E-Procurement: The Italian Experience. E-Government Series
15. McFarlan, F.W. (1984), "Information technology changing the way you compete", Harvard Business Review, Vol. 57 No.2, pp.115-26.
16. Morris C. (2006). A Comparative Study of e-Adoption Activities in South Africa and Mozambique

17. P. T. Joseph, S.J (2005), E-Commerce, An Indian Perspective. Prentice Hall of India
18. Randall D. Harris, (2000) B2B e-Commerce: Business Models and Revenue Generating Activities
19. Regina Haller, (2004).Emerging e-Procurement Models and the Effects on Internal Structures.
20. Somasundaram, R. and Damsgaard, J. (2005) "Policy Recommendations for Electronic Public Procurement" the Electronic Journal of e-Government Volume 3 Issue 3, pp 147-156
21. The Applicability of e-Procurement in the UK Public Sector: An Exploratory Analysis
22. The E-Government Strategy, Republic of Kenya (2004), Government Press
23. The Public Procurement and Disposal Act, Republic of Kenya (2005), Government Press
24. Vaidya K., Ping Yu, Soar J. (2002) Measuring e-Procurement Performance in the Australian Public Sectors: A Preliminary Approach
25. Video Conferencing Presentations by Felipe Goya, E. Talelro, at KDLC, KIA in June 2009
26. Wilson, C. L, "E-procurement and Energy", Power Economics, 2002
27. Internet Resources
 - a. <http://www.ppoa.go.ke/> (last updated on 2-8-2010)
 - b. www.epiq.com (accessed on)
 - c. <http://www.microsoft.com/emea/government/newsletters/procurement.msp>
 - d. http://www.elections.org.za/news_get.asp?News
 - e. www1.worldbank.org/prem/PREMNotes/premnote90.pdf

APPENDICES

Appendix A: Data Collection Instruments

Questionnaire for Head of IT and Procurement Departments

Introduction

This questionnaire is to be used to collect information on the e-procurement practice in the public sector of Kenya, factors that have led to the slow adoption and implementation and the existing models that have been used by those who have attempted to implement e-procurement.

Section I – General Information and current Practice

1. Name (optional)
2. Age bracket :
☐ Below 20
☐ 21 – 30
☐ 31 – 40
☐ Above 40
3. What highest level of education have you attained?
☐ Postgraduate ☐ Graduate
☐ College ☐ High School
☐ Others (specify)
4. Designation
5. Your Organization/Ministry/ Department
6. How long have you served in this organization?
☐ Less than 5 yrs
☐ 5 – 10 Yrs
☐ 10 –20 Yrs
☐ Above 20 Yrs
7. Do you have an IT Department? Yes ☐ No ☐
 a. If yes, who does the head of IT report to?

8. Do you have Internet connectivity in your place of work? Yes ☐ No ☐
9. If Yes, how reliable is the connectivity?
☐ Excellent ☐ Very Poor
☐ Good ☐ Don't Know

☐ Poor

10. Of what use is the Internet connectivity? Tick all that apply

☐ E-mail services

☐ Internet browsing

☐ Online Business transactions

☐ Others. (Specify)

11. Do you have a website for your organization/Department? Yes ☐ No ☐.

12. If yes what is the URL e.g. www.wewe.or.ke

.....

13. What is the website used for

.....

14. Is your website database driven?

☐ Yes

☐ No

☐ Don't know

15. Do you have a procurement department? Yes ☐ No ☐

16. If no, how do you procure your goods and services for the organization/ Department?

.....

.....

17. If yes, who does the head of procurement report to in your organization's structure?

.....

18. Which amongst the following committees do you have in your organization and where do they report to:-

	Committee	Reporting structure	
1	Tender Committee		
2	Procurement Committee		
3	Disposal Committee		
4	Inspection Committee		

19. Is your organization bound by the Public Procurement Acts and regulations? Yes ☐ No ☐

20. How many suppliers are doing business with your organization? (Appox.)

☐ 0 – 10

☐ 10 – 20

☐ 20 – 50

[] Above 100

21. Do you advertise for goods and services to be acquired by the organization? Yes [] No []

22. If yes how is it done?

[] Newspapers

[] Website

[] Telephone

[] Others. (Specify)

23. Are you aware what e-procurement is about?

[] Yes

[] No

If yes briefly explain:-

.....

.....

.....

.....

.....

24. If yes to question 22, do you think e-procurement is a good idea for your organization?

[] Excellent

[] Good

[] Moderate

[] Never

[] Don't know

25. Has your organization attempted to implement any e-procurement system?

Yes [] No []

26. If Yes,

a. Which model did your organization use to implement the system?

b. Which procurement stages do you do online over the Internet?

-
-
-
-
-

27. If No (for Q. 25), Do you have any system with a procurement module? Yes [] No []

28. If Yes (for Q. 27), state the system

Section II – Factors for Slow Adoption

29. In order of priority, tick the reasons why your organization has not or is slow to adopt e-procurement. (Where 1 is very high and 5 is very low).

Reason for slow adoption	Very High	High	Average	Low	Very Low
	1	2	3	4	5
I. Technology issues					
1. Poor/Lack of infrastructure					
2. No Appropriate model					
3. Integration with back office systems					
4. Security issues					
5. Lack of appropriate technology					
6. Lack of Technical Standards					
II. Organizational/Management Related issues					
1. Lack of support					
2. Lack of awareness					
3. Resistance from employees					
4. Lack of funds					
III. Regulatory					
Legal barriers					
IV. High costs involved					
V. External Market Barriers					
VI. Orgn. specific barriers					
1. Too few suppliers					
2. Size of org.					
3. Org. culture - Resistance					

4. Lack of suppliers willing to corporate					
5. Lack of skills					

Other reasons (State).

.....

.....

Section III – The Future of E-Procurement in Your Organization and Perceived Benefits

30. From an overall organizational level perspective, to what extent has your organization used e-Procurement?

Please tick the one that best describes the organization’s level of e-Procurement use in terms of overall procurement transactions.

Level of use of e-Procurement	Tick
No intention to use - <i>Our organization does not intend use e-Procurement within the next 12 months (or in the near future).</i>	
Intention to use- <i>Our organization has an e-Procurement implementation strategy in place or is preparing plans to investigate e-Procurement within the next 12 months (or in the near future).</i>	
Evaluation or Pilot use - <i>Our organization has purchased `trial' capabilities or implemented `pilot' projects and is currently evaluating the possible use of e-Procurement.</i>	
Commitment - <i>Our organization has specific plans and made formal commitments to utilize e-Procurement.</i>	
Limited deployment - <i>Our organization has established a programme of regular but limited use of e-Procurement.</i>	
Generalized deployment -	

<i>Our organization has deployed one or more forms of e-Procurement in its generalized capacity and integrated e-Procurement in the majority of information systems.</i>	
Rejection - <i>Our organization has evaluated and rejected the use of e-Procurement</i>	

31. From your own perspective tick the benefits that can be derived from adoption of e-procurement. (Where 1 is very high and 5 is very low).

Benefits for e-Procurement adoption	Very High	High	Average	Low	Very Low
	1	2	3	4	5
Increased speed, accuracy and transparency					
Wider competition and economy in procurement					
Simply standardize and speed up process and communication					
Real time generation of detailed and accurate data with seamless sharing of information					
Facilitate online pre/ post contracts management including bidding, contracting, ordering and payments with online participation of all stakeholders effectively					
Reduction in cost of Government purchases with reduced cycle time					
Elimination of delays arising out of processing tenders manually					

Reduced paperwork and lower transaction costs					
Problem of mafia and cartel formation arrested.					
Database on goods, services, works and contractors get build up					
Economy of scale achieved by aggregation of requirements					
Quick and easy access to all information and audit trails					
Better monitoring of tenders and contracts					
Reduced inventory costs					
Improved work culture in the department and better vendor relationship					
Winning the trust of people and good governance					
Ability to link directly into existing systems, such as ERP					
Self-invoicing on behalf of clients can add to the bottom line.					
Month-end reconciliation can end the problem of the wrong items being ordered or the wrong price being offered as business processes have been streamlined and all are working off the same catalog					
Maverick spending is reduced					

Other benefits (State).

.....

.....

.....

Thank You.

**UNIVERSITY OF NAIROBI
SCHOOL OF PHYSICAL SCIENCES
DEPARTMENT OF GEOLOGY**

**GROUNDWATER QUALITY IN SAMIA AREA OF BUSIA
COUNTY, WESTERN KENYA**

MANG'ENI FREDRICK BWIRE

I56/60194/2010

**A DISSERTATION SUBMITTED TO THE UNIVERSITY OF NAIROBI IN
PARTIAL FULFILMENT FOR THE DEGREE OF MASTER OF SCIENCE IN
HYDROGEOLOGY AND GROUNDWATER RESOURCE MANAGEMENT**

AUGUST 2012

DECLARATION OF AUTHORSHIP

I hereby declare that this is my original work and has not been submitted by any other person in any Institution for any award. It is hereby presented to the Board of Examiners with permission from my supervisors.

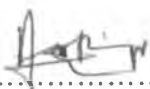
CANDIDATE

Signed..........Date.....23-11-2012

MANG'ENI F BWIRE

MSc. student in Hydrogeology and Groundwater Resource Management
Department of Geology; School of Physical Sciences
College of Biological and Physical Sciences
University of Nairobi

SUPERVISORS

Signed..........Date.....23/11/2012

Dr. D. O. OLAGO

Department of Geology; School of Physical Sciences
College of Biological and Physical Sciences
University of Nairobi

Signed..........Date.....26/11/2012

Prof. J. O. BARONGO

Department of Geology; School of Physical Sciences
College of Biological and Physical Sciences
University of Nairobi

ABSTRACT

This study focuses on the groundwater quality in Samia area, western Kenya. It also investigates the aquifer types and their possible impacts on the water quality. The geology of the area comprises granite-greenstone terrane of Archaean age. The Archaean rocks include the Nyanzian “supergroup” intruded by igneous complexes and calc-alkaline plutons which comprise the Samia series and acid volcanic units which include the banded iron formation of the Samia hills. The lithological composition of these rocks appear to have profound effects on water quality.

The results of this study show that the groundwater is fairly neutral with pH between 6.5 and 7.5. The minor elements like zinc, manganese and iron are within the World Health Organization (WHO) potable water guidelines; similarly fluorides, chlorides, nitrates and sulphate are within safe limits. Chemical analyses of the water indicate geological influence and the hydrochemical facies in the area can be divided into at least two zones;

- Calcium bicarbonate type in the southern and
- Sodium- potassium bicarbonate type in the northern parts of the area.

As the groundwater flows from the recharge to the discharge areas, it increases in Total Dissolved Solids (TDS), which is interrupted by geological structures such as fissures, shears or cracks which could be natural, hence altering the water chemistry. Although the waters from most of the drilled boreholes and wells are suitable for most purposes like domestic, agricultural and industrial, the presence of some mineral components (e.g., Ca and Mg) above the required guidelines in some wells make water from these few boreholes unsuitable for laundry and other domestic purposes including cooking.

It was noted from this study that for all the drilled wells, the water rest level was above the water struck level providing strong evidence for semi-confined aquifer fields. The water struck level in the area ranges from 28 to 68 m while the water rest level ranges from 10.2 to 30.34 m in depth from the ground surface.

ACKNOWLEDGEMENT

I wish to express my greatest gratitude to all the lecturers who taught me during the entire MSc. coursework period. Special thanks go to Dr. D.O Olago and Prof. J. O. Barongo who guided me throughout the entire research period from its inception to its conclusion.

Special appreciation goes to Dr. M Gichaba, Dr. E. Dindi, and Mr. A. Wawsa for accepting to have their input in this project. Wonderful appreciations also go to Prof. Isaac Nyambok and Prof. E. O. Odada who went through the project proposal, suggesting various adjustments which were all taken into consideration.

I also owe thanks to Mr. C. Namenya of the Database Section and Mr. P Njuki of the Drilling Section of the Ministry of Water and Irrigation for sharing their data with me.

I am indebted to the Senior Technologist, Mr. S. Mbiga, for providing field equipments like GPS and maps, and to the Geochemistry Laboratory technicians, Mr. S. Kamano, Ms E. Njeri and Ms J. Mulu for prompt analysis of the field samples despite their huge backlog.

I am so thankful to the Chairman, Department of Geology, and Dr. C. Nyamai for his constant encouragement during the project period and subsidizing the sample analysis fees.

To Mr. S. Atonya of the Geospatial Laboratory and Mr. T. Wanambisi of the Computer Laboratory, both of the Department of Geology, I say thank you. Special category of thanks goes to Mr. C. Mudunyi for his personal assistance during the data collection phase.

Many thanks to all my colleagues in the MSc. course, Mativo K, Mibei Kiptoo, Wekesa G, Deflora K, and Wamalwa R. who assisted me in many ways and to all my friends, I say thank you.

TABLE OF CONTENTS

DECLARATION OF AUTHORSHIP	I
ABSTRACT	II
ACKNOWLEDGEMENT	III
TABLE OF CONTENTS.....	IV
LIST OF FIGURES	VI
LIST OF TABLES	VII
ABBREVIATIONS	VIII
CHAPTER ONE	1
1.0 INTRODUCTION.....	1
1.1 GENERAL INTRODUCTION	1
1.3 LITERATURE REVIEW	2
1.4 STATEMENT OF THE PROBLEM	7
1.5 AIM AND OBJECTIVES	8
1.6 JUSTIFICATION AND SIGNIFICANCE.....	8
CHAPTER TWO.....	9
2.0 THE STUDY AREA.....	9
2.1: LOCATION	9
2.2: CLIMATE/VEGETATION	9
2.3: PHYSIOGRAPHY	11
2.3.1: THE SAMIA HILLS	11
2.3.2: THE MIDDLE GROUND	11
2.3.3: THE SIO LOWLANDS	12
2.4: AGRICULTURE	13
2.5: HYDROLOGY	13
2.5.1: DRAINAGE	13
2.6: GEOLOGY	14
2.7: STRUCTURES.....	16
2.8: HYDROGEOLOGY	17
2.8.1: GROUNDWATER RECHARGE	17
2.8.2: GROUNDWATER FLOW SYSTEM.....	17
2.8.3: MINERAL CONSTITUENTS	18

2.8.4: QUALITY CHARACTERISTICS	19
2.8.5: BOREHOLE SURVEYS	21
2.8.6: POTENTIAL RISK TO AQUIFER.....	22
CHAPTER THREE.....	23
3.0 METHODOLOGY.....	23
3.1 INTRODUCTION.....	23
3.2 LOCATIONS OF BOREHOLES.....	23
3.3 FIELD WORK SURVEY.....	23
3.4 SAMPLE PREPARATION AND ANALYSIS	25
3.5 DATA QUALITY	26
3.7 CHALLENGES	26
CHAPTER FOUR.....	28
4.0 RESULTS AND INTERPRETATION.....	28
4.1 WELL LOGS	28
4.2 PHYSICO-CHEMICAL DATA OF BOREHOLES	31
4.3 CHARGE BALANCE ERROR	34
4.4 WATER QUALITY	41
4.5 CORRELATION COEFFICIENT.....	44
4.6 PRINCIPAL COMPONENT ANALYSIS.....	45
4.7 ROTATED FACTOR ANALYSIS	48
CHAPTER FIVE.....	51
5.0 DISCUSSION.....	51
CHAPTER SIX	54
6.0 CONCLUSION AND RECOMMENDATIONS.....	54
6.1 CONCLUSION.....	54
6.2 RECOMMENDATIONS.....	55
REFERENCES.....	56

LIST OF FIGURES

Fig 1.1 Fetching water in one of the boreholes in the study area	
Fig 1.2 A map of Kenya and its neighborhood showing the location of the area of study	
Fig 2.1 Different features in Samia study area	
Fig 2.2 Physiographic units of Samia study area	
Fig 2.3 Geological map of the study area.....	
Fig 3.1 Locations of boreholes where the water samples were collected	
Fig 4.1 An illustration of Nangina catchment area to groundwater flow and discharge	
Fig 4.2 Hydrogeological profile from the direction indicated in fig.3.1	
Fig 4.3 Piper diagram on the types of water chemistry of the water samples in the area.....	
Fig 4.4 Hydrogeochemical classification system for waters using the piper diagram.....	
Fig 4.5 Schoeller semi-logarithm diagram with some chemical values from Samia area	
Fig 4.6 Hydrogeochemical facies of Samia area	
Fig 4.7 Factor loading plots for 5 factor models	

LIST OF TABLES

Table 1.1 Rainfall figures from a recording station up to 1976	10
Table 2.2 mean hydrochemical parameters of values of the water	18
Table 2.3 Mineralogical composition of common rocks in the study area	19
Table 2.4 Elemental composition of common rocks in the study area	19
Table 2.5. Recommended parameters for chemical analyses	21
Table 2.6 Substances found in groundwater which cause health problems	22
Table 4.1 Some borehole data of the Samia study area	28
Table 4.2 Analytical results of the seven elements from the field water	31
Table 4.3 Physico-chemical data courtesy of Ministry of Water and Irrigation	32
Table 4.4 Milliequivalents data for the boreholes	35
Table 4.5 Table 4.6 Percentage milliequivalent data for the boreholes	36
Table 4.6 Summary of water types available in Samia area	41
Table 4.7 SAR values for the samples in the study area	43
Table 4.8 Correlation co-efficient values of the elemental parameters	44
Table 4.9 Communalities extraction method	45
Table 4.10 Total variance explained	46
Table 4.11 Component matrix. The 5 principal components before rotation	47
Table 4.12 Rotated promax structure matrix	49
Table 4.13 Component correlation matrix	50

ABBREVIATIONS

AAS	-Atomic Absorption Spectrometry
ASL	-Above Sea Level
BIF	-Banded Iron Formation
CBE	-Charge Balance Error
CLR	-Colour
Ec	-Electrical conductivity
GPS	- Global Positioning System
ITCZ	- Inter Tropical Convergence Zone
IYPE	- International Year of Planet Earth
KEFINCO	- Kenya- Finland Company
MDG	- Millennium Development Goals
Mkt	- Market
PCA	- Principal Component analysis
PFA	- Principal factor analysis
Q-L/min	- Discharge in litres per minute
SAR	- Sodium Absorption Ratio
Sch.	- School
T_{ALK}	- Total alkalinity
T/depth	-Total Depth
TDS	- Total dissolved solids.
T_{HARD}	-Total hardness
UNEP	-United Nations Environmental Programme
Vill.	- Village
WHO	- World Health Organization
WRMA	- Water Resource Management Authority
WRL	- Water Rest Level
WSL	- Water Struck Level

CHAPTER ONE

1.0 INTRODUCTION

1.1 GENERAL INTRODUCTION

The Study area (Samia) is located in western Kenya and borders Uganda to the west. The area has several seasonal streams and springs, few shallow wells and many boreholes that provide water for the local population. Due to ever increasing population in the area, as is in the rest of Kenya, the demand for water in Busia County as a whole has surpassed the supply due to its inconsistency in surface availability as a result of adequate but unreliable rainfall that feeds the local streams throughout the year.

The only permanent rivers draining the area are Sio, Wakhungu and Madele-Masanyi Rivers, which also have very low flows during dry seasons. Thus groundwater sources are important (Plate 1.1), even though the water has been found to have varying chemistry. Provision of safe, adequate and reliable supply of clean water within reasonable walking distance is an impetus to economic growth and public health.

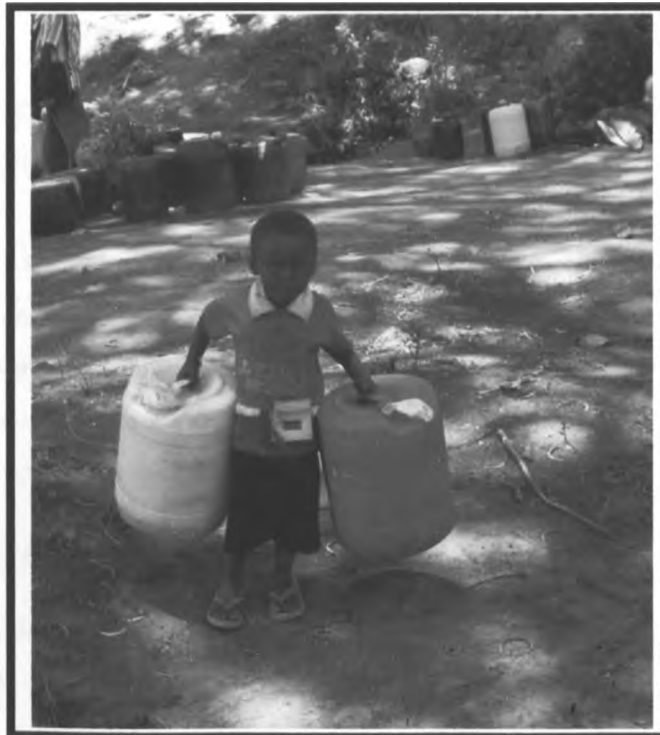


Plate 1.1 Fetching water from one of the boreholes in the study area

Of the water resources, groundwater is the most critical as it is readily available (UNEP 2002) and relatively free from contamination despite its uneven distribution. Groundwater has been exploited for domestic use, livestock and irrigation since the earliest time (Water Well Industry Handbook 1975). In the borehole drilling process, more than one water bearing formation may be encountered by a well and the composition of the respective waters encountered may vary widely. In such a situation, different methods of investigation may be applied in order to characterize the aquifers and waters of unsuitable quality supply. From the 1990s, the Ministry of Water and Irrigation and Kenya-Finland Company (Kefinco) and other donor organizations, sank many boreholes in the study area, sometimes using explosives for cracking the aquifer to increase their capacity during the well development process, depicting the various types of geological formations in reaching the suitable aquifers. This effort was to supplement the surface supply such as springs and streams.

The population of Samia district is approximately 93,500 persons (Kenya National Bureau of Statistics, 2010) and is unevenly distributed, with most people found in the western side of Samia hills. The eastern side of the hills is sparsely peopled owing to slightly lower rainfall and rather poor soils.

1.2 SCOPE OF THE RESEARCH

This research aims at filling the gap in the information available in our water quality data by investigating the elemental composition (water quality) of the groundwater in Samia area (Fig.1.1) to add on the existing information from earlier works.

In particular, the project analyzes the major cations (Na, K, Ca, and Mg) and minor cations (Fe and Mn) in water together with trace elements (Zn) to enable a classification of the ground water according to elemental composition and use the data to suggest possible sources of the elements and treatment of those in excess of the recommended guidelines for various uses.

1.3 LITERATURE REVIEW

Ogege et al. (2002) did investigations on geological controls on groundwater geochemistry in Butula area, Busia County in Kenya, an area adjacent to the study area

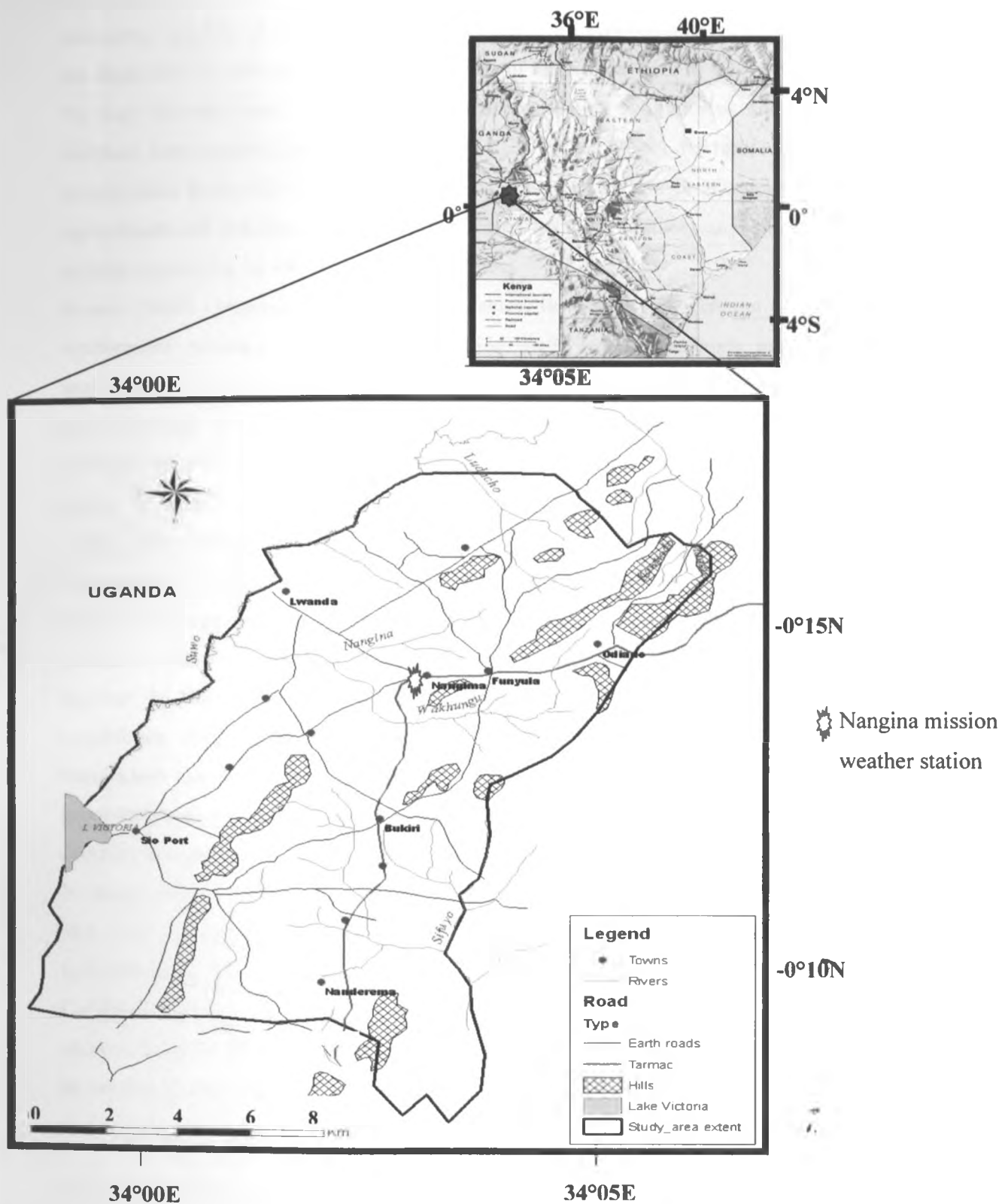


Fig. 1.1: The location of the area of study

and found that the geology of the Area plays the greatest role in the determination of water chemistry. Marleen et al. (2008) investigated the hydrogeochemical processes influencing the high fluoride concentrations in groundwater of the Nairobi area in Kenya and found that the high fluoride concentrations are correlated with high sodium ions and pH and low Calcium ions concentrations. Mohammad and El Nakhal (2004) did research in the Gaza strip (Middle East) and found that groundwater being the only source of water for domestic, agricultural and industrial use has been affected both qualitatively and quantitatively by extensive pumping and the hydrochemical facies.

Arturo (2008) investigated hydrochemical facies, flowpaths and recharge determined by multivariate statistical, isotopic and chloride mass balance methods and found that multivariate statistical methods (MSMs) provide valuable insight into the main hydrochemical species, hydrochemical processes, and water flowpaths important to groundwater evolution using factor analysis. Schroeiber et al. (1999) used hydrochemical facies to delineate groundwater flowpaths in fractured shale and found that the hydrochemical facies identified by correlation of water types between boreholes indicated the general orientation of groundwater flow paths. Casey and David (2007) investigated groundwater age and the importance of chemical fluxes across aquifer boundaries in confined Cretaceous aquifers of North Carolina, USA, and found quantitatively how the aquifers in North Carolina are highly varied in their character and water producing capabilities. Bilqis et al. (2006) did work on Arsenic Mitigation for Water Supply in Bangladesh and found that Bangladesh is faced with the worst arsenic contamination of groundwater in the world, with an estimated 35 to 77 million people at risk of exposure to drinking arsenic-contaminated tubewell water. British Geological Survey (2001) did work in many countries (Tanzania, Zambia) to identify inorganic constituents of significant risk that occur in groundwater. Guler and Thyne (2004) worked on the delineation of hydrochemical facies distribution in the regional groundwater system in southeastern California using Fuzzy C-Means (FCM) and Hierarchical and Cluster Analysis (HCA). Results showed that both FCM and HCA clustering produced cluster centers (prototypes) that can be used to identify the physical and chemical processes creating the variations in the water chemistries.

Aquifers, or more accurately, aquifer systems, are hydraulically connected materials (sands, limestone, and fractured rock) that provide water through a properly constructed well open to those materials.

Ensuring that water is safe to drink is very important to health. Water can contain various contaminants which may be hazardous to humans and animals. There are a few methods to test the potability of water and protect health (Freeze and Cherry, 1979). Partial chemical tests usually check for magnesium, iron, fluoride, chlorine, and other chemical compounds levels in water (Lindsten, 1984). Most tests for water are conducted for independent substances that may be in water. Chemical tests will not show any levels of bacteria or other potential hazards.

Water treatment describes those processes used to make water more acceptable for a desired end-use. These can include use as drinking water, industrial processes, medical and many other uses. The goal of all water treatment process is to remove existing contaminants in the water, or reduce the concentration of such contaminants so the water becomes fit for its desired end-use (Llyod, 1985).

The hydrogeological work that has been done in the study area includes the drilling of wells and keeping their logs by the Ministry of Water and Irrigation headquarters, Nairobi, District Water Officer, Busia and WRMA, Kakamega. The records kept lack complete analysis of major ions and trace elements in the water, but include information such as; the location of borehole, their total depth, water struck level, water rest level and well discharge. Partial chemical analysis records exist for anions for a few wells in the study area.

Kefinco (1990) classified the area as a low groundwater potential zone compared to other areas in the region due to few perennial river sources and springs, a manifestation of potentially low volume of groundwater (Freeze and Cherry, 1979; Mathess, 1982; WHO, 1984). Groundwater character is determined by the interplay between it and precipitation, rivers, soils, vegetation, etc. Generally, groundwater at the discharge zones tends to have higher mineral concentration levels compared to that at the recharge zones due to longer residence time and prolonged contact with the geological factors (Walton, 1970; Todd, 1980). The groundwater yield in the area varies a great deal from the available data (Kefinco, 1990); it varies from 18L/min. at Nakhasiko village to 4L/min. at Namboboto School suggesting that the aquifers in the study area are composed of different rocks. Also, according to Kefinco (1990), sedimentary aquifers have a better yield than the volcanics in the area and the water struck level is shallower in sedimentary aquifers at 22m than in the

volcanics which is deep at about 68m. Chances of getting fresh water in the area are high especially in the central and northern part of the study area. The aquifers in the study area have also different lateral and vertical extent and localized aquifers like in fractured zones are also encountered. Groundwater flow is mainly from the elevated areas in the hills towards the lower areas in the lowlands.

Until the recent geological mapping by Kimomo et al. (2007), information on the geology of Samia area was rather scanty. This is because no detailed geological mapping has been done. The only information available before was of regional nature by authors who worked in the general western Kenya region. These include works by Elliot and Gregory (1895), Odman (1929), Stockley (1943), Yanagi and Suwa (1981), Ichang'i and MacLean (1991), and Kimomo et al. (2007). Their contributions are reviewed in these sections with more emphasis on the most recent work by Kimomo et al. (2007).

Elliot and Gregory (1895) referred briefly to the area indicating a Karagwe-Ankolean age for the quartzites and the dissociated rocks of the Samia hills. Gregory (1921) described "Karagwe-Ankolean sediments" in the Samia hills as "folded into a syncline, at the base of which occur the Ironstones around Sio Port Bay" and mentioned the occurrence of gneiss in the Samia hills and along the River Sio. In 1929, Odman published a paper dealing with the geology of North and Central Kavirondo. He noted the lithological differences in the Samia hills Series-"banded ironstones, shales and greenstones" folded on North-East axes and broken by many small transverse faults. A field study to update the geological data was carried out in 2006 by Kimomo et al.(2007). An attempt was made however, to cover as much as possible in the time available and to arrive at a conception of the geology as a whole.

According to Stockley (1943), the rocks of this area are part of a greenschist belt of Archean age and were formally described as belonging to the 'Nyanzian' system. The Archean rocks comprise a lower sequence of greenschist-grade, predominantly felsic and intermediate volcanics (mainly lavas) with minor mafic volcanics (a rhyolite dated using Rb/Sr methods by Yanagi and Suwa (1981) gave an age of 2710 ± 340 Ma) and an upper, thick turbiditic, sedimentary sequence derived primarily from the erosion of the adjacent volcanics. Proterozoic rocks and structures appear to be absent in the area although some

NW-SE trending faults are evident. Pleistocene and Recent deposits include laterites, soils and alluvial deposits.

The mapping by Kimomo et al. (2007) has highlighted the weakness of the existing chronostratigraphic framework, and supports the type of revision suggested by Ichangi and MacLean (1991) and Ichangi (1992). These authors, working on the greenstone belt of the Winam rift around Migori and Lolgorien commented on the interferences concerning the use of “System” essentially a chronostratigraphic term and subsequently adopted lithostratigraphic nomenclature consistent with the International Sub-commission on Stratigraphic Classification (1976) and the North American Commission on Stratigraphic Nomenclature (1983). They proposed that a Nyanzian Supergroup should comprise all supracrustal rocks in the Nyanza Greenstone belt including those of the “Kavirondo Group.” All supracrustal rocks (the volcanics together with volcanogenic and chemogenic sediments) are therefore viewed as a constituent of a Nyanzian Supergroup.

1.4 STATEMENT OF THE PROBLEM

The primary aim of a water quality investigation is the protection of public health, hence the formation of international standards or guidelines for drinking water quality from World Health Organization (1984) and other agencies such as the United States Environmental Protection Agency (1986). Though water supply projects have been initiated recently as part of development plans and Millennium Development Goals (MDGs), no water treatment plant exists in the area and therefore the quality and use of the groundwater in this respect needs to be investigated as the attention has now shifted to groundwater and water supplies through pipes from surface supply. Knowing which water contaminants are most likely to be a problem and being aware of the warning signs of certain contaminants can make such testing more meaningful and economic.

The demand for water in the study area and Busia County as a whole has surpassed the surface supply from streams, springs and shallow wells, making people concentrate more on the option of exploiting groundwater sources. The water abstracted from some of the wells sunk in the area has noticeable chemistry in that:

- It does not produce lather with soap;

- It has unpleasant strong taste when taken; and
- It is not used in brewing or cooking of certain foods e.g. a mixture of maize and beans though it has not shown any adverse health effects to people in the area.

From the present status, it is clear that no study has been done to establish the chemistry behind these effects and some of the new boreholes drilled in the area continue to experience the same problem.

1.5 AIM AND OBJECTIVES

- The main aim of the study was to investigate the hydrogeochemical facies of the Samia area.

The objectives of the study were:

- (1) to investigate the type of the aquifer, i.e., whether uniform and extensive or small and discrete (discontinuous);
- (2) to define the spatial variation of the major, minor and trace elements present in the water and assess its potability.;
- (3) to classify groundwater based on the chemical constituents; and
- (4) to recommend appropriate water treatment methods.

1.6 JUSTIFICATION AND SIGNIFICANCE

Lack of a water quality database suggests that the users may be exposed to negative health effects due to its consumption. An investigation of water quality parameters would help confirm its suitability for intended use, i.e. whether suitable for domestic, industrial or agricultural and provide future guidelines concerning the development and use of groundwater.

CHAPTER TWO

2.0 THE STUDY AREA

2.1: LOCATION

Samia, herein referred to as 'the study area' is found in the extreme western part of Kenya and to the north-east of Lake Victoria. It is on the Kenyan side of the border with Uganda covered by two topographic sheet maps namely Samia (101/3) and Busia (101/1). It is bounded by longitudes 33°55"E and 34°10"E and by latitudes 0°07"N and 0°20"N but excludes the Siaya area. The study area occupies the southern part of Busia County and covers an area of approximately 180 km² (Fig. 2.1).

The study area can be accessed via Nairobi-Busia highway on which there are two junctions: one at Bumala trading centre approximately 24 km from Busia town to the east and the other at Matayos on the same highway about 15km from Busia town to the east. From Bumala junction, the area is served by a tarmac road to Port Victoria, while from Matayos trading centre, it is an earth road. The study area can also be reached from Siaya town via Ruambwa trading centre to Hakati market. The area is fairly well served with earth roads though access to a large portion may often be difficult, only reached by use of foot paths.

2.2: CLIMATE/VEGETATION

The study area experiences modified equatorial type of climate with moderate amount of rainfall in two seasons related to the movement of the Inter Tropical Convergence Zone (ITCZ) during the long rains from March to June, and frontal rain during the short season from September to December. Rainfall records for a station in the area are given in Table 2.1.

Table 2.1: Rainfall figures from a recording station up to 1976. See fig 1.1 (From Jaetzold and Schmidt, 1982).

Station	Altitude (m)	years	Av. rainfall (mm)	Jan-Feb.	Mar-Apr.	May-Jun	Jul-Aug.	Sep-Oct	Nov-Dec.
Nangina Mission	1219	37 (1945-82)	1421	100 mm	370 mm	234 mm	180 mm	311 mm	226 mm

The annual rainfall records (Table 2.1) show that not only do both rainy seasons have very wide ranges of rainfall, assessed at 60% reliability level, but the second rains are the most variable. Rainfall patterns are modified by the effect of Lake Victoria.

The natural vegetation in the area is tall grass and scattered bush in between hills, scrub and scattered trees on the hills but most of it has been interfered with except on the Samia hills where can still be seen.

2.3: PHYSIOGRAPHY

Physiographically, the study area may be divided into three distinct types (Fig. 2.2) after Kimomo et al. (2007).

- The Samia hills
- the Middle ground and
- the Sio lowlands.

2.3.1: THE SAMIA HILLS

The hills rise sharply from the countryside and include Njalagombe, Samia, Nangosia, Nambuku and Odiado. They extend from Port Victoria (Bunyala) to Murende (Marachi) between Bumala and Matayos junctions (Fig. 2.1). The ridges of the hills are segmented by passes which control communication and are more difficult to access owing to coarse screes, covered by tall grass and in places by thick thorn bush. The highest points are found on Odiado hills (1568 m asl) in the north to Busia hills (1464 m asl) to the south near the shores of Lake Victoria. The hills are most easily examined in the early months of the year when the local people are burning the vegetation.

2.3.2: THE MIDDLE GROUND

In between the hills and the lowlands are ridges, which form the middle ground consisting of gently sloping countryside of elevation 1250-1180 m asl, separated by incised streams. They have moderately deep fertile land with thicker soils and poorly exposed rock units, often volcanic or pyroclastics. The laterite platforms, which form pediments to the outer margins of the ridges, are also locally found.

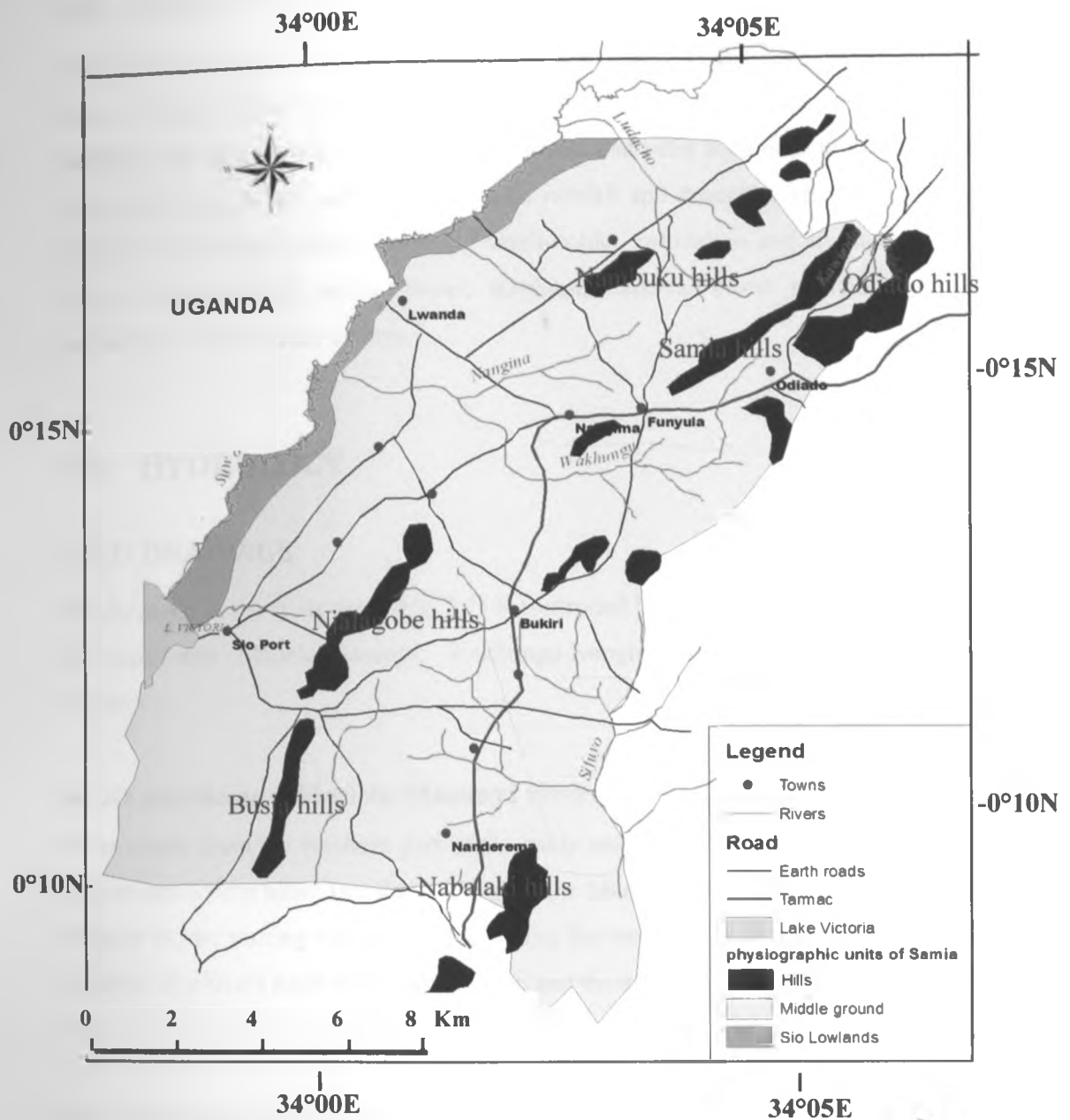


Fig.2.2: Physiographic units of Samia study area.

2.3.3: THE SIO LOWLANDS

These lowlands extend from Sio Port (Bujwanga) past Mluanda up to Nambale (Butula). The elevations are generally lower than 1160 m asl. The lowlands are a broad floodplain, up to 4km wide and are flooded during the peak rainy period. The floodplain is marked by extensive alluvium, colonized by thick reed beds,

2.4: AGRICULTURE

The study area receives moderate rainfall that is adequate for agricultural activities. Most of the activities take place in the middle ground (Fig. 2.2) which is the most inhabited owing to moderate to deep fertile soils that offer little possibility to erosion. Crop production is not stable due to erratic rainfall and droughts. The main activities include subsistence farming involving small-holder cultivation and livestock rearing. Crops grown include maize, beans, sorghum, cassava, sweet potatoes and now horticulture along stream valleys.

2.5: HYDROLOGY

2.5.1: DRAINAGE

The drainage in the study area (Fig. 2.1) is controlled by four major systems namely the Kawuodo and Madele-Masanyi, Wakhungu-Nangina, Ludacho and Sifuyo river systems.

(a) Kawuodo and Madele-Masanyi rivers

These rivers drain the northern part of the study area in a northerly direction, from the Odiado and Samia hills. The Kawuodo joins the Madele-Masanyi at the boundary with Marachi before turning westwards and joining the main Suwo River flowing into Lake Victoria. The rivers have no gauging station and therefore no measure of their discharge can be assessed, although they flow throughout the year with varied regime.

(b) Wakhungu-Nangina rivers

The Wakhungu river drains the central part of the study area from Odiado and Samia hills in the southerly direction before turning westwards, where it is joined by the seasonal Nangina River flowing westwards into the main Suwo River which ends in Lake Victoria. The rivers have no gauging station along their course and therefore no hydrological data is available though the rivers flow throughout the year.

(c) Ludacho river

This river drains the northern section of the study area in the north-west direction before turning west and joining the main Suwo River. The river originates from the Samia hills and flows throughout the year though there is no gauging station along its course.

(d) Sifuyo river

This drains the southern part of the study area in a south easterly direction from the Njalagobe hills but disappears before reaching the main Nzoia River. The river is seasonal with its peak during the rainy season though it has no gauging station.

2.6: GEOLOGY

According to Kimomo et al. (2007), the geology of study area (Fig. 2.3) consists of the rocks of the Archean Nyanzian Supergroup and comprises of rocks of Samia group which include a folded sequence of volcanic, volcanosedimentary, chemogenic, pyroclastic and cataclastic rocks. It is named and defined from the Samia hills, and is equivalent to Pulfrey's (1936a, b). The rocks in this group include andesitic lavas and pyroclastics, dacites, rhyolites and quartz porphyries; banded iron formations (BIF)/chert; minor pillow lavas; and a suite of extensively altered (carbonated) fault-zone rocks which include calc-silicate rocks, ultramafics, conglomerates and breccias. Dolerites, diorites and granite intrude rocks of the group.

The Samia group includes and is characterized by distinctive NE-trending, elongate ridges of Banded Iron Formation (BIF) which occupy much of the western part of the area, from Port Victoria to Matayos. The group is overlain by an extensive cover of laterite thus making the boundaries between it and other groups or formations highly speculative as they are to a great extent based on changes in the soil colour and vegetation.

Chert and BIF form ridges and other high ground, whilst in general rhyolite, dacite, andesite lavas and pyroclastics underlie areas of subdued topography, with flat broad valleys mostly covered with thick dark red clayey soil. Extensive laterite caps (although much is often dissected) are common around the Samia hills.

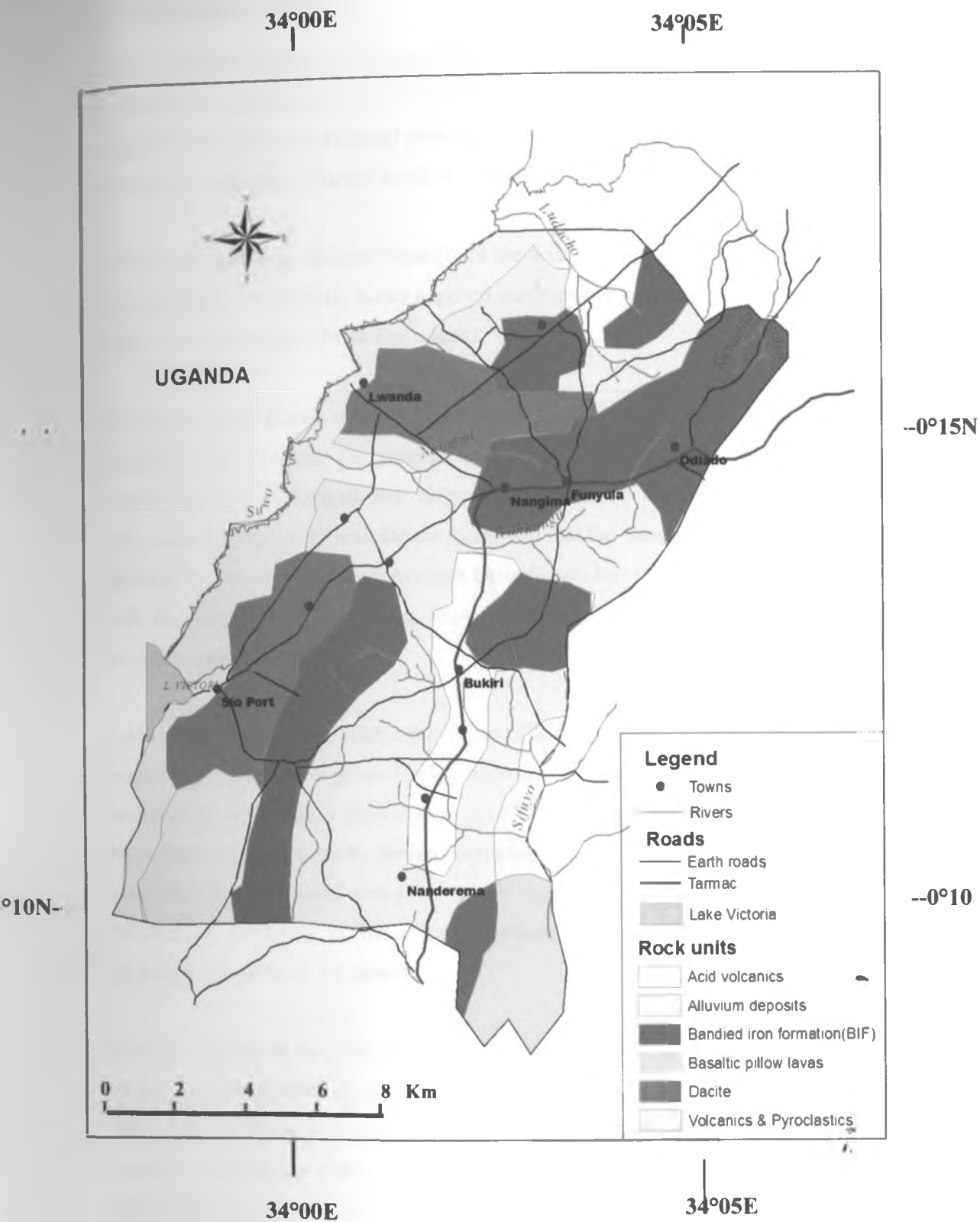


Fig 2.3: Geological map of the study area (after Kimomo et al. 2007).

2.7: STRUCTURES

The structural picture that has been built up from the gross outcrop pattern is linked to anticlinal and synclinal fold pairs with E-W axes, commonly with axial planes dipping steeply to the south. Anticlinal axes appear to 'focus' diapiric uprise of the late-stage potash-rich granites, a feature noted by Suwa (1981).

Because of the poor exposure throughout the area, details of structural styles are taken from a variety of scattered, better exposed outcrops and assumed to be applicable to the area as a whole (larger Busia-Siaya area).

The Samia group shows structural conformity broadly in agreement with the rest of the mapped area. However, the general lack of an axial planar cleavage and the slight rotation of the boudinaged BIF ridge segments, taken together with the significant difference in lithologies, indicate the likelihood that the Samia group was deposited in a different or separate part of an Archean basin (much less tectonically active, contrasting with the progressively deepening, sag-type basin into which most of the other groups were deposited).

Folds of the Samia group within the Samia hills are upright and close, with axial planes trending NE-SW, and in general plunging to the NE. The overall fold pattern is strongly modified by E-W shears (Kimomo et al. 2007). Because of the variation in lithology, the banded ironstones are by far the most competent rocks, the chert next, and the least competent are the altered fine-grained acid pyroclastics. Deformation has produced a cleavage or schistosity in fine-grained, kaolinised pyroclastics which is concordant to the regional structures, i.e. trending E-W.

In detail, sections in the hills show that the more competent ridge-forming BIF's may be weakly boudinaged and slightly rotated. Antiformal-synformal pairs in the southern part of the Samia hills appear to either have limbs sheared out or not exposed (Kimomo et al. 2007). This is true north of the Bulemia area, and complicates the otherwise fairly simple fold structures of the hills.

2.8: HYDROGEOLOGY

2.8.1: GROUNDWATER RECHARGE

The study area receives an average of 1421mm of rainfall which comes in two seasons. Estimates of annual groundwater recharge in the area are absent but are generally higher in the northern part of the area where the streams flowing from those areas are perennial as compared to the southern part where they are seasonal.

2.8.2: GROUNDWATER FLOW SYSTEM

Groundwater flow movement in the area is directed towards the north and southern parts. Groups of springs with small to moderate flow issues from the foot hills of Samia and Odiado hills. In the southern part of the area, groundwater is mostly extracted through boreholes.

The mean hydrochemical parameters of values of the water based on the work by Kefinco (1990) are given (Table 2.1).

Table 2.2: The mean hydrochemical parameters of values of the water based on the work by Kefinco (1990).

Hydrochemical parameter	value
Salinity (TDS)	641.7mg/l
Alkalinity	233.1mg/l
Sulphate	81.6mg/l
Chloride	109mg/l
Total/Hardness	476.6mg/l

The hydrochemical composition is controlled primarily by the dissolution of the andesites, pyroxenes and hornblende which make the bulk of the area geology.

2.8.3: MINERAL CONSTITUENTS

The greater part of the soluble elemental constituents in ground water comes from soluble minerals as a result of rock weathering. The more common soluble constituents include calcium, sodium, magnesium, potassium, fluorides, bicarbonate and sulphate ions (Table 2.2). Another common constituent is chloride ion derived from intruded sea water, connate water, evapotranspiration concentrating salts, and sewage wastes, for example. Nitrate can be a natural constituent but high concentrations often suggest a source of pollution.

Table 2.3 and 2.4 show the mineralogical and elemental composition respectively of the common rocks in the study area as compiled by the author.

Table 2.3 Mineralogical composition of common rocks in the study area.

ROCK	MINERAL COMPOSITION
ANDESITE	Plagioclase, hornblende, pyroxene, biotite
RHYOLITE	Quartz, feldspar
DACITE	Plagioclase, hornblende, pyroxene, biotite, quartz, feldspar
CHERT	quartz
DOLERITE	Plagioclase, pyroxene, feldspar
DIORITE	Plagioclase, feldspar, hornblende

Table 2.4 Elemental composition of common rocks in the study area.

MINERAL COMPOSITION	ELEMENTAL COMPOSITION
Plagioclase, $\text{CaAl}_2\text{Si}_2\text{O}_8$ (anorthite), $\text{NaAlSi}_3\text{O}_8$ (albite)	Na, Ca, Si, Al, O.
Hornblende $(\text{Ca},\text{Na})_{2-3}(\text{Mg},\text{Fe}^{+2},\text{Fe}^{+3},\text{Al})_5\text{Si}_6(\text{Si},\text{Al})_2\text{O}_{22}(\text{OH})_2$	Na, Ca, Mg, Fe, Al, Ti, Si, O, OH
Pyroxene $\text{Si Al Fe}^{3+} \text{ Al Fe}^{3+} \text{ Ti}^{4+} \text{ Cr V Ti}^{3+} \text{ Zr Sc Zn Mg Fe}^{2+} \text{ Mn Mg Fe}^{2+} \text{ Mn Li Ca Na}$	Na, Ca, Mg, Fe Al
Biotite $\text{K}(\text{Mg},\text{Fe})_3(\text{Si}_3\text{Al})\text{O}_{10}(\text{OH},\text{F})_2$	K, Mg, Fe, Al, Si, O, OH, F
Quartz	Si, O

2.8.4: QUALITY CHARACTERISTICS

The measure of total dissolved solids (TDS) is a good indicator of the mineralized character of the water (Water Well Industry Handbook 1975). Ground water having less than 500 mg/L of total dissolved solids is generally satisfactory for domestic and industrial use while ground water having greater than 1000 mg/L of total dissolved solids is generally unsatisfactory for these uses. High total dissolved solids are often indicative of other characteristics such as hardness. Other properties that are especially useful in determining ground water character are hardness, specific conductance and pH (Table 2.2). These constituents can be determined by simple procedures using field equipment. In order to more precisely identify and measure the quality characteristics of ground water, chemical, physical and biological analysis are usually required. Chemical analysis requires the laboratory determinations of the concentrations of common ions found in ground water and are commonly reported in units of milligrams per litre (mg/L). Concentrations may also be expressed as equivalents per million (epm) which is the moles of solute multiplied by the valence of the solute species in 10,000,000 g of water.

Properties of ground water often evaluated in a physical analysis include temperature, turbidity, odor, taste and colour. Biological analysis includes coliform bacteria tests which indicate the sanitary quality of the water for human consumption. A list of recommended parameters for chemical analysis is given (Table 2.5). Additional parameters may be tested if a more detailed analysis is requested or where known or suspected sources of pollution exist.

Table 2.5 Recommended parameters for chemical analyses (guidelines for minimal standards in water well construction, WHO 1984.)

parameter	mg/l
Total alkalinity	50.00
Silver	0.05
Chloride (Cl ⁻)	250.00
Fluoride (F ⁻)	1.50
Iron (Fe)	0.30
Manganese (Mn)	0.05
Lead (Pb)	0.05
Nitrate (NO ₃ ⁻)	45.00
Sulphate (SO ₄ ²⁻)	350.00
Total Dissolved Solids (TDS)	500.00
pH	6.50
Zinc (Zn)	5.00

Some substances even in small concentrations can be troublesome. For example, iron concentrations of 1 to 5 mg/L in ground water and can cause staining to plumbing fixtures and laundry, encrust well screens and clog pipes. Manganese in small concentrations can also cause staining and is even more objectionable as stains are harder to remove than those caused by iron. Chloride contamination is possible in wells located near to the sea where pumping of these wells can move sea water into the freshwater aquifer making water unpotable. Groundwater containing dissolved hydrogen sulphide gas is another common problem which imparts a characteristic "rotten egg" odor and taste to the water. Hydrogen sulphide will combine with other impurities in the water to form iron sulphide (black water), calcium sulphide, sodium sulphide, and so forth. Table 2.6 lists these and other substances found naturally in some ground waters which can cause problems in operating wells.

Table 2.6 Substances found naturally in some groundwater which can cause problems in operating boreholes.

substance	Type of problems
Iron (Fe)	Encrustation, staining of laundry and toilet fixtures
Manganese (Mn)	Encrustation, staining of laundry and toilet fixtures
silica	Encrustation,
Chloride (Cl^-)	Potability, corrosiveness
Fluoride (F^-)	Fluorosis
Nitrate (NO_3^-)	Methemoglobinemia
Sulphate (SO_4^{2-})	Potability
calcium and magnesium	Encrustation,
carbon dioxide	Corrosiveness
Dissolved gases	Corrosiveness
pH	Corrosiveness
Minor constituents	Potability, health aspects

TDS, specific conductance, hardness and pH are good an indicator of the character of the groundwater. Tritium (^3H), which is a radioactive isotope, is good indicator of the age of groundwater (Kenneth 1991). Carbon-isotopic measurements on dissolved inorganic carbon were used to provide information on the relative ages of the ground water pumped from various locations (Alyssa and James 2006). Delineation of the spatial distribution of ground water that is several thousands of years old can provide an important piece of the puzzle in the evaluation of long-term, ground water resource sustainability under pumping conditions. Between 1952 and 1962, large scale atmospheric testing of thermonuclear bombs was carried out and atmospheric contamination occurred. It is therefore apparent that ground water from a location in the northern hemisphere, containing tritium at levels of hundreds or thousands of TU (tritium units), entered the ground water zone after 1953. If the water has less than 5-10 TU it must have entered the ground water zone prior to 1953.

2.8.5: BOREHOLE SURVEYS

These are the only tools for direct observations of rock surface and features occurring at depth. A number of methods are available. Geophysical well logging is a standard

technique, including electrical, caliper, radioactivity, magnetic logging etc. These give information on lithology and structure.

2.8.6: POTENTIAL RISK TO AQUIFER

The term contamination is used for addition of any solute into the hydrological system as a result of man's activity while the term pollution is restricted to a situation when the contamination attains levels that are considered to be objectionable (Freeze and Cherry 1979). There could also be deterioration in water quality due to natural reasons namely dissolution of rock material. Contamination of groundwater can take place from either a wider source like percolation from agricultural fields on account of application of fertilizers and pesticides or from a point source like waste disposal sites. In a polluted atmosphere, many oxidizing compounds (e.g. ozone), acid-forming gases (oxides of nitrogen and sulfur) and particulate material will be higher than in an unpolluted atmosphere. This will be area specific depending on the anthropogenic activities. For example, combustion of fossil fuels results in acid rains containing sulphuric acid (H_2SO_4) and nitric acid (HNO_3), which reduces the pH of source water. This induces greater solubility of aquifer material.

CHAPTER THREE

3.0 METHODOLOGY

3.1 INTRODUCTION

Though there is inadequate hydrogeochemical information concerning the quality of the groundwater in the study area, borehole water has been used and continues to be used extensively. Both primary and secondary data collection techniques were used in the data acquisition. In the primary method, information was sought from the field due to inadequacy of the existing information. This prompted the need to acquire more information directly from the field by collecting samples. The purpose of the fieldwork was thus to fill up the gap in the existing secondary data. Sixty percent of data used was primary and forty percent secondary. The field work lasted six days between 17th - 22nd April 2011 and involved the use of:

- Topographical and geological maps;
- Different kits e.g. GPS, notebooks, camera etc;
- Fixed laboratory analyses (Na, K, Mg, Zn ,Ca, Fe); and
- Software's for processing data and evaluation.

3.2 LOCATIONS OF BOREHOLES

Locating the positions of boreholes sampling sites in the study area was done with the aid of a Global Positioning System (GPS) and local topographical map of the area (Fig. 3.1).

