

Karaganis, Anastassios; Mimis, Angelos

**Conference Paper**

## A Spatial Point Process for Estimating the Probability of Occurrence of a Traffic Accident

46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean", August 30th - September 3rd, 2006, Volos, Greece

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Karaganis, Anastassios; Mimis, Angelos (2006) : A Spatial Point Process for Estimating the Probability of Occurrence of a Traffic Accident, 46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean", August 30th - September 3rd, 2006, Volos, Greece, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118457>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# A Spatial Point Process for Estimating the Probability of Occurrence of a Traffic Accident.

ANASTASSIOS N. KARAGANIS

Panteion University of Social and Political Sciences

[akaragan@panteion.gr](mailto:akaragan@panteion.gr),

ANGELOS MIMIS

Panteion University of Social and Political Sciences

[amimis@panteion.gr](mailto:amimis@panteion.gr),

## Abstract

*In this paper a pioneer method for evaluating the probability of occurrence of a traffic accident in different segments of national road is described. The line segments describing the road have been divided into sections and for each one of those sections it is assumed that the probability of a traffic accident follows an inhomogeneous Poisson distribution. The overall model is build up by using a spatial point process. The Poisson parameter is estimated with the help of a spatial SUR method which exploits the characteristics of national road as described by a specially designed weight matrix. This methodology has been applied to data collected from the Greek national road and has been analyzed in this perspective. Specifically the influence of the presence of daylight has been examined.*

**Keywords:** Spatial point process, spatial econometrics.

## 1. Introduction

In recent years traffic control and safety has become increasingly critical. This is true for both urban traffic and national highway. Research is seeking the determinants of traffic accidents, the methods of prediction and their consequences.

Maher and Summersgill (1996) propose a methodology for fitting of predictive accident models. They relate traffic accidents and traffic volume. They review previous accident studies of UK Transport Research Laboratory. The studies cover

accidents which occur at junctions of various types. They test and accept the advantages of the use of Poisson distribution in the context of the generalized linear model. Finally they extend their analysis with the use of alternative distributions, such as Negative Binomial and quasi-Poisson.

Results of Mayer and Summersgill have been extended by Wood (2002). His study focuses in the direction of goodness of fit testing. A related subject, the safety modeling is also researched.

Finkelstein (1998) applied a stochastic point process model to describe accidents at sea. He uses inhomogeneous Poisson process in the plane, and converts the results of the one-dimensional case into a two dimensional spatial model with the proper parameterization. He studies ship route, and estimates the possibility of accident in that route. Then he estimates the point process which governs this specific route and finally compares the derived point process of all the routes under consideration.

Thomas (1995), examines to what extent the length of road segment influences the statistical description of traffic accidents. She concludes that individual accidents follow the Poisson distribution. Moreover for small levels of aggregation, or in other words for small segments of the highway, Poisson assumption should not be rejected. In general as the level of aggregation is increased, the observed data tend to follow the normal distribution. Large segments are considered those with length bigger than 2100 meters.

Diggle *et al.* (2005) propose an inhomogeneous Poisson point process methodology for spatio-temporal surveillance telephone calls to clinical advice service of a disease. They formulate a multiplicative model consisting of a spatial point process, governing the site of the telephone call, a temporal point process governing the time of the call and an unobserved process describing the severity of the disease incident. They found that the use of point process is advantageous, especially in treating the unobserved component of the model. They also propose an early warning system based on differences between expected and observed calls.

Giuliano (1989) classifies traffic accidents by rate, type, and consequence on the road system. This study focuses mainly on the consequences of accidents, and utilizes duration models.

Svensson and Hyden (2006) investigate the severity of safety related behaviour. They relate the severity hierarchy with road safety. They conclude that severity hierarchy of road incidents describes differences in road user behaviour and predicts the frequencies of the most severe events.

This paper focuses on the use of spatial econometrics analysis for detection of dangerous sites in the national highway system and we evaluate the probability of occurrence of an accident at a specific segment of the road given that the accident has happened.

In the first section of the paper the spatial model process is described and the assumptions followed are analyzed. Then the data collected is described, and the results are illustrated. Finally the importance of the proposed method is given.

## 2. Model specification

Statistically, traffic accidents are defined as a process which is indexed by the exact location and the time (day or night) at which have occurred. The objective of the analysis is to estimate the pattern of dangerous segments of the national road system and the identification of changes of the level of danger during day and night time.

