International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 2 Issue 9 ǁ September. 2014 ǁ PP.35-40

Risk governance for traffic accidents by Geostatistical Analyst methods

Ismail Bulent Gundogdu
Department of Geomatics, Engineering Faculty, Selcuk University.

Abstract : *Geographical Information Systems (GIS) are indispensable tool for administrating big datasets based on location of measured point.*

The values related to space may vary with both time and location. GIS-supported Geostatistical Analyst (GA) can evaluate datasets by analysing the locations of points.

Maps produced using probability and prediction methods must be the base products for city planning. This study develops methods to obtain maps to determine traffic hot zones in Konya, Turkey, by applying GA supported by GIS.

By applying GA, this study differs from previous studies which have determined the hot spots using linear analysis. In this study, unlike preceding studies, the aim is to determine new safe routes and zones with the help of GA.

Another, different aim is to map and determine graduated hot or safe zones using number of mortalities criterion (AC1), number of injured people criterion (AC2), number of accidents with damage only criterion (AC3), and total number of accidents criterion (AC4).

Keywords*: Geostatistical Analyst; GIS; traffic accidents; probability maps; prediction maps; kriging*

I. Introduction

In this paper, geostatistical approaches have been used to study traffic accident datasets in a case study of the city of Konya. The paper attempts to examine the characteristics of the spatial distribution of traffic accident data according to the accident criteria (AC) using kriging methods and to analyse their implications for junction spatial structure. The paper is organised as follows: after the next section on the data and methodology, I set up and interpret the accident mortality numbers, numbers of injured, numbers of accidents with damage only, and the total number of distribution models. Then I analyse the implications of traffic accident data models for urban spatial structure. The last section is the conclusion.

A report from the World Health Organization (WHO) and the World Bank (WB) (2004) on road traffic accidents and injuries estimated 1.2 million people are killed in road crashes each year and as many as 50 million are injured worldwide (Gundogdu, 2010). According to estimations of the WB traffic accidents will be the third most frequent reason of deaths in 2020.

The Turkish Statistical Institute (TUIK) facts reveal that 665 618 traffic accidents, 2629 of which resulted in mortality (3393 deaths), 77 644 of which resulted in injuries (135 441 injuries) and 585 345 of which resulted in economical damage, occurred in Turkey in last seven years.

The most specified analysis and plotting can be carried out using GIS nowadays.

Large number of recent studies such as (Cho 2003; Hansen and Lauritsen, 2010; Hill *et al*., 2011; Parasannakumar *et al*., 2011; Gu *et al*., 2013; Yu and Abdel-Aty, 2013) using the techniques that have been developed over the last fifty years.

No studies of traffic accidents that applied GA were encountered. However, some studies discussed similar problems: risk reduction in urban planning (Wamsler, 2006), transportation planning, discussed in an editorial article (*Journal of Transportation Planning*), the relationship between noise pollution and traffic (Seto et al., 2007), and urban traffic flow (Stathopoulos and Dimitriou, 2008).

This study differs from others in that the mapping is carried out according to the different AC. So, in this phase, GA must be explained in detail.

II. Geostatistical Analyst

In this chapter, the connection with the basis of GA is explained to guide the application. Geostatistics is a branch of practical statistics. Geostatistics was developed by George Matheron of the Centre de Morophologie Mathematicque in Fontainebleau in France.

Evaluation of the traffic accident dataset according to different criteria is also not possible with known statistical methods because the effects of sampling location are not taken into consideration in the calculations of variance and standard deviation for any variable. In GA, weights do not depend on the distances measured or predicted between two points.

According to Boogaart and Schaeben (2002) kriging is defined as BLUP (best linear unbiased predictor) or BLUE (best linear unbiased estimator). The method is named after the mining engineer D.G. Krige who first used it. (Trangmar et al, 1985) described that Kriging is an unbiased estimator of the spatial variation of unsampled points using the attributes of a variogram.