3.3 FIELD WORK SURVEY

Reconnaissance survey was done for two days from 17th -18th April 2011. The survey involved the division of the area into two blocks. The first trip was to the south (Namasali, Bumayenga, Bukiri, Muramba, Bufuma, Naderema, Luchululo and Burudu) and the second trip to the north (Wakhungu, Nangina, Malanga, Saganja, Nakhasiko and Namboboto). This was to:

- verify that the information about boreholes location provided by secondary source was correct and
- familiarize with the area and to assess the cost of the work.

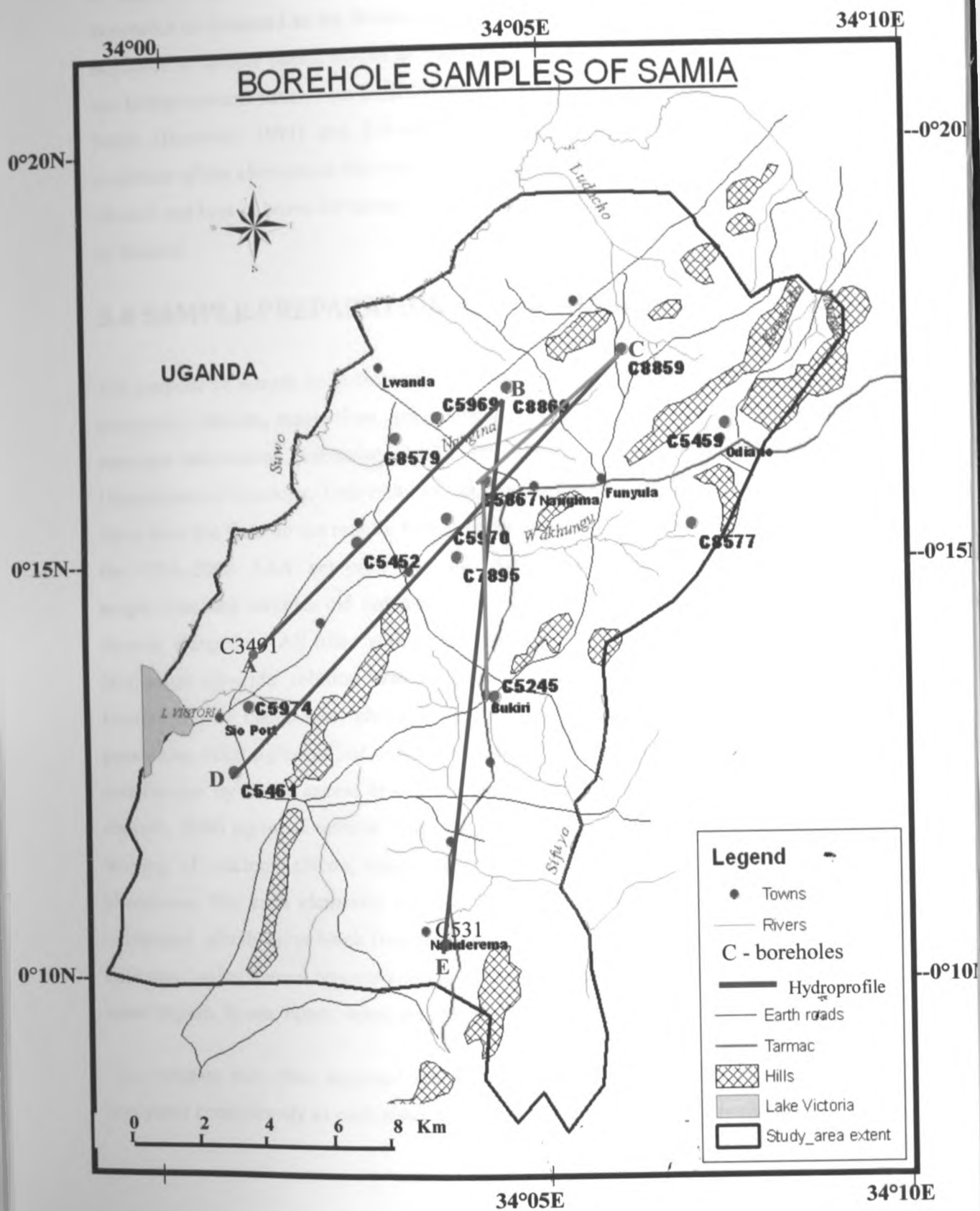


Fig 3.1: Locations of borehole sampling sites and hydrogeological profile traverses.

A major component of the field work involved water sample collection from the boreholes as indicated in the location points (Fig. 3.1). The samples were collected in duplicate in sample plastic bottles of 1000 ml. Contamination was reduced by rinsing the bottles several times with borehole water. The water was filled to the brim of the bottle (Horowitz 1991) and 2-3 ml of concentrated nitric acid added to prevent oxidation of the elements in the water that would lead to precipitation, sealed (capped), labeled and kept in boxes for transportation to the analytical laboratory at the University of Nairobi.

3.4 SAMPLE PREPARATION AND ANALYSIS

The purpose of sample collection was to analyze seven elemental metal ions (sodium, potassium, calcium, magnesium, iron, manganese and zinc) that were missing from the previous information (secondary data). The collected samples were analyzed at the Department of Geology, University of Nairobi. Since the samples collected in liquid form from the field do not require further preparations, they were aspirated directly into the CTA-2000 AAS analyzer. For the analysis of iron, manganese, zinc and magnesium, the samples did not require any dilution or buffering and were analyzed directly using the AAS fitted with an appropriate cathode. In the analysis of calcium, lanthanum chloride solution was added into the liquid sample to overcome the interference of manganese, phosphate and bicarbonate alkalinity. In the analysis of potassium, 1000 $\mu\text{g/ml}$ of sodium ions was added to the sample water to suppress the interference by heavy excess of calcium, magnesium and sodium. For the analysis of sodium, 2000 $\mu\text{g/ml}$ potassium ions was added to the sample water to ensure heavy loading of calcium, silicon, magnesium potassium and phosphate do not affect the absorption. For each elemental analysis, the CTA 2000 AAS instrument had to be calibrated afresh using blank (liquid with no ions present in it) samples first, and then calibrated using known concentration to ensure the accuracy. The known concentrations were 10ppm, 8ppm, 6ppm, 4ppm and 2ppm.

The samples were then aspirated into the machine and the results read off from the computer continuously as each water sample was being aspirated and analyzed.

3.5 DATA QUALITY

The precision of the analytical work depended on the calibration but was within the acceptable limits of error i.e. 0.01 %. The detection limit of the instrument is 0 ppm to 10 ppm. Any concentration above that limit must be diluted appropriately to within the limits and multiplied by the dilution factor. The results are printed in Chapter Four, Table 4.2.

3.6 STATISTICAL ANALYSIS

The data obtained from the primary and secondary sources were analyzed for correlation coefficient using Excel and Principal Component Analysis (PCA is linear combinations of optimally-weighted observed variables) using the Statistical Package for Social Sciences (SPSS) program. Further Principal Factor Analysis (PFA) was carried out on the selected groundwater data parameters from the wells using SPSS program. The results are presented from Table 4.9 to 4.13.

In the correlation coefficient the results range from -1 to +1. As the value approaches -1 means that as the value of X increases that of Y reduces. As the value approaches 0, it means that there is no correlation between the values. As the value approaches +1, it means that the values are very strongly positively correlated meaning that increase in the value of X corresponds to the increase in the value of Y or as the value approaches -1, it means that the values are very strongly negatively correlated meaning that decrease in the value of X corresponds to the increase in the value of Y. This was done in order to bring out the differences and similarities that would enable one to group the wells in terms of the analyzed parameters and to compare them with the results obtained from the Piper diagram.

3.7 CHALLENGES

The challenges faced during the field work included;

- Some boreholes which were well documented in the Ministry of Water and Irrigation's record are non-existent on the ground like Machakha and Musokho boreholes. They could exist in another locality but not in Samia area.

- Some boreholes in the area were in the custody of caretakers, who locked them and therefore one needed permission before water could be drawn, like the case in Muramba, Sagania and Namboboto.
- The fieldwork took place during the unexpected fuel shortage period and therefore, sometimes delaying the commencement of the field work.
- Laboratory analysis was delayed due to the absence of analytical gas, but it was eventually brought in from the Chemistry Department of the University of Nairobi.

CHAPTER FOUR

4.0 RESULTS AND INTERPRETATION

4.1 WELL LOGS

A number of boreholes have been sunk in the area. For some of them, the information available is incomplete and include; the well number, owner, location, total depth, water struck level, water rest level and well discharge. The data of the boreholes logs drilled in the study area are shown in the Table 4.1

Table 4.1 Well logs for some borehole data of the Samia study area. (courtesy of the Ministry of Water and Irrigation Headquarters. District Water Department-Busia, Water Resource Management Authority (WRMA) Kakamega.

Sample source	Well No.	Location	Elevation (m)	T/depth (m)	WSL. (m)	WRL (m)	Q(L/min)
Bukiri mkt	C5245	36N 0620261 UTM 0028448	1207	40	28	14.63	31.35
Muramba sch	C5452	36N 0617707 UTM 0029816	1176	79	68	32.2	**
Wakhungu sch.	C5459	36N 0627999 UTM 0032612	1320	37	22	10.2	**
Bumayenga sch.	C5461	36N 0620003 UTM 0021732	1192	49	33	20	**
Nangina sch.	C5967	36N 0619765 UTM 0032635	1183	64	51	28.10	34
Nakhasiko vill.	C5969	36N 0621695 UTM 0033651	1200	46	32	16	18
Bufuma vill.	C5970	36N 0620041 UTM 0030502	1183	64	43	25	**
Namasali mkt	C5974	36N 0614250 UTM 0025429	1143	**	**	**	**
Burudu Sch	C7895	36N 0621587 UTM 0031257	1197	67	35	19.72	23.53
Malanga sch.	C8577	36N 0625553 UTM 0030671	1242	45	24	17	35
Sagania vill.	C8859	36N 0624799 UTM 0034384	1220	37	26	17.63	26.52
Namboboto sch.	C8869	36N 0618349 UTM 0031972	1152	52	24	30.34	41
Nanderema sch.	C5311	36N 0620003 UTM 0021730	1195	40	30	27.17	**

****No data available**

From Table 4.1 the water struck level ranged from 22- 68 m and is below the water rest level. This rise in the water levels indicates the increase in the piezometric pressure due to semi-confinement of the aquifer. The discharge from the wells also varies greatly from 18 - 41 L/min.

Again, from Table 4.1 and Fig. 3.1, the hydrogeological cross-sections were constructed from the directions as illustrated in Fig. 4.1 and shown in Fig. 4.2 (Kuria and Barongo 2006). From the hydrogeological profiles, it is apparent that the main aquifer in the study area is semi-confined and uniformly follows the topography but is slightly subdued.

The study has also shown that if perched aquifers do not exist in the area then, the predicted groundwater flow direction indicated in Fig. 4.1 holds because the discharge from the well no. C -5967 at Nangina sch. where the flow direction appear to converge is higher than the other wells shown and the water rest level in the well is just thirteen meters from the surface as illustrated in Fig. 4.1. The area around well no. C-5967 at Nangina school is also the source of river Nangina and its tributaries.

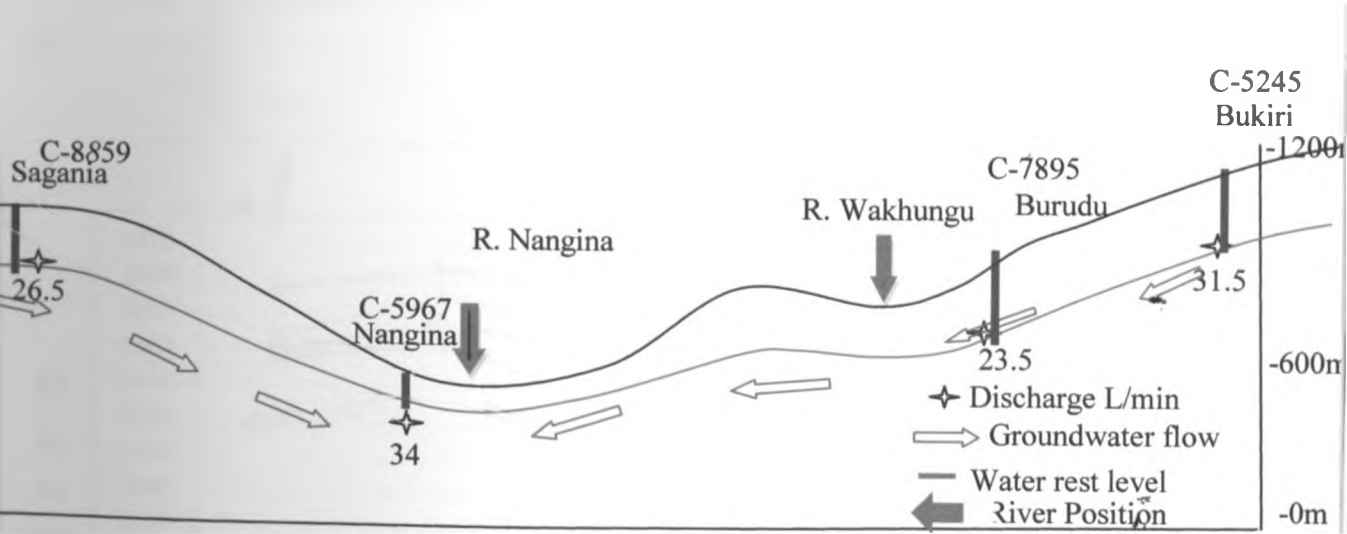


Fig. 4.1: An illustration of Nangina catchment area using groundwater flow direction and discharge values as indicated by green () line section. Not drawn to scale

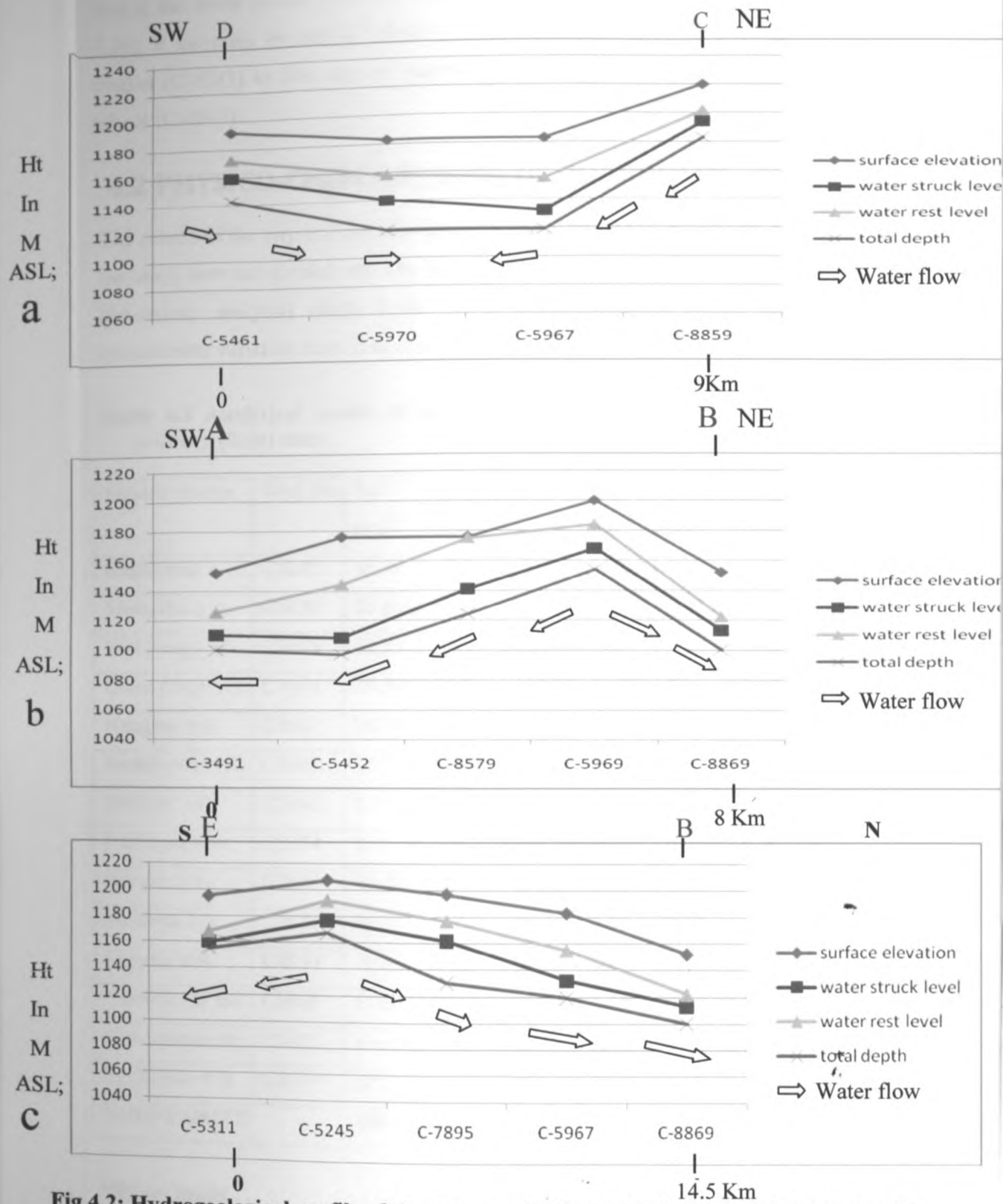


Fig 4.2: Hydrogeological profile of the traverses indicated in Fig 3.1 e.g. D-C (a), A-B (b)

Chemical analysis of the water from some of the wells along the profiles indicate that, as the water travels down the flow path, from area of recharge to discharge (Fig. 4.2c), it increases in total dissolved solids (Table 4.3) e.g. from 156 mg/l at Bukiri market (C-5245) to 246 mg/l at Burudu village (C-7895) to 360 mg/l at Namboboto school (C-8869).

4.2 PHYSICO-CHEMICAL DATA OF BOREHOLES

The results of the physico-chemical analyses of the groundwater samples undertaken in the study area are divided into two sets. Table 4.2 are values from the primary source (laboratory analysis) while Table 4.3 is from secondary source and exhibited considerable variation from sample to sample.

Table 4.2 Analytical results of the seven elements in water samples from the primary (field) data.

Sample source	Well No.	Na ⁺ mg/l	K ⁺ mg/l	Ca ²⁺ mg/l	Mg ²⁺ mg/l	Zn ²⁺ mg/l	Fe ²⁺ mg/l	Mn ²⁺ mg/l
Bukiri mkt	C5245	39.75	4.77	35.60	7.93	0.02	0.06	0.09
Muramba sch	C5452	85.60	3.98	120.00	111.40	0.06	0.00	0.00
Wakhungu sch.	C5459	32.35	2.38	18.92	5.00	0.00	0.00	0.05
Bumayenga sch.	C5461	49.70	4.20	138.50	16.95	0.00	0.10	0.00
Nangina sch.	C5967	26.50	13.80	14.16	3.08	0.00	0.13	0.02
Naakhasiko vill.	C5969	55.15	27.25	168.20	56.20	0.00	0.02	0.00
Bufuma vill.	C5970	113.90	28.00	118.00	35.33	0.016	0.06	0.01
Namasali mkt	C5974	111.20	7.04	326.20	49.70	0.02	0.00	0.12
Burudu Sch	C7895	45.25	5.43	60.52	6.58	0.00	0.05	0.00
Malanga sch.	C8577	46.65	2.23	36.04	9.68	0.02	0.00	0.40
Sagania vill.	C8859	56.40	2.49	27.52	10.80	0.04	0.00	0.00
Namboboto sch.	C8869	41.85	32.10	21.84	5.40	0.01	0.00	0.00
Nanderema sch.	C5311	43.15	4.42	160.00	33.35	0.00	0.18	0.00
Luchululo vill.	C8579	125.20	10.56	235.40	87.88	0.01	0.07	0.00
WHO LIMITS		200	*	300	*	5	0.3	0.1

*No set limits

Table 4.3 Physico-chemical data on groundwater quality in the study area (courtesy of the Ministry of water and Irrigation, District Water Department-Busia).

Loction	Well No.	pH	TDS mg/l	Ec (μ S/cm)	T ^{ALK} mg/l	Cl ⁻ mg/l	F ⁻ mg/l	SO ₄ ²⁻ mg/l	NO ₃ ⁻ mg/l	T ^{HARD} mg/l	HCO ₃ ⁻ mg/l	CO ₃ ⁻² mg/l	CLR
Bukiri mkt	C5245	7.4	156	350	182	8	0.2	4	8.8	176	167	15	20
Muramba sch	C5452	6.9	1920	3200	334	525	0.75	110	12	1610	267	67	0
Wakhungu sch.	C5459	7.3	264	440	224	3	0.1	6	8	200	202	22	25
Bumayenga sch.	C5461	7.5	510	850	350	45	0.25	280	13	400	332	18	0
Nangina sch.	C5967	7.3	90	150	325	15	0	32	2.2	107	292	33	0
Nakhasiko vill.	C5969	7.0	2280	3800	430	538	0	270	2.7	1286	357	73	5
Bufuma vill.	C5970	6.5	708	1180	277	86	0.4	80	1.8	367	174	103	0
Namasali mkt	C5974	7.5	402	670	288	26	0	0	8.7	318	274	14	5
Burudu Sch	C7895	7.4	246	400	0	4	0	4	0	224	0	0	15
Malanga sch.	C8577	6.5	180	300	168	4	0	5	0	163	106	62	5
Sagania vill.	C8859	7.5	180	300	158	3	0	190	0	142	150	8	5
Namboboto sch.	C8869	7.0	360	600	0	10	0	0	0	298	199	41	5
Nanderema sch.	C5311	7.1	1440	2400	288	260	0	150	15.3	1148	251	37	4
Luchululo vill.	C8579	7.0	240	400	239	2	0	15	0	233	198	41	5
WHO LIMITS		6.5-	500	*	*	250	1.5	350	45	*	*	*	-

From the tables, the most abundant cation in the area was calcium with a mean of 105 mg/l, followed by sodium with an average of 62.33 mg/l. Magnesium is third with a mean of 31.38 mg/l and the least abundant cation was potassium with a mean of 10.62 mg/l. Total hardness caused by dissolved calcium and magnesium ions in water ranged from 1071-1610 mg/l. The taste threshold for the calcium ions is between 100-300 mg/l, while that of magnesium is probably less than that of calcium.

Bicarbonate is the most abundant anion with an average of 212.07 mg/l while chloride is the second with a mean of 109 mg/l and sulphate third with a mean of 81.86mg/l. Carbonate with an average of 38.14 mg/l was least abundant. . Nitrate was detected in several wells ranging in concentration from 0-15.3mg /l. Fluoride concentration ranged from 0-0.75 mg/ l. Sulphate proportions in the water range from zero in some areas such as Namasali mkt. to 270 mg/l and 280 mg/l in Nakhasiko village and Bumayenga school respectively, While the concentrations of chloride in water are within the threshold, except in a few areas where the proportions are far above the WHO recommended values of 250 mg/l and reach 525 mg/l and 538 mg/l in Muramba school and Nakhasiko village respectively.

The presence of zinc in the water samples in the study area is in insignificant amounts. It ranged from 0-0.058 mg/l, and seldom does drinking water contain levels above 0.1 mg/l. The concentration of iron in the water sampled is between 0-0.18 mg/l, well within the recommended limits of 0.3 mg/l while Manganese values in the sampled water were very low and ranged between 0-0.4 mg/l. and exceeded the WHO values of 0.1 mg/l at Malanga school and Namasali market.

The following were also observed. The total alkalinity in the waters ranged from 0-430 mg/l. The pH values of the water sampled in the study area showed little variation and ranged from 6.5-7.5. The total dissolved solids in the water from the study area ranged from 180-2280 mg/l. Drinking water with TDS levels greater than the WHO recommended guidelines of 500 mg/l are found in Bufuma village, Muramba School, Nakhasiko village and Nanderema School and is increasingly unpalatable.

The electrical conductivity of the water is a function of TDS. The conductivity ranged from 150- 3800 μ S/cm.

4.3 CHARGE BALANCE ERROR

The cation-anion balance or charge balance error (CBE) was performed for the anions and cations (Murray and Wade, 1996) from the water samples, sometimes can point out some obvious errors in any water quality analysis (Table 4.3). A fundamental property of an aqueous solution that they are electrically neutral, so the total number of equivalents for the positive constituents must equal to total number of equivalents of negative constituents.

The charge balance error (CBE) is obtained from the following equation

$$((\text{Total Cation meq/L} - \text{Total Anion meq/L}) / (\text{Total Cation meq/L} + \text{Total Anion meq/L})) * 100\%$$

“Negative” number for “cation – anion balance” means:

- Excess anions or
- Lack of cations

While “positive” number for “cation – anion balance” means:

- Excess cations
- or, Lack of anions

A positive CBE indicates that one or more of the cations was over determined, one or more of the anions was underdetermined, or both.

Conversely, a negative CBE indicates that one or more of the cations was underdetermined, one or more of the anions was over determined, or both. It is important to note that CBE is a relative error, or percentage (rather than an absolute error). Due to the presence of the denominator the absolute error tends to be larger for more concentrated solutions than for dilute solutions. In other words, the fewer ions in the solution the more sensible is the CBE.

Table 4.4 Milliequivalents data for the boreholes

Borehole No	Ca ²⁺	Mg ²⁺	Na ⁺	K ⁺	T/cation	HCO ₃ ⁻	CO ₃ ²⁻	SO ₄ ²⁻	Cl ⁻	T/anion	CBE
C5245	1.78	0.65	1.73	0.12	4.25	2.74	0.50	0.08	0.23	3.55	8.97
C5452	5.98	9.17	3.72	0.10	18.97	4.38	2.23	2.29	14.79	23.69	11.06
C5459	0.94	0.41	1.41	0.06	2.82	3.31	0.73	0.13	0.08	4.27	-20.45
C5461	6.91	1.40	2.16	0.11	10.58	5.44	0.60	53.76	1.27	61.07	-70.44
C5967	0.71	0.25	1.15	0.35	2.46	4.79	1.10	0.67	0.42	6.98	-47.88
C5969	8.39	4.63	2.40	0.70	16.12	5.85	2.43	5.63	15.15	29.06	-28.64
C5970	5.89	2.91	4.95	0.72	14.47	2.85	3.43	1.67	2.42	10.37	16.51
C5974	16.27	4.09	4.83	0.18	25.37	4.49	0.47	0.00	0.73	5.69	63.40
C7895	3.02	0.54	1.97	0.14	5.67	0.00	0.00	0.08	0.11	0.19	93.52
C8577	1.80	0.80	2.03	0.06	4.69	1.74	2.07	0.10	0.11	4.02	7.69
C8859	1.37	0.89	2.46	0.06	4.78	2.46	0.27	3.96	0.08	6.77	-17.23
C8869	1.09	0.44	1.82	0.82	4.17	3.26	1.37	0.00	0.28	4.91	-8.15
C8579	11.74	7.23	5.44	0.27	24.68	3.24	1.37	0.31	0.06	4.96	66.53
C5311	7.98	2.74	1.88	0.11	12.71	4.11	1.23	3.13	7.32	15.79	-10.81
total					151.74					181.32	

A reasonable balance for routine water quality analysis is generally considered to be less than 10%. If you have a higher number, you should reconsider further constituents for analysis. In the case for Samia the results of CBE are varied for each borehole. There are several reasons that may also create an electrical imbalance such as:

1. Laboratory error where it is simply virtually impossible to analyze every single ions in the water. In most cases it is also not required. In other words, the less ions in the solution the more sensible is the CBE.
2. Sampling procedure.
3. Use of unfiltered water sample which contain particulate matter that dissolves into the sample when we add small acid as preservative.
4. Sometimes some of the constituent may precipitate in the sample container e.g. the carbonates of calcium and magnesium or the sulphate of calcium.

The CBE for the water sample in Samia varied because of two reasons; basically the use of unfiltered (turbid) water that could have dissolved when few drops of the acid were added and the precipitation of the ions in the sample container.

The analytical results for calcium, magnesium, sodium, bicarbonate, carbonate, chloride and sulphate were grouped into cations and anions (Fetter 2001) and converted to percentage milliequivalents shown in Table 4.4.

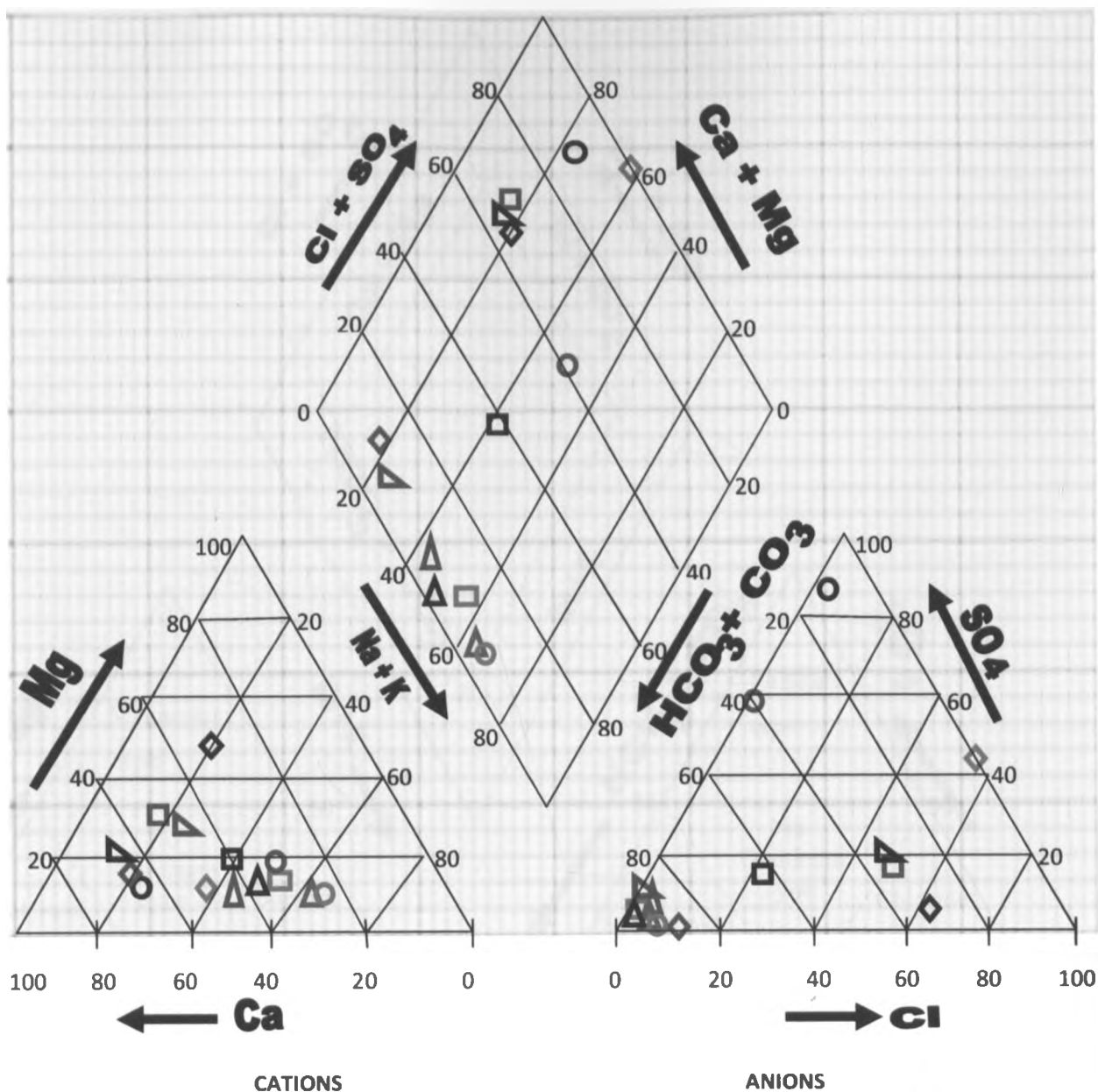
Table 4.5 Percentage chemical data for the boreholes water (in % milliequivalent)

Borehole No.	Ca ²⁺	Mg ²⁺	Na ⁺ + K ⁺	%	HCO ₃ ⁻ + CO ₃ ²⁻	SO ₄ ²⁻	Cl ⁻	%
C5245	41.59	15.17	43.24	100	91.2	2.3	6.5	100
C5452	31.56	48.31	20.13	100	27.9	9.7	62.4	100
C5459	33.33	14.54	52.13	100	95.0	3.1	1.9	100
C5461	65.31	13.23	21.46	100	9.9	88.0	2.1	100
C5967	28.86	10.16	10.16	100	84.38	9.60	6.02	100
C5969	52.05	28.72	19.23	100	28.49	19.37	52.14	100
C5970	40.70	20.12	39.18	100	60.56	16.10	23.34	100
C5974	64.14	16.13	19.73	100	87.17	0.0	12.83	100
C7895	53.26	9.52	37.22	100	0.0	42.11	57.89	100
C8577	38.38	17.06	44.56	100	94.78	2.49	2.73	100
C8859	28.66	18.62	52.72	100	40.32	58.50	1.18	100
C8869	26.14	10.55	63.31	100	94.30	0	5.70	100
C8579	47.57	29.29	23.14	100	92.57	6.22	1.21	100
C5311	62.79	21.56	15.65	100	33.82	19.82	46.36	100

The information from the tables was plotted on a Piper and Scheoller diagram (Fig 4.3). From the plots it is apparent that the similarities and differences between the cations and the anions exist. Some cations and anions are clustered together while others are scattered or not very close. The cationic side shows a wide distribution between calcium, magnesium and sodium-potassium. On the other hand there is a big cluster of anions concentrated towards bicarbonate and carbonate (alkalinity) while little bias in the direction of sulphate. Chloride bias was considerable.

It is apparent from the plots that samples from C-5311, C-5461 and C-7895 were dominated by calcium ions (Fig. 4.4) while Samples from C-8859, C-5967 and C-5459 were dominated by sodium and potassium ions. Magnesium was the dominant ion in C-5452. Samples from C-8579, C-5970, C-5254, C-8577 and C-8869 lie in the centre and showed no bias towards any major cation. The carbonate and hydrogen carbonate anions are the dominant type except for the samples from C- 7895 which are chloride type and samples from C-5461 which are sulphate type. Samples from C-5311 did not show any bias towards any anion suggesting that Samia area is composed of hydrochemical facies or aquifers (groundwater fields) that differ considerably in chemical composition Fig. 4.4.

The Schoeller semi-logarithm diagram provided supplementary support to the piper diagram in order to justify the categorization (Fig. 4.5), the peaks and troughs indicate great variations in the water chemistry from the study area.



Legend	B/H No.	Legend	B/H No.	Legend	B/H No.
▲	C-5245	◻	C-5970	◼	C-5311
◊	C-5452	◈	C-5974	◼	C-8579
◻	C-5459	◈	C-7895	◻	C-5969
○	C-5461	▲	C-8577	○	C-8869
▲	C-5967	○	C-8859		

Fig 4.3: Piper diagram of the types of water chemistry of the water samples in the area.

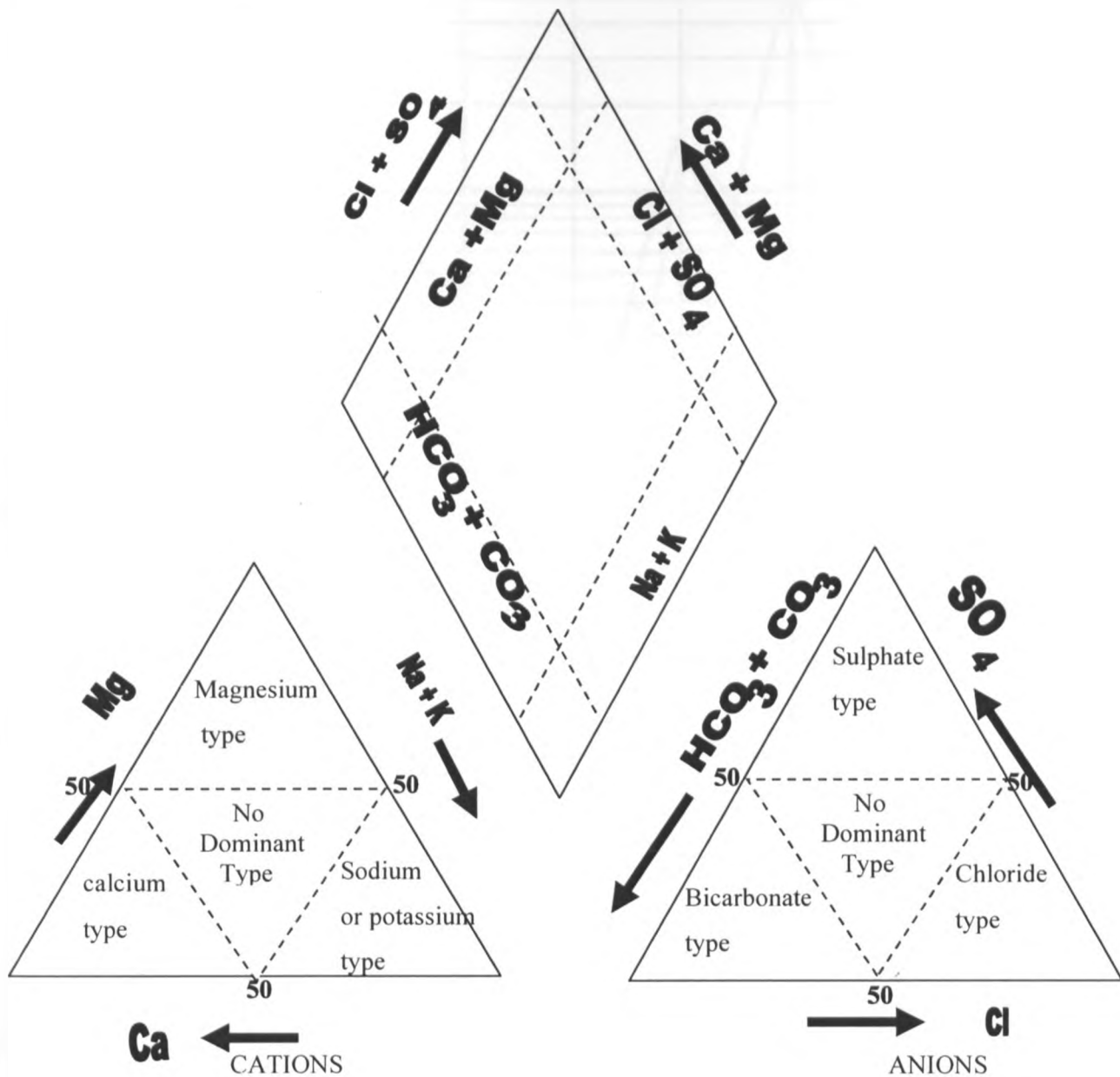
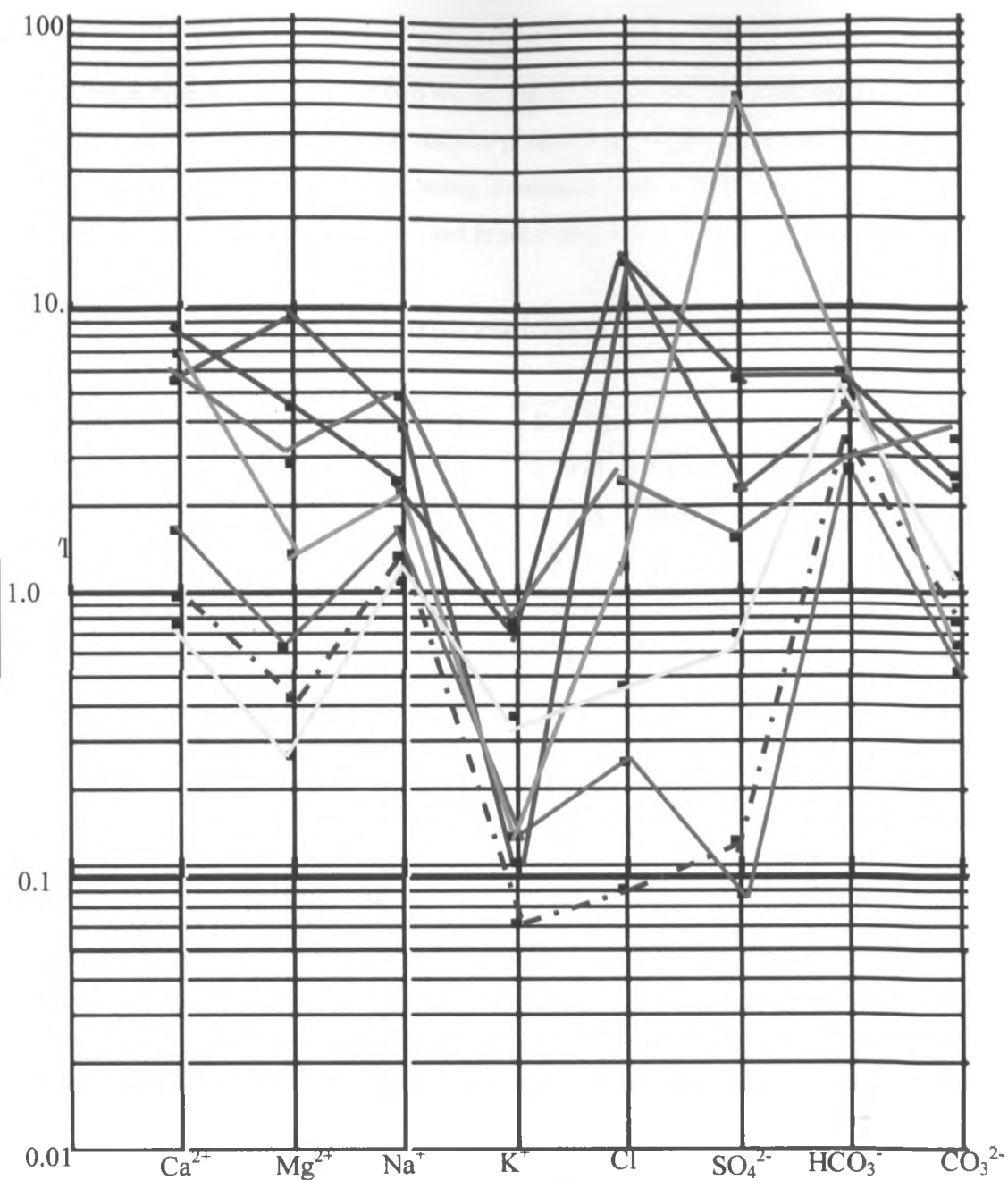


Fig 4.4: Hydrogeochemical classification system for natural waters using the piper diagram.

Units in
Logarithm



B/HOLE No.	Legend	B/HOLE No.	Legend	B/HOLE No.	Legend	B/HOLE No.	Legend
C-5461		C-5967		C-5245		C-5311	
C-5452		C-5970		C-5459			

Fig 4.5: Schoeller semi-logarithm diagram with some chemical values from Samia Area.

The dominant anion in the waters found in the study area is the bicarbonate type which is found in seven boreholes, followed by the chloride type found in four boreholes while

sulphate type is only found in two wells. The dominant cation in the waters in the study area is between the calcium and sodium-potassium, giving rise to calcium and sodium-potassium bicarbonate type as being dominant. The water types in Samia can be categorized as shown in Table 4.7 and represented in Fig. 4.6.

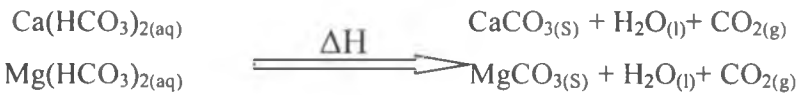
Table 4.6: Summary of water types available in Samia area.

B/H.No	LOCATION	WATER TYPE	B/H.No.	LOCATION	WATER TYPE
C5245	Bukiri mkt	bicarbonate	C5974	Namasali mkt	Calcium bicarbonate
C5452	Muramba sch	Chloride	C7895	Burudu Sch	Calcium chloride
C5459	Wakhungu sch.	Sodium bicarbonate	C8577	Malanga sch.	bicarbonate
C5461	Bumayenga sch.	Calcium sulphate	C8859	Sagania vill.	Sodium sulphate
C5967	Nangina sch.	Sodium bicarbonate	C8869	Namboboto sch.	Sodium bicarbonate
C5969	Naakhasiko vill.	Calcium chloride	C8579	Luchululo vill.	bicarbonate
C5970	Bufuma vill.	bicarbonate	C5311	Nanderema sch	Calcium chloride

4.4 WATER QUALITY

The suitability of the water for domestic purposes which involves drinking and other purposes like cooking and washing cannot be summed by one parameter. As for drinking purposes, water from the study area is good except for wells like C-5452, C-5969 and C-5311 which have total dissolved substances of 1920mg/l, 2280mg/l and 1440mg/l respectively and are above normal (WHO 1984) guideline and therefore unpalatable. These could be overcome by:

Boiling,



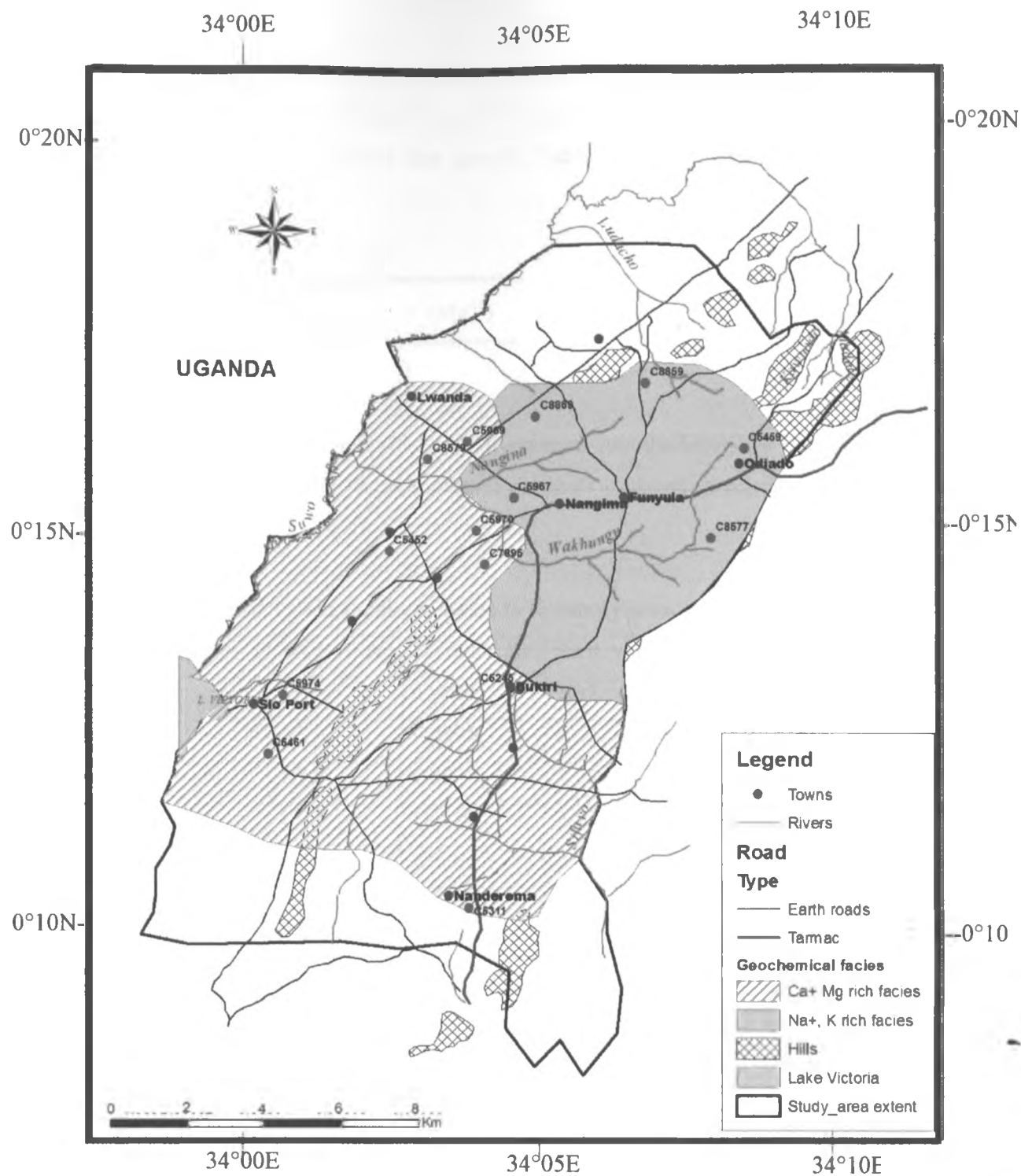


Fig 4.6: Hydrogeochemical facies of Samia area

or treatment with sodium carbonate.



The sodium hazard of soil usually is expressed as the sodium adsorption ratio (SAR). This is the proportion of water soluble Na^+ to Ca^{2+} plus Mg^{2+} in the soil. Ions in the equation are expressed in millequivalents per liter (meq/L, Table 4.3). The formula used to calculate SAR is shown below.

$$\text{SAR} = \frac{\text{Na}^+}{\sqrt{\frac{(\text{Ca}^{2+}) + (\text{Mg}^{2+})}{2}}}$$

The sodium absorption ratio (SAR) of the waters from the boreholes were calculated to determine its suitability for agricultural purposes especially irrigation and the results shown in Table 4.8

Table 4.7: SAR values for the samples in the study area.

B/HOLE No.	SAR Values	B/HOLE No.	SAR Values
C5245	1.57	C5974	1.57
C5452	1.35	C7895	1.48
C5459	1.72	C8577	1.78
C5461	1.06	C8859	2.31
C5967	1.66	C8869	2.08
C5969	0.94	C8579	1.77
C5970	2.36	C5311	0.8

As for the agricultural use, the Sodium Absorption Ratio calculated results are shown in Table 4.8. The SAR values range from 0.94 to 2.36. Low SAR (2-10) indicates little danger from sodium accumulation in the soil which makes it sticky and uncultivable. These make the water in the study area very suitable for agricultural purposes especially irrigation.

In the case of industrial use, the hardness was found to be both permanent and temporary type. Though there are no industries in the study area, the water is often used as a coolant in motor vehicles and posho mills. Most of the hardness in the water is the

bicarbonate type, which is termed as temporary hardness and can be removed by boiling.

4.5 CORRELATION COEFFICIENT

The correlation coefficient analyses were also carried out on all the physico-chemical parameters present from various boreholes in the study area as shown in Table 4.8. From the correlation coefficient, the variability is comparatively small. Sodium, calcium and magnesium are highly correlated with values of 0.71, 0.74 and 0.64.

Potassium which is a major element in the biotite rock (Table 4.8) has little correlation with the rest with values of 0.16 with sodium, 0.02 with calcium and 0.03 with magnesium implying that rocks in the area consists of little biotite. On the other hand sodium is highly correlated with calcium (0.74) and with magnesium (0.71) implying that these elements are derived from the rholite, dolerites and diorites. Chloride is highly correlated with TDS, with a correlation of 0.98. Electrical conductivity and total hardness correlated with a value of 0.97. Correlation between iron and manganese was very high with a value of 0.89 and their source could be andesites and dacites.

Table 4.8: Correlation co-efficient values of the elemental parameters

	Ec	TDS	pH	TH	K	Na	Ca	Mg	Fe	Mn	Cl ⁻	F ⁻	HCO ₃ ⁻	CO ₃ ²⁻	SO ₄ ²⁻	NO ₃ ⁻
Ec	1.00															
TDS	1.00	1.00														
pH	-0.19	-0.20	1.00													
TH	0.97	0.98	-0.21	1.00												
K	0.25	0.25	0.42	0.10	1.00											
Na	0.11	0.12	0.30	0.11	0.16	1.00										
Ca	0.32	0.33	0.05	0.31	0.02	0.74	1.00									
Mg	0.62	0.62	-0.29	0.67	0.03	0.71	0.64	1.00								
Fe	0.10	0.10	0.13	0.08	0.14	0.28	-0.05	-0.12	1.00							
Mn	-0.17	-0.16	0.09	-0.1	0.04	0.29	-0.28	-0.21	0.89	1.00						
Cl ⁻	0.98	0.98	-0.19	0.96	0.19	0.09	0.26	0.65	0.13	-0.08	1.00					
F ⁻	0.21	0.21	-0.21	0.27	0.05	0.29	0.04	0.52	0.20	-0.05	0.25	1.00				
HCO ₃ ⁻	0.53	0.54	0.34	0.11	0.23	0.06	0.41	0.36	0.36	0.25	0.49	0.09	1.00			
CO ₃ ²⁻	0.50	0.50	-0.85	0.43	0.57	0.38	0.01	0.43	0.06	0.05	0.47	0.30	0.11	1.00		
SO ₄ ²⁻	0.61	0.61	0.20	0.55	0.06	0.08	0.16	0.17	0.10	-0.11	0.56	0.02	0.58	0.10	1.00	
NO ₃ ⁻	0.42	0.41	0.34	0.53	0.42	0.08	0.30	0.23	0.05	-0.05	0.36	0.02	0.57	-0.13	0.40	1.00

The analyses also indicated that the temporary hardness of water in the area was due to the presence of the bicarbonate of calcium rather than magnesium with a correlation of 0.41. The major permanent hardness in water is due to the chloride and sulphate of magnesium.

4.6 PRINCIPAL COMPONENT ANALYSIS

Further analysis of the water quality data was performed with principal components analysis (PCA) after Jolliffe (2002) and yielded Table 4.9- 4.13.

Table 4.9: Communalities Extraction Method: Principal Component Analysis.

	Initial	Extraction
Na	1.000	.951
K	1.000	.896
Ca	1.000	.839
Mg	1.000	.907
HCO ₃	1.000	.716
CO ₃	1.000	.953
CLR	1.000	.524
Zn	1.000	.762
pH	1.000	.940
TDS	1.000	.942
Ec	1.000	.943
Fe	1.000	.614
Mn	1.000	.785
T _{ALK}	1.000	.806
Cl	1.000	.952
F	1.000	.554
SO ₄	1.000	.585
NO ₃	1.000	.804
T _{HARD}	1.000	.932

In this case of the nineteen variables, only the first five components account for meaningful amounts of variance (Kaiser criterion, 1960) which explained 81% of the total variation, so only these few components are retained. The remaining components account for only trivial amounts of variance and would not be retained, interpreted or further analyzed (Table 4.11).

Table 4.10: Total Variance Explained

Component	Initial Eigen values			Extraction Sums of Squared Loadings			Rotation
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	
1	7.153	37.645	37.645	7.153	37.645	37.645	6.880
2	2.954	15.550	53.195	2.954	15.550	53.195	4.376
3	2.151	11.319	64.514	2.151	11.319	64.514	2.607
4	1.891	9.954	74.468	1.891	9.954	74.468	2.834
5	1.257	6.614	81.082	1.257	6.614	81.082	2.044
6	.999	5.259	86.340				
7	.957	5.037	91.377				
8	.691	3.638	95.015				
9	.392	2.062	97.077				
10	.319	1.677	98.754				
11	.109	.572	99.327				
12	.086	.454	99.781				
13	.042	.219	100.000				
14	.000	.000	100.000				
15	.000	.000	100.000				
16	.000	.000	100.000				
17	.000	.000	100.000				
18	.000	.000	100.000				
19	.000	.000	100.000				

Before rotation, the first principal component extracted accounts for maximal amount of the total in observed variable, i.e. 37.65%. The second component accounts for 15.55% of a maximal amount of variance in the data set not accounted by the first component.

The remaining components, that are extracted in the analysis display the same characteristics, they become smaller and smaller, i.e. 11.32%, 9.95% and 6.6% respectively.

Each component accounts for a maximal amount of variance in the observed variables that was not accounted for by the preceding components (Table 4.11).

Table 4.11: Component Matrix. The 5 principal components before rotation.

	Component				
	1	2	3	4	5
Na	.363	-.490	.209	.720	-.131
K	.239	-.355	-.752	.030	-.384
Ca	.486	.041	.157	.759	-.035
Mg	.769	-.277	.360	.305	-.131
HCO ₃	.673	.416	-.198	.201	.101
CO ₃	.608	-.607	-.389	-.093	.231
CLR	-.535	.181	.268	-.328	-.161
Zn	.221	-.410	.714	-.165	-.090
pH	-.315	.770	.313	.156	-.355
TDS	.922	.050	-.049	-.275	-.107
Ec	.922	.053	-.045	-.279	-.105
Fe	.099	.564	-.376	.208	.319
Mn	-.307	-.405	.153	-.101	.702
T _{ALK}	.748	.288	-.069	.218	.332
Cl	.912	.012	.038	-.329	-.101
F	.550	-.214	.403	-.195	.073
SO ₄	.599	.427	-.168	-.114	-.056
NO ₃	.441	.609	.371	-.022	.317
T _{HARD}	.910	.074	.118	-.284	-.060

The unrotated component matrix (Table 4.11) which account for 37.6% of total variance show high positive loading for two cations i.e. calcium and magnesium being

0.486 and 0.769 respectively. High loading is also seen for the following anion: chloride being 0.912, bicarbonate as 0.673, carbonate as 0.608, sulphate being 0.599 and 0.530 for fluoride.

There are also high loadings for total hardness of 0.910, stressing the dominance of calcium and magnesium which contribute to the hardness of water. The loading for total alkalinity of 0.748 is contributed by carbonate and bicarbonate. Total dissolved solids correspond to electrical conductivity with the value of 0.922. Potassium, pH and trace elements like zinc, manganese and iron play insignificant role in the water quality in the area.

The second principal component which account for 15.6% of total variance show high positive loading for nitrate as 0.609, sulphate as 0.599 and bicarbonate being 0.416.

Unfortunately, when more than one component has been retained in an analysis, the interpretation of an unrotated factor is usually quite difficult. To make interpretation easier, an operation is performed called rotation. A rotation is a linear transformation that is performed on the factor solution for the purposes of making the solution easier to interpret. To obtain a clear pattern of loadings, promax rotation was employed and shown in Table 4.13.

4.7 ROTATED FACTOR ANALYSIS

The rotated component matrix clearly stresses the importance of magnesium and calcium as being dominant in the water from the wells. The hardness due to these cations is stressed by total hardness of 0.951. The electrical conductivity is also very high due to high content of dissolved substances in the water of 0.960.

In the second principal component, sodium (0.871) calcium (0.870) and magnesium (0.851) play a significant role in the water quality while potassium is insignificant. All the anions show high positive loadings in the water from the area, but bicarbonate and chloride are more dominant showing high positive loadings in the two consecutive loadings of the rotated matrix.

Table 4.12: Rotated promax structure Matrix

	Component				
	1	2	3	4	5
Na	.163	.874	-.21	.293	-.116
K	.193	.134	.137	.267	-.897
Ca	.342	.870	.234	-.007	.125
Mg	.665	.851	-.201	.280	.101
HCO ₃	.666	.478	.614	-.028	.105
CO ₃	.510	.396	-.025	.854	-.397
CLR	-.424	-.574	-.276	-.407	.171
Zn	.202	.242	-.722	.214	.311
pH	-.209	-.195	.217	-.942	.352
TDS	.959	.424	.132	.236	-.039
Ec	.960	.421	.131	.234	-.034
Fe	.112	.028	.772	-.169	.134
Mn	-.372	-.191	-.242	.570	.306
T _{ALK}	.707	.582	.540	.192	.272
Cl	.955	.403	.040	.253	.012
F	.542	.349	-.305	.320	.264
SO ₄	.662	.219	.454	-.117	.048
NO ₃	.501	.222	.371	-.209	.712
T _{HARD}	.951	.432	.060	.209	.119

The rotated factor loading from Table 4.13 were plotted as shown in Fig. 4.7 and names of variables with loadings of less than 0.3 are not plotted because their contributions to the factors are negligible (Reimann et al. 2002).

