Advances in Physics Theories and Applications
ISSN 2224-719X (Paper) ISSN 2225-0638 (Online)
Vol.62, 2017 www.iiste.org

SEMI-STOCHASTIC MIXTURE MODEL FOR PREDICTING THE RATE OF ROAD CARNAGES IN KENYA

Fredrick Ochieng Odhiambo, George Otieno Orwa.

Jomokenyatta University of Agriculture and Technology, Nairobi, Kenya.

P.O.BOX 62000-00200 CITY SQUARE

Nairobi, Kenya

ABSTRACT

In this paper we consider the problem of modeling and predicting the rate of road carnage in Kenya in the presence of randomly changing road conditions. In the literature review, accident prediction rate models are typically regression models and discrete time series models. We study such models and examine their strengths and weaknesses and propose a Semi-stochastic Mixture Model to describe the relation between the highway accidents and the road environment dynamics.The aim of the research paper is to propose a model that captures both the deterministic and stochastic nature of road parameters to explain the cause of high rate of road accidents in Kenya. We apply the proposed model to a simulated data set for the local condition. Our analysis from show that apart from annual average daily traffic (AADT), road curvature is an important component of road carnage.

Keywords: Road system, Semi-stochastic mixture model, road curvature, road carnage,Simulation.

1 INTRODUCTION

Heavy commercial vehicles as means of cargo and human transport play an integral part in Kenyan transport system. Statistics indicate that trucks and other articulated heavy commercial vehicles account for the highest share in total road accidents as well as fatal accidents, Odero, W., Khayesi,M. and Heda, P. M. (2003). Efforts are required to have better understanding of the factors that influence accident.The knowledge about the relationship between the accident and the factors responsible is inadequate, Asingo, P (2004). We have made an effort in this paper to precisely establish what factors are responsible for accidents. The intention here is to predict the rate of road carnage and managed if not eliminate them altogether. The causes of road accidents are such as, roadway geometric design, traffic characteristics, human factor are considered in this paper.

Most of the studies encountered so far have focused on the risk factors such as drinking and driving, restraint systems, and tried to determine their relationship with accident rate. Previous research has shown that accidents involving trucks have a likelihood of producing a severe injury or fatality, Mayes, J. G.(1981). However the relative impacts of various factors; Roadway geometry, traffic characteristics, and other factors have not been seriously factored in. We develop Accident rate Prediction Model (ARPM) for vehicle accidents and use the same in quantifying the factors responsible high rates of accident. Indeed, if such road factors are scientifically identified through acceptable procedures, they can form the basis of policy formulation and decision making with guaranteed positive results.

According to Chin and Quddus(2003), though Multiple Linear Regression models have been applied in the previous studies where the response variables were the numbers of accidents, the Multiple Linear Regression modeling is not appropriate for count data since counts are positive numbers yet the response variable in Multiple Linear Regression analysis is assumed to follow a Normal distribution which covers all numbers on a real interval. This method has limitations of predicting negative numbers of accidents which is not logical. This undesirable statistical property limits Multiple Linear Regression models to describe adequately the random, discrete and non negative accident events. For such reasons, there is need to utilize techniques which can sufficiently describe the specific characteristic of accidents. Such techniques include Poisson Regression, Negative Binomial Regression, Chin and Quddus(2003).

Studies of road accidents in developing countries have indicated that accident rates tends to be particularly high on rural roads, Silyanov V.V.(1973). An analysis of road accidents involving personal injury in Kenya Showed that single vehicle accidents were particularly prevalent on rural roads, being almost 50 percent of the total number of accidents occurring. In this situation it is possible that design features of the road play a significant role, Jacobs D.G(1976). Regression analysis has largely been used to establish and quantify relationships between the dependent variable and one or more independent variables. Some researchers observe the presence of both Poisson and Binomial distribution within the same study, Braga and Bond(2008).

In this paper we take a different perspective than the other approaches, we propose a Semi-Stochastic Mixture Model(SSMM) to describe the relationship between road carnage in Kenya and dynamic environmental conditions. This believed believed to be a more realistic and natural way to describe emerging pattern of accidents that occur on Kenyans road sections. SSMM is envisaged to capture both deterministic and stochastic nature of parameters that characterize road carnage.

We model the accident rate of occurrence as a homogenous process with a rate function that depends depends on several road and environmental parameters. The whole of the Kenyan road system is first considered here taking in consideration the macroscopic and microscopic traffic behavior. The sections with high distribution of black spots per 5 kilometers of road stretch Figure 1.1 is considered is considered as case study.

We now provide an overview of the contributions and findings of this paper. In section 2, we describe the Semi stochastic Mixture model in detail,defining various variables in the model providing their notations. Attempt is also made this section describe the road system using schematic diagrams help identify the queue model for the road section.

In section 3, we describe the accident and road parameters on the road section black spots. In particular we consider the simulated accident data on the Nakuru-Salgaa-Total junction black spot. This is a section of the A104, the busiest road system which span Across East Africa and is plied by heterogenous traffic drawn from countries across East Africa.

In section 4 the SSMM is applied to simulated data set and the prediction results provided. in section 5 we provide summary conclusive remarks. The rate of accident prediction model will be used to improve the location with high frequency of accidents (Black spots) and help reduce the number of accidents.

1.1 Review of models For Road Systems

There exists two major approaches in traffic crash modeling. One is the simple or Multiple linear regression, Equation(1.2) and the Stochastic modeling Equation(1.19) below. Such deterministic models which are widely used are not suitable for an arbitrary and sporadic events like traffic crashes. Much of the early work in the empirical analysis of accident data were done with the use of multiple linear regression models. As has been pointed out, these models suffer from several methodological limitations and practical inconsistencies in the case of accident modeling. To overcome these limitations, researchers have turned to stochastic models, Journal of Science Applied Technology(2012).

Figure 1.1: The sections with high distribution of black spots per 5 kilometers of road stretch

Accidents are random events, non linear forms of models would be more appropriate to use in the calibration of accidents rate prediction model, Maher and Summergil(1996). Poisson model uses the form

$$logE(Y) = a + bx \tag{1.1}$$

Since the Poisson model, Equation(1.8) assumes a non-linear function, it is established to be more effective for crash prediction than linear regression, Maher and Summergil(1996).

However such a model would only be suitable to the extent that there was no over reporting or under reporting. According to Gouvieroux(1999), the Poisson distribution has some severe draw backs, including its equidispersion, this limits its use.

Winkelmann(1996) proposed a Poisson Regression model that takes underreporting into account. In his work, the number of reported events that is y_i results only if absenteeism occuring was assummed to be a Poisson distributed with probability p_i captured by the binomial distribution. Mukopadhyay(1997) derived a mixture of the Negative Binomial and the Binomial distribution. The resulting mixture regression for underreported count is the Negative Binomial Regression model, Mukhopadhyay(1997). Li, Trivedi and Guo(2003) suggested a mixture model of the poisson and negative Binomial regression models that can be used to handle data that is under, over, or accurately reported. According to Li, Trivedi, Guo(2003), misreporting occur when an individual reports the number of events as $y_i, i = 1, 2, 3, ..., n$ Binomial and the Binomial distribution.The resulting mixture regression for underreported count is the Negative Binomial Regression model, Mukhopadhyay(1997). Li,Trivedi and Guo(2003) suggested a mixture model of the Poisson and Negative Binomial Regression models that can be used to handle data that is under, over, or accurately reported. Stochastic modeling has become increasingly important over the last few years, a clear case is the use of the model in life insurance to predict risk and rewards, Wilson, Don(2004). Stochastic modeling builds volatility and variability(randomness) into the simulations and hence provides a better representation of real life in more angles Wilson, Don(2004). A brief discussion of these models are given below.

Review of Multiple linear regression

In this study, a further condition in choosing independent variables was that they should be simple to define and for an engineer working in the field, reasonably easy to measure. As a preliminary investigation of which variables were most closely correlated with accident rate, simple regressions of accident rate on each of the road features individually, were performed. Equations derived were of the form: Equation, (1.2). Where y dependent variable and x is independent variable with a being a constant and b_1 a regression coefficient. However because many of the road design features are inter-related, simple linear regression analysis may give a misleading impression of the relationships that they have with accident rate. Multiple regression, in which the accident rate is expressed as a function of several 'independent' variables simultaneously, is likely to be a better guide. Equations derived were then of the form

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + ... + b_nx_n + e_{ij} \tag{1.2}$$

where

$$y, x_1, X_2, X_n, b_1, b_2, b_n$$

were as above. This model is weak in a number of areas; it made no provision for misreporting, over or under reporting, which is a characteristic of count data, Pararai et al.,(2010). The model assumed that all vehicles in the study had equal chance of failure, naturally this is not the case because the vehicles and their drivers have different disposition to accidents. The parameters used were limited to measurable characteristics only and yet there are other independent variables in accident rates which are not only unpredictable but even when predicted, may not be measurable with any degree of accuracy.

Stochastic models

Stochastic models are logical alternative for events that occur randomly and independently over time. Unlike deterministic models stochastic models assume accident as random event. Okamoto et al.(1989) suggested that the occurrence of traffic crashes follows stochastic distribution. Garber et al.(1990) developed several models to describe the occurrence of crashes in using stochastic modeling techniques, like Poisson Regression (PR), and Negative Binomial Regression (NBR).

Poisson Regression model

Let λ_i be expected number of accident on the stretch corresponding to individual type i λ_i is determined by k exogenous variables or characteristics.

$$X_i = (X_{i,1}, X_{i,2}X_{i,3}, ..., X_{i,k}) \tag{1.3}$$

which represents a priori classification of variables. Since Poisson distribution has only one parameter, namely the mean rate, λ we have little choice but to model λ as a function of x or $\lambda(x)$. We can also write

$$\lambda = exp(\beta X_i) \tag{1.4}$$

Where β is a vector of coefficients ($K \times 1$). The Poisson distribution now becomes

$$P(Y_i = y) = \frac{e^{-exp(x_i\beta)}(exp^{(x_i\beta)})^y}{y!} \tag{1.5}$$

It is important to note that λ_i is not a random variable and that λ_1 is specific to vehicle and driver $No.1$. The model assumes implicitly that the K exogenous variables (factors-measurable) provide enough information to obtain appropriate values of the individual cars and driver probability of accident, Dionee G. Vanasse C(1988). β parameters can be estimated by the maximum likely hood method, Hausman, Hall and Griliches (1984). The model is assumed to contain all the necessary information required to estimate the values of λ_i, there shall be no room for a posterioritarification in the extended Poisson model, Equation(1.8), Deionne G. and Vanasse C(1988). When the vector of explanatory variables does not contain all the significant information, a random variable, error, Equation (1.6) is introduced into the regression component and possibly a constant term. According to Gourieroux and Tragnon(1984), We can write

$$\lambda_i = exp(X_i\underline{\beta} + e_i) \tag{1.6}$$

and better still with a constant component α as

$$\lambda_i = exp(\alpha + X_i\underline{\beta} + e_i) \tag{1.7}$$

.