According to Cressie (1994) a spatial point process is a stochastic model governing the location of events of interest  $\{s_i\}$  in some set  $X$ . In our case, as event of interest is considered the traffic accident and the set  $X$  is taken as a subset of  $R$ , because each site-point of the national highway is represented by the points of  $R$ , which is the set of real numbers. This description embodies the three important features pointed out by Daley and Vere-Jones (2005):

- (i) the number of traffic accidents in each finite segment of the national highway  $(km_i, km_{i+1}]$  of a specific kilometer length has a Poisson distribution;
- (ii) the number of traffic accidents in disjoint segments are independent random variables; and
- (iii) distributions are either stationary or not: they specific to each one segment, besides the fact that they depend on the lengths  $km_{i+1} - km_i$  of the segments.

Thus, the joint distributions are multivariate Poisson. So, the probability of occurrence of a traffic accident in a segment of the national highway of any length, although in the case of this paper the kilometer length for all segments is 1 kilometer, is given by:

$$\Pr\{N(km_i, km_{i+1}] = n_i, i = 0, \dots, KM\} = \prod_{i=0}^{KM} \frac{[\lambda(km_{i+1} - km_i)]^{n_i} e^{-\lambda(km_{i+1} - km_i)}}{n_i!} \quad (1)$$

where  $N(km_i, km_{i+1}]$  stands for variable for the number of traffic accidents in the specific segment,  $km_i$  is the kilometer distance from the beginning of the road,  $KM$  is the total length of the highway and  $\lambda$  is the common Poisson distribution parameter, for a traffic accident to occur at a specific point.

Intensity is defined as the density, which in our case is the average number of traffic accidents per segment as the length of the segment tends to zero, coincides with the common Poisson distribution parameter [Daley, Vere-Jones, page 20, (2005)] and is given by:

$$\lambda = \lim_{(km_i, km_{i+1}] \rightarrow 0} \frac{E[N(km_i, km_{i+1}]]}{(km_i, km_{i+1}]} \quad (2)$$

where  $N(km_i, km_{i+1}]$  is the number of traffic accidents within an infinitesimal region of length  $km_{i+1} - km_i$ , that contains the site where a traffic accident may have occurred. It should be noted that intensity  $\lambda$  tends to a constant value.

The constant  $\lambda$  can be interpreted as the mean rate, or mean density of the points in the process. It also coincides with the intensity of the process as defined, following Khinchin's existence theorem [Daley and Vere-Jones (2005)]. Constant  $\lambda$  also means that traffic accidents are characterized by complete spatial randomness, or in other words that all the segments of the highway are equally dangerous.

The mean  $M(km_i, km_{i+1}]$  and variance  $V(km_i, km_{i+1}]$  of the number of traffic accidents falling in the segment  $(km_i, km_{i+1}]$  are given by:

$$M(km_i, km_{i+1}] = \lambda(km_{i+1} - km_i) = V(km_i, km_{i+1}] \quad (3)$$

Relation (3) is useful for the construction of a diagnostic test for the stationary-homogeneous Poisson process. Thus according to (3) the ratios of means and variances over a range of different segment lengths should be approximately equal, for a stationary Poisson process and equal to the mean rate [Daley and Vere-Jones, page 20, (2005)], otherwise there is strong evidence of the existence of an inhomogeneous Poisson process.

An alternative to homogeneous Poisson process, i.e. complete spatial randomness, is the inhomogeneous Poisson process, which is also the generalization of Poisson point process [Cressie, page 620, (1993)]. The intensity of an inhomogeneous Poisson process is a function of spatial location. An inhomogeneous Poisson process has a space-varying intensity  $\lambda(x)$ , with  $x$  indicating the location of the traffic accident. The process is defined exactly as in the homogeneous Poisson case -equation (1), with the

quantities  $\lambda(km_{i+1} - km_i) = \int_{km_i}^{km_{i+1}} \lambda dx$  replaced by the quantity

$$\lambda_i(km_i, km_{i+1}] = \int_{km_i}^{km_{i+1}} \lambda(x) dx \quad (4)$$

That means that the Poisson parameter  $\lambda$  is constant within each segment but differs between segments, as is required by (2). Thus, the joint distributions are still Poisson, and the independence property still holds.