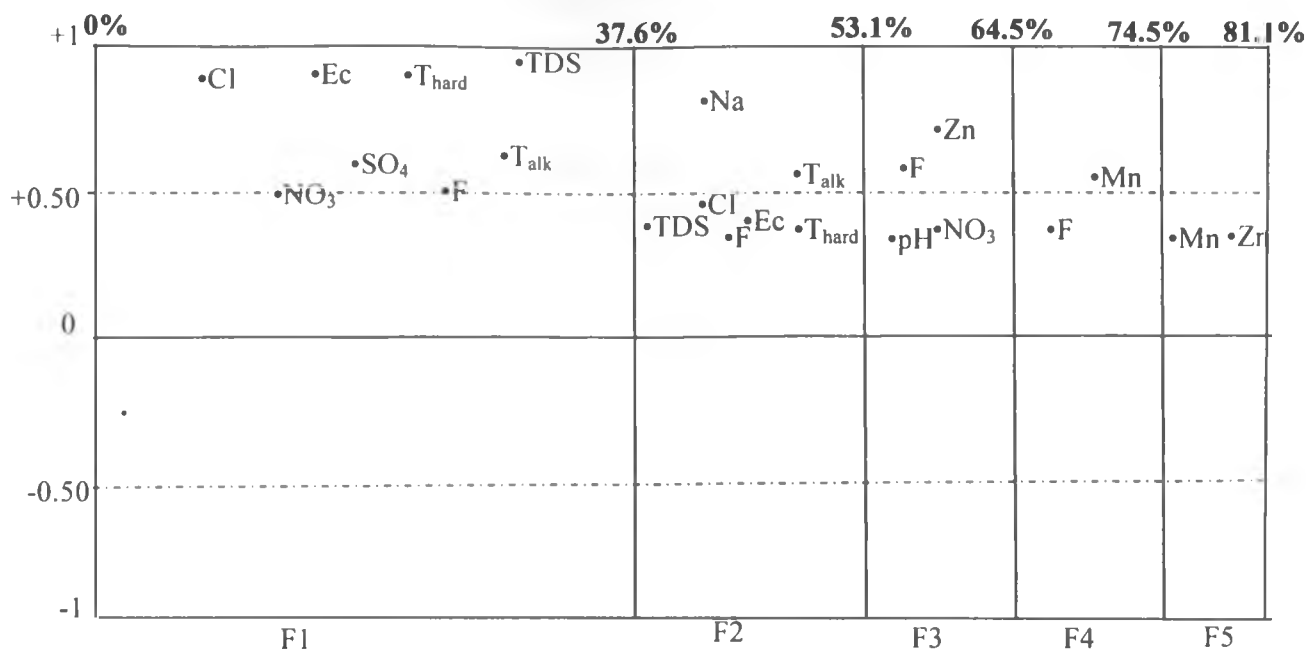


Fig 4.7: Factor loading plots for 5 factor models (PFA and promax rotation)

Component correlation matrix was constructed to display the correlation of the variables in the components (Table 4.13). There was high correlation between the first two components of 0.512 and little or trivial correlation with the remaining.

Table 4.13: Component Correlation Matrix

Component	1	2	3	4	5
1	1.000	.512	.206	.196	.068
2	.512	1.000	.086	.242	.047
3	.206	.086	1.000	-.206	.024
4	.196	.242	-.206	1.000	-.148
5	.068	.047	.024	-.148	1.000

From the plotted rotated factor loading (Fig. 4.7) and the component rotated matrix (Table 4.13) it is evident that the correlation between the first and second principal components is very high. It is clear that the cations that determine the water quality in the area are calcium and magnesium from the total hardness while the dominant anions in the water come from bicarbonate and carbonate from total alkalinity and chloride.

CHAPTER FIVE

5.0 DISCUSSION

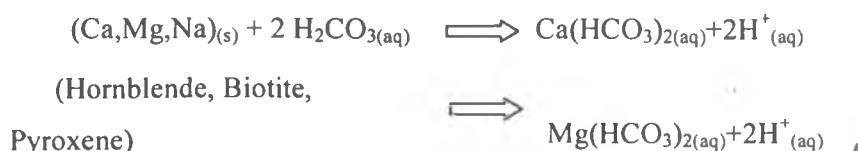
From the analysis of the water samples from boreholes in the study area using various tools as shown in the text, the following was established:

- The waters from the area are suitable for domestic use but with notable exception from a few wells like C-5452, C-5969 and C-5311 which have total dissolved substances of 1920 mg/l, 2280 mg/l and 1440 mg/l respectively and are above normal guideline and therefore unpalatable. The hardness of water in the area is due the bicarbonate, chloride and sulphate of magnesium and calcium which could be removed by boiling or chemical treatment.

- As for agricultural use, the calculated SAR values range from 0.94 to 2.36 and are within the safe levels from sodium hazard of soil accumulation and hence suitable for irrigation.

- The study has also established that the use of water for cooling purposes such as motor vehicles and posho mills should be done cautiously or the water from some wells where the mills are located could block the cooling systems due to incrustation by precipitates of carbonates of calcium and magnesium and therefore not advised.

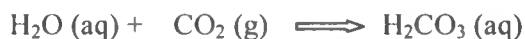
- From the principal component analysis (Table 4.9-4.13 and Fig. 4.7), the major elements that determine the water chemistry in the study area are: calcium, magnesium (total hardness), bicarbonate and carbonate (total alkalinity), sodium, sulphate and chloride which comprises 37.6% of the variables in the determination of the water chemistry and could be derived from similar rocks for the cations and similar process for the anions. The latent variable sources of the cations could be from the weathering of the minerals of calc-alkaline 'Nyanzian system' comprising andesite, rhyolite, dacite, dolerite, diorite and pyroclastics which make the geology of the area.



The principle minerals in them are plagioclase feldspars ($\text{NaAlSi}_3\text{O}_3$ to $\text{CaAl}_2\text{Si}_2\text{O}_3$), biotite ($\text{K}(\text{MgFe})\text{AlSi}_3\text{O}_{10}(\text{F.OH})_2$), pyroxene ($\text{CaMgFeAl}_2(\text{SiAl})_2\text{O}_6$) and hornblende ($(\text{Ca,Na})_{2-3}(\text{Mg,Fe,Al})_5(\text{Al,Si})_8\text{O}_{22}$) (Table 2.4).

Variables such as potassium and zinc contribute but to a lesser extent on the water chemistry. Other variables like manganese, iron and pH are rather insignificant in the water chemistry. Irons' insignificant contribution is due to its precipitation as hydrated iron (iii) hydroxide termed BIF.

Extensive banded iron formation (BIF) which is a sedimentary rock is from the weathering of biotite, pyroxene and hornblende and contributed to iron in the water chemistry in the area but the bulk of it being chemically precipitated as red-yellow iron hydroxides.



Biotite, Pyroxene)



Low presence of nitrate in the sampled water indicates that pollution from the influence of anthropogenic use of nitrogenous fertilizers for farming is non-existent or has not affected the aquifers.

From the information exhibited by the well logs and plotted on the hydrogeological profiles (Table 4.2 and Fig. 4.1), the rise in the levels between the water struck level and the rest level it is evident that the main aquifer in the study area is under pressure (piezometric), an indication of semi-confined aquifer and follows the topographical shape but is slightly subdued.

The groundwater flow in the study area is influenced both by the local and regional geology and trends towards NE-SW as can be seen by the flow direction shown in the hydrogeological profiles in (Fig. 4.2 (a) and (c)) and illustrated in Fig. 4.1. This is because the groundwater at discharge zones tends to have higher concentration levels compared to recharge zones (Walton, 1970; Todd; 1980).

The study has also established that the area consists of hydrochemical facies that are small and discrete due to varied concentrations of different anions and cations as evident from the percentage milliequivalent (Table 4.6) and plotted on Piper diagram (Fig. 4.3).

Based on the analysis and interpretation of the results (Table 4.7 Fig. 4.6) it has been confirmed that the study area consists of two major hydrochemical facies.i.e sodium bicarbonate/sulphate rich and calcium bicarbonate/chloride rich facies.

Referring to the CBE analysis (Table 4.4) it evident that in eleven out of fourteen wells there is high levels of dissolved ions, above $\pm 10\%$ possibly calcium and magnesium which dominate the waters of the study area and cause hardness therefore, their removal from the water is very important if the waters have to be used especially for laundry. The hardness which is both of temporary and permanent type could be removed by boiling and treatment with chemicals in a water treatment plant.

The aspect of pollution could not be overlooked especially when discussing water quality for public use. Low presence of nitrate in the sampled water indicates that pollution from the influence of anthropogenic use of nitrogenous fertilizers for farming is non-existent or has not affected the aquifers.

CHAPTER SIX

6.0 CONCLUSION AND RECOMMENDATIONS

6.1 CONCLUSION

The groundwater resources in Samia area of Busia County are generally neutral with pH values between 6.5 and 7.5. The bedrock geology and local geological structures such as faults and fissures play an important groundwater zones with good water for tapping. The study has also established that the area is underlain by at least two major groundwater fields;

- The Calcium rich fields and
- The Sodium-Potassium fields.

The major anions in these fields are the carbonate and bicarbonate type though considerable chloride and sulphate occur and should not be ignored because they contribute to the permanent type of hardness which is difficult to remove. Trace elements like zinc, manganese and iron occur in insignificant levels. Fluoride concentration in the area is quite low but chloride levels reach unacceptable levels in some wells.

Although the waters are suitable for most purposes, be it domestic, agricultural or industrial the question of their hardness especially for industrial, laundry and cooling purposes need to be addressed for their concentration levels in some wells is very high. The treatment of the water in the area can be achieved by addition of sodium or potassium carbonate which removes all types of hardness.

Drillers' Well logs (Table 4.1 and Fig. 4.1 and 4.2) indicate that the ground water profile is closely related to the topography of the area but slightly subdued, and is controlled by regional geology of the East-West trend.

The study established that the question of pollution in the area from both natural and anthropogenic use of nitrogenous fertilizers has not been observed in the groundwater systems.

6.2 RECOMMENDATIONS

- Large scale drilling of boreholes for water supply projects should be encouraged in the northern part of Samia which is also the water tower of the area because the water is soft.
- Treatment of water from some wells should be done to reduce the hardness before it is put to use.
- Most wells in the area are shallow and need to be extended to greater depth if geophysical exploration would recommend so.
- Information on wells dug in the area and the rest of the country should be declassified to enable access to the data for public scrutiny as well as research purposes.
- Drillers log records should include different geological formations penetrated and their extent to enable meaningful modeling and interpretation of geochemical actions in the aquifer instead of having water struck level, rest level, total depth and discharge which is scanty.
- Manpower should be trained to start gauging stations which are valuable hydrological resources data and to rehabilitate weather stations since the only one in the area is no longer operational.
- More geophysical exploration needs to be carried out in the study area in order to search for deeper aquifers that could hold fresh water.

REFERENCES

- Alyssa A. D. and James L. O. (2006): Carbon-14 dating of groundwater in the Palouse Basin of the Columbia river basalts a Hydrology Program, Department of Geological Sciences, University of Idaho Moscow, ID 83844-3022, USA.
- Arturo W. (2008): "Groundwater hydrochemical facies, flowpaths and recharge determined by multivariate statistical, isotopic and chloride mass balance methods ETD collection for University of Texas, El Paso. Paper AA13320185.
- Bilqis A. H, Sombo Y, Akira S, Sufia K, Mahbooba K, Yamen H, Sanowar H, Shoriful I. and Obaid H. (2006): Arsenic Mitigation for Water Supply in Bangladesh: Appropriate Technological and Policy Perspectives Water Qual. Res. J. Canada, Volume 41, No. 2, 226-234.
- British Geological Survey (2001): Natural Environmental Research Council Technical Report WD/OO/35.
- Casey K. and David P.G. (2007): Groundwater age and the importance of chemical fluxes across aquifer boundaries in confined cretaceous aquifer of North Carolina, USA, Department of Marine, Earth & Atmospheric Sciences, NCSU, Raleigh, NC, in Radiocarbon, Vol. 49, Nr 3, 2007, p. 1181-1203.
- Elliot G. F. S. and Gregory J.W. (1895): "The Geology of Mount Ruwenzori and some Adjacent Regions of Equatorial Africa" Quart. Journ. Geol. Soc. London, Vol. L1.
- Fetter C.W (2001): Applied Hydrogeology. 4th edition. Prentice Hall, New Jersey. ISBN 10: 0130882399.
- Freeze R. A. and Cherry J. A. (1979): Groundwater, Prentice Hill INC, Englewood Cliffs, New Jersey, USA, 604p.
- Gregory J. J. W. (1921): The Rift Valley and Geology of East Africa. Seeley, Service and Co., London. 128.
- Guler C. and Thyne G.D. (2004): Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy c-means clustering, Water Resources Research Vol. 40, W125003, doi, 1029/2004, WR003299, 2004.
- Horowitz J, Arthur, A primer on sediment trace element chemistry. US Geological Survey Handbook. CRC Press; 2nd edn. ISBN-13: 978-0873714990.
- Ichang'i D.W. and Maclean W. H. (1991): The Archean volcanic facies in the Migori segment, Nyanza greenstone belt, Kenya; stratigraphy, geochemistry and mineralisation. J. Afr. Earth Sci. 13: 277-290.
- Ichang'i D.W. (1992): A lithostratigraphic framework for the Migori segment of the Archean Nyanza greenstone belt, Tanzanian craton, East Africa. Conference paper Accra: Geol. Soc. of Africa.
- Jaetzold R. and H. Schmidt (1982): Farm management handbook of Kenya; Vol. II/A: West Kenya. Nairobi; Ministry of Agriculture/GTZ.
- Jolliffe T (2002): Principal component analysis. 2nd edn. Springer series in statistics. Aberdeen, UK.
- Kaiser H. F. (1960): The varimax criterion for analytic rotation in factor analysis. Vol. 23, No 3 (1958), 187-200, DOI: 10.1007/BF02289233.
- Kefinco (Kenya-Finland: Western Water Supply Programme; 1981-1995): Water Supply Development Plan (1990); Ministry of Water Resources, Kenya and Ministry of Foreign Affairs, Finland. Kakamega, Kenya. Drawing No. 5-2 Govt. Printer.
- Kenya National Bureau of Statistics (2010): 2009 Population and Housing census results. Vol 1A.
- Kenneth R. B. (1991): Tritium as an Indicator of Ground-Water Age in Central Wisconsin. Groundwater, Vol 29 No 3:398-404.
- Kimomo S.M, Ndigirigi H, M and Njonjo H, W (2007): Geology of the Busia-Siaya Mineral Resources. Mines and Geology Department. Report No. 113.

- Kuria Z. and Prof. Barongo J. (2006): Hydrogeology, unpublished work.
- Lindsten D. C.(1984): "Technology transfer: Water purification, U.S. Army to the civilian community". The Journal of Technology Transfer 9 Vol.1: 57–59.
- Llyod J. W (1985): Natural Inorganic Hydrogeochemistry. Oxford University Press, USA. 294p.
- Marleen C Kilonzo F and Walraevens K (2008): Hydrochemistry and sources of high fluoride in groundwater of Nairobi area, Kenya. Hydrological sciences- journal 56 (6).1230-1240
- Mathess G (1982): The Properties of Groundwater. John-Wiley and Sons, New York. DOI:10.1016/0016-7037(82)90218-6.
- Mohammad R. and El-Nakhal H.A. (2004): Hydrogeochemistry and carbonate saturation model of groundwater, Khanyounis Governorate—Gaza Strip, Palestine Hydrological Sciences journal 49(3) Vol. 47, Issue 7, pp 898-906.
- Murray K. and Wade P. (1996): Checking anion-cation charge balance of water quality analyses. ISSN 0378-4738- Water SA Vol. 22 No 1, pp 27-32.
- North America Commission on Stratigraphic Nomenclature (1983): North American Stratigraphic Code. A.A.P.G Bull. 67: 841-875.
- Odman O. H. (1929) "Preliminary Report on the Archaean Geology of Western Nzoia Province, Kenya Colony," Geol. Foren., Stockholm. Libris Id: 3204706.
- Ogege J.I, Odada E.O. and Olago D. O (2002): Geological Controls on Groundwater Geochemistry in Butula area, Busia District, Kenya. African Journal of Science and Technology (AJST) Science and Engineering Series vol.3(2):24-33.
- Pulfrey W. (1936 a): The Geology of an Area in the Kavirondo District, Kenya Colony. Geol. Mag 58. pp. 129–336,
- Pulfrey W (1936 b): Preliminary report on the geology of the No. 1 area, north and central Kavirondo. Rep. Min. Geol. Dep. Kenya, Vol 5: 31.
- Reimann C, Filzmoner P, Garrett R. and Dutter R. (2002): Statistical data analysis explained; Applied environmental statistics. R. John Wiley and Sons Ltd, England.
- Schroeiber M. E. Moline G.R and Bahr J.M. (1999): Ground Water monitoring and Remediation. Journal article, vol.19: 95-109.
- Stockley, G. M. (1943): The Pre-Karoo stratigraphy of Tanganyika. Geol. Mag 80; 161–170.
- Suwa K. (1981): Petrochemical and petrographical notes on some Nyanzian volcanic rocks, West Kenya .In: Sixth Prelim. Rep. Afr. Studies Nagoya Univ. (Earth Sciences 4) (Suwa, K. Ed). 15-32. Department of Earth Sciences, Nagoya Univ. Japan.
- Todd D.K. (1980): Groundwater Hydrology, 2nd edn. John-Wiley and Sons Inc.London. 539p.
- UNEP (2002): Vital Water Graphics: An overview of state of the worlds Fresh and Marine waters, [CD].Nairobi GRID Arendal.
- United States Environmental Protection Agency (EPA 1986): Quality Criteria for Water, EPA 440/5-86-001.
- Water Well Industry Handbook (1975): Johnson division, UOP Inc.
- Walton W. C. (1970): Groundwater Resource Evaluation. McGrawHill London. pp. 664p
- WHO (1984): Guidelines for drinking water quality, Health Criteria and other Supporting information, VII. World Health Organization Geneva, Switzerland.
- Yanagi, T. and Suwa, K. (1981): Rb/Sr radiometric dating on the Precambrian Rocks, in the western part of Kenya in; Sixth Prelim. Rep. Afr, Studies, Nagoya Univ. (Earth science 4).



UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

Telecommunication Traffic Forecasting Model

By

Maina, Charles Mwangi

P58/72894/2009

Supervisor: Mr. Andrew Mwaura

October 2011

A Research Project submitted in partial fulfilment of the requirements for the degree of Master
of Science in Computer Science

University of NAIROBI Library



Declaration

I certify that all work presented in this report is my original work and has not been submitted for any other academic award at any other university. The work presented was carried out under the supervision of Mr. Andrew Mwaura of the University of Nairobi, School of Computing and Informatics. All the work in this Research Project is my own except where acknowledged in the text.

Signature  Date..... 31/10/2011

Charles M. Maina

This Research Project has been submitted in partial fulfilment of the requirements for the award of a Master of Science Degree in Computer Science of the University of Nairobi with my approval as the University Supervisor.

Signature.....  Date..... 20/Nov/2011

Andrew Mwaura

Acknowledgement

Apart from the hard work and efforts I have put in, the successful completion of this project depended highly on the support, guidance, and motivation from a number of people. I would like to take this opportunity to extend my gratitude to all those people who have been a source of great support to me in the successful completion of this research project.

I acknowledge, with deep gratitude and appreciation, the tremendous support and assistance received from my supervisor, Mr. Andrew Mwaura in this research project. I am very grateful for his efforts, his tireless guidance and invaluable ideas throughout this research. In particular his commitment to share helpful discussions and the many reference materials he provided remain a big source of encouragement to me. He made me realize that everything is possible through sheer hardwork and determination.

Also much appreciated are the Supervisors of the Panel Dr. Waiganjo for his guidance, encouragement and advice and especially during all the presentations and the motivation to continue on. His reference materials were an eye-opener. The materials gave me in-depth insights that assisted me to catapult on my project. To Mr. Moturi for his encouragement and pointing out on areas of improvement on the project and the challenge to continue on. To Mr. Ogutu for his advice and very encouraging comments towards the realization of my research project.

I am also grateful to Mr. Kimani who is an Engineer at Telkom Kenya for introducing me to the Switching and Transmission concepts. Our discussion assisted me in the problem interpretation.

Last but not least I am grateful to Telkom Kenya staff at the Switching & Transmission Department for their positive response when I required any assistance during data collection.

Finally, I am most humbled by God's grace that has been my strength all through.

Dedication

My greatest thanks and appreciation go to my family for their permanent love and support. I realize how lucky I am to have them. I dedicate this research project to my wife Terry and to my parents Jidraph Maina and Mary Maina. Appreciation to my wife Terry for her encouragement and constant support throughout my research. She always gave me all the time I needed to complete my studies without any hassle. My parents provided a big source of warmth and encouragement.

Abstract

A problem of limited resources at the International Switch existed and this led to congestion. Congestion occurred when the call load exceeded the capacity of available routes or trunks, or of common control equipment for example, call registers that temporarily store call data while the call is being processed. Congestion also occurred when the call load could not be handled in available real time. There were some identified calls drop rates of about 15% due to limited Bandwidth resources at the Switch and this posed challenges of lost revenue.

Telecommunications system designers have studied the problem of system resource congestion for years. These same problems were evident at the switch where International Traffic got transited from all local operators to International operators.

We discuss the results of the study done on forecasting the Telecommunication Traffic from Kenya local operators to the International destinations through a common gateway, the International Switch (IGW). The idea was to determine the volumes of traffic each operator transit through the International Switch and to pick the operator giving maximum revenue to the company and to optimize the resources available at the International Switch based on the outcome of the prediction scenarios.

Success on data collection eventually led to the formulation of data cycles each cycle containing 7-days. The derived seasonal indices of the days in the cycle were found to be considerably applicable in the computations of the trend and seasonal forecasts. Forecasting was determined by developing a Linear Regression model and then fitting an observed dataset of four months (Jan-April) into the model with two independent variables namely duration and period being used as the key elements. Time series were used due to their intelligent nature of learning from additional data. An Artificial neural network forecasting model was also developed. The system could forecast the operator's traffic volumes based on the past historical trends and provide decision-making for the Carrier Connect team to implement.

The research reveals that in making predictions regarding Telecommunication Traffic, a Prediction Model can be developed and be implemented whose objective is to predict traffic volumes. The results also reveal that the Linear Regression model shows better optimal results than the Neural Network system in the Forecasting of Telecommunication traffic.

List of Terminology and Abbreviations

IGW- International Gateway

ANN- Artificial Neural Network

NGN- Next Generation Network

PSTN- Public Switched Telephone Network

ISDN- Integrated Services Digital Network

**EURESCOM - European Institute for Research and Strategic Studies in
Telecommunications**

MA- The Moving Averages

ASI- Adjusted Seasonal Index

XML- Extensible Mark-up Language

IDE- Integrated Development Environment

Table of Contents

CHAPTER ONE	3
1: INTRODUCTION	3
1.1. Problem statement	5
1.2. Research Objectives.....	5
1.3. Project Justification	5
1.4. Significance of the research to Telecommunication Industries.....	6
1.5. Project Scope	7
1.6. Assumptions and Limitations of the Research.....	7
CHAPTER TWO	8
2: LITERATURE REVIEW	8
2.1. Telecommunication Switches.....	8
2.1.1. NGN reference model.....	10
2.1.2. Packet network.....	10
2.1.3. Call Server	10
2.1.4. Call router	10
2.1.5. Media Gateways	11
2.1.6. Media Server	11
2.1.7. Application Server	11
2.1.8. Application creation environment.....	12
2.2. Prediction Methods and techniques	12
2.2.1. Decision trees for Prediction	12
2.2.2. CART	13
2.2.3. Linear regression	14
2.2.3.1 Regression Models.....	15
2.2.4. Clustering and Nearest Neighbor	16
2.2.5. Delphi Survey Method	16
2.2.6. Scenario Method	17
2.2.7. Statistical demand analysis	17
2.2.8. Technology Foresight Method	17
2.2.9. Assessment method	18
2.2.10. Rule induction	18
2.2.11. Neural Networks	18
2.2.11.1. Kohonen Networks.....	20
2.2.11.2. Feed Forward Artificial Neural Networks	22
2.2.11.3. The Artificial Neural Networks	22
2.2.11.4. The Artificial Neuron	23
2.2.12. Training Artificial Neural Networks	24
2.2.13. Supervised Learning	24
2.2.14. Back Propagation	25
2.2.15. Applications of Neural Networks	25
2.2.16. Sigmoid Function.....	26

CHAPTER THREE 27

3: METHODOLOGY 27

3.1. Data Collection 27

3.1.1. Steps of extracting the data..... 27

3.2. Using Linear Regression 28

3.2.1. Trend Forecast 28

3.2.2. Seasonal Forecast 28

3.2.3. Observations on data..... 31

3.2.4. Elements of the research..... 32

3.2.5. Data Cycles 32

3.2.6. Traffic Volumes 32

3.2.7. Modeling the data 34

3.2.8. Fitting the data into a regression model 34

3.2.9. Regression Model 36

3.2.10. Examples of other Linear Regression Equations 36

3.3. Assumptions of the linear model 38

3.4. Goodness of fit of the Regression Model 39

3.5. Time series 39

3.5.1. Outliers 39

3.5.2. Predicting..... 40

3.5.3. Modeling Seasonality and trend..... 40

3.5.4. Seasonal Index..... 40

3.5.5. Seasonal Adjustment..... 41

3.5.6. Forecasting 41

3.5.7. Advantages of Time Series 41

3.5.8. Trend Line..... 42

3.6. System Analysis 42

3.6.1. Approach 43

3.7. System Design 43

3.7.1. Design of the Prototype Web-based System..... 46

3.7.1.1. Traffic Forecasting 46

3.7.1.2. Bandwidth Optimization 47

3.8. System Development 47

3.8.1. Telecommunication Forecasting System..... 46

3.8.2. Telecommunication Forecasting Model..... 49

3.8.3. Telecommunication Bandwidth Optimization..... 50

3.8.4. System Testing..... 52

3.9. Using Neural Networks..... 52

3.9.1. Neural Network Traffic Forecasting Model 53

3.9.2. Types of the Layers..... 53

3.9.3. The Neural Network/Brain 53

3.9.4. Neural Network Program Structure 53

3.9.5. The Organization of the Classes 54

3.9.6. Inputs..... 57

3.9.7. Neuron Training..... 57

3.9.8. Prediction/Thinking	57
3.9.9. Results from Neural Networks	58
3.9.10. Storing the Brain State	62
3.9.11. Advantages of using Neural Networks	62
3.9.12. Weaknesses of Neural Network	63
3.9.13. Applications of Neural Networks	63
3.9.14. NN System Testing	64
CHAPTER FOUR.....	65
4: Results and Discussions	64
4.1. Results and Discussions.....	65
4.2. Comparison between Actual Traffic versas Predicted	65
4.3. Comparison of two cycles, one for 7 days and another one for 14 days	69
4.3.1. Observations	71
4.3.2. Noisy Data	73
4.4. Challenges	74
CHAPTER FIVE.....	75
5: Conclusions and Recommendations	75
5.1. Conclusions	75
5.2. Recommendations	76
5.3. Future Research	76
GLOSSARY	77
REFERENCES	78
APPENDICES.....	80
6.1. APPENDIX A	80
6.1.1. Prediction Algorithms Code	80
6.2 APPENDIX B	81
6.2.1. Bandwidth Optimization Code	81
6.3 APPENDIX C.....	83
6.3.1. Neural Network Code.....	83
6.4. APPENDIX D	84
6.4.1. Predicting the Neural Network	84

CHAPTER ONE

1: Introduction

Voice traffic has increased tremendously due to the rise in the number of operators. Increase in voice traffic affects several other aspects of the network such as reduced bandwidth capacity. Network management functions within telephone switching systems are designed to maintain a high level of service during periods of unusual traffic on systems that are engineered for less than full capacity. During these periods, machine performance naturally degrades due to resource contention and network delays. Even when normal traffic levels are restored the system continues to struggle to achieve normal performance levels.

Interconnect billing deals with the business of transiting the inbound and outbound calls through the international switch so that all carriers of the calls can be compensated for the portion they carry. The interconnect portion is from switch to switch or Originating Point to Interconnection (POI) to Terminating Point to Interconnection (POI). This portion is usually handled through high capacity trunks or circuits, by long distance carriers that aggregate the traffic at the switch and send it out via trunks over the appropriate routing for connection to the terminating user. The routing of the calls can vary depending on the cost of bandwidth or capacity available. As the call goes from country to country or carrier to carrier, the compensation becomes more complex. Sometimes, thousands of rates must be stored for calculation of the compensation to all providers of the call service.

A switch is a device that channels incoming data from any of multiple input ports to the specific output port that will take the data toward its intended destination. In the traditional circuit-switched telephone network, one or more switches are used to set up a dedicated though temporary connection or circuit for an exchange between two or more parties. telephone switching systems, where phone calls are "switched" meaning they cross through a switching matrix to route calls from an origination point to a destination point.

Carrier connect and traffic managers are inconvenienced by the fact that they are not able to balance bandwidth at the International switch level and for this reason an urgent measure was implemented that was geared towards helping in predicting the

traffic capacity based on the generated traffic volumes for each operator. There was need to define how to balance the load based on the operators generating maximum returns to the company. A priority for capacity upgrade at the switch was also determined.

One of the characteristics of Telecomms traffic is the superimposing of many cyclical effects. For instance, there are hourly trends corresponding to the business day, daily trends according to which day of the week (some working days are typically busier than others and weekends have very little traffic) trends according to the day of the month (end of month can be busier) and seasonal trends. Each of these trends are cyclical with differing periodicities.

Neural Networks have been widely used as time series forecasters, most often these are feed-forward networks which employ a sliding window over the input sequence. Typical examples of this approach are market predictions, meteorological and network traffic forecasting. Traffic growth rates have been tracked for many years and the patterns have varied significantly dependent upon the day in the week, holiday seasons and change in tariffs.

The accurate traffic model and prediction of the traffic plays an important role in resource allocation. Traffic is dynamic and usually affected by immeasurable parameters and variables. In general, the traffic stochastic parameters vary slowly. Hence, they can be estimated using historical data. The traffic model built on historical data enables forecasters to predict the future traffic pattern (Li & Zekavat, 2008).

Neural network based traffic prediction approach is complicated to implement. The accuracy and applicability of the neural network approach in traffic prediction is limited (Hall & Mars, 2000). However, Artificial Neural Network (ANN) can capture the non-linear nature of traffic and the relationship between the output and input (Khotanzad & Sadek, 2003).

The contribution of this study to the Telecommunication Industries was to offer a solution to the Switching and Transmission teams in Telecommunication industries to make better decisions on how the future traffic was likely to be and allocate resources at the switch in a balanced percentage proportions that will maximize returns.

1.1. Problem statement

“A problem might be defined as the issue that exists in the literature, theory, or practice that leads to a need for the study” (Creswell, 1994, p. 50). This project aims at examining the traffic volumes that is received at the main International Switch for transit and the resources that telecommunication providers use for transiting traffic. Telecommunication providers face a big challenge of some traffic routes being highly congested resulting to large number of transit calls getting dropped at the International switch (IGW). Problems of capacity and load balancing exist whereby upgrading of the bandwidth at the switch is done based on the units of data a particular operator is transiting through the switch. At certain intervals if all operators were engaged and the bandwidth utilization at the switch is 100% for all of them then it means the routes carrying the traffic gets congested and calls from operators will be dropped at the switch. Due to lack of good models to predict traffic growth the calls drop rate had increased drastically.

1.2. Research Objectives

The purpose of this project was: -

1. To obtain and analyze operators’ generated traffic data and resource allocation from a telephone network interconnection host.
2. To apply prediction of resource usage using both neural networks and linear regression with the aim of forecasting, capacity upgrade and optimal resource allocation for increased revenue.
3. To propose an optimal load balancing technique based on the results obtained from the prediction results.
4. To develop a web-based prototype system that does the prediction, bandwidth optimization and resource allocation based on the prediction scenarios obtained.

1.3. Project Justification

Historically, telecommunications networks were deployed to carry voice traffic i.e. their architectures were optimized for handling the stochastic nature of PSTN traffic on a circuit switched basis. The cost pressure of competition led networks to evolve towards digital transmission and switching, thus making the networks more universal. In the last decade all carriers in the developed world had consequently seen a surge in data traffic conveyed by means of packet switched technology.

Based on the volumes transported it was foreseeable that data traffic would become prevalent. In Telecommunication Companies, huge infrastructure had to be designed, network flow and place element network in right node so it would be effective to serve subscribers traffic and also due to broadband services. The term broadband refers to a telecommunications signal of greater bandwidth. The broader the band, the greater the capacity for traffic.

Traffic forecast was needed for deciding network dimensioning and it was therefore useful to estimate traffic growth and traffic offered in a network. Besides that, we needed to calculate traffic characteristic generated by mobile subscribers, how the traffic distribute, amount in busy hour, and mobile subscribers' behaviour in some periods of the year.

And finally we needed to estimate about subscribers' growth in the future. The result is to get traffic tendency in long time-span, so as to get effective decision for network investment.

Today telecommunication companies are getting into interconnection billing with operators across the world. Once an agreement has been reached the host companies transiting the traffic can open up some channels and allocate bandwidth for traffic to flow. In this case the challenges accompanying interconnection traffic routing is that at times the routes gets congested and overflow routes are not immediately available to cater for excess traffic. In the view of this researchers needed to come up with a method of prediction that would enable for optimum allocation of resources at the switch without inconveniencing operators or dropping calls at the switch and at a much more revenue to the company routing the traffic.

1.4. Significance of the research to Telecommunication Industries

- Telecommunication Industries will be in a position to utilize prediction for optimum allocation of resources at the switch and especially bandwidth to avoid dropping of calls due to congestion of the routes.
- It will help Telecommunication companies in business planning and demand forecasting.

1.5. Project Scope

This research Project concentrated more on the Outgoing Route Traffic i.e. Traffic from Local Operators to International Operators. The local operators being Safaricom, Airtel and Econet.

1.6. Assumptions and Limitations of the Research

1. The research assumes that all blocked calls are immediately cleared at the switch
2. Assumption that all blocked calls stay in the system until they can be handled. If calls cannot be immediately answered, they enter a queue.
3. There is no single time when the switch will be idle so for monitoring you must have a timeframe
4. The assumption that traffic to be used will be outgoing traffic during busy hours.

CHAPTER TWO

2: Literature Review

Telecommunication services are indispensable because they are the driving forces of economic growth and integration. Intense competition in the telecommunication market has emerged over the last decades and is expected to grow continuously and rapidly. Therefore it is essential for those network companies involved in the telecommunication market i.e. network carriers or service providers to gain advantage of their core competencies by adopting appropriate strategies or positioning themselves properly in the market. In particular telecommunication companies are forced to search for new solutions to reduce their operational costs, in order to eliminate the unnecessary expenditure on the existing infrastructure which often consists of multiple parallel and even overlapping networks, and to provide differentiated services with superior quality (Crimi, 2005). The expected “network convergence” may offer a solution to the problem, as it will involve providing integrated services on a common network infrastructure. Network carriers have attempted to address network convergence by increasing the number of services and/or applications on existing networks. Examples include network carriers and service providers making a great effort in integrating residential broadband services and voice services into the existing copper phone line network by adopting advanced network technology.

At the same time, the migration from traditional voice traffic to IP traffic in our networks has already become a prevailing trend in telecommunication network development (Stiller, 2000).

2.1. Telecommunication Switches

Softswitch

According to the International Softswitch Consortium (ISC) a softswitch can be defined as follows: A softswitch is, "an all-encompassing term for next generation communications systems that employ open standards to create integrated networks with a decoupled service intelligence capable of carrying voice data traffic more efficiently and with far greater value-added service potential than is possible with existing circuit

switched networks. The migration from circuit switching to packet/frame/cell-switched networks controlled by softswitches will transition the industry from a closed environment to open standards-based systems, achieving the openness that started years ago with Intelligent Networking (IN) initiatives" (ISC, 2002, p. 3). The concept of the softswitch was initiated by the Bell Laboratory in 1997.

A softswitch based architecture separates the service and call control layer, which is integrated in a PSTN world and allocates these functions to different platforms. Gateways play a prominent role within a softswitch-based network. KOLESAR & LEVIN (2002, p.11) and CAWLEY (2001) argue that in a packet based architecture the only candidate for a bottleneck is the final link to the customer's terminal device, i.e. the link between the last router and the end-user.

"The Next Generation Network (NGN) is a packet-based network that is able to provide telecommunication services to users and able to make use of multiple broadband, and in which service-related functions are independent of the underlying transport-related technologies. It enables unfettered access for users to networks and to competing service providers and services of their choice. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users." (ITU-T, 2004). NGN is an IP based platform capable of handling voice data traffic simultaneously and in which the transport layer and the control layer are separated. Moreover, in contrast to the centralized architecture of the circuit switched PSTN/ISDN networks where the switch with all its hardware, call control and services software is located in the geographical market it serves, in the NGN architecture the functional elements can be fully distributed within the network. The NGN represents a major research direction in telecommunications related research, and has drawn a great deal of industry-wide attention since it was first introduced.

NGN research is also aligned with the well publicized research and development of IPv6 as part of the NGN infrastructure (Iekel-Johnson et al., 2008). Over the last decade the explosion of data traffic in telecommunication networks has been impressive. The shift from simple voice communication to rich content interactions (video and image) over the Internet, even in terms of the simple voice communication, voice carried by packets over mobile networks and the Internet has seen a dramatic increase while the voice traffic over conventional PSTN has dropped in recent years (Lee & Knight, 2005).

Network carriers and service providers are aware of those market changes and therefore turn to constantly deploying broadband services within the existing network infrastructure; as a result many parallel and service-specific networks are established in order to meet those increasing demands (Olszewski et al., 2003).

2.1.1. NGN reference model

Research carried out by the European Institute for Research and Strategic Studies in Telecommunications (EURESCOM, 2001a,b) analyze NGNs on a highly abstract level, which allows a clear characterization of the essential elements of future networks and their functions. The main building blocks of a NGN switch is: -

2.1.2. Packet network

The packet network contains of mainly network elements for data transport and the routing of data packets. They perform the basic functions of packet networks, namely the transport of data packets from point A to point B.

2.1.3. Call Server

The key element of a NGN apart from the transport function is the call server, which is responsible for signaling and controlling resources in the network. The control functions can be specified as follows: call control, media gateway control and service control. The call server is the essential element within a NGN needed to establish a telephone call.

It manages and controls the call set-up using signalling protocols (call control). Moreover, the call server communicates with the media gateways to ensure the physical call set-up (media gateway control) and it controls functions being stored on media, message or application servers (service control). Call control is a very complex task and encompasses, according to EURESCOM, the following functions: call routing, call signalling, third party call set-up, third party call control and CLASS services.

2.1.4. Call router

Call routing means that signalling information has to be conveyed across the network. To this end addresses and routing tables are used containing information to handle the steering of a connection. "Call signalling" enables the exchange of signalling information to set up a connection within the network on the basis of signalling protocols.

In particular, the call server usually supports a SS7 signalling gateway to enable the exchange of signalling information between PSTN and packet network. The signalling gateway function can either be an integral part of the call server or is located on a dedicated signalling gateway platform (EURESCOM, 2001, p. 38). Call server can provide functions such as "third party call control" and "third party call set-up". These functions enable the provision of a multitude of extended call services on the network such as call center services (SINNREICH & JOHNSTON, 2001, p. 73 and p. 141).

2.1.5. Media Gateways

Media gateways are physically located at the interface between different networks. This can mean the interface between the PSTN/ISDN and the packet network or the interface between packet networks, which are supported by different protocols. Media gateways are important because of the presence of access networks, which are not based on packet-switched technology. Media gateways, thus, convert the media information flows of one network into those of another based on the specific requirements of the latter network. If, for example, a PSTN access line is connected to a NGN the media gateway provides for the conversion of frequencies into data packets (transcoding).

2.1.6. Media Server

Media servers provide functions enabling interaction between a caller and an application through telephony devices. Examples are answering a call, playing announcements, or reading an e-mail by using speech synthesis.

Media servers consequently consist of software and hardware responsible for functions like speech recognition, speech synthesis, audio generation as well as conversion of speech to text and vice versa.

2.1.7. Application Server

Application servers are primarily capable of introducing, executing, controlling and managing applications. EURESCOM specifies several functionalities an application server should be able to provide, including Control of network elements performing authentication, authorization and accounting capabilities for service provision, Support for registration, Security services (encryption, authentication and authorization to ensure access to services, Support for developing services, communication with other internal or

external applications and external network resource control servers like media server and message server, data base support to store, subscriber and service data. Broadly speaking, application servers contain the intelligence for the provision, execution and management of applications.

2.1.8. Application creation environment

The application creation environment support the life cycle of an application, which is composed of a series of phases. EURESCOM (2001, p. 48) in this context specifies the needs analysis and conception, application creation, acceptance testing, application deployment, application provisioning and operation, application removal. The application creation environment should provide the tools for addressing the identified activities.

The components enabling the creation of applications are to provide an integrated environment to develop and validate applications to run on an application server. It is important to note that the structure of NGNs is based on a distributed functional structure i.e. different functions are provided by distinct network elements geographically distributed at different locations.

Within NGNs the transport function is separated from the control and signalling functions. In particular these tasks are performed in separate network elements. NGNs use a common platform for the provision of voice and data services. The different services and applications are based on a common "bearer protocol". The bearer protocol can terminate directly on a user device such as a PC, server or telephone and, thus, defines a "terminal point".

Access to this terminal point therefore takes place directly at the user premise or with a service provider (DENTON, 1999, p. 16). To enable the provision of services and applications within a NGN there are a multitude of protocols. Voice services are supported for example by SIP and H.323.

2.2 Prediction Methods and techniques

2.2.1. Decision trees for Prediction

A decision tree is a predictive model that can be viewed as a tree. Specifically each branch of the tree is a classification question and the leaves of the tree are partitions of the dataset with their classification.

Although some forms of decision trees were initially developed as exploratory tools to refine and preprocess data for more standard statistical techniques like logistic regression. They have also been used and more increasingly often being used for prediction. This is interesting because many statisticians will still use decision trees for exploratory analysis effectively building a predictive model as a by-product but then ignore the predictive model in favor of techniques that they are most comfortable with. Sometimes veteran analysts will do this even excluding the predictive model when it is superior to that produced by other techniques. With a host of new products and skilled users now appearing this tendency to use decision trees only for exploration now seems to be changing. One disadvantage of using is that too much information can be an issue. Managers face too much information when making a decision. Rather than making a decision and advancing their company's mission or vision, managers spend more time looking at decision trees. Decision trees can require more analysis than other analysis methods and slow down the decision-making process.

2.2.2. CART

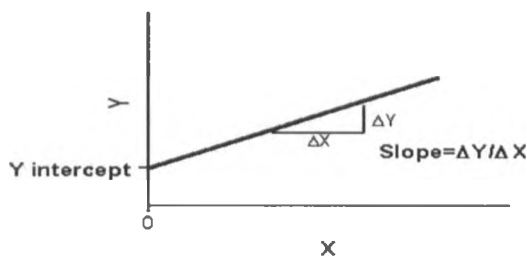
CART stands for Classification and Regression Trees and is a data exploration and prediction algorithm developed by Leo Breiman, Jerome Friedman, Richard Olshen and Charles Stone and is nicely detailed in their 1984 book "Classification and Regression Trees" ([Breiman, Friedman, Olshen and Stone 1984]). These researchers from Stanford University and the University of California at Berkeley showed how this new algorithm could be used on a variety of different problems e.g. the detection of Chlorine from the data contained in a mass spectrum. Predictors are picked as they decrease the disorder of the data.

In building the CART tree each predictor is picked based on how well it pulls apart the records with different predictions. For instance one measure that is used to determine whether a given split point for a given predictor is better than another is the entropy metric. The measure originated from the work done by Claude Shannon and Warren Weaver on information theory in 1949. They were concerned with how information could be efficiently communicated over telephone lines. Interestingly, their results also prove useful in creating decision trees.

2.2.3. Linear regression

Linear regression is a statistical procedure for predicting the value of a dependent variable from an independent variable when the relationship between the variables can be described with a linear model. A linear regression equation can be written as $Y_p = mX + b$, where Y_p is the predicted value of the dependent variable, m is the slope of the regression line, and b is the Y-intercept of the regression line.

In statistics, linear regression is a method of estimating the conditional expected value of one variable y given the values of some other variable or variables x . It can also be said to be a statistical technique used to explain or predict the behavior of a dependent variable. The variable of interest, y , is conventionally called the dependent variable (endogenous variable/ output variable). The other variables x are called the independent variables (exogenous variables/ input variables).



Generally, a regression equation takes the form of $Y = a + bx + c$, where Y is the dependent variable that the equation tries to predict, X is the independent variable that is being used to predict Y , a is the Y-intercept of the line, and c is a value called the regression residual. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships.

Linear regression contains one predictor and a prediction. The relationship between the two can be mapped on a two dimensional space and the records plotted for the prediction values along the Y axis and the predictor values along the X axis. The simple linear regression model then could be viewed as the line that minimized the error rate between the actual prediction value and the point on the line (the prediction from the model). The simplest form of regression seeks to build a predictive model that is a line that maps between each predictor value to a prediction value. Of the many possible lines that could be drawn through the data the one that minimizes the distance between the line and the

data points is the one that is chosen for the predictive model. On average if you guess the value on the line it should represent an acceptable compromise amongst all the data at that point giving conflicting answers. Likewise if there is no data available for a particular input value the line will provide the best guess at a reasonable answer based on similar data.

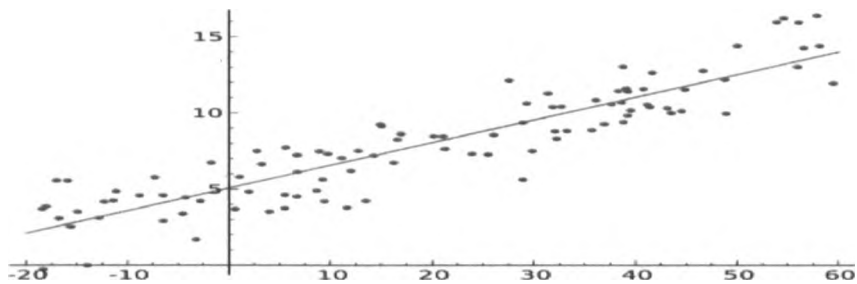


Fig. An illustration of linear regression on a data set.

In statistics, regression analysis includes any techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables — that is, the average value of the dependent variable when the independent variables are held fixed. Regression analysis is widely used for prediction and forecasting and is used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships.

2.2.3.1. Regression models

Regression models predict a value of the Y variable given known values of the X variables. Prediction within the range of values in the dataset used for model fitting is known informally as **interpolation**. Prediction outside this range of the data is known as **extrapolation**. Linear Extrapolation Method is a technique that uses trendline adjusted with indexes and forecasting is done beyond the known trend. Performing extrapolation relies strongly on the regression assumptions. The further the extrapolation goes outside the data, the more room there is for the model to fail due to differences between the assumptions and the sample data or the true values.

Regression models involve the following variables:

- The **unknown parameters** denoted as β ; this may be a scalar or a vector.
- The **independent variables**, $\mathbf{X}$.
- The **dependent variable**, Y .

A regression model relates Y to a function of $\mathbf{X}$ and β .

$$Y \approx f(\mathbf{X}, \beta)$$

The approximation is usually formalized as $E(Y | \mathbf{X}) = f(\mathbf{X}, \beta)$. To carry out regression analysis, the form of the function f must be specified. Sometimes the form of this function is based on knowledge about the relationship between Y and $\mathbf{X}$ that does not rely on the data. If no such knowledge is available, a flexible or convenient form for f is chosen.

2.2.4. Clustering and Nearest Neighbor

Clustering and the Nearest Neighbor prediction technique are among the oldest techniques used in data mining. Nearest Neighbor is a prediction technique that is quite similar to clustering (or grouping of records together). Its essence is that in order to predict what a prediction value is in one record look for records with similar predictor values in the historical database and use the prediction value from the record that is “nearest” to the unclassified record. Clustering is used mostly for consolidating data into a high-level view and general grouping of records into like behaviors.

2.2.5. Delphi Survey Method

The Delphi Survey technique is a popular method used in prediction. It involves a panel of experts that judge the timing, probability, importance and implications of factors, trends, and events regarding the problem in question. The basic idea of the Delphi method is to create a list of statements or questions, to have the experts give their ratings/answers, to make a report and send it out to everyone and have the experts revise their answers and make the second report.

2.2.6. Scenario Method

Scenario analysis is a process of analyzing possible future events by considering alternative possible outcomes or scenarios using expert judgment, rather than quantitative forecasts. Each scenario developed tells a story about an "alternative" future and encourages those who explore it to "stretch" their imaginations and suspend their preconceived judgments in order to develop a perspective on complex events contained in that world. Existing or newly formulated strategies and policies can then be tested against the future scenarios. The method is a powerful planning tool because it is based on the idea that the future is unpredictable; it enables one to ask what the future might hold and to identify relevant actions that can be taken today, no matter how the future turns out.

2.2.7. Statistical demand analysis

Statistical techniques are driven by the data and are used to discover patterns and build predictive models. Some factors are mapped which determine previous traffic growth and make traffic mapping into current map, so one can compare it when we design new traffic map.

2.2.8. Technology Foresight Method

Technology foresight is a prediction methodology for determining the most likely technological developments in the future. It is defined as "Systematic attempts to observe the long-term future of science, technology, the economy and society, with the aim to identify the emerging technologies that will probably produce the greatest economic and social benefits" (Creswell, 1994). In technology foresight projects the researchers and users of science and technology in the innovation system are brought together to develop a common vision of the future developments.

The timeframe of foresight usually ranges between 10 and 30 years. From the timeline given by this method, theorists may craft the rest of society, taking into consideration the tenet that social, political and economic patterns revolve around technological advancements.

2.2.9. Assessment method

This prediction method uses analogy method to observe same precondition which can grow to same condition in the future. This is more in comparison method.

2.2.10. Rule induction

Rule induction systems are highly automated and are probably the best of data mining techniques for exposing all possible predictive patterns in a database. They can be modified for use in prediction problems but the algorithms for combining evidence from a variety of rules comes more from rules of thumbs and practical experience. Rule induction systems when used for prediction are like having a committee of trusted advisors each with a slightly different opinion as to what to do but relatively well grounded reasoning and a good explanation for why it should be done. The business value of rule induction techniques reflects the highly automated way in which the rules are created which makes it easy to use the system but also that this approach can suffer from an overabundance of interesting patterns which can make it complicated in order to make a prediction that is directly tied to return on investment (ROI).

2.2.11. Neural Networks

Neural network is a mathematical model or computational model based on biological neural networks. It is an emulation of biological neural system. It can also be said to be an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases neural network is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Neural network is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems.

During World War II a seminal paper was published by McCulloch and Pitts which first outlined the idea that simple processing units (like the individual neurons in the human brain) could be connected together in large networks to create a system that could solve difficult problems and display behavior that was much more complex than the simple pieces that made it up.

Since that time much progress has been made in finding ways to apply artificial neural networks to real world prediction problems and in improving the performance of the algorithm in general.

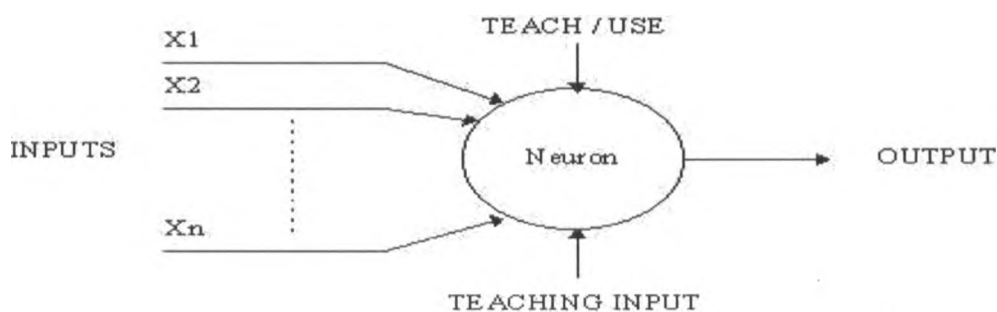
In many respects the greatest breakthroughs in neural networks in recent years have been in their application to more real world problems like customer response prediction or fraud detection. A good example of the Neural Networks is the use of Kohonen Neural Networks. Generally Kohonen Neural Network is known as one of the algorithms for Data Mining and Knowledge Discovery in Data.

Knowledge Discovery is very difficult task, but it is possible to use Kohonen Network for this purpose. The most important thing is to properly define the structure of the rules, which can be stored in the system, and then data have to fit to this construction, so training will try to find in the system some rules which will satisfy. Kohonen Networks are also used for classification, compression, pattern recognition and diagnostics. For Classification Kohonen Network can be used in a supermarket to group clients based on their purchases. For instance let's consider a shop with electronic shopping carts. The neural network can recognize all carts and what each customer has in his or her cart. After recognition, we can observe representative carts and in reality, what a typical Mr. "X" bought. But the network is more flexible than the statistics, because it can divide clients into groups: women will probably have different purchases to men, and young people will have different purchases to older ones, the same between single and married people, and so on. If our aim is to help the customer shop quickly, the manager will place products from a particular group close to each other; or if our target is to sell everything from stock, we can place all elements from a recognised group in different parts of the shop and in the paths between those products we can place things that are expensive or hard to sell. Kohonen Networks can also be used for compression. Taking the assumption that information can be lost, if there is a huge amount of data that can be used to train a network, then it can be used and from the network we can identify the most important information about data.

In compressed storage we can put just the most important information. For this usage it is important that the number of recognized details is chosen to determine the length of the output independently of the length of the input. Classification is similar to pattern recognition, but the important work of the network is done after training.

We train network using some data - this data can define classes (groups). After training we put new data as the input of the network. Diagnosis is a combination of the pattern recognition and classification. But the core of the system is to define diagnostic rules in the system using Kohonen Network. We define internal parameters in neural network to fit to different let's say diseases. Parameters can be the same as are typically when we control health in hospital. Then we put to the network data of other patient, and we can observe what is a hypothesis of his or her disease. Another possibility is that we describe the state of the patient then train the network with some data.

As an example of advance medicine application of the Kohonen network is used for image segmentation of the scans from the magnetic resonance [Reyes-Aldasoro]. The target in this example is to separate on the image cranium, cerebrum and other parts of human head. Kohonen network has been used to solve the Travelling Salesman Problem (TSP) [Brocki, 1984]. In his work the author assumes that every neuron represents single place that salesman visits. And then dedicated definitions, of the tasks and proper training of the networks, lead him to solutions which are quite good.



2.2.11.1. Kohonen Networks

Kohonen Networks are also used for classification, compression, pattern recognition and diagnostics. For Classification Kohonen Network can be used in a supermarket to group clients based on their purchases. For instance let's consider a shop with electronic shopping carts. The neural network can recognize all carts and what each customer has in his or her cart. After recognition, we can observe representative carts and in reality, what a typical Mr. "X" bought. But the network is more flexible than the statistics, because it can divide clients into groups: women will probably have different purchases to men, and young people will have different purchases to older ones, the same between single and

married people, and so on. If our aim is to help the customer shop quickly, the manager will place products from a particular group close to each other; or if our target is to sell everything from stock, we can place all elements from a recognised group in different parts of the shop and in the paths between those products we can place things that are expensive or hard to sell. Kohonen Networks can also be used for compression. Taking the assumption that information can be lost, if there is a huge amount of data that can be used to train a network, then it can be used and from the network we can identify the most important information about data.

In compressed storage we can put just the most important information. For this usage it is important that the number of recognized details is chosen to determine the length of the output independently of the length of the input. Classification is similar to pattern recognition, but the important work of the network is done after training. We train network using some data - this data can define classes (groups). After training we put new data as the input of the network. Diagnosis is a combination of the pattern recognition and classification. But the core of the system is to define diagnostic rules in the system using Kohonen Network.

We define internal parameters in neural network to fit to different let's say diseases. Parameters can be the same as are typically when we control health in hospital. Then we put to the network data of other patient, and we can observe what is a hypothesis of his or her disease. Another possibility is that we describe the state of the patient then train the network with some data. As an example of advance medicine application of the Kohonen network is used for image segmentation of the scans from the magnetic resonance [Reyes-Aldasoro].

The target in this example is to separate on the image cranium, cerebrum and other parts of human head. Kohonen network has been used to solve the Travelling Salesman Problem (TSP) [Brocki, 1984].

In his work the author assumes that every neuron represents single place that salesman visits. And then dedicated definitions, of the tasks and proper training of the networks, lead him to solutions which are quite good.

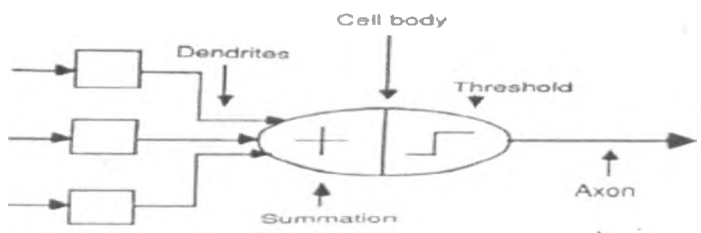
2.2.11.2. Feed Forward Artificial Neural Networks

Feedforward artificial neural networks consist of an architecture where there are no connections that loop back to nodes that have already propagated their output signal. They have the property of being static, producing only one output pattern for each input pattern. In an implementation where all nodes are numbered in ascending sequence this means that there are no connections from higher numbered nodes to lower numbered nodes. Single-layer feed forward networks consist of input and output layers only.

Multi-layer feed forward networks contain at least one hidden layer of nodes that receives connections from the previous adjacent layer of nodes. Connections between nodes in adjacent layers are common, but shortcut connections that circumvent one or more hidden layers may also exist. Artificial neural networks with an architecture that includes feedback connections are recurrent or feedback neural networks. As a result they are dynamic systems, entering more than one state for each new input pattern.

2.2.11.3. The Artificial Neural Network

An Artificial Neural Network is a system. A system is a structure that receives an input, process the data, and provides an output. The ANN library I have chosen to implement is a multilayer feedforward ANN, which is the most common kind of ANN. In a multilayer feedforward ANN, the neurons are ordered in layers, starting with an input layer and ending with an output layer. Between these two layers are a number of hidden layers. Connections in these kinds of network only go forward from one layer to the next. Multilayer feedforward ANNs have two different phases. A training phase (sometimes also referred to as the learning phase) and an execution phase. In the training phase the ANN is trained to return a specific output when given a specific input, this is done by continuous training on a set of training data. In the execution phase the ANN returns outputs on the basis of inputs.



The neuron model

2.2.11.4. The Artificial Neuron

A single artificial neuron can be implemented in many different ways. The general mathematic definition is as showed in equation

$$y(x) = g \left(\sum_{i=0}^n w_i x_i \right)$$

x is a neuron with n input dendrites $(x_0 \dots x_n)$ and one output axon $y(x)$ and where $(w_0 \dots w_n)$ are weights determining how much the inputs should be weighted.

g is an activation function that weights how powerful the output (if any) should be from the neuron, based on the sum of the input. If the artificial neuron should mimic a real neuron, the activation function g should be a simple threshold function returning 0 or 1.

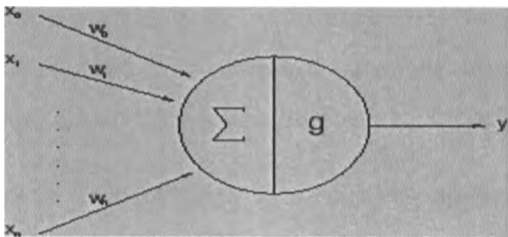
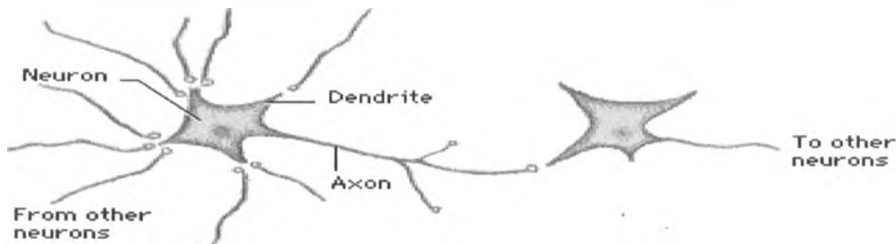
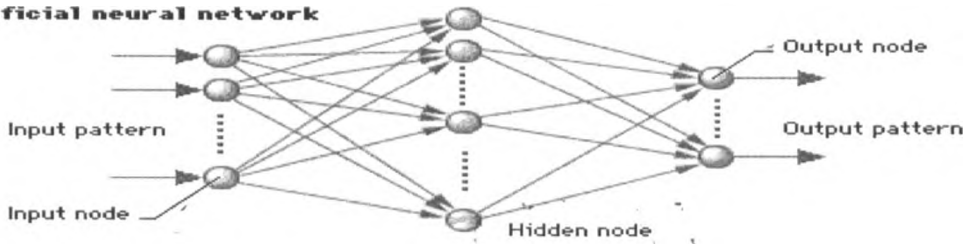


Figure 2: An artificial neuron.

Neural connections in animals



Artificial neural network



2.2.12. Training Artificial Neural Networks:

The purpose of neural network training is to produce appropriate output patterns for corresponding input patterns. It is achieved by an iterative learning process that updates the neural network weights based on the neural network response to a set of training input patterns. Learning can be categorized as supervised, reinforcement, unsupervised or hybrid. Supervised learning occurs when the correct output pattern is known and used during training. In reinforcement learning the correct output is not known but a measure of the correctness of a network output response can be computed. Unsupervised learning does not require a correct output to be available during training. Hybrid learning combines supervised and unsupervised approaches. Different learning rules form the basis of different training algorithms and their applicability is dependent on the neural network architecture and the learning category being used. Error-correction learning rules in supervised learning use the difference between the correct and actual output patterns to adjust connection weights with the aim of reducing this difference (error). The most popular supervised training algorithm for multi-layer feed forward neural networks is backpropagation. Based on the error-correction learning rule, backpropagation is a gradient descent algorithm that updates connection weights by reducing output error.