Calculating the variance is the most important and distinctive peculiarity of kriging in comparison with other techniques. The value is the criterion for the reliability of the estimated value. If the calculated variance is smaller than the variance of certain values, the estimated value is reliable for the unsampled point or area. (Pearce *et al.*, 2009; Laguardia, 2011; Pan and Zhu, 2011) state that kriging is better than other interpolation techniques using obtained geostatistical results.

Geostatistical Analyst provides many tools to help determine the parameters to be used and the defaults are also provided so that a surface can be created quickly. Kriging is a moderately quick interpolator that can be exact or smoothed depending on the measurement error model.
The ordinary Kriging formula is generally given by:

$$Z^*(u) = \sum_{a=1}^{n(u)} \lambda_a(u) Z(u_a) + \left[1 - \sum_{a=1}^{n(u)} \lambda_a(u)\right] m$$

where, $Z^*(u)$ is the ordinary Kriging estimate at spatial location u, $n(u)$ is the number of the data used at the known locations, $Z(u_a)$ are the n measured data at locations u_a located close to u, m is mean of distribution, $\lambda_a(u) =$ weights for location u_a computed from the spatial covariance matrix.
Ordinary Kriging assumes a constant but unknown mean and estimates the mean value as a constant in the searching neighbourhood (Kumar et al., 2007).

The variogram is the function which characterises the dependence of variables between two different points in space. According to Çetin and Tülücü (1998), in contrast to classical statistical methods, semivariogram models are interested in all data points whether regularly distributed or not with respect to both time and location.

III. Materials and method

Konya is almost exactly in the middle of Turkey, it can be designated as a junction point where all territories of the country are linked to each other. The city is not only the largest province of Turkey in the context of territorial size but also has the longest road network, with a 2966 km state road in a country with a total of just 62344 km of roads.

Traffic accident data

Accident records for ten years have been obtained from Konya Police Department and input into the database using MS Access to set up the GIS-based application. To define junctions, digitisation and analyses were performed using maps with a scale of 1/25 000.

Figure 1 shows 103 digitised junction points on the roads. In this process, besides all AC, dates, times, and types of vehicle data were recorded in the main dataset. In this period, a total of 2662 accidents occurred; there were 17 accidents with mortalities, 361 accidents with injuries, and 1903 accidents with only damage.

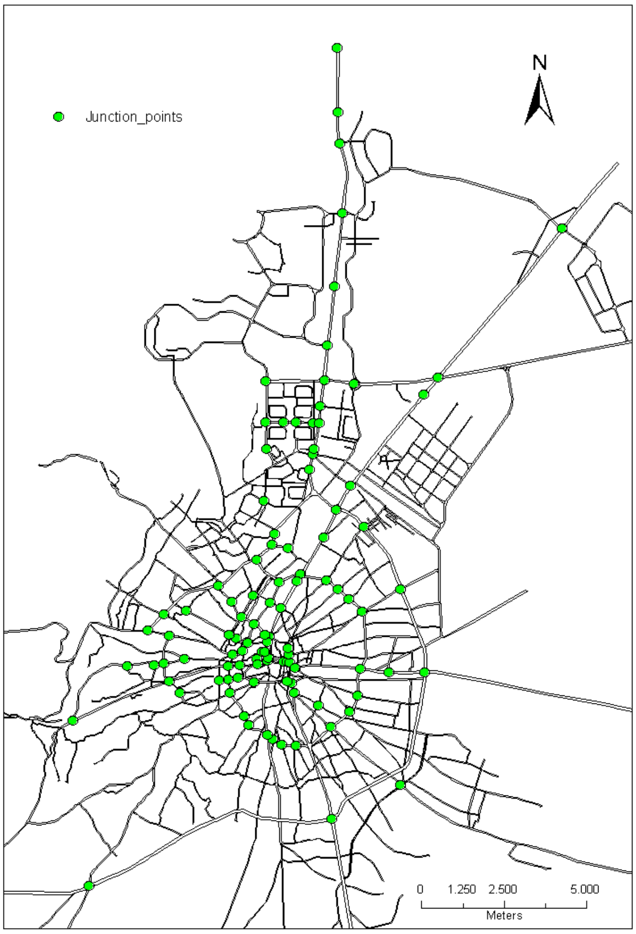

Figure 1: 103 junction points in the city centre.

