Application of GIS Spatial Interpolation Methods in Auto Insurance Risk Territory Segmentation and Rating

Samuel Ruugia
skruugia@gmail.com
and
Christopher Moturi
moturi@uonbi.ac.ke

Abstract

Evolution in the field of Geographic Information Systems (GIS) has given rise to sophisticated scientific techniques for collection, analysis and visualization of location based data. These GIS analysis processes are used to reveal some critical patterns of occurrences. Due to inaccurate analysis and covering of insurance risks in Kenya, several companies have closed down prompting the Insurance Regulatory Authority (IRA) and Association of Kenyan Insurers (AKI) set up maximum and minimum premium rates on insurance risks. The set premiums discounts are given to the insured based on records of their annual claims. The main problem here is that the rates cover the entire nation without considering the distribution of risk in various regions. The objective of the paper is to show that GIS can be used to analyse and generate auto insurance risk territories for insurance companies from which an insurance rating model can be developed. We used GIS analysis methods such as inverse distance weighting (IDW) interpolation, data smoothing and clustering techniques and data on auto insurance accidents and crime, geo-coded police stations, roads, socio-economic, aerial and satellite imagery for Nairobi County. A risk territory map showing the distribution of auto insurance risk and other related maps were generated. A prescriptive insurance rating model was then developed that uses generated risk territories to calculate varying rates for auto insurance premiums rates for the *respective regions. This research shows that GIS techniques can be used for better visualization of risk at a given location for accurate risk analysis and uptake.*

Keywords: Auto Insurance Risk Territory, Spatial Interpolation, Inverse Distance Weighting (IDW), Prescriptive Auto Insurance Rating Model

Introduction and Problem Definition

Insurance has a strong geographic component, from managing policy holder's addresses, risk location and handling of claims logistics. Location intelligence is vital in risk profiling and offers a more complete and accurate picture of risk (Esri, 2012). There are two primary benefits that can be realized as a result of understanding a risk based on its location (Esri, 2012): Insurers can make sure their clients are as safe as possible by offering risk mitigation advices, and insurers can accurately rate risks to ensure they collect appropriate premiums for the losses that are likely to be incurred.

Kenyan insurance industry stakeholders know that to be successful in this industry it is critical that a risk is accurately valued and rated in order to charge the right premium. Over or under insurance is critically lethal for business. According to AKI (2012), profit in the industry is an aspect of accurate pricing mechanism where both can be expressed by the following equations.

Profit = Earned Premium + Invested Income – Claims – Expenses

While Price mechanism can be expressed as follows:

Price Considers: Expected Claims + Admin Costs + Risk Premiums + Profit Margin

Some of these parameters like claims and risks clearly have geographical location based data. Determining of insurance premiums also referred to as rating is a comprehensive process that involves an actuarial function, insurance marketer and investments performance of the insurers (AKI, 2012). Guidelines on how to apply these variables to arrive at the optimum pricing of insurance products is provided for by IRA's guidelines of insurance risk of 2013 (IRA, 2013).

The Insurance Regulatory Authority (IRA), is a statutory government agency established under the Insurance Act (Amendment) 2006 Cap 487 of the Laws of Kenya, with the mandated to regulate, supervise and develop the insurance industry in Kenya. The Act was revised in 2013 to provide updated regulations for the Kenyan Insurance industry (IRA, 2013). Another insurance advisory agency that seeks to stabilize the insurance industry in Kenya is the Association of Kenyan Insurers (AKI) that was established in 1987 as an independent nonprofit making consultative and advisory body. The main objectives of AKI is to provide, promote and champion excellence in the Kenyan insurance industry, therefore AKI works hand in hand with IRA in developing and regulating the insurance industry in Kenya.

IRA and AKI conducted a joint research in 2009 to determine the minimum premium rates for all insurance classes. However, depending on an insurer's investment performance and market forces, an insurer can give various discounts and riders that make the products more profitable and competitive. These discounts are proportionate of the Profit Margin variable where in privately owned and comprehensively insured motor vehicles forms the No Claims Discount (NCD) (IRA, 2009). No Claims Discount (NCD) forms the basis of this research project. The NCD basically reduces the base premium rate of a certain period if the insured makes no insurance claims within a specific period that spans into years.

The price of premium payable per risk is subject to a price mechanism framework constituting expected claims, administration costs, risk premium and profit margin (AKI, 2012). A joint research conducted by IRA and AKI in 2009 arrived at various flat minimum premium rates for various classes of risks that apply uniformly within the country (IRA, 2009). It also gave a guidance on how insurers can offer No Claim Discounts (This reduces the base premium rate in case of no insurance claims launched annually) to their private motor insured clients. The motor underwriting guidelines indicates the maximum and minimum premium rates for this class of insurance at 7.5% of sum insured and the distribution of No Claim Discount (NCD) that reduces the base premium rate as follows: 2nd Year 10% NCD, 3rd Year 20% NCD, 4th Year 30% NCD, 5th Year 40% NCD, 6th Year 50% NCD and minimum rate of 3.8% of the sum insured or Kshs.15,000 whichever is higher for comprehensively insured private motor vehicles.

These rates apply regionally without considering the risks location aspect. This research shows that GIS spatial interpolation method can be used to show the spread of auto risk in a given region from which a prescriptive insurance rating model can be developed to calculate premium rates in respect to various territories.

Literature Review

The power of pooling high-resolution satellite imagery, geospatial information, and particularly census and statistical information for decision making and policy formulation has been recognized, while accessibility to such information is seen as an important development tool (Galigeo, 2012). Traditional maps have long been the primary tool for users to view and access geospatial information. In the past decades, rapid advances in geospatial information technology are enabling better access and integration of location-based information and transforming the traditional role of maps to include new tools for analysis and management with nations and leading organization conducting research in such areas as security, health, transport, agriculture, and insurance. Due to committed research in GIS, expert systems have been developed that enable mining of crucial intelligence from location data. In order to achieve a successful location analytics system there are two basic components that are required: data and technical components (Galigeo, 2012). Data models and technical components can be found in the Business Intelligence repository and the GIS Server. These data are geographic or spatial in nature.

Data Models

Every GIS data set has two components: the feature component - the spatial representation of real-world features (e.g. trees, roads, rivers, towns); the attribute component - the non-spatial data that describe the features (Goodchild et al., 2005). Complex shapes such as polygons are better represented in vector data. As the pixels in raster layer get smaller (i.e. finer resolution or fine scale), the better they would be representing complex features.

Location Analytics

Technical components necessary for implementing a Location Analytics include:

a) Business Intelligence (BI) platform - Contains modules that enable mining of various patterns or knowledge from the GIS database
b) GIS Platform and Geospatial Web Server - Contain algorithms for analysis of data to get various outputs which is usually in form of maps
c) Interactive Map Viewer - Connects to the existing BI and GIS systems respectively and assists in displaying the results of analysis in various formats.

The above components have been successfully implemented to analyze both data in a database and GIS system to mine locations intelligence. In order to carry out an insurance data GIS analysis the following data sources was need to be considered (Siebert, 2006).

A number of methods have been developed for representing geographic information for analysis. GIS for insurance has seen tremendous research being carried out by the giants of reinsurance, Munich Re and Swiss Re setting up full-fledged departments to research on insurance risk location intelligence. This has also been mainly featured in annual conferences on Insurance and Risk Management.