A **neural network** has to be configured such that the application of a set of inputs produces (either 'direct' or via a relaxation process) the desired set of outputs. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly, using prior knowledge. Another way is to '**train**' the **neural network** by feeding it with some teaching patterns and letting it change its weights according to some learning rule.

2.2.13. Supervised learning

The network is trained by providing it with input and matching output patterns. These input-output pairs can be provided by the system which contains the neural network (self-supervised). Supervised learning in neural networks generally involves three data sets:

- A Training Set. Input patterns used during training.
- A Test Set. Input patterns used to test the network after training is complete.
- A Validation Set. Input patterns used to determine when to stop training to prevent the network from becoming too specific to the training data and as a result not generalizing well.

This is achieved by testing the network on the validation set every n iterations of training on the training data. Training is stopped when the validation set error is at its minimum. The ability to perform well on data that was not used during training (test set) is generalization.

2.2.14. Back Propagation

Back Propagation is a common method of teaching artificial neural networks how to perform a given task. The back propagation algorithm is used in layered feedforward ANNs. This means that the artificial neurons are organized in layers, and send their signals “forward”, and then the errors are propagated backwards. The back propagation algorithm uses supervised learning, which means that the algorithm is provided with examples of the inputs and outputs we want the network to compute, and then the error (difference between actual and expected results) is calculated. In order to train a neural network to perform some task, we must adjust the weights of each unit in such a way that the error between the desired output and the actual output is reduced. This process requires that the neural network compute the error derivative of the weights (**EW**). In other words, it must calculate how the error changes as each weight is increased or decreased slightly. The back propagation algorithm is the most widely used method for determining the **EW**.

It is important to note that the error and the learning rate are used to modify weights through back propagation starting with the last neuron going backward. The idea of the back propagation is to reduce this error, until the ANN *learns* the training data. The inputs are then passed again.

2.2.15. Applications of Neural Networks

1) Neural Networks for clustering: Neural networks of various kinds can be used for clustering and prototype creation. Neural networks are used in a wide variety of applications. They have been used in all facets of business from detecting the fraudulent use of credit cards and credit risk prediction to increasing the hit rate of targeted mailings. They also have a long history of application in other areas such as the military for the automated driving of an unmanned vehicle at 30 miles per hour on paved roads to biological simulations such as learning the correct pronunciation of English words from written text.

2) Neural Networks for Outlier Analysis: Sometimes clustering is performed not so much to keep records together as to make it easier to see when one record sticks out from the rest. For instance most wine distributors selling inexpensive wine in Missouri and that ship a certain volume of product produce a certain level of profit. There is a cluster of stores that can be formed with these characteristics. One store stands out, however, as producing significantly lower profit. On closer examination it turns out that the distributor was delivering product to but not collecting payment from one of their customers. A sale on men's suits is being held in all branches of a department store for southern California.

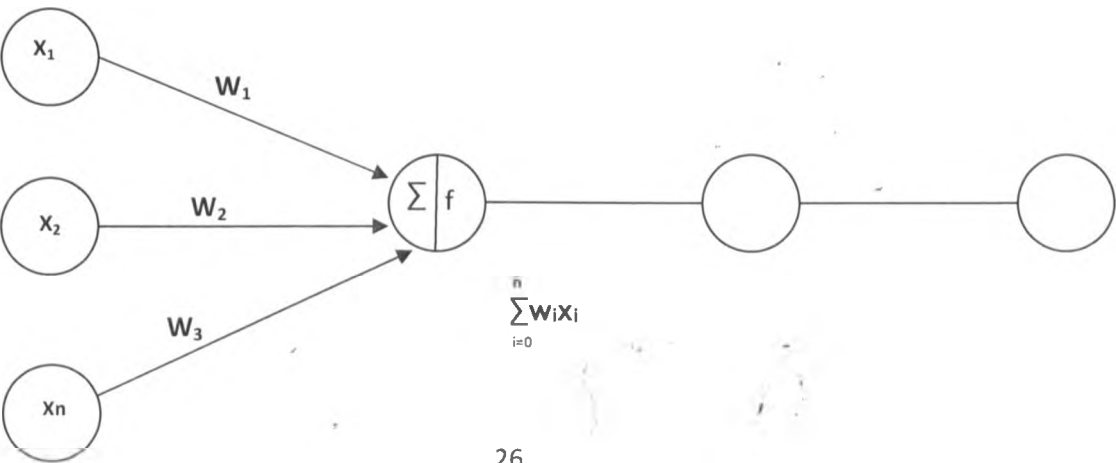
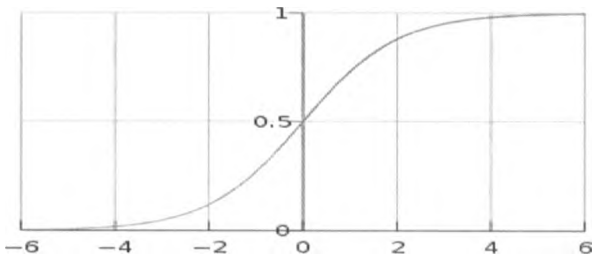
All stores with these characteristics have seen at least a 100% jump in revenue since the start of the sale except one. It turns out that this store had, unlike the others, advertised via radio rather than television.

2.2.16. Sigmoid Function

A logistic/sigmoid function computes a continuous output:

$$f(x) = \frac{1}{1 + e^{-kx}}$$

The basic formula for the sigmoid function above caters only for 'real' values of x from -1 to +1 and produces an output f(x) over the range 0 to +1.



CHAPTER THREE

3: Methodology

The goal was to predict the operators' traffic volumes and which operator gives maximum returns to the company. In doing this Time Series forecasting and Linear Regression analysis methods were used to fit a predictive model to an observed data set of durations and periods (day of the week) values. The approach was to first predict the Call Data Volumes to forecast which operator will give more volumes and hence yielding more returns/ revenue to the company and later optimize the resources allocated to the operator through checking the bandwidth allocated to respective routes. An Artificial neural network forecasting model was also developed. A prototype for each prediction method was developed that depicted how various prediction scenarios were achieved.

3.1. Data Collection

The sources of collection of data were: -

- The International Switching and Transmission section of Telkom Kenya (IGW).
- The IT Billing and Mediation section, Telkom Kenya
- Carrier Connect Services Department, Telkom Kenya

3.1.1. Steps of extracting the data:

- Using SQL Developer ETL tools data was loaded into the statistical software in this case Excel 2007. All the independent variables were properly organized starting with Operator, Calldate, Day, Duration
- The days were arranged in such a way that a mapping was formed Days were 1-Sunday, 2-Monday, 3-Tuesday, 4-Wednesday, 5-Thursday, 6-Friday, 7-Saturday
- The Moving Averages (MA) of the durations for 1 Cycle (7 Days) were obtained
- The Adjusted Seasonal Index (Ratios to Moving Averages) were obtained by getting the averages of all moving averages
- The Adjusted Seasonal Index was obtained by getting the ratio of Unadjusted Seasonal Index/ Averages of 1 cycle e.g. Adjusted Seasonal Index for all Mondays
- The Seasonally Adjusted Series were also obtained by the Product of Duration by Adjusted Seasonal Index.

3.2. Using Linear Regression

Linear regression uses the values from an existing data set consisting of measurements of the values of two variables, X and Y, to develop a model that is useful for predicting the value of the dependent variable, Y for given values of X. In this case X is the period while Y is the Forecasted Duration.

3.2.1. Trend Forecast

-To achieve the trend forecast a function is used that gets the value of the trend forecast

(=regression!\$B\$17+operator!K2*regression!\$B\$18)

- The Dependent Variable (Seasonally Adjusted Forecast) which represents the forecasted Duration is achieved by obtaining the Product of the Adjusted Seasonal Index and the trend forecast for that day.

3.2.2. Seasonal Forecast

The Actual Forecast (Seasonally Adjusted Forecast) is achieved through obtaining the Product of the Adjusted Seasonal Index and the trend forecast for that day.

Seasonal Forecast (i) = $N2 \cdot Q2$ - Trend forecast (i) * Adjusted Seasonal Index (i) for i= Period in cycle e.g. i= Monday, Tuesday etc.

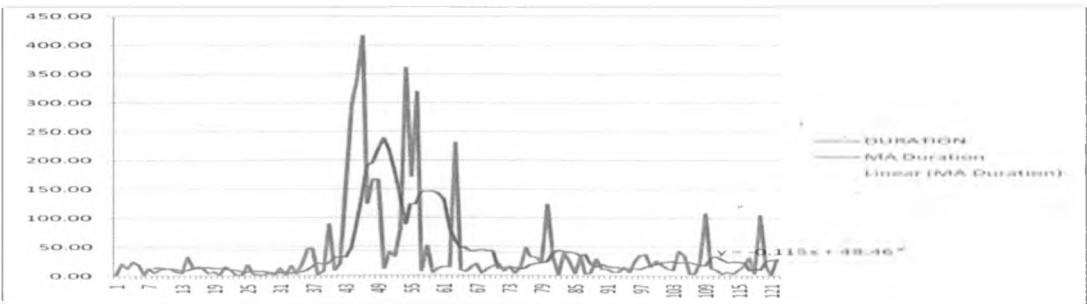
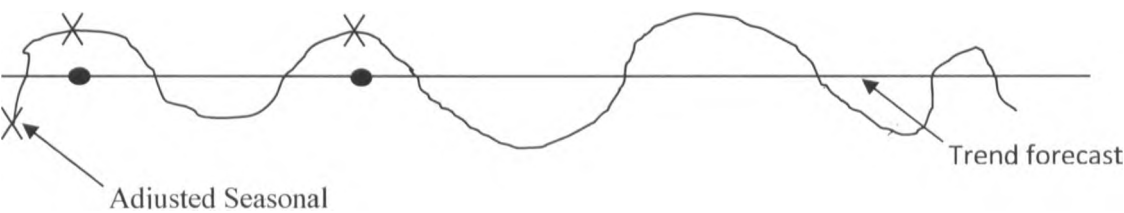


Fig. A graph showing seasonal variations of the durations vs Moving Averages over time for selected days. A linear regression equation is formed and is shown on the trend line. The equation is based on the data available. There are some few outliers which are an exceptional as observed on the graph.

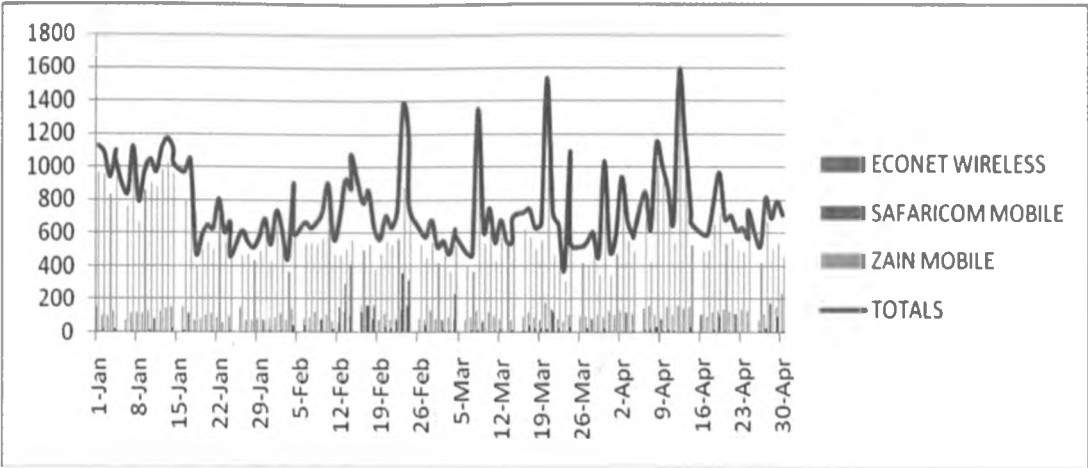


Fig. Operators' Traffic Data for the month of January-April 2011 showing seasonal changes and curves

Note:

- To cater for seasonal fluctuations we multiply the trend forecast with the adjusted seasonal index. This helps to simulate and smoothen the curve or the pattern. Time series are intelligent because they keep on learning from additional data. When the data grow the adjusted seasonal index changes since we have MA of more data. You extend the trend line for $j+1, j+2, \dots, j+n$

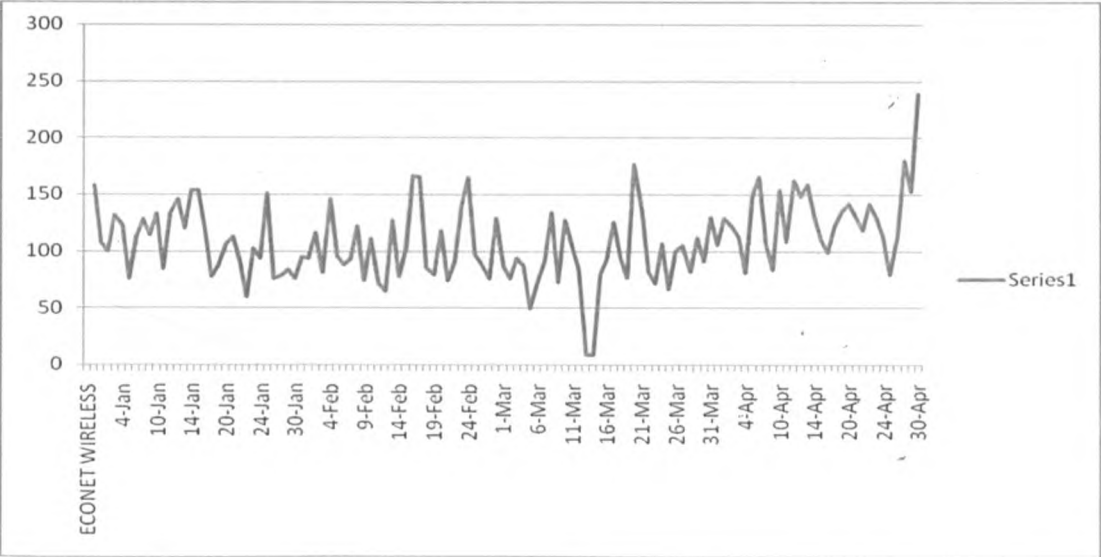


Fig. Econet Data for the month of January-April 2011 showing seasonal changes and curves

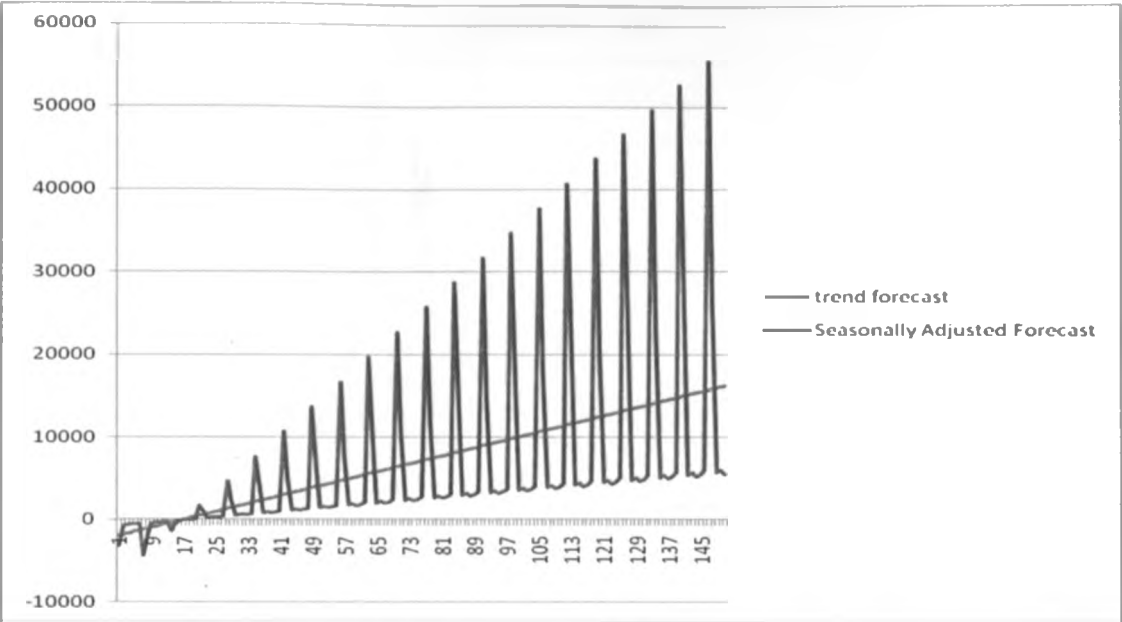


Fig. A graph showing Trend Forecast vs. Seasonally Adjusted Forecast. The trend line shows a linear increasing trend

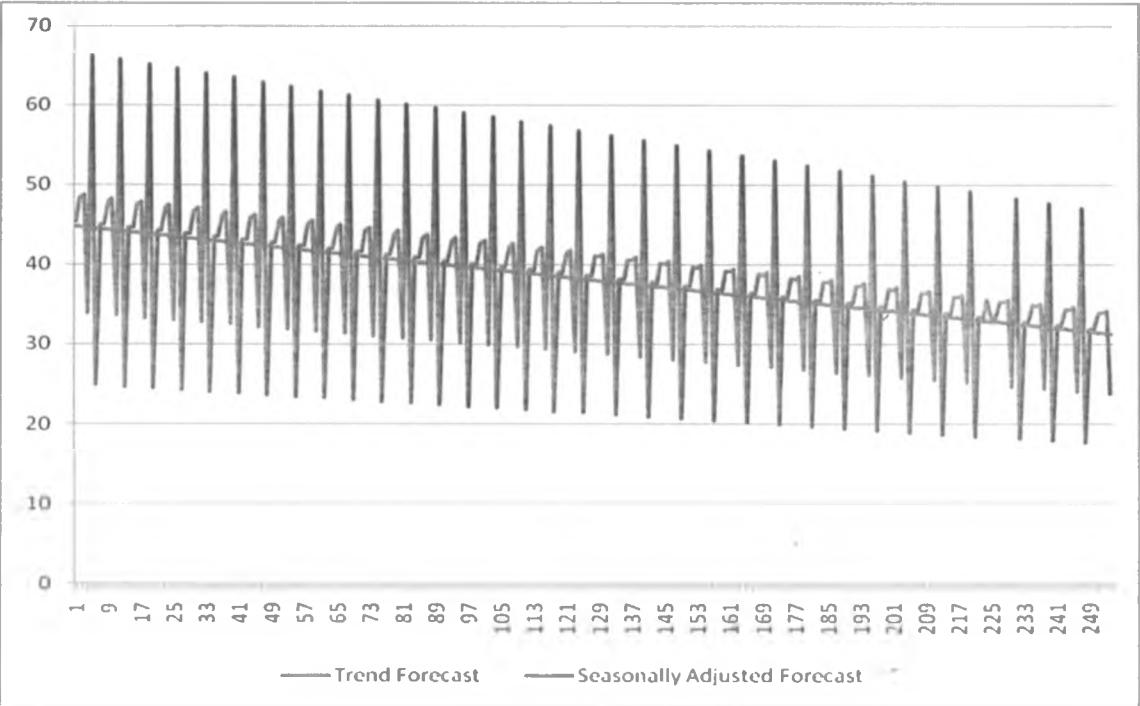


Fig. A graph showing Trend Forecast vs. Seasonally Adjusted Forecast for Safaricom. The trend line shows a linear decreasing trend

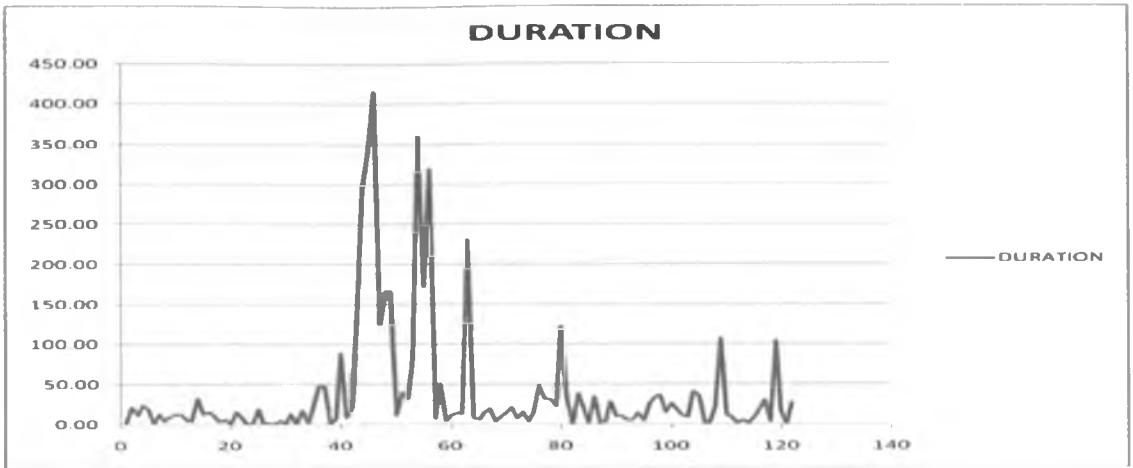


Fig. The seasonal effect of the durations with few outliers

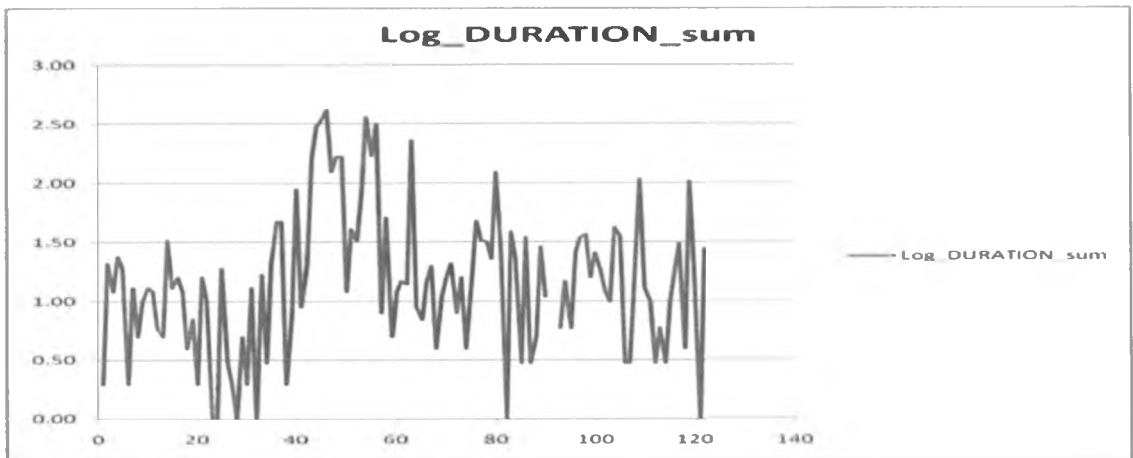


Fig. Logarithmic effect of the durations showing some seasonal fluctuations

3.2.3. Observations on data:

Data Patterns

- Linear pattern – Pattern goes into a straight line
- Cycles- (Up and down movement of the increasing trend)
- Predicted (Seasonal Forecast): is the expected traffic after removing the trend effect.
- Holding other factors constant like:
 - If it's a holiday then it's a normal holiday.
 - Under normal circumstances means without offers, promotions etc
 - Eliminate point events from analysis they are outliers. Points events means e.g. earthquakes

3.2.4. Elements of the research

The frequency of the data has been sub-divided into daily figures. The study used the analysis of the previous call durations to forecast the volumes from three different operators namely **Safaricom, Airtel and Econet.**

A regression equation takes the form of $Y=a+bx+c$, where Y is the dependent variable that the equation tries to predict, X is the independent variable that is being used to predict Y, a is the Y-intercept of the line, and c is a value called the regression residual. The values of a and b are selected so that the square of the regression residuals is minimized.

We want to predict revenue (Y)

Hypothesis Regression Model

$$Y= a+b_1operator +b_2day+b_3duration +b_4route$$

3.2.5. Data Cycles:

Using time series data was organized into cycles.

1 week Data= 1 cycle meaning we get 7 days moving averages

Note that the cycle starts at the 8th day i.e. after 7 days are over. There are no moving averages on the first cycle

Steps:

- 1) Take the average of the first cycle of either 7-days or 14-days
- 2) Move a step down in the next cycle to generate cycle 2 and take another moving averages

3.2.6. Traffic Volumes:

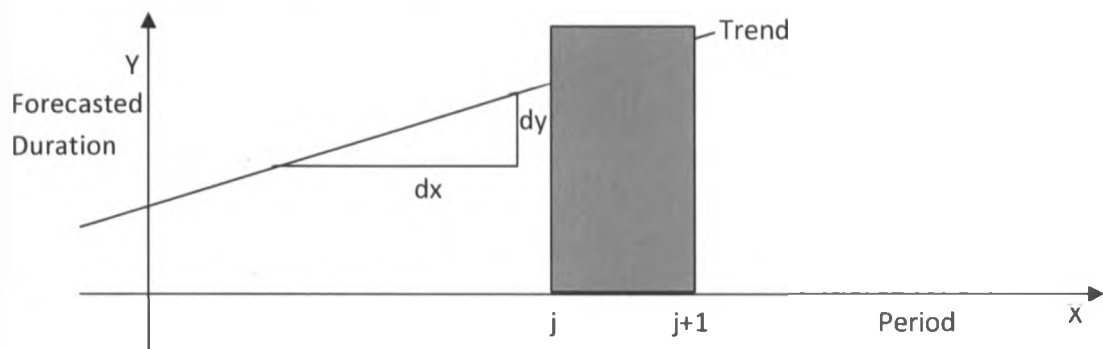
The traffic volumes of the Telecommunication operators were obtained. In order to obtain time series analysis, the traffic was collected for the four months period (January-April) and observations made.

Data was re-organized to obtain:

- 1) Moving Averages (**MA**) of the durations for 7-days----**1 Cycle** and Moving Averages (**MA**) of the durations for 14-days----**2 Cycles**
 - 2) Ratios to Moving Averages (**Ratio to MA**)
 - 3) Unadjusted Seasonal Index (**USI**) i.e. Averages of all moving averages
 - 4) Adjusted Seasonal Index (**ASI**) i.e. Ratios of averages of one cycle ...Seasonal Index is a value that holds seasonal variations, which are mainly caused by the change in seasons or short term movements in a periodic manner.
These fluctuations are repeated with more or less the same intensity within a specific period of one year or shorter. The simple averages of the monthly or quarterly values over a period of years are known as the seasonal averages. When these seasonal averages are expressed as percentage or grand average, they are called seasonal indices.
 - 5) Seasonally Adjusted Series (**SAS**) i.e. Normal observations holding everything else constant. It was achieved through obtaining the Product of the Durations and Adjusted Seasonal Index.
 - 6) Trend Forecast- To achieve the trend forecast a function was used that gets the value of the trend forecast (=regression!\$B\$17+econet!K2*regression!\$B\$18)
 - 7) Seasonally Adjusted Forecast (**SAF**) - It was achieved through obtaining the Product of the Adjusted Seasonal Index and the trend forecast for that day. Each period represent a day in the dataset and all periods have values which is automatically generated based on the day you pick from the system Calendar
- Seasonal Forecast (i) =Trend forecast (i) * Adjusted Seasonal Index (i)** for **i=** Period in the cycle e.g. **i=** Monday, Tuesday etc.

3.2.7. Modeling the data

The Linear Approach



Where j = Period up to when we have data

$j+1$ = are the periods we want to forecast for

periods $j+1, j+2, \dots, j+n$

Note: Time series are intelligent because they keep on learning from additional data. You extend the trend line for $j+1, j+2, \dots, j+n$

3.2.8. Fitting the data into a regression model

A model is a simplified representation or abstraction of the reality. All the dependent and independent variables were identified and then equations were established to describe their relationship. The data collected was analyzed using a statistical approach. The variables were as follow: -

- Independent Variables- Operator, Date (Day), Duration
- Dependent Variable- Forecasted Duration (seasonal forecast)- Y

A15		SAFARICOM												
	A	B	C	D	E	F	G	H	I	J	K	L	M	
	OPERATOR	DATE	C_DAY	DUR	MA	Ratio to MA	Unadj Seas Index	Adj Seas Index	Adj Seas Index	Seas Adj Series	Period	Trend Forecast	Seasonally Adj Forecast (Predicted Dur)	
2	SAFARICOM	1-Jan-11	7	263					1.113957	292.970592	1	238.2857	265.4399497	
3	SAFARICOM	2-Jan-11	1	217					1.121591	243.3852215	2	239.8901	269.05856	
4	SAFARICOM	3-Jan-11	2	342					0.515645	176.3506479	3	241.4945	124.5254752	
5	SAFARICOM	4-Jan-11	3	249					0.721554	179.6668954	4	243.0989	175.4089351	
6	SAFARICOM	5-Jan-11	4	197					1.218251	239.9955193	5	244.7033	298.1101258	
7	SAFARICOM	6-Jan-11	5	241					1.162416	280.1422869	6	246.3077	286.3120341	
8	SAFARICOM	7-Jan-11	6	138					0.029239	4.034955512	7	247.9121	7.248653954	
9	SAFARICOM	8-Jan-11	7	296	235.2857	1.258045	1.25804493	1.113956624	1.113957	329.7311606	8	249.5165	277.9505395	
10	SAFARICOM	9-Jan-11	1	304	240	1.266667	1.26666667	1.121590882	1.121591	340.9636282	9	251.1209	281.6548884	
11	SAFARICOM	10-Jan-11	2	147	252.4286	0.582343	0.58234295	0.515645169	0.515645	75.79983988	10	252.7253	130.3165671	
12	SAFARICOM	11-Jan-11	3	183	224.5714	0.814885	0.8148855	0.721553797	0.721554	132.0443448	11	254.3297	183.5125393	
13	SAFARICOM	12-Jan-11	4	296	215.1429	1.37583	1.37583001	1.218251367	1.218251	360.6024047	12	255.9341	311.7920257	
14	SAFARICOM	13-Jan-11	5	301	229.2857	1.312773	1.31277259	1.162416128	1.162416	349.8872546	13	257.5385	299.3668614	
15	SAFARICOM	14-Jan-11	6	308	237.8571	1.294895	1.29489489	0.029238808	0.029239	9.005552882	14	259.1429	7.57702826	

Fig A table showing seasonal indices for Safaricom and forecasted results for Safaricom Mobile for one cycle in the month of January 2011

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	Regression Statistics								
4	Multiple R	0.10477457							
5	R Square	0.01097771							
6	Adjusted R Square	-0.071440814							
7	Standard Error	86.30689554							
8	Observations	14							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	1	585.8044	585.8044	0.133195	0.721493			
13	Residual	12	52759.25	4396.604					
14	Total	13	53344.86						
15									
16		Coefficients	Standard Err	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	238.6813187	37.43148	6.323055	3.81E-05	155.1251	318.2375	155.1251	318.2375
18	X Variable 1	1.604395604	4.396104	0.364958	0.721493	-7.97389	11.18268	-7.97389	11.18268
19									
20									
21									
22	RESIDUAL OUTPUT								
23									
24	Observation	Predicted Y	Residuals						
25	1	238.2857143	24.71429						
26	2	239.8901099	-22.8901						
27	3	241.4945055	100.5055						
28	4	243.0989011	5.901099						
29	5	244.7032967	-47.7033						
30	6	246.3076923	-5.30769						
31	7	247.9120879	-109.912						
32	8	249.5184835	46.48352						
33	9	251.1208791	52.87912						
34	10	252.7252747	105.725						

Fig. Regression model Coefficients

3.2.9. Regression Model:

$Y = MX + C$ Where Y= Forecasted Variable (Dependent Variable).....Forecasted Duration

M= Gradient (Seasonal Change).....Represented in this case by X variable*Period

C= Intercept.....y intercept

$Y = (X\text{-variable} \times \text{Period}) + \text{Intercept}$

$Y = (1.604395604 \times \text{Period}) + (236.6813187)$

3.2.10. Examples of other Linear Regression Equations that can be formed from the regression coefficients

OPERATOR	DATE	DAY	DUR (Min)	MA	Ratio to MA	Unadjusted Seasonal Index	Adjusted Seasonal Index
ECONET	1-Jan-2011	7	158.00				
ECONET	2-Jan-2011	1	108.00				
ECONET	3-Jan-2011	2	100.00				
ECONET	4-Jan-2011	3	132.00				
ECONET	5-Jan-2011	4	123.00				
ECONET	6-Jan-2011	5	76.00				
ECONET	7-Jan-2011	6	113.00				
ECONET	8-Jan-2011	7	128.00				
ECONET	9-Jan-2011	1	115.00	115.71	0.99382716	1.116395045	0.360770647
ECONET	10-Jan-2011	2	133.00	111.43	1.193589744	1.164164976	0.376207825
ECONET	11-Jan-2011	3	85.00	112.43	0.756035578	1.062051674	0.343209218
ECONET	12-Jan-2011	4	133.00	117.14	1.135365854	1.098401069	0.354955772
ECONET	13-Jan-2011	5	146.00	110.43	1.322121604	1.199600394	0.387659022
ECONET	14-Jan-2011	6	121.00	111.86	1.081736909	10.954228	3.539933238
ECONET	15-Jan-2011	7	154.00	121.86	1.263774912	5.06647018	1.637264278

Fig A table showing seasonal indices for Econet for one cycle in the month of January 2011

A Linear Regression Equation is derived from the generated regression statistics: -

The model is developed based on degree of aggregation e.g. duration, period etc.

$Y = MX + C$ Where Y= Forecasted Variable (Dependent).....Forecasted Duration

$M = \text{Gradient (Seasonal Change).....Represented in this case by X variable*Period}$

$C = \text{Intercept.....y intercept}$

$Y = (X\text{-variable*Period}) + \text{Intercept}$

$Y = (121.0511*Period) + (-2061.543115)$

$Y = -2061.54 + 121.0511*Period$

A3		Regression Statistics							
	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	Regression Statistics								
4	Multiple R	0.215601438							
5	R Square	0.04648398							
6	Adjusted R Square	0.040084544							
7	Standard Error	24057.65666							
8	Observations	151							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	1	4204053404	4204053404	7.26376155	0.00784468			
13	Residual	149	86236855777	578770844.1					
14	Total	150	90440909181						
15									
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	-2061.543115	3935.09781	-0.523886118	0.601136136	-9837.348113	5714.261882	-9837.348113	5714.261882
18	X Variable 1	121.0511272	44.91465297	2.695136648	0.00784468	32.29918258	209.8030719	32.29918258	209.8030719
19									
20									

Fig. Regression model Coefficients

Regression Model formed from the above equation:

The Linear equation derived $Y = -2061.54 + 121.0511*Period$ forms the model for the data and will be used any time in obtaining the Trend Forecast.

OPERATOR	DATE	DAY	DUR [MINS]	Ratio to MA	Unadj Seasonal Index	Adj. Seasonal Index
ZAIN	1-Jan-2011	7	969.00			
ZAIN	2-Jan-2011	1	958.00			
ZAIN	3-Jan-2011	2	832.00			
ZAIN	4-Jan-2011	3	943.00			
ZAIN	4-Jan-2011	4	895.00			
ZAIN	6-Jan-2011	5	762.00			
ZAIN	7-Jan-2011	6	1,000.00			
ZAIN	8-Jan-2011	7	667.00			
ZAIN	9-Jan-2011	1	854.00	0.940084919	1.130797165	0.048361421
ZAIN	10-Jan-2011	2	901.00	1.041274558	1.042577621	0.039469731
ZAIN	11-Jan-2011	3	879.00	1.033596506	1.048113496	0.039818989
ZAIN	12-Jan-2011	4	981.00	1.140318831	1.0684139	0.04888835
ZAIN	13-Jan-2011	5	1,024.00	1.203088285	1.040854257	0.039844588
ZAIN	14-Jan-2011	6	956.00	1.107213766	6.897134938	0.060713887
ZAIN	14-Jan-2011	7	857.00	0.951316207	6.290487505	0.037782221

Fig A table showing seasonal indices for Zain for one cycle in the month of January 2011

3.3. Assumptions of the linear model

- 1) Period should not have lapses (i= 1,2, ...n)
- 2) No Moving Average on first cycle
- 3) Assumption = for cycles to apply data must be available from day 1-30. If data is not available we change the cycles from days to weeks.
- 4) Both the independent (X) and the dependent (Y) variables are measured at the interval or ratio level. You may also use categorical variables with two categories i.e. binary variable. If you have a categorical variable with more than two categories they have to be converted into ratios
- 5) The relationship between the independent (X) and the dependent (Y) variables is linearly true
- 6) Errors in prediction of the value of Y are distributed in a way that approaches the normal curve (Central limit theorem) –This is true in this model because there were more than 30 observation and therefore more than 30 error values
- 7) Errors in prediction of the value of Y are all independent of one another.
- 8) The distribution of the errors in prediction of the value of Y is constant regardless of the value of X.

3.4. Goodness of fit of the Regression Model:

To determine the goodness of fit of the regression model or rather a very good model r^2 is used. r^2 is a measure of association. It represents the percent of the variance in the values of Y that can be explained by knowing the value of X. r^2 varies from a low of 0.0 (none of the variance is explained), to a high of +1.0 (all of the variance is explained).

3.5. Time series

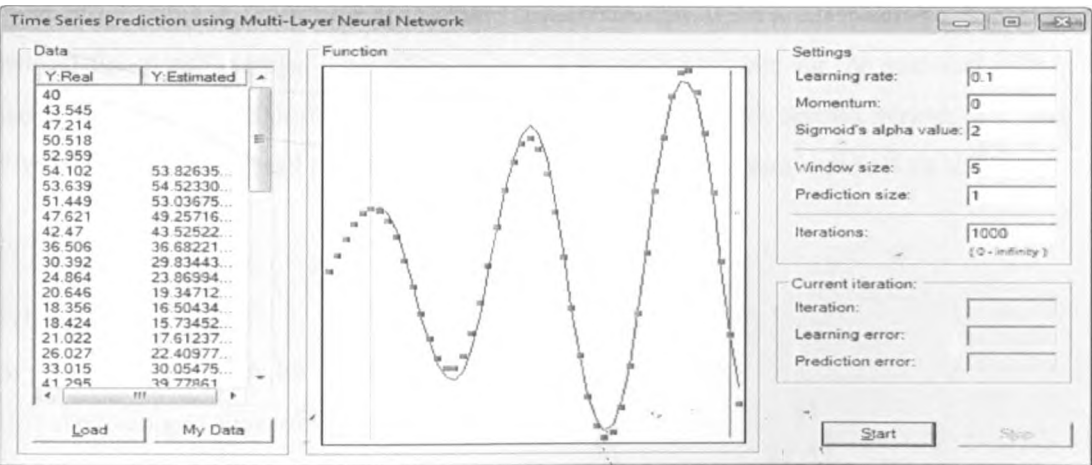
Time series forecasting is the use of a model to forecast future events based on known past events to predict data points before they are measured. Time series can also be said to be a sequence of data points from successive times taken at uniform intervals. This can be represented as follows:

$$\{ \cdots, x(t_{i-2}), x(t_{i-1}), x(t_i), x(t_{i+1}), x(t_{i+2}), \cdots \}$$

This data is used to analyze patterns and other meaningful characteristics. This data can then be used to forecast future values based on the already available data. The goals of time series analysis are: forecasting, modeling, and characterization.

3.5.1. Outliers

These are variations from the normal i.e. within given value limits for a set of data. They can be one-time, seasonal pulses or a sequential set of outliers with nearly the same magnitude and direction (level shift) or local time trends. A pulse is a difference of a step while a step is a difference of a time trend. To be able to mark a value as an outlier, one must identify the markers to identify the upper and lower “usual value” limits.



3.5.2. Predicting

Extending backward from time t , we have time series $\{x[t], x[t-1], \dots\}$. From this, we now want to estimate x at some future time

$$x[t+s] = f(x[t], x[t-1], \dots)$$

s is called the horizon of prediction.

In the training phase, some previous data values of the time series are presented to the network. The network is trained to predict the next value of the time series based on this data. More training samples mean better prediction models attained.

Window size—the number of values from the history used to predict the future one. The larger the window size, the better model gotten. A larger window size, on the other hand, will decrease the amount of training samples you need, so it is a trade-off value.

3.5.3. Modeling Seasonality and trend

Seasonality is a pattern that repeats for each period.

3.5.4. Seasonal Index

Seasonal index represents the extent of seasonal influence for a particular segment of the year. Seasonal index represents the extent of seasonal influence for a particular segment of the year. The calculation involves a comparison of the expected values of that period to the grand mean.

A seasonal index is how much the average for that particular period tends to be above (or below) the grand average. Therefore, to get an accurate estimate for the seasonal index, we compute the average of the first period of the cycle, and the second period, etc., and divide each by the overall average. The formula for computing seasonal factors is:

$$S_i = D_i / D,$$

Where:

S_i = the seasonal index for i^{th} period,

D_i = the average values of i^{th} period,

D = grand average,

i = the i^{th} seasonal period of the cycle.

A seasonal index of 1.00 for a particular month indicates that the expected value of that month is $1/12$ of the overall average. A seasonal index of 1.25 indicates that the expected value for that month is 25% greater than $1/12$ of the overall average. A seasonal index of 0.80 indicates that the expected value for that month is 20% less than $1/12$ of the overall average.

3.5.5. Seasonal Adjustment

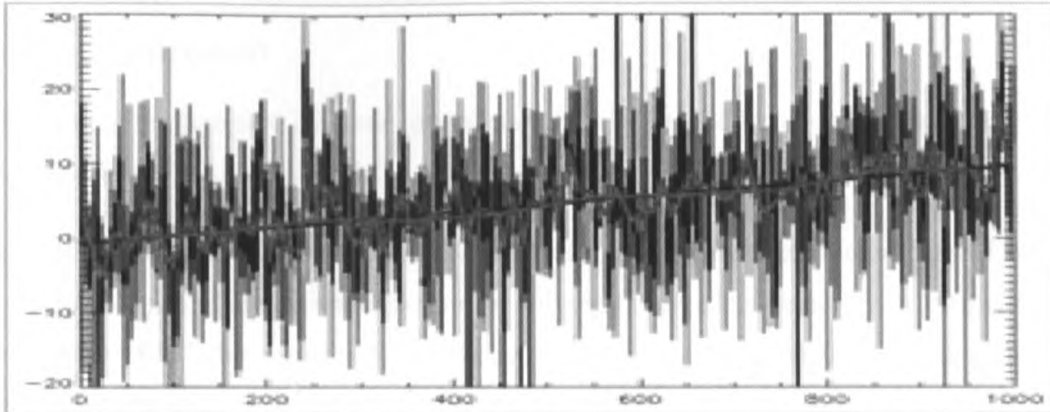
Seasonal Adjustment is the process of removing recurrent and periodic variations over a short time frame, e.g., weeks, quarters, months. Therefore, seasonal variations are regularly repeating movements in series values that can be tied to recurring events. The deseasonalized data is obtained by simply dividing each time series observation by the corresponding seasonal index.

3.5.6. Forecasting

Incorporating seasonality in a forecast is useful when the time series has both trend and seasonal components. The final step in the forecast is to use the seasonal index to adjust the trend projection. One simple way to forecast using a seasonal adjustment is to use a seasonal factor in combination with an appropriate underlying trend of total value of cycles.

3.5.7. Advantages of Time Series

- Time series data have a natural temporal ordering which makes it distinct from other common data analysis problems, in which there is no natural ordering of the observations
- Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations
- A time series model will generally reflect the fact that observations close together in time will be more closely related than observations further apart
- Time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values



Time series: random data plus trend, with best-fit line and different smoothings

3.5.8. Trend line

A trendline represents a trend, the long-term movement in time-series data. A trend is a regular, slowly evolving change in the series level. It tells whether a particular data set (say duration) have increased or decreased over the period of time. When using a trendline to forecast, it gives an assurance that conditions are the same throughout your forecasting period (purely linear). This is in most cases not true. Data (once in series) exhibits seasonal patterns where there curves. To reflect this reality in the data we use the Adjusted Seasonal Index which is supposed to shape the straight line (trend line) to show the curves.

3.6. System Analysis

The analysis involved a detailed study of the Switching operations and the way Call Data Records (CDRS) are received from various operators leading to specifications of a new prototype system. During analysis, I collected data on the available files for four months. The data was resident in the mediation server.

Using the following steps it became easy to draw the exact boundary of the new prototype system under consideration: -

- Keeping in view of the problems and the requirements
- Working out the pros and cons including all the areas of the prototype system.

3.6.1. Approach

Modeling of the complex, dynamic interrelationships of data collected incorporating rules into the data and the model build.

Evaluation of potential courses of action for solving the prediction problem-

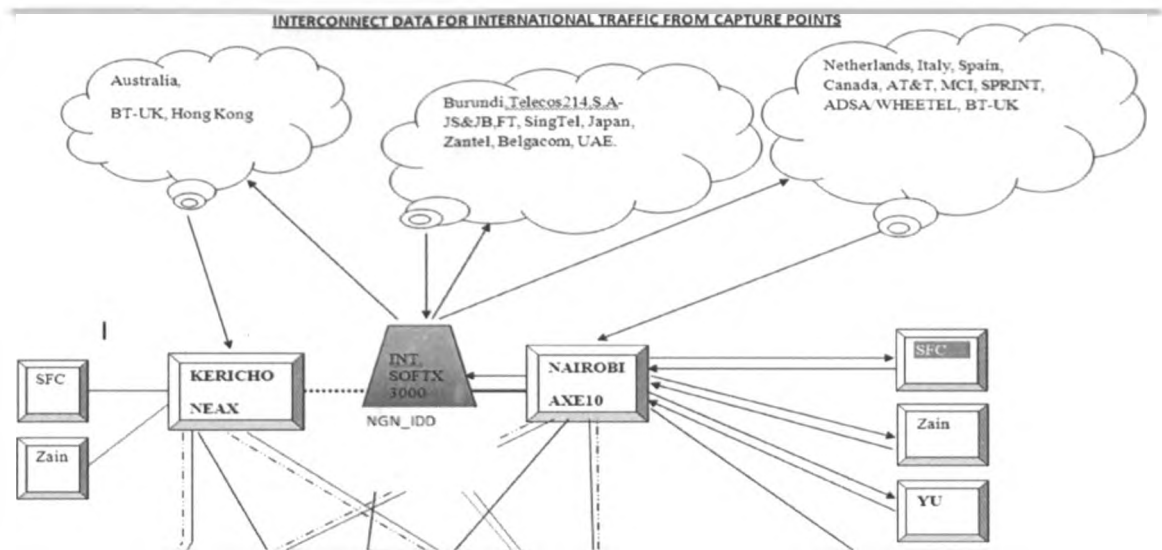
Evaluation of potential courses of action for solving the problem began by investigating the sensitivity of the results of the modeling to assumptions made by the model (sensitivity analysis).

3.7. System Design

Using the **waterfall methodology** a prototype system was designed using the following Platforms: -

- i) **Oracle 10g Database:** The Oracle database served as the backend database for the storage of information. The information stored included the relevant traffic trends. Other information available was based on the need and preference of the user. Oracle is one of the best management system database used in the development of information systems and it has a large capability of storing data i.e. it provides a large data warehouse with a large database geared towards the business intelligence requirements of an organization. The data warehouse integrates data from the various operational systems and is typically loaded from these systems at regular intervals.
- ii) **VB.Net:** This served as the platform for developing the front-end GUI and provided a well-designed user-interface which assisted the user in accessing the relevant data. In so doing the interface acted as a tool for use by all stakeholders in providing information to the other parties.
- iii) **Crystal Reports version 10.6:** Crystal Reports provided a platform for various reports that would be posted on the reporting dashboard to be used by the respective teams.

INTERCONNECT SCHEMA



Traffic Flow to the Switch

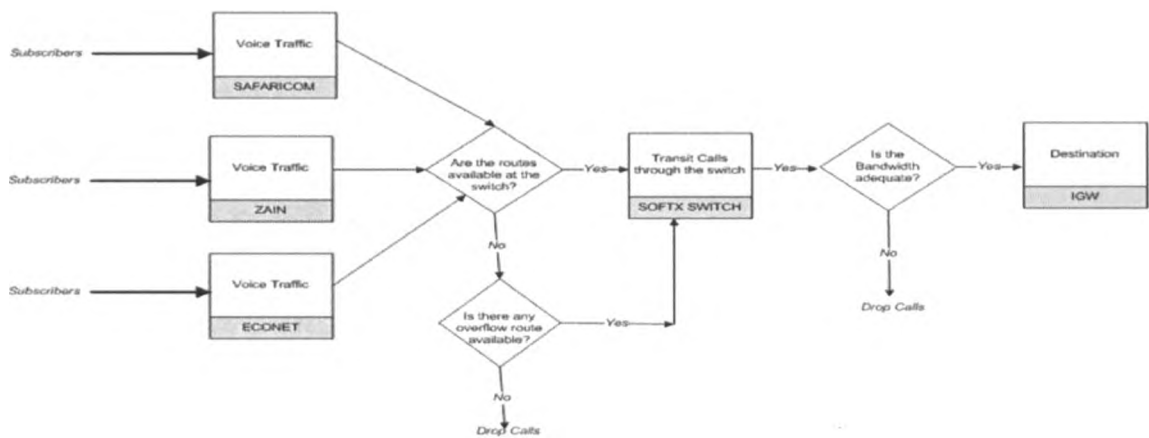


Fig. Traffic flow to the switch from local operators to international destinations

VOICE INTERCONNECTION TOPOGRAPHY

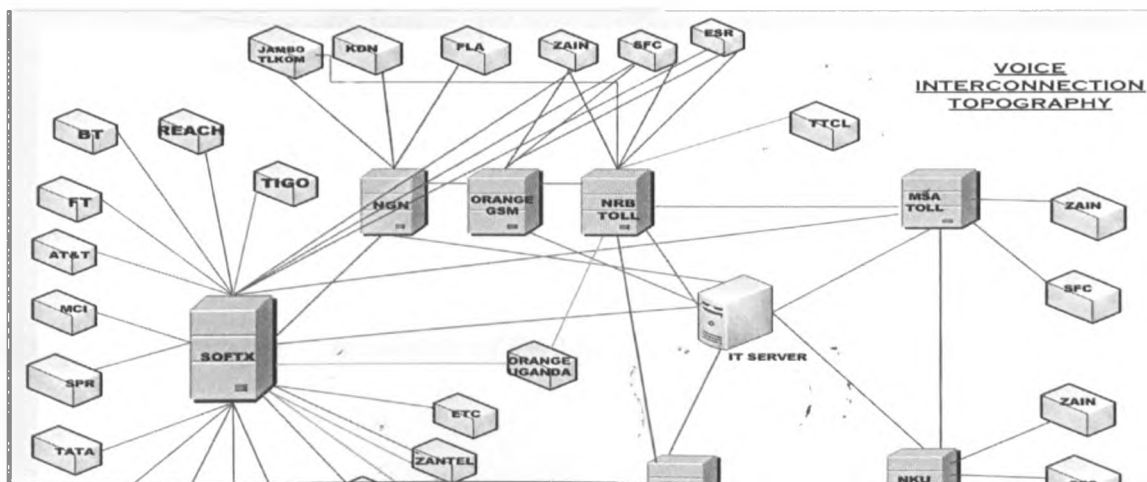


Fig. Voice Interconnection Topography at the International Switch

3.7.1 Block Diagram

Block Diagram displaying the components of the Telecommunication Traffic Forecasting Model

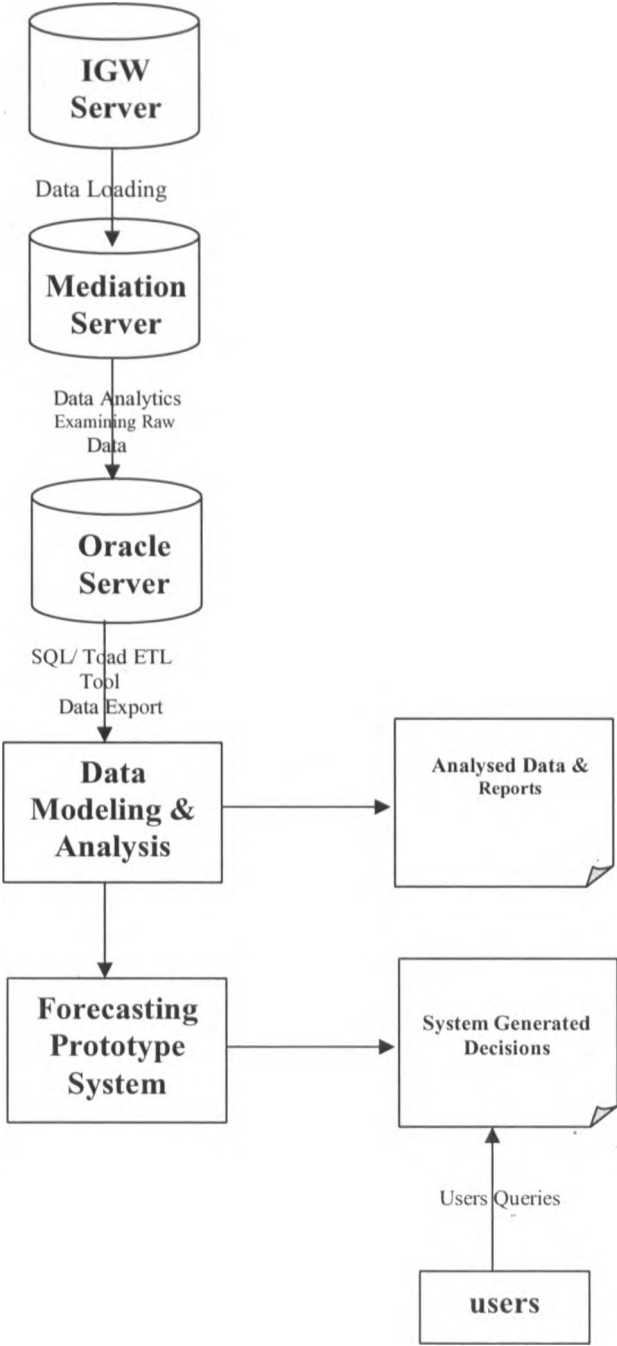


Fig A Block Diagram showing the Components of the Telecommunication Forecasting System model

3.7.1. Design of the Prototype Web-based System

The prototype design is web-based due to the fact that users can check the status of the bandwidth, resources allocated and update where applicable. The possible prediction scenario was carried out by accepting predictor values at the input nodes and performing calculations on those values to produce new values in the output node. The value at this node (Output node) represents the prediction from the neural network model. In my case the representation for the predictors values are:-

Input Nodes = Operator, Day, Duration (Independent variables)

Output Nodes = Forecasted Duration (dependent variable)

3.7.1.1. Traffic Forecasting

Traffic is predicted based on the period specified. The system retrieves the Seasonal Indices for the period and then using the trend forecast it predicts the traffic volumes.

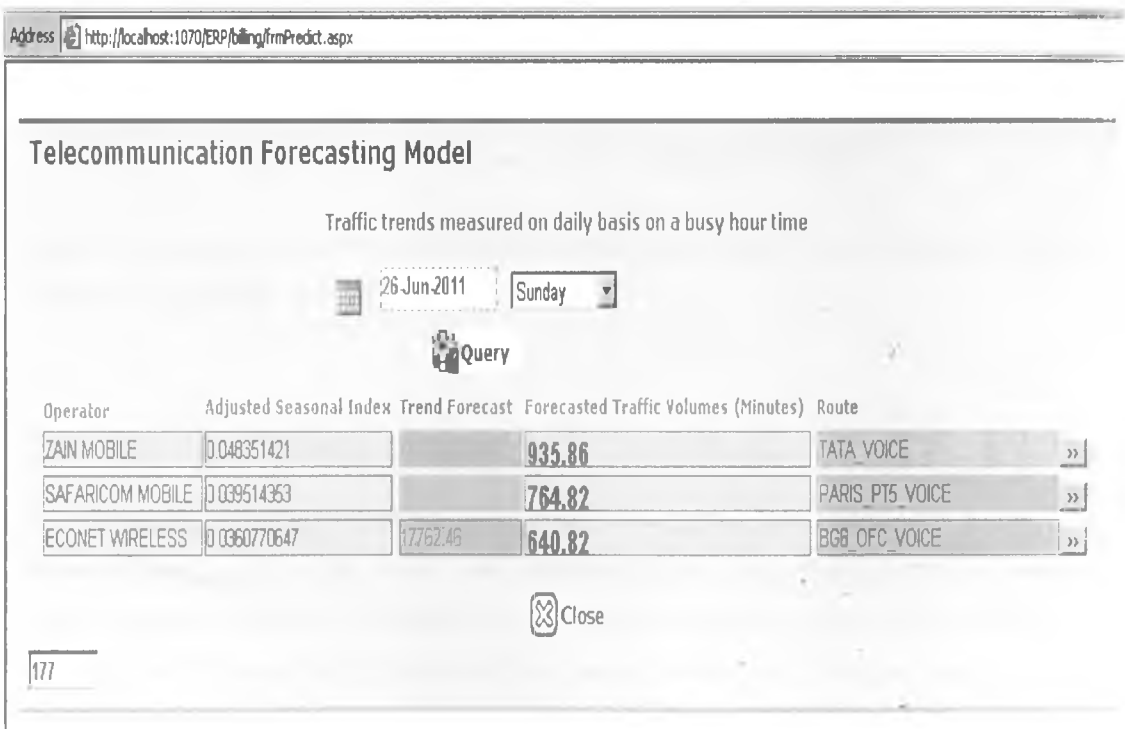


Fig A prototype module showing the predicted traffic volumes for three operators Zain, Safaricom and Econet for the period stipulated

3.7.1.2. Bandwidth Optimization

Bandwidth optimization is done based on the routes generating maximum revenues to the company.

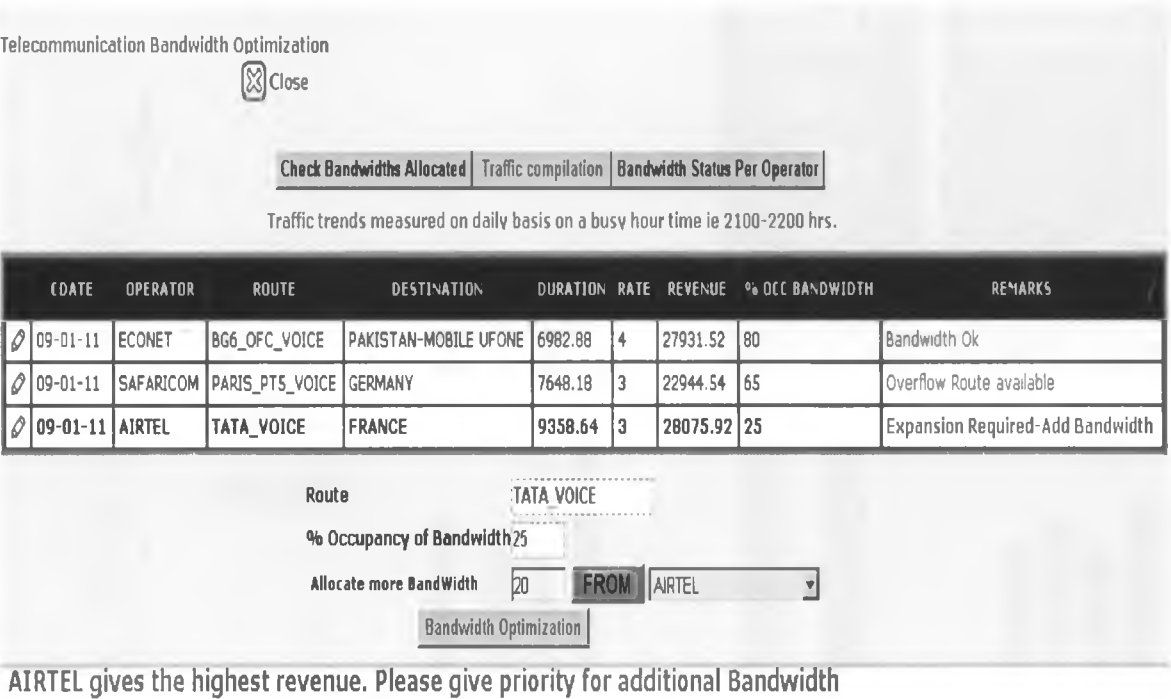


Fig A prototype module showing Bandwidth optimization for the corresponding routes of the three operators Zain, Safaricom and Econet

3.8. System Development

A web-based prototype system was developed based on the linear regression model. The objective was to utilize the model and available data to forecast on the traffic volumes each operator was transiting through the international gateway to the outside world. The system utilized transit rate to calculate the expected revenue and this gave way for better optimization.

Oracle database was used for the storage of data. The statistical forecasting system was developed in Visual Basic.Net which provided a Graphical User Interface (GUI).