IV. Methodology

First of all, the characteristics of all data must be evaluated. As mentioned above, in this phase all analyses were carried out, including not only the number of total accidents, but also other AC separately. Table 1 shows the histogram values for all AC, and Figure 2 shows a QQPlot graphic for AC4 as an example.

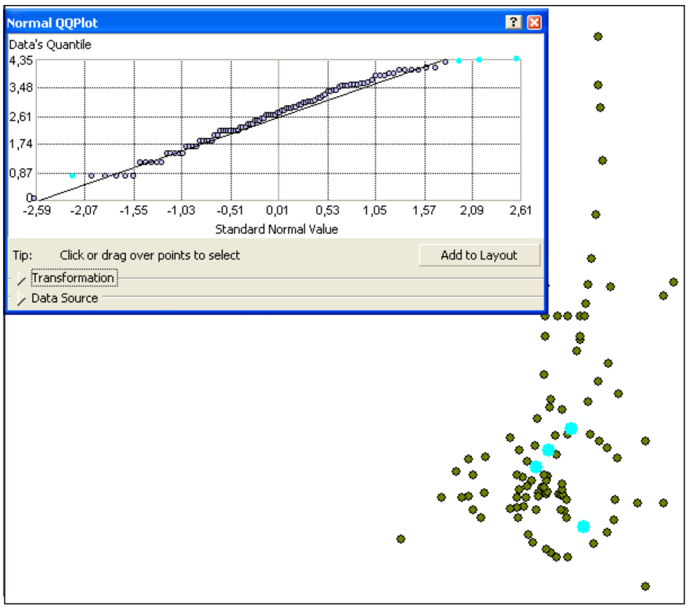

Figure 2: Example of a QQPlot graphic for AC4.

Table 1: Histogram values calculated by AC4.

AC	Min	Max	Mean	Std. Dev.	Skewness	Kurtosis	1st Quantile	Median	3rd Quantile
AC1	0	1	0.029	0.1689	5.6	32.363	0	0	0
AC2	0	22	3.0485	3.4052	2.849	14.249	1	2	4
AC3	0	67	17	16.111	1.244	3.779	4.25	11	25.5
AC4	1	76	20.078	18.123	1.2805	3.9217	6.25	13	30.25

If the mean and median values are close, it can be said that the data have a normal distribution. A histogram shows whether distributions of data are symmetrical or not. Symmetrical data can be realised with a skewness value close to zero. According to the values in Table 1, our data are not normally distributed or exactly symmetrical.

Two effects may influence the results. One of them is the global trend and the other is anisotropy. When a study is carried out on a two dimensional surface, sometimes the semivariogram and covariance functions must be investigated not only according to distance but also according to direction. This is called anisotropy. Anisotropy is determined by random errors, and it is different from a global trend.

In addition, unlike the circular isotropic model, an ellipse must be chosen in the searching neighbourhood to determine the neighbourhood number and location. In interpolation, the major axis must have the same direction as the anisotropic direction. This is taken into consideration in the calculation of the values in Table 2.

According to Ahmadi and Sedghamiz (2007), the ratio of nugget to sill can indicate statistical autocorrelation. If the value is smaller than 0.25, there is strong autocorrelation; if the value is between 0.25 and 0.75 there is moderate spatial autocorrelation; otherwise there is no spatial autocorrelation. Table 2 shows all geostatistical values constituted with all AC. All values which are important to show AC and to plan for the future are in shown the table. To carry out mapping with the help of the Table2, the rules must first be considered primarily as below.

