

# Adapting a Local Indicator of Spatial Association to Identify Hot Spots in Traffic Safety

Elke Moons<sup>1</sup>, Tom Brijs<sup>1</sup>, Geert Wets<sup>2</sup>

<sup>1</sup>Transportation Research Institute (IMOB), Hasselt University, Science Park 1/15, 3590 Diepenbeek, BELGIUM;  
E-mail: [elke.moons@uhasselt.be](mailto:elke.moons@uhasselt.be)

<sup>2</sup>Transportation Research Institute (IMOB), Hasselt University, Science Park 5/6, 3590 Diepenbeek, BELGIUM

## Abstract

In the last years, traffic safety has become a 'hot' topic in the media, for policy makers, for academics and for the broad audience. In general, traffic safety analysis can be split up in four phases: identification of hot spots, ordering, profiling of the hazardous locations and finally selecting the locations that need to be handled. This paper will focus on phase 1. Several spatial measures have been developed to identify hot spots, however, not all measures are equally well suited in the field of traffic safety. E.g. a high-accident location is always of interest for hot spot analysis, but a low-accident location may be of interest if its neighbouring locations show an opposite behaviour regarding the number of accidents. In this paper we will present an adaptation of a local measure of spatial association, more in specific of Moran's I and apply it to identify hot spots on highways.

**Key Words:** spatial statistics, local indicator of spatial association, Moran's I, hot spot analysis, traffic safety

## 1. Introduction

In the past decades, traffic accident figures have become a topic of increasing interest in the media, as well as for policy makers. As opposed to most other European countries, Belgium's score concerning traffic safety is still below par. The number of people having a fatal accident per 1 billion vehicle kilometers equals 11.1 in Belgium in 2006 (IRTAD, 2008). When compared to our neighboring countries, this figure is about 31% higher than France, 44% more than The Netherlands and even 50% higher when compared to Germany. Putting these figures in an international context shows that also there Belgium performs rather poorly (The United States have a figure of 9 persons killed per 1 billion vehicle-kilometer, Australia has a value of 7.9 and Japan of 10.3) (Note that these latter figures are of 2003). Therefore, it seems only logical that traffic safety has become top priority in the National Safety Plan. Furthermore, in the States General of Traffic Safety (2007), the ambitious goal has been set to reduce the number of individuals killed in traffic per year from 1,000 in 2006 to 500 by 2015.

A key topic in traffic safety analysis is determining the reason for a site to be hazardous, also referred to as hot spot analysis (HSA). In general, HSA can be split up into four phases. At first, one needs to identify the dangerous locations/zones. Next, a ranking of these locations needs to be established. The severity of the crash, determined by the severity of the injuries, can be taken into account here (see e.g. Brijs *et al.*, 2006, 2007; Vistisen, 2002). Afterwards, one tries to come up with an explanation why some sites are hot spots and others are not (i.e. profiling of hot spots). This can be verified through an analysis of manoeuvre diagrams, conflict observations, information from traffic accident records, characteristics of the environment, of the infrastructure, etc. (Geurts *et al.*, 2005; Jianming and Kockelman, 2006). And, finally, the hot spots to be treated need to be selected (Miranda-Moreno *et al.*, 2007). The latter is very often a policy decision and the choice may be based on different aspects: e.g. based on limited financial supplies, or a cost-benefit analysis (Kar and Datta, 2004). Only the first phase, identification, will be discussed in this paper, although the technique could be applied for the purpose of ranking as well. To reduce the number of collisions, it is important to know where concentrations of crashes occur. Consequently, the geographical aspect is of large importance to determine and to handle the most unsafe traffic sites in a scientifically sound and practical way.

According to Hauer (1996), there is no univocal definition of a hot spot. Some researchers rank locations based on the number of crashes per vehicle-kilometer driven (VKD) or per number of vehicles, others use the ‘crude’ number of crashes (crashes per km/year or per year), and some use a combination of both. Although one acknowledges the importance of the geographical aspect, very often statistical – non-spatial – regression models are used to model the number of crashes.

