PREDICTION UNDER BAYESIAN APPROACH OF CAR ACCIDENTS IN URBAN INTERSECTIONS

Oscar A. Rosas-Jaimes
Researcher, Universidad Autónoma del Estado de México,
Ciudad Universitaria, Cerro de Coatepec s/n, C.P. 50100, Toluca, México
e-mail: oscar.rosasjaimes@yahoo.com

Araceli C. Campero-Carmona
Graduate Student, Universidad Autónoma del Estado de México,
Ciudad Universitaria, Cerro de Coatepec s/n, C.P. 50100, Toluca, México,
e-mail: arkeli9502@yahoo.com.mx

Oscar L. Sánchez-Flores
Researcher, Universidad Autónoma del Estado de México,
Ciudad Universitaria, Cerro de Coatepec s/n, C.P. 50100, Toluca, México
e-mail: osanchez@uaemex.mx

Submitted to the 3rd International Conference on Road Safety and Simulation, September 14-16, 2011, Indianapolis, USA

ABSTRACT

The increase in the number of automobiles in circulation has brought by direct consequence an increase of the interactions among different factors that contribute to the exposure of crashes, particularly in the urban intersections. For example, according to data of the city council of Toluca, in the State of Mexico, 80% of these accidents happens in intersections of urban routes (Hinojosa, 2003). This motivates a fundamental interest in the study of this phenomenon, from where possible policies or instruments could be obtained to reduce this class of incidents.

In this communication three final distributions of Bayes Rule are shown (Gregory, 2005) that predict the probability of occurrence of accidents in urban intersections of the Metropolitan Zone of the City of Toluca. The first one is a specification of a Translated Poisson Mixture Model, which presents a weighed average of occurrence rates of observed car accidents. The second one is the One-variable Poisson-Gamma Model, which introduces an effect of random type to the error term of the flow variables, which is an independent variable. Finally, the third one is a Two-variable Possion-Gamma Model, which in addition to the randomness mentioned previously, considers the relation between the frequency of accidents and the vehicular flow.

From statistical data of a year of reference, the parameters of these three models are calibrated, with which the final distribution is considered that predicts the occurrence of car accidents in intersections for various temporary horizons. In order to determine which of these three proposed models is more precise, observed data against estimated values by these models are compared, using the correlation coefficient and the GEH as measuring tools for carrying out this comparison.

The results of these analyses show that the Two-variable Poisson-Gamma Model is the one that better adjusts to the observed data, reason why it is used to compute an estimation of the number of accidents expected in a set of intersections, and thus to hierarchize their danger, applying then to a case of study.

Keywords: transportation, safety, Bayes prediction, car accidents.

INTRODUCTION

Car accidents occurring in urban intersections are classified as a danger indicator for safety in cities. Their study has turned into a quest for their understanding. Their nature is mostly probabilistic and the models for their forecasting in transportation networks has raised a big amount of research in this topic (Lord, 2002; Lord 2004; Mitra et al., 2002, Kononov and Janson, 2002; Hauer et al., 2002; Miranda, 2005; Miaou and Jin, 2005).

During the last three decades (Abbes and Jarret, 1981) researchers have used different approaches for modeling data and calculate expected rates of accidents, which typically are represented as a Poisson process, i.e. when the number of trials for an event is large, but its probability of occurrence is small (Ross, 1989).

Poisson-based models and analysis have been made in to obtain accuracy in the study of urban car accidents, particularly in the intersections of streets. Poisson log-normal schemes (Miranda et al., 2005) have been used to locate dangerous sites, for example, and Poisson-gamma approaches, also called negative binomial, are preferred when the random effects must be considered (Miaou and Lord, 2003; Miranda et al., 2005). Due to the sources and methodologies of data acquisition, a number of technical problems must be faced, as when over-dispersion of accident data is present, making necessary the use of zero-inflated Poisson-based models (Qin et al,2002; Lord et al., 2005; Miranda et al., 2005). Bayesian approaches are related to those models, but taking into account the dependence or independence of the events under study, which can be related to the history of those events (Rios et al., 1999; Miaou and Jin, 2005).

Besides the logic in the choice of any model, it is important to have in mind their prediction capability. A comparison among these mathematical tools shows that none of them has the best result, and the accuracy depends strongly in the data themselves, their obtaining, and the conditions in which the events happen. The reader will see that we present a set of locations of intersections that can be considered dangerous in relation with a degree of probability of occurrences of car accidents. These intersections have been located through a Geographic Information System (GIS), conformed by the combination of the roadway network, the number of accidents and ortophotos as the main layers. We describe the nature, collection and process of the data utilized herein, in the next section of this paper.

A brief description of some of the models that best behave to describe these crash events are discussed in the section titled General Structure of Models. We propose some models to calculate a forecast of urban crashes in intersections, considering random effects and a simple relation among number of crashes and traffic flows. All the chosen intersections to perform the calculations in this work are traffic-lighted and the events are collision-type, with similar

conditions in pavement. We compared the performance of three schemes: Mixed Poisson distributions, One-variable Poisson-Gamma, and Two-variable Poisson-Gamma.

Results of series of real data obtained in the city of Toluca urban area are showed in a respective section of this text. We present comparisons between real data an estimations, giving a correlation about its similarity and goodness of our calculations. This is uselful to perform a calibration on our models in order to express in the best possible way the reality of the phenomena involved.

As it will be seen, the two-variable Poisson-gamma model best approaches to our data, composed not only by number of accidents in intersections, but by flows of vehicles. Then we proceed to use it to analyse intersections in an important roadway in the City of Toluca. At the end, we include some conclusions to the results obtained.

DATA SETS

Data Acquisition

The involved data for this work is referred to vehicular crashes that have happened in the road network of the city of Toluca, Mexico. This information has been provided by the municipality and spans the period from 2000 to 2005. It is significant to consider that the quality of the information of the group of data of road accidents is little reliable, due to a series of systematic errors and inconsistencies as registry omissions, arrangement between involved parts without authority participation or other unregistered crash events. Unfortunately, this is a common situation for these records. The rest of this data collection has been categorized as poorly precise. Unfortunately, this is a common situation, because typically only 40% of accidents are recorded in most cases globally (Lee, 2002). However, from a thorough review of the available statistics, data of the year 2000 present the best accuracy of all.

The data utilized have been analyzed and stored in a Geographic Information System (GIS), reporting a count of 11 530 accidents that have happened in highway stretches, intersections or pedestrian crossings, of these 9 204 have been registered and located in 1 686 intersections that present one road accident at least, taking into account a defined radius of 15 m from the center formed by the main way and the secondary one, according to a established criterion (Miaou and Lord 2003), emphasizing that more than 82% are road accidents classified as collisions or crashes, another 18% includes another type of accidents like running-overs, run-off-road or crashes with trains. From the 9204 accidents occurred in intersections, 3% resulted in casualties and 97% without fatality. Intersections individually have registered from 0 to 43 accidents per year.