Figure 1: Waterfall Development Infrastructure overview

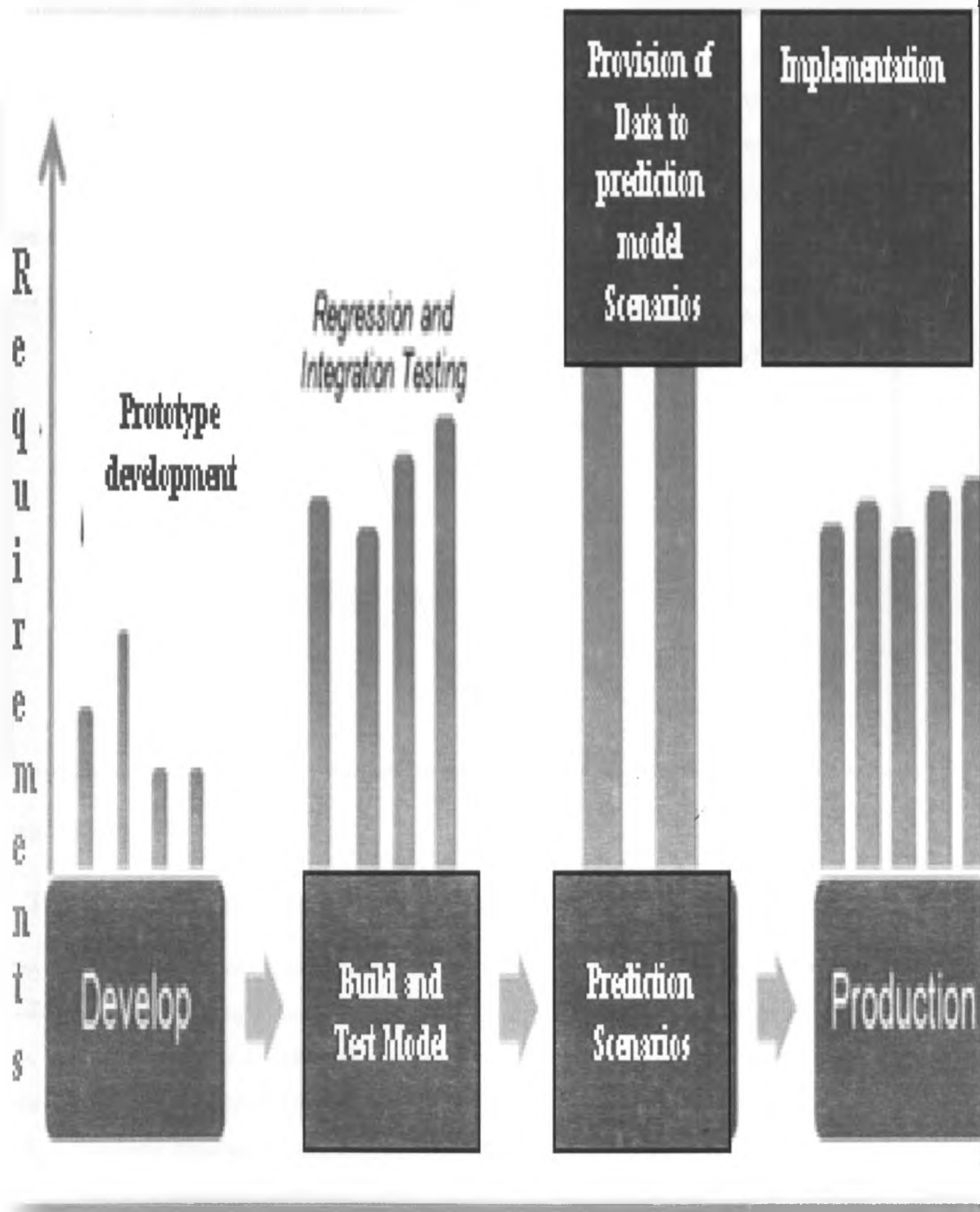


Fig. Telecommunication Forecasting Model- Prototype Development Stages

3.8.1. The Telecommunication Forecasting System (TFS)

The TFS was developed to predict the traffic volumes transiting through the Main International Switch to the outside world. The transit rate is then used to obtain the operator giving the company maximum revenue. A decision to increase the capacity of bandwidth is made based on the revenue stream to the company. The solution to the optimization of bandwidth is based on an algorithm where the source code was done in VB.Net and integration done through call procedures. The system runs on Windows XP environment

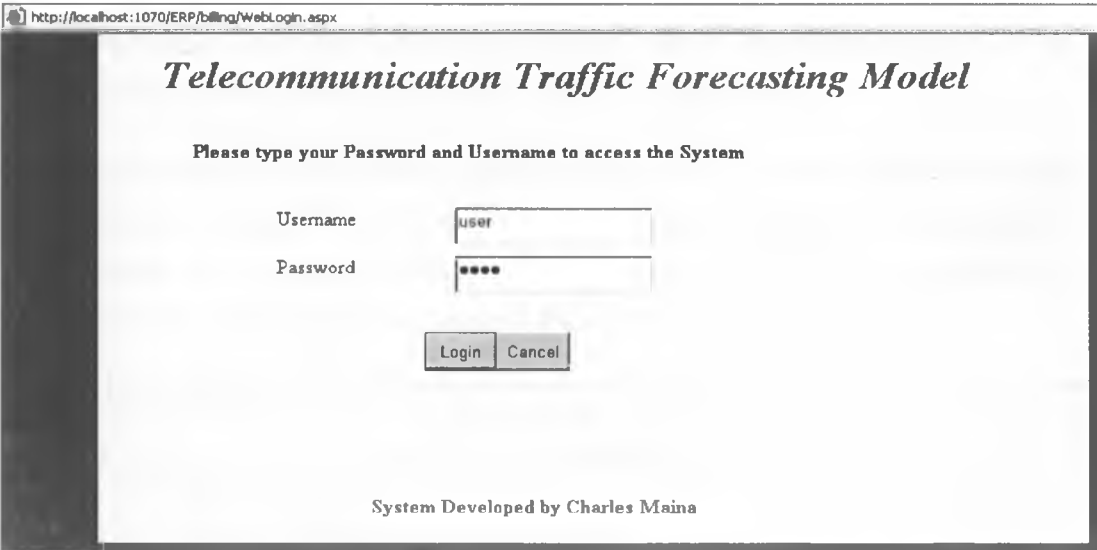
The image shows a web browser window with the address bar displaying "http://localhost:1070/ERP/billing/WebLogin.aspx". The main content area has a title "Telecommunication Traffic Forecasting Model" in a bold, italicized font. Below the title is a prompt: "Please type your Password and Username to access the System". There are two input fields: "Username" with the text "user" and "Password" with four dots. Below these fields are two buttons: "Login" and "Cancel". At the bottom of the page, it says "System Developed by Charles Maina".

Fig. Authentication Interface

3.8.2. Telecommunication Forecasting Model

This subsystem allows various parameters for the model to be set. The system works through obtaining the moving averages (MA) for one cycle and the consequent ratios to MA. This yields an Adjusted Seasonal Indices for one cycle consisting 7 days. This means one can obtain Seasonal Indices for all the Days in the Cycle from Monday to Sunday. The Seasonal Indices for the 1 cycle (seven days) are stored in a hashtable and retrieved any time a forecast is being done. The linear model is represented by the coefficients of dependent variable Y and independent variable X and is drawn through a linear equation that forms a trend line based on the seasonal changes resulting from the available data. Each operator exhibit their own model based on the observed data for 4-months i.e. Jan- April data.

The Seasonal Forecast (Forecasted Duration) is achieved by introducing the Adjusted seasonal index and multiplying by the Trend forecast for that particular day. The Adjusted Seasonal Index are occurring seasonal variations, which are mainly caused by the change in seasons, are short term movements in a periodic manner. These fluctuations are repeated with more or less the same intensity within a specific period of one year or shorter. An index number which measures the seasonal variations is called seasonal index number. Note that the Adjusted Seasonal Index for each day is generated by the system since it is stored in the hashtable. The Trend Forecast (TF) is achieved out of the linear model that is formed based on the data available in the dataset for each operator

($Y = -2061.54 + 121.0511$). The period (Day of the cycle) is factored. The Route forecasted to be generating maximum duration is picked for bandwidth optimization.

The simple averages of the monthly or quarterly values over a period of years are known as the seasonal averages. When these seasonal averages are expressed as percentage or grand average, they are called seasonal indices. Adjusted Seasonal Index is obtained by getting the Ratios of all Unadjusted Seasonal Indices for one cycle.

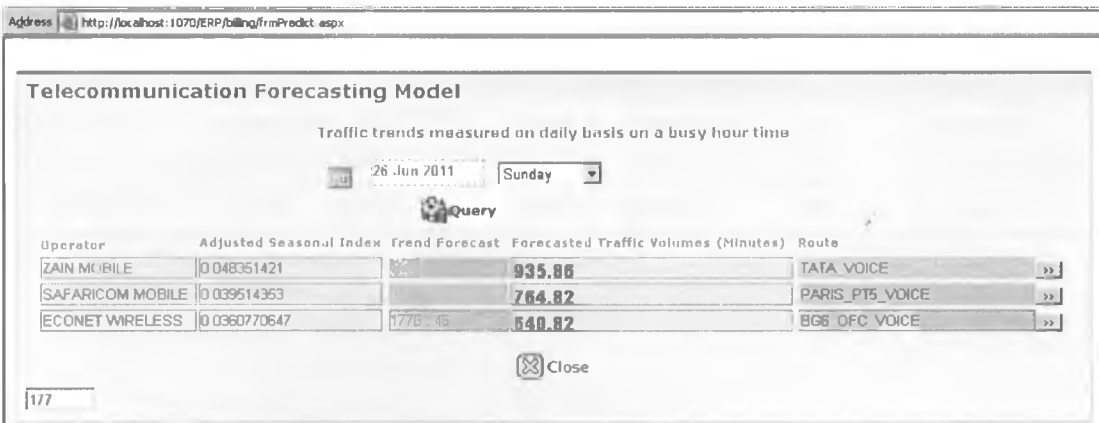


Fig. Telecommunication Forecasting Model Interface

3.8.3. Telecommunication Bandwidth Optimization

To optimize bandwidth an independent variable Rate is introduced and is multiplied by the duration of the operator’s Routes and Revenue output values are realized. Bandwidth Optimization is done based on this fact. The threshold set for minimum bandwidth is 50% and the system generates decisions based on rules.



Check Bandwidths Allocated Traffic compilation Bandwidth Status Per Operator

Traffic trends measured on daily basis on a busy hour time i.e 2100-2200 hrs.

DATE	OPERATOR	ROUTE	DESTINATION	DURATION	RATE	REVENUE	% OCC BANDWIDTH	REMARKS
09-01-11	ECONET	BG6_OFC_VOICE	PAKISTAN-MOBILE UPHONE	6982.88	4	27931.52	80	Bandwidth Ok
09-01-11	SAFARICOM	PARIS_PTS_VOICE	GERMANY	7648.18	3	22944.54	65	Overflow Route available
09-01-11	AIRTEL	TATA_VOICE	FRANCE	9358.64	3	28075.92	25	Expansion Required-Add Bandwidth

Route TATA_VOICE

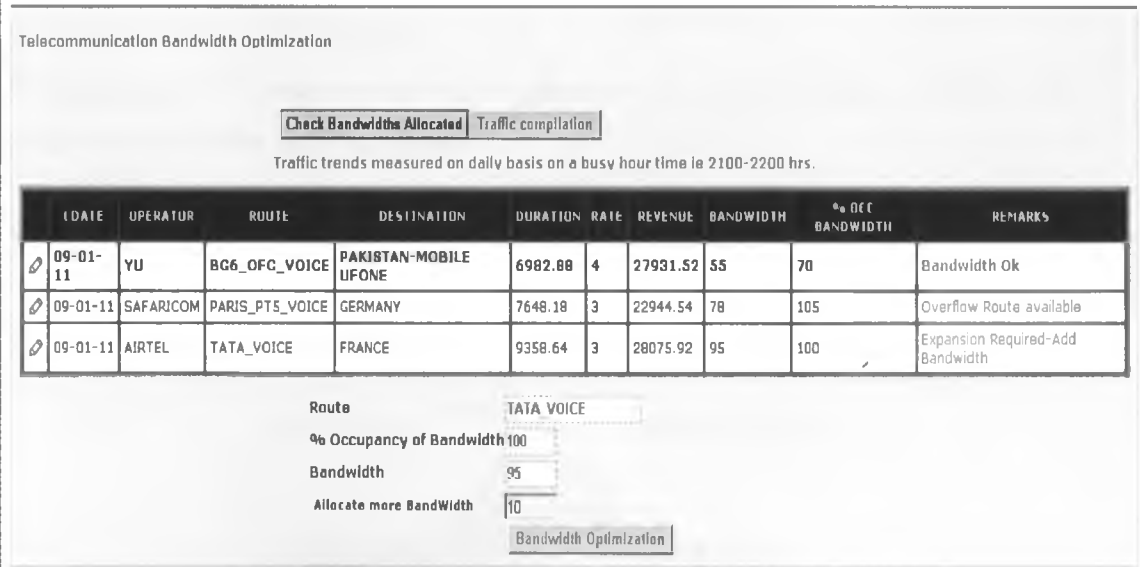
% Occupancy of Bandwidth 25

Allocate more BandWidth 20 FROM AIRTEL

Bandwidth Optimization

AIRTEL gives the highest revenue. Please give priority for additional Bandwidth

Fig. Telecommunication Bandwidth Optimization Interface displaying the operator generating highest revenue.



You have allocated 10% of Econet Bandwidth to Airtel Route TATA_VOICE Your bandwidth is now upgraded

Fig. Telecommunication Bandwidth Optimization Interface displaying allocated Bandwidth to a particular Route

3.8.4. System Testing

I tested the prototype system as follow: -

- 1) Using Linear Regression I tested the system using 1 cycle of data for one operator and using the regression model I computed the predicted values based on the seasonal indices and trend forecast. One of the days used for Test is Friday 3rd June 2011 which showed a forecasted duration of **484 Minutes**. Then I tested for the same day on the developed Forecasting model which produced good results.
- 2) I performed test for each operator transiting traffic through the switch at different periods.
- 3) For each scenario the system will use a prediction function that will predict the operator giving maximum returns to the company.
- 4) For Decision-support a rule-based algorithm will be used that will generate the Bandwidth utilization of the operators that have reached a 100% utilization and recommend for Bandwidth upgrade.
- 5) Summaries of the above tests will be generated for decision-making and this will conclude that the tests were 'successful'.

3.9. Using Neural Networks

3.9.1. Neural Network Traffic Forecasting Model

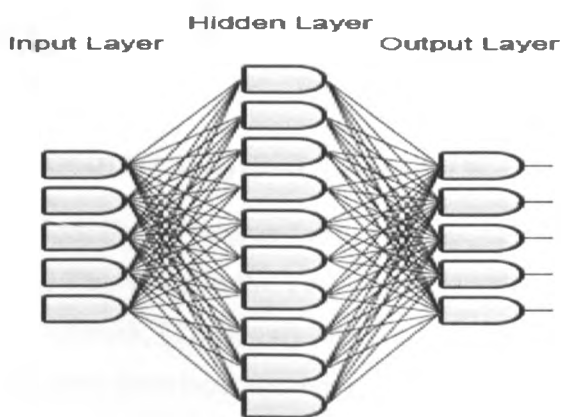


Figure 4: Schematic of a neural network with three layers of neurons.

3.9.2. Types of the Layers:

1) **The input Layer**- This layer introduces input values into the network.

No activation function or other processing

2) **The hidden Layer**- This layer performs classification of features.

3) **The output Layer**- In this layer outputs are passed on to the world.

3.9.3. The Neural Network/Brain

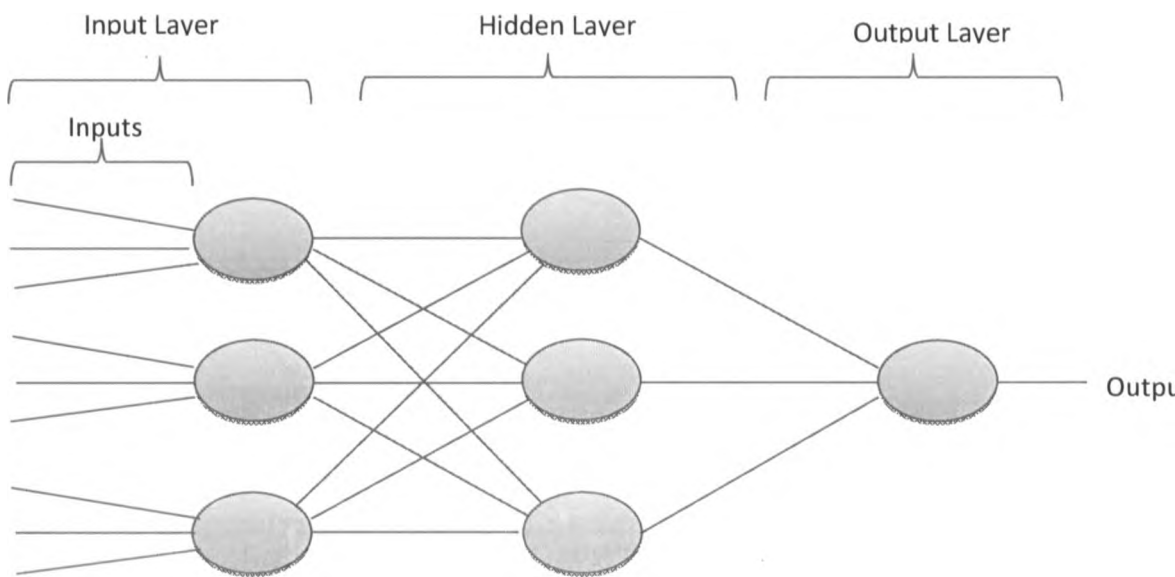


Fig: The Neural Network/Brain

3.9.4. Neural Network Program Structure

Basically the neural network (in the program) receives a set of 3 inputs (**The Operator**, **The Day of the week** and **The Week in the Year**), and attempts to predict the amount of traffic that will be in the network, (**Thus, only one output**). XML has been used to store the initial brain state. We have three input neurons and one output neuron.

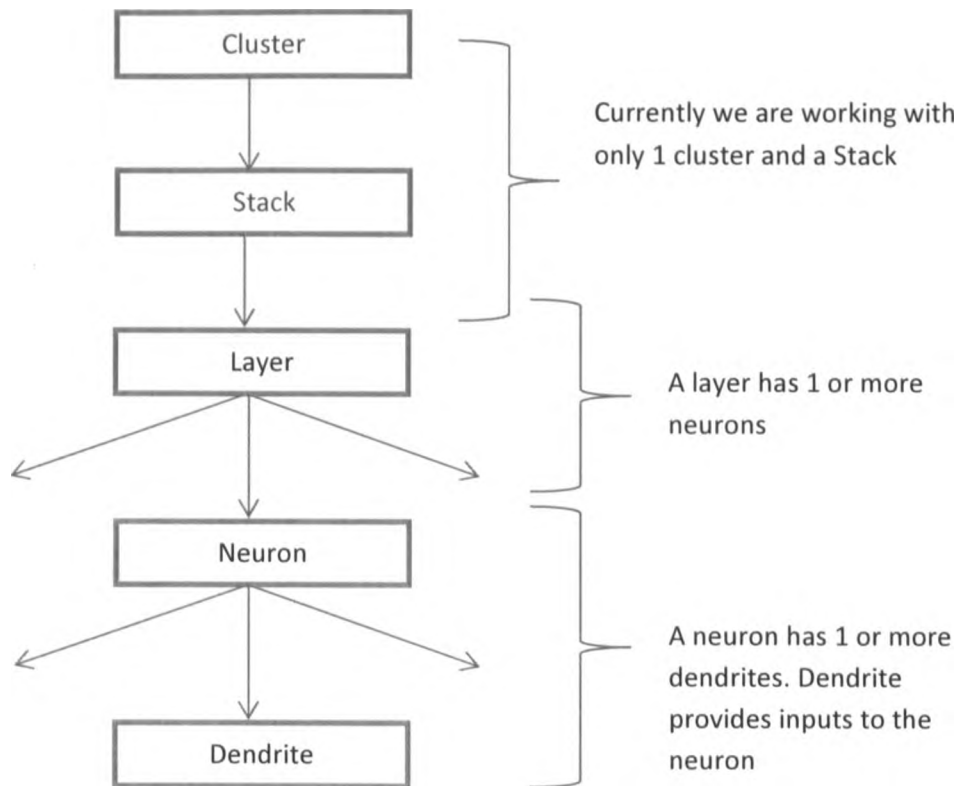


Fig: Neural Network Program Structure

3.9.5. The Organization of the Classes

Brain

Stores all the Neuron in XML file format

Neuron Cluster

Properties

- Name – the name of the cluster
- Brain – the brain the cluster belongs to
- Stacks – the list of stacks in the cluster

Methods

The methods in this class exist in all the classes and are described later in the appendixes

Neuron Stack

Properties

- Name – the name of the stack
- Cluster – the cluster the stack belongs to
- Layers – the list of neuron layers in the stack

Methods

The methods are described later

NeuronLayer

The neuron layers are organized such that they are grouped into stacks, which are then arranged into clusters. This program only works with 1 stack and 1 cluster.

The Neuron Layer class has the following properties:

- Name –stores the name of the layer
- Neuron Stack –is of type neuron stack. This identifies the stack the layer belongs to
- Neurons – this is a collection/list type property. It stores the list of neurons belonging to the layer

Neuron

The Neuron class stores the dendrites that belongs to that neuron. It also stores the axon value(output). It retrieves the dendrite by position. It also adds a new dendrite to the list of the dendrites.

Properties

- Name – the name of the neuron
- Layer – the layer the neuron belongs to
- Dendrites – the list of dendrites which provide inputs to the neuron
- Axon Value – the value of the output from the neuron

Methods

- `NewDendrite(Name, From_Neuron as a neuron)` – adds a dendrite to the neuron, indicating the source neuron
- `NewDendrite(Name, From_Neuron as a string)` – adds a dendrite to the neuron, indicating the path to the source neuron. Differs from the above that this one provides the path to the source from the stack level, while the other one points to a given neuron (the rest of the path information can be retrieved from this neuron)
- `Think` – uses the inputs provided to compute the value of the output. Thinking happens only in the neuron.
- `BackPropagate` – uses the training value to adjust the weights on the dendrites

Dendrite

The Dendrite class handles the weights and implements all the functionality for providing the inputs to the neuron. For a given dendrite it indicates where it is getting its inputs from and which its parent neuron is.

Properties

- `Name` – the name of the dendrite
- `ForNeuron` – the neuron it belongs to
- `FromNeuron` – the neuron the dendrite receives input from
- `FromNeuronPath` – the path of the neuron that provides the input to the dendrite
- `Weight` – the weight value of the dendrite
- `Value` – the input value

Similar Methods for all classes

- `AsXML` – generates a string representing the class elements in XML format i.e. for a neuron layer, all the neurons in the layer, the dendrites and their values and the weights on the dendrites, and the neurons the inputs are coming from.
- `FromXML` – retrieves the neural network values stored in the *brain* and loads them to memory. Rather it extracts the data from xml file and loads in the memory.
- `Think` – for the list of child elements in the given class, the process is propagated to each of the elements. The `Think()` method generates the axon value.

It goes through the whole network and computes the axon value (output) for a particular neuron. Sigmoid function is used to get the axon value.

- Interconnect – this method links up the children in the different layers so that the predecessor and successor neurons are known
- For each class two methods are implemented for retrieving the current element, e.g. for a neuron: *ANeuron(index as an integer)* and *ANeuron(name as a string)* These methods have the following format: *A + Name_of_the_class*
- *New + (Name of the children elements)* – for each class that has a list of children elements, this method adds all the children into the list
- NamePath – returns the position of the given element e.g. neuron in string identifying the path from the highest level in a string format. The path returned uses the name attributes of the classes and separates them using a dot (.). an example of a path is: *'C1.S1.Input.N4'*

3.9.6. Inputs

Outputs were normalized so that their values range from **-1 to 1**. This was done by identifying the range of values the inputs can fall into. For example, the inputs for operators are, 1 for Safaricom, 2 for Zain and 3 for Econet. I used one of the operators to be -1, the second one 2 and the third one 1.

3.9.7. Neuron Training

A training data set is provided. Each training data item consists of a combination of an operator, a date and the total call duration for that given operator.

To train, the values are supplied to the neural network. The data provided is used to compute the duration. Depending on the variation from the expected (i.e. the error), the error is propagated backwards from the output neuron to all the rest as the weights are modified. The prediction process is repeated with all the above steps. This is done until the generated value is equal to or almost the same value as the expected.

3.9.8. Prediction/Thinking

To predict, the neural network is provided with all the inputs required, i.e.:-

- The operator (1-Safaricom, 2-Airtel, 3-Econet)
- The day of the week (as a number from 1 – Sunday, to 7 - Saturday)
- The week of the year (1 to 52)

These inputs are then normalized to make the values range from -1 to 1. This is done by making the centre of the range between the lowest and the highest values zero, the lowest value becomes -1 and the highest becomes 1. Other values are the computed to fit this setup. When the three inputs are ready, they are passed to the neural network. A Sigmoid function is used to determine the output.

3.9.9. Results from Neural Networks

The result below from the Neural Network model confirms the applicability of a neural network model to make predictions of Telecommunication Operators' traffic volumes. The predictions themselves depend on which time period is used for training and testing the model, how the data is fed into the neural system. The more promising results in terms of Forecasting were obtained with relatively short time periods.

Neural Network Demo

Train

Predict

Operator

Zain

10

Train

Call Date	Training Duration	Forecasted Duration	
☐ 1-1-2011	969	-	
☐ 2-1-2011	958	-	
☐ 3-1-2011	832	-	
☐ 4-1-2011	943	-	
☐ 5-1-2011	895	-	
☐ 6-1-2011	762	-	
☐ 7-1-2011	1000	-	
☐ 8-1-2011	667	-	
☐ 9-1-2011	854	-	

Fig. Training module. In the Training module the neural network weights are adjusted to map inputs to outputs.

Actual Observations versus Neural Networks Observations for the				
Actual Observations			Neural network Observations	
Day	Duration		Day	Forecasted Duration
1-Jan-11	969		1-Jan-11	973
2-Jan-11	958		2-Jan-11	-
3-Jan-11	832		3-Jan-11	-
4-Jan-11	943		4-Jan-11	942
5-Jan-11	895		5-Jan-11	-
6-Jan-11	762		6-Jan-11	761
7-Jan-11	1000		7-Jan-11	-
8-Jan-11	667		8-Jan-11	-
9-Jan-11	854		9-Jan-11	-

Neural Network Demo

Train

Predict

Operator

Zain

10

Train

Call Date	Training Duration	Forecasted Duration	
☒ 1-1-2011	969	973.208959019461	
☐ 2-1-2011	958	-	
☐ 3-1-2011	832	-	
☒ 4-1-2011	943	942.956049774968	
☐ 5-1-2011	895	-	
☒ 6-1-2011	762	761.677532695061	
☐ 7-1-2011	1000	-	
☐ 8-1-2011	667	-	
☒ 9-1-2011	854	1499.87741958263	

Fig. Training module. In the Training module the neural network weights are adjusted to map inputs to outputs.

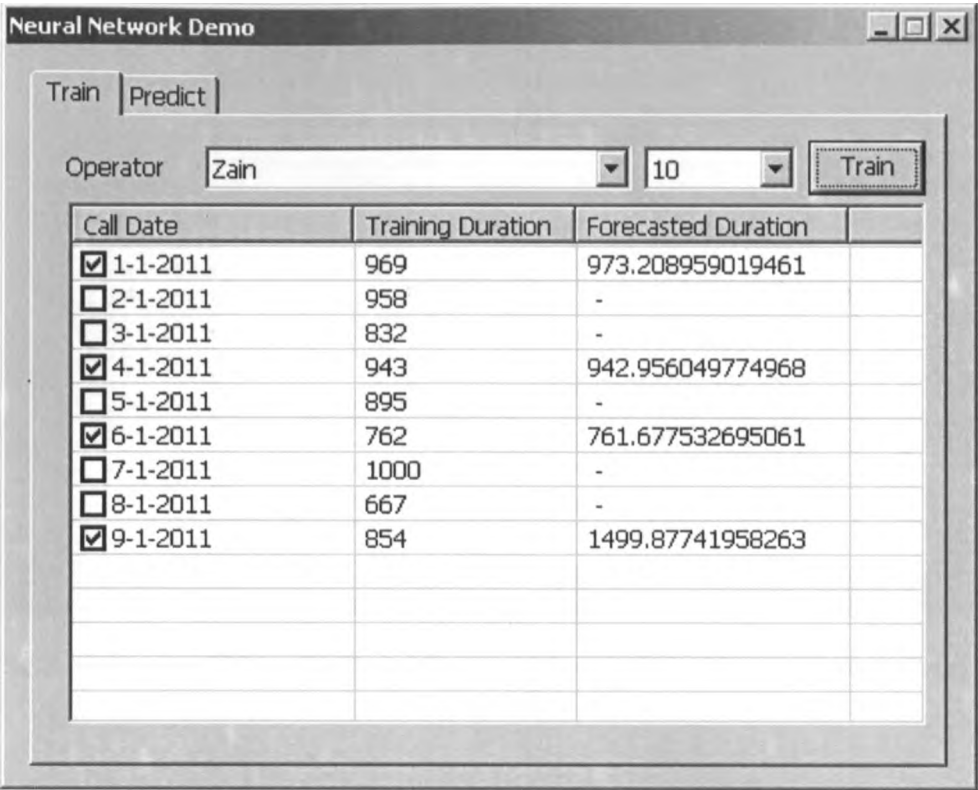


Fig. Prediction module showing Actual Observation versus Neural Networks observations

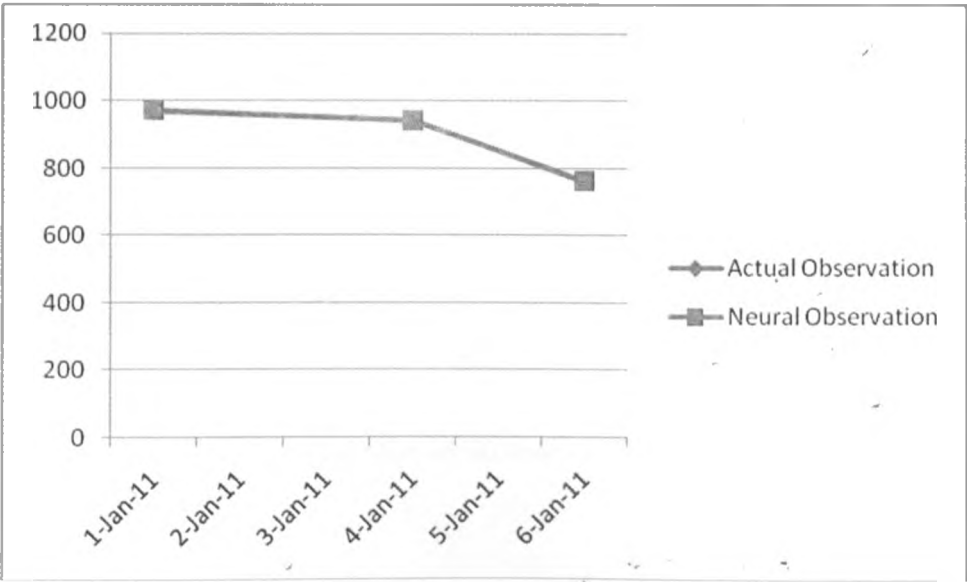


Fig. Graph showing Actual Observation versus Neural Networks observations for Operator Zain

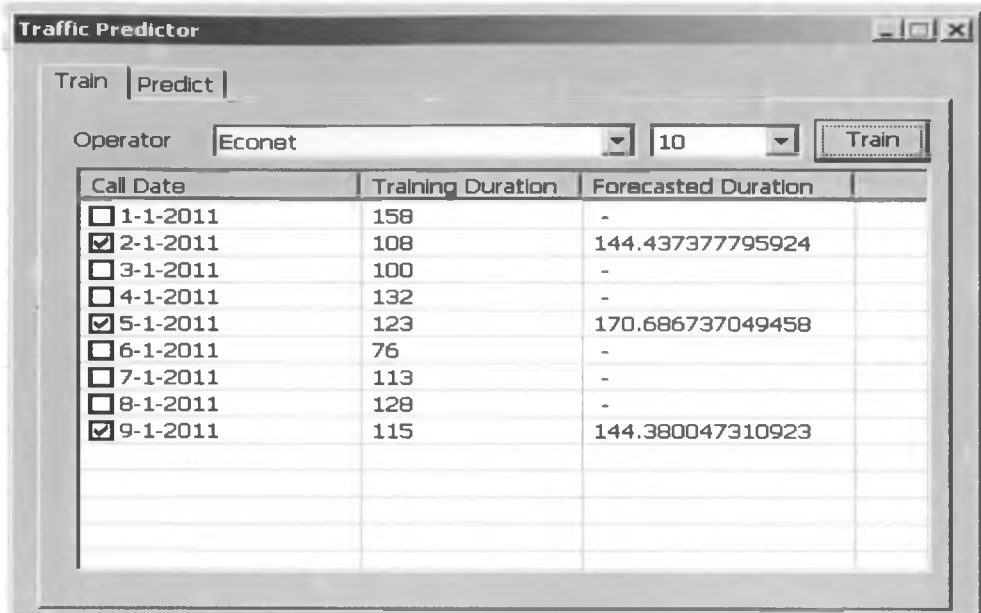


Fig. Prediction module showing Actual Observation versus Neural Networks observations

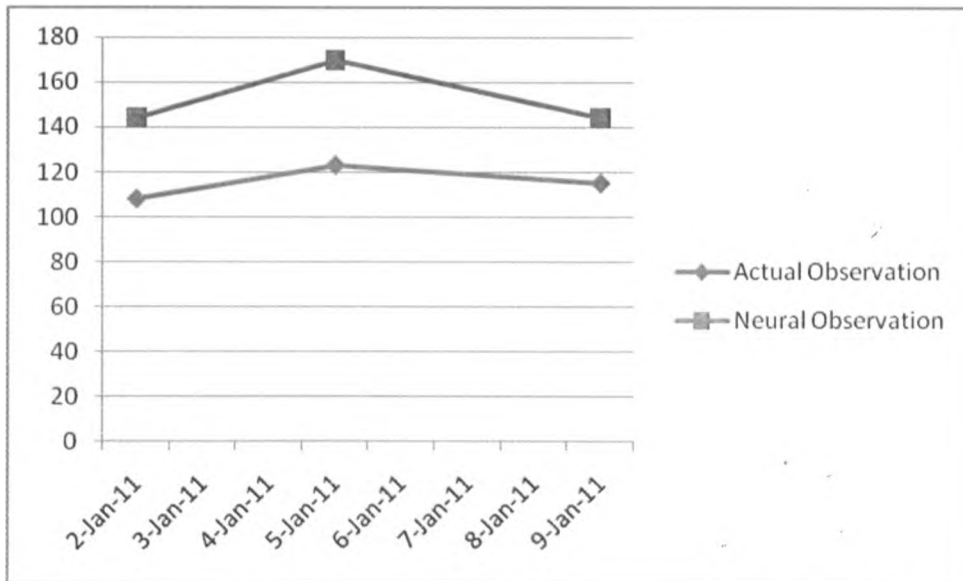


Fig. Graph showing Actual Observation versus Neural Networks observations for Operator Econet

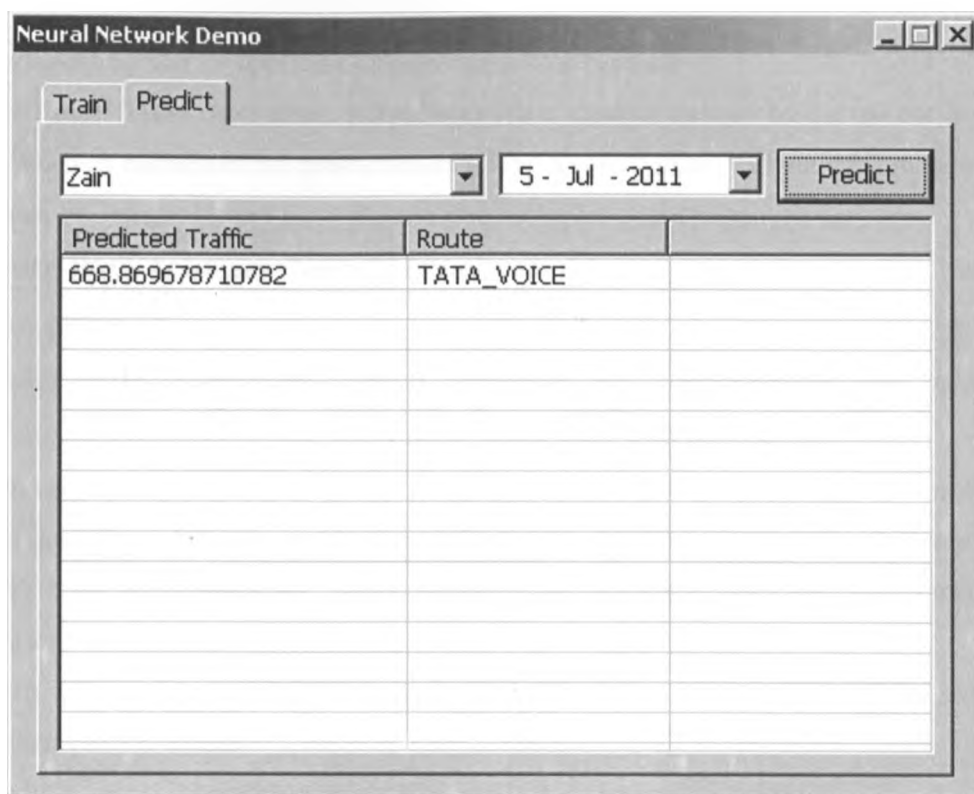


Fig. Prediction module

3.9.10. Storing the Brain State :

1. Creating a neural network based on available inputs and outputs
2. With the training sets, back propagating the outputs to modify the weights
3. After full training, the brain state is stored in an xml file.

3.9.11. Advantages of using Neural Networks

Neural Network Strengths

- i) Neural networks have a remarkable ability to derive meaning from complicated data and can be used to extract patterns and detect trends that are too complex to be noticed by other computer techniques.
- ii) **Adaptive learning:** Neural networks have the ability to learn how to do tasks based on the data given for training or initial experience.
- iii) **Self-Organization:** Neural network can create its own organization or representation of the information it receives during learning time.
- iv) NN can process numeric and categorical data.

v) Appropriate for applications requiring a time element.

vi) Real Time Operation: Neural network computations may be carried out in parallel. When an element of the neural network fails, it can continue without any problem by their parallel nature. Neural networks can also be implemented in parallel hardware

vii) High Accuracy: Neural networks are able to approximate complex non-linear mapping

viii) Noise Tolerance: Neural networks are very flexible with respect to incomplete, missing and noisy data. Work well with noisy data.

ix) Independence from prior assumptions: Neural networks do not make prior assumptions about the distribution of the data, or the form of interactions between factors.

x) Ease of maintenance: Neural networks can be updated with fresh data, making them useful for dynamic environments.

xi) When an element of the neural network fails, it can continue without any problem by their parallel nature.

xii) A neural network learns and does not need to be reprogrammed.

3.9.12. Weaknesses of Neural Network

1. NN Lack explanation capabilities.
2. NN may not provide optimal solutions to problems.
3. The neural network needs training to operate and overtraining can be a problem.
4. Requires high processing time for large neural networks.

3.9.13. Applications of Neural Networks

1. **Neural Networks for clustering:** Neural networks of various kinds can be used for clustering and prototype creation. Neural networks are used in a wide variety of applications. They have been used in all facets of business from detecting the fraudulent use of credit cards and credit risk prediction to increasing the hit rate of targeted mailings. They also have a long history of application in other areas such as the military for the automated driving of an unmanned vehicle at 30 miles per hour on paved roads to biological simulations such as learning the correct pronunciation of English words from written text.

2. Neural Networks for Outlier Analysis: Sometimes clustering is performed not so much to keep records together as to make it easier to see when one record sticks out from the rest. For instance most wine distributors selling inexpensive wine in Missouri and that ship a certain volume of product produce a certain level of profit. There is a cluster of stores that can be formed with these characteristics. One store stands out, however, as producing significantly lower profit.

On closer examination it turns out that the distributor was delivering product to but not collecting payment from one of their customers. A sale on men's suits is being held in all branches of a department store for southern California. All stores with these characteristics have seen at least a 100% jump in revenue since the start of the sale except one. It turns out that this store had, unlike the others, advertised via radio rather than television.

3.9.14. NN System Testing

I tested the NN prototype system as follow: -

- 1) Using Neural Network I tested the system using data from each operator.
- 2) In the training phase the ANN is trained to return a specific output when given a specific input, this is done by continuous training on a set of training data. In the execution phase the ANN returns outputs on the basis of inputs.
- 3) Basically the neural network (in the program) receives a set of 3 inputs (**The Operator, The Day of the week and The Week in the Year**), and attempts to predict the amount of traffic that will be in the network, (**Thus, only one output**). XML has been used to store the initial brain state. We have three input neurons and one output neuron

CHAPTER FOUR

4.1. Results and Discussions

A15 SAFARICOM													
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	OPERATOR	DATE	C_DAY	DUR	MA	Ratio to MA	Unadj Seas Index	Adj Seas Index	Adj Seas Index	Seas Adj Series	Period	Trend Forecast	Seasonally Adj Forecast (Predicted Dur)
2	SAFARICOM	1-Jan-11	7	263					1.113957	292.970592	1	238.2857	265.43994
3	SAFARICOM	2-Jan-11	1	217					1.121591	243.3852215	2	239.8901	269.058
4	SAFARICOM	3-Jan-11	2	342					0.515645	176.3506479	3	241.4945	124.52547
5	SAFARICOM	4-Jan-11	3	249					0.721554	179.6668954	4	243.0989	175.40893
6	SAFARICOM	5-Jan-11	4	197					1.218251	239.9955193	5	244.7033	298.11012
7	SAFARICOM	6-Jan-11	5	241					1.162416	280.1422869	6	246.3077	286.312034
8	SAFARICOM	7-Jan-11	6	138					0.029239	4.034955512	7	247.9121	7.2486539
9	SAFARICOM	8-Jan-11	7	296	235.2857	1.258045	1.25804493	1.113956624	1.113957	329.7311606	8	249.5165	277.95053
10	SAFARICOM	9-Jan-11	1	304	240	1.266667	1.26666667	1.121590882	1.121591	340.9636282	9	251.1209	281.65488
11	SAFARICOM	10-Jan-11	2	147	252.4286	0.582343	0.58234295	0.515645169	0.515645	75.79983988	10	252.7253	130.31656
12	SAFARICOM	11-Jan-11	3	183	224.5714	0.814885	0.8148855	0.721553797	0.721554	132.0443448	11	254.3297	183.51253
13	SAFARICOM	12-Jan-11	4	296	215.1429	1.37583	1.37583001	1.218251367	1.218251	360.6024047	12	255.9341	311.79202
14	SAFARICOM	13-Jan-11	5	301	229.2857	1.312773	1.31277259	1.162416128	1.162416	349.8872546	13	257.5385	299.36686
15	SAFARICOM	14-Jan-11	6	308	237.8571	1.294895	1.29489489	0.029238808	0.029239	9.005552892	14	259.1429	7.577028
16													

Fig A Table showing seasonal indices for Safaricom and forecasted results for Safaricom Mobile for one cycle in the month of January 2011

From the graph it can be noted that the trend forecast can be determined and looking at the pattern it shows a consistent trend. The Seasonally Adjusted Forecast (Predicted Duration) can be achieved by taking the product of the Seasonal index and the trend forecast for that cycle. This continues based on the data within the cycle. This means that taking the period in the cycle we can forecast for the future. This is done for all the operators with the data having all the variables required.

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	Regression Statistics								
4	Multiple R	0.10477457							
5	R Square	0.01097771							
6	Adjusted R Square	-0.071440814							
7	Standard Error	68.30689554							
8	Observations	14							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	1	585.6044	585.6044	0.133195	0.721493			
13	Residual	12	52759.25	4396.604					
14	Total	13	53344.86						
15									
16		Coefficients	Standard Err	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	236.6813187	37.43148	6.323055	3.81E-05	155.1251	318.2375	155.1251	318.2375
18	X Variable 1	1.604395604	4.398104	0.364958	0.721493	-7.97389	11.18268	-7.97389	11.18268
19									
20									
21									
22	RESIDUAL OUTPUT								
23									
24	Observation	Predicted Y	Residuals						
25	1	238.2857143	24.71429						
26	2	239.8901099	-22.8901						
27	3	241.4945055	100.5055						
28	4	243.0989011	5.901099						
29	5	244.7032967	-47.7033						
30	6	246.3076923	-5.30769						
31	7	247.9120879	-109.912						
32	8	249.5164835	46.48352						
33	9	251.1208791	52.87912						
34	10	252.7252747	105.725						

Fig. Regression model Coefficients

Seasonal Forecast (i) = $N2 \times Q2$ - Trend forecast (i) * Adjusted Seasonal Index (i) for i= Period in cycle e.g. i= Monday, Tuesday etc. Each period represent a day in the dataset and all periods have values which is automatically generated based on the day you pick from the system Calendar

Example of Duration Forecast for Safaricom for Friday, 3rd June 2011 (D)

From the system the Period is 154

$Y = MX + C$ Where Y= Forecasted Variable (Dependent).....Forecasted Duration

M= Gradient (Seasonal Change).....Represented in this case by X variable*Period

C= Intercept.....y intercept

$Y = (X\text{-variable} \times \text{Period}) + \text{Intercept}$

$Y = (1.604395604 \times \text{Period}) + (236.6813187)$

$Y = (1.604395604 \times 154) + (236.6813187)$

= **484 Minutes**

Actual Traffic Versas Predicted traffic			
OPERATOR	DATE	Actual Traffic	Predicted Traffic
SAFARICOM	1-Jan-11	263	265.4399497
SAFARICOM	2-Jan-11	217	269.05856
SAFARICOM	3-Jan-11	342	124.5254752
SAFARICOM	4-Jan-11	249	175.4089351
SAFARICOM	5-Jan-11	197	298.1101258
SAFARICOM	6-Jan-11	241	286.3120341
SAFARICOM	7-Jan-11	138	7.248653954
SAFARICOM	8-Jan-11	296	277.9505395
SAFARICOM	9-Jan-11	304	281.6548884
SAFARICOM	10-Jan-11	147	130.3165671
SAFARICOM	11-Jan-11	183	183.5125393
SAFARICOM	12-Jan-11	296	311.7920257
SAFARICOM	13-Jan-11	301	299.3668614

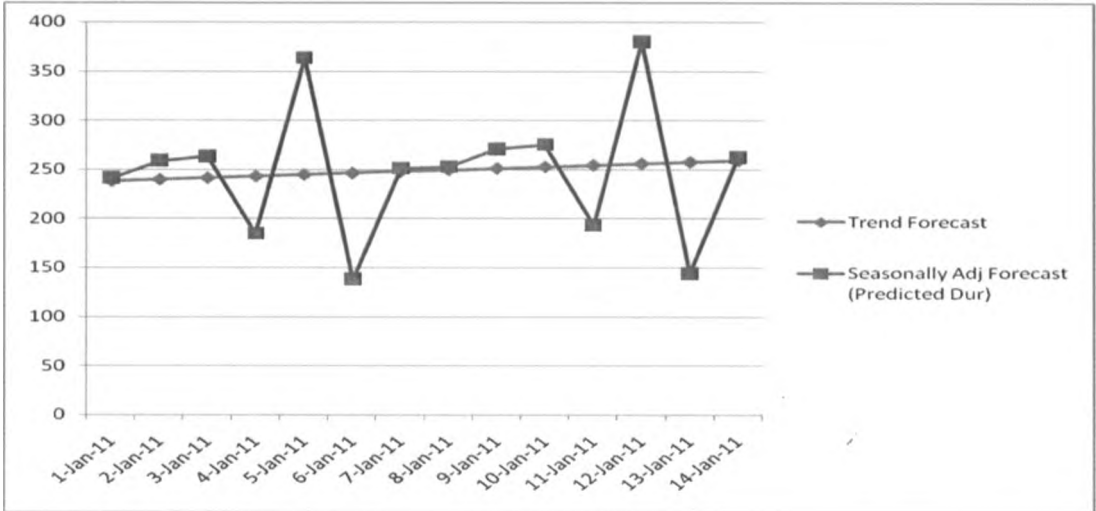


Fig A Trend forecast versus Seasonal Forecast for Safaricom Mobile for one cycle of the month of January

From the graph it can be noted that the trend forecast is consistent though due to seasonal variations prediction is represented as seasonally adjusted Forecast.

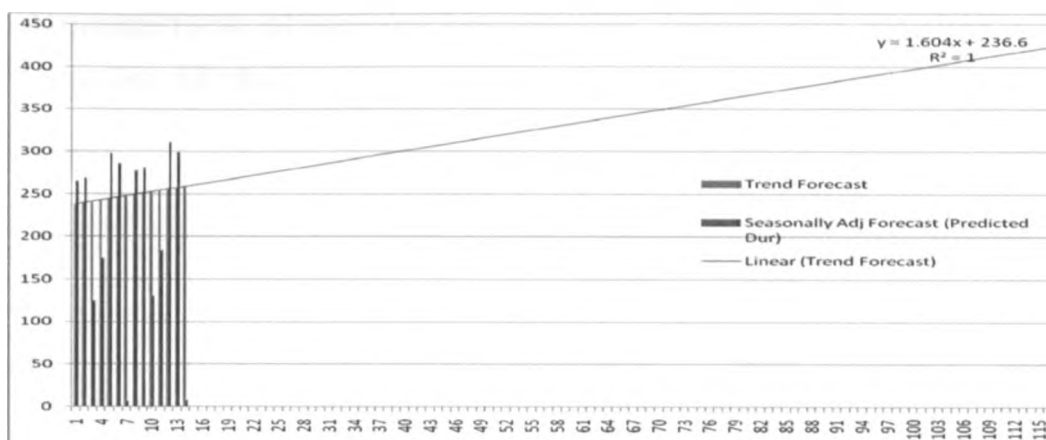


Fig Trend forecast versus Seasonal Forecast for Safaricom Mobile for one cycle of the month of January

4.2. Comparison between Actual Traffic versus Predicted Traffic from Linear Regression

Actual Traffic Versus Predicted traffic			
OPERATOR	DATE	Actual Traffic	Predicted Traffic
SAFARICOM	1-Jan-11	263	265.4399497
SAFARICOM	2-Jan-11	217	269.05856
SAFARICOM	3-Jan-11	342	124.5254752
SAFARICOM	4-Jan-11	249	175.4089351
SAFARICOM	5-Jan-11	197	298.1101258
SAFARICOM	6-Jan-11	241	286.3120341
SAFARICOM	7-Jan-11	138	7.248653954
SAFARICOM	8-Jan-11	296	277.9505395
SAFARICOM	9-Jan-11	304	281.6548884
SAFARICOM	10-Jan-11	147	130.3165671
SAFARICOM	11-Jan-11	183	183.5125393
SAFARICOM	12-Jan-11	296	311.7920257
SAFARICOM	13-Jan-11	301	299.3668614

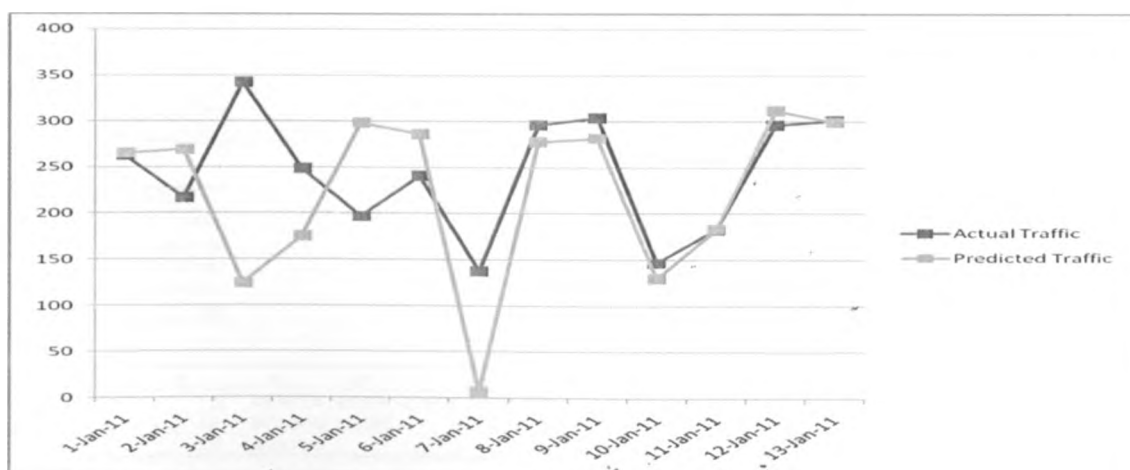


Fig Comparison between Actual Traffic versus Predicted Traffic from Linear Regression with some outliers (Data for one cycle)

4.3. Comparison of two cycles, one for 7 days and another one for 14 days

14-Days cycle

	A	B	C	D	E	F	G	H	I	J	K	L
1	OPERATOR	CALL_DATE	DAY	DUR	MA	Ratio to MA	Unadj Seas Index	Adj seas Index	Seas Adjusted Series	Period	Trend Forecast	Seas Adj Forecast
2	SAFARICOM	1-Jan-11	1	2				0.786141023	13.23921881	1	15.01159379	11.80122969
3	SAFARICOM	2-Jan-11	2	21				0.956534341	18.82703655	2	16.0621279	15.36397694
4	SAFARICOM	3-Jan-11	3	12				1.009434774	13.67364665	3	15.56450648	15.71135408
5	SAFARICOM	4-Jan-11	4	24				0.628067382	3.179900574	4	16.22800171	10.19227855
6	SAFARICOM	5-Jan-11	5	19				1.362795103	9.320860782	5	15.95154536	21.73868791
7	SAFARICOM	6-Jan-11	6	2				0.842789659	1.10266476	6	15.01159379	12.65161602
8	SAFARICOM	7-Jan-11	7	13				1.121734423	18.7893804	7	15.61979775	17.52126482
9	SAFARICOM	8-Jan-11	8	5				1.393742673	11.57470992	8	15.1774676	21.15348426
10	SAFARICOM	9-Jan-11	9	10				1.130508368	1.118583561	9	15.45392394	17.47079034
11	SAFARICOM	10-Jan-11	10	13				0.800839454	0.928962093	10	15.61979775	12.5089503
12	SAFARICOM	11-Jan-11	11	17				1.715756578	16.80673104	11	15.56450648	18.9148681
13	SAFARICOM	12-Jan-11	12	6				0.979190728	3.702051125	12	15.23275887	14.91577624
14	SAFARICOM	13-Jan-11	13	5				0.749407866	1.120439279	13	15.1774676	11.37411361
15	SAFARICOM	14-Jan-11	14	38	12.64285714	0.840169492	1.06003105	1.023557677	33.77746335	14	16.72562313	17.11368888
16	SAFARICOM	15-Jan-11	1	18	13.86268889	0.968085106	0.829493798	0.786141023	10.21983329	15	15.61979775	12.37986378
17	SAFARICOM	16-Jan-11	2	16	13.07142857	1.234043716	1.00928368	0.956534341	15.30454946	16	15.78567156	15.09953693
18	SAFARICOM	17-Jan-11	3	12	13.07142857	0.918082787	1.065101376	1.009434774	12.113211729	17	15.56450648	15.71135408
19	SAFARICOM	18-Jan-11	4	4	11.64285714	0.343558282	0.66270298	0.628067382	2.5112269527	18	15.12217633	9.497748694
20	SAFARICOM	19-Jan-11	5	7	10.78571429	0.649006623	1.437948223	1.362795103	9.539563722	19	15.28880014	20.83447955
21	SAFARICOM	20-Jan-11	6	2	10.78571429	0.185490464	0.889266398	0.842789659	1.685793119	20	15.01159379	12.65161602
22	SAFARICOM	21-Jan-11	7	16	11	1.454545455	1.183593936	1.121734423	17.54775077	21	15.78567156	17.70733118
23	SAFARICOM	22-Jan-11	8	10	11.35714286	0.880503145	1.470602437	1.393742673	13.91742673	22	15.43592394	21.53879388
24	SAFARICOM	23-Jan-11	9	1	10.71428571	0.093333333	1.192851732	1.130508368	1.130508368	23	14.95630252	16.90522318
25	SAFARICOM	24-Jan-11	10	1	11.07142857	0.101449275	0.845002794	0.800839454	0.800839454	24	14.95630252	11.9759715
26	SAFARICOM	24-Jan-11	11	19	11.07142857	1.834482759	1.282273439	1.215256528	21.08987407	25	15.95154536	19.38521988
27	SAFARICOM	26-Jan-11	12	3	10.14285714	0.295774648	1.033189482	0.979190728	2.937572183	26	15.06888506	14.75135414
28	SAFARICOM	27-Jan-11	13	2	9.85714286	0.201408444	0.790733944	0.749407866	1.494855733	27	15.01159379	11.24980647
29	SAFARICOM	28-Jan-11	14	1	7.642857143	0.190841121		1.023557677	1.023557677	28	14.95630252	15.30868827
30	SAFARICOM	29-Jan-11	1	5	7.071428571	0.707070707		0.786141023	3.930705113	29	15.1774676	11.9316299
31	SAFARICOM	30-Jan-11	2	2	6.071428571	0.329411765		0.956534341	1.913068683	30	15.01159379	14.35910498
32	SAFARICOM	31-Jan-11	3	13	6.142857143	2.11627907		1.009434774	13.12265206	31	15.61979775	15.76716701
33	SAFARICOM	1-Feb-11	4	1	5.938571429	0.168674499		0.628067382	0.628067382	32	14.95630252	9.393555787

Fig A graph showing seasonal indices for Safaricom and forecasted results for Safaricom Mobile for 14-days cycle

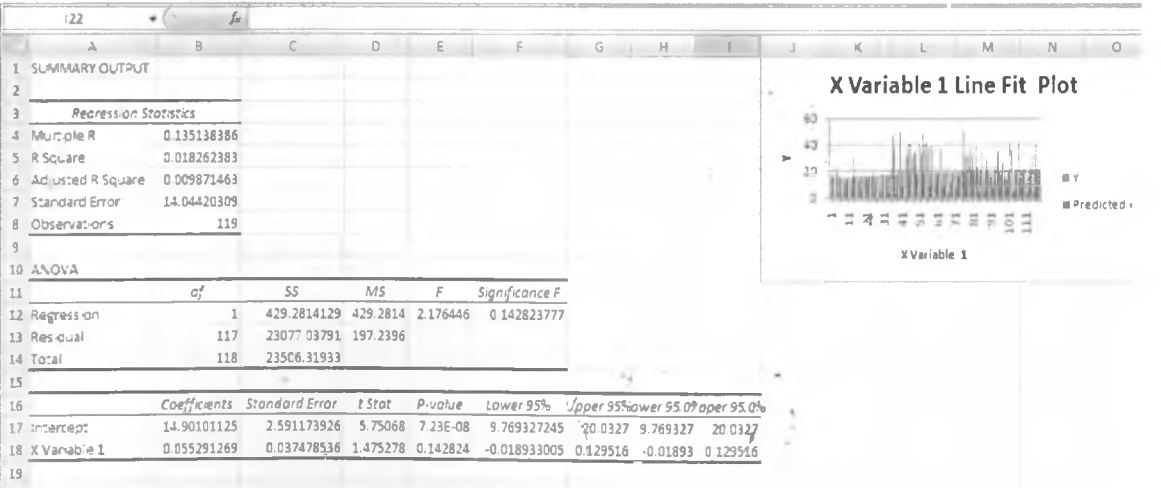


Fig. Regression model Coefficients

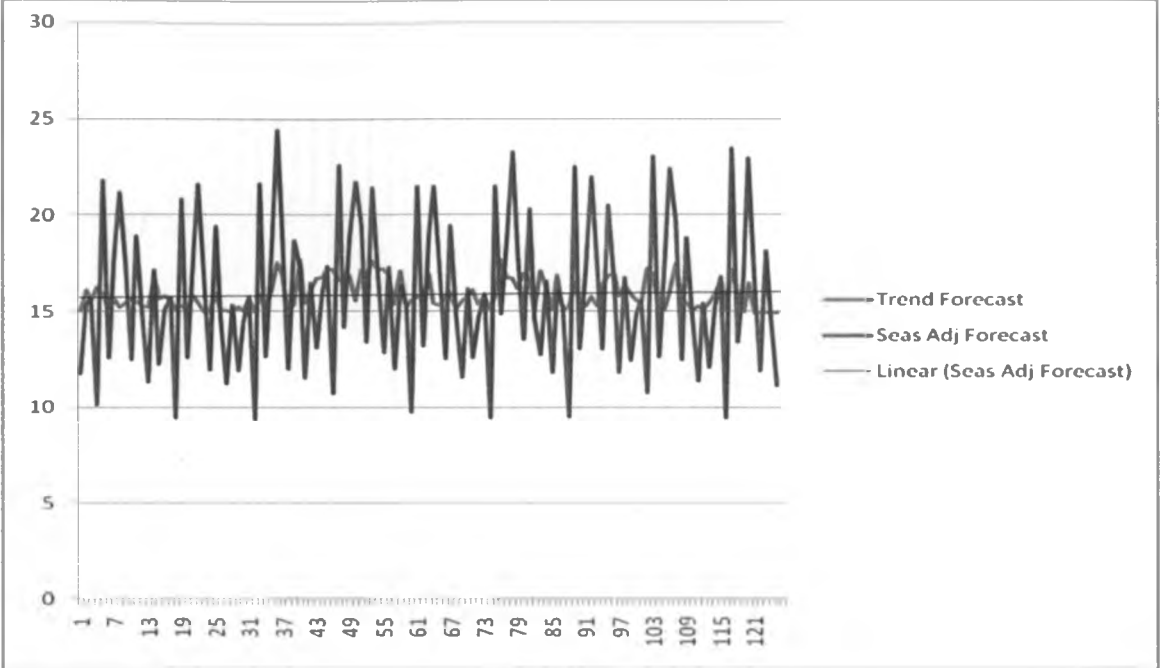


Fig Trend forecast versus Seasonal Forecast for Safaricom Mobile for 14-days cycle

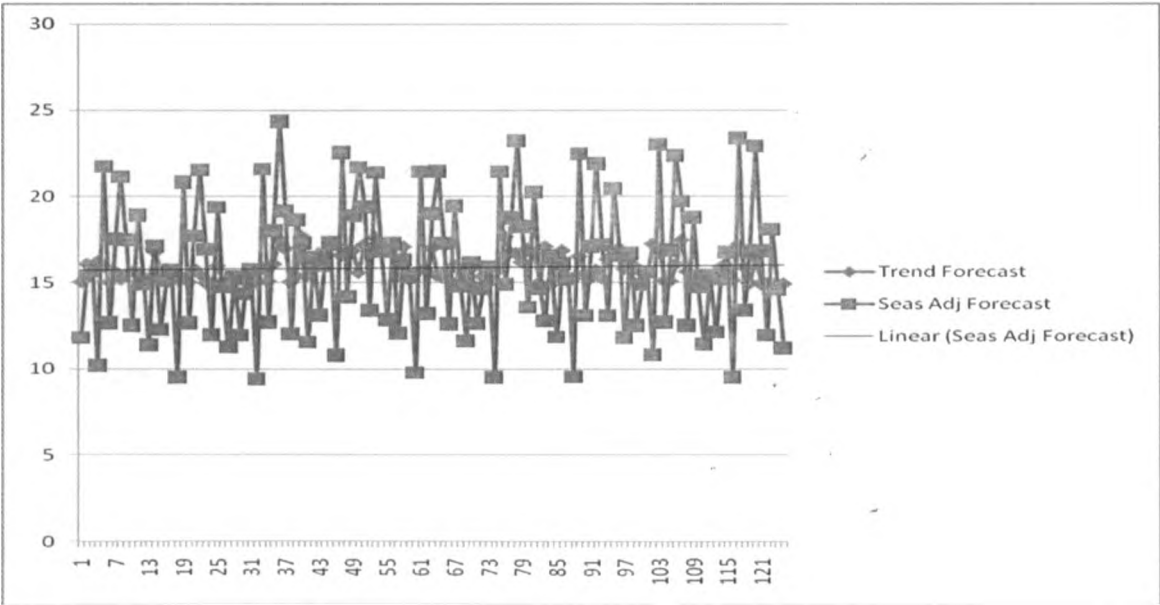


Fig Trend forecast versus Seasonal Forecast for Safaricom Mobile for 14-days cycle

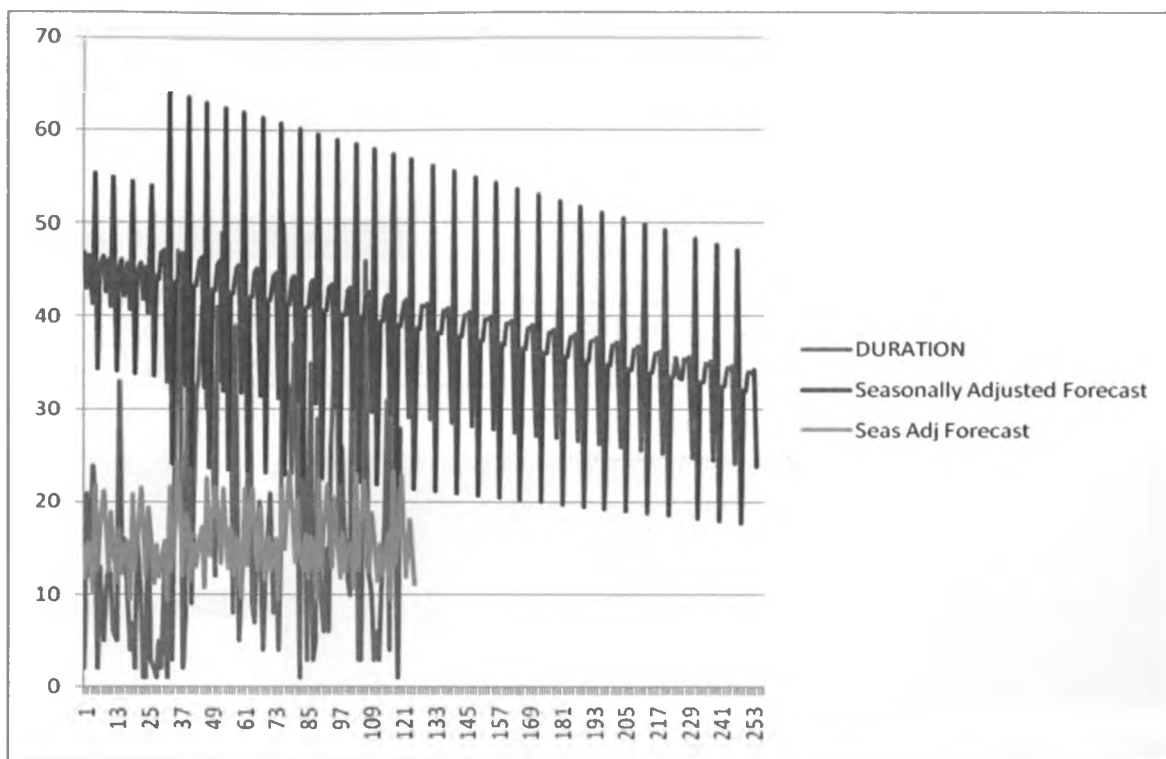


Fig Seasonal Forecast for Safaricom Mobile for 7-days cycle and 14-days cycle. It is based on the moving averages. Every pick is a cycle. The cycle comes after 7 days and 14 days cycle respectively.