When analyzing hot spots within in a certain time frame, one needs to acknowledge that at a large number of locations no crashes occurred for that period of time. This is recognized in the literature as an abundance of zeroes. This sparseness of data causes estimation problems in most prediction models. Negative binomial models have been applied to solve this problem and in the recent past this was often countered by using Zero Inflated Poisson (ZIP) models. Though, recently this was criticized in the literature, because there is no theoretical underpinning to believe that there exists a location that is inherently safe. This can be solved on the one hand by enlarging the time frame or the geographical window or by using a better set of explanatory variables and/or by taking non-observed heterogeneity effects into account to explain the model or by applying methods for small area estimation (e.g. Poisson-lognormal models) (Lord *et al.*, 2005, 2007).

Next to applying a frequentist’s approach, traffic safety literature also leans towards the use of (empirical) Bayesian models, since they can make use of prior information in an efficient way. The Poisson-gamma model is an example that is widely used in traffic safety literature (Hauer, 1997; Hauer *et al.*, 2003a, 2003b; Cheng and Washington, 2005). Researchers prefer to use it to the Poisson regression model, because this model deals with the problem of overdispersion. The Poisson distribution, underlying the regression model, assumes that the mean and variance are equal to each other and since the mean number of crashes usually is very low, accident data often show a larger variance. Very recently (Park and Lord, 2007), research was conducted on the effect of low means and small sample sizes in traffic safety research, leading to the conclusion that the Poisson-lognormal models often achieve better results than Poisson-gamma models.

All the techniques discussed above concentrate all characteristics of a location in one point, hereby ignoring the existing geographical relationship between the different locations. However, it seems only logical that the structure of the underlying road network can play an important role in determining hazardous locations. E.g. crossroads, on and off ramps on a highway may have direct implications on the number of crashes on a location nearby. Next to that, there is a recent trend to examine road segments instead of dangerous locations, because of the obvious spatial interaction between accident locations that are close to one another. Spatial techniques allow accounting for this. These methods exist in one dimension and in two dimensions, but often they are not suited to be used alongside a network. This paper displays the use of a local indicator of spatial association (LISA) to identify hot spots on motorways, hereby taking into account the structure of the road network. However, as indicated above, due to the nature of road crashes (sparseness), the normal use of the indicator has serious flaws and adaptations are required to apply it in this context. Section 2 gives an overview of LISA’s in general together with the required adaptations and the changing distributional properties when applied in a traffic safety context. It also denotes the implications of changing some factors. Section three gives a description of the data, together with the results on motorways in Flanders, Belgium. Section 4 ends with conclusions and some ideas for future research.

## 2. Method

Recently, there is a tendency to use spatial data analysis techniques next to statistical (Bayesian) regression models in HSA (Flahaut *et al.*, 2003). This enables to account for the spatial character of a location. In this paper, a spatial autocorrelation index is used. It aims at evaluating the level of spatial (inter-)dependence between the values  $x_i$  of a variable  $X$  under investigation, among spatially located data. In other words, spatial autocorrelation is a spatial arrangement where spatial independence has been violated (Levine, 2000). If the idea of temporal autocorrelation is extended, then a simple representation of spatial dependence can be formulated as follows:

$$x_i = \rho \sum_j w_{ij} x_j + u_i$$

where  $\rho$  measures the spatial autocorrelation between the  $x_i$  ‘s,  $w_{ij}$  are the weights that represent the proximity between location  $i$  and  $j$  (they are often denoted in terms of distance,  $w_{ij} = d_{ij}^{-p}$  for some power  $p$ ) and  $u_i$  are independent and identically distributed error terms with mean zero and variance  $\sigma^2$ . However, in contrast to temporal autocorrelation,

the spatial neighborhood is multidirectional, making it more complex and leading to specific indices for spatial autocorrelation. Specifically, spatial correlation analysis assesses the extent to which the value of the variable  $X$  at a certain location  $i$  is related to the values of that variable at contiguous locations. This assessment involves analyzing the degree to which the value of a variable for each location co-varies with values of that variable at nearby locations. When the level of co-variation is higher than expected, neighboring locations have similar values (both high or both low) and autocorrelation is positive. Opposite, when the level of co-variation is lower than expected, high values of the variable are contiguous to low values and the autocorrelation is negative. The lack of significant positive or negative co-variation suggests absence of spatial autocorrelation (Flahaut *et al.*, 2003).