Vehicular flows for year 2000, vary from 11 893 to 37 607 vehicles per day, showing an entrance of vehicles to an intersection from 333 000 to 1 053 000 vehicles per month, figures that emphasize the importance of the nonlinear relation that exists between the data of accidents and the vehicular flows, which seems to show an inverse relation, i.e. the smaller the flow, the greater is the amount of accidents which have been registered for a determined intersection, this

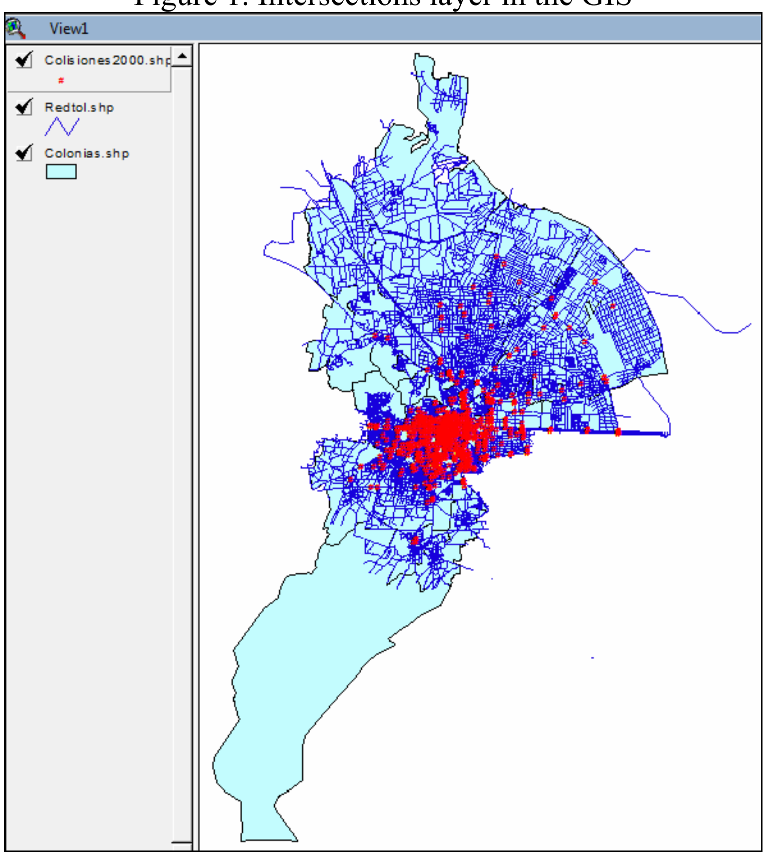
phenomenon is according with the relation among the levels on service, speed, capacity and drivers decisions (Lee et al., 2002).

From the cited set of data, 20 intersections have been selected, under three main criteria: they are those that present the biggest number of accidents; they are those which correspond with records of sufficiently complete data; and they are those which corresponds with a high impact in the distribution of passengers, drivers and merchandise in the road network of Toluca zone.

Data Process

The procedure to register and to report the road accidents in the city of Toluca, is coordinated by the municipal transit office, which designates a specialized group for monitoring zones in 24 hour duties. The accident report is based on the Regulation of Transit and in two forms: Presentation and Agreement, both are official documents before the attorney general. The assigned transit agents go to the event's place and verify if there are injured individuals, deceases or disagreement among the affected ones and proceed to the filling of the forms, making the involved parts available to the attorney general; when there are only minor material damages and it exists a conciliation among the involved ones or its insurers, the agreement is written down in the respective report. From these forms data are brought together.

Figure 1. Intersections layer in the GIS



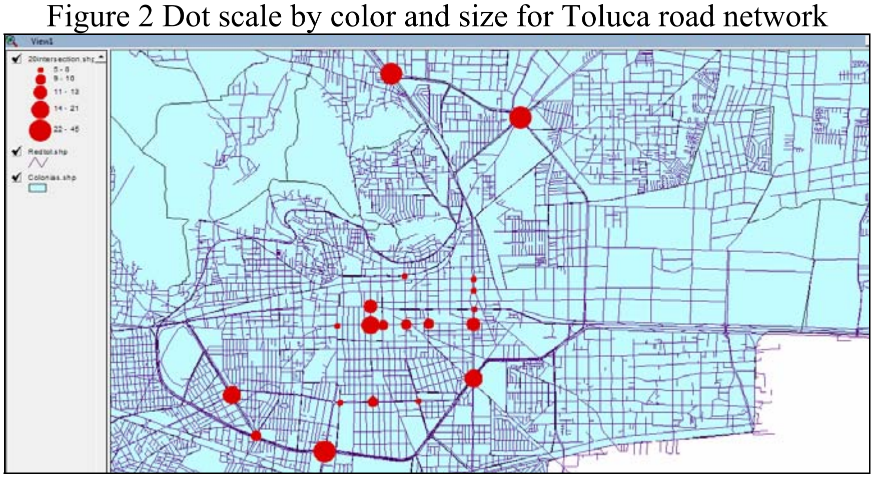
With the obtained information, data bases are filled to build a GIS, based in a set of orthophotos of the zone of Toluca which are the background of two layers, the first referring to the road network and the second to accidents. This last layer is a product of the Information System for Transport Planning and Administration (Sistema de Información para la Planeación y Administración del Transporte - SIPAT), an agreement between the Faculty of Engineering of the State of Mexico University (UAEMex) and the Municipality of Toluca, which is conformed by 36 different categories (neighborhoods, blocks, basic geostatistical areas, routes, map courses, infrastructure, parkings, among other) which constitutes a snapshot of the city transport system.

For GIS integration, the fields ID, date, hour, notes, location, sector or zone name, accident type, result, vehicle participation, injured persons, number of casualties, sex and hospital are standardized, then out of place points are correctly located, data bases are joined by similar attributes (using software like Arcview® or Excel®), missing data and records are captured and finally the space coincidence of each layer with respect to the mosaic of orthophotos are verified. Figure 1 shows the aspect of the final GIS.

The study area has been selected for being a zone of a greater urbanization in relation to other municipalities This city is conformed by 117 colonies and 7567 blocks, counting on almost 17.000 intersections in all the road network, with an extension of 420,14 km ², a population considered in 2008 of 1 million 714 thousand 831 inhabitants. Toluca stands out in economy like a great commercial center (Municipal Side, 2005).

The selection of 20 intersections of greater accident rates is made using as a main tool the GIS with the following criteria of selection: first, the data must be of the same administration and jurisdiction where notable changes through time do not exist as for vehicular volume or as for infrastructure during consecutive periods, (Persaud, 2002); second, to make the regression of the average and the forecasting of accidents at micro level; third, use of buffers in each crossroad to make the count of accidents; and fourth, making of dot scale with degradation of color and size as Figure 2 shows, to distinguish crash frequency.