The α State of road at any one given time of the observation period is assumed to remain the same for all the drivers and vehicles yielding a random α_1.

Equivalently 1.9, can be re-written as

$\lambda_i = exp(X_i\underline{\beta} \times \alpha_i \times u_i)$

where $u_i = exp(e_i)$. If we assume $\underline{u}_i$ follow a gamma distribution with

$E(u_i) = 1$ and $var(u_i) = \frac{1}{a}$ The probability specification becomes,

$$Pr(Y_i = y) = \frac{\gamma(y + a)}{y!\gamma(a)}[\frac{exp(\alpha + x_i\underline{\beta}}{(a)}]^y \tag{1.8}$$

This is a negative binomial distribution with parameters a and $exp(\alpha + x_i\underline{\beta})$ and $E(Y_i) = exp(\alpha + x_i\underline{\beta})$

$$Var(Y_i) = exp(\alpha + x_i\underline{\beta})[1 + \frac{exp(\alpha + x_i\beta)}{a}] \tag{1.9}$$

Clearly $Var(Y_i)$ in the Equation(1.9) is a non linear increasing function of $E(Y_i)$.

Poisson Distribution model

Denote the number of accidents $A_{i,t}$ for vehicle i at time t , hence

The rate of accidents

$$\lambda_{it} = exp(\beta_o + X_{i,t}{}'\beta + e) \tag{1.10}$$

According to Gouvieroux(1999), the Poisson distribution has some severe draw backs such as its equidispersion and this limits its use. It also assumes that the vehicles have same accident frequency, this may not be the case on different road stretches. It is only plausible that the rate may differ across subgroups of the data sometimes defined by geography and time.

Negative Binomial model

If it is assumed that the parameter λ vary among individual road stretches then a combination of them at a point of connectivity results into a more general model that will allow λ to vary among individuals and hence perhaps a better λ for the mixture of $\lambda\prime_s$. If we assume λ is also a random variable and follows a gamma distribution with parameter a and $\frac{1}{b}$ as proposed by Greenwood and Yule(1920), Bichsel(1964) and Seal(1969), the distribution of the number of accidents during a given period will be given by

$$\Pr(Y_1 = y) = \frac{\gamma(y + a)}{y!\gamma(a)} \times \frac{1/b}{(1 + 1/b)^{y+a}} \tag{1.11}$$

This corresponds to a negative binomial distribution with

$$E(Y_i) = \overline{\lambda}$$

and

$$Var(Y_i) = \overline{\lambda}[1 + \frac{\overline{\lambda}}{a}] \tag{1.12}$$

, where

$$\overline{\lambda} = ab$$

and the parameters a and $\frac{1}{b}$ can be estimated by the method of moments or by maximum likely hood method. The model assumes that variable Y_i are independent car accidents and λ vary between individuals.

The negative binomial regression model allows for over dispersion in the model and can be used to quantify various parameters more effectively.

Li, Trivedi and Guo (2003), suggested a mixture model of the Poisson and negative binomial regression models that can be used to handle data that that is under, over and accurately reported.

Generalized Poisson Regression model

Famoye(1993), derived a generalized Poisson regression(GPR) model given by Equation (1.21)

$$P(Y = y_i) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_i} \frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \exp\left[\frac{-\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right] \tag{1.13}$$

for $y_i \geq 0$ and μ_i is the log-link function.

Poisson-Gamma Model

This has properties similar to Poisson model, Equation (1.6) with dependent variable y_i modeled as a Poisson variable with mean μ_i and with the model error assumed to follow a gamma distribution. Taking into account over-dispersion that is commonly observed in discrete or count data, Lord et al.,(2005) considered a Poisson distribution,

$$g(y_i; \lambda_i) = \frac{e^{-}\lambda_i \times \lambda_i}{y_i!} \tag{1.14}$$

When $\alpha < 0$, the GPR can be used for under dispersed data.

Mixture Model of the Poisson and Negative binomial regression Models

Li, et al.(2003) Suggested a mixture model of the Poisson, Equation (1.8) and negative binomial regression model, Equation(1.9) that can be used to handle data that is under, over and accurately reported. The negative binomial regression took care of of the accurate counts while the Poisson regression model took care of the under reported and over reported data.

zero inflated negative binomial regression models

Transportation safety analysts have typically justified the use of Zero Inflated (ZI) models because of the improved statistical fit compared to traditional Poisson and NB models. Zero Inflated regression models are two regime models. First probability model governs whether a count number is zero or positive number, known as inflated model. Then the positive part of the distribution is described by suitable stochastic distribution, known as base model. The other is structural zeros(true zeros) which are inevitable and are part of the counting process. Beyond this, we base our choice on the model providing the closest fit between the observed and predicted values.

In practice, even after accounting for zero inflation,the non-zero part of the count distribution is often over-dispersed, Green W.H(1994). Green W.H(1994), described an extended version of the negative binomial model for excess zero count data referred to as the Zero-Inflated Negative Binomial (ZINB).

$$P(Y = y_i) = (1 - P_i)\frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})}\frac{(\alpha\mu_i)^{y_i}}{(1 + \alpha\mu_i)^{y_i+1/\alpha}}$$

is more appropriate than the ZIP. It has been established that the ZIP parameter estimates can be severely biased if the non-zero counts are over-dispersed in relation to the Poisson distribution.

The Poisson regression model has been traditionally considered as the starting point in modeling crash data, in this case we assume the mean of accident occurrence is equal to the its variance, Miaou S. P.(1994) (this is the equal dispersion).

Then if $f_1(y_1|\underline{\vartheta_1})$, a probability distribution assume by Trunk T_i be Poisson, then

$$P\left(y_i\right) = \frac{\exp\left(-\theta_i\right)\theta_i{}^{y_i}}{y_i!} \tag{1.15}$$

where $y_i$is the observed number of counts for trunk i=1,2,. . .,n and λ_i is the mean of the Poisson distribution. Such assumption ignores the heterogeneous nature of traffic, also in much of the crush data , the variance is greater than the mean (over-dispersion), clearly some trunks T_i therefore shall display over-dispersion resulting from extra variation in crash means across the sites and excess zero counts. This way, applying a Poisson regression model for some trunk T_i is inaccurate and may result in underestimation of the standard error of the regression parameter and narrow Confidence intervals, leading to a biased selection of co variates, Miaou, S. P.(1994).

The NB model takes the unobserved heterogeneity of the Poisson mean into account by allowing the variance to differ from the mean, that is

$$var[y_i] = E\left[y\right]1 + \alpha E\left[y_i\right] = E\left[y_i\right] + \alpha E\left[y_i\right]^2 \tag{1.16}$$

for α=0, the negative binomial reduces to the Poisson model.We are carefull not to let α=0, in this second scenario. The NB model assumes that unobserved heterogeneity across road sections follows a gamma distribution, while crashes within sites are Poisson distributed.

For the road sections, The competing models were the Negative binomial model and the traditional Poisson model. The suitability of one over the other is determined by the statistical signicance of the estimated coefficient α, Mehdi H.et al(2012).

When α is not significantly different from zero, the Negative Binomial(NB) model is the correct choice, Mehdi H. et all (2012). However with excess zeros in crash data with resulting over dispersion, NB model cannot handle the overdispersion due to high amounts of zeros. A better model would suitably characterize the distribution along such a node that exhibits excess zeros. The Zero inflatio(ZI) models are suitable for such road trunks, particularly some trunks will exhibit the Zero inflated Poisson (ZIP) model, while still others will exhibit Zero inflated negative Binomial(ZINB) model

Clearly therefore given that trunks T_1 has a Poisson distribution model parameter λ_i, T_2,trunk has a Negative Binomial model parameters T_3, a Zero inflated Poisson model and still . . ,T_K, has Zero infalted Negative Binomial model parameters.

The trunks T_1,T_2, T_3,. . . ,T_K contributes traffic to a road segment S_i sampled randomly from different models, consequently a mixture of distributions ex at S_i.

Let P_i be the probability of excess zero for the section i and $(1 - P_i)$ be probability of accident (non-zero) counts derived from the Poisson probability distribution. For count $y_i = 0$ The probability density for the ZIP model is

$$P\left(Y = y_i\right) = P_i + \left(1 - P_i\right)\exp\mu_i \tag{1.17}$$

and for the non zero road carnage counts count $y_i{\geq}0$ The probability density for the ZIP model is where y is the number of road carnages for trunc T_i and μ_i is its expected crash frequency.μ_i is a function of road trunc covariates, thus μ_i= $exp(X_i'\beta_i')$. The probability of being in the zero-crash-state P_i is often fitted using logistic regression model, as follow:

$$logit\left(P_i\right) = log\left(\frac{P_i}{1 - P_i}\right) = Z_0^i\gamma \tag{1.18}$$

$$logit(P_i) = log(\frac{P_i}{1 - Pi}) = Z_0^{i'}\gamma = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + ... + \gamma_N Z_N$$

where $Z = (Z_1, Z_2, Z_3, ..., Z_N)$ is a function of the exploratory variables and $\gamma = \gamma_1, \gamma_2, ..., \gamma_N$ is the estimable coefficient.

and $log(\mu_i) = X_i'\beta$

For the trunks T_j with the Zero Inflated Negative Binomial probability density with mean μ_j and dispersion parameter, α_i.The Probability density function for For count $y_i = 0$ The probability density for the ZINB model given, μ_i and dispersion parameter is , α ,is given by:

$$P(Y = y_i) = P_i + (1 - P_i)\frac{1}{(1 + \alpha\mu_i)^{\frac{1}{\alpha}}}$$

for the road carnage count $y_i = 0$

and

$$P(Y = y_i) = (1 - P_i)\frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})}\frac{(\alpha\mu_i)^{y_i}}{(1 + \alpha\mu_i)^{y_i + 1/\alpha}}$$

for the road carnage count $y_i > 0$.

To estimate the parameters of the Zero-Inflated models we use the maximum likelihood method.

Road Sections and Road Trunks

The road section $S_i i = 1, 2, .., n$ draw traffic from the access trunks $T_i i = 1, 2, ..., k$.