The best map can be produced with the trial and error method by changing parameters. Transformations are not applied but there is anisotropy.

Finally, before the mapping, the performance of values can be controlled by cross validation. Cross validation investigates which model has the best performance.

Figure 3 shows prediction maps with the numbers of (a) mortalities, (b) injured persons, (c) accidents with damage only, and (d) total accidents. Representations of maps illustrated using Equal Area techniques have generally been used for probability maps to clarify extreme values so far. Therefore, all maps are produced using this method.

Furthermore, probability maps can be produced in respect of threshold values that are exceeded or not exceeded. A related example is shown in Figure 3 (e) where the number is 10, which is the below the mean value for the total number of accidents.

The indicator prediction value shows the probability of exceeding a threshold value. The most important maps to interpret to prevent traffic accidents are of this type. Figure 3(f) was produced using covariance values. These maps illustrate the probability of exceeding 10.

Table 2: All calculated statistical values necessary for interpretations.

Accident criteria / Statistical values	Mortality Number (AC1)	Injured Number (AC2)	Accident with damage (AC3)	Total Number (AC4)
Trend	Arcsin.	-	-	Log
Anisotropy	✓	✓	✓	✓
Major range	20097.4	23584.3	927.791	958.977
Minor range	7538.64	8943.27	200.871	207.655
Nugget	0.017025	5.5656	0	0
Sill	0.017687	8.3892	347.7	310.86
Nugget sill ratio	0.96	0.66	0	0
Ordinary kriging results				
Mean	0.001294	0.1055	-0.4731	-0.3556
Root-Mean-Square	0.179	3.926	16.7	18.95
Average Standard Error	0.1458	2.658	16.13	18.1
Mean Standardized	0.01359	0.02456	-0.02577	-0.0153
Root-Mean-Square Standardized	1.209	1.345	1.026	1.032

For the correct interpretation of the values shown in the table, the following rules must be taken into consideration.