4.3.1. Observations

- The bigger the cycle the better the estimate

7-Days Cycle- The curves

As observed in the graph the predictor variables are way off the actual observation. It has a lot of noise and therefore not a good predictor.

14-Days Cycle- The curves

As observed from fig above the 14-days cycle is better than the 7-days cycle. The fitted curve traverses the actual observations curve (the actual and the forecasted lies within the same range - overlaps). Additionally, the cycles are smooth.

R-squared

The squared multiple correlation coefficient, r^2 was used to determine the goodness of fit of the current regression. The bigger the R squared the better the fitness (in a regression model). However the R squared coefficient is also dependent on the number of variables used in the model. In our case only one independent variable is used, the duration.

Seasonal Forecast for days within 3-months dataset as compared to days within 4-months dataset

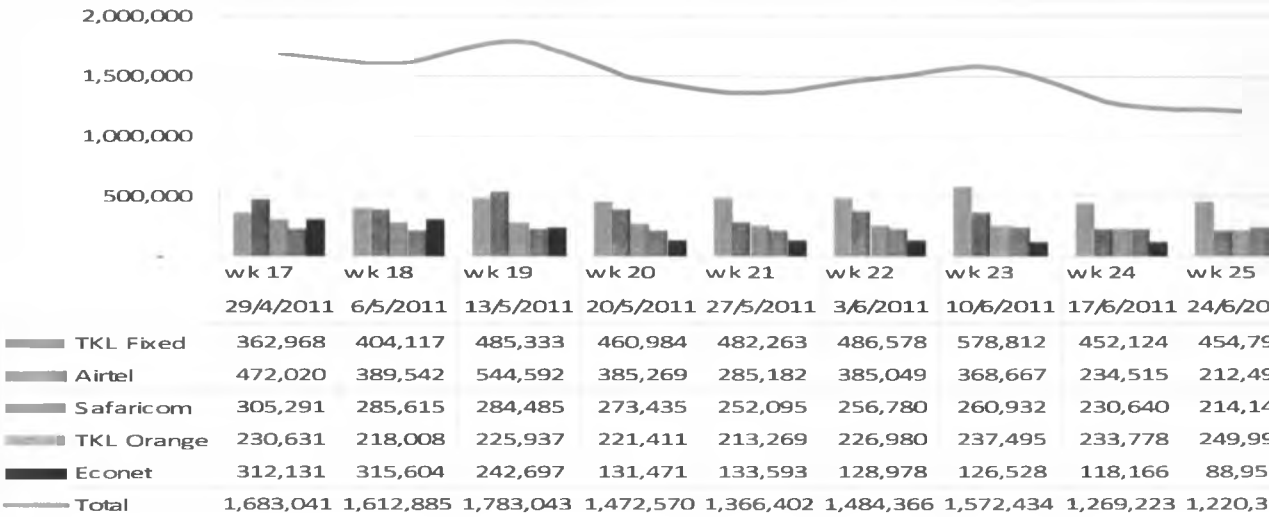
Days within 3-months Dataset		Days within 4-months Dataset	
Day	Forecasted Duration	Day	Forecasted Duration
22nd Jan	21	27th April	23
8th Feb	18	29th April	16

18	SAFARICOM	17-Jan-11	3	12	13.07142657	0.918032787	1.065101376	1.009434774	12.11521729	17	15.56450648	15.71185408
19	SAFARICOM	18-Jan-11	4	4	11.64285714	0.343558282	0.68270298	0.628067382	2.512269527	18	15.12217633	9.497745694
20	SAFARICOM	19-Jan-11	5	7	10.78571429	0.649006623	1.437948223	1.362795103	9.589565722	19	15.28805014	20.83447988
21	SAFARICOM	20-Jan-11	6	2	10.78571429	0.185430464	0.889266398	0.842789659	1.685579319	20	15.01159379	12.65161602
22	SAFARICOM	21-Jan-11	7	16	11	1.454545455	1.183593936	1.121734423	17.94775077	21	15.78567156	17.70739116
23	SAFARICOM	22-Jan-11	8	10	11.14285714	0.888888889	1.437948223	1.881762777	18.91774899	22	15.31579326	21.33679326
24	SAFARICOM	23-Jan-11	9	1	10.71428571	0.093333333	1.92851732	1.190908368	1.150508368	23	14.95630252	16.90822511
25	SAFARICOM	24-Jan-11	10	1	9.857142857	0.101449275	0.845002794	0.800839454	0.800839454	24	14.95630252	11.97739715
26	SAFARICOM	24-Jan-11	11	19	10.85714286	1.834482759	1.282273439	1.215256528	23.68937402	25	15.95154536	19.38521901
35	SAFARICOM	3-Feb-11	6	3	6.714285714	0.446808511		0.842789659	2.528368978	34	15.06688506	12.69821493
36	SAFARICOM	4-Feb-11	7	21	7.071428571	2.969696967		1.121734423	23.55642289	35	16.0621279	18.01744178
37	SAFARICOM	5-Feb-11	8	47	9.714285714	4.838235294		1.393742673	65.50590562	36	17.4997009	24.3900799
38	SAFARICOM	6-Feb-11	9	37	12.28571429	3.011627907		1.130508368	41.82880962	37	16.94678821	19.15848588
39	SAFARICOM	7-Feb-11	10	2	12.35714286	0.161849711		0.800839454	1.601678908	38	15.01159379	12.02187658
40	SAFARICOM	8-Feb-11	11	1	11.14285714	0.091156015		1.215256528		39	15.06688506	11.97739715
41	SAFARICOM	9-Feb-11	12	50	14.92857143	3.349282297		0.979190728	48.95953638	40	17.6655747	17.29756695
42	SAFARICOM	10-Feb-11	13	9	15.42857143	0.583333333		0.749407866	6.744670798	41	15.39863267	11.53985646
113	SAFARICOM	24-Apr-11	4	10	10	1.123		0.70039491	17.11701013	114	15.07024903	15.200514294
116	SAFARICOM	25-Apr-11	3	31	17.5	1.771428571		1.009434774	31.29247799	115	16.61504059	16.77179975
117	SAFARICOM	26-Apr-11	4	4	14.78571429	0.270531401		0.628067382	2.512269527	116	15.12217633	9.497745694
118	SAFARICOM	27-Apr-11	5	41	15.14285714	2.70754717		1.362795103	55.87459923	117	17.16795328	23.39640267
119	SAFARICOM	28-Apr-11	6	18	16.21428571	1.110132159		0.842789659	15.17021387	118	15.89625409	13.39719857
120	SAFARICOM	29-Apr-11	7	1	16.07142857	0.062222222		1.121734423	1.121734423	119	14.95630252	16.77699938
117	SAFARICOM	26-Apr-11	4	4	14.78571429	0.270531401		0.628067382	2.512269527	116	15.12217633	9.497745694
118	SAFARICOM	27-Apr-11	5	41	15.14285714	2.70754717		1.362795103	55.87459923	117	17.16795328	23.39640267
119	SAFARICOM	28-Apr-11	6	18	16.21428571	1.110132159		0.842789659	15.17021387	118	15.89625409	13.39719857
120	SAFARICOM	29-Apr-11	7	1	16.07142857	0.062222222		1.121734423	1.121734423	119	14.95630252	16.77699938
121	SAFARICOM	30-Apr-11	8	28	16.57142857	1.689655172		1.393742673	39.02479484	120	16.44916679	22.92590568
122												
123												

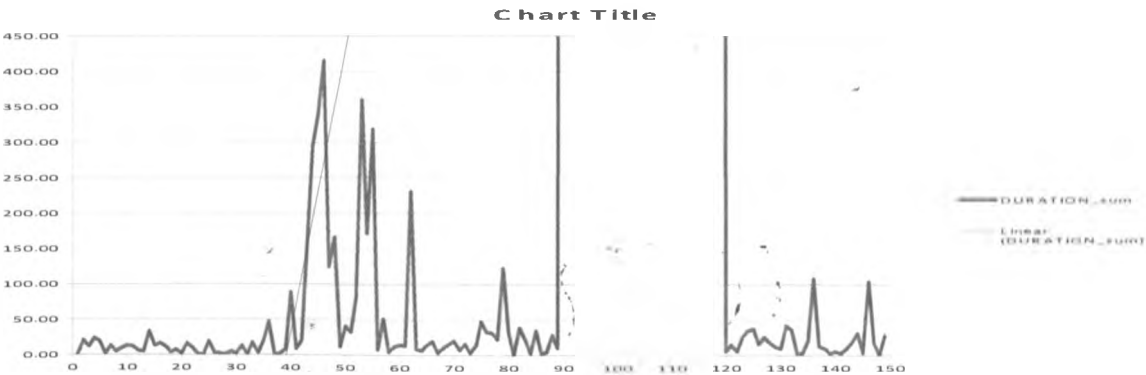
OUTGOING TRAFFIC (ORIGINATING from Kenya Networks):

Weekly Traffic:

Week Ending	29/4/2011	6/5/2011	13/5/2011	20/5/2011	27/5/2011	3/6/2011	10/6/2011	17/6/2011	24/6/2011
Origin	wk 17	wk 18	wk 19	wk 20	wk 21	wk 22	wk 23	wk 24	wk 25
TKL Fixed	362,968	404,117	485,333	460,984	482,263	486,578	578,812	452,124	454,792
Airtel	472,020	389,542	544,592	385,269	285,182	385,049	368,667	234,515	212,492
Safaricom	305,291	285,615	284,485	273,435	252,095	256,780	260,932	230,640	214,141
TKL Orange	230,631	218,008	225,937	221,411	213,269	226,980	237,495	233,778	249,990
Econet	312,131	315,604	242,697	131,471	133,593	128,978	126,528	118,166	88,950
Total	1,683,041	1,612,885	1,783,043	1,472,570	1,366,402	1,484,366	1,572,434	1,269,223	1,220,366



4.3.2. Noisy Data



4.4. Challenges

- Training the Neural Network with a very large data set was a problem such that after sometime, the predicted values increase continuously
- Encountered noisy data when coming up with a linear regression model. A bit of normalization was done.
- Complexity of data received at the switch that necessitated for conversions and transformation.

CHAPTER FIVE

Conclusions and Recommendations

5.1. Conclusions

From the research on Telecommunication Forecasting Traffic model, the study concluded with the following comments: -

The study has presented an opportunity in the world of research on Telecommunication Traffic forecasting using Linear Regression and Neural Networks. This research study has led to a good regression model that has enabled for a web-based system to be built. Both Linear Regression and Neural Networks are suitable for Telecommunication Forecasting.

The study made use of three local operators for traffic originating from Kenya. Using these forecasting techniques it is possible to include many operators to make forecasts that can be used in various decision-making forums. The use of a prediction system has been realized from this study where telecommunication traffic volumes of the next months were predicted. On comparison with the actual observations, the errors were minimal.

In this research study the approach used to predict the telecommunication traffic volumes was to use historical data and experience to extrapolate forward. Results have shown that it is possible to predict the future traffic volumes. The study has revealed some advantages of Linear Regression over Neural Networks. Neural Networks is able to extract hidden information from the historical data. The forecaster requires some telecommunication knowledge in order to identify the correct model and gather the correct independent variables.

The advantage of Linear Regression is that it implements a statistical model that, when relationships between the independent variables and the dependent variable are almost linear, it shows optimal results.

5.2. Recommendations

- The scope within which the research was carried out had extracted just a portion of the data from the Telecommunication Switch (data for four months: Jan-April).
- There is need for telecommunication operators to utilize forecasting models in order to at least realize the expectation of future traffic.
- The forecasting models have not been fully utilized by Telecommunication companies. The models can be used to improve the forecasting of traffic and resources harmonization in the Telecommunication industries.

5.3. Future Research

- There is need for continuous research in this field to come up with more models of prediction.
- The system that was developed utilized a world wide web on a windows environment that handled the user interface. Future work can improve this methodology to make use of IDE in cloud computing.
- Subjecting the results to more statistical analysis can enhance the comparative analysis discussed in this study.

Research Questions answered

1. How does the traffic volume affect the level of bandwidth? From the module on Bandwidth Optimization it's found that the more the traffic volumes the more occupancy of bandwidth required at the switch.
2. To what extent can telecommunications companies predict operators' traffic volumes and balance the bandwidth based on operator giving maximum returns to the company? From the module on prediction it's possible to predict operator's traffic volumes and balance the bandwidth based on operator giving maximum returns to the company.
3. Is it possible for Telecommunication companies to yield maximum returns by applying prediction? It's possible since forecasting will ease decision making.

Glossary:

Traffic: Telephone calls, or any requests for service over a telecommunication switch.

Trunk: The logical connection of the switch to the physical line over which inter-switch traffic arrives.

Call Processing: The switching system application that connects telephone calls from Calling party to Called Party.

Control: Changes to normal call processing instituted during congestion or overload to allow the system to either handle additional traffic, or to protect itself from insanity.

Congestion: When the system receives more requests for service than it has peripheral resources to handle them.

Network Management: Capabilities used to monitor and control, in real time, a network of telecommunications resources.

Overload: When the switch has more work for the processor than it can handle in the available resources.

REFERENCES:

- [1.] Fausett, Laurene (1994). *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*, Prentice-Hall, New Jersey, USA.
- [2.] Villiers J. and Barnard E. (1993). Backpropagation neural nets with one and two hidden layers. *IEEE Transactions on Neural Networks*, 4(1), 136-141.
- [3.] Zurada J.M. (1992). "An introduction to artificial neural networks systems", St. Paul: West Publishing.
- [4.] Long J. S., (1997), *Regression Models for Categorical and Limited Dependent Variables*, SAGE, Thousand Oaks.
- [5.] Edwards, T., Tansley, D, S. W., Davey, N., Frank, R. J.(1997). Traffic Trends Analysis using Neural Networks. *Proceedings of the International Workshop on Applications of Neural Networks to Telecommunications 3*. pp. 157-164.
- [6.] Gershenfeld N. A. and A. S. Weigend, *The Future of Time Series*. In *Time Series Prediction: Forecasting the Future and Understanding the Past*, Gershenfeld A. N., and A. S. Weigen, eds, pp 1-70. 1993.
- [7.] Breiman, Friedman, Olshen and Stone (1984). "Classification and Regression Trees"
- [8.] Carl W. (2003). "2nd Serving of Percentile Billing", Article Issue No.14
- [9.] CAVE M. (1999). "Interconnection and the Internet: Competition and regulation issues at local access and backbone levels", OECD, Working Party on Telecommunication and Information Services
- [10.] CAWLEY R. (2001). "Traffic Exchange: Global Networks", paper presented to the OECD Workshop on Internet Traffic Exchange, June 7-8, Berlin.
- [11.] CAVE M. (1999). "Interconnection and the Internet: Competition and regulation issues at local access and backbone levels", OECD, Working Party on Telecommunication and Information Services
- [12.] Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks
- [13.] Larose D.T. (2003). *Discovering Knowledge in Data: An Introduction to Data Mining* European Institutes for Research and Strategic Studies in Telecommunications (EURESCOM) (2001): "Next Generation Networks: The Service Offering Standpoint, Technical Information
- [14.] James Kolsky 2001. "Statistics and Data Mining in the analysis of massive data sets"

- [15.] Nisbet and Watt (1980). Research Studies, Routledge Publishers
- [16.] Gilbert Kandie (2003). A Decision Support System for Forecasting Exchange Rates and Investment Optimization, School of Computing and Informatics, University of Nairobi
- [17.] Salant, Dillman D. A. (1994). How to conduct your own survey, John Wiley & Sons Inc.
- [18.] SPETA J.B. (2002). "A common carrier approach to Data interconnection", in: Federal Communications Law Journal, vol. 54, no. 2; pp. 225-279.
- [19.] Steve Cerruti and Carl Wright (2003). "A 2nd Serving of Percentile Billing" ISSN: 1532-1886

APPENDICES

6.1. APPENDIX A

6.1.1. Prediction Algorithms Code:

Name :	Charles Maina
Program :	Forecasting Model
Purpose :	The Telecommunication Forecasting Model predict Traffic Volumes from various Operators based on the period indicated.
Date :	25th June 2011

```
Imports Oracle.DataAccess.Client
Imports Oracle.DataAccess.Types
Imports CrystalDecisions.Shared
Imports CrystalDecisions.CrystalReports.Engine
Imports System.data
```

```
Partial Class billing_frmPredict
    Inherits System.Web.UI.Page
```

```
Public x As Integer
Public h,h2,h3 As Double
```

```
Protected Sub cmdQueryPredVals_Click(ByVal sender As Object, ByVal e As
System.Web.UI.ImageClickEventArgs) Handles cmdQueryPredVals.Click
```

```
Dim reporting_date As Date
Dim td As String
```

```
reporting_date = DateValue(DateTimePicker1.Text)
td = Weekday(reporting_date)
```

```
td = DateDiff("d", "1/1/2011", reporting_date)
```

```
x = (Convert.ToInt32(td) + 1)
```

```
txtPeriod.Text = x
```

```
'-----hashtable values for Zain-----
Dim value3 As New Hashtable
```

```
value.Add("1", " & SeasonalIndex &") 'Seasonal Indices for Mapped Days---Sun to
Saturday
```

```
txtAdjSeasIndex1.Text = value3.Item(ddDays.SelectedValue)
h = value.Item(ddDays.SelectedValue)
```

```
Dim conn As OracleConnection = New OracleConnection("User ID =system;Password
=maina;Data Source = localdb;Persist Security Info = False; Pooling=false")
conn.Open()
```

```
Dim dr As OracleDataReader
```

```
Dim str_zain As String = "select operator,route from tbl_zain_load2 where c_day =
'" & ddDays.SelectedValue & "'"
```

```
Dim mycmd As New OracleCommand(str_zain, conn)
dr = mycmd.ExecuteReader
```

```
If dr.Read = True Then
```

```
txtOperatorZain.Text = dr("operator")
txttrend_forecast1.Text = (x * 116) / - 2061.54 'dr("trend_forecast")
txtDuration1.Text = Math.Round(h3 * ((x * 121) - 2061.54), 2)
txtroute1.Text = dr("route")
```

```
End If
```

```
Session("Route") = txtroute3.Text
```

```
Response.Redirect("frmGenerateBillingData.aspx")
```


6.2. APPENDIX B

6.2.1. Bandwidth Optimization Code:

Name :	Charles Maina
Program :	Bandwidth Optimization Module
Purpose :	The Telecommunication Bandwidth Optimization Module for allocating Bandwidth to operator's Routes giving maximum returns to the company

```
Partial Class frmGenerateBillingData
    Inherits System.Web.UI.Page
    Dim strReportPath As String
    Dim strReport As String

    Public Sub MyTrafficData()
        Dim conn As OracleConnection = New OracleConnection("User ID =system;Password
=maina;Data Source = localdb;Persist Security Info = False; Pooling=false")
        conn.Open()

        Dim MyDA As New OracleDataAdapter
        Dim cmd As New OracleCommand

        conn = New Oracle.DataAccess.Client.OracleConnection(ConnClass.conn_billing)
        cmd = New Oracle.DataAccess.Client.OracleCommand("", conn)
        MyDA = New Oracle.DataAccess.Client.OracleDataAdapter
        MyDA.SelectCommand = cmd
        cmd.CommandType = Data.CommandType.StoredProcedure

        Dim ds As New Data.DataSet

        ds.Tables.Add("tbl_traffic_summary")
        'Load Traffic
        With cmd
            .Parameters.Clear()
            .CommandText = "pkg_billtraffic.get_trafficdata"
            .Parameters.Add(New OracleParameter("pcur", OracleDbType.RefCursor)).Direction
= Data.ParameterDirection.Output
        End With

        MyDA.Fill(ds.Tables("tbl_traffic_summary"))
        Session("datagrid") = ds.Tables("tbl_traffic_summary")
        ViewBillDataSheetGrid.DataSource = ds.Tables(0)
        ViewBillDataSheetGrid.DataBind()

    End Sub

    Dim dr As OracleDataReader

    Dim str As String = "select operator from vw_loadbalance where revenue in (select
max(revenue) from vw_loadbalance)"

    Dim mycmd As New OracleCommand(str, conn)
    dr = mycmd.ExecuteReader

    Dim ds As New DataSet
    Dim adp As New OracleDataAdapter
    Dim dset As New Data.DataSet
    With mycmd
        .Parameters.Clear()
        .CommandText = str
    End With

    adp.SelectCommand = mycmd
    ' Fills what is in the dataset
    adp.Fill(dset)
    Dim rt As String = dset.Tables(0).Rows(0).Item("operator").ToString
```

```

lblPriorityRoute.Text = rt + " gives the highest revenue. Please give priority for
additional Bandwidth "
End Sub

Try
    Session("Route") = ViewBillDataSheetGrid.SelectedRow.Cells(3).Text
    Session("Operator") = ViewBillDataSheetGrid.SelectedRow.Cells(2).Text
    txtRoutedat.Text = Session("Route")
    Session("bandwidth") = ViewBillDataSheetGrid.SelectedRow.Cells(8).Text
    txtBandwidth_ge.Text = Session("bandwidth")
Catch ex As Exception

End Try

conn.Close()
conn.Dispose()

End Sub

Protected Sub cmdViewTraffic_Click(ByVal sender As Object, ByVal e As System.EventArgs)
Handles cmdViewTraffic.Click
    MyDataGridData()
End Sub

Protected Sub txtRoutedat_TextChanged(ByVal sender As Object, ByVal e As
System.EventArgs) Handles txtRoutedat.TextChanged
    Session("Route") = ViewBillDataSheetGrid.SelectedRow.Cells(3).Text
    txtRoutedat.Text = Session("Route")
End Sub

Protected Sub txtBandwidth_TextChanged(ByVal sender As Object, ByVal e As
System.EventArgs) Handles txtBandwidth_ge.TextChanged
End Sub

Protected Sub cmdOptimize_Click(ByVal sender As Object, ByVal e As System.EventArgs)
Handles cmdOptimize.Click
    updatebandwidth()
End Sub

Public Sub updatebandwidth()

    Dim conn As OracleConnection = New OracleConnection("User ID =system;Password
=maina;Data Source = localdb;Persist Security Info = False; Pooling=false")

    'dataset holds the data in the database, OracleDataAdapter holds the data
session,OracleDataReader reads what is in the dataset
    Dim MyDA As New OracleDataAdapter
    Dim cmd As New OracleCommand

    conn = New Oracle.DataAccess.Client.OracleConnection(ConnClass.conn_billing)
    cmd = New Oracle.DataAccess.Client.OracleCommand("", conn)
    MyDA = New Oracle.DataAccess.Client.OracleDataAdapter
    MyDA.SelectCommand = cmd
    cmd.CommandType = Data.CommandType.StoredProcedure

    conn.Open()

    Dim ds As New Data.DataSet
    ds.Tables.Add("tbl_traffic_summary")
    'Load Traffic
    With cmd
        .Parameters.Clear()
        .CommandText = "pkg_billtraffic.update_bwidth"
        .Parameters.Add(New OracleParameter("pcurrbwidth", Val(txtallocatedbw.Text) +
Val(txtBandwidth_ge.Text)))
        .Parameters.Add(New OracleParameter("proute", txtRoutedat.Text))
        .ExecuteNonQuery()
    End With
    lblresponse.Text = "You have allocated " & txtallocatedbw.Text & " Els of Bandwidth to "
    & txtRoutedat.Text & ". Your Bandwidth is now upgraded"
End Sub

```

6.3. APPENDIX C

6.3.1. Neural Network: Code

Name : Charles Maina

Program : Training Module

Purpose : Loading and training data for the Telecommunication operators

```
'load the data from the excel spreadsheet
xlApp = New Excel.Application
xlWorkBook = xlApp.Workbooks.Open("D:\Projects\Charles\Data\Data Inputs.xls")
'Using the selected records, this method trains the neural network
PrivateSub TrainTraffic()
Dim S As NeuronStack
Dim Value AsDouble, i AsInteger
Dim C As Control, Name AsString
    S = Brain.ACluster(1).AStack(1)
Dim counter AsInteger = 0
Dim theOperator AsDouble = 0.0
Dim weekDay AsDouble = 0.0
Dim weekYear AsDouble = 0.0
'Set up the input variables
'Get the operator to predict traffic for
If cboOperators.SelectedIndex = 0 Then'Safaricom selected
    theOperator = -1
ElseIf cboOperators.SelectedIndex = 1 Then'Zain selected
    theOperator = 0
ElseIf cboOperators.SelectedIndex = 2 Then'Econet selected
    theOperator = 1
EndIf
    predictedValue = 0
For j AsInteger = 0 To lvwData.Items.Count - 1
If lvwData.Items(j).Checked = TrueThen
    Name = "N1"
    predictedValue = 0
Do While Not Math.Round(Convert.ToDouble(predictedValue), 0) >=
Math.Round(Convert.ToDouble(lvwData.Items(j).SubItems(1).Text), 0)
Value = GetOutputValue(Convert.ToDouble(lvwData.Items(j).SubItems(1).Text))
S.ALayer("Output").ANeuron(Name).BackPropagate(Value)weekDay =DatePart(DateInterval.
Weekday,Convert.ToDateTime(lvwData.Items(j).Text))
If weekDay = 1 Then
    weekDay = -(1)
ElseIf weekDay = 2 Then
    weekDay = -(0.67)
ElseIf weekDay = 3 Then
    weekDay = -(0.33)
ElseIf weekDay = 4 Then
    weekDay = (0)
ElseIf weekDay = 5 Then
    weekDay = (0.33)
ElseIf weekDay = 6 Then
    weekDay = (0.67)
ElseIf weekDay = 7 Then
    weekDay = (1)
EndIf
weekYear = DatePart(DateInterval.WeekOfYear, Convert.ToDateTime(lvwData.Items(j).Text))
weekYear = GetWeekYearValue(weekYear)
'Predict
    PredictTraffic(theOperator, weekDay, weekYear)
    lvwData.Items(j).SubItems(2).Text = predictedValue
    DumpBrainState()
Loop
    counter += 1
    lblCount.Text = "Training With Data Set "& counter & " Of "&
GetSelectedDataSets()
EndIf
Next
```

6.4. APPENDIX D

6.4.1. Predicting the Neural Network:

Name :	Charles Maina
Program :	Predicting Module
Purpose :	Predicting using a given set of inputs

```
'Predict using a given set of inputs
PublicSub Think()
Dim Sum AsSingle, D As Dendrite

If Dendrites.Count = 0 ThenReturn

ForEach D In Dendrites
IfNot IsNothing(D.FromNeuron) Then
    D.Value = D.FromNeuron.AxonValue
EndIf
    Sum += D.Value * D.Weight
Next
    AxonValue = -2 / (1 + Math.Exp(2 * Sum)) + 1
EndSub

PublicSub BackPropagate(ByVal DesiredValue AsSingle)
'Dim LocalGradient As Single, Adjustment As Single
Dim ErrorTerm AsSingle
Dim D As Dendrite
Const LearningRate AsSingle = 0.01

    ErrorTerm = DesiredValue - AxonValue

ForEach D In Dendrites
    D.Weight += D.Value * ErrorTerm * LearningRate
Next
EndSub

PublicSub AutoTrain()

Dim T As TrainingCase, Td As TrainingCaseDendrite
Dim ErrorTerm AsSingle
Const LearningRate AsSingle = 0.01

'Learn from each training case in this one cycle
ForEach T InMe.TrainingCases

'Preset my dendrite's source axons' values to reflect
ForEach Td In T.Dendrites
    Td.ForDendrite.FromNeuron.AxonValue = Td.Value
Next

'Given the training inputs, generate an output
    Think()
'Morph toward the desired output
    ErrorTerm = T.AxonValue - Me.AxonValue
ForEach Td In T.Dendrites
    Td.ForDendrite.Weight += Td.Value * ErrorTerm * LearningRate
Next

Next
EndSub

EndClass
```



UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

Polyepitope design optimization system: A web-based application prototype.

By

MAINGI, JULIUS KAIRITHIA

P56/P/8144/03

SUPERVISOR

LAWRENCE MUCHEMI

March 2012

Submitted in partial fulfilment of the requirements of the Master of Science in Information Systems

Declaration

This project, as presented in this report is my original work and has not been presented for any other University award

Signed: 

Date: 9th May 2012

Name: JULIUS KAIRITHIA MAINGI

Reg/No: P56/P/8144/03

This project has been submitted as part fulfilment for Master of Science in Information systems of the University of Nairobi with my approval as the University supervisor

Signed: 

Date: 9/05/2012

LAWRENCE M.GITHIARI

Project Supervisor

School of Computing and Informatics

University of Nairobi

Dedication

This project is dedicated to my wife Mrs Eunice M. Kairithia

Acknowledgment

First, I thank the Almighty God for the strength, grace and good health throughout this research project and during my studies at the School of Computing and Informatics.

Secondly, my sincere gratitude goes to my supervisor Mr. Lawrence Muchemi who has been very supportive and helpful in providing academic guidance throughout the entire project. I appreciate the panelists of panel 2 September 2011 group.

The School of Computing and Informatics for the knowledge and skills imparted to me in course of my studies in this school.

I would also like to appreciate my family members. I acknowledge the support of my friends especially Dr Moses Kindiki, Dr Muriithi Kigunda, Hon Dr Kilemi Mwiria, Mr Henry Mwenda, Mr Daniel Muriuki, Mr Muriuki Tarichia, and Mr. Geoffrey Muriira Karau.

Abstract

Immunogenic epitopes in a polyepitope construct may induce immunity against multiple antigenic targets only if epitopes are correctly processed and presented. Applications that combine multiple supertypes and integrate variables that measure the quality of polyepitopes can be used to make better polyepitopes yet there are no publicly accessible tools. In this research project, the researcher developed a web-based polyepitope optimization prototype. The prototype runs on 3-tier application architecture: MYSQL was used for the database tier, PHP for the application tier that runs on Apache HTTP server and the presentation layer was implemented on a web browser. The researcher collected epitopes data and information from online databases; the Immune Epitope Database (IEDB), HIV molecular immunological database and National Center for Biotechnology Information (NCBI). The epitopes were analyzed using Microsoft Excel 2003 and uploaded onto a local database. Polyepitopes were generated, optimized and tested on an online web server NetChop 3.1 which was used to validate one of the polyepitope quality measurement variables- proteasome cleavage predictions. The results of data analysis indicate that the immune system recognize epitopes in clusters and the main clusters are 9-mers(9 amino acid long), 15mers,20mers and 25mers. The polyepitope optimization results indicate that polyepitope optimization algorithms that integrate proteasome cleavage prediction, transporter associated with antigen processing (TAP) binding prediction and Major Histocompatibility Complex (MHC) binding predictions can be used to develop better polyepitopes which can be used to make better vaccines within a shorter time and at a lower cost.

Key words: polyepitope, epitopes, optimization, mer, TAP, proteasome, MHC, prototype.

TABLE OF CONTENTS

Declaration	ii
Dedication	iii
Acknowledgment	iv
Abstract	v
List of tables	ix
List of abbreviations	x
CHAPTER ONE: INTRODUCTION	1
1.0 Introduction	1
1.1 Background of the study	1
1.3 Statement of the problem	4
1.4 Research Objectives	4
1.4 Significance of the study	4
1.5 Definition of terms	5
CHAPTER TWO: LITERATURE REVIEW AND THEORY	6
2.0 Introduction	6
2.1 The cell	6
2.1.1 The cell Surface	7
2.1.2 The amino acids	8
2.1.3 Protein synthesis	8
2.2 The human immune system	9
2.2.1 MHC Class I Pathway	10
2.2.2 MHC Class II Pathway	10
2.3 Vaccine design and development	11
2.4 Types of vaccines	11
2.5 Polyepitope construct	11
2.6 Approaches to poly epitope optimization	12
2.6.1 Quantitative matrix (QM)	12
2.6.2 Quadratic programming	13
2.6.3 Integer linear programming	13
2.6.4 Monte Carlo metropolis simulations algorithm	13
2.7 Epitope based tools and algorithms	14
2.8 Theoretical framework	14
2.8.1 Systems theory	14

2.9 Genetic algorithms.....	14
2.9.1 Genetic algorithms implementation	15
2.9.2 Simple genetic algorithm pseudo code.....	15
CHAPTER THREE: RESEARCH METHODOLOGY.....	16
3.0 Introduction	16
3.1 Prototype design	17
3.1.1 ERD Model.....	17
3.1.2 Relational model.....	17
3.1.3 Polyepitope construct definition.....	18
3.2 Prototype implementation	19
3.2.1 Database implementation	19
3.2.2 Populating the database	22
3.2.3 Epitope data integrity	22
3.2.4 Acquisition of polyepitope data	23
3.2.5 Optimized polyepitope formation framework	23
3.2.6 Three- tier application	23
3.2.7 Polyepitope optimization.....	24
3.2.8 Supertypes formation	26
3.2.9 Polyepitopes formation.....	29
3.2.10 Proteasome cleavage prediction	29
3.2.11 TAP prediction	29
3.2.12 MHC binding prediction	30
3.3 Prototype testing.....	31
CHAPTER FOUR: RESULTS AND DISCUSSION	32
4.0 Introduction	32
4.1 Results	32
4.2 Epitopes data for organisms with vaccines	33
4.2.1 T-cell epitopes data for organisms with vaccines.....	33
4.2.2 B –cell Epitopes data for organisms with vaccines	35
4.3 Epitopes data for organisms without vaccines	38
4.3.1 T-cell epitopes data for organisms without vaccines	38
4.3.2 B-cell epitopes data for organisms without vaccines	40
4.4 Optimized polyepitopes.....	42

4.5 Random polypeptides 43

4.6 Discussion..... 44

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION 45

5.0 Introduction 45

5.1 Conclusion 45

5.2 Recommendations 46

5.3 Limitations..... 46

REFERENCES 47

APPENDIX A: USER MANUAL 50

APPENDIX B: OTHER DIAGRAMS..... 52

APPENDIX C: SAMPLE CODE 57

List of tables

Table 1: Web-Based tools for vaccine design 3

Table 2: T- Cell Epitopes data for organisms with vaccines..... 33

Table 3: B cell Epitopes data for organisms with vaccines..... 35

Table 4: T-Cell Epitopes data for organisms without vaccines..... 38

Table 5: B cell Epitopes data for organisms without vaccines..... 41

List of figures

Figure 1: The Animal Cell.....	6
Figure 2: The DNA.....	7
Figure 3: The cell surface.....	7
Figure 4: Protein synthesis	8
Figure 5: ERD	17
Figure 6: Eptopedb.sql code.....	19
Figure 7: Optimized polyepitope formation framework	23
Figure 8: Tiers	24
Figure 9: Polyepitope construct.....	25
Figure 10: HLA- A Tree.....	26
Figure 11: HLA-B Tree	27
Figure 12: Flowchart 1: Classification of Epitopes into supertypes.....	28
Figure 13: Testing report.....	31
Figure 14: Polyepitope Optimization Page	31
Figure 15: T- Cell Epitopes data for organisms with vaccines	35
Figure 16: B cell Epitopes data for organisms with vaccines	37
Figure 17: T cell Epitopes data for organisms with vaccines.....	40
Figure 18: B cell Epitopes data for organisms without vaccines	42
Figure 19: Optimized polyepitopes	43
Figure 20: Random polyepitopes	43

List of abbreviations

WHO-World Health Organization

T.B. – Tuberculosis

B.C.G.-Bacille Calmette Guérine

NCBI – National Center for Biotechnology Information

CHAPTER ONE: INTRODUCTION

1.0 Introduction

Polyepitope vaccines are novel approaches of developing vaccines. This chapter gives the background of the project, statement of the problem, research objectives, gives the significance of the study and gives the definitions of terms used in the project.

1.1 Background of the study

Pathogens have been a threat to humans for a long time. RAC (2002) indicates that there are more than 400 microbial agents associated with disease in healthy adult humans. The World Health Organization (WHO) report (2004a) notes that an estimated 11 million (19%) of the 57 million people that died in the year 2002 of infectious or parasitic infections. The three main infectious diseases are, HIV/AIDS, Tuberculosis (TB) and malaria each of which causes more than a million deaths. AIDS, which is caused by HIV, is now the leading cause of deaths among the young adults world wide.

WHO (2004b) report notes that tackling HIV/AIDS is the world's most urgent public health challenge, more than 20 million people have died from AIDS and an estimated 34- 46 million people are now infected with the HIV. There is no vaccine or a definite cure for AIDS.

The T.B. is another emerging public health threat. TB caused by *Mycobacterium Tuberculosis* (mtb) and is spread from person to person by airborne droplets expelled from lungs when a person with TB coughs sneezes or speaks. Mtb infected person can develop TB if the immune system is impaired by for example HIV infection. In 1995 according to a WHO (2004) report 31 million people died of TB world wide.

Currently there is only one licensed vaccine against TB in the United States of America but it is not recommended for use, this vaccine *Bacille Calmette Guérine* (BCG) is reportedly highly variable in its efficacy to prevent adult pulmonary TB. BCG may have lower efficacy in poor tropical societies where population is more exposed to other *Mycobacterium* in the environment. According to NIAID (2000) report there is need to develop an improved anti-TB vaccine for adequate control of multi-resistant TB (MDR-TB) among other reasons.

The immune system is the system that defends the body against disease causing organisms, malfunctioning cells, and foreign particles. The major task of the immune system is to defend the host against infections, an assignment that is essential to the survival of the organism. The challenge arises because pathogens and diseases are constantly evolving or mutating to escape

the human defence mechanism. In order for the immune system to defend the host it must identify the invading pathogen and initiate a process to eliminate the antigen. The specific sequences in an antigen that are recognized by the components of the immune system are known as epitopes. Vaccines are developed to mimic the working of the adaptive immune system.

It is not cost effective to develop vaccines against all those microbes. Also, it takes a very long time to establish the virulence of a pathogen and therefore it might be dangerous to use partial or complete organism.

Immunological bioinformatics can be used to make an important contribution to the rapid design of novel vaccines by identifying the most immunogenic regions of the pathogen. These regions can subsequently be used as candidates for rational vaccine design.

Polyepitope vaccines are new approaches in vaccine design and development. According to Suhrbier (2002) a short epitope sequence is often capable of inducing protective immunity against a large and complex pathogen. A polyepitope construct may induce immunity against multiple antigenic targets, multiple strain variants or even multiple pathogens such immunization is highly relevant to induce protection against organisms like *HIV* or *Epstein Barr* virus where the immune escape is an important issue or for cancer treatment where immunization with subdominant epitopes might be effective in breaking tolerance. The development of polyepitopes is based on the knowledge of the immunogenic peptides also called epitopes. Such epitopes form the key components of the polyepitope. The number of epitopes that can be included in a poly epitope are limited due to practical and economic reasons. A polyepitope that is effective in one person may be ineffective in another patient because every person possesses a set of MHC class I and MHC class II molecules of different specificity. This makes the selection of optimal set of epitopes an important and interesting optimization problem. There has been exponential growth of tools and databases that store epitopes. This has not been coupled with a corresponding increase in epitope-based vaccines. The epitopes are available in public database. The concatenation of epitopes does not lead to the generation of an optimal polyepitopes. There is need to develop polyepitope optimization tools and carry out more research in this area. Some of the databases are shown in **Table 1** below

Table 1: Web-Based tools for vaccine design

Database	Principal investigators	url	Description
SYFPETHI	Rammansee	www.syfpeth.de	Natural MHC ligands
MHCPEP	Brusic, Harrison	Wehih.wehi.edu.au/mhcpep	MHC binding Peptides
JenPep	Flower	www.jenner.ac.uk/jenpep2	MHC and TAP ligands, B cell epitopes
HIV Molecular Immunological database	Korber	Hiv-web-lanl.gov/content/immunology	HIV epitopes
EPIMHC	Reihertz	Immunax.dfci.havard.edu/tools/db_query_epimhc.html	Mhc ligands
Immune epitope database	Lund O	www.immuneepitope.org http://iedb.org	Epitopes Prediction tools

Source: Lund et al (2005)

SYFPETHI database contains more than 4000 peptides sequences known to bind to MHC class I and II molecules based on previous publications on T cell epitopes and MHC Ligands according to Rammansee et al (1999). JenPep has more than 8000 entries and contains quantitative binding data of peptides to MHC and TAP as well as B and T epitopes as noted by Blithe et al (2002). HIV Molecular Immunological database contains a CTL and T helper epitopes and antibody binding sites for HIV -1 epitopes according to Korber et al (2001a). This epitope database does not contain epitopes for HIV-2. EPIMHC is an MHC ligand database can be searched based on sequence length, class, and species and whether a ligand is an epitope or not, ultimately it is being graded to a publicly accessible immune epitope database containing linear and structural antibody epitopes and T cell epitopes. MHCPEP database contains peptides that bind MHC

The web based tools described above have been developed to provide epitopes data for researchers in immune system and immunological bioinformatics. There is need to develop tools that facilitates research in polyepitope vaccines and promote sharing of research data on poly epitopes.

1.3 Statement of the problem

Polyepitope optimization algorithms that integrate proteasome cleavage, TAP binding and MHC class I binding predictions can be used to make better polyepitopes yet there are no publicly accessible applications on the World Wide Web which can be used by researchers. In this project, the researcher developed a web-based polyepitope application prototype that can be used to optimize the polyepitopes and make the results publicly accessible to researchers.

1.4 Research Objectives

The objectives for this research project were:

1. to compare epitope lengths for pathogens with vaccines and pathogens without vaccines
2. to suggest most appropriate length(s) of epitopes for use in a poly epitopes for malaria, HIV-I and T.B
3. to develop and maintain epitopes in an epitopes database.
4. to optimize polyepitope design
5. to develop a web-based application prototype

1.4 Significance of the study

In this research project, the researcher created a polyepitope optimization prototype that generates publicly accessible results. The prototype can be used to make better polyepitopes which can minimize the cost and time required to develop epitope-based vaccines. The prototype will spur interest in computer scientists to develop tools that can be used in bioinformatics, immunological bioinformatics and computational biology. The prototype will benefit researchers in bioinformatics, immunological bioinformatics and computational biology and vaccine developers.

1.5 Definition of terms

Pathogen-something that can cause disease, such as a bacterium or a virus

Microbe-a microscopic organism, especially one that transmits a disease

Vaccine-substance introduced into the body in order to achieve protection through an immunological reaction against specific micro-organisms that cause a number of infectious diseases.

Efficacy-ability to produce the necessary or desired results

Mycobacterium- a rod like Gram-positive aerobic bacterium that can form branching structures resembling filaments. Some cause diseases in humans, for example tuberculosis or leprosy.

Immune system-the system that defends the body against pathogens, malfunctioning cells, and foreign particles by recognizing and killing pathogens and tumor cells

Epitope-The specific sequences in an antigen that are recognized by the components of the immune system

Antigen-Any agent perceived as foreign by the body's immune system

Immunization- Process of rendering people immune to an infectious organism by inoculating them with a form of the organism that does not cause severe disease but does provoke formation of protective antibodies

Peptide- one of a group of organic chemicals found in most living tissues, with a wide range of biological functions. The chemicals are relatively low-weight polymers of amino acids, as contrasted with the high-weight proteins. The acids are linked together by so-called peptide bonds between their carboxyl (COOH) and alpha amino (NH₂) groups.

Ligands-ion or molecule that is bonded by a coordinate bond to a central transition metal ion in a complex

Immunoglobulin- antibody

TAP-transporter associated with antigen processing

Protein - one of a large group of nitrogen-rich compounds of high molecular weight that are essential and abundant constituents of living organisms

Lymphocytes-white blood cells

Antibody- any normally occurring protein molecule that is produced in the body of cells called lymphocytes and that act primarily as a defence against invasion by foreign substances.

Immunogenic- creating immunity or immune response

Polyepitope- also called **polytope** is a vaccine construct made up of a sequence epitopes and linker amino acid sequences.

Proteasome-large complex in the nucleus and cytosol where selective protein degradation occur

CHAPTER TWO: LITERATURE REVIEW AND THEORY

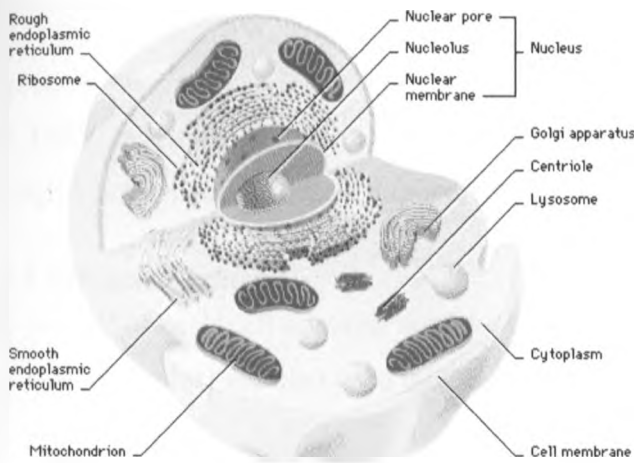
2.0 Introduction

This chapter gives an overview of the animal cell, the immune system, MHC class I and MHC class II pathways, systems theory, give a definition of vaccine, types of vaccines, design and development of vaccines, approaches to polyepitope optimization and discusses genetic algorithms.

2.1 The cell

The cell is smallest unit of an organism that can function independently. All living organisms are made of cell(s). Cells are alive and abilities such grow, move, reproduce and respond in an informed way to the environment. The cell contains many molecules- the main one is DNA which found in the nucleus.

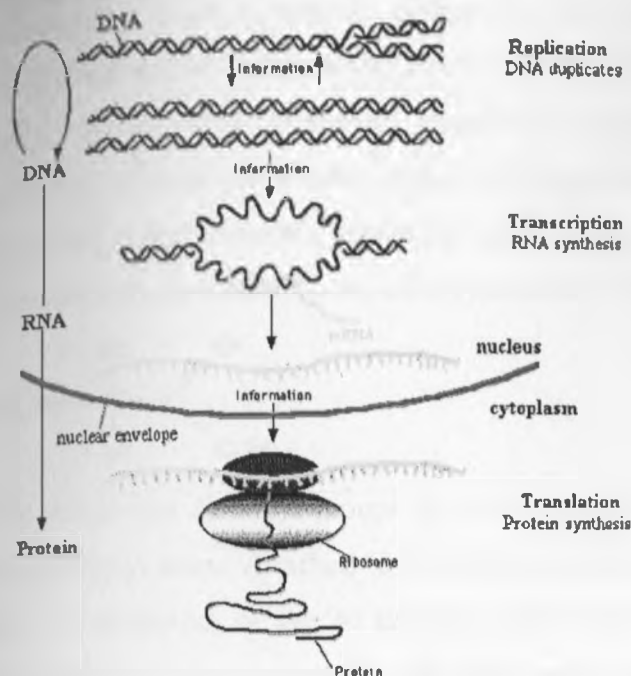
Figure 1: The Animal Cell



Source: Encarta, (2005) Microsoft encyclopaedia

The DNA (deoxyribonucleic acid) is the molecule that acts as the mechanism of biological inheritance in almost all living creatures. It is found in nearly all cells and contains the coded instructions that control the workings of the cell. DNA is passed from parents to offspring, and contains the coded instructions that enable the offspring to develop from a single cell into an adult body. The central concept in molecular biology is illustrated using the **Figure 2**.

Figure 2: The DNA



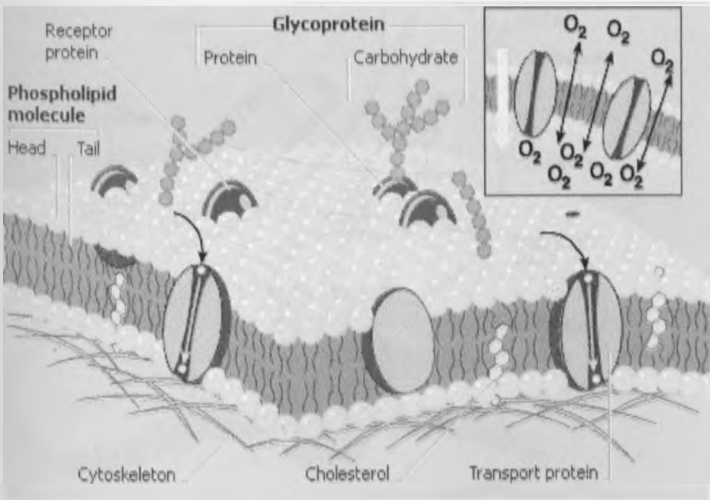
Source: www.

The DNA contains all the information required to manufacture proteins needed by the cell through a three-phase process: duplication, transcription and translation.

2.1.1 The cell Surface

The surface of each cell contains receptors. Receptors can use by pathogens to gain entry into the host cell. Receptors are also used by the cells of the immune system.

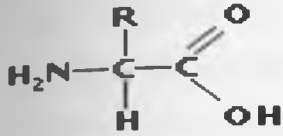
Figure 3: The cell surface



Source: Encarta, (2005) Microsoft encyclopaedia

2.1.2 The amino acids

These are a class of organic compounds that contain both the amino (NH₂) and carboxyl (COOH) groups. Of these acids, 20 serve as the building blocks of proteins. Known as the alpha, amino acids, they comprise alanine(A), arginine(R), asparagine(N), aspartic acid(D), cysteine(C), glutamic acid(E), glutamine(Q), glycine(G), histidine(H), isoleucine(I), leucine(L), lysine(K), methionine(M), phenylalanine(F), proline(P), serine(S), threonine(T), tryptophan(W), tyrosine(Y), and valine(V). All 20 are constructed according to a general formula:

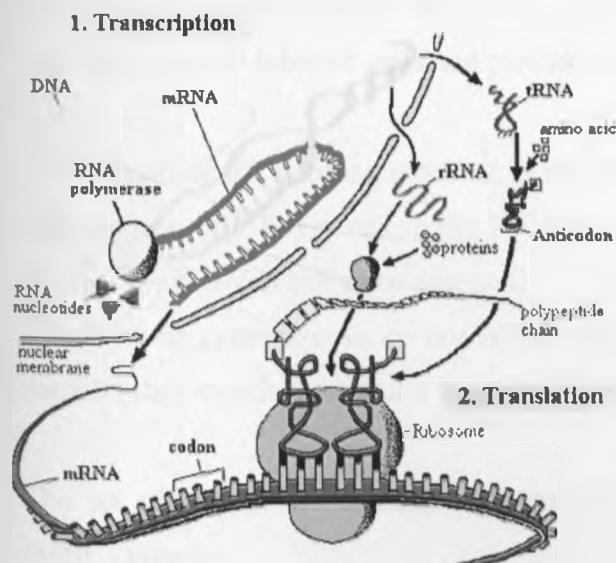


The amino and carboxyl groups are both attached to a single carbon atom, which is called the alpha carbon atom. Attached to the carbon atom is a variable group (R); it is in their R groups that the molecules of the 20 standard amino acids differ from one another. The acidic amino acids DE and are coloured red the basic amino acids-HKR and are coloured blue hydrophobic amino acids are ACFILMPVW are coloured black neutral are GNQSTY and are coloured green

2.1.3 Protein synthesis

Protein is the complex organic nitrogen-rich substance found in the cells of all animals and plants. Twenty (20) different amino acids can linked together in linear polymers, known as polypeptide chains. The following **figure 4** illustrates the synthesis of protein

Figure 4: Protein synthesis



Source: www

2.2 The human immune system

The human immune system is the body system that is primarily responsible for destroying the disease-causing agents that it encounters. The responsibility of the immune system is immense, and it must encompass a vast diversity in order to react appropriately with the thousands of different and potentially disease-causing antigens that invade the body.

The immune system has six major components, three of which are different kinds of cells and three of which are soluble proteins. The three broad categories of immune cells are granulocytes, monocyte/macrophages, and lymphocytes. The three kinds of proteins in the immune system, found dissolved in the liquid portion of the blood, are immunoglobulins, cytokines, and complement proteins. There are thousands of different kinds of immunoglobulins, called antibodies and each of them combines exactly with one specific kind of antigen and helps remove it from the body. All six components can be found circulating in the blood in some form.

The vertebrate immune systems have two basic branches. The main difference between the two lies in the means by which they recognize pathogens. Innate immune system which according to Fearon and Locksley (1996) distinguishes harmful and innocuous (harmless) agents according to the carbohydrate signals.

The adaptive immune system is induced by lymphocytes and can further be divided into: humoral immunity which is mediated by antibody molecules secreted by B lymphocytes that can neutralize pathogens outside cells and cellular immunity which is mediated by T lymphocytes that eliminate the infected cells and provide help to other immune responses.

At the heart of the immune response is the ability of the immune system to distinguish between self and non-self. Every cell in the body carries the same set of distinctive surface proteins that distinguish between self from non-self.

The immune systems cells do not attack own body tissues which carry the same pattern of self markers they coexist peacefully in a state of self tolerance(unless due to autoimmune disease).

The set of unique self markers on human cells is called Major Histocompatibility Complex (MHC) proteins.

There are two classes of MHC proteins: MHC Class I, which are on all the cells and MHC class II proteins which are only on certain cells. Any non- self substance capable of triggering immune

response is called an antigen. An antigen can be a whole non-self cell, a bacterium, a virus, an MHC protein marker or even a protein from a foreign.

The distinctive markers on antigens that trigger immune response are called epitopes. The cells of the immune system recognize MHC proteins when they distinguish between self and non-self. MHC proteins serve as a scaffold–framework that presents peptides from foreign proteins to the immune cells. MHC marker proteins are as distinct as blood types in humans MHC class I bear markers out of possible 200 variations and MHC class II bear 8 markers out of 230 possibilities.

The MHC class I protein molecules which are also called HLAs (human leukocyte antigens) are encoded by 3 different loci on the genome called A, B and C each of the genes is highly polymorphic and for each locus hundreds of different alleles exists. Each allele binds a very specific set of peptides. The alleles can clustered into 9 super types. Alleles within the same super type exhibit the same peptide specificity.

Majority of peptides binding to the HLA complex have a length of between 8 to 10 amino acids (8mers to 10mers). For 9mers, positions 2 and 9 are very important for binding to most class I HLAs. There are two main pathways of processing and presenting antigens to T-Cells. The presentation of peptides to the immune system cells is done by MHC molecules. There are MHC class I and MHC class II molecules. The processing steps are simple sequence analysis performed by the components of the immune system and these steps can be modelled using bioinformatics approaches.