## 2.1 Background on Local Indicators of Spatial Association (LISA)

Global measures of spatial autocorrelation exist for several decades and mainly stem from the work of Moran (1948). Moran's  $I$  is most often used and its usefulness for transport fluxes and traffic accident analysis has been thoroughly discussed in the literature (Black, 1992; Black and Thomas, 1998). Next to the global measure, that gives an idea about the study area as a whole, it may also be interesting to limit the analysis to a smaller part of it. It might happen that smaller parts of the study area show spatial autocorrelation, but that it has not been picked up by the global measure. Though, also when there is global autocorrelation present, the local indices can be useful to point at the contribution of smaller parts of the investigated area.

The use of these local indices is more recent (Ord and Getis, 1995; Anselin, 1995; Flahaut *et al.*, 2003). Each basic spatial unit  $i$  is now characterized by one value of the index, it denotes the individual contribution of the location in the global autocorrelation measured on all  $n$  locations.

These local indices are considered to be *Local Indicators of Spatial Association* (LISA's) if they meet two conditions:

- It needs to measure the extent of spatial autocorrelation around a particular observation, and this for each observation in the data set;
- The sum of the local indices needs to be proportional to the global measure of spatial association.

Following some notations used in Goodchild (1986), we have:

- $s_{ij}$  representing the similarity of point  $i$ 's and point  $j$ 's attributes,
- $w_{ij}$  representing the proximity of point  $i$ 's and point  $j$ 's locations, with  $w_{ii} = 0$  for all points,
- $x_i$  representing the value of interest of variable  $X$  for point  $i$ , and
- $n$  representing the total number of points.

The equation for a global spatial association coefficient (SAC) then takes the general form:

$$SAC \approx \frac{\sum_{i=1}^n \sum_{j=1}^n s_{ij} w_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}$$

while the corresponding LISA may look as follows:

$$SAC_i \approx \sum_{j=1}^n s_{ij} w_{ij} .$$

## 2.2 Local Moran Index

### 2.2.1 General Use

The global version of Moran's  $I$  was first discussed in Moran (1948), however, in this paper its local version will be applied. The LISA version of Moran's  $I$  that satisfies the two requirements as stated in 2.1 can be written down as follows:

$$I_i = \frac{n}{(n-1)S^2} (x_i - \bar{x}) \sum_{j=1}^n w_{ij} (x_j - \bar{x})$$

with  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ , the variance of the observed values,  $\bar{x}$  the average value of  $X$  and the rest of the notation

is as was denoted in Subsection 2.1. Note that the proximity measure  $w_{ij}$  (i.e. the distance between location  $i$  and location  $j$ ) is not determined by means of a bird's-eye view. Crashes occur on a road network, and it may happen that locations are very close to each other in space, though, via the network, they cannot be reached easily (e.g. because one of them is located in a one-way street). Therefore, the distance traveled alongside the road network will be considered for  $w_{ij}$ . This is the first extension that is used in this paper in comparison to the 'normal' use of LISA's.

A nice property of Moran's  $I$  is the fact that it looks relative with respect to an average value. Because of computational issues, it is often impossible to compute the index for the study area as a whole in one time, and it needs to be split into smaller parts (as will be the case in Section 3). By plugging in the average of the entire study area as  $\bar{x}$  (instead of just the average of the smaller part), all results can easily be combined. So,  $\bar{x}$  might serve as a reference value for the study area under investigation (see also 2.2.2).

Anselin (1995) derives the mean and variance of  $I_i$  under the randomization assumption for a continuous  $X$ -variable. The expected value of  $I_i$  is, for example (Schabenberger and Gotway, 2005):

$$E[I_i] = \frac{-1}{n-1} \sum_{j=1}^n w_{ij} .$$

The exact distributional properties of the autocorrelation statistics are elusive, even in the case of a Gaussian random field. The Gaussian approximation tends to work well, but the same cannot necessarily be said for the local statistics (Schabenberger and Gotway, 2005). Anselin (1995) recommends randomization inference, e.g. by using a permutation approach. However, Besag and Newell (1991) and Waller and Gotway (2004) note that when the data have heterogeneous means or variances, a common occurrence with count data, such as crashes, the randomization assumption is inappropriate. Instead, they recommend the use of Monte Carlo testing.

### 2.2.2 Adaptations

Four adaptations to previous uses of local Moran's  $I$  in traffic safety have been proposed here. First of all, it is important to use the index in a correct way. One needs to account for zero observations as well (instead of only taking into account locations with at least one crash, see e.g. Flahaut *et al.*, 2003). Otherwise the average value would be determined in a wrong way, it would clearly be overestimated. Moreover, all locations with crashes would be judged to be of a too high importance.