Figure 2 Dot scale by color and size for Toluca road network



On its own, Table 1 shows more detailed data of accidents of the 20 intersections of the greatest rate of accidents for year 2000. Streets names of each of these intersections are shown, and the counting of number of accidents per month is listed.

Table 1 Data of monthly vehicular accidents, year 2000

No.	Streets Name	Jan	Feb	Mar	April	May	June	July	Aug	Sep	Oct	Nov	Dec
1	Benito Juárez y Miguel Hidalgo	0	2	1	0	2	0	0	1	2	2	2	1
2	Isidro Fabela e Independencia	0	0	0	3	1	1	0	1	0	0	0	1
3	Isidro Fabela y Alfredo del Mazo	1	2	3	4	2	5	8	3	3	6	2	2
4	Isidro Fabela y Lerdo de Tejada	0	0	0	0	0	0	1	0	3	1	1	2
5	Isidro Fabela y Miguel Hidalgo	1	1	0	0	0	0	0	3	0	1	0	2
6	José López Portillo y Alfredo del Mazo	1	3	2	2	1	4	3	8	1	4	10	6
7	José María Morelos e Isidro Fabela	0	0	0	1	3	0	2	0	1	2	1	3
8	José María Morelos y Vicente Villada	0	0	0	0	2	0	0	1	0	0	0	2
9	José María Morelos y López Rayón	0	1	2	1	1	2	0	2	0	0	0	0
10	José María Morelos y Benito Juárez	1	0	2	2	2	0	0	0	5	2	2	1
11	José María Morelos y Pino Suárez	1	0	0	0	2	1	1	1	1	1	1	1
12	Santos Degollados y Pino Suárez	1	1	3	0	0	0	0	1	0	0	1	1
13	Tollocan e Isidro Fabela	3	1	1	0	1	1	3	4	1	2	2	2
14	Tollocan y Cristobal Colón	0	5	4	2	5	2	0	2	2	3	6	3
15	Tollocan y Pino Suárez	2	1	3	0	3	1	2	4	2	1	1	1
16	Tollocan y Vicente Guerrero	0	0	1	0	1	1	2	3	1	2	2	1
17	Venustiano Carranza y Benito Juárez	0	0	1	2	0	1	4	0	1	0	1	0
18	Venustiano Carranza y Jesús Carranza	0	1	0	0	0	0	2	0	0	2	1	1
19	Venustiano Carranza y Pino Suárez	0	1	0	1	1	2	2	1	0	0	0	0
20	Venustiano Carranza y Vicente Guerrero	1	0	0	2	1	3	2	2	2	3	0	1

Table 2 Data of monthly vehicular flows (1000 vehicles/intersection/month), year 2000

No.	Name	Jan	Feb	Mar	April	May	June	July	Aug	Sep	Oct	Nov	Dec
1	Benito Juárez y Miguel Hidalgo	693	708	693	685	685	693	685	662	685	662	655	700
2	Isidro Fabela e Independencia	462	462	467	457	457	467	472	467	472	478	452	472
3	Isidro Fabela y Alfredo del Mazo	525	550	556	538	538	538	550	556	550	531	538	531
4	Isidro Fabela y Lerdo de Tejada	484	458	463	458	453	469	484	484	453	458	490	479
5	Isidro Fabela y Miguel Hidalgo	484	458	463	458	453	469	484	484	453	458	490	479
6	José López Portillo y Alfredo del Mazo	1,030	1,019	1,007	1,041	1,053	996	1,019	996	1,041	985	1,007	1,030
7	José María Morelos e Isidro Fabela	875	875	866	905	836	905	856	905	895	836	885	875
8	José María Morelos y Vicente Villada	662	670	647	685	655	693	670	655	700	670	685	647
9	José María Morelos y López Rayón	555	588	581	581	588	581	555	568	588	568	555	568
10	José María Morelos y Benito Juárez	709	749	693	733	749	693	701	725	717	693	717	717
11	José María Morelos y Pino Suárez	562	581	569	581	556	581	562	569	550	556	556	569
12	Santos Degollados y Pino Suárez	645	668	675	698	660	645	698	690	668	668	683	683
13	Tollocan e Isidro Fabela	525	537	543	556	525	556	549	537	525	525	562	562
14	Tollocan y Cristobal Colón	720	754	745	762	770	754	762	720	728	737	745	779
15	Tollocan y Pino Suárez	601	581	614	581	574	581	594	594	588	614	594	581
16	Tollocan y Vicente Guerrero	337	345	333	337	337	333	353	357	349	357	333	353
17	Venustiano Carranza y Benito Juárez	770	779	745	737	779	754	737	728	745	779	779	779
18	Venustiano Carranza y Jesús Carranza	622	643	615	622	643	650	601	622	650	650	615	629
19	Venustiano Carranza y Pino Suárez	623	609	602	589	623	582	596	582	582	602	576	576
20	Venustiano Carranza y Vicente Guerrero	650	679	642	650	650	672	650	679	635	672	657	665

On the other hand, Annual Average Daily Traffic (AADT) data have being conformed from vehicles registered hourly through traffic flow sensors installed, including private and public vehicles entering to each crossroad, to obtain monthly flows for year 2000 shown in Table 2.

All these data collections are the inputs for our probability models.

GENERAL STRUCTURE OF MODELS

Model 1: Mixed Poisson distributions

This model for vehicular accident forecast calculates data of crashes where the number of accidents assumes that X_{it} occurs in a period of time t in the site i independently to another site, considering an initial uniform distribution and a Poisson likelihood function

$$P(X_{it}|\mu) \sim Poisson\ (\mu) \tag{1}$$

Vehicular flow is assumed constant for all the sites, the intersections are of equal geometry and all accidents are of the same type, i.e. the observed data set $(x_1, x_2, \ldots, x_n)$ has accident rates $\mu_i = (\mu_1, \mu_2, \ldots, \mu_n)$, under $E(x_i|\mu) = \mathrm{var}(x_i|\mu) = \mu = \overline{X}$.