GIS Analysis Methods, Tools and Algorithms

There are usually three main stages involved in Geo-analytics (Guagliardo, 2004):

a) *Data Capture, Storage and Management* - of spatial or attribute data. This is usually done using digitization of analogue maps, geo-coding, satellite imagery or remote sensing, data from existing databases or warehouses, aerial photos or airborne laser scanners, existing reports. Databases can range from advanced relational database management systems like Oracle, MS SQL Server, and Sybase to simple ones like Access and Spreadsheets.

b) *Mapping and Visualization Tools* - to communicate the results of analysis. Some of the applications are AccessMod, SIGEPi, Arc-GIS, Maptitude-GIS, Galigeo, and Quantum-GIS among others.

c) *Algorithms* - for main Analysis of Vector and Raster data**.** A number of algorithms are applied for analysis such as buffering, classification, overlay, proximity analysis, shortest path, data smoothing, clustering and raster Cost-distance analysis.

GIS can be used to develop insurance rating territories by applying spatial smoothing and clustering techniques (Mishra, 2011). Spatial smoothing applies data from all areas within a set radius from a specific location, weighted by its distance to the center, this helps to achieve a desired level of credibility and reduce sampling error. To develop new rating territories clustering techniques can then be used. Apart from its use as an effective analytical tool, GIS is a great communication medium for visualizing analytical results on a map tocommunicate to a wider audience. Advanced GIS and spatial econometric techniques are useful in analyzing spatial interactions and developing a predictive model that accounts for local factors to determine annual premium rating estimates.

GIS Spatial Interpolation Methods

Spatial interpolation is a GIS mathematical method of estimating values at geographic locations where no values have been provided by using values known at a certain location (Azpurua & Ramos, 2010). Spatial interpolation assumes the attribute data are continuous over space allowing estimation of the attributes at any location within the data boundary. It also assumes that attributes are spatially dependent, indicating that values closer together are more likely to be similar than the values farther apart. The main objective of spatial interpolation is to create a surface that is intended to best represent empirical reality. The most common interpolation methods available in GIS software are Spline, Inverse Distance Weighting (IDW) and Kriging.

a) Inverse Distance Weighting (IDW)

The IDW method gives a surface that does not exceed the highest or the lowest values of the z-value (George & Penka, 2010). IDW calculates grid cell values by averaging the data of the points from each neighbouring cell. Points closest to the central one are estimated and their weights will have greater influence in the interpolation process. IDW refers to definite interpolation method since it is directly based on the surrounding measured values. IDW is the most applied GIS analysis tool (Eldrandaly & Abu-

Zaid, 2011). It uses the Tobler's First Law of Geography by estimating unknown measurements as weighted averages over the known measurements at nearby points, giving the greatest weight to the nearest points (Longley et al., 2011). The general equation for IDW method is shown below in equation (1)

$$z_0 = \frac{\sum_{i=1}^{n} z_i \frac{1}{d_i^k}}{\sum_{i=1}^{n} \frac{1}{d_i^k}} \tag{1}$$

z_0= estimated value at point 0, z_i= z value at a known point i, $d_i$is the distance between point i and point 0,n is the number of known points used in estimation, and k isthe specified power which controls the degree of local influence (Chang, 2010).

b) Splines

Splines interpolation method estimates values by using a mathematical function that reduces the total surface curvature creating a smooth surface that precisely passes through the sampled points (Azpurua & Ramos, 2010). Merits of splining methods are that they can generate accurate surfaces from only a few sampled points and they retain small features. A demerit is that they may have different minimum and maximum values than the data set provided. Splines are also sensitive to outliers. This is mainly due to the inclusion of the original data values at the sample points.

c) Kriging

Kriging works same as IDW that apply a weighting which assigns more influence to the nearest points when interpolating values for locations that values are not known. Kriging is non-deterministic; it extends the proximity weighting approach of IDW to include random components. The main disadvantage of Kriging is that it requires more computing and modelling time and also requires more data from the user (Azpurua & Ramos, 2010).

Kriging estimates the value of unknown real-valued function, *E* at a point (*x, y*), given values at some other points $\{(x_1, y_1), (x_2, y_2)\ldots(x_3, y_3)\}$. The predicted value E(x, y) is a linear combination that makes Kriging estimator which is represented by the equation below:

$$E(x,y) = \sum_{j=0}^{n} \lambda_j E(x_j, y_j)$$

Developing an Insurance Rating Model

In insurance rating, premium rates should be designed in a way that after subtracting commissions, expenses, risk and profit loads it should be sufficient enough to cover claims costs(Ana & ACAS., 2010). A consistent benchmarking Premium rating model should consist of the following parameters:

P	=	Pure Premium or loss cost rate
V	=	Variable expenses such as commission
F	=	Fixed expenses such as overhead costs
Q	=	Risk or contingency load
U	=	Underwriting profit load

$$\boxed{R=P + V * R + F*R + Q*R +}$$

Note that V, F, Q and U are expressed as percentages of the final premium rate (R). Alternatively to cover claims cost, Premium rate should be sufficient enough after subtracting commissions, expenses, risk and profit loads.

$$\boxed{P=R - V * R - F*R - Q*R -}$$

The above equation equates to

$$P = R^* (1 - V - F - Q - U)$$

Therefore the premium rate is the pure premium or loss cost rate divided by 1 minus all costs included risks and profit loads.

$$R = \frac{P}{(1 - V - F - Q - U)}$$

Prescriptive GIS Premium Discounting Rating Model

A prescriptive model usually consists of an objective function, decision variables and constraints (Winston, 2003). The objective function is the value that one seeks to minimize or minimize, decision variables are the variables that influence the value of the objective function and they are also the variables that can be controlled. Finally constraints define the conditions under which the model operates. The model currently being used by IRA was as a result of a five year analysis of claims incurred by the insurance companies and the insured who never made claims for the same period. As a result IRA, set the rate limits or constraint at 7.5% for new clients and 3.8% for who do not make any claim for 6 years. The main problem with this model is that it applies a flat rate on the premium and does not consider the geographic location aspect of risk. Due to inability to view risks at different locations a vehicle operator have to wait for 6 years to enjoy the lowest discounted rates.

Case Studies of GIS in Auto Insurance Risk Segmentation and Rating

The current trend in the market is the creation of powerful GIS application servers consisting of a range of analysis tools that can be applied over a wide range of applications (Caliper Corporation, 2013). Geo-analytics has been successfully applied in many areas of interest to mine location based intelligence in various fields of interest. Some of the most recent developments in GIS have been done by Caliper Corporation developed an intelligent mapping solution for Business, Government and Education, called Maptitude-GIS. This package provides the necessary spatial and attribute data analysis tool, visualization and presentation tools in one package that are required to start analysis on the go. This is the future of GIS analysis. The challenge at the moment is that a number of countries' spatial and attribute data has not been represented in the package. The kind of data collected through telematics can be used by insurers to develop accurate premium pricing that mitigates losses from claims. Actuaries can use this data for developing new pricing variables and also develop existing one.

Auto Location Insight

Morgando (2011) conducted a research that resulted in the development of an Auto Location Insight that can be used to rate auto insurance risks within the region. The author is a senior statistician at Explore Information Services in Eagan, Minnesota and organization whose core business is to aggregate, analyze, and deliver location-based risk intelligence to insurance carriers in the United States. Auto Location Insight, that uses ArcGIS Server and Microsoft Server, assists insurance carriers assess location-based risk for automobile policyholders. Data was collected from their garaging address and the likely commute routes. Analysis was done using ArcGIS Server and Esri Street-Map Premium dataset from providers and Esri Partners Navteq and Tele Atlas. Street addresses were geocoded and displayed on a map, analysis of the data to find the shortest or fastest distance between locations was done. Commute time is highly accurate due to the information provided with the street networks. Historical traffic data, such as the average travel speed for roadways to create more accurate arrival time projections avoiding congestion based on day and time, can be applied to the modeling. More geo-referenced data including traffic, weather, and crime can be analyzed.Depending on the risk attributes, certain geographic areas can be classified from low to high risk.