The dangerous points are identified through a multiplicative decomposition of the space and time intensity of traffic accidents point process, which includes separate terms for spatial intensity estimated for day and night, and the overall probability of the traffic accident occurrence. Hence the intensity of traffic accidents Poisson distribution for the general inhomogeneous case is:

$$\lambda_{ji} = \lambda_j(x) P_{accident}, \quad x \in (km_i, km_{i+1}], \quad j = Day, Night \quad (5)$$

The overall probability of traffic accident is the ratio of the number of traffic accidents to the total volume of traffic. It is considered irrelevant in our case study and thus ignored, because it is assumed to be the same for all the segments. So, in our model the parameter of special interest is the intensity of the Poisson distribution for each segment of the highway, which in the case of an inhomogeneous process will be different in each segment, while on the homogeneous case intensity parameter will be the same for all segments.

The proposed model for space-time variation combines a linear model for the observed data, alongside a model for the spatial effect representing the unobserved impact of the prevailing road conditions in any specific road segment and its neighboring segments as well.

The study of the data of the traffic accidents collected, which are presented in the following section; show a strong difference between accidents which take place during daytime and nighttime. Based on this fact a system of predicting the probability of occurrence of an accident during day and night has been devised.

In order to estimate  $\lambda_{ji}$ , a spatial SUR formulation has been used with a spatially dependent error term. All the effects influencing the process have been taken into consideration, and are either included in an explanatory proxy variable or in the residuals [Shankar *et al.* (1997)]. The proxy variable includes the effects of the geometrical and physical characteristics of the road while the other non-controllable factors such as locally prevailing weather conditions, driving behavior, and traffic volume are included in the residuals. The time has been divided into two periods, day and night, and an equation of a spatial SUR model has been applied for each one of them. Further due to the spatial nature of the data, the concept of spatial autocorrelation is treated, in analogy with time series autocorrelation (Bartels 1979 pages 20-21), data has been also included into the model which is given in matrix form by:

$$\begin{aligned}\lambda_{Day} &= \beta_{11}\mathbf{1} + \beta_{12}X_{Day} + \varepsilon_{Day} \\ \lambda_{Night} &= \beta_{21}\mathbf{1} + \beta_{22}X_{Night} + \varepsilon_{Night}\end{aligned}\quad (6)$$

The error term in matrix form can be specified as

$$\begin{aligned}\varepsilon_{Day} &= r_{Day}W\varepsilon_{Day} + v_{Day} \\ \varepsilon_{Night} &= r_{Night}W\varepsilon_{Night} + v_{Night}\end{aligned}\quad (7)$$

The vectors in equations (6) have 94 rows, describing the road segments of analysis. The number of accidents occurred is described in the  $\lambda$  vectors, which are the dependent variables of the model. The independent variable  $X$  is the proxy variable, while  $\mathbf{1}$  is the vector of ones used for the estimation of the constant term. Further the  $\beta_{ij}$  are the scalar coefficients of the regressions. In equation (7) the spatial dependence structure of the error term is described with the help of the 94x94 contiguity matrix  $W$ . This matrix describes algebraically the spatial structure of the national highway i.e. defines the contiguous-neighboring segments by the Cartesian product of the segment space. As spatially contiguous segments, are defined, those which are successive. Each segment is not considered contiguous to itself. All segments have two neighbors while the first and the last have just one. The elements of matrix  $W$  which correspond to contiguous segments take the value of one while all the others are zero. The 94x94 contiguity matrix  $W$  has the following form:

$$W = \begin{bmatrix} 0 & 1 & 0 & 0 & ] & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & ] & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & ] & 0 & 0 & 0 & 0 \\ ] & ] & ] & ] & ] & ] & ] & ] & ] \\ 0 & 0 & 0 & 0 & ] & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & ] & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & ] & 0 & 0 & 1 & 0 \end{bmatrix}\quad (8)$$

In the model the contiguity matrix  $W$  is row standardized by dividing each element of the contiguity matrix  $W$  by the row sum. That is based on the previous work done by Tiefelsdorf (2000). Vector  $v$  is a normally distributed random variable with zero mean and constant variance.