2.2.1 MHC Class I Pathway

In this pathway a protein is cleaved (broken) down by a protein complex called proteasome into peptide fragments, binds to TAP in order to be translocated to ER, bind to MHC class I molecule and thereafter transported to the cell surface. For a peptide to be an epitope it must go through this process more efficiently than other peptides in the cell. The specificity of given molecules can be predicted from the amino acid protein sequences and this can be used to select epitopes to include in a vaccine and help to understand the immune system

2.2.2 MHC Class II Pathway

Presentation of MHC class II molecules follows a different path, after synthesis and translocation and processing peptides are transported to the cell surface. The specificity of given molecules

can be predicted from the amino acid protein sequences and this can be used to select epitopes to include in a vaccine and help to understand the immune system

2.3 Vaccine design and development

A Vaccine is a substance introduced into the body in order to achieve protection through an immunological reaction against specific micro-organisms that cause a number of infectious diseases. The process of making vaccines is lengthy and expensive. According to Andre (2002) the current cost of vaccine making process from concept stage to market stage cost between \$200 and \$500 million.

2.4 Types of vaccines

There are different types of vaccines these are: live vaccines which are able to replicate in the host but attenuated (weakened) micro-organisms and do not cause disease, inactivate or killed micro-organisms (subunit) and genetic vaccines. Genetic vaccines carry one or more epitopes rather than whole organisms

2.5 Polyepitope construct

Polyepitope also called polytope is a vaccine construct made up of a sequence epitopes and linker amino acid sequences. The number of epitopes included in a poly epitope is usually kept small up to a dozens epitopes. Linker or flanking amino acid sequences are inserted between to improve the quality of the polyepitope. There are three types of epitopes that make the polyepitope construct. The B epitopes which are recognized by antibodies, The T-helper (Th) epitopes are recognized by T-helper cells and the T-CTL epitopes are recognized by the cytotoxic T lymphocytes.

In order to predict epitopes using computational methods the following steps are used. Analyse of protein sequence. For B-cell epitopes, predict secondary structure, hydrophobicity, mobility/flexibility and for T-cell epitopes, search for anchor residues (amino acids) and predict proteasomal cleavage. For a 9-mer (9 amino acid length) epitopes, positions 2 and 9 are important. For example in measles T cell epitope whose id is 32029 and the sequence is K L M P N I T L L amino acids positions 2(L) and 9(L) are important. The amino acids in those positions are the anchor residues. **Figures 10 and 11** give detailed information on which residues are important for each of the selected super types. Good polyepitopes constructs are those that ensure large population coverage and the number epitopes in a polyepitope that are correctly processed and presented represent most super types. The super types **A2, A3 and B7** cover 83%

to 88% percent of the global population while super types **A1, A2, A3, A24, B7, B27, B44, B58** and **B62** cover 98.1% to 100% of the global population.

According to Suhrbie (2002) a short epitope sequence is capable of inducing protective immunity against a large and complex pathogen. Including several such immunogenic epitopes in a polytope construct may induce immunity against multiple antigenic targets, multiple strain variants or even multiple pathogens. Such immunization is highly relevant to induce protection against organisms like HIV or *Epstein-Barr* virus where immune escape is an important issue, or for cancer treatment where immunization with subdominant epitopes might be effective in breaking tolerance.

According to Flower et al (1999) the design of a polyepitope can be done in an effective way by modifying sequential order of different epitopes and liker amino acid(s) that will favour proteasome cleavage and transport/translation in ER by TAP complex.

2.6 Approaches to poly epitope optimization

There are several ways that have been used to computationally build and optimize a polytope constructs. There include:

2.6.1 Quantitative matrix (QM)

This approach of polyepitope optimization has been used in TPREDICT application. The QM based approach which predicts MHC-binding prediction is illustrated using the peptide **RLRPGGKKK**. At first, this peptide is encrypted with sparse encoding: it is represented as an array with shape $i*j$ ($20*9$) where each nonzero element at position (i,j) codes for amino acid i at position j :

```
# A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y
{{0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0},
 {0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0},
 ...
 {0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0}}
```

If amino acid properties are used to parametrize peptide, selected scale "P" of properties is multiplied with obtained sparse matrix element wise to produce an array $\{\{P1\},\{P2\},...\{P9\}\}$ (all zero elements are removed). This array is multiplied with predictive model matrices element wise. According to the type of selected model, either multiplicative or additive, elements of resulted matrix are either multiplied or summed up. The resulting score after correction (either multiplication with corrective coefficient or summation with corrective constant according to the

type of used model) is compared to the selected threshold; if score is greater than threshold, peptide is considered to be a binder.

2.6.2 Quadratic programming

This method has been used by Lilliana et al (2003). In this model assigns epitopes a binding strength greater than the minimum binding strength of known epitopes. The non epitopes are assigned binding strength less than the minimum binding strength of known epitopes. The high binders have a higher binding strength than the low and medium binders and that the medium binders have a higher binding strength than the low binders.

2.6.3 Integer linear programming

Toussaint N et al (2008) created a mathematical model for the selection of optimal set of peptides for a epitope-based vaccine. The main idea behind this approach is that starting from target antigens, a list of properties of interest, and a target population the information necessary to determine an optimal set of epitopes is derived. Given this information, a mathematical framework can be used to find a set of epitopes that is optimal to the target population and properties of interest. Using this approach a set of optimal epitopes for inclusion in a polyepitope can be defined, however the approach does not take into the fact that epitope-epitope interaction (epitasis) in a polyepitope

2.6.4 Monte Carlo metropolis simulations algorithm

This method relies on repeated random sampling to compute their results. This method is most suited to calculation by a computer and tends to be used when it is infeasible to compute an exact result with a deterministic algorithm. The method is also used to complement the theoretical derivations. The Monte Carlo method was coined in the 1940s by John von Neumann, Stanislaw Ulam and Nicholas. PolyCTLDesigner developed by Denis V. Atonets, is an application developed optimizes a polyepitope construct for HIV-1 and does not run on a web server. The system utilizes Monte Carlo Metropolis Simulations an algorithm developed by Metropolis et al (1953). Metropolis Monte Carlo metropolis simulations algorithm was first used by scientists on the Manhattan project to calculate the probability with which a neutron from one fission Uranium1 atom would cause another to fission according to Feynman (1985). There are plans to upgrade PolyCTLDesigner according to its developer.

2.7 Epitope based tools and algorithms

There are various epitope-based tools and algorithms available to researchers among them are: The SYFPETHI database which contains more than 4000 peptides sequences known to bind to MHC class I and II molecules based on previous publications on T cell epitopes and MHC Ligands according to Rammansee et al (1999). JenPep has more than 8000 entries and contains quantitative binding data of peptides to MHC and TAP as well as B and T epitopes as noted by blithe et al (2002).

HIV Molecular Immunological database contains a CTL and T helper epitopes and antibody binding sites for HIV -1 epitopes according to Korber et al (2001a). This epitope database does not contain epitopes for HIV-2.

EPIMHC is an MHC ligand database can be searched based on sequence length, class, species and whether a ligand is an epitope or not ultimately being graded to a publicly accessible immune epitope database containing linear and structural antibody epitopes and T cell epitopes. MHCPEP database contains peptides that bind MHC.

The immune epitope database (IEDB) is a publicly accessible web based tool that contains epitope database. It contains linear and conformational antibody and T cell epitopes.

2.8 Theoretical framework

2.8.1 Systems theory

The researcher used systems theory to model the immune system components and processes. Systems theory was first advanced by von Bertalanffy when he presented his idea of a General System Theory in a philosophy seminar at the University of Chicago in 1937. Generally a system consists of complex interacting components together with the relationships among them that permit the identification of a boundary-maintaining entity or process. The theory was used the theory to understand the components and processes of the immune system.

In this research project, the researcher used the models of Toes et al (2001) to create an algorithm for proteasome cleavage prediction and Peters et al (2003) to develop an algorithm for TAP binding prediction algorithm

2.9 Genetic algorithms

A genetic algorithm (GA) is a search technique used in computing to find exact or approximate solutions to optimization and search problems. Genetic algorithms are categorized as global search heuristics. Genetic algorithms are a particular class of evolutionary algorithms (also

known as evolutionary computation) that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover (also called recombination)

2.9.1 Genetic algorithms implementation

Genetic algorithms are implemented as a computer simulation in which a population of abstract representations (called chromosomes or the genotype of the genome) of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem evolves toward better solutions. Traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. If the algorithm has terminated due to a maximum number of generations, a satisfactory solution may or may not have been reached.

A typical genetic algorithm requires two things to be defined: the genetic representation of the solution domain and the fitness function to evaluate the solution domain.

2.9.2 Simple genetic algorithm pseudo code

- Choose initial population
- Evaluate the fitness of each individual in the population
- Repeat until termination: (time limit or sufficient fitness achieved)
 - Select best-ranking individuals to reproduce
 - Breed new generation through crossover and/or mutation (genetic operations) and give birth to offspring
 - Evaluate the individual fitness of the offspring
 - Replace worst ranked part of population with offspring

CHAPTER THREE: RESEARCH METHODOLOGY

3.0 Introduction

This chapter presents the research methodology used in this project. The objectives of this project were to compare the lengths of epitopes of sample pathogens with vaccines to those without, suggest most appropriate length(s) to use a polyepitope construct for malaria, HIV-1 and TB. There has been exponential growth of biological data as noted in literature review. Some of this data is stored in private database and part of it is stored in publicly accessible databases such as IEDB and HIV molecular immunological database.

There was need to create a local database to store part of this data for better analysis of such data depending on a researcher's interest. In this project a local database was created to store data downloaded from IEDB and HIV molecular immunological database.

Epitopes data was collected in December 2011 from two sources. The main source of data was the immune epitope database (IEDB) <http://iedb.org>. The epitopes data were collected from this source excluding the HIV epitopes. The HIV epitopes data were collected from the HIV molecular immunological database. The B –cell epitopes were filtered by species criteria (human only) and then analyzed.

The search criteria used for epitopes includes specifying B –cell or T – cell epitopes, the host organism and the source organism. If the host name or source organism was not on the list a search was done using the pathogen name from the NCBI database which is linked to IEDB to facilitate the search.

Pathogens were divided into two clusters – those pathogens that have vaccines and pathogens that do not have vaccines. Seven out of fourteen pathogens with vaccines were selected from NIAID A-C list. For each pathogen selected, B- cell epitopes and T- Cell epitopes were downloaded from IEDB and the HIV molecular immunological database. The epitopes were downloaded in CSV format and saved in Microsoft Excel for B–cell and T-cell. For each pathogen B –Cell and T- cell epitopes were saved in a separate workbook.

The epitopes were analysed using =LEN() and=COUNTIF(\$m\$2: \$m\$n, p) functions. The first function was used to calculate the epitope length of each epitope downloaded. The second

function was used to calculate the frequencies of epitopes whose absolute range was \$m\$2 to \$m\$n and lengths p1, p2 ...px.

The database was populated with epitope data using PhpMyAdmin graphical user interface application. The data which had been stored in csv format in Microsoft excel workbooks and uploaded using batch processing.

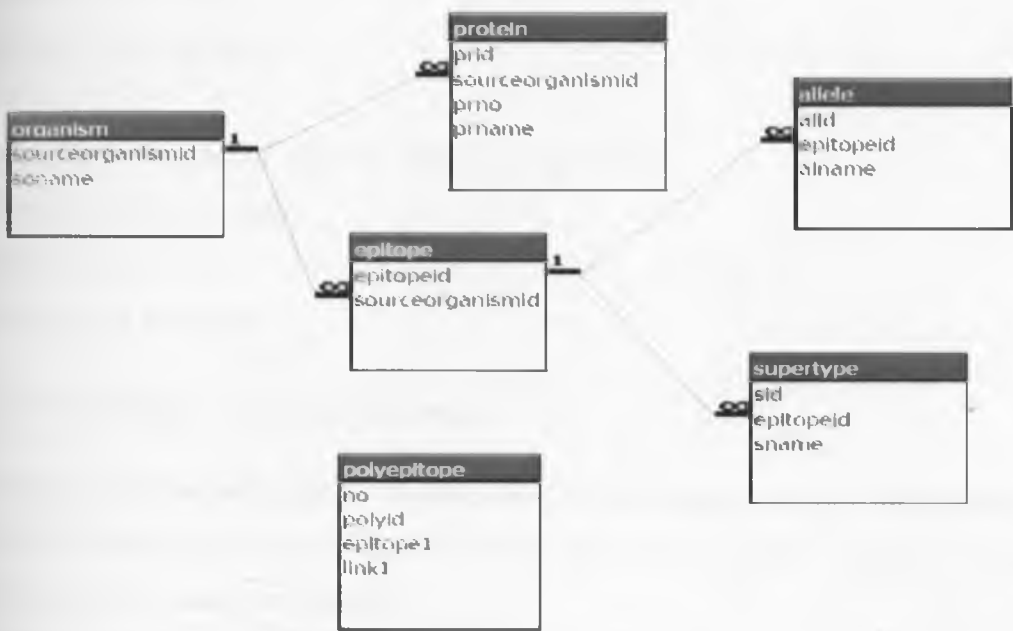
3.1 Prototype design

3.1.1 ERD Model

The epitope database stores epitope data and generate required reports the following entities were identified: Organism, Epitope, Protein, Super type, Allele and polytepitope

The entity relationship diagram for the database is shown below

Figure 5: ERD



3.1.2 Relational model

The ERD created was converted to a relational model shown below

Organism (sourceornanismid, soname)

Primary Key sourceornanismid

Foreign Key -

Protein (prid, sourceorganismid,prno,prname)

Primary Key prid

Foreign Key sourceorganismid

Epitope (epitopeid, sourceorganismid,objectdescription,
linearsequence,sma,smname,type)

Primary Key epitopeid

Foreign Key sourceorganismid

Supertype (sid, epitopeid,sname)

Primary Key sid

Foreign Key epitopeid

Allele (alid, epitopeid,alname,mhcclass)

Primary Key alid

Foreign Key epitopeid

Polyepitope (no,polyid, epitope1,link1, epitope2,link2,
epitope3, ,rank, run date)

Primary Key no

Foreign Key epitopeid

3.1.3 Polyepitope construct definition

In this project the poly epitope is made up of three epitopes and two link sequences between the first and the second epitope and between the second and the third epitope. Each epitope is made up nine (9) amino acid sequences.

The link amino acid used is **ADL**. In a fully implemented system a variety of linkers of varying lengths would have to be selected. This linker is used in this because it promotes cleavage in the flanking (link) regions. The polyepitope of this length can cover 83% to 88% of the human population if the epitopes from super types **A2, A3 and B7** are correctly processed and presented to MHC class molecules.

3.1.4 Goals

The main goal of designing a polypeptide construct is to ensure all epitopes contained in the polypeptide are correctly processed and presented to the immune system cells. There are two pathways to processing and presentation of antigens to T- lymphocytes as stated in chapter one. First, the MHC class I pathway which is used to present endogenous antigens to CD8+ T-cell. In order to be presented a peptide must be generated by proteasome, the proteasome cleaves (degrades) the peptide into small peptides. The peptides presented, must bind to TAP in order to be translocated to Endoplasmic Reticulum(ER) and be transported to the cell surface. The most selective step is binding to MHC class I molecules. In order for a peptide to be immunogenic it must go through proteasome cleavage, bind to TAP and bind to MHC class I molecules more efficiently than other peptides. Second, the presentation of MHC class II molecules follows a different path, after synthesis and translocation and processing peptides are transported to the cell surface. The specificity of given molecules can be predicted from the amino acid protein sequences and this can be used to select epitopes to include in a vaccine and help to understand the immune system

3.2 Prototype implementation

The steps below describe the implementation of the prototype.

3.2.1 Database implementation

The database was implemented using MYSQL database, innodb engine was used. The innodb engine supports usage foreign keys and transactions. MYSQL database was select because it is fast and it is an open source software and therefore accessible. The support of foreign keys ensures data integrity. MYSQL database was also selected because it works well with Apache web server and compatible with PHP scripting language

The following SQL script was used to create the database.

Figure 6: Epitopedb.sql code

```
/* EPITOPEDB VERSION 1 created in December 2011 */
/* EPITOPEDB VERSION 1 */
Create database if not exists `EPITOPEDB`;
USE `EPITOPEDB`;
DROP TABLE IF EXISTS `organism`;
CREATE TABLE `organism` (
  `sourceorganismid` int(10) NOT NULL,
```

```

'soname' varchar(30) default NULL,
PRIMARY KEY ('sourceorganismid')
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS 'species';
CREATE TABLE 'species' (
'spid' int(5) NOT NULL ,
'sourceorganismid' int(10) NOT NULL,
'spname' varchar (40) NULL,
PRIMARY KEY ('spid'),
KEY 'sourceorganismid' ('sourceorganismid'),
FOREIGN KEY ('sourceorganismid') REFERENCES 'organism' ('sourceorganismid')
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS 'protein';
CREATE TABLE 'protein' (
'prid' int(2) NOT NULL,
'spid' int(5) NOT NULL ,
'prno' int(10) NOT NULL,
'prname' varchar(20) NOT NULL,
PRIMARY KEY ('prid'),
KEY 'spid' ('spid'),
FOREIGN KEY ('spid') REFERENCES 'species' ('spid')
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS 'epitope';
CREATE TABLE 'epitope' (
'epitopeid' int(7) NOT NULL ,
'sourceorganismid' int(10) NOT NULL,
'objectdescription' varchar (20) NULL,
'linearsequence' varchar (20) NULL,
'sma' varchar (20) NULL,
'smanme' varchar (50) NULL,
'type' enum('T','B') NOT NULL,
PRIMARY KEY ('epitopeid'),
KEY 'sourceorganismid' ('sourceorganismid'),
FOREIGN KEY ('sourceorganismid') REFERENCES 'organism' ('sourceorganismid')

```

```

) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS `supertype`;
CREATE TABLE `supertype` (
  `sid` int(3) NOT NULL,
  `epitopeid` int(7) NOT NULL ,
  `sname` varchar(5) NOT NULL ,
  PRIMARY KEY (`sid`),
  INDEX (epitopeid),
  FOREIGN KEY (epitopeid) REFERENCES epitope (epitopeid)
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS `allele`;
CREATE TABLE `allele` (
  `alid` int(3) NOT NULL,
  `epitopeid` int(7) NOT NULL ,
  `alname` varchar(7) NOT NULL ,
  `mhcclass` varchar(2) NOT NULL ,
  PRIMARY KEY (`alid`),
  INDEX (epitopeid),
  FOREIGN KEY (epitopeid) REFERENCES epitope (epitopeid)
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS `polytope` ;
CREATE TABLE `polytope` (
  `epid` int( 5 ) NOT NULL ,
  `A2` varchar( 20 ) NULL ,
  `A3` varchar( 20 ) NULL ,
  `B7` varchar( 20 ) NULL ,
  `O` varchar( 20 ) NULL ,
  PRIMARY KEY ( `epid` )
) ENGINE = InnoDB DEFAULT CHARSET = latin1;
DROP TABLE IF EXISTS `polytope2`;
CREATE TABLE `polytope2` (
  `epitopeid` int(7) NOT NULL ,
  `epitope` varchar(20) NOT NULL ,
  `organism` varchar(50) NOT NULL ,

```

```

'supertype' char(2) NOT NULL ,
'counter1' int(4) NOT NULL ,
'counter2' int(4) NOT NULL ,
'counter3' int(4) NOT NULL ,
PRIMARY KEY ('epitopeid'),
FOREIGN KEY (epitopeid) REFERENCES epitope (epitopeid)
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS `polytope3`;
CREATE TABLE `polytope3` (
  'epitopeid' int(7) NOT NULL ,
  'epitope' varchar(20) NOT NULL ,
  'organism' varchar(50) NOT NULL ,
  'supertype' char(2) NOT NULL ,
  'counter' int(4) NOT NULL ,
PRIMARY KEY ('epitopeid'),
  FOREIGN KEY (epitopeid) REFERENCES epitope (epitopeid)
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
DROP TABLE IF EXISTS `polytope`;
CREATE TABLE `polytope` (
  'pid' int(7) NOT NULL ,
  'organism' varchar(50) NOT NULL ,
  'polytope' char(33) NOT NULL ,
  PRIMARY KEY ('pid')
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

```

3.2.2 Populating the database

The database was populated with epitope data using PhpMyAdmin graphical user interface application. The data which had been stored in csv format in Microsoft excel workbooks and uploaded using batch processing. The column headers in the comma separated value (csv) Excel files should be the same columns in the respective database table and same order maintained.

3.2.3 Epitope data integrity

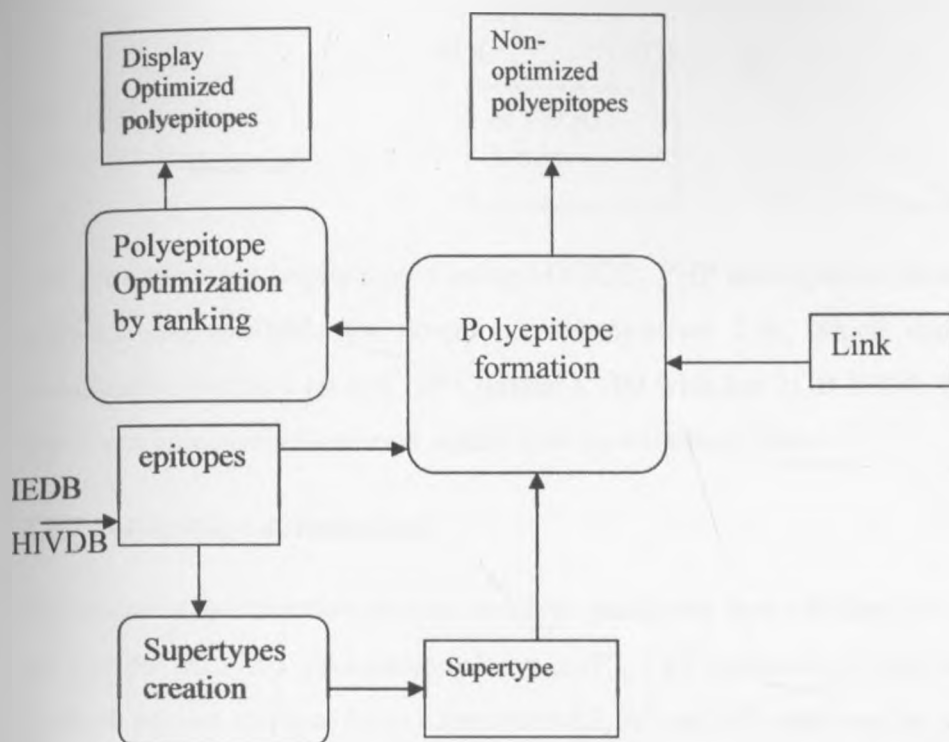
The data downloaded from the sources stated earlier is of high integrity. The IEDB data is only uploaded to the database after it has been curated by highly experienced IEDB curators. Similarly HIV molecular database is maintained by Los Alamos Laboratory.

3.2.4 Acquisition of polyepitope data

In order to design the polyepitope all the required data is stored in an epitope database. This data includes epitope; species super type, allele, species, epitope type, molecule, protein and organism.

3.2.5 Optimized polyepitope formation framework

Figure 7: Optimized polyepitope formation framework



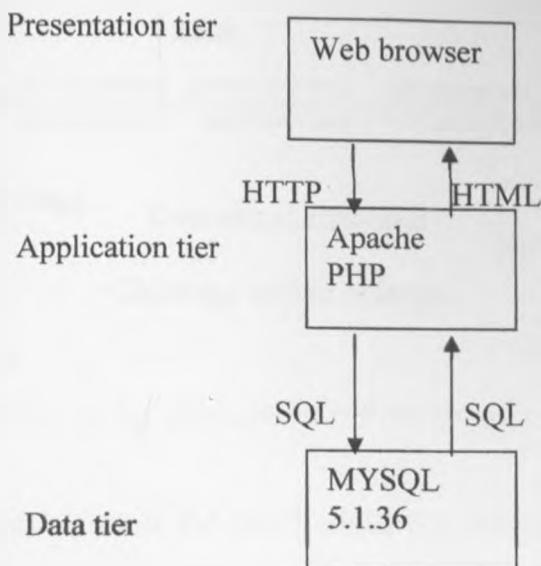
The processes of the Polyepitope optimization prototype are to: select epitopes and organism, specify linker amino acid sequence, and cluster formation (classify epitopes into supertypes A2, A3, B7 and O). The supertype O does not exist but was used in this project to denote supertypes of epitopes which do not fit the three supertypes that were select. Polyepitope formation and calculating the quality of the polyepitope optimization are the main processes.

3.2.6 Three- tier application

The three tier application architecture has been chosen because it would facilitate faster sharing of research data.

The presentation tier runs on a web browser, the HTTP server that has been use is Apache 2.2.11, server side scripting language used is PHP and the data tier runs on MYSQL 5.1.36.

Figure 8: Tiers

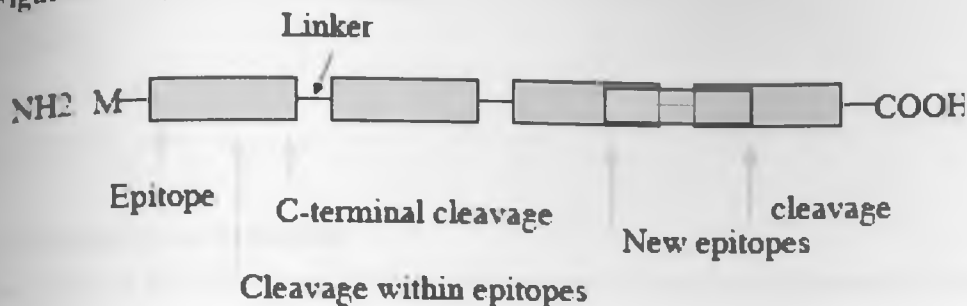


The prototype was implemented using MYSQL, PHP and Apache because they are open source software and available for download. Wampserver 2.0i, which contains the software, was downloaded installed on and HP Compaq C700 with has 1GB RAM, HDD 120 GB, 1.46, 1.47 Duo Core processors Notebook which runs on windows Vista.

3.2.7 Polyepitope optimization

Polyepitopes optimization process involves predicting how efficiently a peptide goes through the three steps described: proteasome cleavage (P), TAP transport (T) and MHC binding (M). In this research project epitopes from supertypes A2, A3 and B7 were used to construct the polyepitope. The aim was to reduce internal cleavage sites within each epitope to zero. Cleavage should occur at C- termini between P1 and P1' positions. The right end of a peptide is the C- terminal and the left end is the N terminal. The new epitopes processed and presented in fusing regions spanning over epitopes should tend to zero. The length of the linker and also the number of poor C-terminal sites also determine the quality of the polyepitope. The predicted score should be less than 0.9 as noted by Lund et al (2005). Minimal variables were used in this prototype, in a fully implemented system all the predictive variables would have to be included.

Figure 9: Polyepitope construct



Source: www

The quality (Q) of the polyepitope = Proteasome cleavage score (P) + TAP score (T) +MHC score (M)

In this research project the MHC scores (M) was held constant, each polyepitope consists of epitopes from supertypes A2, A3 and B7.

The local epitope database contains epitopes of varying lengths and only 9mer epitopes were selected and selected epitopes were pathogen specific. Malaria (*plasmodium falciparum*), HIV and Mycobacterium Tuberculosis epitopes were considered. The following example illustrates the polyepitope quality calculation.

EXAMPLE

Given the polyepitope below where the array indexes are 0 to 32, where N1 is the element at index 0, N2 element at index 1, N3 element at index 2 and C element at index 32, also A2C9 is element at index 8, A3C9 is element at index 20 (R) and B7C9 is element at index 32 the sequences coloured blue are the linkers.

Polyepitope: DVKDTKEALADILEELRQHLLRADLYPGIKVKQL

TAP SCORE (T)= N1+ N2+ N3+C where N1 is D, N2 is V, N3 is K and C is L(the element at index 32 of the array., substitute with values from peters et al(2003) table 1

Therefore T= 1.37+(-0.50)+0.09+(-0.94)

$$T= 0.12$$

PC SCORE (P) = A2C9+ A3C9+ B7C9 where A2C9 is element at index 8 i.e. (L), A3C9 is element at index 20 i.e. (R) and B7C9 is element at index 32 i.e. (L) of the array, substitute values with Toes et al table (use P1 column values and divide each by 1000)

$$P=1.185+0.275+1.185$$

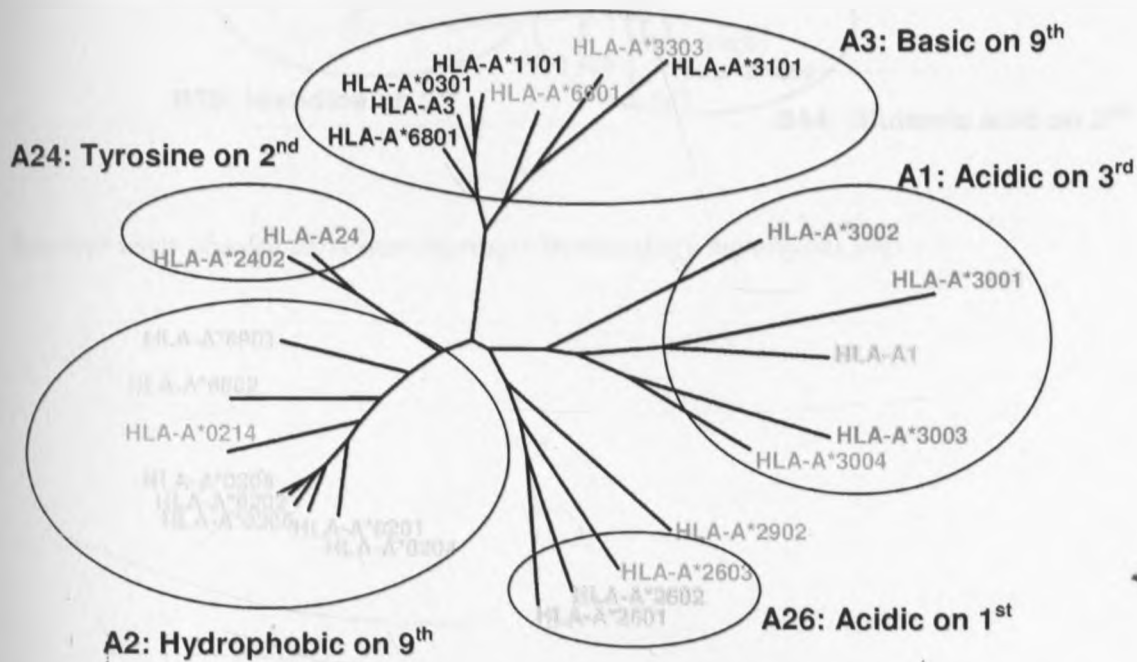
$$P=2.645$$

$$Q=2.765$$

The rules for the formation of the supertypes are derived from **Figures 10 and 11** below.

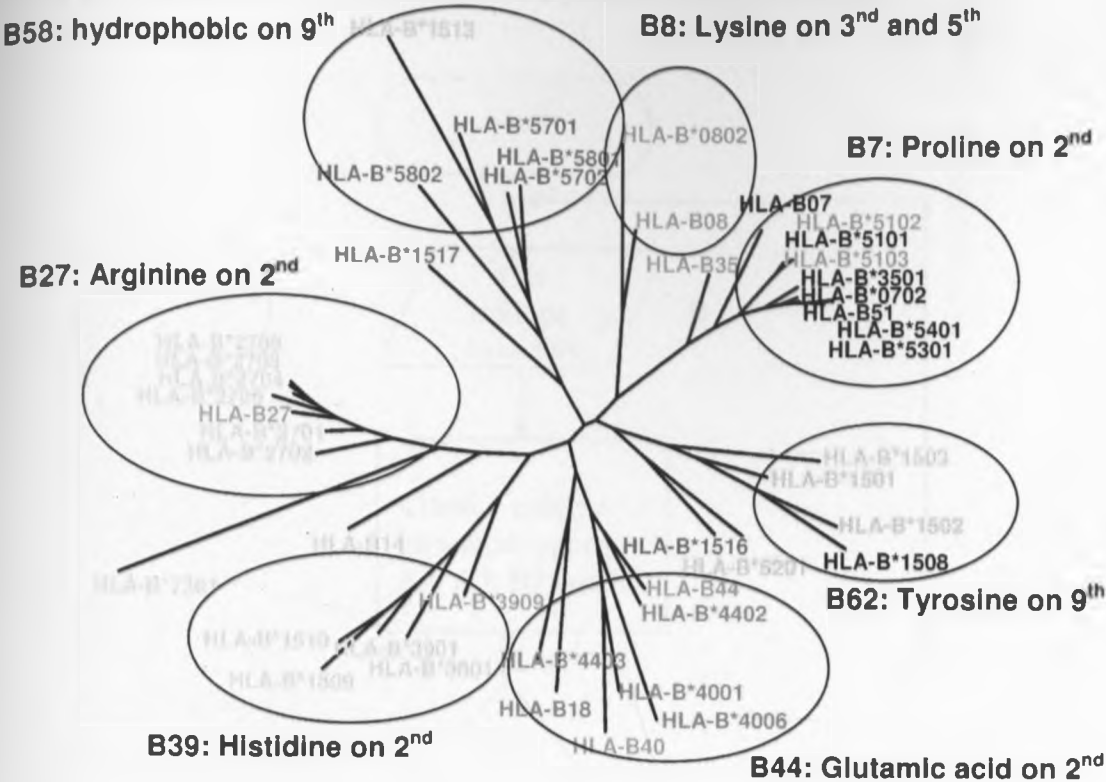
In a 9-mer epitopes the anchor residues (amino acids) for the supertypes are as follows: if epitopes are to belong to supertype **A2** the amino acid on the 9th position must be hydrophobic which means they must contain either of these amino acids **ACFILMPVW**. If epitopes are to belong to supertype **A3** the amino acid on the 9th position must be basic which means they must contain either of these amino acids **HKR** and if epitopes are to belong to supertype **B7** the amino acid on the 2nd position must be Proline (**P**). The supertypes are to generate rules for supertype clusters which determine MHC binding.

Figure 10: HLA- A Tree



Source: www.cbs.dtu.dk/researchgroups/immunology/supertypes.php

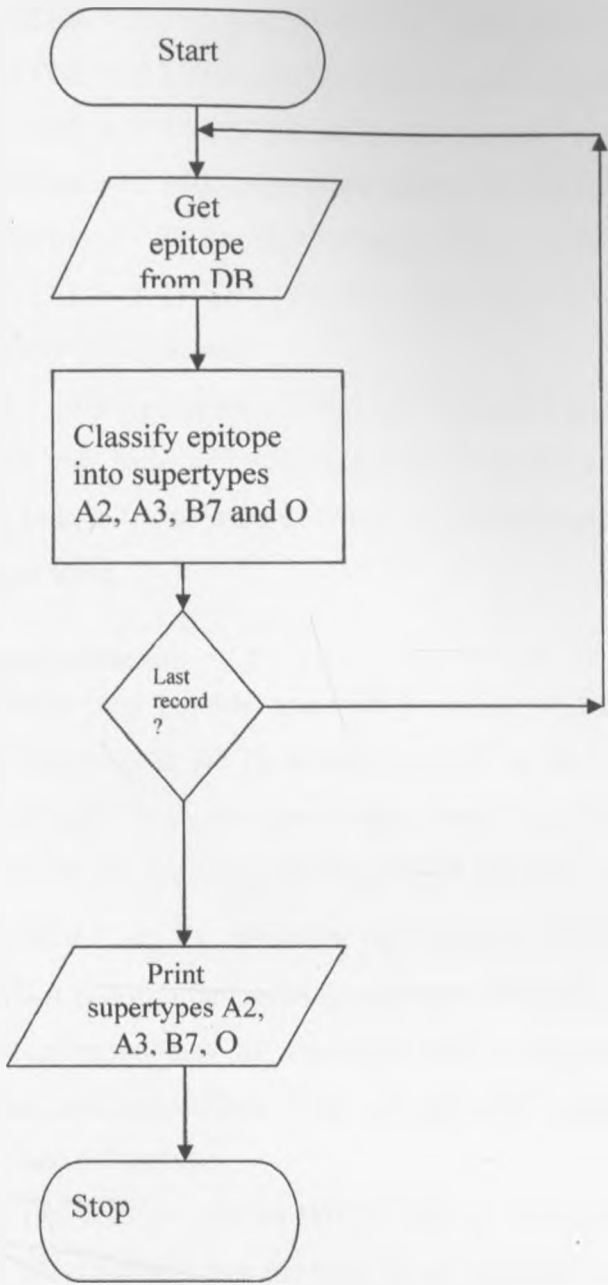
Figure 11: HLA-B Tree



Source: www.cbs.dtu.dk/researchgroups/immunology/supertypes.php

The following flow chart illustrates the formation of supertypes from selected 9mer epitopes. The supertypes were stored in the supertypes table.

Figure 12: Flowchart 1: Classification of Epitopes into supertypes



The epitopes are classified into supertypes A2, A3, and B7 otherwise it's classified into supertype O which does not exist. Supertype O is used only in this research project prototype to indicate epitopes which do not fall into A2, A3 and B7 categories.

3.2.9 Polyepitopes formation

Cross tabulation query was use to transform the supertypes into A2, A3 and B7 columns. The cross tabulation query is show below. (See polyepitope2.php script for further clarification)

```
SELECT SID, EPITOPEID, ORGANISM,  
MAX(CASE WHEN CLUSTER = '1' THEN SUPERTYPE END) AS A2,  
MAX(CASE WHEN CLUSTER = '1' THEN EPITOPE END) AS EPITOPE,  
MAX(CASE WHEN CLUSTER = '2' THEN SUPERTYPE END) AS A3,  
MAX(CASE WHEN CLUSTER = '2' THEN EPITOPE END) AS EPITOPE,  
MAX(CASE WHEN CLUSTER = '3' THEN SUPERTYPE END) AS B7,  
MAX(CASE WHEN CLUSTER = '3' THEN EPITOPE END) AS EPITOPE  
FROM SUPERTYPE GROUP BY epitopeid;
```

The supertypes were by the script (polyepitope2.php) on an html page which was exported to Excel by right clicking on the page and selecting export to excel. The Excel function = concatenate (A2, link, A3, link, B7) was used to create the polyepitope. The polyepitopes were the stored in the polyepitopes table.

3.2.10 Proteasome cleavage prediction

Given an N amino acid (AA) long peptide and with positions within the peptide P_1 - P_n and flanking positions FN , FC where N is the N- terminal and C is the C- terminal, a peptide is produced if positions FN - P_1 and FN - P_n are cleaved and none of internal positions are cleaved. Cleavage occurs between P_1 and P_1' according to Berger and Schechter (1970). Residues on the left of flanking region are called P_1 , P_2 , P_3 .. while the right side are called P_1' , P_2' , P_3' ... The P_1 position is the most important position determining cleavage as noted by Altuvia and Margalit (2000) although the flanking region may be important Mo et al(2000). In this project, each polyepitope is 33 mer long and consists of 9mer A2 epitope+3mer link (ADL), 9mer A3 epitope+ 3 mer link (ADL) and B7 epitope.

Proteasome cleavage score (P) = $A2C9 + A3C9 + B7C9$. The C- termini scores of each epitope is calculated using Toes et al (2001) table 1 (see the table in the Appendix B).

3.2.11 TAP prediction

The TAP prediction score (T) of amino acid sequence (AA) N_1 , N_2 , N_3 C is given by $T = MAX_1N_1 + MAX_2N_2 + MAX_3N_3 + MAX_9C$ where N_1 , N_2 , N_3 are N termini positions and C is the C- terminal of the polyepitope and MAX_p is the matrix score. The TAP scores where based on the Peters et al (2003) consensus scoring matrix which gives high accuracy TAP scores (see the table in the Appendix B). The above scoring matrix was derived on data for 9mer peptides and is

also used to predict TAP score for peptides longer than 9mers because binding of peptides to TAP is determined mainly C-terminal and the three N-terminal residues N1,N2,N3. In position positions N1 and N2 basic residues are favoured i.e. HKR and N3 can either be W or Y. P at position 2 is the strongest destabilizing residue found, nearly abolishing binding as indicated by Uebel et al (1997). This implies B7 epitope should not be the leftmost epitope in a polyepitope.

3.2.12 MHC binding prediction

There are a number of methods that have been used to predict MHC binding according to Schirle et al (2001). Majority of peptides binding to HLA complex are 8mer to 10mer peptides. For 9mers, peptides positions 2 and 9 are very important for binding to most class I HLAs these positions are referred to as anchor positions according to Rammansee et al (1999). In this research project MHC predictions are based on the anchor positions for the supertypes A2, A3 and B7 as shown in Figures 10 and 11. All the polyepitopes contain A2, A3 and B7 epitopes therefore MHC variable was held constant.

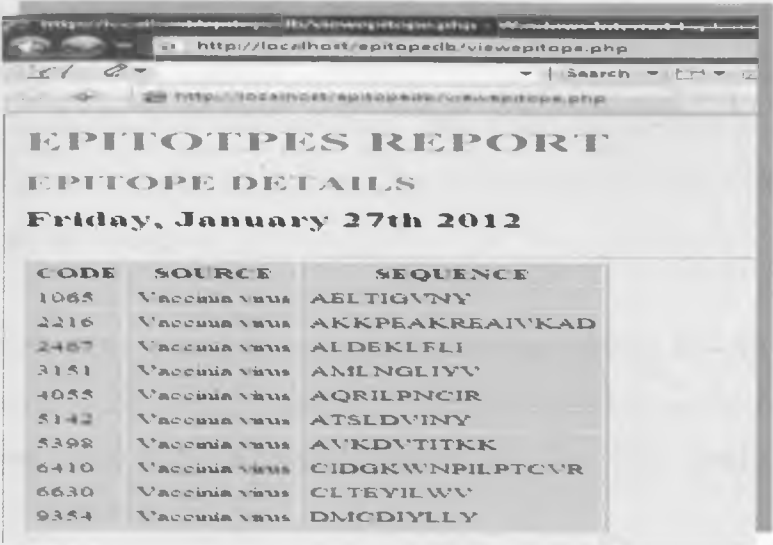
3.3 Prototype testing

The optimized polypeptides generated by the prototype were tested on NetChop 3.1 server. The results of random polypeptides and optimized polypeptides were compared and conclusions drawn.

The web server has been tested and the following epitopes report was generated using the SQL script. **SELECT e.epitopeid, o.soname, e.linearsequence FROM epitope e, organism o WHERE e.sourceorganismid=o. sourceorganismid limit 10;** the following report was generated by the prototype.

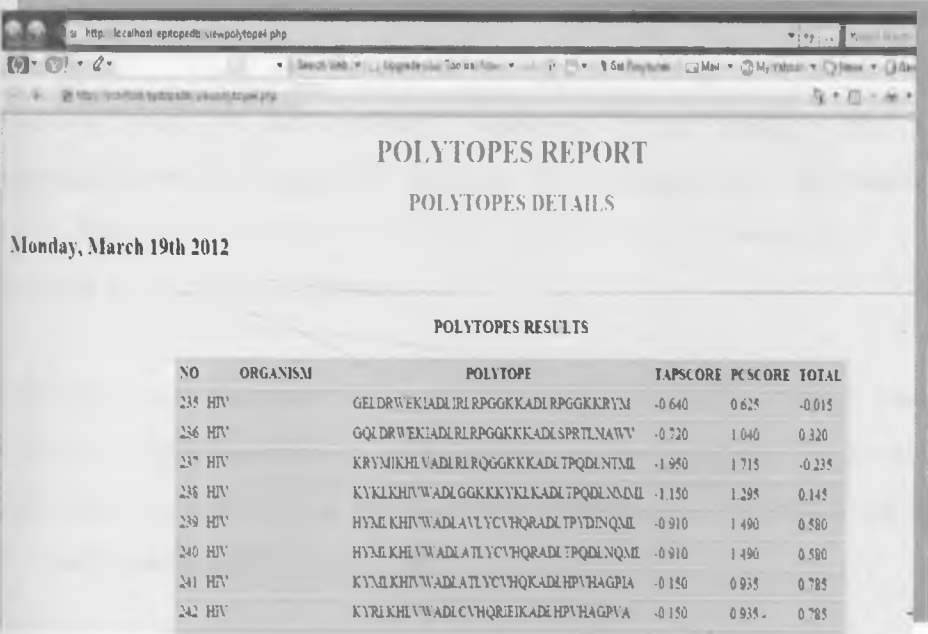
Figure 13: Testing report

<http://localhost/epitopedb/viewepitope.php>



EPITOPES REPORT		
EPITOPE DETAILS		
Friday, January 27th 2012		
CODE	SOURCE	SEQUENCE
1065	Vaccinia virus	AELTIGVNY
2216	Vaccinia virus	AKKPEAKREAIVKAD
2467	Vaccinia virus	ALDEKLFLI
3151	Vaccinia virus	AMLNGLIVV
4055	Vaccinia virus	AQRILPNCIR
5142	Vaccinia virus	ATSLDVINY
5398	Vaccinia virus	AVKDVITITKK
6410	Vaccinia virus	CIDGKWNPIPTCVR
6630	Vaccinia virus	CLTEYILWV
9354	Vaccinia virus	DMCDIYLLY

Figure 14: Polypeptide Optimization Page



POLYTOPES REPORT				
POLYTOPES DETAILS				
Monday, March 19th 2012				
POLYTOPES RESULTS				
NO	ORGANISM	POLYTOPE	TAPSCORE	PCSCORE TOTAL
235	HTV	GELDRWEKIADLRIRPGGKKADLRPGGKKRYM	-0.640	0.625 -0.015
236	HTV	GQDRWEKIADLRIRPGGKKADLRPTLNQWV	-0.720	1.040 0.320
237	HTV	KRYMIKHLVADLRIRPGGKKADLRTPQDLNQM	-1.940	1.715 -0.225
238	HTV	KYRLKHLVADLRIRPGGKKADLRTPQDLNQM	-1.150	1.295 0.145
239	HTV	HYMLKHIVWADLRIRPGGKKADLRTPQDLNQM	-0.910	1.490 0.580
240	HTV	HYMLKHIVWADLRIRPGGKKADLRTPQDLNQM	-0.910	1.490 0.580
241	HTV	KYMLKHIVWADLRIRPGGKKADLRTPQDLNQM	-0.150	0.935 0.785
242	HTV	KYRLKHLVADLRIRPGGKKADLRTPQDLNQM	-0.150	0.935 0.785

CHAPTER FOUR: RESULTS AND DISCUSSION

4.0 Introduction

This chapter presents the results and discusses the results according to researcher's specific objectives. The objectives of the project were to compare epitope lengths for pathogens with vaccines and pathogens without vaccines, suggest most appropriate length(s) of epitopes for use in a polyepitopes for malaria, HIV-I and T.B, to develop an epitopes database, to optimize polyepitope design and to develop a web-based application prototype.

4.1 Results

Epitopes data was collected in December 2011 from two sources. The main source of data was the immune epitope database (IEDB) <http://iedb.org>. The epitopes data were collected from this source excluding the HIV epitopes. The HIV epitopes data were collected from the HIV molecular immunological database. The B –cell epitopes were filtered by species criteria (human only) and the analyzed.

The search criteria used for epitopes includes specifying B –cell or T – cell epitopes, the host organism and the source organism. If the host name or source organism was not on the list a search was done using the pathogen name from the NCBI database which is linked to IEDB to facilitate the search.

Pathogens were divided into two clusters – those pathogens that have vaccines and pathogens that do not have vaccines. Seven out of fourteen pathogens with vaccines were selected from NIAID A-C list. For each pathogen selected, B- cell epitopes and T- Cell epitopes were downloaded from iedb and HIV websites. The epitopes were downloaded in CSV format and saved in Microsoft Excel for B –cell and T-cell. For each pathogen B –Cell and T- cell epitopes were saves in a separate workbook.

The epitopes were analyzed using `=LEN ( )` and `=COUNTIF ($m$2: $m$n, p)` Excel functions. The first function was used to calculate the epitope length of each epitope downloaded. The second function was used to calculate the frequencies of epitopes whose absolute range was `$m$2: $m$n` and lengths `p1, p2 ...px`.

4.2 Epitopes data for organisms with vaccines

The number of epitopes for organisms with vaccines that were downloaded from IEDB and HIV molecular immunological database were two thousand seven hundred and twenty four (2724) as the analysis results show in tables 2 and 3 and figures 15 and 16.

4.2.1 T-cell epitopes data for organisms with vaccines

One thousand eight hundred and seventy four (1874) T-cell epitopes for organisms without vaccines were downloaded and the results are shown in table 2 and figure 15.

Table 2: T- Cell Epitopes data for organisms with vaccines

Microbe	9 mer		15 mer		20 mer		25 mer		C	Total
Hepatitis B	F	%	F	%	F	%	F	%	133 50%	266
	48	18.05	59	22.18	25	9.40	1	0.38		
Vaccinia	128	44.14	27	9.31	2	0.69	00	00	157 54%	290
Clostridium Tetani	29	13.39	5	2.33	132	61.40	00	00	166 77%	215
Hepatitis A	8	47.06	0	0	8	47.06	0	0	16 94%	17
Salmonella Typhi	4	100	0	0	0	0	0	0	4 100%	4
Mycobacterium Tuberculosis	176	18.74	116	12.35	281	29.93	93	9.90	666 71%	939
Measles	20	13.99	65	45.45	31	21.68	00	00	106 74%	143
TOTAL										1874

Key

F – Frequency

% - percentage of epitopes

Total – the sum of epitopes

Mer- the length of epitope

e.g.

- 5mer epitope has five amino acids,
- 6 mer epitope has six amino acids etc

C- Cumulative total

The number T cell epitopes for *Hepatitis B*, *Vaccinia*, *Clostridium Tetani*, *Hepatitis A*, *Salmonella Typhi*, *Mycobacterium Tuberculosis* and *Measles* were one thousand eight hundred and seventy four (1874). Seven out of twenty three pathogens with vaccines were selected from NIAID A-C list; only one subtype of each microbe was selected.

Two hundred and sixty six (266) T-cell *Hepatitis B* epitopes were downloaded. Upon careful scrutiny of the epitopes, the researcher discovered that the epitopes appeared in clusters and the main clusters formed around 9mer, 15mer, 20mer and 25mer.

For *Hepatitis B* T cell epitopes , 59 epitopes whose length was 15 amino acids(15mer) this accounted for 22.18 % , 9mer were 48(18.08%), 20mer were 25(9.40%) and 25mer was 1(0.38%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 133(50%).

The frequencies and percentages of *Vaccinia* epitopes were as follows: 9mer were 128(44.14%) ,15mer 27 (9.31%), 20 mer were 2(0.69%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 157(54%).

The frequencies and percentages of *Clostridium Tetani* epitopes were as follows: 9mer were 29(13.39%) ,15mer 5 (2.33%), 20 mer were 132(61.40%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 166(77%).

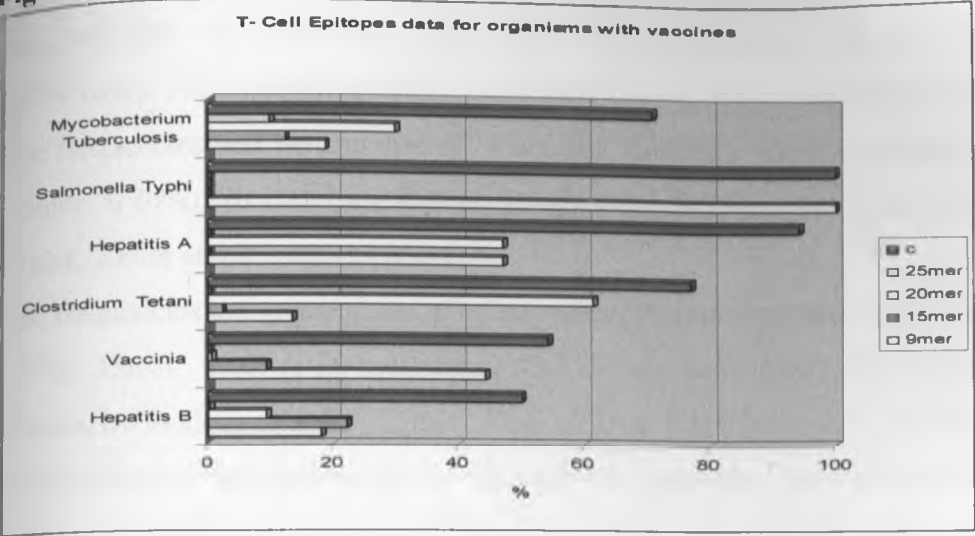
The frequencies and percentages of *Hepatitis A* epitopes were as follows: 9mer were 8(47.06%) ,15mer 0 (0%), 20 mer were 8(47.06%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 16(94%).

The frequencies and percentages of *Salmonella Typhi* epitopes were as follows: 9mer were 4(100%) ,15mer 0 (0%), 20 mer were 0(0%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 4(100%).

The frequencies and percentages of *Mycobacteria Tuberculosis* epitopes were as follows: 9mer were 176(18.74%), 15mer 116 (12.35%), 20 mer were 281(29.93%), 25mer were 93(9.90%).

The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 666(71%) and The frequencies and percentages of measles epitopes were as follows: 9mer were 20(13.99%) ,15mer 65(45.45%), 20 mer were 31(21.68%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 106(74%).

Figure 15: T- Cell Epitopes data for organisms with vaccines



4.2.2 B –cell Epitopes data for organisms with vaccines

Eight hundred and fifty (850) B-cell epitopes for organisms with vaccines were downloaded and the results are shown in table 3 and figure 16.

Table 3: B cell Epitopes data for organisms with vaccines

Microbe	9 mer		15 mer		20 mer		25 mer		23 mer	C	Total
	F	%	F	%	F	%	F	%			
Hepatitis B	5	4.42	14	12.39	9	7.96	16	14.16		44 39%	113
Vaccinia	0	0	0	0	0	0	0	0	0	0	0
Clostridium Tetani	0	0	0	0	0	0	0	0	43 91.49%	43 91%	47
Hepatitis A	0	0	3	3.39	77	89.53	0	0	0	80 93%	86
Salmonella Typhi	3	100	0	0	0	0	0	0	0	3 100%	3
Mycobacterium Tuberculosis	2	0.49	229	55.72	59	14.36	7	1.17	13 1.38%	295 71%	441
Measles	0	0	152	95.00	0	0	0	0	0	152 95%	160
TOTAL											850

For *Hepatitis B*, the B cell epitopes, 14 epitopes whose length was 15 amino acids (15mer) this accounted for 12.39 %, 9mer were 5(4.42%), 20mer were 9(7.79%) and 25mer was 16(14.16%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 133(39%).

The frequencies and percentages of *Vaccinia* epitopes were as follows: 9mer were 0(0%), 15mer 0 (0%), 20 mer were 0(0%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 0(0%).

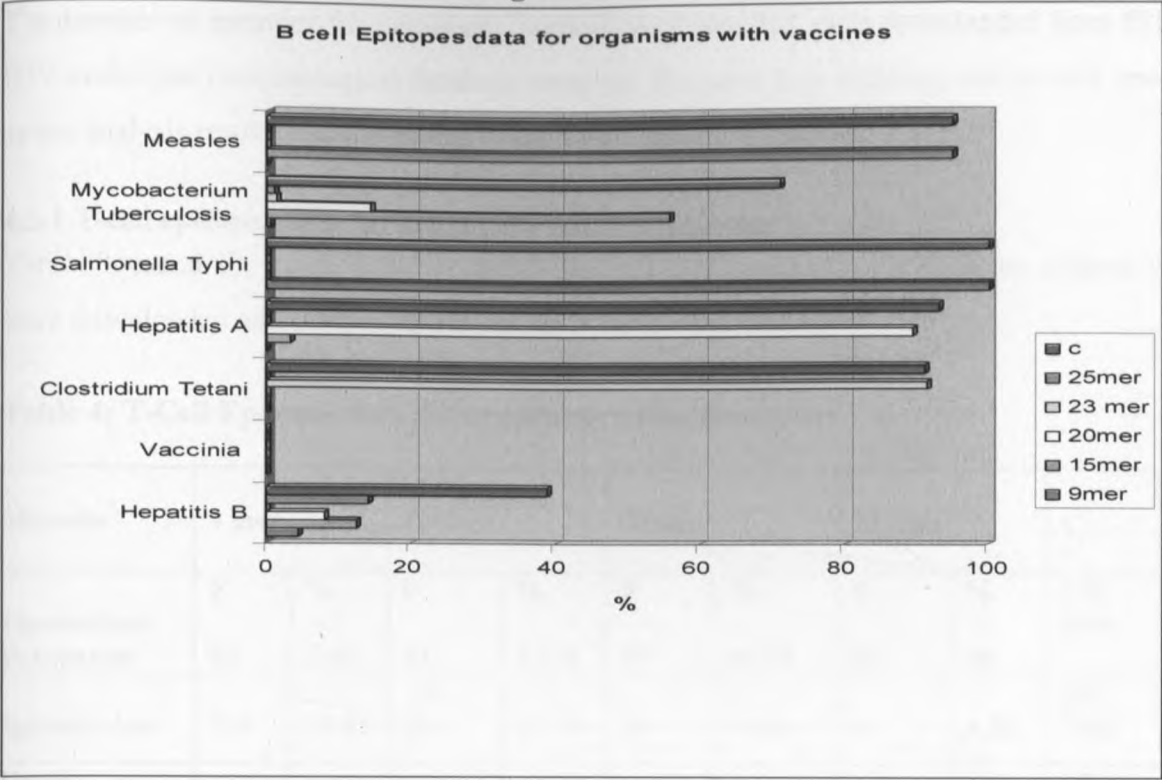
The frequencies and percentages of *Clostridium Tetani* epitopes were as follows: 9mer were 0(0%), 15mer 0 (0%), 20 mer were 0(0%), 23mer were 34(91.49%) ,25mer were 0(0%) .The cumulative total(C) of 9 mer, 15mer, 20mer, 23mer and 25mer was 43(91%).

The frequencies and percentages of *Hepatitis A* epitopes were as follows: 9mer were 0(0%), 15mer 3 (3.39%), 20 mer were 77(89.53%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 80(93%).

The frequencies and percentages of *Salmonella Typhi* epitopes were as follows: 9mer were 3(100%) ,15mer 0 (0%), 20 mer were 0(0%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 3(100%).

The frequencies and percentages of *Mycobacteria Tuberculosis* epitopes were as follows: 9mer were 2(0.49%) ,15mer 229 (55.72%), 20 mer were 59(14.36%), 25mer were 7(1.17%), 23mer 13(1.38%) . The cumulative total(C) of 9 mer, 15mer, 20mer, 23mer and 25mer was 295(71%) and The frequencies and percentages of *measles* epitopes were as follows: 9mer were 0(0%) ,15mer 152(95.00%), 20 mer were 0(0%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 152(95%).

Figure 16: B cell Epitopes data for organisms with vaccines



4.3 Epitopes data for organisms without vaccines

The number of epitopes for organisms without vaccines that were downloaded from IEDB and HIV molecular immunological database were six thousand four hundred and seventy one (6471) as the analysis results show in tables 4 and 5 and figures 17 and 18.

4.3.1 T-cell epitopes data for organisms without vaccines

Three thousand six hundred and seventy (3670) T-cell epitopes for organisms without vaccines were downloaded and the results are shown in table 4 and figure 17.

Table 4: T-Cell Epitopes data for organisms without vaccines

Microbe	9 mer		15 mer		20 mer		25 mer		C	Total Epitope s
	F	%	F	%	F	%	F	%		
Plasmodium Falciparum	12	5.02	41	17.15	97	40.59	00	00	150 63%	239
Epistein- barr	159	39.85	71	17.79	56	14.04	6	1.50	295 74%	399
Dengue	56	34.36	58	35.58	10	6.13	0	0	124 76%	163
Hepatitis C	250	17.78	371	26.39	205	15.58	23	1.67	849 60%	1406
HIV-1	661	49.99	00	00	00	00	00	00	661 49.9%	1341
Mycobacteriu m Bovis	2	1.64	58	47.54	27	22.13	00	00	87 71%	122
TOTAL										3670

The total number of T cell epitopes downloaded for pathogens without vaccines was three thousand six hundred and seventy (3670).

The parasite that cause malaria –*Plasmodium Falciparum* , T cell epitopes , 9mer epitopes were 12(5.05%), 15mer were 41(17.15%), 20mer were 97(40.59%) and 25mer was 0(0%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 150(63%).

The frequencies and percentages of *Epstein-Barr* epitopes were as follows: 9mer were 159(39.85%), 15mer 71 (17.79%), 20 mer were 56(14.04%), 25mer were 6(1.50%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 295(74%).