Geo-analytics for Improved Risk Segmentation in Auto Insurance

A research on application of Geo-analytics for improving auto insurance risk segmentation was carried out by Mishra (2011). The author realized that garaging location was a critical rating variable when establishing premiums for private owned automobile insurance. The region where a client or insured lives affects not just claims for theft, vandalism, and damage from weather (hail, flood, etc.) but it also influences and shapes the overall client driving behavior. General road conditions, topography, signage, and other distractions can have a profound effect on occurrence of accidents hence the resulting claims. Commute time, distance to shopping centers, availability of public transit, and housing affordability are some of the factors that determine the choice of where to live. Some of those same factors also have a big impact on where, how often, and how far an insured drives. Annual mileage is a strong predictor of risk for auto insurance and is animportant rating variable in many rating plans. However, many insurers hitherto rely on self-reported annual mileage estimates provided by their policyholders or agents, with little or no validation. Mishra (2011) applied GIS smoothing and clustering techniques to generate a map for spatially smoothed average commute time in California.

Telematics Devices for 'Pay as you Drive' for Insurance Premiums

Currently motor insurance is a mature and largest class of general property and casualty insurance, however realizing underwriting profits for this class of risk has been elusive for some time now (Anusuya, 2013). Telematics devices monitor wirelessly a vehicles geographic coordinates and driver's performance in real time. Knowing the 'where' of the risk usually doesn't tell us of all the risk, this needs to be coupled with other factors like the 'how', behavior aspects that drive the risk. Careless driving usually exhibited by drivers has mostly contributed to high cost of claims payments resulting from employee injuries, lost productivity, fines, asset damage, litigation, poor fuel economy and excess CO_2 emissions.

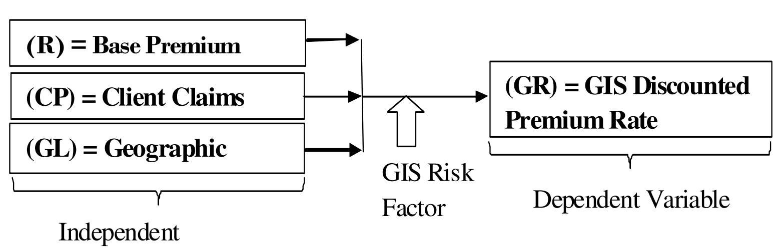
A good example is that of General Motors who have come up with an offering like 'OnStar' a subscription based on in-vehicle security, hands-free calling, turn-by-turn navigation and remote diagnostics. Another product called 'Sync' has been installed in all Ford, Lincoln and Mercury products and within Asia Pacific. 'GBook Alpha' is a standard product in most Toyota vehicles. The vehicle and driver data collected by telematics GPS devices is used by insurers to develop accurate premium pricing and reduce claims losses (Anusuya, 2013). Actuaries will use data collected by these devices to determine new pricing variables and improve the quality of existing pricing variables. For underwriters this data can be used to validate annual mileages, commute distances, garaging location and other variables linked to leakages. Marketers can use this data to develop innovative products and additional services to reinforce a company's brand and market positioning. This data can also be used in assessment of claims as support information.

Telematics have resulted in the emergence of highly innovative schemes in motor insurance. 'Pay as you drive' is a product that allows insurance policyholders to pay premiums based on the car's usage. These GPS devices used by this product tracks the vehicle's position, speed and other information via GPS and transmits the data to a center of interest on real time basis.

Conceptual Framework

From the above cases it was noted that garaging addresses, likely commute routes, data about location or region where one lives, and crime data, can be considered as the main data that affect the auto insurance risk in relation to a geographic location. It is evident that GIS tools have a wide range of applications in insurance. They are extremely useful for risk segmentation. Application of these GIS tools makes it easy for one to move away from artificial constructs such as Zip codes and redefine rating territories that are truly representative of the unique risks and exposures within each region or territory (Mishra, 2011). Like most of the segmentation based systems, it is evident that we will have to apply market based client data to come up with the region based rating territories and rating model. This will be due to the main factors that determine the off the shelf based system such as, objectives of research, availability of data and budget available for research and development (Futterman, 2009). From the research on the available or commonly used surface interpolation techniques it was determined that Inverse Distance Weighting (IDW) technique is the most appropriate model in generating of varying geographic territories in respect to varying attribute data related to a spatial location. IDW generates segments that precisely match the dataset provided. From the territories or segments created a prescriptive or optimization model was developed to help in determining premium rates in respect to a territory. The model will appear as follows.

Fig.1: *Proposed Conceptual Framework for the GIS Discounted Premium Rating Model*



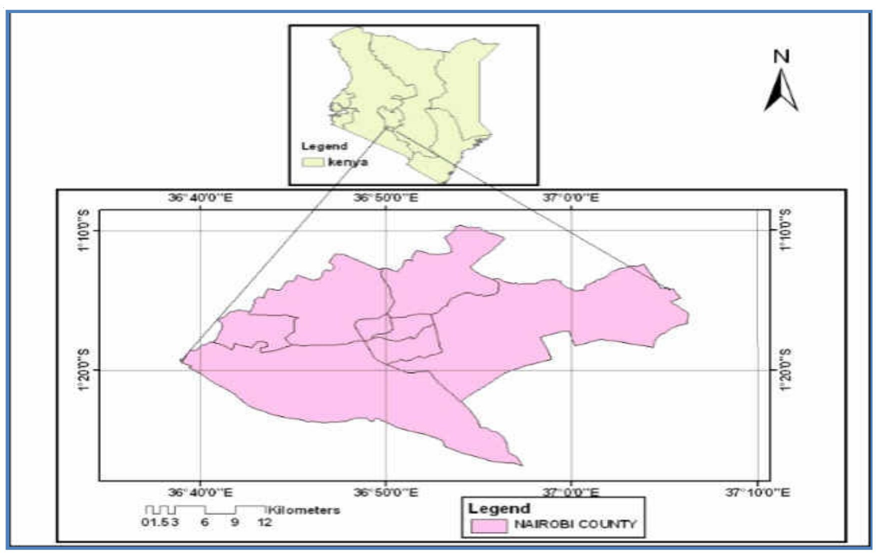
Source: *(Authors)*

Research Methodology

Research Design

The area of study was Nairobi County which also has Nairobi City the capital city of Kenya, which was founded in 1899 as a supply depot for Kenya Uganda Railway. The county occupies an area of 696 square kilometers (270 square miles). Nairobi County has the following main administrative districts: Nairobi West, Nairobi South, Nairobi North and Westlands. Nairobi is the most populous city in East Africa, with a current estimated population of 3,138,369 according to the 2009 Census (KNBS, 2013). Nairobi City is at an altitude 1,795 meters (5,889 ft.) above sea level, it enjoys a moderate subtropical highland climate. Below is a map of Nairobi County and its location in the Kenyan Market.

Fig. 2: *Map Showing Nairobi County the Area of Study*



Source: *(Authors)*

This study was established to help Nairobi County motor insurance industry improve the model of rating motor insurance risk premiums by generating a GIS risk factor that can be loaded on the existing model. Both Spatial and attribute data mostly secondary data was collected and used for conducting this research. A GIS spatial interpolation method was used to create rating territories using police stations with respective insurance related auto crimes and accidents data as control points. The reason for choosing police stations as the analysis control points was because the traffic rules in Kenya require that an when an accident occurs, it must be reported to the nearest police station. Insurance companies also pay claims based on reports from the police station where the accident was reported. The reports are known as police abstracts. This simply means a police station with the highest number of reported incidences is located in a region with the highest auto risks. A GIS spatial interpolation method was used to generate risk territories in a risk distribution map. Smoothing and clustering algorithms were applied to achieve the required level of accuracy. The resulting risk map gave the distribution of auto risks within the region. Probability ratios were developed from the risk map territory limits to form the basis for development of a prescriptive Auto insurance rating model.