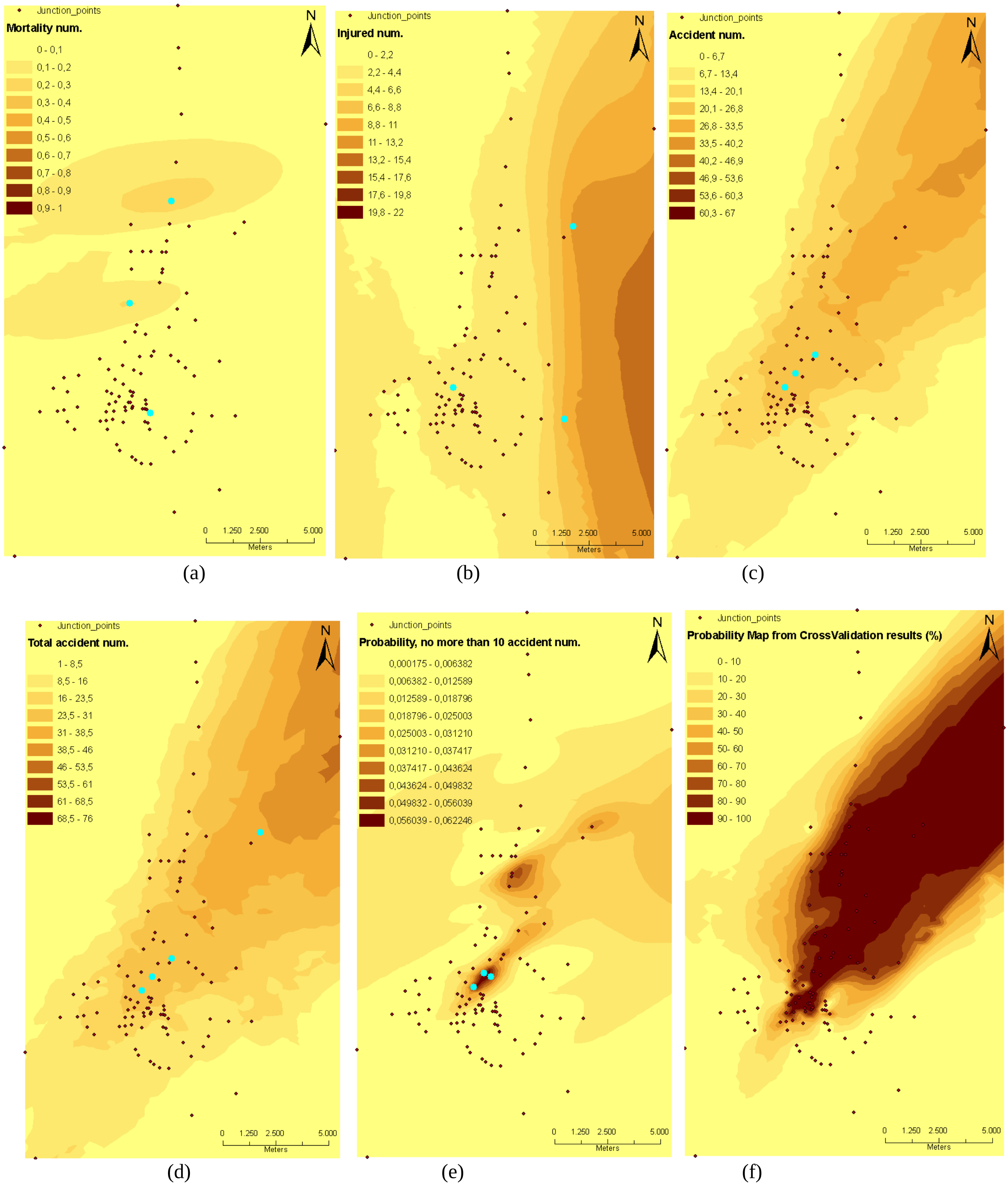

Figure 3: Result maps: (a) prediction map for number of mortalities, (b) prediction map for number of injured persons, (c) prediction map for number of accidents with damage only, (d) prediction map for total number of accidents, (e) probability map with no more than 10 accidents, (f) probability map using cross validation results.

V. Conclusions and future work

Deterministic and geostatistical methods are examined with two groups in GA. These methods are basically similar as data for nearby points are close.

On the other hand, when the mathematical and statistical methods are used together, geostatistical methods are not only used as interpolation methods but also give an opportunity for interpretation. The method also gives information about the reliability of predicted values. Therefore, deterministic applications were not performed in this study. This study has tried to understand the characteristics of traffic accidents. Nevertheless, in order to generalise, similar tests must be applied for more areas.

The study gives different results from others concerning GA applications for traffic accidents in the literature. In addition, it differs from other studies in that it uses not only the number of total accidents but also different traffic accident criteria.

According to the resulting values, accident data are not exactly normally distributed or symmetrical. This shows that the frequency and intensity of accidents are out of order, and some junctions are problematic.

When the QQPlot graphic is considered with all outlier AC values, there are problematic junctions: three junctions (6, 20, 71) arising from AC1, three junctions (25, 38, 52) arising from AC2, three junctions (30, 49, 51) arising from AC3, and four junctions (25, 30, 50, 51) arising from AC4. These problematic junctions are shown in light colour in Figures 3(a), (b), (c), and (d) separately.

Three junctions (30, 49, 51) which meet four criteria, of a total of nine problematic junctions, must be improved urgently.

When the trend graphic is observed, global trends are seen for all AC.

In the Semivariogram/Covariance Cloud, most of the data are clustered in the first part of the variogram. This distribution shows that the numbers of accidents are correlated with each other, and there are no, or less, location influences.

Anisotropy was proved for all AC in the study as on weather pollution maps. There are anisotropies for all AC except AC1. The reason is that this is the biggest artery lying in a northeast to southwest direction. The most straight and longest north artery with high traffic speed triggering the all AC. Precautions about reducing the speed must be put on the agenda urgently.