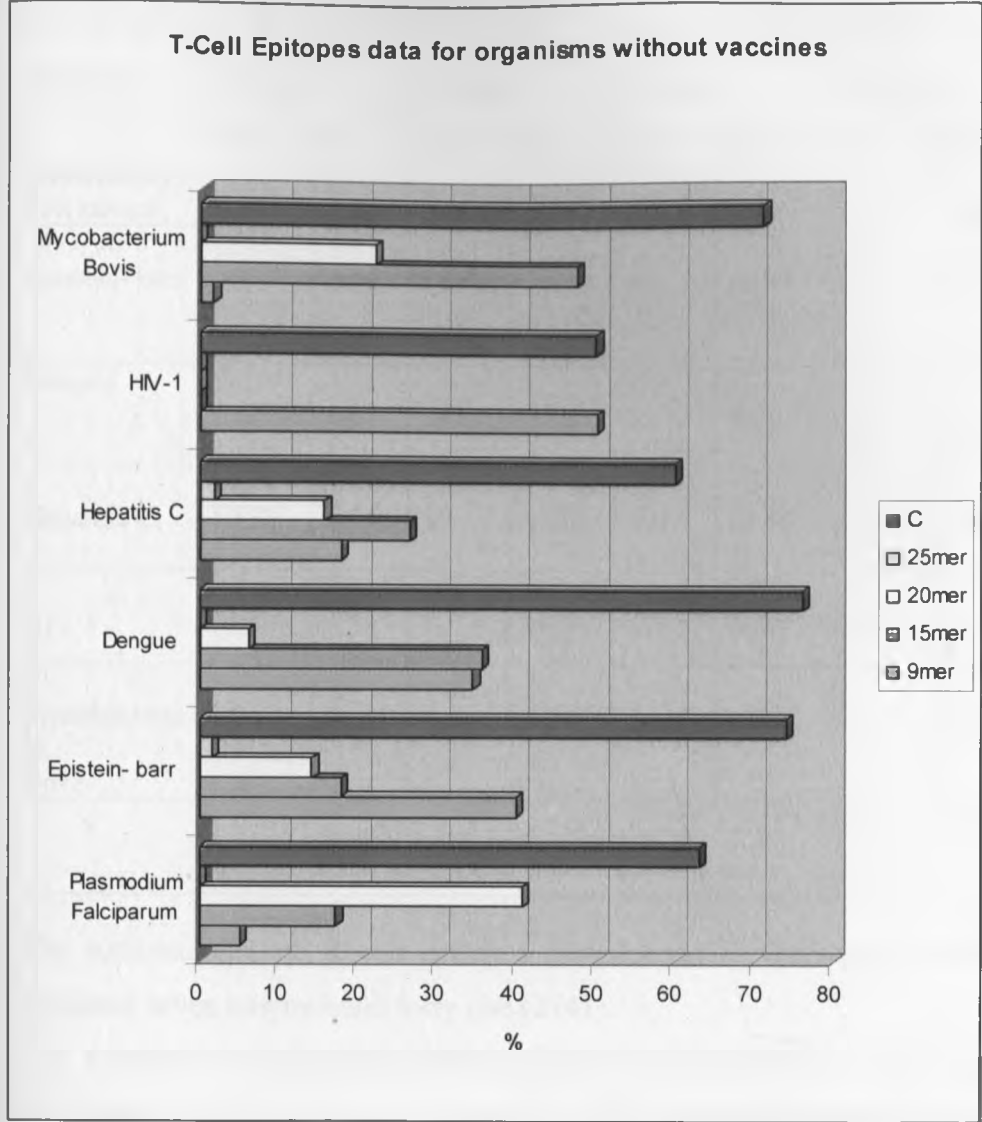
The frequencies and percentages of *Dengue* epitopes were as follows: 9mer were 56(34.36%), 15mer 58 (35.58%), 20 mer were 10 (6.13%), 25mer were 0(0%).The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 124(76%).

The frequencies and percentages of *Hepatitis C* epitopes were as follows: 9mer were 250(17.78%), 15mer 371 (26.39%), 20 mer were 205(15.58%), 25mer were 23(1.67%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 16(94%).

The frequencies and percentages of *HIV-1* epitopes were as follows: 9mer were 661(49.99%), 15mer 0 (0%), 20 mer were 0(0%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 661(49.9%).

The frequencies and percentages of *Mycobacterium Bovis* epitopes were as follows: 9mer were 2(1.64%), 15mer 58 (47.54%), 20 mer were 27(22.13%), 25mer were 0(0%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 87(71%).

Figure 17: T cell Epitopes data for organisms with vaccines



4.3.2 B-cell epitopes data for organisms without vaccines

Two thousand seven hundred and forty one (2741) B-cell epitopes for organisms without vaccines were downloaded and the results are shown in table 5 and figure 18.

Table 5: B cell Epitopes data for organisms without vaccines

Microbe	9 mer		15 mer		20 mer		25 mer		C	Total
	F	%	F	%	F	%	F	%		
Plasmodium Falciparum	20	3.95	105	20.57	24	4.74	24	4.74	173 34%	506
Epistein- barr	3	1.96	16	10.46	41	26.80	0	0	60 39%	153
Dengue	19	6.91	14	5.09	3	1.09	0	0	36 13%	275
Hepatitis C	152	8.94	69	4.06	391	22.99	35	2.06	647 38%	1701
HIV-1	4	5.13	8	10.26	5	6.41	0	0	17 22%	78
Mycobacteriu m Bovis	0	0	1	3.57	4	14.29	0	0	5 18%	28
TOTAL										2741

The total number of B cell epitopes downloaded for pathogens without vaccines was two thousand seven hundred and forty one (2741).

The *Plasmodium Falciparum*, 9mer epitopes were 20(3.95%), 15mer were 105(20.57%), 20mer were 24(4.74%) and 25mer was 24(4.74%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 173(34%).

The frequencies and percentages of *Epstein-Barr* epitopes were as follows: 9mer were 3(1.96%), 15mer 16 (10.46%), 20 mer were 41 (26.80%), 25mer were 0(0%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 60(39%).

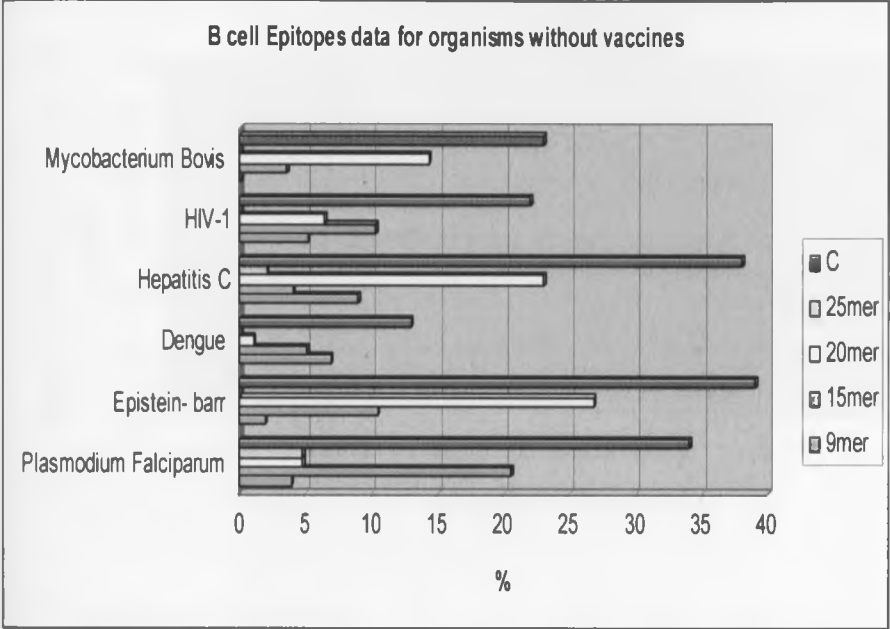
The frequencies and percentages of *Dengue* epitopes were as follows: 9mer were 19(6.91%), 15mer 14 (5.09%), 20 mer were 3 (1.09%), 25mer were 0(0%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 36(13%).

The frequencies and percentages of *Hepatitis C* epitopes were as follows: 9mer were 152(8.94%), 15mer 69 (4.06%), 20 mer were 391(22.99%), 25mer were 35(2.06%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 647(38%).

The frequencies and percentages of *HIV-1* epitopes were as follows: 9mer were 4(5.13%), 15mer 8 (10.26%), 20 mer were 5(6.41%), 25mer were 0(0%) . The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 17(22%).

The frequencies and percentages of *Mycobacterium Bovis* epitopes were as follows: 9mer were 0(0%), 15mer 1 (3.57%), 20 mer were 4(14.29%), 25mer were 0(0%). The cumulative total(C) of 9 mer, 15mer, 20mer and 25mer was 5(18%).

Figure 18: B cell Epitopes data for organisms without vaccines



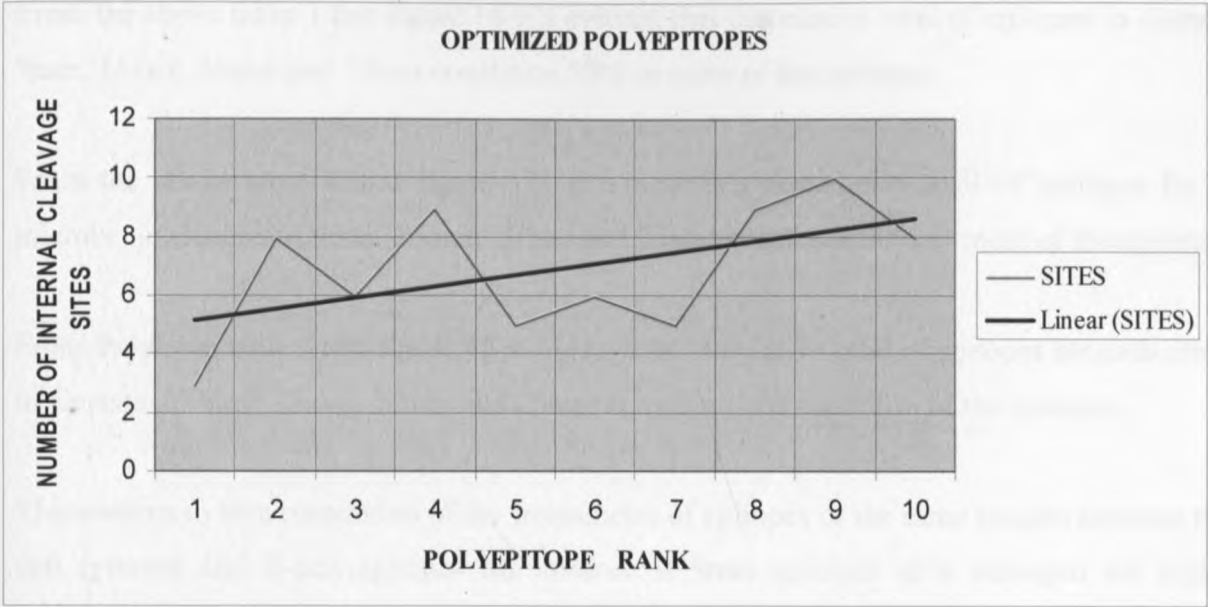
The percentages of epitopes of lengths 9mer, 15mer, 20mer and 25mer are less than 50% of the B cell epitopes in circulation as shown if figure 18. The cumulative totals of the stated clusters for each pathogen is denoted C and is shown on the blue bars.

4.4 Optimized polyepitopes

The optimized polyepitopes that were generated by the prototype were ranked using the total score and the top ten polyepitopes were selected and tested on the NetChop 3.1 server. The NetChop 3.1 makes preateasome predictions based artificial neural networks. The researcher choose this tool to validate polyepitope results generated by the prototype because it available on the World Wide Web and makes reliable predictions, however the tool measures only one of the two variables implemented on the prototype. The results are shown in figures 19 and 20.

For each of the top ten polyepitopes selected the internal cleavage sites were counted. The best polyepitope was assigned the first position, the second, second position and so on as shown above. It was expected that as the quality of the polyepitope increases the number of internal cleavage sites reduce and tend to zero. The best polyepitopes had only three internal cleavage sites. The best polyepitopes should not have any internal cleavage sites (cleavage sites within each epitope). This was consistent with the results generated by NetChop 3.1 server –this is shown by the trend line -linear sites in figure 18.

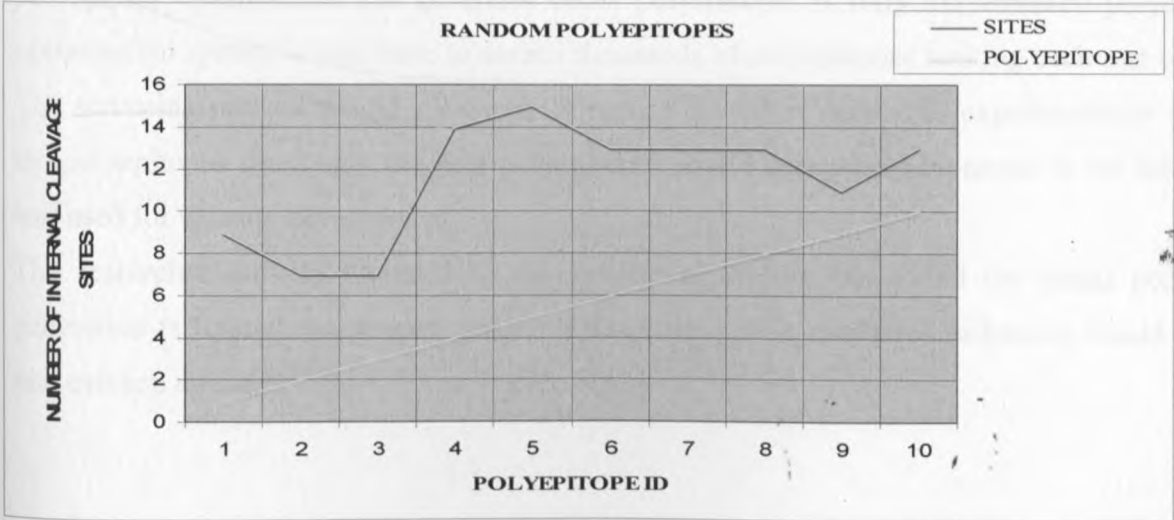
Figure 19: Optimized polyepitopes



4.5 Random polyepitopes

Randomly generated polyepitopes were also tested on the NetChop 3.1 server and the results are shown in figure 19 shown below.

Figure 20: Random polyepitopes



The average number of internal cleavage sites for the top ten optimized polyepitopes is 6.9 and the ten random polyepitopes have 11.5, this is consistent with the expected results- polyepitope optimization should reduce internal cleavage sites to zero.

4.6 Discussion

From the above table 2 and figure 15 it's clear that cumulative total of epitopes in clusters of 9mer, 15mer, 20mer and 25mer constitute 50% or more of the epitopes.

From the above table 3 and figure 16 it's evident that cumulative total of epitopes in clusters of 9mer, 15mer, 20mer and 25mer constitute 50% or more of the epitopes.

From the above table 4 and figure 17 it's clear that cumulative total of epitopes for each microbe, in clusters of 9mer, 15mer, 20mer and 25mer constitute 50% or more of the epitopes.

From the above table 5 and figure 18 it's clear that cumulative total of epitopes for each microbe in clusters of 9mer; 15mer, 20mer and 25mer constitute less than 50% of the epitopes.

There seems to be a correlation of the frequencies of epitopes of the same lengths between the T-cell epitopes and B-cell epitopes for instance if 9mer epitopes of a pathogen are high the corresponding 9mer B cell epitopes tend to be high for the pathogens which have vaccines.

Polypeptide optimization prototype generated polyepitopes with fewer internal cleavage sites compared the randomly generated polyepitopes. The average internal cleavage sites for the optimized polyepitopes was 6.9 and for the random polyepitopes 11.5. This indicates that polyepitope optimization can generates better polyepitope. A fully implemented polyepitopes optimization system would have to screen thousands of polyepitopes ranking each one of them. The screening process would substantially reduce the effort needed to experimentally validate the polyepitopes since only the best polyepitopes would best would be tested in the laboratory and used for vaccine development.

The researcher initially intended to use genetic algorithms but found the initial population generation is limited to epitopes generated from the public databases mutations would lead to non existent epitopes.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

5.0 Introduction

This chapter presents the conclusion, recommendations and limitations according to researcher's specific objectives. The objectives of the project were to compare epitope lengths for pathogens with vaccines and pathogens without vaccines, suggest the most appropriate length(s) of epitopes for use in a polyepitopes for malaria, HIV-I and T.B, to develop an epitopes database, to optimize polyepitope vaccine design and to develop a web-based application prototype.

5.1 Conclusion

The epitopes occur in clusters and the main clusters are clusters contain epitopes of the following lengths 9mers, 15mers, 20mers and 25mers. The clusters of 9mer, 15mer, 20mer and 25mer are important in the immune system response generation.

This suggests that the immune system memory remembers pathogenic attacks using the clusters stated above. This implies that polyepitope vaccines constructs design could be made better by taking into account the epitope clusters, however further research is needed to establish finer detail of epitope cluster.

Generally, the epitopes of pathogens with vaccines have more than 50% of the epitopes in circulation being made up of lengths 9mer, 15mer, 20mer and 25mer and also have 50% of the B cell epitopes of the same lengths. The clusters of epitopes of pathogens without vaccines constitute more than 50% percent of B cell epitopes do not feature the main epitope clusters stated above that is 9mers, 15mers, 20mers and 25mers. The epitopes of pathogens without vaccines have more clusters compared to those with vaccines

The immune system response is made better when both T- cell and B- cell epitopes are involved in generating immune response. The absence or representation of 9mers, 15mers, and 20mers and 25mers epitopes in pathogens without vaccines may suggest the unavailability of vaccines for the pathogens.

Polyepitopes optimization algorithms can be used to create better polyepitopes which can be use to make better epitope-based vaccines at a lower cost and within a shorter time.

5.2 Recommendations

Epitopes clusters seem to play a role in the generation of immune response and also in the creation of polyepitope vaccine design, further research is needed on epitope clusters and the immune signal generation and validation of results in wet lab research.

There is to search for 9mer, 15mer 20mer and 25mer epitopes for malaria (*plasmodium falciparum*), HIV-1 and mycobacterium tuberculosis (TB)

The need to carry out collaborative research by biological and computer scientist on the design of polyepitopes vaccine constructs and other types of vaccines

5.3 Limitations

The unavailability of publicly accessible tools integrate proteasome cleavage, TAP binding and MHC binding predictions that can be used to test the quality of polyepitope constructs/vaccines. The of polyepitope construct validation requires wet lab experiments which were not conducted in this research.

Time was also limited and the researcher decided to develop a prototype which though functions lacks some feature-for instance security features and incorporates minimal variables for polyepitope quality computation.

REFERENCES

- Altuvia, Y. and Margalit, H., Sequence signals for generation of antigenic peptides by the proteasome: implications for proteasomal cleavage mechanism. *J. Mol. Biol.* 2000. 295: 879–890.
- Andre, F. How the research-based industry approaches vaccine development and establishes priorities. *Dev Bio. (Basel)*, 110:25-29 2002.
- Blithe M.J, Doytchnova I.A, and Flower D. R, 2002, Jenpep: a database of quantitative functional peptide data for immunology. *Bioinformatics*, 18:433-439
- Devis V.A, 2009 PolyCTLDesigner < <http://tepredict.sourceforge.net/PolyCTLDesigner.html>>
- Encarta (2005) Microsoft Encyclopaedia, Microsoft Corporation. 1993-2004
- Feason D.T and Locksley RM. 1996, The instructive role of innate immunity in the acquired immune response. *Science*, 272:50-53
- Flower D.R, McSparron H, Blyth M.J, 2003, Computational vaccinology: quantitative approaches *Norvatis found.symp.*, 254: 102-120.
- Kesmir, C., Nussbaum, A. K., Schild, H., Detours, V. and Brunak, S., Prediction of proteasome cleavage motifs by neural networks. *Protein. Eng.* 2002. 15: 287–296.
- Korber T.M, Brander C, Haynes B.F, 2001a. *HIV Molecular Immunology 2001*. Theoretical Biology and Bio-physics group, Los Alamos National Laboratory, Los Alamos, NM,
- Liliana F., Bjarni H., Oliver K., Russell S., Stephen H., and Sorin I. 2003 Epitope Prediction Algorithms for Peptide-based Vaccine Design: Proceedings of the Computational Systems Bioinformatics (CSB'03) computer society

- Mette V., Claus L., Kasper L., Søren Buus, Søren Brunak, Ole Lund and Morten N. An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions *Eur. J. Immunol.* 2005. 35: 2295–2303
- Metropolis N, Rosenbluth A.W, Teller A.H and Teller E. 1953, Equation of state calculation by fast computing machines. *J. chem. Phys.*, 21:1087-1092 NIAID, 2000, *The Jordan report 2000: accelerated development of vaccines* <http://www.niaid.nih.gov/publications/pdf/jordan.pdf>
- Mo, X. Y., Cascio, P., Lemerise, K., Goldberg, A. L. and Rock, K., Distinct proteolytic processes generate the C and N termini of MHC class I-binding peptides. *J. Immunol.* 1999. 163: 5851–5859.
- Lund O, Morten N, Claus L, Can K and Søren B, 2005, *Immunological Bioinformatics*, MIT press, London.
- Peters, B., Bulik, S., Tampe, R., Endert, P. M. V. and Holzhutter, H. G., Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J. Immunol.* 2003. 171: 1741–1749.
- RAC, 2002, appendix B, http://www4d.nih.gov/oba/rac/guidelines_02/APPENDIX_b.htm
- Rammensee J. H, Emerich N.P, Bachor A.O and Stevanovic S, 1999, SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, 50:213-219.
- Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A. And Stevanovic, S., SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 1999. 50: 213–219.
- Richard Feynman ,1985 ,Surely You're Joking Mr. Feynman! Bantam
- Suhrbier ,A 2002, 'Polytope vaccines for the codelivery of multiple CD8 T-cell epitopes', *Expert rev. Vaccines*, 1:207-213.
- Toes RE, Nussbaum AK, Degermann S, Schirle M, Emmerich NPN, et al. (2001) Discrete

cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J Exp Med* 194: 1-12.

Toussaint NC, Doñnes P, Kohlbacher O (2008) A Mathematical Framework for the Selection of an Optimal Set of Peptides for Epitope-Based Vaccines. *PLoS Comput Biol* 4(12): e1000246. doi:10.1371/journal.pcbi.1000246

Thomas C and Carolyn B, 2005 *Database systems: a practical approach to design, implementation and management*, Addison Wesley, England

WHO, 2004a, the world health report, changing history,
<http://www.who.int/entity/whr/2004/annex/topic/en/annex_2_en.pdf>

WHO, 2004b, the world health report, changing history,
<http://www.who.int/entity/whr/2004/en/01-contents_en.pdf>

Whitten JL, Lonnie D, and Kelvin C, 2001, *System analysis and design methods 5th ed*, McGraw Hill, New York.

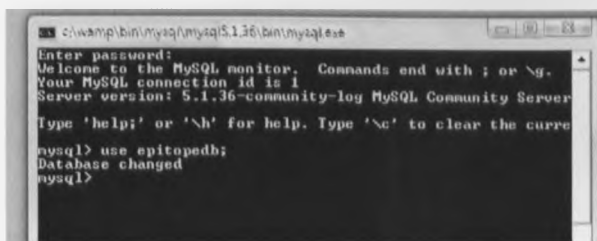
APPENDIX A: USER MANUAL

1. Install the software

- i) Download and install the Wampserver 2.0i. install micomedia dreamweaver 8 on a computer running windows xp or vista operating systems
- ii) Copy the folder named EPITOPEDB and paste in the folder in C:\wamp\ www
- iii) double click the wapserver icon that appears on the desktop

2. Create Mysql Database

- i) Run the epitopdb.sql script on the Mysql placed in epitopeDb folder



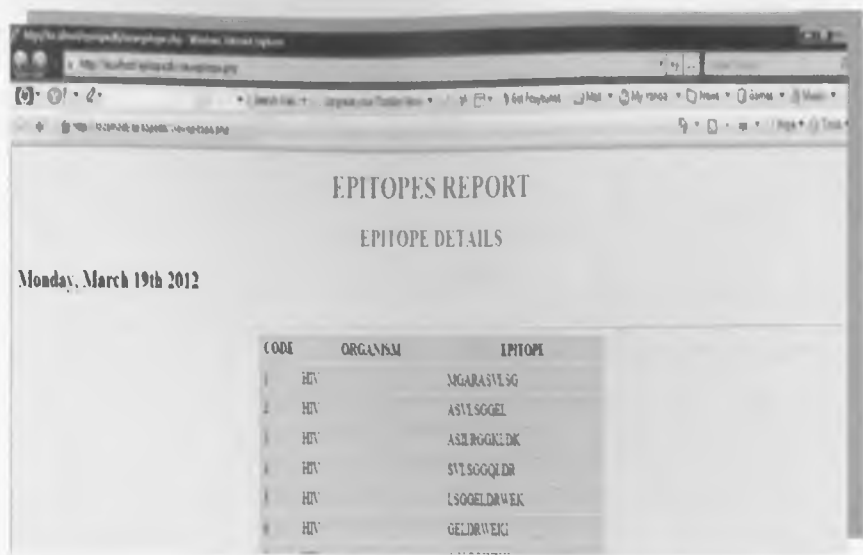
3. Start the prototype

- i) Open a web browser e.g. Internet explorer
- ii) Type the url for the index page on the address bar:
<http://localhost/epitopedb/index.php>
- iii) Press enter



4. Select the require script by clicking on the links on the home page

Click **viewepitope** link produce the following report.



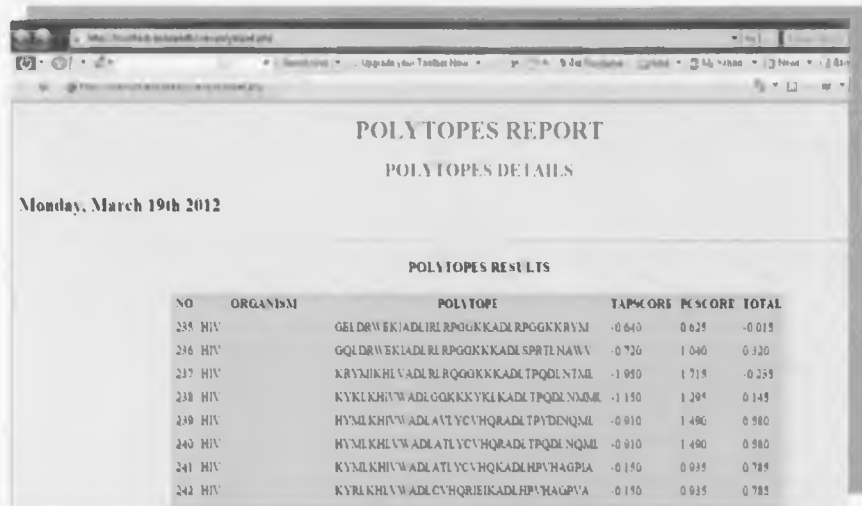
EPITOPES REPORT

EPITOPE DETAILS

Monday, March 19th 2012

CODE	ORGANISM	EPITOPE
1	HIV	NGARASVLSG
2	HIV	ASVLSGGEL
3	HIV	ASHPGGKEDK
4	HIV	SVLSGGQEDR
5	HIV	LSGGELDRWEK
6	HIV	GELDRWEK

Click **polyepitope** link to produce the following report.



POLYTOPES REPORT

POLYTOPES DETAILS

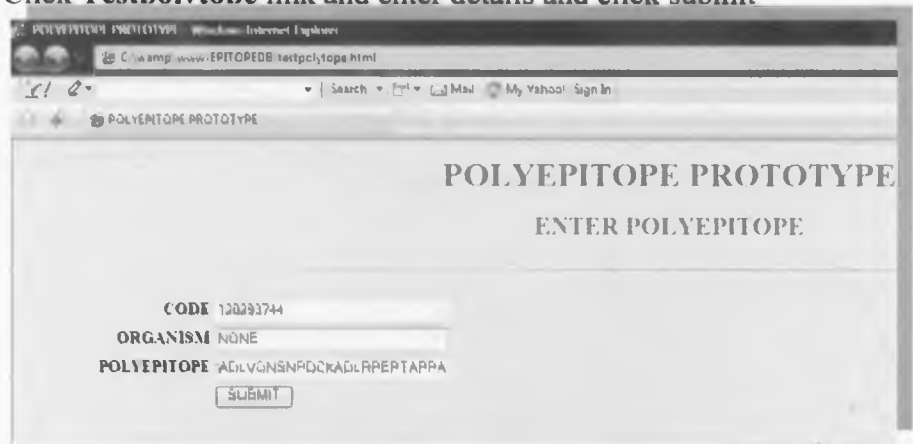
Monday, March 19th 2012

POLYTOPES RESULTS

NO	ORGANISM	POLYTOPE	TAPSCORE	PCSCORE	TOTAL
235	HIV	GELDRWEKADLRIRPGGKKADLRPGGKKRYM	-0.640	0.625	-0.015
236	HIV	GQEDRWEEKADLRIRPGGKKADLRSPRTNAAV	-0.720	1.040	0.320
237	HIV	KRYMLKHVWADLRIRPGGKKADLRTPQDLNQML	-1.050	1.715	-0.235
238	HIV	KYKMLKHVWADLRIRPGGKKADLRTPQDLNQML	-1.150	1.295	0.145
239	HIV	HYMLKHVWADLRIRPGGKKADLRTPQDLNQML	-0.910	1.485	0.580
240	HIV	HYMLKHVWADLRIRPGGKKADLRTPQDLNQML	-0.910	1.490	0.580
241	HIV	KYMLKHVWADLRIRPGGKKADLRTPQDLNQML	-0.150	0.935	0.785
242	HIV	KYMLKHVWADLRIRPGGKKADLRTPQDLNQML	-0.150	0.935	0.785

To Test dummy polyepitope

Click **Testpolvtone** link and enter details and click submit



POLYEPITOPE PROTOTYPE

ENTER POLYEPITOPE

CODE 120293744

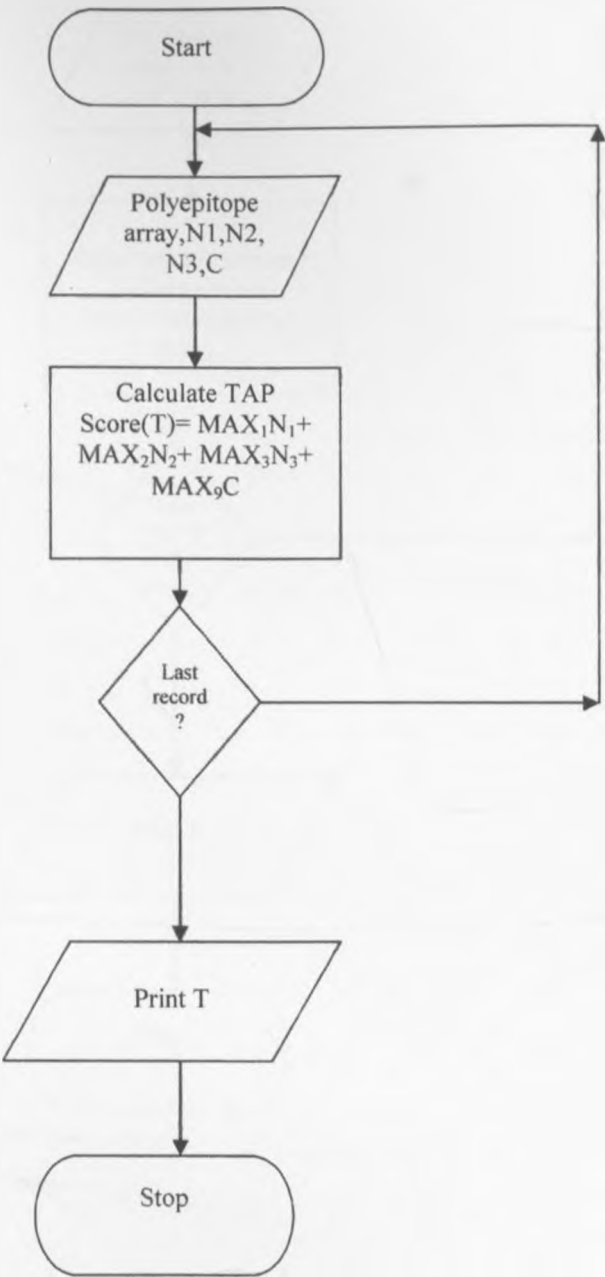
ORGANISM NONE

POLYEPITOPE ADLVGNSINPQCKADLRPEPTAPPA

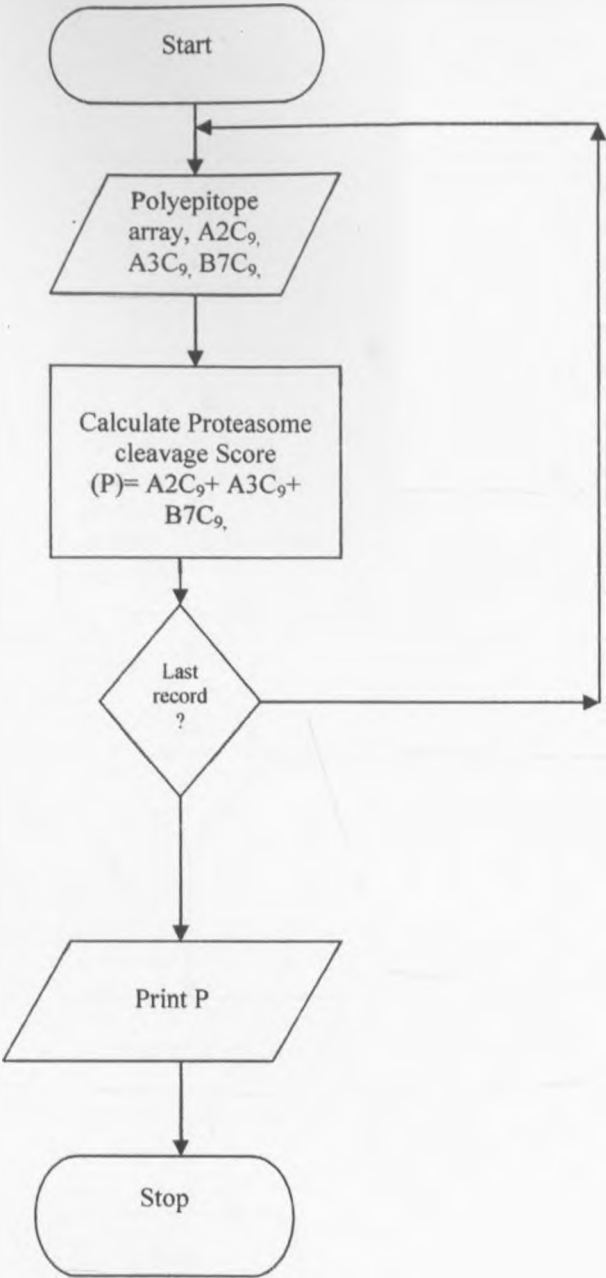
SUBMIT

APPENDIX B: OTHER DIAGRAMS

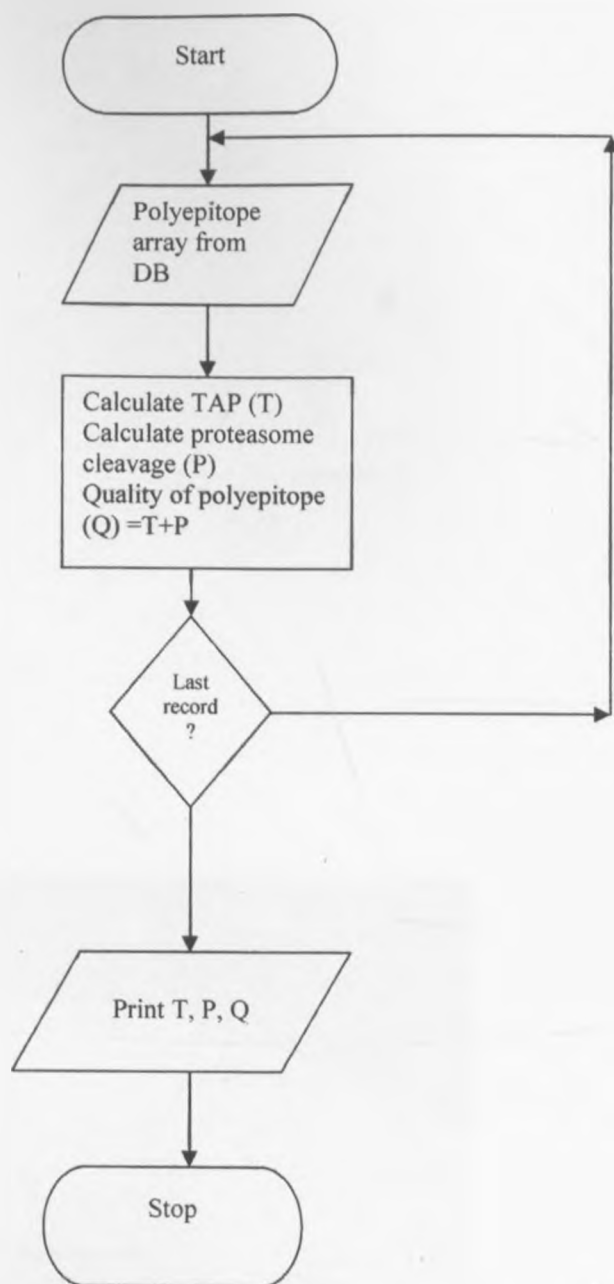
Flowchart 2: Calculating TAP score (T)



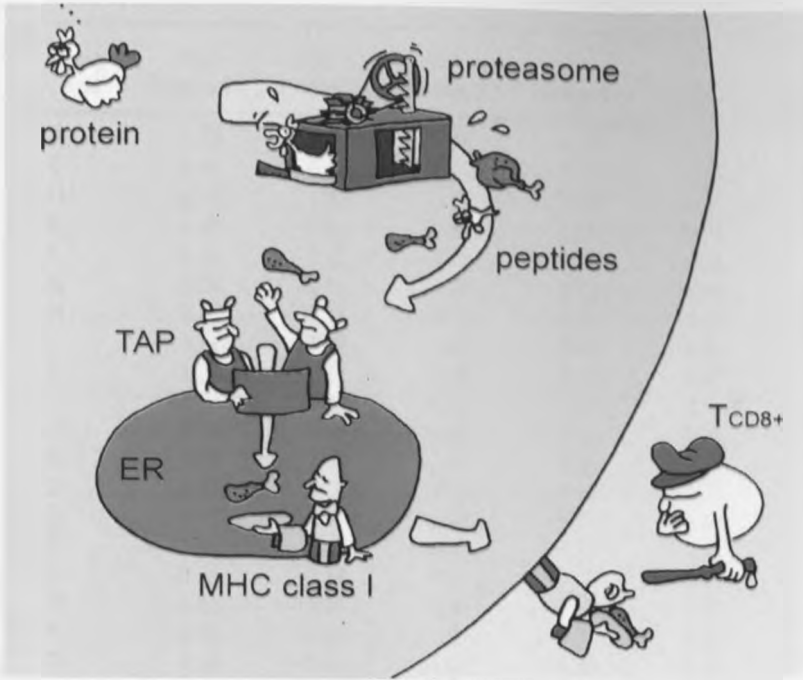
Flowchart 3: Calculating Proteasome cleavage score (P)



Flowchart 4: Polyepitope optimization

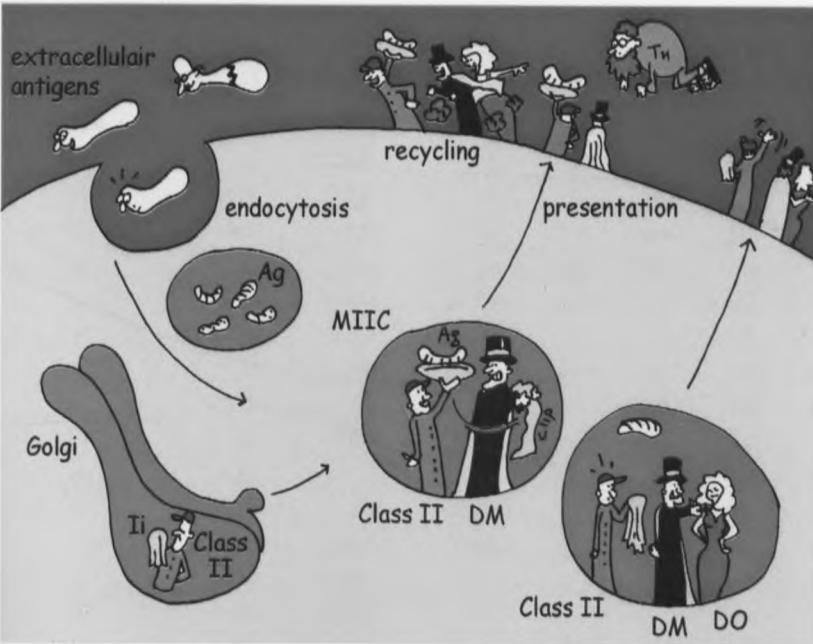


MHC class I pathway



Source: Lund (2005) and www

MHC class II pathway



Source: Lund (2005) and www

Peters et al (2003) Consensus Scoring Matrix

Amino Acid	(N1) Position 1	(N2) Position 2	(N3) Position 3	Position 4	Position 5	Position 6	Position 7	Position 8	(C) Position 9	Amino Acid
A	1.56	-0.25	0.10	0.24	-0.10	0.17	0.27	0.00	0.55	A
C	0.05	-0.01	-0.02	0.11	0.09	0.05	0.00	-0.13	0.00	C
D	1.37	1.42	1.83	-0.23	0.33	0.32	1.07	0.32	1.83	D
E	1.65	0.02	1.51	0.05	0.54	-0.13	0.64	0.44	1.58	E
F	1.03	0.45	1.05	-0.50	-0.26	0.08	-0.50	0.17	-2.52	F
G	0.28	1.14	1.70	0.45	0.66	0.12	1.41	-0.38	1.41	G
H	0.21	0.33	-0.23	-0.21	-0.11	-0.06	-0.19	0.39	0.55	H
I	0.11	0.49	-0.62	-0.09	-0.42	-0.75	-0.94	0.45	-0.52	I
K	-1.03	-0.41	0.09	-0.23	-0.08	-0.26	0.44	0.12	-0.45	K
L	-0.50	0.09	0.11	0.11	-0.34	0.02	-0.73	0.01	-0.94	L
M	-0.38	0.46	-0.55	0.35	0.26	0.30	0.64	-0.11	-0.29	M
N	-1.43	0.69	1.01	0.38	0.49	-0.27	0.16	0.33	1.33	N
P	1.43	3.00	0.22	-0.04	-0.72	-0.13	-0.84	0.03	-0.09	P
Q	0.47	-0.97	0.39	0.15	0.15	-0.07	0.34	0.26	0.12	Q
R	-1.34	-1.47	-0.42	-0.27	-0.32	-0.75	-0.09	-0.42	-1.47	R
S	-0.56	-0.34	0.11	0.27	0.45	0.31	0.87	-0.51	2.26	S
T	-0.12	0.04	0.43	0.23	0.43	0.49	0.39	-0.46	0.72	T
V	-0.49	-0.50	-0.71	0.27	0.37	-0.02	-0.29	0.10	-0.30	V
W	0.54	-0.64	-1.65	-0.18	0.78	0.31	-0.50	-0.63	-0.87	W
Y	0.50	0.67	-1.80	-0.18	0.13	0.28	-0.87	0.02	-2.91	Y

Toes et al (2001) table 1

AA

	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'	P6'
A	865	595	1,020	485	620	825	1,545	330	710	590	930	365
C	0	0	0	0	130	0	0	0	0	0	0	0
D	580	460	255	215	185	805	300	1,185	315	235	170	400
E	530	465	385	455	250	605	300	275	460	150	465	440
F	50	275	430	165	430	145	190	250	165	575	220	50
G	300	575	250	230	545	175	185	415	760	185	320	780
H	175	125	10	150	280	65	165	110	65	35	385	90
I	505	125	385	335	180	510	60	845	220	450	405	170
K	220	580	240	945	455	80	365	115	370	675	600	655
L	540	550	535	525	230	1,185	345	900	1,120	435	670	565
M	10	0	95	10	15	35	70	40	75	65	170	60
N	120	285	235	140	240	95	390	215	195	160	270	730
P	365	475	550	365	95	40	480	20	255	240	330	315
Q	45	260	25	20	340	60	75	60	155	305	35	75
R	125	45	45	425	200	275	595	95	155	145	80	60
S	575	365	380	90	535	395	500	215	170	530	490	555
T	735	400	140	365	365	140	150	430	140	360	250	105
V	270	380	365	920	840	450	275	410	695	705	355	695
W	40	15	35	30	20	30	0	50	0	0	50	10
Y	85	100	255	265	90	220	145	175	110	295	100	15
sum (total)	6,135	6,135	6,135	6,135	6,135	6,135	6,135	6,135	6,135	6,135	6,135	6,135

APPENDIX C: SAMPLE CODE

// COMPUTE EPITOPE SUPERTYPES

// Computation rules based on A2, A3 and B7 supertypes

```
<?php
```

```
//connect
```

```
$conn=mysql_connect("Localhost","root","") or die(mysql_error());
```

```
// select database
```

```
mysql_select_db("epitopedb",$conn) or die(mysql_error());
```

```
//query database and select only 9 MER epitopes
```

```
$result=mysql_query("SELECT epitopeid,linearsequence FROM epitope where  
length(linearsequence)=9",$conn) or die(mysql_error());
```

```
?>
```

```
<h3 style = "color: cyan" align="center">
```

```
    SUPERTYPES RESULTS</h3>
```

```
    <table border = "1" cellpadding = "3" cellspacing = "2" style = "background-color:  
#ADD8E6"
```

```
        align="center">
```

```
        <tr>
```

```
        <th>NO</th>
```

```
        <th>Sequence</th>
```

```
        <th>Supertype </th>
```

```
        </tr>
```

```
<?php
```

```
// fetch records in result set
```

```
    while($row=mysql_fetch_array($result)){
```

```
// put result set into hash/associative array
```

```
    $epitopeid = $row["epitopeid"];
```

```
    $linearsequence = $row["linearsequence"];
```

```
// process supertypes
```

```
        if($linearsequence[1]=="P"){
```

```
            $Supertype="B7";
```

```
        }
```

```
        elseif($linearsequence[8]=="H"){
```

```
            $Supertype="A3";
```

```
        }
```

```
        elseif($linearsequence[8]=="K"){
```

```
            $Supertype="A3";
```

```
        }
```

```
        elseif($linearsequence[8]=="R"){
```

```
            $Supertype="A3";
```

```
        }
```

```
        elseif($linearsequence[8]=="A"){
```

```
            $Supertype="A2";
```

```
        }
```

```
        elseif($linearsequence[8]=="C"){
```

```
            $Supertype="A2";
```

```
        }
```

```
        elseif($linearsequence[8]=="F"){
```

```
            $Supertype="A2";
```

```

    }
    elseif($linearsequence[8]=="I"){
        $Supertype="A2";
    }
    elseif($linearsequence[8]=="L"){
        $Supertype="A2";
    }
    elseif($linearsequence[8]=="M"){
        $Supertype="A2";
    }
    elseif($linearsequence[8]=="P"){
        $Supertype="A2";
    }
    elseif($linearsequence[8]=="V"){
        $Supertype="A2";
    }
    elseif($linearsequence[8]=="W"){
        $Supertype="A2";
    }
    else
    {
        $Supertype="O";
    }

```

// Print output

```

    echo"<tr>";
    echo"<td>$epitopeid</td>";
    echo"<td>$linearsequence</td>";
    echo"<td>$Supertype</td>";
    echo"</tr>";
    }
    echo"</table>";
mysql_close( $conn );
?>

```

// CALCULATE TAP SCORE

<body background="image001.jpg">

<h1 style = "color: #4974B3" align="center">POLYTOPES REPORT </h1> <h2 style = "color: #4974B3" align="center">POLYTOPES DETAILS</h2>

<h2><?php echo date('l, F dS Y');?></h2>

<hr width="100%" size="2" color="#FF0000" align="center">

<?php

//connect

\$conn=mysql_connect("Localhost","root","") or die(mysql_error());

// select database


```
mysql_select_db("epitopedb",$conn) or die(mysql_error());
//query database
$result=mysql_query("SELECT pid,organism,polytope FROM polytope4",$conn) or
die(mysql_error());
?>
```

```
<h3 style = "color: black" align="center">
```

```
POLYTOPES RESULTS</h3>
```

```
<table border = "1" cellpadding = "3" cellspacing = "2" style = "background-color:
#ADD8E6" align="center">
```

```
<tr>
```

```
<th>NO</th>
```

```
<th>ORGANISM</th>
```

```
<th>POLYTOPE</th>
```

```
<th>TAPSCORE</th>
```

```
<th>PCSCORE</th>
```

```
<th>TOTAL</th>
```

```
</tr>
```

```
<?php
```

```
// fetch records in result set
```

```
while($row=mysql_fetch_array($result)){
```

```
    $pid = $row["pid"];

```

```
    $organism = $row["organism"];

```

```
    $polytope = $row["polytope"];

```

```
    // process TAP scores

```

```
    // compute N1 OR position 1 score

```

```
    if($polytope[0]=="A"){

```

```
        $N1=-1.56;

```

```
    }

```

```
    elseif($polytope[0]=="C"){

```

```
        $N1=0.05;

```

```
    }

```

```
    elseif($polytope[0]=="D"){

```

```
        $N1=1.37;

```

```
    }

```

```

elseif($polytope[0]=="E"){
    $N1=1.65;
}

elseif($polytope[0]=="F"){
    $N1=1.03;
}

elseif($polytope[0]=="G"){
    $N1=0.28;
}

elseif($polytope[0]=="H"){
    $N1=0.21;
}

elseif($polytope[0]=="I"){
    $N1=-0.11;
}

elseif($polytope[0]=="K"){
    $N1=-1.03;
}

elseif($polytope[0]=="L"){
    $N1=-0.50;
}

elseif($polytope[0]=="M"){
    $N1=-0.38;
}

elseif($polytope[0]=="N"){
    $N1=-1.43;
}

elseif($polytope[0]=="P"){
    $N1=1.43;
}

elseif($polytope[0]=="Q"){
    $N1=0.47;
}

```

```

elseif($polytope[0]=="R"){
    $N1=-1.34;
}

elseif($polytope[0]=="S"){
    $N1=-0.56;
}

elseif($polytope[0]=="T"){
    $N1=-0.12;
}

elseif($polytope[0]=="V"){
    $N1=-0.49;
}

elseif($polytope[0]=="W"){
    $N1=0.54;
}

elseif($polytope[0]=="Y"){
    $N1=0.50;
}

else
{
    $N1=0;
}

```

// compute N2 score

```

if($polytope[1]=="A"){
    $N2=-0.25;
}

elseif($polytope[1]=="C"){
    $N2=-0.01;
}

elseif($polytope[1]=="D"){
    $N2=1.42;
}

elseif($polytope[1]=="E"){
    $N2=0.02;
}

```

```

    }

    elseif($polytope[1]=="F"){
        $N2=-0.45;
    }

    elseif($polytope[1]=="G"){
        $N2=1.14;
    }

    elseif($polytope[1]=="H"){
        $N2=0.33;
    }

    elseif($polytope[1]=="I"){
        $N2=-0.49;
    }

    elseif($polytope[1]=="K"){
        $N2=-0.41;
    }

    elseif($polytope[1]=="L"){
        $N2=0.09;
    }

    elseif($polytope[1]=="M"){
        $N2=-0.46;
    }

    elseif($polytope[1]=="N"){
        $N2=0.69;
    }

    elseif($polytope[1]=="P"){
        $N2=3.00;
    }

    elseif($polytope[1]=="Q"){
        $N2=-0.97;
    }

    elseif($polytope[1]=="R"){
        $N2=-1.47;
    }

```

```

elseif($polytope[1]=="S"){
    $N2=-0.34;
}

elseif($polytope[1]=="T"){
    $N2=-0.04;
}

elseif($polytope[1]=="V"){
    $N2=-0.50;
}

elseif($polytope[1]=="W"){
    $N2=-0.64;
}

elseif($polytope[1]=="Y"){
    $N2=-0.67;
}

else
{
    $N2=0;
}

```

// Compute N3

```

if($polytope[2]=="A"){
    $N3=-0.10;
}

elseif($polytope[2]=="C"){
    $N3=-0.02;
}

elseif($polytope[2]=="D"){
    $N3=1.83;
}

elseif($polytope[2]=="E"){
    $N3=1.51;
}

elseif($polytope[2]=="F"){
    $N3=-1.05;
}

```

```

elseif($polytope[2]=="G"){
    $N3=1.70;
}

elseif($polytope[2]=="H"){
    $N3=-0.23;
}

elseif($polytope[2]=="I"){
    $N3=-0.62;
}

elseif($polytope[2]=="K"){
    $N3=0.09;
}

elseif($polytope[2]=="L"){
    $N3=-0.11;
}

elseif($polytope[2]=="M"){
    $N3=-0.58;
}

elseif($polytope[2]=="N"){
    $N3=1.01;
}

elseif($polytope[2]=="P"){
    $N3=0.22;
}

elseif($polytope[2]=="Q"){
    $N3=0.39;
}

elseif($polytope[2]=="R"){
    $N3=-0.42;
}

elseif($polytope[2]=="S"){
    $N3=0.11;
}

elseif($polytope[2]=="T"){
    $N3=0.43;
}

elseif($polytope[2]=="V"){
    $N3=-0.71;
}

elseif($polytope[2]=="W"){

```

```

                                $N3=-1.65;
        }
        elseif($polytope[2]=="Y"){
                                $N3=-1.80;
        }
        else
        {
                                $N3=0;
        }

```

// compute C-terminal score

```

        if($polytope[32]=="A"){
                                $C=0.55;
        }
        elseif($polytope[32]=="C"){
                                $C=0.00;
        }
        elseif($polytope[32]=="D"){
                                $C=1.83;
        }
        elseif($polytope[32]=="E"){
                                $C=1.58;
        }
        elseif($polytope[32]=="F"){
                                $C=-2.52;
        }
        elseif($polytope[32]=="G"){
                                $C=1.41;
        }
        elseif($polytope[32]=="H"){
                                $C=0.55;
        }
        elseif($polytope[32]=="I"){
                                $C=-0.52;
        }
        elseif($polytope[32]=="K"){

```

```

        $C=-0.45;
    }
    elseif($polytope[32]=="L"){
        $C=-0.94;
    }
    elseif($polytope[32]=="M"){
        $C=-0.29;
    }
    elseif($polytope[32]=="N"){
        $C=1.33;
    }
    elseif($polytope[32]=="P"){
        $C=-0.09;
    }
    elseif($polytope[32]=="Q"){
        $C=0.12;
    }
    elseif($polytope[32]=="R"){
        $C=-1.47;
    }
    elseif($polytope[32]=="S"){
        $C=2.26;
    }
    elseif($polytope[32]=="T"){
        $C=0.72;
    }
    elseif($polytope[32]=="V"){
        $C=-0.03;
    }
    elseif($polytope[32]=="W"){
        $C=-0.87;
    }
    elseif($polytope[32]=="Y"){
        $C=-2.91;
    }

```



```

    }
    else
    {
        $C=0;
    }

// compute TAP SCORE
    $TAP=number_format(($N1+ $N2+$N3+ $C),3);
// calculate proteasome cleavage
// proteasome cleavage epitope 1 C-terminal score
    if($polytope[8]=="A"){
        $c_terminal1=0.825;
    }
    elseif($polytope[8]=="C"){
        $c_terminal1=0.000;
    }
    elseif($polytope[8]=="D"){
        $c_terminal1=0.805;
    }
    elseif($polytope[8]=="E"){
        $c_terminal1=0.605;
    }
    elseif($polytope[8]=="F"){
        $c_terminal1=0.145;
    }
    elseif($polytope[8]=="G"){
        $c_terminal1=0.175;
    }
    elseif($polytope[8]=="H"){
        $c_terminal1=0.065;
    }
    elseif($polytope[8]=="I"){
        $c_terminal1=0.510;
    }
    elseif($polytope[8]=="K"){

```

```

        $c_terminal1=0.080;
    }

    elseif($polytope[8]=="L"){
        $c_terminal1=1.185;
    }

    elseif($polytope[8]=="M"){
        $c_terminal1=0.035;
    }

    elseif($polytope[8]=="N"){
        $c_terminal1=0.095;
    }

    elseif($polytope[8]=="P"){
        $c_terminal1=0.040;
    }

    elseif($polytope[8]=="Q"){
        $c_terminal1=0.060;
    }

    elseif($polytope[8]=="R"){
        $c_terminal1=0.275;
    }

    elseif($polytope[8]=="S"){
        $c_terminal1=0.395;
    }

    elseif($polytope[8]=="T"){
        $c_terminal1=0.140;
    }

    elseif($polytope[8]=="V"){
        $c_terminal1=0.450;
    }

    elseif($polytope[8]=="W"){
        $c_terminal1=0.030;
    }

    elseif($polytope[8]=="Y"){

```

```

        $c_terminal1=0.220;
    }
    else
    {
        $c_terminal1=0;
    }
// proteasome cleavage epitope 2 C-terminal score

    if($polytope[20]=="A"){
        $c_terminal2=0.825;
    }
    elseif($polytope[20]=="C"){
        $c_terminal2=0.000;
    }
    elseif($polytope[20]=="D"){
        $c_terminal2=0.805;
    }
    elseif($polytope[20]=="E"){
        $c_terminal2=0.605;
    }
    elseif($polytope[20]=="F"){
        $c_terminal2=0.145;
    }
    elseif($polytope[20]=="G"){
        $c_terminal2=0.175;
    }
    elseif($polytope[20]=="H"){
        $c_terminal2=0.065;
    }
    elseif($polytope[20]=="I"){
        $c_terminal2=0.510;
    }
    elseif($polytope[20]=="K"){
        $c_terminal2=0.080;
    }

```

```

}
elseif($polytope[20]=="L"){
    $c_terminal2=1.185;
}
elseif($polytope[20]=="M"){
    $c_terminal2=0.035;
}
elseif($polytope[20]=="N"){
    $c_terminal2=0.095;
}
elseif($polytope[20]=="P"){
    $c_terminal2=0.040;
}
elseif($polytope[20]=="Q"){
    $c_terminal2=0.060;
}
elseif($polytope[20]=="R"){
    $c_terminal2=0.275;
}
elseif($polytope[20]=="S"){
    $c_terminal2=0.395;
}
elseif($polytope[20]=="T"){
    $c_terminal2=0.140;
}
elseif($polytope[20]=="V"){
    $c_terminal2=0.450;
}
elseif($polytope[20]=="W"){
    $c_terminal2=0.030;
}
elseif($polytope[20]=="Y"){
    $c_terminal2=0.220;
}

```

```

else
{
    $c_terminal2=0;
}

// proteasome cleavage epitope 3 C-terminal score
if($polytope[32]=="A"){
    $c_terminal3=0.825;
}

elseif($polytope[32]=="C"){
    $c_terminal3=0.000;
}

elseif($polytope[32]=="D"){
    $c_terminal3=0.805;
}

elseif($polytope[32]=="E"){
    $c_terminal3=0.605;
}

elseif($polytope[32]=="F"){
    $c_terminal3=0.145;
}

elseif($polytope[32]=="G"){
    $c_terminal3=0.175;
}

elseif($polytope[32]=="H"){
    $c_terminal3=0.065;
}

elseif($polytope[32]=="I"){
    $c_terminal3=0.510;
}

elseif($polytope[32]=="K"){
    $c_terminal3=0.080;
}

elseif($polytope[32]=="L"){

```

```

        $c_terminal3=1.185;
    }
    elseif($polytope[32]=="M"){
        $c_terminal3=0.035;
    }
    elseif($polytope[32]=="N"){
        $c_terminal3=0.095;
    }
    elseif($polytope[32]=="P"){
        $c_terminal3=0.040;
    }
    elseif($polytope[32]=="Q"){
        $c_terminal3=0.060;
    }
    elseif($polytope[32]=="R"){
        $c_terminal3=0.275;
    }
    elseif($polytope[32]=="S"){
        $c_terminal3=0.395;
    }
    elseif($polytope[32]=="T"){
        $c_terminal3=0.140;
    }
    elseif($polytope[32]=="V"){
        $c_terminal3=0.450;
    }
    elseif($polytope[32]=="W"){
        $c_terminal3=0.030;
    }
    elseif($polytope[32]=="Y"){
        $c_terminal3=0.220;
    }
    else
    {

```

```

        $c_terminal3=0;
    }

// calculate polytope proteasome cleavage
$pcscore=number_format(($c_terminal1+$c_terminal2+$c_terminal3),3);
//calculate total
$total=number_format(($TAP+$pcscore),3);
// print output
    echo"<tr>";
    echo"<td>$pid</td>";
    echo"<td>$organism</td>";
    echo"<td>$polytope</td>";
    echo"<td>$TAP</td>";
    echo"<td>$pcscore</td>";
    echo"<td>$total</td>";
    echo"</tr>";
    }
    echo"</table>";
mysql_close( $conn );
?>
</pre>
</body>

```