Second, as already indicated in the previous paragraph, any reference value can be used for  $\bar{x}$ . If one wants to compare different countries to each other, a global average can be computed and in this way all countries can be compared to that global average. From a traffic safety point of view, it might be interesting to compare e.g. to the average for that type of road, the average of a region or a country.

Third, within the field of application, this local measure of spatial association can be regarded as being a traffic safety index, since for each basic spatial unit (BSU) of road the local Moran index can be regarded as a measure of association between the BSU under study and the neighboring BSU's that are similar to the one under study concerning the number of crashes. A negative value of the local autocorrelation index at location  $i$  indicates opposite values of the variable at location  $i$  compared to its neighboring locations. A positive value, on the contrary, points at similar values at location  $i$  and its neighborhood. This means that location  $i$  and its weighted neighborhood can both have values above the average value or both can have values below the average. In the application area of traffic safety, however, one is only interested in locations that have:

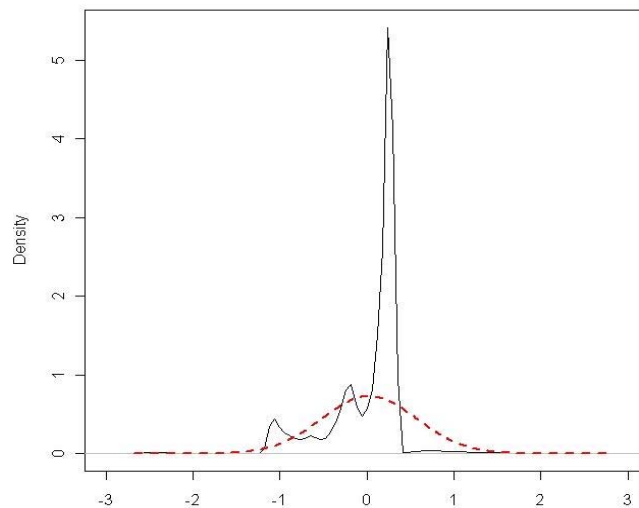
1. a high number of crashes in regard to the total average number of crashes (i.e.  $x_i - \bar{x} > 0$ ),
2. and where the neighborhood also shows more crashes than was expected on average (i.e.  $\sum_j w_{ij} (x_j - \bar{x}) > 0$ ).

One might argue that it is also important to look at locations with a high number of crashes at location  $i$  and a very low number in the surrounding area (i.e. a spike). In this case, very negative values of Moran's  $I$  would occur. However, although conceptually appealing, this gives very contradictory effects as will be illustrated by the following example.

Suppose that the global average over a certain area equals one crash ( $\bar{x}=1$ ). Then, if 5 crashes occurred at location  $i$  and none in its surrounding, this would lead to a negative value of Moran's I and possibly a significant negative autocorrelation. However, adding one crash to every surrounding point of location  $i$ , hence making the surrounding area more hazardous, would lead to a Moran's I of zero, indicating no significant autocorrelation. This would mean that a more dangerous location has a less significant Moran's I when compared to a more 'safe' location. This is really counterintuitive, so therefore it was opted to look only at points where a high number of crashes is contiguous with high values in the neighborhood (the location and its surrounding area reinforce each other in a positive way).

Finally, since the distributional properties of Moran's I are intangible, a Monte Carlo approach was opted for to arrive at cut-off values for the local Moran's I above which the location can be considered to be a hot spot.

To this end, the total number of crashes for the study area will be spread randomly over the total available locations. Note that locations are allowed to have more than one crash, otherwise, high concentrations of crashes cannot be determined. This simulation was repeated 500 times to end up with an approximate distribution of the Moran index for the particular situation at hand. A real world example for 506 crashes at 3,252 locations is shown in Figure 1. It is obvious that a Gaussian approximation would not work well in these circumstances. The black solid curve indicates the simulated density for Moran's I, while the red dashed curve shows the Gaussian approximation with the mean and variance as they are expected to be under randomization.



**Figure 1:** Simulated density of local Moran's I.

To determine the hot spots, we decided to look only at locations that show a positive reinforcement with their neighbors in the calculation of the local autocorrelation index. So, these values were filtered out and from their distribution, the 95% percentile ( $P_{95}$ ) was calculated. This value was utilized as the cut-off value to determine an accident hot spot in the study area. Therefore, the true Moran index values were calculated at each location. Next, the local Moran indices with similar high values between the location under study and its contiguous locations are selected and each of these index values was compared to the cut-off value. If the observed local value exceeds the cut-off value (i.e. if  $I_i > P_{95}$ ), then location  $i$  is considered to be hazardous.