Data Sources and Tools

The datasets used in this research was obtained from different sources and have various characteristics as summarized in the table 1 below:

Table 1: *Datasets and Data Sources*

Dataset	*Characteristics*	*Data Source*
Current Max and Min Auto Insurance Premium rates	Percentage of Sum Insured	IRA(Insurance Regulatory Authority
Nairobi County Demographic Data	Data from 2009 National population census	Ministry Of Planning
Road network (Classes A, C &D)	Shape files	Survey of Kenya
Nairobi County Administrative boundaries	Shape files	Survey Of Kenya
Police Divisions HQ Stations	Discrete Coordinates	Nairobi City Council
Insurance related motor Crime and Accident Data	Crime and Accident incidences reported that required claims settlement	Kenya National Police, IRA and Kenya Reinsurance Corporation

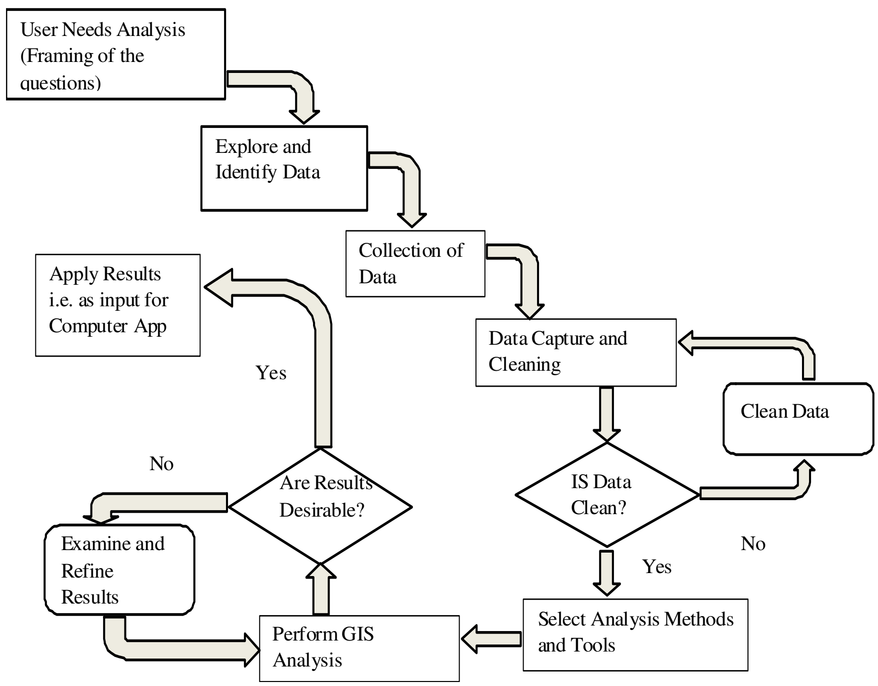
Source: *(Authors)*

Model Verification and Validation

For model verification and validation a set of collected data was used to generate the required data from the GIS tool and this data was used on the developed model which gave desirable output. Data for motor vehicle crime and accidents was collected for six years covering the period 2008-2013. Data for five years that is 2009-2013 was used for analysis and model development while data for 2008 was used for model verification and validation.

Methodology Overview

Fig. 3: *GIS Analysis Methodology Overview*



Source: *(Authors)*

During user assessment study user information needs concerning the application of GIS in auto insurance risk segmentation and rating were identified. In data identification various datasets necessary for carrying out the analysis were identified such as; police motor crime and accidents data, demographic data, road network, insurance liability, land cover and built up areas. Data collection was from sources such as Nairobi City County, Kenya National Police Service, Survey of Kenya, Kenya

Reinsurance Corporation and Kenya Open Data Initiative. This involved spatial data and attribute data. Data cleansing was done by editing unnecessary data and calculating the necessary averages required for the GIS analysis. ArcMap and Inverse Distance Weighting spatial interpolation tool were identified for the GIS analysis. Various input parameters like input point features, interpolation power factor, search radius; number of analysis points, classification factor and Z-Values were varied to calibrate the GIS model to give the desired output.

A set of separate observed data was used to test the used GIS model and the developed insurance rating model in order to validate the models.

Detailed Procedure for the Analysis

ArcGIS was used for the overall analysis and presentation of the research findings. The main processing and analysis steps followed for the application of the method implemented in preparation for the ArcGIS analysis were:

a) Standardization of databases to UTM 37 S projection and datum WGS-1984.

b) Thiessen polygon analysis to identify risk areas based on the nearest police station.

c) Assignation to trajectories and roads to be done using 'surface tools for points, lines and polygons'.

d) Application of spatial smoothing from all areas within a variableradius (dictated by number of incidences reported by each police station) from a police station, weighted by its distance to the center.

e) Application of interpolation techniques to develop risk territories, interpolates a surface from points using an Inverse Distance Weighted Technique (IDW).

f) Application of IDW algorithms to discern risk scores, this applies Thiessen polygon analysis and Nearest Neighbor analysis, smoothening algorithms.

Creation of a GIS Database

In preparation for visualization of the auto risk and maps showing related data, a geo-database was created using ArcCatalog SQL. ArcMap attribute edit option was used to capture attribute data into the database for the respective shape files:

i. Nairobi County administrative boundary
ii. Nairobi County divisions administrative boundaries
iii. Nairobi County major roads network
iv. Police stations point co-ordinates
v. Nairobi County divisions demographic data
vi. Nairobi County auto crime and accident data linked to police stations

Generation of Maps for Datasets

All spatial data were collected as shape files which are in the format identified by the ArcMap software. The following maps were created: Nairobi County administrative boundaries and divisions, Nairobi County police division headquarters, Nairobi County road network, Nairobi County population distribution, and Nairobi County auto insurance risk map.

Analysis and Generation of a Nairobi County Auto Insurance Risk Territory Maps

The main objective of this study was to generate an auto insurance risk map for Nairobi County to enable proposing a rating model improvement on the model currently used by the insurance companies. It was also part of the research questions to find out what patterns can be learnt from the relationship of auto risk and associated data sets. The questions asked were: what would be the relationship between roads network and auto insurance risk? What is the relationship between demographic distribution and auto insurance risk? ArcTool module of the ArcGIS was used for analysis.

a) Buffering

Variable buffers were created using police division headquarter stations against the auto crime and accident data reported per police headquarter station. The extent of the buffers was determined by the total number of auto risk insurance reported per station. ArchMap presents a special algorithm for enabling creation of fixed and variable buffers. Due to the varying nature of risk, variable buffers were created using the police headquarters as epi centers.

b) Use of Inverse Distance Weighting (IDW) to Generate Risk Territories

IDW was used to develop the final Nairobi County risk territory map. IDW is a local exact and deterministic technique (Burrough & McDonnell, 2009). IDW estimates data values for each point by calculating a distance weighted average of all the points within the search radius. IDW was chosen in regard to Splines and Kriging for its ease of understanding, good execution time and its ability to honor control point data. IDW precisely produces segments within the data given at the control points. The generated map was analyzed for its risk distribution within Nairobi County.