As a result, in the light of calculated values, result maps were produced with the aim of helping to create new city models. Maps which have statistical value and can be visualised will be a better help to city planners.

References

[1.] Ahmadi, S.H., Sedghamiz, A., 2007. Geostatistical analysis of spatial and temporal variations of groundwater level, Environ Monit Assess 129, pp:277-294

[2.] Boogaart, K.G., van der, Schaeben, H., 2002. Kriging of regionalized directions, axes and orientations I. Directions and axes, Mathematical Geology, Vol. 34, No. 5, 479-503

[3.] Cho, W.K.T., 2003. Contagion effects and ethnic contribution Networks. American journal of political science. 47 (2) pp.368-387

[4.] Gu, X., Sun, G., Li, G., Huang, X., Li, Y., Li, Q. 2013. Multiobjective optimization design for vehicle occupant restraint system under frontal impact. Structural & Multidisciplinary Optimization. 47-3, p465-477.

[5.] Gundogdu, I.B., 2010. Traffic management in Konya by Junction Risk Factor. Municipal Engineer. 163, pp:99-105.

[6.] Hansen, D., Lauritsen, J M., 2010. Identification of black spots for traffic injury in road intersections dependence of injury definition. Injury Prevention Vol. 16, pp:261-267

[7.] Hill, L., Rybar, J., Baird, S., Concha-Garcia, S., Coimbra, R., Patrick, K. 2011. Road safe seniors: Screening for age-related driving disorders in inpatient and outpatient settings. Journal of Safety Research.42-3, pp165-169.

[8.] Kumar, A., Maroju, S., and Bhat, A., 2007., Application of ArcGis geostatistical analyst for interpolating environmental data from observations. Wiley Interscience DOI 10.1002/eg. 10223

[9.] Laguardia, G., 2011 Representing the precipitation regime by means of Fourier series. International Journal of Climatology. 31-9, p1398-1407

[10.] Pan, F., Zhu, P., 2011. Design optimisation of vehicle roof structures: benefits of using multiple surrogates. International Journal of Crashworthiness, 16-1, pp85-95.

[11.] Pearce, JL; Rathbun, SL; Aguilar-Villalobos, M; Naeher, LP. 2009 Characterizing the spatiotemporal variability of PM2.5 in Cusco, Peru using kriging with external drift Atmospheric Environment, 43; 12; p2060-p2069

[12.] Prasannakumar ,V., Vijith, H., Charutha, R., Geetha, N. 2011. Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment In International Conference: Spatial Thinking and Geographic Information Sciences 2011, Procedia - Social and Behavioral Sciences. 21:317-325

[13.] Seto, E.Y.W., Holt, A., Rivard, T., Bhatia, R., 2007. Spatial distribution of traffic induced nnoise exposures in a US city: an analytical tool for assessing the health impacts of urban planning decisions. International Journal of Health Geographics. 6:24.

[14.] Stathopoulos, A., Dimitriou, L., 2008. Fuzzy modelling approach for combined forecasting of urban traffic flow. Computer-Aided Civil and Infrastructure Engineering. 23, 521-535.

[15.] Trangmar, B.B., Yost, R.J., and Wehara, G., 1985. Application of geostatistics to spatial studies of soil properties. Advances in Argonomy, Vol 38, 65-91

[16.] Wamsler, C., 2006. Main streaming risk reduction in urban planning and housing: a challenge for international aid organisations. Blackwell Publishing. USA.

[17.] World report on road traffic injury prevention: summary, http://www.who.int/world-health-day/2004/infomaterials/world report/en/summary _en_rev.pdf.

[18.] Yu, R., Abdel-Aty, M. 2013. Investigating the different characteristics of weekday and weekend crashes. Journal of Safety Research. Vol. 46, pp91-97