Generalized *a posteriori* expression can be written in the form

$$P(\mu_i|X_{it}) = \frac{\mu_1^x e^{-\mu_1}}{x!} + \frac{\mu_2^x e^{-\mu_2}}{x!} + \cdots + \frac{\mu_n^x e^{-\mu_n}}{x!} = \sum\left(\frac{\mu^x e^{-\mu}}{x!}\right) \tag{2}$$

As can be seen, expression (2) represents a weighted average of Poisson distributions for different μ_n for each x_n observations. In addition, (2) can be re-written in a normalized fashion

$$P(X_{it}) = P(X_{it}|x_1) = \frac{\mu_1^{x_1} e^{-\mu_1}}{x_1!} P(\mu_1|x_1) + \frac{\mu_2^{x_1} e^{-\mu_2}}{x_1!} P(\mu_2|x_1) + \cdots + \frac{\mu_n^{x_1} e^{-\mu_n}}{x_1!} P(\mu_n|x_1) \tag{3}$$

We can apply (3) to any of the intersections of our GIS. As an ilustrative example, intersection number 6, formed José López Portillo y Alfredo del Mazo streets, presents 7 different monthly rates for year 2000. Initially we can consider equal probability for each one. So, the probability of having 1 accident in this intersection is

$$\begin{aligned} P(1|x=1) = {} & \frac{1^1 e^{-1}}{1!}(0.143) + \frac{2^1 e^{-2}}{1!}(0.143) + \\ & + \frac{3^1 e^{-3}}{1!}(0.143) + \frac{4^1 e^{-4}}{1!}(0.143) + \frac{6^1 e^{-6}}{1!}(0.143) + \\ & + \frac{8^1 e^{-8}}{1!}(0.143) + \frac{10^1 e^{-10}}{1!}(0.143) \end{aligned} \tag{4}$$

Which value is $P = 0.1253$.

The regression of the average has been made with monthly parameters $\mu_i = (\mu_1, \mu_2, \ldots, \mu_n)$ and real data using a linear regression $\mu_i = 0.514 * X_{it} + 0.409$, thus, the considered annual average is $\hat{\mu} = 2.3366$, the estimation of the probability takes place using a Poission distribution, it is normalized and the percentage of probability for each real rate of accidents registered in the intersection is obtained. Table 3 shows the results for each occurrence of accidents in intersection 6.

Table 3 Mixtures of Poisson Probabilities, intersection 6

Month	Real accidents	Estimated accident rates	Mixtures of Poisson	Normalized	% Probabilty
January	1	0.9231	0.1253	0.10	10.3
February	3	1.9511	0.1124	0.09	9.2
March	2	1.4371	0.1261	0.10	10.4
April	2	1.4371	0.1261	0.10	10.4
May	1	0.9231	0.1253	0.10	10.3
June	4	2.4651	0.0969	0.08	8.0
July	3	1.9511	0.1124	0.09	9.2
August	8	4.5211	0.0563	0.05	4.6
September	1	0.9231	0.1253	0.10	10.3
October	4	2.4651	0.0969	0.08	8.0
November	10	5.5491	0.0388	0.03	3.2
December	6	3.4931	0.0733	0.06	6.0
	$\overline{X} = 3.75$	$\hat{\mu} = 2.3366$	- - - - -	$P = 1.0$	$P[\%] = 100.00$

Model 2: One-variable Poisson-gamma model

This Poisson-like model allows introducing a random effect, making use of the gamma function, relaxing the limitation of having the variance equalized to the mean in a fixed way. This model assumes that the number of accidents X_{it} happens in a period of time t in the site i independently to another accident, with initial gamma distribution and Poisson likelihood

$$P\left(X_{i,t} \middle| \mu\right) \sim Poisson\ (\mu) \tag{5}$$

The intersections are considered of equal geometry and all accidents are of the same type, with

$$\mu_i = e^{\beta_2 x_{it}} \cdot e^{\varepsilon_i} \tag{6}$$

where μ represents the rate of accidents per time unit and follows a Poisson distribution, independent of the rest of intersections and periods of time $e^{\beta_2 x_{ti}}$, represents the historical data of accidents, and $e^{\varepsilon_i} \sim Gamma\ (\alpha_0, \beta_0)$ represents the random effect. The gamma function can be approximated and standardized by

$$Gamma\ (\alpha_1, \beta_1) = \left(n\overline{X} + \alpha_0, n + \beta_0\right) \tag{7}$$

For which annual parameters are:

$$\hat{\mu} = \frac{\alpha_1}{\beta_1} = \frac{n\overline{X} + \alpha_0}{n + \beta_0} \approx \overline{X} \tag{8}$$

$$\hat{\sigma}^2 = \frac{\alpha_1}{\beta_1^2} = \frac{n\overline{X} + \alpha_0}{(n + \beta_0)^2} \tag{9}$$

Consider now α_1 and β_1 as fixed values and a predictive negative binomial function

$$P(Y|\mu,\phi) = C_{\alpha-1}^{Y+\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{Y} \tag{10}$$

Observe, however, that $\mu_i = e^{\beta_2 x_{it}} \cdot e^{\varepsilon_i}$ has been adopted in this model as a first assumption. Annual parameter $\hat{\mu}$ tends to $\overline{X}$, therefore in order to take advantage of the Poisson-like expression (5) that includes the random effect approximated by (6), it is better to introduce $\hat{\mu}$ for the probability estimation in this model.

Then, if $\alpha_1 = \phi = n\overline{X} + \alpha_0$ and $\beta_1 = \dfrac{\phi}{\overline{X}}$, it is possible to obtain

$$P(y_i|\hat{\mu}_i,\phi) = \frac{\Gamma(y_i+\phi)}{y_i!\Gamma(\phi)}\left(\frac{\phi}{(\phi+\hat{\mu})}\right)^{\phi}\left(\frac{\hat{\mu}}{\phi+\hat{\mu}}\right)^{y_i} \tag{11}$$

where ϕ is known as the dispersion parameter. For the same intersection, for example, we have

$$\hat{\mu} = \frac{\alpha_0}{\beta_0} = 3.75 \tag{12}$$

$$\hat{\sigma}^2 = \frac{\alpha_0}{\beta_0^2} = 8.3864 \tag{13}$$

From where it is possible to resolve for α_0 and β_0 and get an initial distribution for the gamma function (7), obtaining to the same time for (8) and (9)

$$\hat{\mu} = \frac{\alpha_1}{\beta_1} = \frac{46.68}{12.45} = 3.75 \tag{14}$$

$$\hat{\sigma}^2 = \frac{\alpha_1}{(\beta_1)^2}(\beta_1+1) = \frac{46.68}{(12.45)^2}(13.447) = 4.05 \tag{15}$$

Table 4 One-variable Negative Binomial Probabilities, intersection 6

Month	Real accidents	Estimated accidents	Negative Binomial	Normalized	% Probability
January	1	1.287	0.1123	0.07	7.1
February	3	2.134	0.2100	0.13	13.2
March	2	1.657	0.1861	0.12	11.7
April	2	1.657	0.1861	0.12	11.7
May	1	1.287	0.1123	0.07	7.1
June	4	2.75	0.1814	0.11	11.4
July	3	2.134	0.2100	0.13	13.2
August	8	7.621	0.0187	0.01	1.2
September	1	1.287	0.1123	0.07	7.1
October	4	2.75	0.1814	0.11	11.4
November	10	12.72	0.0031	0.00	0.2
December	6	4.574	0.0765	0.05	4.8
	$\overline{X} = 3.7500$	$\hat{\mu} = 3.4882$	- - - - -	$P = 1.0$	P[%] =100.00

If we take $\alpha_1 = \phi = 46.68 \approx 47$, finally we have, for 1 accident in intersection 6

$$P(1|\hat{\mu}_i, \phi) = \frac{\Gamma(1+47)}{1!\Gamma(47)} \left(\frac{47}{(47+3.4882)} \right)^{47} \left(\frac{3.4882}{47+3.4882} \right)^{1} \quad (16)$$

That is to say, $P = 0.1123$.