Use of GIS Model Output to Develop a GIS Insurance Rating Model

The GIS model building framework was used to develop a GIS model that gave output values that were used to develop a rating model. The four steps of GIS model development were followed: defining goals of the model, identifying the model constituent elements and defining them; implementation and calibration of the model by varying the input numeric variables; model validation and verification by using a different set of observed data apart from the set used to develop it; and finally using the model output as input for the computer program prototype.

Development of a GIS Insurance Risk Rating Model

A prescriptive or optimization model that uses the output of the GIS model as its input to generate an auto insurance premium rate was developed. The model comprised of an objective function, decision variables and constraints. Objective function is the output of the model or the function that is either maximized or minimized. Decision variables are those variables that act as input and can be varied to alter the value of the objective. Constraints define the conditions under which the model operates.

Results and Discussions

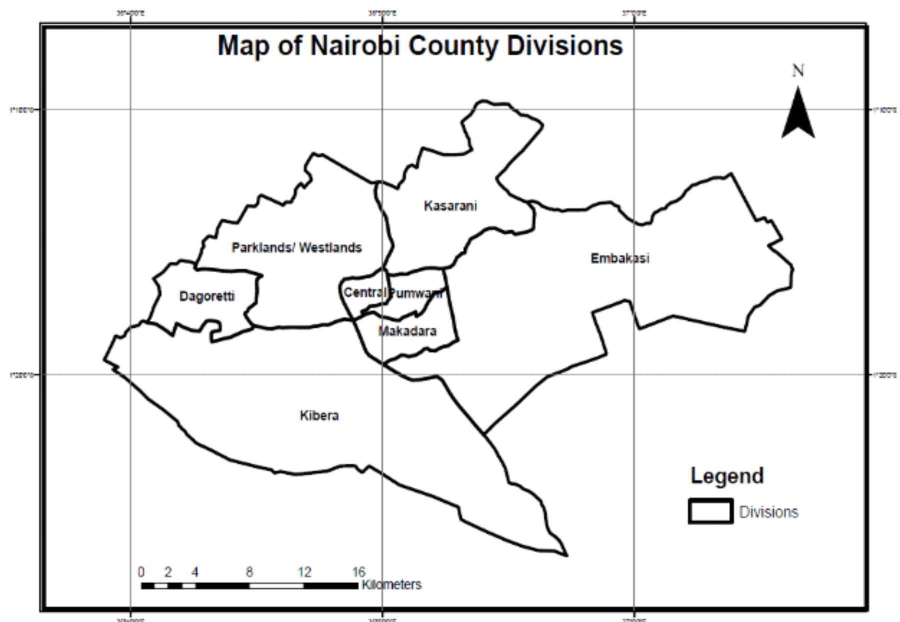
Results

From the data collected various maps were generated and respective observations made.

Administrative Boundaries and Divisions for Nairobi County

Nairobi County consists of the following divisions: Parklands/Westlands, Kasarani, Embakasi, Kibera, Makadara, Pumwani, Kibera and Dagoretti. This study noted that there is some slight change in matching of the Nairobi County administrative divisions and police divisions. For the purpose of this study police division headquarters were used as the control points for risk analysis.

Fig. 4: *Nairobi County Administrative Division Boundaries*



Source: *(Authors)*

Demographic Data for Nairobi County

According to the government of Kenya 2009 census Nairobi County has an estimated population of 3,138,369 people (KNBS, 2013). This constitutes about 10% of the entire population of Kenya. Data from 2009 National Census was used to generate a population distribution map for Nairobi County. The data is as follows:

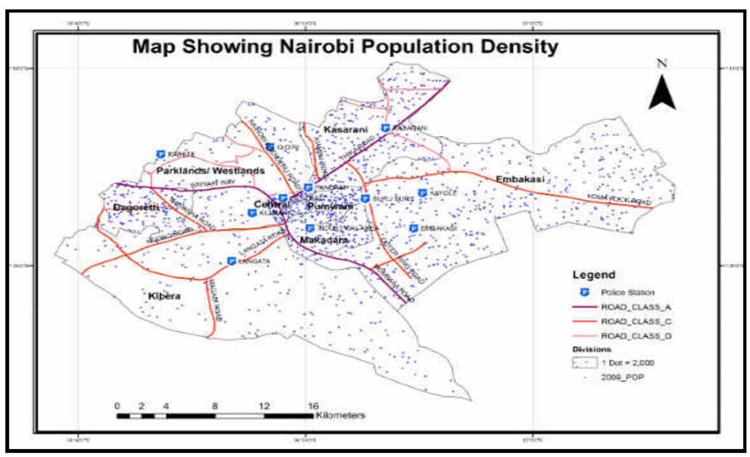
Table 2: *Nairobi County Divisions Population and Risk Distribution*

Division	*Population*	*5 Year Average Risk*
Central Nairobi	274,607	229
Makadara	218,641	210
Kasarani	525,624	354
Embakasi	925,775	659
Pumwani	261,855	276
Westlands	247,102	484
Dagoretti	329,577	187
Kibera	355,188	147

Source: *(Authors)*

From the above data a population distribution map was generated as follows. This map also comprises of the Nairobi county road network that was collected as shape files.

Fig. 5: *Population Distribution for Nairobi County*



Source: *(Authors)*

The map in Fig 5 shows that Kasarani and Embakasi have the highest population per square kilometer. One of the patterns learnt in this analysis is that the number of motor vehicle crime and accidents data from the police data increased with the number of population. Emphasis was laid on the division with the highest number of risk incidences reported i.e. Kasarani. A sudden increment of incidences by over 44% from 2012 to 2013 was observed (see table 3 below). Information from Kenya roads network indicated that a super highway was just completed spanning through the division. Good roads would have translated to less incidences but something else may be causing the surge in incidences i.e. luck of education to the road users on how to safely use a super highway. Further research can be done on this area.

Table 3: *Nairobi County 5 Year Average Police Motor Accident and Crime Data*

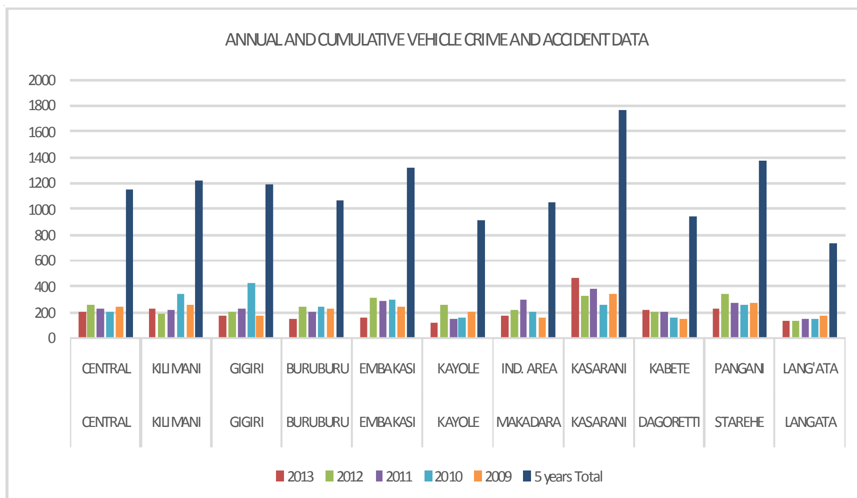
Head Quarter (Police Divisions)	*2013*	*2012*	*2011*	*2010*	*2009*	*5 Year Total*	*5 Year Average*
Central	207	257	236	198	246	1144	228.8
Kilimani	231	189	210	336	255	1221	244.2
Gigiri	173	206	223	420	176	1198	239.6
Buruburu	145	242	208	241	228	1064	212.8
Embakasi	163	316	289	304	248	1320	264
Kayole	124	264	146	166	209	909	181.8
Industrial Area	178	211	305	203	157	1054	210.8
Kasarani	467	324	381	261	337	1770	354
Kabete	219	209	195	166	146	935	187
Pangani	235	342	269	261	274	1381	276.2
Lang'ata	129	136	148	146	175	734	146.8

