Available online at www.sciencedirect.com
ScienceDirect
Procedia Social and Behavioral Sciences 20 (2011) 786–795
Procedia
Social and Behavioral Sciences

14th EWGT & 26th MEC & 1st RH

An Entropy (Shannon) based Traffic Safety Level Determination Approach for Black Spots

Yetis Sazi Murat[a,*]

[a]*Pamukkale Üniversitesi, Mühendislik Fakültesi, İnsaat Müh. Bölümü, Kınıklı Kampüsü, 20070, Denizli, Türkiye*

Abstract

Black Spots are one of the crucial subjects in traffic safety. Determination of black spots and their safety levels would be useful for prevention of future traffic accidents. But it is not an easy task. Many parameters have considerable effects on the phenomenon. On the other hand, safety level determination has uncertainty. Therefore deterministic approaches are incapable in classification. This study deals with determination of black spots' safety levels using Shannon Entropy Approach considering accident types and effective factors on accident occurrence. Geometrical and physical conditions, traffic volumes, average speeds and average accident rates at around black spots are considered as effective factors on occurrence of accidents. Entropy values are calculated using these parameters. Safety levels are classified as five groups based on calculated entropy values. Traffic accident data for three years (2004-2006) are used in development and testing the model. Validation of entropy approach is tested by Chi-Square and truth value methods, and encouraging results are obtained.

Selection and/or peer-review under responsibility of the Organizing Committee.
Keywords: Black spot; traffic; safety level; determination; entropy.

1. Introduction

Most of researches about traffic safety in the literature are focused on black spots. The locations and the reasons of black spots have been investigated in many studies. Meuleners et al. (2008) investigated efficiency of the safety audit program applied in Western Australia. They determined that traffic signal control, geometrical design and safety program for pedestrians have 15% improvement on occurrence of accidents. Retting et al (2001) considered urban traffic accidents and black spots and made some recommendations for reducing number of accidents. Flahaut et al (2003) developed some models to define locations of black spots using autocorrelation index and Kernel methods. Results of the study show that both of the methods can be used for definition of black spots' locations. Geurts et al (2005) analyzed black spots and the accidents occurred around these locations. The reasons of accidents are classified considering frequency levels. It is resulted that frequency analysis can be used instead of conventional statistical methods. Cheng and Washington (2005) compared Ranking, Definition by Confidence Interval and

* Corresponding author. Tel:+90-258-2963357 ; fax: +90-258-2963460
E-mail address: : ysmurat@pau.edu.tr

Selection and/or peer-review under responsibility of the Organizing Committee
doi:10.1016/j.sbspro.2011.08.087

Empirical Bayesian approaches. They determined that results of the Empirical Bayesian approach are better than the others. Moreno et al. (2007) used two different approaches of Bayesian Method for black spot definition. Vorko and Jovic (2000) searched school students' accidents using entropy classification approach. They classified the accidents into four groups considering injury types and locations of accidents. Wang and Nihan (2004) dealt with vehicle-bicycles accidents occurred at signalized intersections and developed a risk estimation model for these types of accidents. The accidents are classified considering direction maneuvers of vehicles and a probability based method is used. Three negative binomial regression models are improved and maximum likelihood approach is used in estimation of models' parameters. It is stated that the negative binomial regression approach can be used instead of poisson regression approach. Saplıoğlu and Karaşahin (2006) regarded traffic accidents of Isparta city by Geographical Information Systems (GIS). Black spots in the city center are determined by GIS and an increasing trend in number of black spots by years is also ascertained. Detailed statistical analyses are reported in the study. Abdel-Aty and Pange (2007), considered collision data for individual and grouped accidents. Estimation of accident occurrence locations is concentrated in the research. Hawas (2007), developed an accident estimation model for urban road networks using fuzzy logic approach.

Although the previous studies have mainly focused on black spots, estimation of number of accidents and reasons of accidents, they did not deal with the safety level determination of black spots. On the other hand, use of entropy approach in safety level determination is not regarded. Determination of black spots safety level has an importance on developing counter-measures for these locations. Furthermore, investment priorities for these locations can easily be made considering safety levels. Therefore this study is employed. The main aim of the present study is to explore use of Shannon entropy approach in safety level determination of black spots.