The regression of the average $\mu_i = e^{\beta_2 x_{it}} \cdot e^{\varepsilon_i}$, has been calculated with the parameter $\beta_2 = 0.2519$ (parameter considered with an estimation using WinBUGS14®), the considered annual average is $\hat{\mu} = 3.4882$ and $\phi = 47$ for normalized expression (11). Table 4 shows the results for intersection 6 using this model.

Model 3: Two-variable Poisson-gamma model

This Poisson-Gamma model allows to introduce the existing relation between the number of accidents and the vehicular flows, considering the average of accidents as an exponential function, in such a way that integration to conform the final distribution is of high complexity, being most suitable for numerical calculation. This model asumes that the number of accidents X_{it} occurs in a period of time *t* at the intersection *i* independently of another intersection like a Poisson process, with initial gamma distribution and a Poission distribution as likelihood function. In addition, we consider that involved intersections are of equal geometric characteristics and the accidents are all of the same type.

$$P(X_{i,t}|\mu) \sim Poisson\ (\mu) \quad (17)$$

For this case, the accident rate of a Poisson distribution is structured as

$$\mu_i = e^{\left(\beta_3 * X_{it} + \ln(F^{\beta_4}) + \varepsilon_i\right)} \quad (18)$$

Where X_{it} represents the number of crashes occurring monthly in the i^{th} intersection at time t. *F* represents the number of motor vehicles that enter an intersection conformed by a main route crossed by a secondary route, including all type of maneuver (straight, left turn, and right turn) measured in vehicles per month and μ_i represents the rate of accidents by unit of time, β_2 and β_3 are parameters to consider and as before, $e^{\beta_3 x_{ti}}$ represents the historical data of accidents and the expression $e^{\varepsilon_i} \sim Gamma\ (\alpha_0, \beta_0)$ represents the random effect, in addition the expression $e^{\ln(F^{\beta_4})}$ represents vehicular flows.

As it has already been seen, the gamma function can be approximated and standardized by

$$Gamma\ (\alpha_1, \beta_1) = \left(n\overline{X} + \alpha_0, n + \beta_0\right) \quad (19)$$

for which:

$$\hat{\mu} = \frac{\alpha_1}{\beta_1} = \frac{n\overline{X} + \alpha_0}{n + \beta_0} \quad (20)$$

$$\hat{\sigma}^2 = \frac{\alpha_1}{\beta_1^2} = \frac{n\overline{X} + \alpha_0}{(n + \beta_0)^2} \quad (21)$$

Considering then α_1 and β_1 as fixed values and a predictive function

$$P(Y|\mu,\phi) = C_{\alpha-1}^{Y+\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{Y} \tag{22}$$

As in the previous case, $\mu_i = \exp^{\beta x_{it}} \cdot \exp^{\varepsilon_i}$ has been adopted in this model as a first assumption.

Therefore, making the same assumptions it is possible to obtain

$$P(y_i|\hat{\mu}_i,\phi) = \frac{\Gamma(y_i+\phi)}{y_i!\Gamma(\phi)}\left(\frac{\phi}{(\phi+\hat{\mu})}\right)^{\phi}\left(\frac{\hat{\mu}}{\phi+\hat{\mu}}\right)^{y_i} \tag{23}$$

being ϕ the so called dispersion parameter.

Since this case has the accident rate from an expression as (18), then $\hat{\mu} = 3.5195$ and

$$P(1|\hat{\mu}_i,\phi) = \frac{\Gamma(1+47)}{1!\Gamma(47)}\left(\frac{47}{(47+3.5195)}\right)^{47}\left(\frac{3.5195}{47+3.5195}\right)^{1} \tag{24}$$

That is to say, the probability of having 1 accident in the intersection 6 is $P = 0.1099$

Table 5 Two-variable Poisson Gamma Probabilities, intersection 6

Month	Real accidents	Estimated accidents	Negative Binomial	Normalized	% Probability
January	1	1.344	0.1099	0.07	7.0
February	3	2.199	0.2092	0.13	13.2
March	2	1.718	0.1838	0.12	11.6
April	2	1.719	0.1838	0.12	11.6
May	1	1.344	0.1099	0.07	7.0
June	4	2.815	0.1821	0.12	11.5
July	3	2.199	0.2092	0.13	13.2
August	8	7.597	0.0194	0.01	1.2
September	1	1.344	0.1099	0.07	7.0
October	4	2.814	0.1821	0.12	11.5
November	10	12.52	0.0032	0.00	0.2
December	6	4.621	0.0781	0.05	4.9
	$\overline{X} = 3.7500$	$\hat{\mu} = 3.5195$	- - - - -	$P = 1.0$	$P[\%] = 100.00$

The regression of the average $\mu_i = e^{\left(\beta_3 * X_{it} + \ln(F^{\beta_4}) + \varepsilon_i\right)}$ has been made with the parameters $\beta_3 = 0.2454$ and $\beta_4 = 0.01076$, (parameters considered with WinBUGS14®), the considered annual average is $\hat{\mu} = 3.5195$, that aproximates better to the average sample, the calculation of the probability takes place using the Poisson-Gamma distribution, which parameters are $\hat{\mu} = 3.5195$ and $\phi = 47$. Table 5 shows the results of estimated accidents with this model for intersection 6.