### 3. Analysis and Results

#### 3.1 The Data

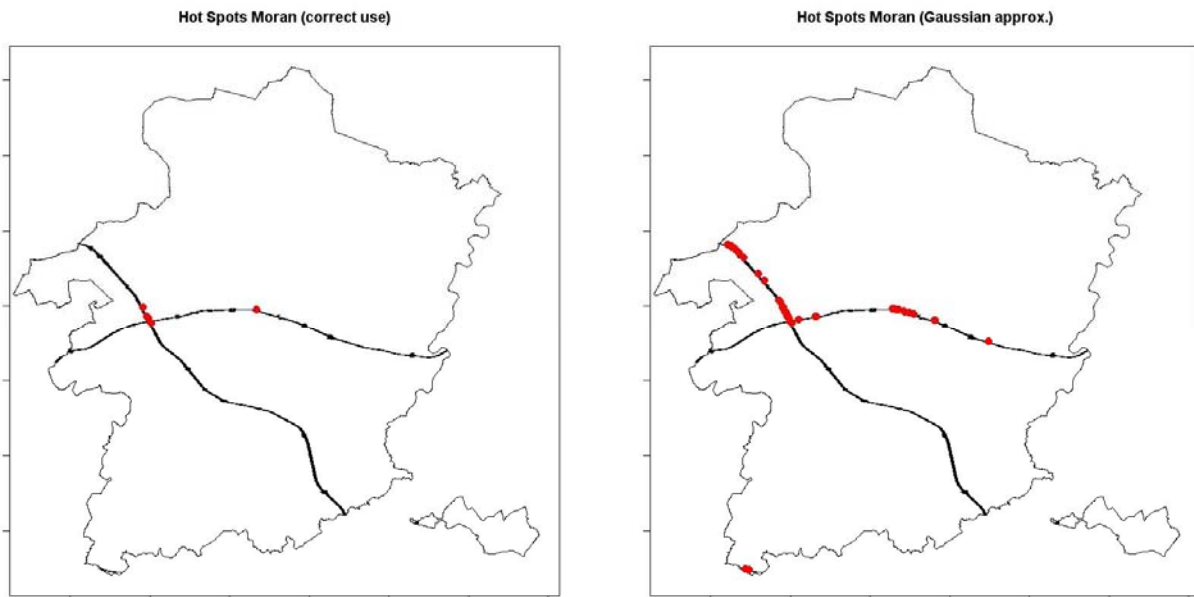
The analyses were carried out on the province of Limburg in Flanders, the upper, Dutch-speaking, part of Belgium. The basic spatial unit is defined to be about 100m. Crashes occurring at motorways are assigned to the closest hectometer pole, so they are regarded as BSU. The weights that were used are the inverse of the squared distance, where the distance was determined from one BSU to the next one on the network. The number of neighbors is also distance based. For each BSU, BSU's within a 1km range of the BSU under investigation are included as neighbors. So each

point, not located near the end of any motorway, has approximately 20 neighboring points. Nearby the junction of both motorways (see Figure 2), it may happen that BSU's from the second motorway are within the predefined number of neighbors for a location at the first motorway. To account for them in a proper way, distances need to be network-based. In the city environment, this would be even more important since there are much more small roads within the neighborhood of each other.

Because we want to compare the results of Limburg with other provinces in Flanders, the number of crashes for Flanders was set as a reference value in both analyses. This means that for Moran's I, the average number of crashes in Flanders was used for  $\bar{x}$ . Limburg has 3,252 hectometer poles alongside its two motorways (E313 and E314). 506 crashes occurred on Limburg highways from 2004-2006.