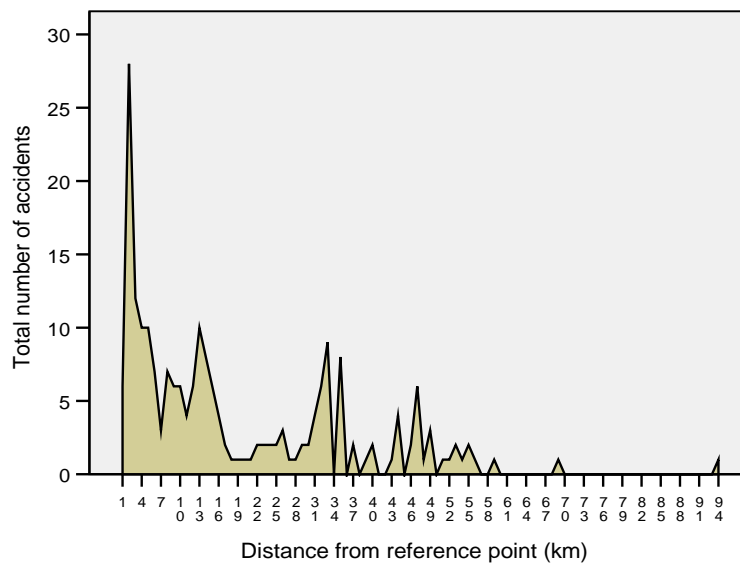
This model description fulfills two purposes. The first is to incorporate into the model the spatial impact on the intensity, while the second is to establish criteria which identify whether the process governing the phenomenon is stationary or inhomogeneous.

### 3. Data

The data have been collected by the Traffic Police Division on a 94 km segment of the national road. This segment starts from Antirio, which is referred and as a reference point in our study, and finishes at Agrinio. The road has one traffic line on each side and there is no parapet in the middle of the road. Also in the first 20 km there are straight parts of about 1 km, followed by dangerous turns. The data used in the analysis concern the period 2001-2004, and the information provided include place, time, and severity of the accident.

As it was expected by the condition of the road and can be seen in Figure 1, accidents tend to occur in the first part of the segment.

Figure 1  
Total number of accidents

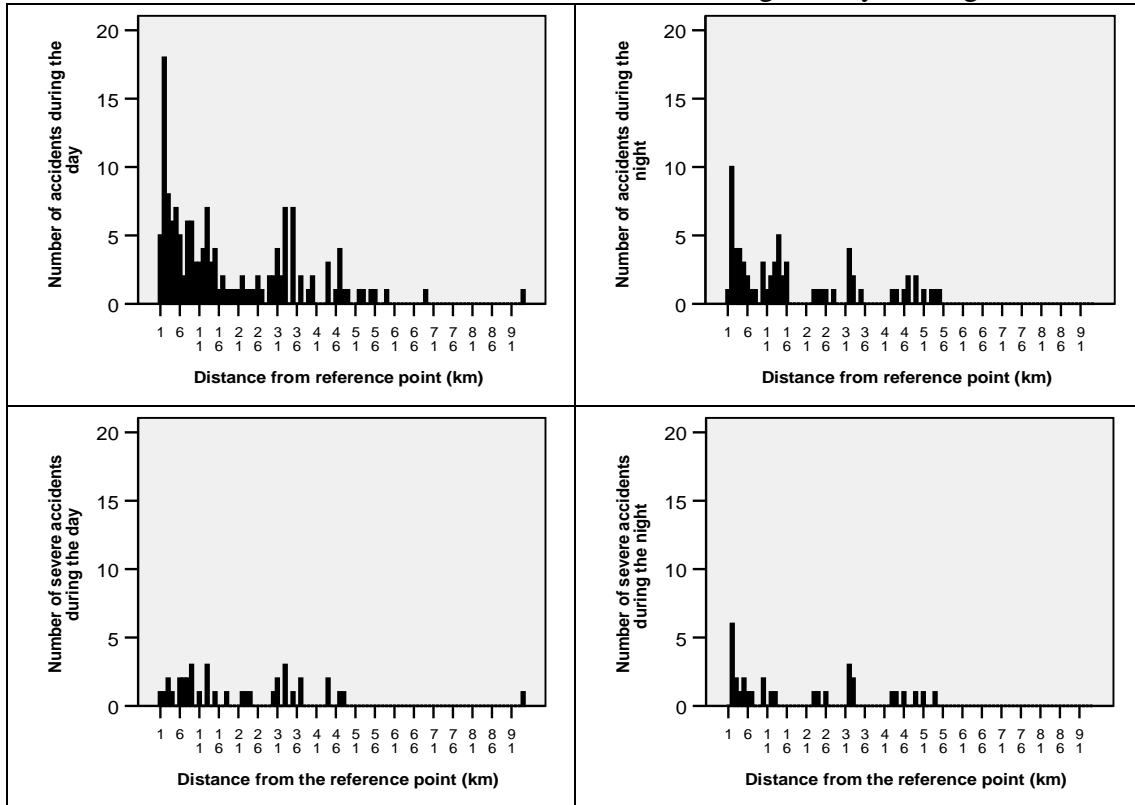


Source: Hellenic Police Headquarters, Traffic Police Division

Figure 2 shows that the number of accidents, at the same location, is higher during the day comparing to the night, due to a different mean traffic volume and different prevailing road conditions. Figure 2 depicts also the relation between the total number of accidents and the severity of them. Traffic accident severity, according to the literature [Svensson and Hyden (2006)], encapsulates differences in road physical characteristics and predicts the frequencies of the most severe events.