Suppose standard Poisson distribution mean λ_1,is the suitable model for the road carnage along road trunk T_1 then

$$P(Y_i = y) = \frac{e^{-exp(x_i\beta)}(exp^{(x_i\beta)})^y}{y!} \tag{1.19}$$

where

$$\lambda_i = exp(\beta X_i) \tag{1.20}$$

and suppose road carnage along road trunk T_2 is modeled suitably by the standard Negative Binomial (NB) distribution, then

$$P\left(Y = y_i\right) = \left(\frac{\mu_2}{1 + \alpha\mu_2}\right)^{y_i}\frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!}\exp\left[\frac{-\mu_2\left(1 + \alpha y_i\right)}{1 + \alpha\mu_2}\right] \tag{1.21}$$

where y_i is considered the number of road accident on trunkT_2 per annum. If Zero Inflated Poisson (ZIP) distribution models the distribution of accidents along the joining trunk T_3 , then

$$P(Y = y_i) = P_i + (1 - P_i)\frac{1}{(1 + \alpha\mu_i)^{\frac{1}{\alpha}}}$$

for the road carnage count $y_i = 0$.

Similarly if Zero Inflated Negative Binomial were the suitable model for the road trunk joining at T_4, then

$$P(Y = y_i) = (1 - P_i)\frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})}\frac{(\alpha\mu_i)^{y_i}}{(1 + \alpha\mu_i)^{y_i + 1/\alpha}}$$

for the road carnage count $y_i > 0$.

Where $\mu_j$is the mean and α_i is the dispersion parameter.

The Mixture distribution at road section

Clearly, the traffic at any main highway section S_i is sampled from the four independent accident count models, hence at S_i, the distribution of the accident is sampled from the Poisson model , Negative Binomial(NB) model , Zero-Inflated Poisson model and Zero inflated Negative Binomial (ZINB) model. Hence,

$$P(Y_i = y) = \frac{e^{-exp(\mu_i\beta)}(exp^{(\mu_i\beta)})^y}{y!} + \left(\frac{\mu_i}{1 + \alpha\mu_i}\right)^{y_i}\frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!}\exp\left[\frac{-\mu_i\left(1 + \alpha y_i\right)}{1 + \alpha\mu_i}\right] \tag{1.22}$$

$$+P_i + (1 - P_i)\frac{1}{(1 + \alpha\mu_i)^{\frac{1}{\alpha}}} + (1 - P_i)\frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})}\frac{(\alpha\mu_i)^{y_i}}{(1 + \alpha\mu_i)^{y_i + 1/\alpha}}.$$

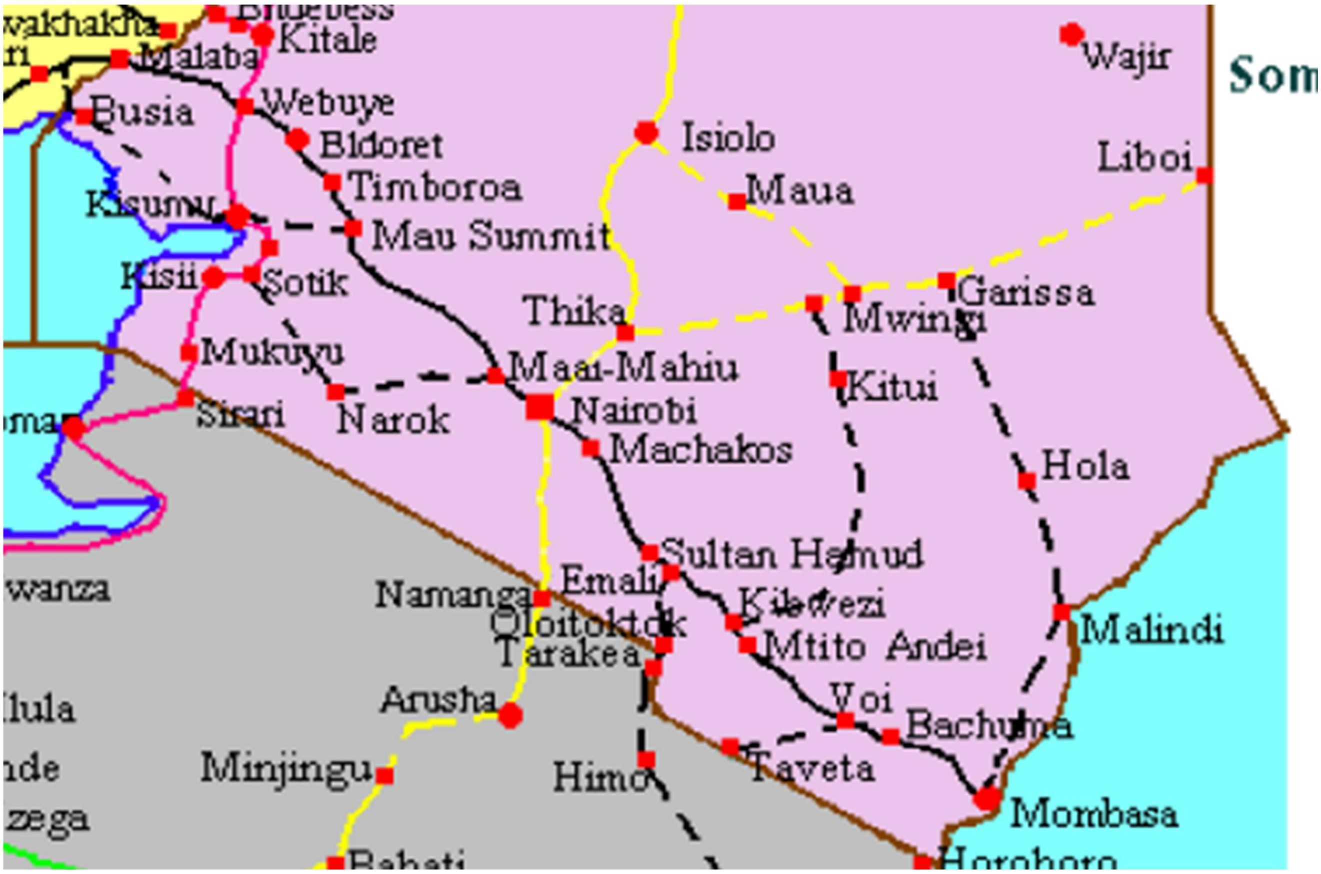


Figure 1.2: The Kenya road system of the major road A104 with rate of accident $A_i = f(AADT, SS, LW, SW, NL, HC, VC, TD)$

or perhaps a product of the individual Trunk distributions.

Clealy such a function would be cumbersome to manupulate for accurate probabilities of accidents on any section S$_i$ and ignores the unique nature of the very point S$_i$ of the road section. Better model is therefore developed.

1.2 Review of the Kenya Road system

In this chapter, the Kenya road system Figure 1.2, is reviewed and described using the schematic diagrams and a suitable model is proposed.

Road systems depicts a complex process that results in a built up of heterogenous traffic whose model of rate of accident can no longer reflect the individual component distribution at the time of joining the main high way. For the Kenyan situation, there are series of accident black spots $B_1, B_2, \ldots B_n$ along the major highways.

Close study of the Kenyan road system show that it is marked by two major road networks- the Mombasa to Malaba, Mombasa to Busia(Mackinon Scarter road)-Historicaly the first 1000 kilometer ox-cart earth track form Mombasa to Busia whose construction was started in 1890. The other road begins from the border between Kenya and Tanzania to the south and all the way to Moyale at the border between Kenya and Ethiopia to the North. Other county roads join in severally at distintinct points. Vehicles that join the road arbitrarily must adjust to the phenomenon changes which include High vehicle densities, very wide roads and high speeding traffic.

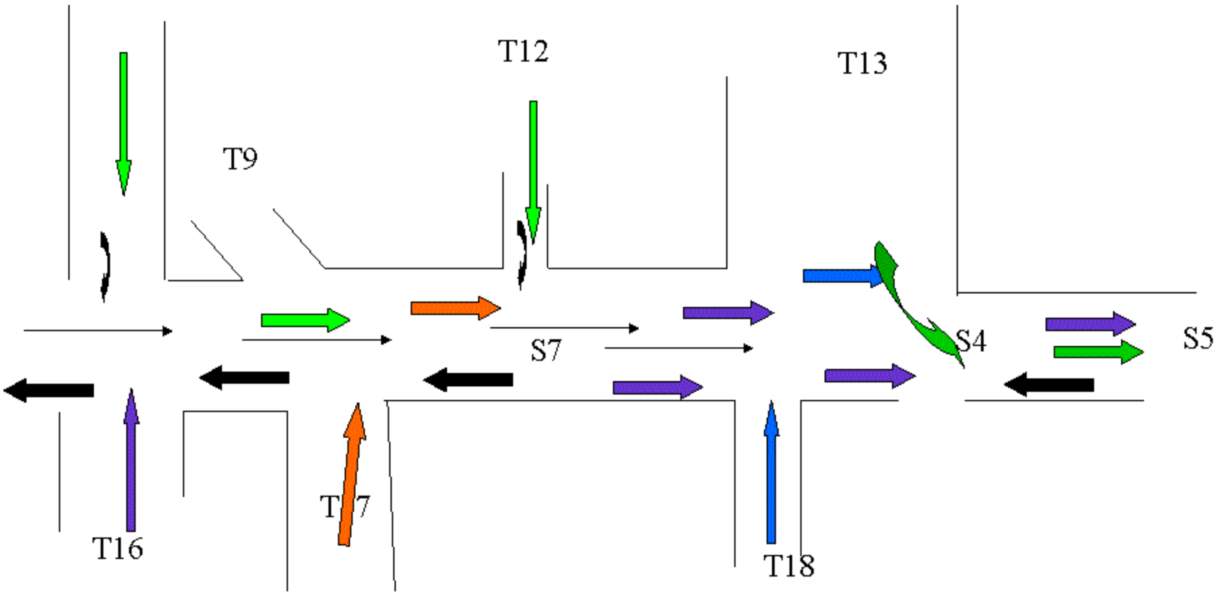


Figure 1.3: Two way traffic road section joined in by many minor feeder roads

Schematic illustration of the road system

The black spots $B_1, B_2, . . . B_n$ along the sections of the road sections S_i are therefore identified on the basis of an existing clasification by the Kenya National Roads Authority(KeNHA).

The black sports are examined to determine the nature of distribution of accidents within the proximity of the black sports, and the rate of accidents on every section of a black sport within a stretcth of 1 kilometres of the black sport, 500metres either direction .

Trunks T_1,T_2, T_3,. . . ,T_K, that contribute traffic into the highway figure 1.3. The traffic along the road section stretch S_i is in effect a built up from a combination individual distributions. These are either, Poisson , Negative Binomial(NB) model, Zero-Inflated Poisson model or Zero inflated Negative Binomial (ZINB) model.