### 3.2 Analysis

As stated above, since crashes form a Poisson process instead of a Gaussian process and because of the sparseness (most locations have a zero crash count), the distribution of the local autocorrelation statistics proves to be far from Gaussian. Moreover, count data often suffer from the problem of overdispersion and means and variances tend to be heterogeneous, so therefore, a Monte Carlo approach is recommended to derive the distribution of the local autocorrelation statistic. The 506 crashes are spread randomly over the 3,252 hectometer poles to determine the distribution of the local version of Moran's I. This density is illustrated in Figure 1. Since we decided to look only at the locations that show a positive reinforcement with their contiguous location in the calculation of the local autocorrelation index, these values need to be filtered out of the 500 x 3,252 values. From these values, the 95% percentile was calculated and this value, i.e.  $P_{95} = 4.32$  was utilized as cut-off value to determine which location is a hot spot and which not. Only 5 of the 3,252 locations proved to be hot spots on motorways in Limburg. For reasons of comparison, the standard used Gaussian approximation was also applied to investigate the difference in results. Using also the 95% percentile (of the Gaussian distribution!) as cut-off value, now 46 locations proved to be hot spots according to this method. The previous 5 are part of them, however, about 8/9 (!) of the points are falsely identified as belonging to the 5% most extreme Moran index values. From a policy point of view, this might lead to a wrong allocation of funds to ameliorate traffic safety and thus it indisputably shows the importance of using the right distributions. Figure 2 shows the resulting hot spots. The figure shows the province of Limburg and its two motorways. The hot spots are indicated in red, while the underlying road network is drawn in black. The left figure shows the correct application, while the right figure shows the results based on the Gaussian approximation.



**Figure 2:** Hot spots for the local Moran index

## 4. Conclusions and Discussion

The aim of this paper was to apply a local indicator of spatial association, more in particular Moran's I, to identify hazardous locations on motorways. First of all, it needs to be acknowledged that crashes occur on a network, and this should be accounted for by using the correct network-based distances between locations under study.

Because of the nature of road crashes, locations with zero counts are very frequent. Moreover, accident data in general stem from a Poisson random process, rather than a Gaussian random process, so the normal use of the indicator seemed very elusive. To account for these characteristics, a simulation procedure was set up to arrive at the distribution of the local indicator, so as to determine the 5% most extreme observations. The motorway network of the province of Limburg in Belgium has been considered as the application setting. To construct the distribution of the LISA, a Monte Carlo simulation experiment was set up where the reported number of crashes was spread randomly over the population of possible locations to arrive at the distribution of the local indicator. Sites that showed a local index above the 95% cut-off value of the density are then regarded as hot spots.

For comparison purposes, the same analysis was carried out using the Gaussian approximation for Moran's I instead of the simulated distribution. Now more than nine times as much locations were defined as hot spots, including the ones obtained by a correct use. This illustrates that blindly using the Gaussian approximation certainly is not an option, and one absolutely needs to take into account the nature of the data under study. For policy makers, this is a very relevant result, since they usually do not have access to an unlimited budget to treat hot spots. To allocate their funds in the best possible way, it is important to know which locations are true hot spots.

A next step to be taken now is to investigate how these hot spots can be combined into hot zones. A possible way forward has been suggested by Loo (2008).

Applying this and other (spatial) techniques (such as network-based K-function) to identify hot spots on other road types (city environment, local roads) and to compare the different results is certainly an important avenue for future research. Furthermore, it also seems of importance to investigate the impact of the number of neighbors/the weight function on these different settings in order to come up with some rules of thumb to be used for analyses in the future.

## Acknowledgements

The authors would like to thank the Strategic Analysis Department of the Federal Police for providing the data used in this study.

## References

- Anselin, L. (1995), "Local Indicators of Spatial Association-LISA," *Geographical Analysis*, 27(2), 93-115.
- Besag, J. and Newell, J. (1991), "The Detection of Clusters in Rare Diseases," *Journal of the Royal Statistical Society, Ser. A*, 154, 327-333.
- Black, W. R. (1992) "Network Autocorrelation in Transport Network and Flow Systems," *Geographical. Analysis*, 24(3), 207-222.
- Black, W. R. and Thomas, I. (1998), "Accidents on Belgium's Motorways: a Network Autocorrelation Analysis," *Journal of Transport Geography*, 6(1), 23-31.
- Brijs, T., Van den Bossche, F., Wets, G. and Karlis, D. (2006), "A Model for Identifying and Ranking Dangerous Accident Locations: a Case Study in Flanders," *Statistica Neerlandica*, 60(4), 457-476.
- Brijs, T., Karlis, D., Van den Bossche, F. and Wets, G. (2007), "A Bayesian Model for Ranking Hazardous Road Sites," *Journal of the Royal Statistical Society, Ser. A*, 170, 1-17.
- Cheng, W. and Washington, S. P. (2005), "Experimental Evaluation of Hotspot Identification Methods," *Accident Analysis and Prevention*, 37, 870-881.
- Flahaut, B., Mouchart, M., San Martin, E. and Thomas, I. (2003), "The Local Spatial Autocorrelation and the Kernel Method for Identifying Black Zones - A Comparative Approach," *Accident Analysis and Prevention*, 35, 991-1004.
- Geurts, K., Thomas, I. and Wets, G. (2005), "Understanding Spatial Concentrations of Road Accidents using Frequent Itemsets," *Accident Analysis and Prevention*, 37(4), 787-799.