Source: *(Authors)*

Motor Vehicle Crime and Accident Data for Nairobi County Police Divisions

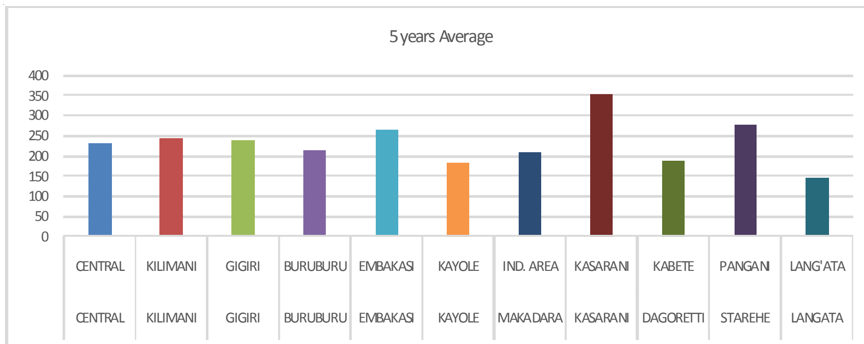
The incidents reported here are those that required police abstract for insurance companies to settle the claims. They comprise of motor accidents to other vehicles, pedestrians and other road users, vehicle theft and vandalism. They are assumed to approximate the magnitude of risk in a specific location in relation to the risk incidences reported in that area. Fig 6 contains the 5 year average auto insurance crime and accident data per police stations being covered.

Fig. 6: *Annual and Cumulative Motor Vehicle Crime and Accident Data for Nairobi County*



Source: *(Authors)*

Fig. 7: *5 Year Average Motor Vehicle Crime and Accident Data for Nairobi County*



Source: *(Authors)*

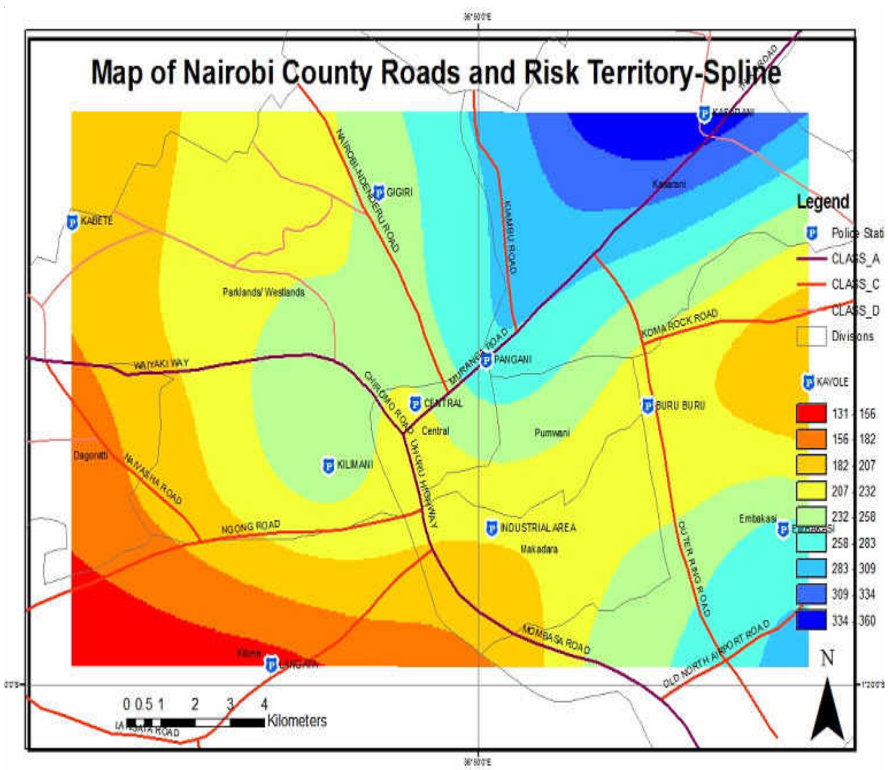
From Fig 6 and Fig 7 it can be observed that Kasarani and Embakasi Police Divisions tops in the total vehicle crimes and accidents data report and it can also be seen that on the demographic map both divisions have the highest number of population respectively. We can therefore argue that increased population may increase the number of auto vehicle crime and accident data.

Nairobi County Auto Insurance Risk Segmentation Maps

Using the data collected Nairobi County roads network, police divisional headquarters, auto crime and accident data, ArcCatalog was used to create the database, ArcTool was used to generate maps showing segmentation and distribution of this risk in relation to police stations spatial locations. Three spatial interpolation methods were considered; Splines, Kriging and Inverse Distance Weighting (IDW). 5 years police Auto accident and crime data (Z-Values), Geo-coded police stations, roads and division boundaries were used to generate the maps.

Splines

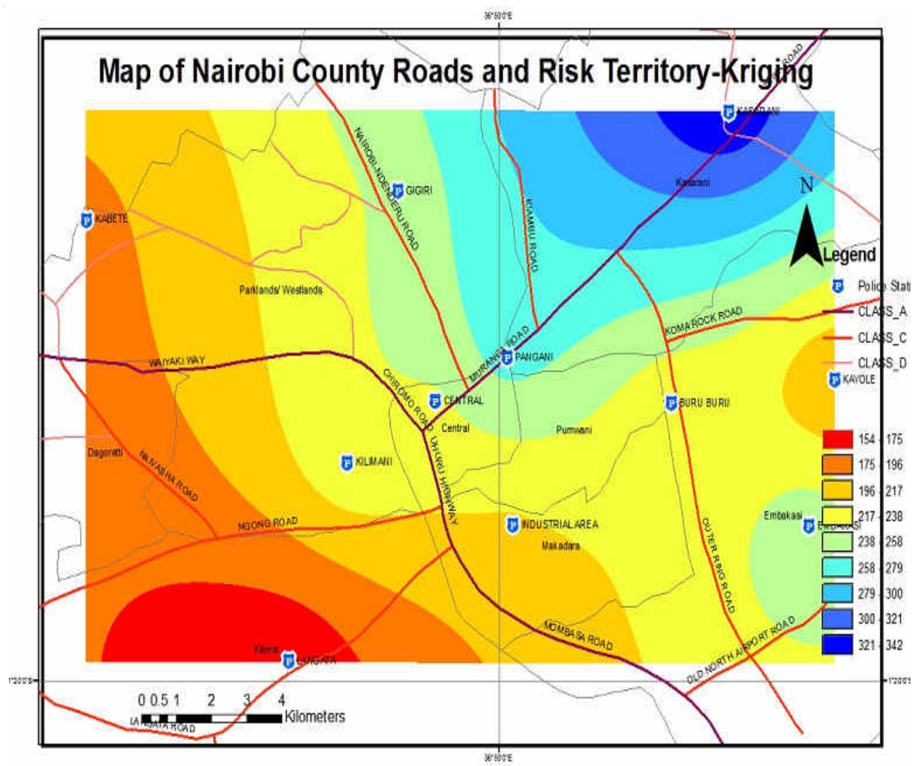
Fig. 8: *Nairobi County Auto Risk Distribution Map Generated Using Spline Method*