Figure 2

Number of accidents and severe accidents during the day and night



Source: Hellenic Police Headquarters, Traffic Police Division

#### 4. Results

The model has been implemented in Gauss 5.0 and it is based on the CML library (Constrained Maximum Likelihood Estimation library, Aptech Systems 2001). This library is a set of procedures for the estimation of the parameters of models via the maximum likelihood method with general constraints on the parameters, along with an additional set of procedures for statistical inference.

The spatial SUR model has been applied and the results are illustrated in Table 1. It should be noted, that all estimates are statistically significant. Especially, statistical significance of the slope coefficient verifies the hypothesis of inhomogeneous Poisson spatial process, due to the dependence of intensity to the physical characteristics of the road. Moreover, the higher value of the estimated coefficient for the constant of the day equation reflects the bigger traffic volume during the day.

Table 1  
Maximum likelihood estimation results

Parameters	Estimates	Std. err.	t
Day constant, $\beta_{11}$	0.8846	0.2600	3.4023
Day severity, $\beta_{12}$	1.7148	0.2175	7.8841
Night constant, $\beta_{21}$	0.4217	0.1178	3.5798
Night severity, $\beta_{22}$	0.8968	0.1163	7.7111
Day lamda, $r_{Day}$	-0.1723	0.0700	2.4614
Night lamda, $r_{Night}$	-0.2440	0.0657	3.7139
Log-likelihood	-118.621		

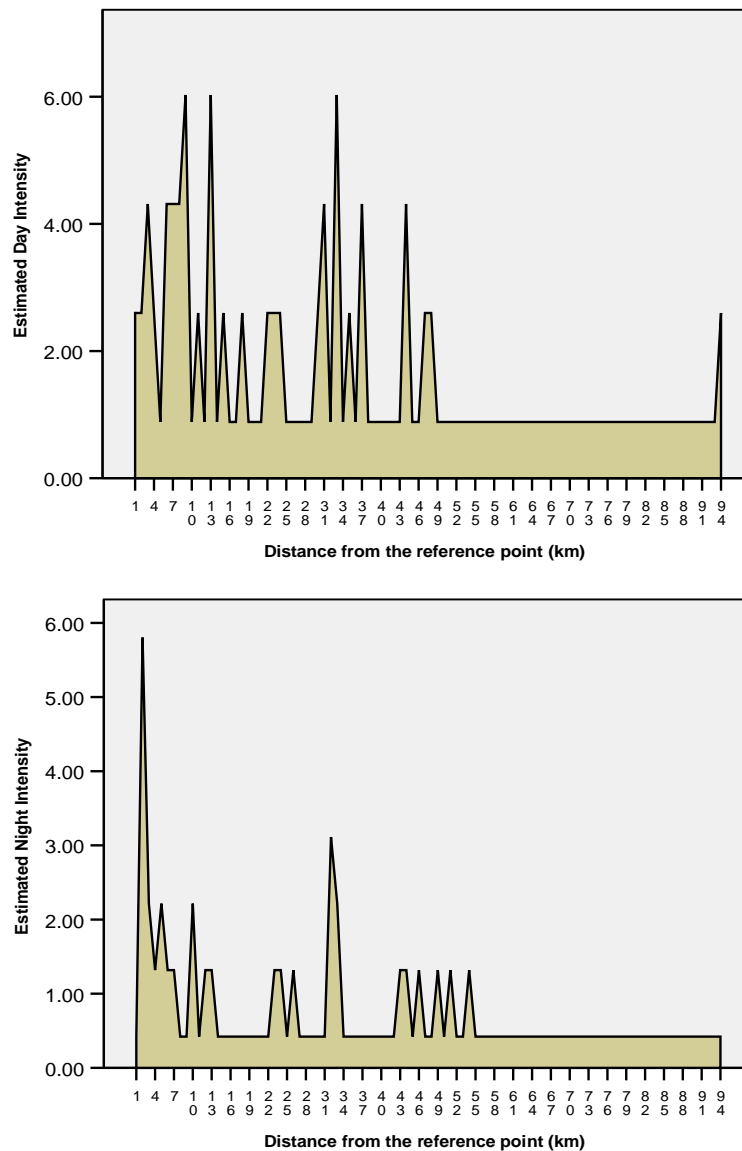


Table 2 verifies the use of SUR because the cross equation covariances are statistically significant.