It is asumed that to any road section S_i, the trunks T_1,T_2, T_3,. . . ,T_K, contribute to the traffic dynamics, the situation worsens when traffic proceed from a one way road system into a two way road system as illustrated below, figure 1.4.

Their exists conflicts among major streams of traffic, Ruskin and Wang (2002). Sections immediately after some major intersections are of significant impact and have varried capacities from other sections. The myth of purely behavioral explanation on part of the driver/rider for the growing burden of road traffic accidents in Kenya is inadequate, Manyara G C(2014). Structural and physical components of the road system is therefore considered .

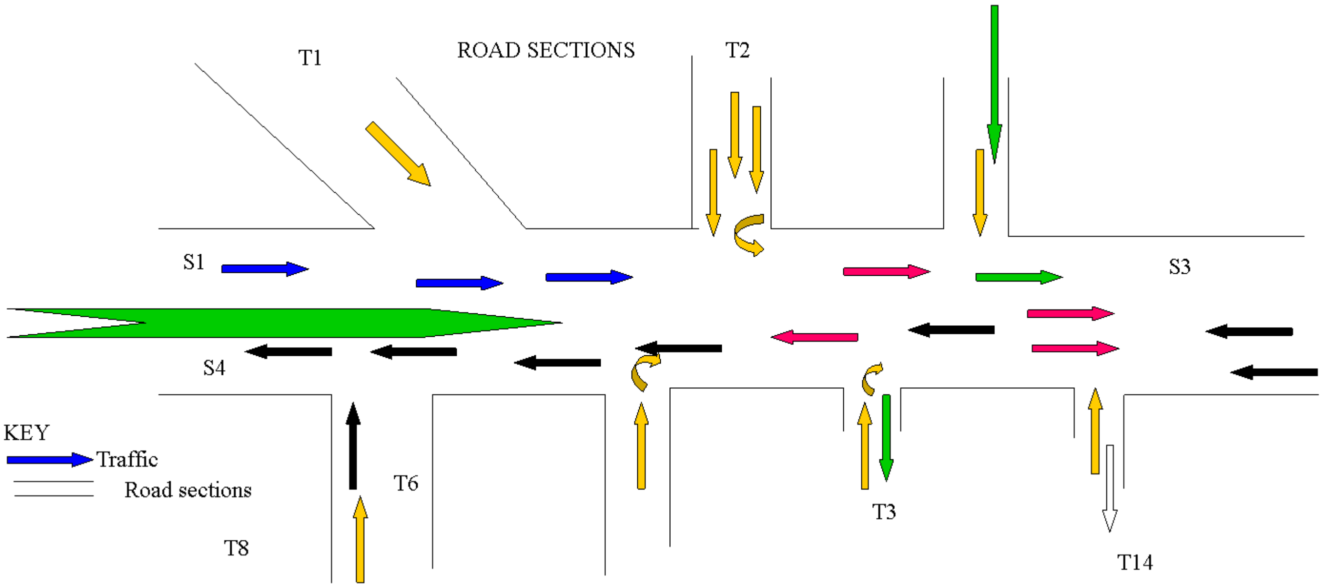


Figure 1.4: One way road segments merge into a two way

Road carnage contributing conditions

Within the model of road carnage, eight categories of contributing condition are presented. Equally eight black spot are considered sellected across the country road system. We therefore obtain an nxn matrix. According to (Hussein M. , 2009) if number of equation equals the number of unknowns, the system is defined and the solution can be obtained.

Conceptual framework for road carnage in kenya

On basis of the work conducted on the first part of this chapter, a conceptual framework, figure 1.5, for determining the contribution of road parameters to rate of road carnages in kenyan road system as a whole was developed. This involved identifying potential road crash related road parameters that could exist on a stretch of road and that could be used as variable in the road carnage prediction and control.The framework contains both the methods with which to collect and analyse road carnage related data and a few suggested carnage mitigation measures to reduce, eradicate or manage the road users own contribution to the cited cases.

The road system

The road network studied is composed of nodes and links. A node in this case seperates parts of the same road with different characteristics for example, a dual carriege way may narrow to a single stream or even several intersections flowing into a carriege way an a point and exits at another point, Chao Yang(2013). In the case of Kenya such a road sytem is evidently turning into killer roads figure 1.6.

The road system and traffic schematic diagrams generally depicts a process where customers (road users) arrive at a road sections according to some distribution with a mean arrival rate λ_i.

Many sections of the road system are largely single lanes, clearly suggesting a single server system. Customers arrive at this sections from a population size that cannot be determined in advance, thus an infinite population. A part from a few cases like the, most modern Thika Superhighway and very few sections of the roads , customers (traffic) arrive in a single lane for traffic towards a specifc direction. Such traffic form a single queue.

The number of servers represented by number of lanes (servers) are for greater length, single lanes. The capacity of the system is finite. The population, P from where the system draw clients is infinite. i.e $p = \infty$.

Queue Discipline

It is assumed that given the single lanes characterising greater lengths of road sections, the queue discipline is first come first served with, occassional priorities being given to special vehicles on emergency operations and Hospital Ambulances, i.e the discipline is FIFO. With this scenario every driver try to be ahead of the other , " beat the jam". The Service time for these customers are therefore independent with some exponential distribution mean say, μ.

The study proposed an $M/M/1$ process for the road system. Kendall G D (1953).

1.3 Development of the mathematical Model for the Kenyan road system.

Model components

Every dorminant loop of the road system is examined and the following representative variables are selected as independent variables.

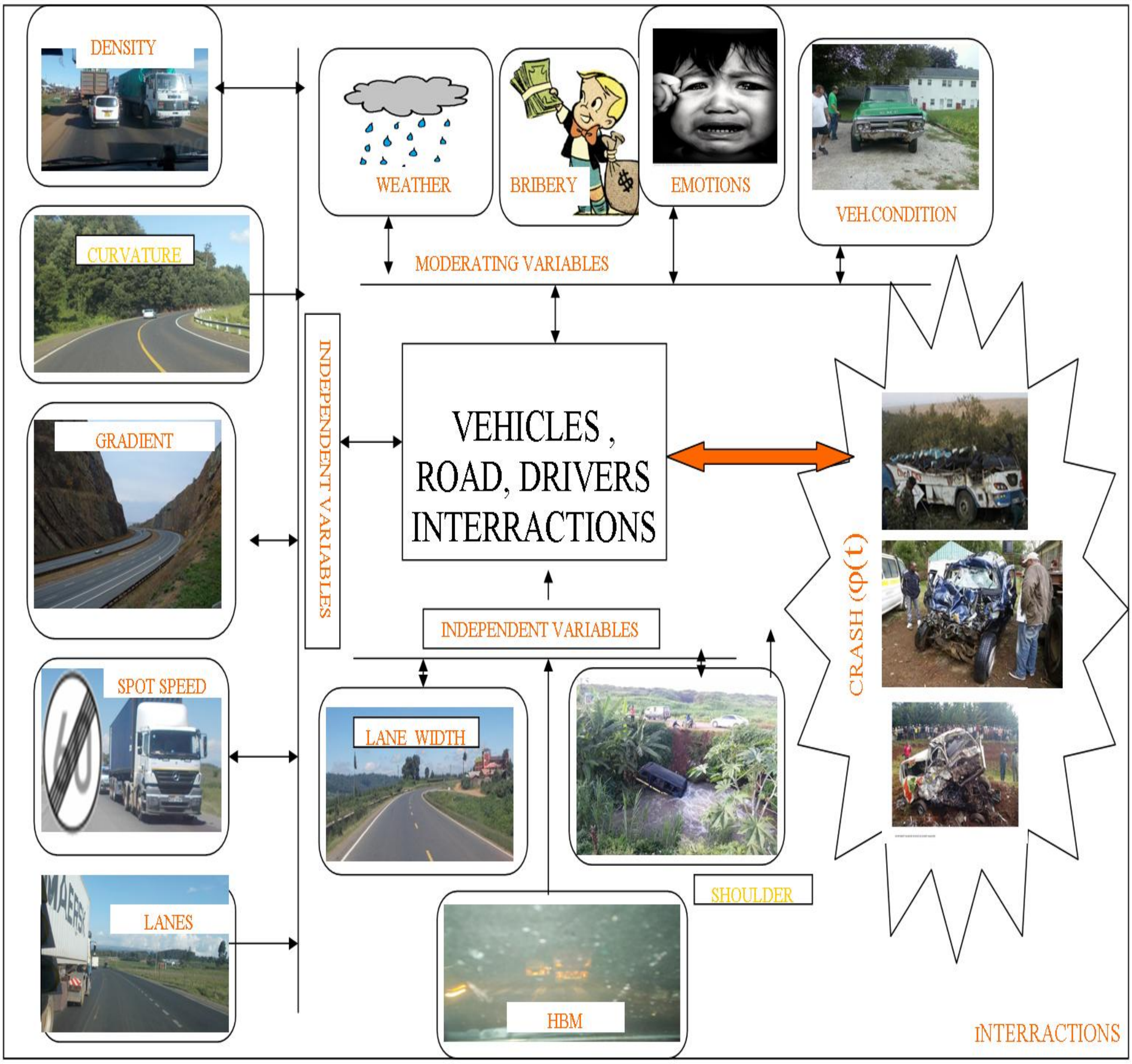


Figure 1.5: Conceptual framework for road carnage in kenya

Figure 1.6: Killer roads

i. Shoulder width (SW)

Shoulders provides an area along the highway for vehicle to stop, swerve particularly during emergency. Slow moving vehicles, pedestrians can use the shoulder and keep the carriageway free for heavy and fast moving vehicles, Zegeer et al. (1981)

A report by Zegeer et al. (1981) on the "Effect of lane and shoulder widths on accident reduction on rural two-lane roads" indicated that a paved shoulder widening of 2 feet per side reduces accidents by 16%. Shoulder width has been a parameter with significant influence on safe operations of traffic and hence selected as a variable. The area under consideration shows a wide variation in the shoulder width from 0 to 3.10 m.

ii. Lane width (LW)

Traffic flow tends to be restricted when lane width reduces. This is because vehicles have to travel closer together in lateral direction. Lane width, therefore, is treated as an important parameter. It has been found that accident rate reduces as lane width increases.

iii. Traffic volume (AADT)

Traffic volume is believed to have considerable impact on the crash rate Williams Ackaah, Mohammed Salifu (2011). For this study annual average daily traffic (AADT) is used as a parameter to indicate traffic volume. Traffic volume, converted in to passenger car unit (pcu), for various segments was collected and included in the model.