Source: *(Authors)*

Kriging

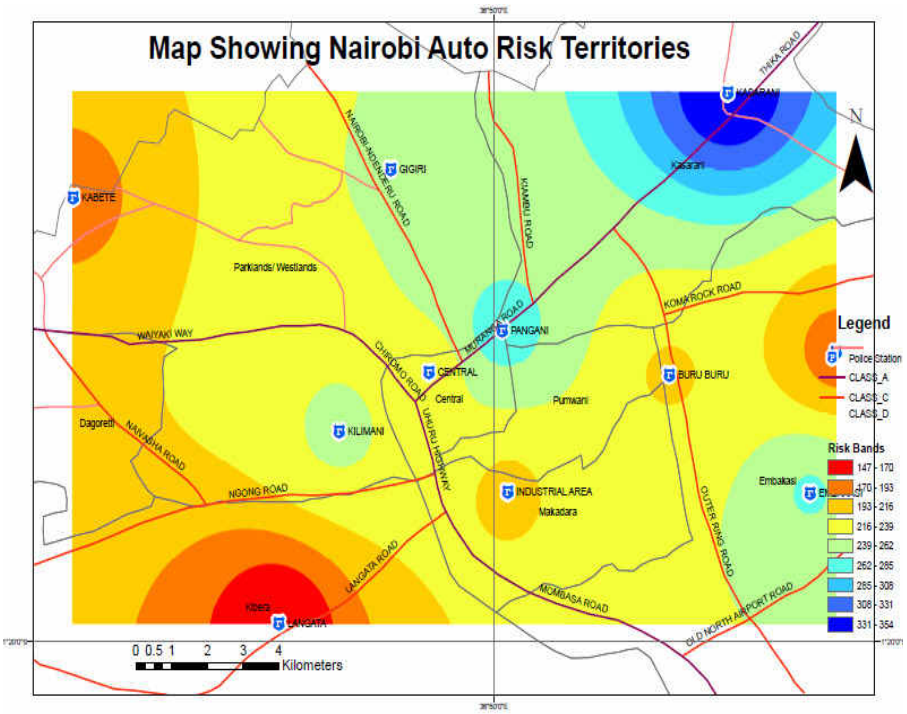
Fig. 9: *Nairobi County Auto Risk Distribution Map Generated Using Kriging Method*



Source: *(Authors)*

IDW

Fig.10: *Nairobi County Auto Risk Distribution Map Generated Using IDW Method*



Source: *(Authors)*

Table 4 below shows the comparison between Splines, Kriging and IDW risk territory limits generated by the three spatial interpolation methods shown in the maps above. These limits were compared with the actual incidences reported where the lowest was 147 reported at Langata and 354 reported at Kasarani (refer to Table 3 above). The limits are created by using average incidences also referred to as the Z-Values reported showing the probability of an incidence occurring in a certain region.

Table 4: *Comparison between Splines, Kriging and IDW Auto Risk Territory Limits*

Splines	*Kriging*	*IDW*
131-156	154-175	147-170
156-182	175-196	170-193
182-207	196-217	193-216
207-232	217-238	216-239
232-258	238-258	239-262
258-283	258-279	262-285
283-309	279-300	285-308
309-334	300-321	308-331
334-360	321-342	331-354

Source: *(Authors)*

From Table 4 above it can be observed that IDW has created risk territories that most appropriately match the 5 year average incidences reported at the control points (Police Stations HQs). In this regard the map generated by IDW was taken to best represent the distribution of auto risk within the county. IDW also creates bull's eyes around the control points conforming to the main assumption that the area around the control points or police station usually reflect the magnitude of the risk since all incidences are first reported at the nearest police stations (refer to figure 10 above). From the map it can be observed that Kasarani division topped all other divisions in form of motor vehicle accidents reported. Roads are the auto risk trajectories as one moves from one region of risk to another. Taking for instance drivers who live in Kasarani and works in the Central district area their probability of getting involved in an accident reduces as they leave Kasarani for CBD and increases as they go back to Kasarani. A driver living in Lang'ata has a lower probability of getting involved in an accident as compared to the one who lives in Gigiri. From this understanding the limits of IDW are used to create probability ratios can be used in an insurance prescriptive premium rating model.

Development of a GIS Risk Rating Factor

From the GIS risk segments generated by the IDW method, unique risk probability ratios were generated by dividing each segment upper limit with the highest number of risk incidences reported. Then this fraction is multiplied by the difference between the IRA (Insurance Regulatory Authority) regulated maximum and minimum premiums rates to get the GIS discounted premium rate to be loaded on the minimum regulated premium rate. This ensures that the final premium never operates at a loss since it is within the regulated limits. The reason for using upper limits is to ensure that the values adds to 1 at the highest risk point or region hence maximizing on the risk load and minimizing on the risk discount. The IDW segments represent risk buffers or territories. Using the map one can approximate the probability of a risk occurring depending on which IDW risk buffer it falls under. Table 5 below shows how GIS generated risk probability ratios can be used to calculate the final GIS discounted premium rates. This is further compared by the rates generated by current IRA rating model.

Table 5: *GIS Motor Insurance Risk Rating Table*

Head Quarter (Police Stations)	*5 Year Average*	*GIS Risk Band Or Territory*	*GIS Risk Factor (%)*	*% of discountable rate (7.5-3.8=3.7)%*	*GIS loaded discounted rating (%)*	*Rating by Current IRA Model (%) - Comparison*
Lang'ata	147	147-170	48.02	1.78	5.58	7.5
Kayole	182	170-193	54.52	2.02	5.82	7.5
Kabete	187	170-193	54.52	2.02	5.82	7.5
Ind. Area	211	193-216	61.02	2.26	6.06	7.5
Buruburu	213	193-216	61.02	2.26	6.06	7.5
Central	229	216-239	67.51	2.5	6.3	7.5
Gigiri	240	239-262	73.94	2.74	6.54	7.5
Kilimani	244	239-262	74.01	2.74	6.54	7.5
Embakasi	264	262-285	80.51	2.98	6.78	7.5
Pangani	276	262-285	80.51	2.98	6.78	7.5
Kasarani	353	331-354	100	3.7	7.5	7.5

Source: *(Authors)*

Development of a GIS Auto Risk Rating Model

Using the data generated in the Table 5 above a prescriptive or optimization auto insurance rating model was developed for rating of the auto insurance risk in Nairobi County. An optimization model usually gives the best behaviour or value of operation that an organization derives most benefit at each point. The components of an optimization model comprise of the following: Objective Function(s), Decision Variables and Constraints (Winston, 2003).

a) Model Objective Function

This is the value that we seek to minimize or maximize. For this study the value that we seek to vary or optimize is the **GIS Discounted Rate = GR**

b) Decision Variables

These are the variables that influence the value of the objective function. These are also variables that we can control. For this study these variables are identified as follows:

GIS Highest risk Territory Upper Limit = $\mathbf{I_{max}}$
Maximum Regulated Rate = $\mathbf{R_{max}}$
Minimum Regulated Rate = $\mathbf{R_{min}}$
GIS Risk Territory Upper Limit Residential= $\mathbf{GR_{max}}$
GIS Risk Territory Upper Limit Work= $\mathbf{GW_{max}}$
GIS Risk Territory Lower Limit Residential = $\mathbf{GR_{min}}$
GIS Risk Territory Lower Limit Work = $\mathbf{GW_{min}}$

c) Model Constraints

Constraints define the conditions under which the model must operate**.**For this study a certain conditions must be fulfilled by the model. The GIS Discounted Rate (GR) cannot be more than Maximum Regulated Rate (R_{max}) and cannot be less than the Minimum Regulated Rate (R_{min}). All Variables must be greater than Zero. From the above conditions we can derive them mathematically as follows:

$GR \geq R_{min}$, $GR \leq R_{max}$

$X_i \geq 0$ for $i = GR, GR_{max}, GW_{max}$,

$GR_{min}, R_{max}, R_{min}$ and I_{max}

GIS Auto Insurance Rating Model

After letting GR be our objective function our complete optimization model that can be used to provide various GIS auto insurance risk premium rates can be expressed as follows for a given risk territory i.