RESULTS

Table No. 6 Estimation of regression parameters for the three models

No.	Intersection Name	Mixtures of Poisson $\mu_i = (\mu_1, \mu_2, \ldots, \mu_n)$		One-variable Poisson-Gamma $\mu_i = e^{\beta_2 x_{it}} \cdot e^{\varepsilon_i}$			Two-variable Poisson-Gamma $\mu_i = e^{(\beta_3 * X_{it} + \ln(F^{\beta_4}) + \varepsilon_i)}$			
		R_2	r	β_2	R_2	r	β_3	β_4	R_2	r
1	Benito Juárez y Miguel Hidalgo	0.405	0.636	0.297	0.954	0.977	0.257	0.016	0.992	0.996
2	Isidro Fabela e Independencia	0.000	0.000	0.261	0.941	0.970	0.219	0.022	0.961	0.980
3	Isidro Fabela y Alfredo del Mazo	0.000	0.000	0.285	0.921	0.960	0.277	0.011	0.925	0.962
4	Isidro Fabela y Lerdo de Tejada	0.400	0.633	0.294	0.961	0.980	0.260	0.019	0.968	0.984
5	Isidro Fabela y Miguel Hidalgo	0.037	0.192	0.294	0.961	0.980	0.253	0.022	0.969	0.984
6	José López Portillo y Alfredo del Mazo	0.410	0.640	0.252	0.904	0.951	0.245	0.011	0.908	0.953
7	José María Morelos e Isidro Fabela	0.281	0.531	0.326	0.969	0.984	0.304	0.012	0.975	0.988
8	José María Morelos y Vicente Villada	0.405	0.636	0.267	0.982	0.991	0.230	0.015	0.992	0.996
9	José María Morelos y López Rayón	0.000	0.000	0.266	0.954	0.977	0.224	0.018	0.989	0.995
10	José María Morelos y Benito Juárez	0.000	0.000	0.312	0.985	0.992	0.293	0.015	0.915	0.956
11	José María Morelos y Pino Suárez	0.0250	0.158	0.107	0.867	0.931	0.042	0.018	0.914	0.956
12	Santos Degollados y Pino Suárez	0.000	0.000	0.253	0.927	0.963	0.219	0.016	0.956	0.978
13	Tollocan e Isidro Fabela	0.000	0.000	0.329	0.962	0.981	0.302	0.019	0.962	0.981
14	Tollocan y Cristobal Colón	0.000	0.000	0.316	0.976	0.988	0.304	0.014	0.926	0.962
15	Tollocan y Pino Suárez	0.000	0.000	0.329	0.962	0.981	0.302	0.018	0.962	0.981
16	Tollocan y Vicente Guerrero	0.415	0.644	0.298	0.931	0.965	0.242	0.030	0.970	0.985
17	Venustiano Carranza y Benito Juárez	0.000	0.000	0.301	0.968	0.984	0.280	0.014	0.952	0.976
18	Venustiano Carranza y Jesús Carranza	0.184	0.429	0.235	0.951	0.975	0.192	0.016	0.984	0.992
19	Venustiano Carranza y Pino Suárez	0.000	0.000	0.221	0.936	0.968	0.173	0.018	0.981	0.990
20	Venustiano Carranza y Vicente Guerrero	0.000	0.000	0.331	0.965	0.982	0.302	0.015	0.972	0.986

Table 6 shows the parameters considered for the regression of the monthly average μ_i for the three proposed models, observing that the fitness R^2 for Mixtures of Poisson is found between 0.000 and 0.415 using linear regression. In a similar way, fitness R^2 for One-variable Poisson-Gamma model displays is better in relation to the previous one between 0.867 and 0.982. This can be attributed to the random effect expressed as $e^{\varepsilon_i} \sim Gamma\ (\alpha_0, \beta_0)$. Nevertheless, the Two-variable Poisson-Gamma model shows an improvement in the fitness of the model with R^2 between 0.914 and 0.992. This difference of fitness relative to the previous model is attributed to the covariance introduced like monthly vehicular flow, which means that the regression of the rate considered per month includes additional information. This seems to be confirmed by the considered rate $\hat{\mu} = 3.5195$, which approximates much to the rate sample $\overline{X} = 3.75$.

According to the previous analysis, the best model is the one with two explanatory covariants, i.e. the Two-Variable Poisson-Gamma model is selected to be used for the estimation of the rate of accidents considered for each one of the intersections included for our study. In this way, the probabilities for 1, 2, … crashes occurring in X_{it} in a period of time t at an intersection i independently to another intersection, given rates μ and the parameter of dispersion ϕ are displayed in Table 7. These probabilities are the maximum for each case, and they are the result of calculating each rate in each intersection. That is to say, if an intersection in this Table shows 3 accidents in it, then this is the number of crashes with the biggest probability to occur. These results have a 97.5% of confidence.

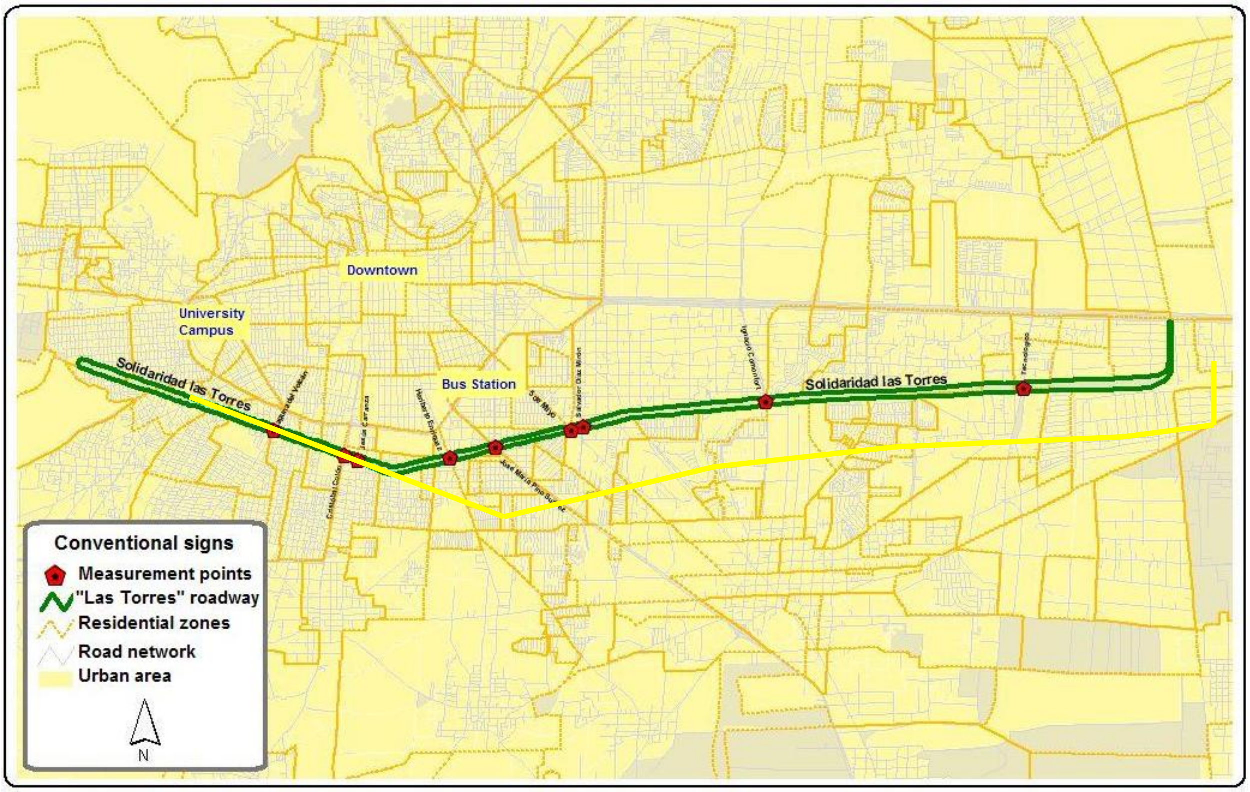
Table 7 Number of accidents with greater probability of occurrence