Curvature(C),

When the curves are disregarded the consequences become unbearable. Motorists on keep left traffic policy are on a greater danger when negotiating a right bend.

Figure 1.7: Danger signals by continuos yellow line mean nothing to drivers

Number of lanes (L)

Road sections with few lanes pose risks to motorists. Similar right bend curves pose greater risk of road carnage in Kenya .

Kenya is known for having some of the most dangerous roads in world,WHO(2015), dangers seem evident on curves, consequences of ignorring curvatures can be fatal even to the well trained drivers figure1.5.

Curvatures with near Zero Shoulder width compounds the problem

Shoulders provides an area along the Highway for vehicle to stop, particularly during emergency or swerve in the event of an unlikely encounter . Slow moving vehicles, pedestrians can use the shoulder and keep the carriageway free for heavy and fast moving vehicles. A right bend curve without a reasonable shoulder figure 1.7, clearly remain a death trap.

Lane Width(W)

Traffic flow tends to be restricted when lane width reduces. This is because vehicles have to travel closer together in lateral direction. Lane width is hence treated as an important parameter. It has been found that accident rates reduce as lane width increases.

Spot speed(S)

Speed and travel time are the most common indicator of the performance of a traffic facility. Spot speed is one of the major parameter that is used as an indicator of traffic performance. Spot speed of a location has considerable impact on the traffic safety of the area. The data collected shows a wide variation in the spot speed from 25kmph to 60kmph.

Clearly therefore the rate of accident alon trunk T_i, $i = 1, 2, .., n$, and consequently sections S_i, are a factor of the following variables, Table 1:

1. Road Shoulder width (S),
2. Lane Width (W)
3. Total Traffic Volume(AADT),
4. Roadway Number of Lanes (Lane),
5. Vertical Alignment,(VC)
6. Grade on Curve (C)
7. Sight Distance (Feet/m),
8. Shoulder Width (Foo)(SW),

Let the number of road accidents along trunk i be denoted by A_i, clearly A_i is a function of the above variables for different road stretches, hence A_i can be expressed as

$$A_i = f(AADT, AD, W$$

$$NL, SC, VA, C, SD, S, VC) \tag{1.23}$$

and therefore motivativated by IHSDM, a tool that was developed by the USA Federal Highway Administration(FHWA) for assing the safety impacts of a geometric design, Koorey et al (2009) , also Zeegar et al (1992). The six variables are used in the model development with additional two variables of serious concern to the Kenya's case.

$A_i = f(AADT, SS, LW, SW, NL, HC, VC, TD)$. The Similar study conducted by Huang Su P.(2013) used, A= f(P,V,R,E) where A is variable representing number of accidents. P represents Physical features, V represents the vehicle condition, R is used to represent road condition parameters where as E was used to represent prevailing Environmental factors such as the weather variations. The variables for the model in this study are more detailed.

Boundary condition for the variables

Roadway Number of Lanes (Lane) $\geq$1
Lane Width(Foot/m) $>$1
Vertical Alignment $>$1
Grade on Tangent (%)
Grade on Curve (%),
Sight Distance (Feet/m)
Horizontal Alignment Degree on Curve (Degree),
Shoulder, Width (Foot/m),
Surface Condition (Good/Bad),

If the rate of accident along the road stretch T_1 is denoted by $y_1^{/}$, and the rate of accident along the road stretch T_2 denoted by $y_2^{/}$, proceeding this way, the rate of road accident along road stretch T_i can be denoted by $y_i^{/}$ where

$$y_i^{/} = a_{11}y_1 + a_{12}y_2 + ... + a_{1i}y_i$$

for $i = 1, 2, ..., n$. Considering varried trunks that join into a main higway, a linear system of equation is proposed.

$$y_1^{/} = a_{11}y_1 + a_{12}y_2 + ... + a_{1n}y_n$$

$$y_2^{/} = a_{21}y_1 + a_{12}y_2 + ... + a_{2n}y_n$$

$$y_3^{/} = a_{31}y_1 + a_{12}y_2 + ... + a_{3n}y_n$$

$$y_4^{/} = a_{41}y_1 + a_{12}y_2 + ... + a_{4n}y_n$$

$$y_5^{/} = a_{51}y_1 + a_{12}y_2 + ... + a_{5n}y_n$$

....

$$y_n^{/} = a_{n1}y_1 + a_{n2}y_2 + ... + a_{nn}y_n \qquad (1.24)$$

where $y_i^{/}$ stands for the derivative with respect to t. In the sequel, all the coefficients of a_{ij} of the system are assumed to be real numbers.

Summary of model variables

The system can be reduced to a single homogenous linear constant-coefficient n$^{\text{th}}$ order equation. The equation $sql1$ is conventionally written as ,

$$y_1^{/} = ay \qquad (1.25)$$

where y =$(y_1, y_2, ..., y_n)^T$ is the column vector of the unknowns and $a = (a_ij)$ is the matrix of the equation coefficient.
Let $y_k = (y_{k1}, y_{k2}, ..., y_{kn})^T$ be linearly independent particular solution of the monogamous system, equation (2.2) .The general solution of the homogenous system $y_1^{/} = ay$ expressed as a vector of variables and constants

$$y = C_1y_1 + C_2y_2 + ... + C_ny_n \qquad (1.26)$$

The system is assumed to be linear differential, constant coefficient, homogenous and non- autonomous. The Picard's existence and uniqueness theorem is adopted here. It is assumed that the system of equation is of the form

$$\underline{y}^{/} = f(t, \underline{y}) \qquad (1.27)$$

and

$$f(t, \underline{y}) = A(t)\underline{y} \qquad (1.28)$$

where

$$A(t) = [a_ij(t)] \qquad (1.29)$$

an nxn matrix of functions.

Theorem 1. *(Existence and uniqueness)*

Suppose that n×n matrix function function A(t) and the n×1 matrix $\underline{q}$(t) are both continuous on an interval, I in R.
Let $t_0\varepsilon$ I. Then for every choice of the vector $\underline{y_0}$ initial value problem

$$\underline{y}^{/} = f(t, \underline{y}) \qquad (1.30)$$

has a unique solution $\underline{y(t)}$ which is defined on the same interval I, Coddington E.(1961).
For the system (1.24) and applying the existence and uniqueness theorem, there is a solution

$$y(t) = [y_1(t), y_2(t), ..., y_n(t)]^{/} \qquad (1.31)$$

Table 1: variables for modelling

VARIABLE	ABREVIATION	NOTATION
Annual rate of accident at spot i	ARRA	R_i'
Average daily traffic	AADT	y_1
Curvature	C	y_2
Sight distance	SD	y_3
Hours before midnight	HBM	y_4
Shoulder width	S	y_5
Spot Speed	SS	y_6
Number of Lanes	L	y_7
Lane Width	LW	y_8

and the functons $y_1(t)$,$y_2(t)$,$y_3(t)$,...,$y_n(t)$ have a laplace transform. The laplace transform is adopted here to determine the general solution.

$$Y_1(s) = L(y_1)$$

$$Y_2(s) = L(y_2)$$
$$Y_3(s) = L(y_3)$$
$$Y_4(s) = L(y_4)$$
$$Y_5(s) = L(y_5)$$
$$............$$

$$............$$

$$Y_n(s) = L(y_n) \tag{1.32}$$

The equation above yields a system of algebraic equation of the form

$$sY_1(s) - y_1(0) = \alpha_1 y_1(s) + \beta_1 y_1(s) + \gamma_1 y_1(s) + \rho_1 y_1(s) + ... + \psi_1 y_1(s)$$
$$sY_2(s) - y_2(0) = \alpha_2 y_2(s) + \beta_2 y_2(s) + \gamma_2 y_2(s) + \rho_2 y_2(s) + ... + \psi_2 y_2(s)$$
$$sY_3(s) - y_3(0) = \alpha_3 y_3(s) + \beta_3 y_3(s) + \gamma_3 y_3(s) + \rho_3 y_3(s) + ... + \psi_3 y_3(s)$$
$$sY_4(s) - y_4(0) = \alpha_4 y_4(s) + \beta_4 y_4(s) + \gamma_4 y_4(s) + \rho_4 y_4(s) + ... + \psi_4 y_4(s)$$

$$..$$

$$sY_n(s) - y_n(0) = \alpha_n y_n(s) + \beta_n y_n(s) + \gamma_n y_n(s) + \rho_n y_n(s) + ... + \psi_n y_n(s) \tag{1.33}$$

Let

$$\underline{Y} = [y_1, y_2, ..., y_n]^{/} \tag{1.34}$$

the sytem of equation (1.34) can be written in matrix form

$$s\underline{Y}(s) - \underline{y}(0) = AY(s) \tag{1.35}$$

which is then easily written as the matrix equation

$$(sI - A)\underline{Y}(s) = \underline{Y}(0) \tag{1.36}$$

Assuming that $(sI - A)$ is invertible we can solve equation (1.34) for $\underline{\text{Y}}$(s) and apply the inverse laplace to the entries $\underline{\text{Y}}$(s) to find the unknown function, $\underline{\text{y}}$(t).
$(sI - A)$ is then written in a square matrix form as

$$S - \alpha_1 - \beta_1 - \gamma_1 - \rho_1 - \psi_1 - \delta_1 - \omega_1 - \phi_1$$
$$-\alpha_2 + S - \beta_2 - \gamma_2 - \rho_2 - \psi_2 - \delta_2 - \omega_2 - \phi_2$$
$$-\alpha_3 - \beta_3 + S - \gamma_3 - \rho_3 - \psi_3 - \delta_3 - \omega_3 - \phi_3$$
$$-\alpha_4 - \beta_4 - \gamma_4 + S - \rho_4 - \psi_4 - \delta_4 - \omega_4 - \phi_4$$
$$-\alpha_5 - \beta_5 - \gamma_5 - \rho_5 + S - \psi_5 - \delta_5 - \omega_5 - \phi_5$$
$$-\alpha_6 - \beta_6 - \gamma_6 - \rho_6 - \psi_6 + S - \delta_6 - \omega_6 - \phi_6$$
$$-\alpha_7 - \beta_7 - \gamma_7 - \rho_7 - \psi_7 - \delta_7 + S - \omega_7 - \phi_7$$
$$-\alpha_8 - \beta_8 - \gamma_8 - \rho_8 - \psi_8 - \delta_8 - \omega_1 + S - \phi_8 \tag{1.37}$$

let

$$p(s) = det(sI - A) = S^8 - Tr(A)s + detA \tag{1.38}$$

Clearly p(S) is a non zero polynomial function of degree 8 and hence det(sI-A).
For the purpose of laplace transform the interestes is focused on (sI-A), Equation (1.35) is solved for $\underline{\text{Y}}$(S) to get

$$\underline{Y}(S) = (sI - A)^{-1}\underline{y}(0) \tag{1.39}$$

but

$$(sI - A)^{-}1 = \frac{1}{p(s)}[C_{ij}(s)] \tag{1.40}$$

$$= [\frac{C_{ij}}{p(S)}]$$

if $Z_1(s), Z_2(s), Z_3(s)...Z_8(s)$ be the first upto the eigth columns of $(sI - A)^{-}1$ respectively, clearly each entry of $Z_1(s), Z_2(s), Z_3(s)...Z_8(s)$ is a rational function of S.