2. Methodology

2.1. Conventional Approaches

Traffic safety planning includes determination of locations that have a risk about accident occurrence. The conventional methods used for this purpose are accident frequency method, accident ratio method, ratio-quality control method and accident risk index method (FHWA, 1991). The data required for the corresponding methods are given in Table 1.

Table 1 The risky location determination methods and required data

Data	Method			
	Accident Frequency	Accident Ratio	Ratio-quality control	Accident risk index
Accident summary	X	X	X	X
Traffic volume		X	X	X
Accident severity				X
Average number of accident	X	X	X	X
Statistical constants			X	
Other regional features				X
Roadway features				

As seen on Table 1, only the accident risk index method needs more information and data about accident location. The accident risk index method considers classification of problematic locations. In this method, parameters related to direct measurements (accident rate, number of accidents, accident severity) and indirect measurements (volume/capacity ratio, sight distance ratio, irregular manoeuvres, roadwork, driver expectations etc) are considered. Value of each parameter is calculated using pre-defined conversion curves. In addition to this process, weight coefficients are also pre-determined. Accident risk index is ascertained by multiplying parameter values and weights. The following equation is used for this calculation:

$$ARI = \frac{\sum_{i=1}^{n}\left[w_i (PV_i)\right]}{\sum_{i=1}^{n} w_i} \quad (1)$$

where:
ARI : Accident risk index
w_i : weight coefficient for parameter i
PV_i : value of parameter i

This method is not practical and objective. On the other hand, some measurements (such as indirect measurements) can not be obtained in an easy way. Additionally, calculation procedure has some deficiencies. The conversion process and assignment of weights can be subjective and include uncertainties. Although it is known that the roadway features has great importance on accident occurrence, it is not considered in this method. To remove these deficiencies an efficient and systematic approach is required. Therefore Shannon Entropy approach is preferred.

2.2. Shannon Entropy Approach

Definition of entropy is based on information theory. In information theory, entropy is a measure of the uncertainty associated with a random variable. In this concept, the term usually refers to the Shannon entropy, which quantifies the expected value of the information contained in a message, usually in units such as bits. Equivalently, the Shannon entropy is a measure of the average information content one is missing when one does not know the value of the random variable.

Shannon defined the entropy as expected value of alternative conditions for a variable using a mathematical expression. Using this definition and expression, entropy of a stochastic process can easily be determined if the probability of process known. Based on the definition by Shannon value of entropy function is always be positive. Because of many attributes, the entropy concept is accepted as an objective criterion that can be used in measuring information content of any statistical process. Four main entropy values (marginal, common, conditional and trans-information) are used in the method for information content.

Shannon is the first researcher who define marginal entropy; H(X). Marginal entropy is defined in the following equation.

$$H(X) = -K\sum_{i=1}^{N} p(x_i)\log p(x_i) \quad (2)$$

where:
x shows discrete random variable,
K shows a constant based on the unit used in entropy calculation,
N shows number of basic events that has p(xi) probabilities.

2.3. Entropy Calculations of Black Spots

Accident records for the years from 2004 to 2006 are considered and accidents are located on digital map of Denizli city using corresponding coordinates. The centers of black spots that have been determined by clustering approach are considered in entropy calculations (Murat et al, 2008). Figure 1. (a) shows the centers of black spots determined by clustering approach and Figure 1 (b) shows a detailed sample.