No.	Intersection Name	Accident number	% Maximum probability
1	Benito Juárez y Miguel Hidalgo	1	10.60
2	Isidro Fabela e Independencia	1	10.19
3	Isidro Fabela y Alfredo del Mazo	3	10.99
4	Isidro Fabela y Lerdo de Tejada	1	10.80
5	Isidro Fabela y Miguel Hidalgo	1	10.81
6	José López Portillo y Alfredo del Mazo	3	13.20
7	José María Morelos e Isidro Fabela	2	11.55
8	José María Morelos y Vicente Villada	1	10.28
9	José María Morelos y López Rayón	1	10.40
10	José María Morelos y Benito Juárez	2	11.30
11	José María Morelos y Pino Suárez	1	8.90
12	Santos Degollados y Pino Suárez	1	9.98
13	Tollocan e Isidro Fabela	2	9.96
14	Tollocan y Cristobal Colón	3	12.12
15	Tollocan y Pino Suárez	2	9.98
16	Tollocan y Vicente Guerrero	1	10.31
17	Venustiano Carranza y Benito Juárez	1	11.00
18	Venustiano Carranza y Jesús Carranza	1	10.01
19	Venustiano Carranza y Pino Suárez	1	9.85
20	Venustiano Carranza y Vicente Guerrero	2	10.81

CASE OF APPLICATION

As the two-variable Poisson-gamma model has shown the best performance, we have carried out an analysis of the intersections of one of the main routes as it is the avenue Solidaridad las Torres crossing with its secondary roadways (Figure 3).

Figure 3. “Solidaridad las Torres” avenue.



Selection of this route, commonly known by the locals as “Las Torres” due to lines of pylons run parallel to this it, is a main path, which communicates to the east with the exit to the Valley of Mexico and to the west with exits towards Zitacuaro and Bravo Valley, but at the same time presents intersections to enter to the city of Toluca.

Table 8 shows the name of each intersection and the monthly frequency of vehicular accidents. Vehicular flows for the same intersection set are in Table 9. In order to conform the data that will serve to feed a two-variable Poisson-gamma model, the GIS under the following criteria is used: First, “Las Torres” roadway is selected for identification of data of year 2000 available. In this case the sampling bias is smaller than in the 20 intersections with the biggest number of accidents shown before, and the analysis is treated at a microscopic level, because the use of manual accountability was necessary for the count of accidents.

Table 8.Monthly accident data, "Las Torres" avenue, year 2000

No.	Intersection Name	Jan	Feb	Mar	April	May	June	July	Aug	Sep	Oct	Nov	Dec
1	Torres y Paseo Tollocan	1	1	2	2	3	2	1	0	0	4	0	0
2	Torres y Tecnológico	1	0	2	0	2	2	0	2	0	0	3	0
3	Torres y Ignacio Comonfort	1	4	0	3	0	0	3	0	0	0	0	0
4	Torres y Salvador Díaz Mirón	0	0	3	2	3	1	0	0	1	3	3	1
5	Torres y 5 de Mayo	4	1	1	0	0	0	5	1	0	3	0	1
6	Torres y José María Pino Suárez	2	1	1	1	6	3	0	2	4	3	1	3
7	Torres y Heriberto Enriquez	0	0	0	0	0	0	0	0	0	0	0	0
8	Torres y Jesús Carranza	0	0	2	0	3	1	1	3	2	1	1	0
9	Torres y Cristobal Colón	0	1	2	1	1	0	1	2	3	3	0	3
10	Torres y Laguna del Volcan	1	2	1	0	2	0	4	2	0	1	4	0

Table 9. Monthly vehicular flows x 10 000, "Las Torres" avenue, year 2000

No.	Intersection Name	Jan	Feb	Mar	April	May	June	July	Aug	Sep	Oct	Nov	Dec
1	Torres y Paseo Tollocan	75	77	75	75	75	75	75	72	75	72	71	76
2	Torres y Tecnológico	57	60	61	59	59	59	60	61	60	58	59	58
3	Torres y Ignacio Comonfort	66	67	65	69	65	69	67	65	70	67	69	65
4	Torres y Salvador Díaz Mirón	88	90	89	90	87	90	88	89	86	87	87	89
5	Torres y 5 de Mayo	102	105	107	110	104	102	110	109	105	105	108	108
6	Torres y José María Pino Suárez	194	198	200	205	194	205	203	198	194	194	207	207
7	Torres y Heriberto Enriquez	59	61	61	62	63	61	62	59	59	60	61	63
8	Torres y Jesús Carranza	68	66	70	66	65	66	68	68	67	70	68	66
9	Torres y Cristobal Colón	65	67	65	65	65	65	68	69	68	69	65	68
10	Torres y Laguna del Volcan	59	60	57	57	60	58	57	56	57	60	60	60

Table 10. Probability of accidents, "Las Torres" avenue. Flow data of year 2000

No.	Intersection Name	Number of Accidents	% Probability
1	Torres y Paseo Tollocan	2	17.40
2	Torres y Tecnológico	1	33.45
3	Torres y Ignacio Comonfort	3	15.30
4	Torres y Salvador Díaz Mirón	3	15.63
5	Torres y 5 de Mayo	1	18.01
6	Torres y José María Pino Suárez	1	13.70
7	Torres y Heriberto Enriquez	0	0.00
8	Torres y Jesús Carranza	2	19.65
9	Torres y Cristobal Colón	3	19.15
10	Torres y Laguna del Volcan	0	10.06

Vehicular flows in Table 9 are the sums of all the streams entering to each intersection in dirtections east to the west, west to east, north to south and south to north. The measurements were made through installed sensors and directional manual counting from 6:00 to 21:00 hours, considering that the flow of registered traffic includes private and public vehicles.

Table 10 is constructed with data in Tables 8 and 9.

Table 10 uses accident statistics and flow measurement for the same year an intersection sets. With it, we can try to predict some tendencies. For example, vehicular flow data for year 2008 are available with enough precision degree (Table 11).