- Goodchild, M.F. (1986), *Spatial Autocorrelation*, CATMOG 47, Norwich: Geobooks, University of East Angolia.
- Hauer, E. (1996), "Identification of Sites With Promise," in *Transportation Research Record: Journal of the Transportation Research Board*, No. 1542, Transportation Research Board of the National Academies, Washington D.C., pp. 54-60.
- Hauer, E. (1997), *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*, Elsevier Science Ltd, Oxford.
- Hauer, E., Kononov, J., Allery, B. and Griffith, M. S. (2003a), "Screening the Road Network for Sites With Promise," in *Transportation Research Record: Journal of the Transportation Research Board*, No. 1784, Transportation Research Board of the National Academies, Washington D.C., pp. 27-32.
- Hauer, E., Douglas, W. H., Council, F. M. and Griffith, M. S. (2003b), "Estimating Safety by the Empirical Bayes Method: A Tutorial," in *Transportation Research Record: Journal of the Transportation Research Board*, No. 1784, Transportation Research Board of the National Academies, Washington D.C., pp. 126-131.
- International Traffic Safety Data and Analysis Group (IRTAD) (2008), "Selected Risk Values for the Year 2006," <http://cemt.org/IRTAD/IRTADPublic/we2.html>. Accessed October 27, 2008.
- Jianming, M. and Kockelman, K. M. (2006), "Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity," in *Electronic proceedings of the 85<sup>th</sup> Annual Meeting of the Transportation Research Board*, Washington, D.C..
- Kar, K. and Datta, T. K. (2004), "Development of a Safety Resource Allocation Model in Michigan," in *Transportation Research Record: Journal of the Transportation Research Board*, No. 1853, Transportation Research Board of the National Academies, Washington D.C., pp. 64-71.
- Levine, N. (2000), *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations*, vol. 1.1. Ned Levine and Associates/National Institute of Justice, Annandale, VA/Washington, D.C..
- Loo, B. P. Y. (2008), "The Identification of Hazardous Road Locations: A Comparison of Blacksite and Hot Zone Methodologies in Hong Kong". Paper forthcoming in *Journal of Sustainable Transportation*.
- Lord, D., Washington, S. P. and Ivan, J. N. (2005), "Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory," *Accident Analysis and Prevention*, 37, 35-46.
- Lord, D., Washington, S.P. and Ivan, J. N. (2007), "Further Notes on the Application of Zero-Inflated Models in Highway Safety," *Accident Analysis and Prevention*, 39, 53-57.
- Miranda-Moreno, L. F., Labbe, A. and Fu, L. (2007), "Bayesian Multiple Testing Procedures for Hotspot Identification," *Accident Analysis and Prevention*, 39(6), 1192-1201.
- Moran, P. (1948) "The Interpretation of Statistical Maps," *Journal of the Royal Statistical Society*, Ser. B, 10, 243-251.
- Ord, J. K. and Getis, A. (1995), "Local Spatial Autocorrelation Statistics: Distributional Issues and Applications," *Geographical Analysis*, 27, 286-306.
- Park, E. S. and Lord, D. (2007), "Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity," in *Electronic proceedings of the 86<sup>th</sup> Annual Meeting of the Transportation Research Board*, Washington, D.C..
- Schabenberger, O. and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Chapman and Hall/CRC Press.
- Staten-Generaal van de Verkeersveiligheid – States General of Traffic Safety (2007), *Verslag van de Federale Commissie voor de Verkeersveiligheid*, 38p, (in Dutch).
- Vistisen D. (2002), "Models and Methods for Hotspot Safety Work," *Phd Dissertation*, Technical University of Denmark.
- Waller, L. A. and Gotway, C. A. (2004), *Applied Spatial Statistics for Public Health Data*, John Wiley and Sons.