Taking

$$Z_1(t) = L^{-1}Z_1(s), Z_2(t) = L^{-1}Z_2(s), Z_3(t) = L^{-1}Z_3(s), ..., Z_8(t) = L^{-1}Z_8(s) \tag{1.41}$$

The entries of Zwill be of the form

$$C_1e^{r_1t} + C_2e^{r_2t} + C_3e^{r_3t} + C_4e^{r_4t}... + C_8e^{r_8t}$$

if $P(S)$ has distinct roots then $r_1, r_2, r_3, r_4r_5, r_6, r_7$, and r_8 are all unique values, with $C_1, C_2, C_4, C_5, C_6, C_7$ and C_8 being appropriate constants.

Consider equation (2.1) it can be shown that

$$\underline{Y}(s) = \underline{Z}(s)\underline{y}(0) \tag{1.42}$$

hence

$$\underline{Y}(s) = y_1(0)Z_1(s) + y_2(0)Z_2(s) + ... + y_8(0)Z_8(s)$$

and applying laplace transform

$$\underline{y}(t) = y_1(0)Z_1(t) + y_2(0)Z_2(t) + ... + y_8(0)Z_8(t)$$

let

$$\underline{Z}(t) = [Z_1(t), Z_2(t), .., Z_8(t)]$$

which implies that

$$\underline{Z}(t) = L^{-1}((sI - A)^{-1}) \tag{1.43}$$

The solution y(t) of the system (1.32) can be expressed in matrix form $\underline{y}(t) = \underline{Z}(t)\underline{y}(0)$ whic implies that $= \underline{Z}(t)\underline{y}_0$ given $\underline{y}^{/} = Z(t)\underline{y}$ with the initial condition $\underline{y}(0) = \underline{y}_0$ and $Z(t) = L^{-1}((sI - A)^{-1})$. According to Existence and uniqueness theorem , the the unique solution to the initial value problem equation(1.32) is

$$y(t) = e^{At}y_0$$

where

$$e^{At} = L^{-1}((sI - A)^{-1})$$

.

The difficulty in manually determining $((sI - A)^{-1})$ is at this point acknowledged, especially for the $8X8$ matrix in this study. The matlab program is therefore adopted for the subsequent data manipulation.

In this applications, time t plays the role of the independent variable, and the associated system of differential equations is conventionally written in the following notation:

Clearly for the prediction of the rate of road carnage for the road system depicted by the scenario , the model is

$$X_n = C_1A_{n_1}e^{\lambda_1t} + C_2A_{n_2}e^{\lambda_2t} + C_3A_{n_3}e^{\lambda_3t} + ... + C_mA_{n_m}e^{\lambda_mt} \tag{1.44}$$

The equation can be extended to non linear systems. For the purpose of this study linears system are considered.

Table 2: Simulated data for the road parameters with IVP

BS	R_i	T	C	SD	HBM	S	SS	L	LW
BS1	3	4999	46	2	9	2	48	3	3
BS2	5	4800	50	3	10	2	50	3	2.5
BS3	2	5000	51	5	6	0	40	3	3
BS4	7	4866	53	5	6	1	50	2	2.7
BS5	3	5006	52	3	6	2	48	3	3
BS6	7	4945	52	8	8	1	50	4	3
BS7	8	5117	52	4	8	1	6	3	2.7
BS8	5	4500	55	2	2	2	30	3	2

Empirical Study

This model was deloped using MatlabR2013b (The Mathworks inc., Rice University) and R statistics version 2.12.1 (The R foundation for Statistical computing). A Time step of 0.125 was used to simulate 2013 to 2015 data. The probability disribution function used to generate random test data for parameters R_i is poisson, where as for AADT(T), Curvature(C), Gradient(G), Hours before Midnight(HBM) normal distribution was used similaly for the Shoulder(S), Spot speeds(SS), Lane width(W) and number of lanes (L) were generated by a discrete uniform distribution. The following 8 by 8 matrix was threfore obtained for the eight variables corresponding to the the 8 black spots across the 8 former administrative boundaries as per the KNTSA classification of black spots. With $R_i(0)$, $i = 1, ..., 8$, as the initial values for every black spot segment BS_i, $i = 1, ..., 8$.The $R_i$is generate using a poisson distribution λ=3.

This 8 by 8 matrix is hence solved using matlab for its eigenvalues and vectors which were used to obtain the constants fot the general sotution,(1.44) .Our problem indeed becomes an initial value problem.

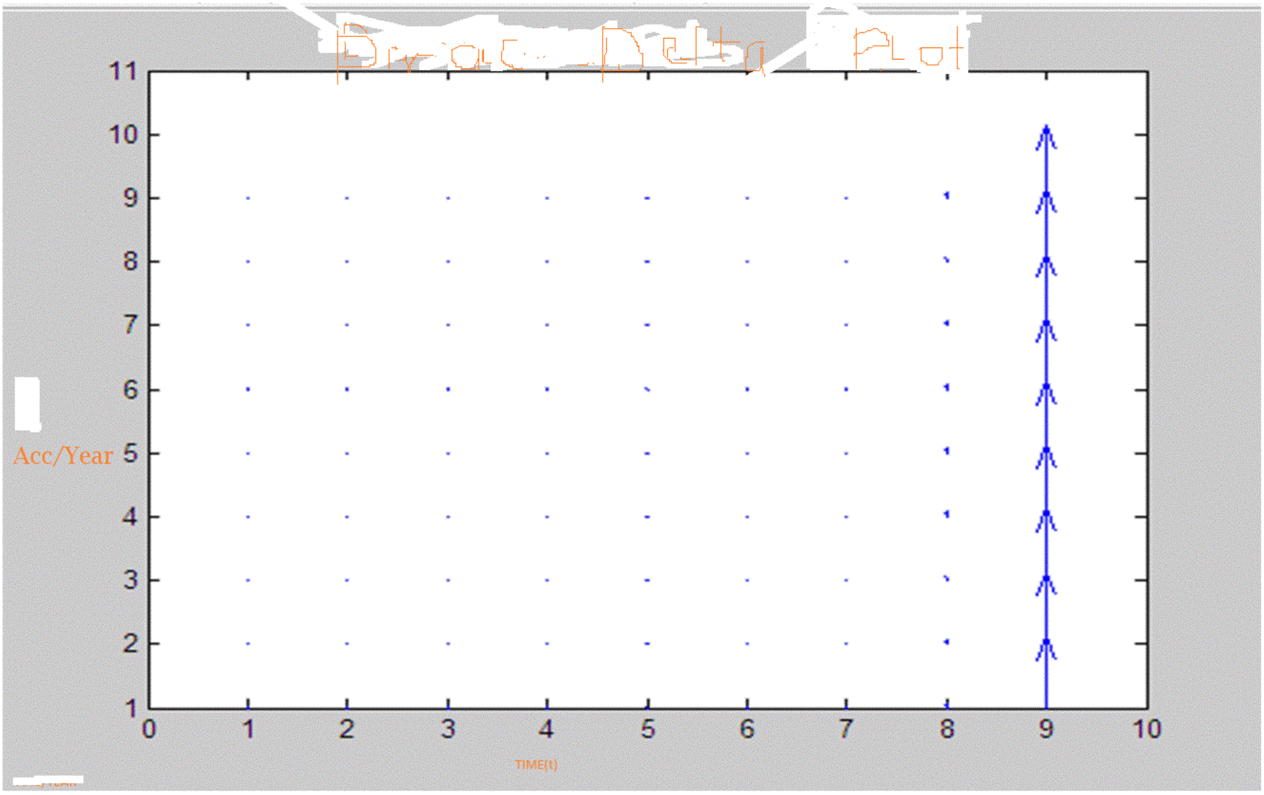
2 RESULT

Using the road parameters and the initial values provided, the constants for the specific equation is ψ_1=43.36, ψ_2=37.6, ψ_3=21.60, ψ_4=23.48, ψ_5=1.25, ψ_6=1.25, ψ_7=4.7, ψ_8=4.2 , with e =2.7 hence we obtain the specific equation for the initial values provided as

$$\omega(t) = \psi_1 * 0.35 * \epsilon.^{6.8*t} + \psi_2 * 0.26 * \epsilon.^{0.07*t} + \psi_3 * 0.18 * \epsilon.^{-0.024*t} + \psi_4 * 0.162 * \epsilon.^{0.0006*t} + \psi_5 * 0.16 * \epsilon.^{-0.0002*t} + \psi_6 * 0.16 * \epsilon.$$

$$+ \psi_7 * 0.16 * \epsilon.^{0.0012*t} + \psi_8 * 0.06 * \epsilon.^{-0.0001*t} \tag{2.1}$$

Figure 2.1: Simulation result showing a shock at t=9



given the initial condition we obtain the phase diagram figure below with

The Figure 5.0.10 above reveal that right from the onset values are zero or near zero but at t =9 , the accident values can be seen as indefinite. A clear indication of a shock that characterises accident occurences.

A futher simulation was conducted with lower curvature and keeping the AADT at a mean of 3000 and a curvature(C) with a low mean of 12 metres.

and using initial values $R_1(0) = 6$, $R_2(0) = 7$,$R_3(0) =5$,$R_4(0) =3$,$R_5(0) =3$,$R_6(0) =2$, $R_7(0)=2$, $R_3(0) =4$, simulated as poisson distribution mean $\lambda=5$, $n = 8$

The model obtained is

$$\omega(t) = \psi_1 * -0.325 * \epsilon.^{379.5*t} + \psi_2 * -0.0999 * \epsilon.^{-0.324*t} + \psi_3 ** -0.0999\epsilon.^{-0.324*t} + \psi_4 * 0.1019 * \epsilon.^{0.3150*t} + \psi_5 * 0.1019 * \epsilon.^{0.3}$$

$$+\psi_7 * 0.0195 * \epsilon.^{0.0762*t} + \psi_8 * 0.0195 * \epsilon.^{0.0762*t} \tag{2.2}$$

with the following constants ψ_1=-11.89, ψ_2=0.58, ψ_3=0.581, ψ_4= -0.399, ψ_5= -0.399, ψ_6=-2.807, ψ_7=1.89, ψ_8=1.89 with e =2.7

When ploted the following curves are obtained. The constants are obtained using table 3.