Table 11. Monthly vehicular flows x 10 000, "Las Torres" avenue, year 2008

No.	Name	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	Torres y Paseo Tollocan	246	252	246	243	243	246	243	235	243	235	233	249
2	Torres y Tecnológico	186	195	197	190	190	190	195	197	195	188	190	188
3	Torres y Ignacio Comonfort	233	235	227	241	230	243	235	230	246	235	241	227
4	Torres y Salvador Díaz Mirón	264	273	267	273	261	273	264	267	258	261	261	267
5	Torres y 5 de Mayo	275	285	288	297	281	275	297	294	285	285	291	291
6	Torres y José María Pino Suárez	337	345	349	357	357	357	353	345	337	337	361	361
7	Torres y Heriberto Enriquez	188	197	195	199	201	197	199	188	190	192	195	203
8	Torres y Jesús Carranza	248	240	253	240	237	240	245	245	242	253	245	240
9	Torres y Cristobal Colón	218	223	216	218	218	216	228	231	226	231	216	228
10	Torres y Laguna del Volcan	160	162	155	153	162	156	153	151	155	162	162	162

We can make a calculation in the two-variable Poisson-gamma model by entering these flow data, leaving unchanged the rest of the values, to obtain the results in Table 12.

Table 12. Table 10. Probability of accidents, "Las Torres" avenue. Flow data of year 2008

No.	Intersection name	Number of accidentes	% Probability
1	Torres y Paseo Tollocan	2	17.20
2	Torres y Tecnológico	1	33.94
3	Torres y Ignacio Comonfort	3	14.03
4	Torres y Salvador Díaz Mirón	3	14.80
5	Torres y 5 de Mayo	3	28.65
6	Torres y José María Pino Suárez	1	13.80
7	Torres y Heriberto Enriquez	0	0.00
8	Torres y Jesús Carranza	2	18.92
9	Torres y Cristobal Colón	3	15.69
10	Torres y Laguna del Volcan	1	21.08

Even though some concern can araise with the suposition made, it is possible to observe that interesting outcomes has been obatined. For example, from 2000 to 2008 the flows have increased, but in most of the intersections the probability of having a number accidents remains almost the same. However, one intersection, "Torres and 5 de Mayo ", has a considerable increment. This is remarkable due to its respective flow is not incremented as the others flows in the rest of the intersections, and could lead to think that the relations of flows in "Las Torres" Avenue have made it more dangerous as time has past.

This result can help to direct a decision making process, by focusing on those intersections that present an increment in their number accident calculated with this prediction tool. Series of calculations that can be adjusted by a series of respective observations can constitute a data base of tendencies, useful for those professionals dedicated to the urban transit.

CONCLUSIONS

We present three Poisson-like models under a bayesian approach that can be used to predict the number of accidents occuring in individual intersections from data collected. This is useful to obtain a priori information of events that can lead to economical and personal damages in order to design strategies to avoid them.

This models have been analyzed to compare their respective goodness. In this way, Mixtures of Poisson model has shown a fittnes to real data very poor, due to their lack of flexibility about rates that are mainly variable, and because the use of only one kind of variable.

On the other hand, a Poisson-like approach, with the same variable, has a better behavior with the real data, improving forecating performance due to the introduction of a random effect.

However, if a second variable is introduce to the data set, as the vehicular flows, present in the respective intersections, and process by a two-variable approach, then the correlation with actual data is improved significatively.

Once the best result for data fitness is reached, the corresponding approach is utilized then to perform estimations and to define the 20 most dangerous intersections in the city of Toluca through the help of a GIS. This could be a first step to focus on this set of intersections in order, fpor example, to establish actions to conduct studies in order to reduce the danger in these urban intersections.

This bayesian approach permits to understand the accident events in an statistical manner. Not only leads to identify intersections with high degree of danger, but it also estimates monthly and annual rates about the probability of having a determined number of accidents, with which it is possible to better explained this phenomenon for people in charge of taking decisions about traffic safety or urban planning.

ACKNOWLEDGEMENTS

We wish to thank to the Municipality of the city of Toluca, Mexico, for the data and means given to the realization of this work.

REFERENCES

Abbess C., Jarrett D. (1981). Accidents al blackspots: estimating the effectiveness of remedial treatment with spacial reference to the regression to mean effect. Traffic engineering and control. 535-542

Gregory, P. (2005). *Bayesian logical data analysis for the phisycal sciences*. Cambrige University.

Hauer E., Harwood J., Council F. and Griffith M. (2002). “Estimating safety by the empirical Bayes method”. *Journal of the transportation research board.* 02-2181, 126-131.

Hinojosa, R. (2003). “Administración y análisis de datos de accidentes de tránsito en la ciudad de Toluca, Estado de México, utilizando como herramienta un sistema de información geográfica”. Engineering Masters Dissertation, Universidad Autónoma del Estado de México, Toluca, México.

Kononov J. and Janson B. (2002). “Diagnostic methodology for the detection of safety problems at intersections”. *Journal of the transportation research board.* 02-2148, 51-56.

Lee, C., Saccomanno, F., Hellinga, B. (2002). “Analysis of Crash Precursors on Instrumented Freeways”. *Transportation Research Record*. 1784, 02-3790, 1-8.

Lord D. (2002). “Application of accident prediction models for computation of accident risk on transportation networks”. *Journal of the transportation research board*. 02-3598, 17-26.

Lord D., Washington, S. M., Ivan, J. N, (2005). “Poisson-gamma and zero inflacted regression models of motor vehicle crashes: balancing statistical fit and theory”. *Journal of the transportation research board*. 37, 35-46.

Miaou S. and Lord D. (2003). “*Modeling traffic crash-flow relationship for intersections”*. Dispersion parameter, functional form, and Bayes versus empirical Bayes methods. Transportation Research Record 1840, 03-3363, 31-41

Miaou S. and Jin J. (2005). “Bayessian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence”. Elsevier, accident analysis and prevention. 37, 699-720.

Miranda L., Fu L., Saccomanno F. and Labbe A. (2005). “Alternative risk models for ranking locations for safety improvement”. *Journal of the Transportation Research Board* 1908, 1-8.

Mitra S., Chor Chin H. and Abdul M. (2002) “Study of intersection accidents by maneuver type”. *Journal of the transportation research board.* 02-3436, 43-50

Qin X., Ivan J. and Ravishanker N. (2002) “Selecting exposure measuresin crash rate prediction for two-line highway segments”. Elsevier, accident analysis and prediction. 36, 183-191

Rios I., Martin J., Rios D. and Ruggeri F. (1999) " Bayesian forecasting for accident proneness evaluation" *Scandinavian Actuarial Journal*. (Dec. 3, 1999), 134-156.

Persaud, B; Lord, D. and Palmisano, J. (2002) "Calibration and Transferability of Accident Prediction Models for Urban Intersections". *Transportation Research Record* 1784, 02-3293, 57-64.

Ross, S. (1989) *A First Course in Probability.* Academic Press.