Variiances, covariances	Estimates	Std. err.	T
S <sub>11</sub> (day equation)	4.1324	0.6022	6.8622
S <sub>12</sub> (cross equation)	1.1796	0.2618	4.5057
S <sub>21</sub> (cross equation)	1.1796	0.2618	4.5057
S <sub>22</sub> (night equation)	0.7579	0.1362	5.5646

Figure 3 depicts the prediction results of the aforementioned SUR model. The first part of the figure shows that the dangerous segments are more spread during the day. Moreover by comparing both figures, it can be seen that the level of danger of specific segments fluctuates less during the night. Finally, the dangerous spots during the night are slightly shifted to the left in comparison to the corresponding dangerous spots of the day distribution.

Figure 3  
Estimated intensities ( $\lambda_{ji}$ )



## 5. Conclusions

In this study a methodology for evaluating the probability of occurrence of a traffic accident is proposed. The model is build up by using a spatial point process, where the Poisson parameter is estimated with the help of a spatial SUR model. The model has been applied to data collected from the Greek national road, in order to predict the dangerous spots during day and night. The model has successfully captured the peculiarities of the road such as the existence of straight parts, followed by dangerous turns.

In a future extension of the model, more factors affecting the dangerousness of the road will be included and their effect on the results will be studied. Further the effect of differently defined weight matrices  $W$  will be tested on the perspective of incorporating the spatial effects more accurately. Finally this method will be applied to predict the probability of occurrence of a traffic accident during seasons around the year conditioned on day and night time.

## Acknowledgements

The authors would like to thank Ioulia Galifa for compiling the data, and Hellenic Police Headquarters (Traffic Police Division) for providing the data.

## References

- Anselin, L.(1988), "*Spatial Econometrics: Methods and Models*", Kluwer Academic Publishers, Dordrecht.
- Aptech Systems (2001), "*Maximum Likelihood Estimation*".
- Daley, D.J. and Vere-Jones D. (2005), "*An Introduction to the Theory of Point Processes – Volume I: Elementary Theory and Methods*", 2<sup>nd</sup> Edition, Springer, New York.
- Diggle, P., Rowlingson, B., and Su T.L. (2005), "Point Process Methodology for on-Line Spatio-Temporal Disease Surveillance", *Environmetrics*, Vol. 16, pp. 423-434.
- Finkelstein M.S. (1998), "A Point-Process Stochastic Model with Application to Safety at Sea", *Reliability Engineering and System Safety*, Vol. 60, pp. 227-233.
- Giuliano G. (1989), "Incident Characteristics, Frequency, and Duration on a High Volume Urban Freeway", *Transpn. Res. –A*, Vol. 23A(5), pp.387-396.
- Shankar, V., Milton, J. and Mannering, F. (1997), "Modeling Accident Frequencies as Zero-Altered Probability Processes: An Empirical Inquiry", *Accid. Anal. And Prev.*, Vol. 29(6), pp. 829-837.
- Maher, M., and Summersgill, I. (1996), "A Comprehensive Methodology for the Fitting of Predictive Accident Models", *Accident Analysis and Prevention*, Vol. 28(3), pp. 281-296.
- Svensson, A. and Hyden, C. (2006), "Estimation the Severity of Safety Related Behaviour", *Accident Analysis and Prevention*, Vol. 38, pp. 379-385.
- Thomas I. (1996), "Spatial Data Aggregation: Exploratory Analysis of Road Accidents", *Accident Analysis and Prevention*, Vol. 28(2), pp. 251-264.
- Tiefelsdorf M., (2000), *Modelling Spatial Processes*, Springer.

- Walter, C., McBratney, A.B., Rossel, R.A.V., and Markus, J. (2005), "Spatial Point-Process Statistics: Concepts and Application to the Analysis of Lead Contamination in Urban Soil", *Environmetrics*, Vol. 16, pp. 339-355.
- Wood G.R. (2002), "Generalised Linear Accident Models and Goodness of Fit Testing", *Accident Analysis and Prevention*, Vol.34, pp. 417-427.