The figure 2.2 depicts a graph is zero elsewhere except at about t=1.625.

Similarly the figure 2.3 depicts a graph that is zero elsewhere except at about t = 9.

R_i	T	C	SD	HBM	S	SS	L	LW
6	3000	20	1	4	2	20	2	3
7	2500	15	1	3	2	30	2.5	3.5
5	1500	10	2	2	0	15	3	2
3	2000	15	1	6	2	10	1	1
3	3000	5	1	7	2	25	0	2.5
2	2000	8	1	8	1	10	0	2
2	1000	10	0	9	0	20	2	2
4	500	12	1	9	2	15	2	2

Table 3: Change of the data set to reflect low curves

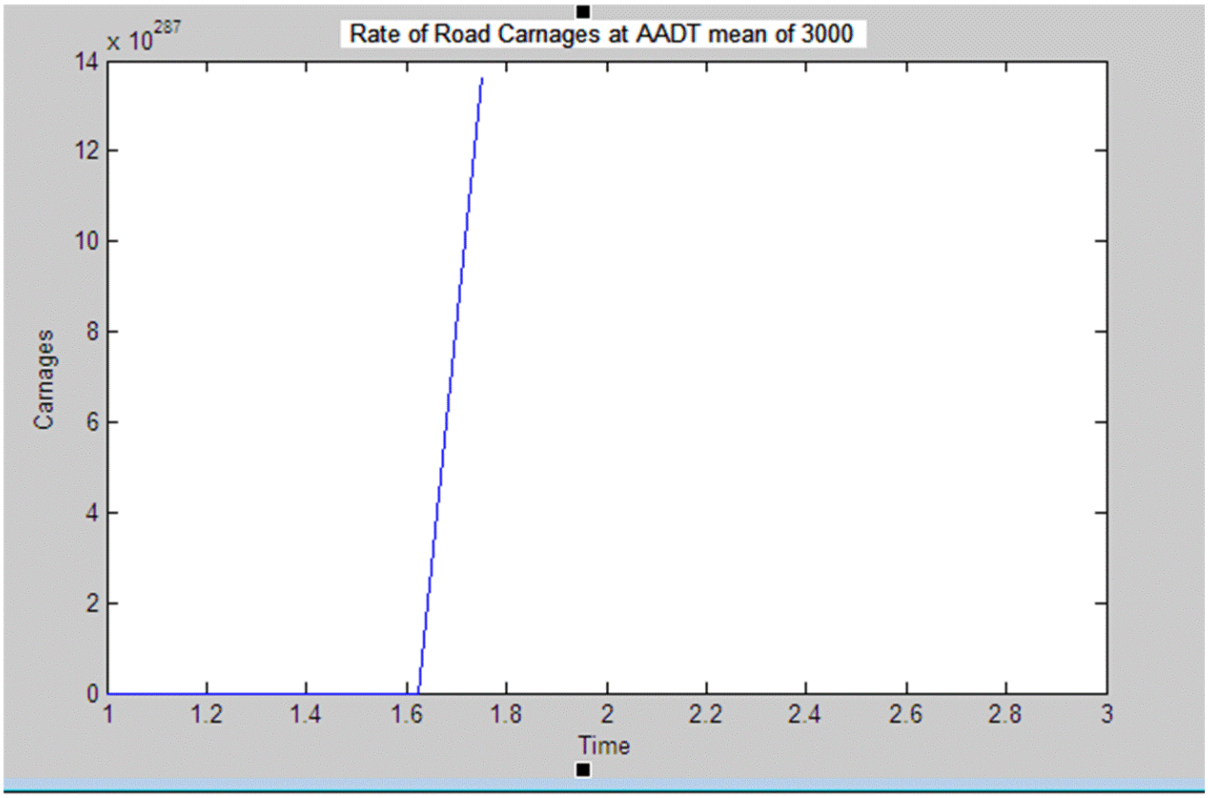


Figure 2.2: A plot with near infinite at about 1.6 on a 3 year time span .

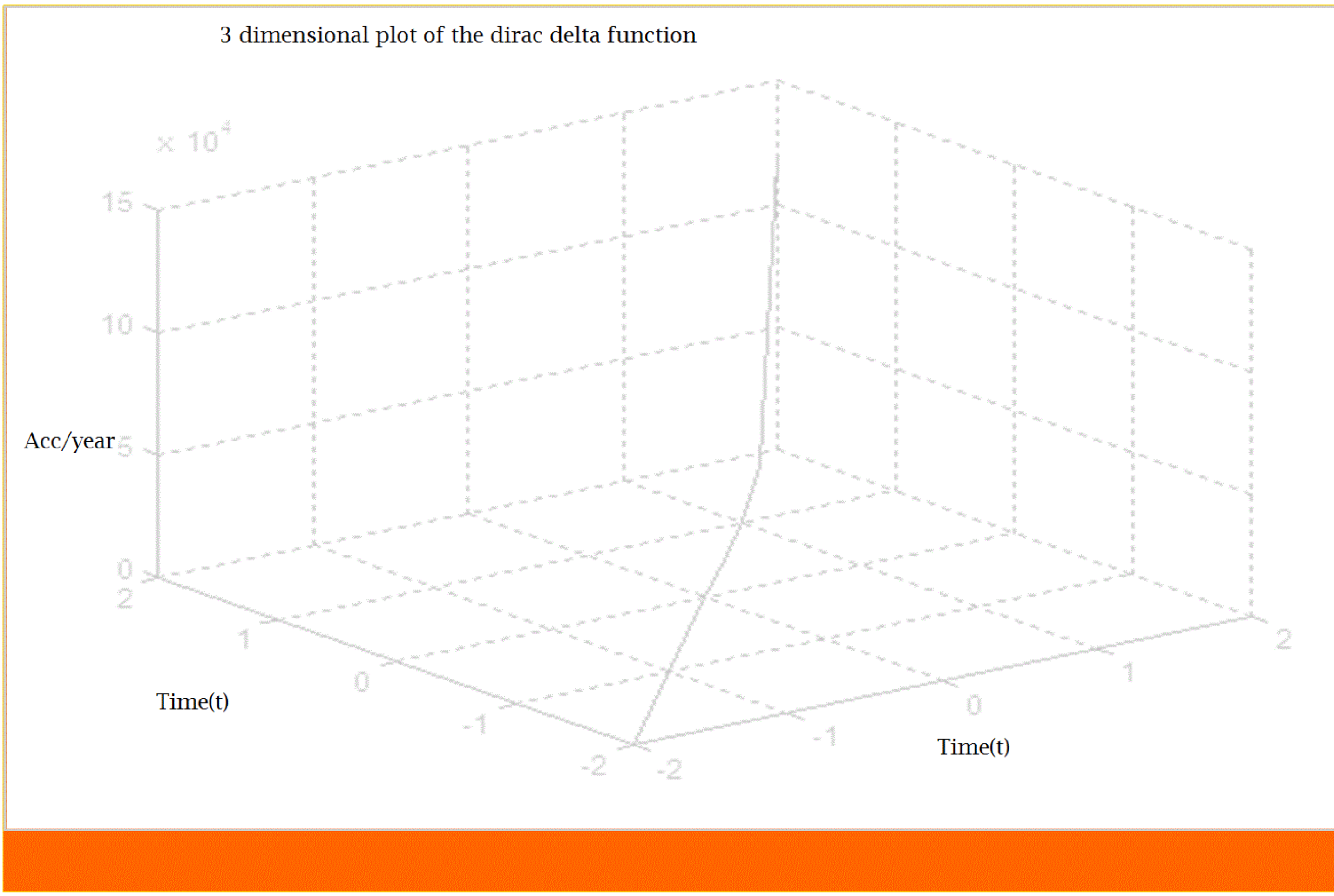


Figure 2.3: The 3 dimensional plot of the δerac-δelta fucntion

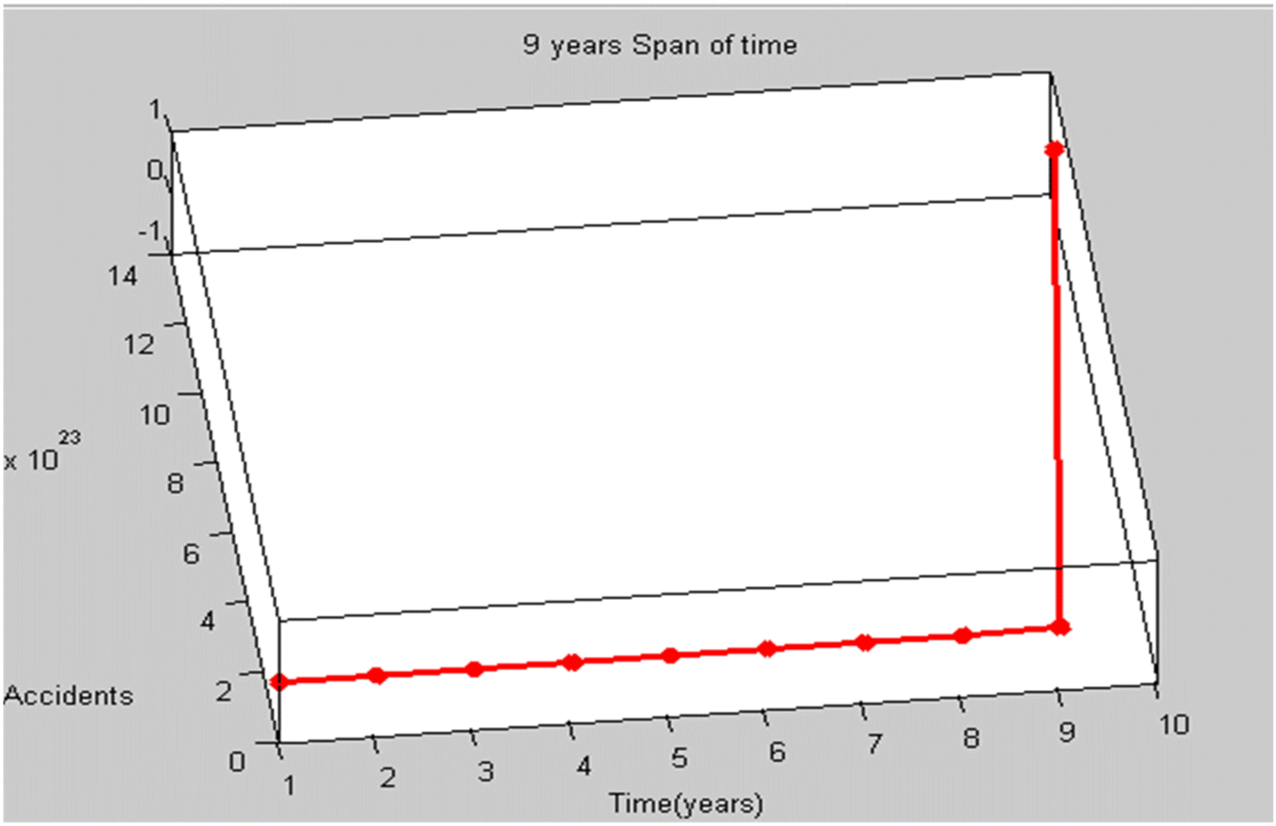


Figure 2.4: The 3 dimensional plot of the δerac-δelta fucntion

The 3 dimensional plot of the δerac-δelta fucntion figure 2.4, equally show a curve that is very tall with narrow spikes and representing a large force exerted on a system over a very short time frame. Such a pattern characterize forcing functions, for this study we adopt an impulse function or commonly reffered to as delta function, usually denoted by $\delta(t)$.Straud K .A.(2003).