The model above takes into consideration the fact that a resident of Nairobi County may be working in one division and living in another. The model takes into consideration both the division of residence and division of work.

$$\boxed{GR_i = (((GR_{max} / I_{max}) + (GW_{max}/I_{max}))/2)\,(R_{max} - R_{min})) + R_{min}}$$

Computer System Prototype to Demonstrate usage of the GIS Auto Rating Model

In order to demonstrate how the above model can be used in a business setup a simple model application prototype was developed using Oracle Developer and Oracle Database 11g server (See Appendix for sample code).

Overview of Results and Analysis

This research has successfully shown that given a certain geographic region, one is able to segment the geographic region into territories of varying risk magnitude. Unique GIS risk probability ratios can be generated and used in calculating of GIS regulated premium rates. This research has also shown that the magnitude of risk increases with increase in population. Good roads also do not translate to low risk on motor accidents and crime incidences. This has been shown by the Kasarani Division which is connected by recently completed Thika Super Highway registering one of the highest risk incidences. GIS can also be used to analyse the spread of motor risk not only based on the police stations as control points but also in relation to the roads which in this case are the risk trajectories.GIS Mapping is a critical tool in visualizing location intelligence. It provides a number of advantages in a graphical view that may not have been possible with other tools in an easily translatable and appealing way. There are a number of advantages that GIS mapping of insurance risk provides which have been experienced in this research.

a) Aggregating of various data into a single view
b) Graphical mapping of insurance risk
c) Identifying territories with varying insurance risk
d) Facilitating ways of mitigating motor risk
e) Assisting insurance companies vary uptake of risk
f) Recommendation of proper recording of data
g) Can be used for analysis of risk related resources
h) Can assist in distribution and deployment of security resources to prevent crime
i) Can be used to develop a curriculum for road users
j) Can be used as a basis for other important research

Conclusions and Recommendations

Conclusions

This study sought to demonstrate the use of GIS in understanding and rating of auto insurance risk within Nairobi County. The research successfully segmented Nairobi County into various regions of varying auto risk magnitude. It was shown that auto insurance risk increased as one moved from Southern part to Northern part of the County that is from Lang'ata to Kasarani. It was also noted that Kasarani, apart from having the highest insurance auto risk, also had the highest human population distribution and the best road connection to the Central Business District that is the Thika Road Super Highway. Various regions of the County can be classified into varying risk magnitude falling under varying risk territory bands. A GIS auto rating model was developed from the GIS risk territory probability ratiosthat can be used to calculate discounted auto insurance rates.

Recommendations

Having demonstrated that GIS can be used to visualize varying geographical nature of risk, the following recommendations are made.

a) Adoption of the proposed auto rating model as the initial step to improve rating of auto insurance risk in Nairobi County.

b) IRA and AKI to incorporate use of GIS in carrying out insurance risk research and eventual development of risk rating models. A policy on capturing of spatial risk data will require to be developed.

c) Insurance and re-insurance companies to employ use of GIS risk analysis and visualization to help in determining their uptake and distribution of risk.

d) Security agencies to use GIS in auto crime and accident risk analysis to assist in security resources planning and deployment for purposes curbing motor crime and traffic offences.

Considering that this study applied GIS in identifying the magnitude and distribution of risk on one class of insurance and covering a limited geographical area, there is need to extend the work to other geographical regions and all classes of insurance products within the country. GIS can greatly help insurers understand various risks hence enabling them strategize on risk take up, insurance product developments, and accurate rating that increases insurance penetration, take-up and business profitability.

References

IRA. (2009), *Motor underwriting guidelines Private car(individual).* Nairobi: Insurance Regulatory Agency.

AKI. (2012), AKI Journal. *Association of Kenyan Insurers Journal* (June).

Ana, J.M., and ACAS. (2010), A step by step guide to designing insurance rating models. London: Matblas.

Anusuya, D. (2013, July). Be insured with risk mapping. *Geospatial World*.

Azpurua, M., and Ramos, K.D. (2010), A Comparison of Spatial Interpolation Methods. *Progress In Electromagnetics Research, 14*.

Burrough, and McDonnell. (2009), *Spatial Interpolation Techniques.*

Caliper Corporation. (2013), The Intelligent Mapping Solution for Business, Government, and Education. *Maptitude Geographic Information System*.

Chang, K. (2010), *Introduction to Geographic Information Systems.* McGraw-Hill, New York.

Eldrandaly, K.A., and Abu-Zaid, M.S. (2011), Comparison of Six GIS-Based Spatial Interpolation Methods for Estimating Air. *Journal of Environmental informatics, 18*(1).

Esri. (2012, March), GIS for Insurance. *GIS Best Practices Series*.

Futterman, M. (2009), Using Market Segmentation for Better Customer Service and More Effective Strategic Planning. *White Paper for Public and Academic Libraries*, p5.

Galigeo. (2012), How to implement Location Analytics. *Galigeo Solutions*.

George, V., and Penka, K. (2010), Analysis of the Problems of Interpolation in GIS medium. Nessebar, Bulgaria.

Goodchild, Maguire, D.J., and Rhind., D.W. (2005), *Geographic Information Systems and Science* (abridged ed.). Carlifonia: Wiley.

Guagliardo, M.F. (2004), Spatial accessibility of primary care: concepts, methods and challenges. *International Journal of Health Geographics*.

IRA. (2009), *Private Moto Insurance Policy.* Nairobi: Insurance Regulatory Authority.

IRA. (2013), *Guidelines on insurance risk.* Nairobi: Insurance Regulatory Authority, Kenya.

IRA. (2013). *The Insurance Act.* Nairobi: Insurance Regulatory Authority.

KNBS. (2013), *Statistical Abstract.* Nairobi: Kenya National Bureau of Statistics.

Longley, P., Goodchild, M., Maguire, D., and Rhind, D. (2011), *Geographic Information Systems and Science.* New York: Wiley.

Mishra, S. (2011), *Geoanalytics for Improved Risk Segmentation in Auto Insurance.* San Fransisco: QPC(Quality Planning).

Morgando, D. (2011, Spring). Insurance Underwriters Turn to GIS for More Accurate Risk Analysis preceeding of ESRI. *ArcNews*.

Siebert, A. (2006), Geographical Information Systems (GIS) and their use in insurance underwriting. *Aquaculture Insurance and Risk Management*, 11.

Winston, W. (2003), Kriging. An Introduction to model building.

Appendix: GIS Prototype Sample Code

```
// Calculates GIS Risk Factor
declare
    gis_uwqtr number(1);
    gis_uwyear number(4);
begin
:gis_rate.gis_rate:= (((:gis_rate.risk_gis_upper/:gis_rate.max_incident)+
(:gis_rate.risk_gis_upper2/:gis_rate.max_incident))/
2)*(:gis_rate.ira_max_rate-:gis_rate.ira_min_rate)+:gis_rate.ira_min_rate;
end;

// Calculate Annual Premium
declare
    gis_uwqtr number(1);
    gis_uwyear number(4);
begin
:gis_rate.annual_prem := (:gis_rate.sum_insured*:gis_rate.gis_rate)/100;
end;
```