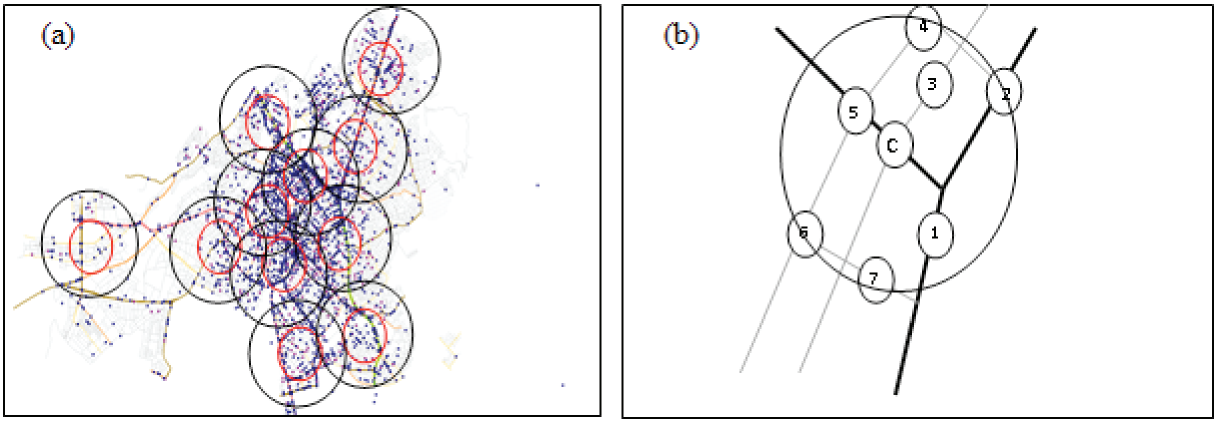

Figure 1 (a) The Black Spots' Centers Determined by Fuzzy Clustering Approach (b) Sample Detail

In figure 1 (b) a sample depicted. In this figure, C shows the center of black spots and the numbers from 1 to 7 show sample black spots around the center. In entropy calculations, geometrical and physical conditions, traffic volumes, average speeds and average accident rates at around black spots are considered. Entropy value of the center is determined considering entropy values of corresponding black spots around the center. As seen on Fig.1 (b), the entropy value of C is calculated using entropy values of 7 black spots. These 7 black spots represent the locations where accidents (more than one) have been occurred same addresses (i.e. different locations of same streets or boulevard).

2.3.1. Geometrical and Physical Condition of Black Spot

One of the main factors on traffic accidents are related to geometrical and physical conditions of accident location (ITE, 1993). Geometrical and physical condition includes many random variables that can be changed in different location. Therefore, this parameter is taken into consideration in entropy calculations. Traffic accident reports are used for determining geometrical and physical conditions of black spots. These reports are obtained from Local Police Department. The variables considered for this parameter are; illumination condition, lane line, walkway, shoulder, traffic sign, road work, obstacle on sight, direction (way), type of pavement, lane width, road surface characteristic, vertical curve, horizontal curve. The values of these variables given in accident reports are considered and sum of the values are calculated. The high value represents negative conditions. For instance, if there is a shoulder at the location, 1 point is assigned as the value, if it is not, 2 point is assigned. If there is a shoulder but it can not be available for use, 3 point is assigned for the location. The lane widths are measured by police department from accident locations at the time when the accident occurred. It is known that, the larger lane width increase the freedom of maneuver and also decrease the concentration of drivers. Therefore it is considered in calculations. Sample calculations are given in Table 2.

2.3.2. Traffic Volume

It is known that traffic accidents are increased by an increase of traffic volume. This situation is seen especially at the intersection that has geometrical problems. On the other hand, traffic volume is varied based on function of the road and the time period. It can be increased because of some special events (concert, basketball, football match etc). Based on these features, it can be assumed as random variable. Therefore traffic volume is considered in entropy calculations of black spots. Traffic volumes are observed at the black spots around the centers. On the other hand, results of previous observations are also used. Samples about observed traffic volumes at black spot centers are given in Table 3.

Table 2 Sample Calculations about Geometrical and Physical Conditions for Sevindik Section of 25th Street Black Spots

Year	Location	Illumination	Lane Line	Walkway	Shoulder	Traffic Sign	Roadwork	Obstacle on Sight	Direction (way)	Pavement type	Lane width	Road Surface	Horizontal curve	Vertical curve	Geometrical and Physical Condition
2004	S**	1	2	1	2	2	2	2	1	2	10.5	1	1	1	28.5
	S	1	1	2	2	1	2	2	1	2		1	1	1	17
	S	1	1	2	2	1	2	2	1	2	12	1	1	1	28
	S	1	2	2	2	2	2	2	1	2		1	1	1	19
	S	1	1	1	2	2	2	2	2	2	9	1	1	1	27
	S	1	1	1	2	2	2	2	1	2	14	2	1	2	33

Table 3 Samples about observed traffic volumes at black spot centers

Location	**Average Traffic Volume (veh/h)**
Ucgen Intersection	1677
Karayolları Intersection	1837
Cınar Square	1234
Kiremitci Intersection	652
Yeni Adliye Intersection	555
İstasyon Intersection	1809
Sevindik Intersection	1375
Emniyet Müd. Intersection	1564
Ulus Caddesi Intersection	572
Devlet Hastanesi Intersection	728
25. Cadde Intersection	955