The delta function is suitable for representing impulsive force between two bodies in collision , in this case two vehicle crashes or a vehicle crashing on to another object that could be a road guardrail , road bank or vehicle rolling off the road.

The nature of road carnages is revealed by figure 2.4 , which characterizes a linear system that originaly was stable and is excited by a sudden large force, Marcel B.F(2000).

The impulses that characterize carnages can be modelled by providing the initial values(IVP), Marcel B.F(2000).

Mathematically such a time variable function is described by the dirac delta function symbolised by

$$\delta(t-\tau) \tag{2.3}$$

Clearly the simulation results depict a Dirac delta distribution with zeo or near zero elsewhere except one point $\tau = a$ in our two cases, $\tau = 9$ for the Function 5.0.10 and $\tau = 1.65$ for the function 5.0.12. At these points the function can be thought of as either undefined or having an indefinite value. The

Simulation Results

The simulation results exhibit a distribution with zero or near zero everywhere except one point, $t = \tau$ in our two cases, $\tau = 9$ for the function 2.23 and $\tau = 1.65$ for the function 2.24. At this these point the function can be thought as either undefined or having an indefinite value.

revealing impulse response. An impulse in this case is the sudden change in the vehicular flow of activities (traffic) in the system as a result of external forces.

An action of a force acting instantaneously at a time τ and imparting a unit impulse to some mass is evident, Malhan J . A (2002)

If the external forcing function $\Psi(t) = \delta(t-\tau$ Malhan J . A (2002), is adopted , with $\int_{t_0}^{t_1} f(t)dt = I(t)$. This is used to represent a momentum imparted to a system over a short time interval $t_0 \leq \tau \geq t_1$. The function $\delta(t-\tau)$ is called the dirac delta function with the property that $\int_0^\infty \delta(t-\tau)dt = 1$ and that it is zero elsewhere except at $t = \tau$ where it is undefined.

The function can be used to represent the impulse that characterize vehicle crash phenomenon , this time at a point $t = \tau$ after t_0 .

The intergral of the distribution functions about the value $t = \tau$ is unity. ie area under the "curve" equals 1, quite synonimous with probability distributions.

Conclusion

We have proposed a semi-stochastic mixture model to predict the rate of road carnages in kenya. This model has an additive risk function with eight components. The first component models the impact that annual average daily traffic(y_1), has on the rate of road carnages, the second spot speed(y_2) , third, lane width (y_3), fourth is road shoulder width(y_4), fifth factor is number of lanes, sixth factor is Horizontal curvature(y_6),the seventh factor is the vertical gradient(y_7) and finaly the Hours before midnight (y_8). Our analysis of the simulated data set show that in addition to other critical facors in determining rate of road carnages the road curvature variability is important in explaining risk of road carnage. Other studies have explained the greater contribution of other road variables in explaining rates of road carnage, to the best of our knowledge this is one of the study that shows road curvatures as an important variable to consider when modelling rate of road accident carnages. Future reseach should focus in in investigating the explanation for this relationship.

Acknowledgment

I wish to acknowledge my supervoisors, supervisors Dr. George Otieno Orwa and Prof. Romanus Odhiambo for their valuable support. The National Council of Science and Technology (NACOSTI) - Kenya for the funding the research undertaking.

References

1. Assum T.(1998), Appraisal of Road Safety Initiatives in Five Selected Countries in Africa,SSATP Working Paper No. 33

2. Asingo, P (2004), Implimentation of road Transport safety measures in kenya, policy issues and challenges, Working paper no 545(UON)

3. Akaike, H., "Information Theory and an Extension of the Maximum Likelihood Principle" Proceedings of the Second International Symposium on Information Theory, eds. B. N. Petrov and F. Csaki, Budapest: Akademia Kaido, 1973, pp. 267—281.

4. Bohning, D., Dietz, E. and Schalattmann, P. (1997). Zero-inflated count models and their applications in public health and social science. Beitrag zur.

5. Braga A. and Bond B.(2008). Policing crime and Dis-order hot spots; A randomised controlled trial.Criminology, 46(3): 517-608

6. Cameron, A., T. Li, P. Trivedi, and D. Zimmer(2004) "Modelling the Differences in Counted Outcomes Using Bivariate Copula Models with Application to Mismeasured Counts," Econometrics Journal 7, pp. 566—584.

7. Cameron, A., and P. Trivedi(1986) "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators," Journal of Applied Econometrics 46, pp. 347—364. 160

8. Chin H.C., and Quddus, M.A., (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis & Prevention. 35, pp. 253-259.

9. Cheung YB (2002). Zero-inflated models for regression analysis of count data: a study of growth and development. Statistics in Medicine 21: 1461-1469.

10. F. Otsyeno (2011), East African Orthopaedic Journal, vol 5

11. Coddington E.(1961) An Introduction to Odinary Differential Equation by Prenticehall page 256.

12. Greenwood, M. Yule G.U.(1920). An inquiry into the nature of frequency distributions of multiple Happenings. Journal of the Royal Statistical society.

13. Garber, N.J., Wu, L(1989), "Stochastic models relating crash probabilities with geometric and corresponding traffic characteristics data", Research report No UVACTS-5-15- 74 Center for transportation studies at the University of Virginia.

14. Jacobs G.D.(1976). Transport and Road research laboratory. A Study of accident rates on rural roads in Developing countries .

15. Jovanis, P. & Chang H.(1986)., "Modeling the relationship of accidents to miles traveled", Transportation Research Record 1068, pp42-51.

16. Khayesi, M. (1997) Matatu workers in Nairobi, Thika and Ruiru towns, Kenya: Research report, Institute for Development Studies, Nairobi.

17. Karlaftis , M.G. & Golias, I.(2002) “Effect of road geometry and traffic volumes on rural roadway accident rates”, Accident Analysis and Prevention vol.34, Issue 3, pp. 357-365.

18. Landge, V.S., Jain S.S. & Parida M.,(2006) “Modeling traffic accidents on two lane rural highways under mixed traffic conditions”, 87th Annual Meeting of Transportation Research Board.

19. Li, T. ,Trivedi, PK, and Guo(2003). Modeling response bias in count: A structural approach with an application to the national crimne victimization survey data. Sociological Methods and Research 31, 514-544.

20. Manyara G C(2014).Combating road traffic accidents in kenya: A challenge for an emerging economy.

21. Masaeid, H.R. & Nelson, C.D., “Pedestrian accidents and their relationship to street geometrics and operation variables in Jordan”, Indian Highways, Vol. 24, No. 9, Sept.

22. Miaou, S.P(1994). “The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions”, Accident Analysis and Prevention vol. 26, issue 4, 1994, pp. 471.

23. Mukhopadhyay, K.(1997). Regression model for Underreported counts.Indian University indiana

24. Okamoto(1989) “A Method to cope up with random errors of observed accident rates” Safety literature, Vol.4, page 317-332, 1989.

25. Odero, W., Khayesi,M. and Heda, P. M. (2003), Road traffic injuries in Kenya: Magnitude, causes and status: Injury Control and Safety Promotion, Vol.10 (1-2)

26. Pararai, M., Famoye, F.and Lee, C.(2006). Generalized Poisson Regression model for under reported counts. Advances and applications in Statistics 26,485-494

27. Sharma1 A.K., Landge V. S. (2013), Zero inflated negative binomial for modeling heavy vehicle crash rate on indian rural highway: International Journal Advances in Engineering and Technology (ISSN 2231-1963) January 2013.

28. Sharma et al.(2012), Pedestrian Accident Prediction Model for Rural Road: International Journal of Science and Advanced Technology (ISSN 2221-8386) Volume 2 No 8 August 2012.

29. Summersgill, I. Maher (1996), A Comprehensive Methodology for the Fitting Predictive Accident Models. Accident Analysis & Prevention, Vol. 28, No. 3, pp.281-296.

30. Sia at al.(2002), Traffic impact Model, Department of Civil Engineering, University of Toronto, Toronto, Ontario Cananda M5S 1A4.

31. Silyanov V. V.(1968), Carriageway marking tests in the USSR. Traffic Engineering and Control, 409 – 12.

32. Stroud K. A. (2003),Futher Engineering Mathematics, Third edition by palgrave publishers.

33. Greene WH (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. working paper, Department of Economics, Stern School of business, New York University, New York.

34. Haddon Jr W.(1980), Advances in the epidemiology of injuries as a basis for public policy. Public Health Report, 95:411–421.

35. Ipsos Synovate(2012), Press Release, 19th October 2012

36. Winkelmann, R(2003), Seemingly Unrelated Negative binomial Regression, Oxford bulletin of Economics and Statistics.

37. Zegeer, C V., Deen, R. C. & Mayes, J. G.(1981) "Effect of lane and shoulder widths on accident reduction on rural two-lane roads". TRR 806 1981.pp 33-43

38. Mehdi Hossein Pour et al. / Procedia -Social and Behavioral Sciences 53 (2012) 1218 – 1227

39. Valera I and Gomez R.(2015), Modeling Adoption and Usage of Competing Products. MPI for Software Systems 1406.0516v4 [cs.SI]

40. Williams Ackaah, Mohammed Salifu (2011), "Crash prediction model for two lane rural highways in Ashanti region of Ghana", International Association of Traffic and Safety Sciences(IATSS) Research, http://dx.doi.org/10.1016/j.iatssr.2011.02.001

41. www.ntsa.go.ke