2.3.3. Average Speed

Speed and careless driving are the main contributors of traffic accidents. Most of the accidents are occurred because of these factors. Drivers can increase or decrease their speeds based on roadway and traffic conditions of locations. Speed can be affected by traffic density, number of lane, lane width, sight distance and climatic conditions. Speed has an importance on accident severity. Hence average speed values around the black spot centers are regarded. The data are obtained from Local Police Department of Denizli city. Average speed values collected at different days by Police Department are given in Table 4.

Table 4 Measured speed values at Denizli city

Location	**Number of Vehicles**	**Average Speed (km/h)**
29 Ekim Boulevard	396	83
Acıpayam Road	132	88
Acıpayam Road	353	84
Ankara Road	660	87
Ankara Road	106	87
Yeniyol	374	87
Ankara Road	23	87
Acıpayam Road	115	89
Yeni askeri Road	47	88
Fatih Sultan Street	44	87
Bahçelievler Kontrol	25	61

2.3.4. Accident Rate

It is known that, accident rate and number of accidents are including randomness. Accident rate is another parameter used in safety level determination of black spots' centers. Accident rate is determined using the addresses where accident occurs. The sections where accidents are accumulated are ascertained. First, the number of accidents that occur in the same address is aggregated, and then the total numbers of accidents at the black spots are determined. The proportion of aggregated value to the total is considered as accident rate of black spot. Table 5 depicts accident rate samples for black spots and locations at around.

The parameters summarized above are used in entropy calculations. The entropy calculations are made for the sections around the black spots and average values are determined for the centers of black spots. Steps of calculations are defined in the following.

Step 1. Aggregation of the values of parameters.

$$T=GP+TV+AS+AR \tag{3}$$

Where:
GP: Geometrical and Physical Condition value
TV: Traffic Volume value
AS: Average Speed value
AR: Accident Rate value
T: Total value.

Step 2: Determination of probability of parameters.

$$PGP=GP/T \tag{4}$$

$$PTV=TV/T \tag{5}$$

$$PAS=AS/T \tag{6}$$

$$PAR=AR/T \tag{7}$$

Step 3: Calculation of information contents and entropy.

$$E=-[\log(PGP)*PGP+\log(PTV)*PTV+\log(PAS)*PAS+\log(PAR)*PAR] \tag{8}$$

Table 5 Accident Rate Samples around the Black Spots

Location	Black Spot No	Annual Accident Rate
Üçgen Intersection	U1	0.10
	U5	0.25
Karayolları Intersection	K3	0.50
	K7	0.40
Çınar Square	Ç2	0.35
	Ç6	0.09
Kiremitçi Mah. Intersection	Ki2	0.44
	Ki4	0.73
Yeni Adliye Intersection	Ya1	0.87
	Ya3	0.50
İstasyon Intersection	İ4	0.09
	İ7	0.16
Sevindik Intersection	S3	0.11
	S5	0.41
Emniyet Müd. Intersection	E4	0.21
	E6	0.32
Ulus Caddesi Intersection	U1	0.80
	U4	0.46
Devlet Hastanesi Intersection	H3	0.20
	H6	0.125
Cadde Intersection	25c2	0.52
	25c3	0.14

The entropy values are calculated using the steps defined above. Sample calculations and used values for some sections are given in Table 6. Entropy values of black spots are ascertained using the Shannon Entropy values of the sections (around the black spots) given in Table 6.

Total 120 sections around the black spots centers are assigned in the analysis. 73 of the data are used in training, 47 data are used in testing the approach. The numbers of classes are adjusted considering total number of data and statistical approaches. The following formula defined by Bayazit and Oguz (1985) is used for determination of class numbers.

$$2^k \geq n. \qquad (9)$$

where; k : number of class.

n: total number of data

Using Eq.8 and 73 data, 5 classes are determined for safety level definition. Definition of these classes are given as; definitely unsafe (1), unsafe (2), approximately safe (3), safe (4), definitely safe (5). Intervals of the classes are ascertained using Eq. 9.

$$RCI = (MAX-MIN)/NC \qquad (10)$$

where:

RCI means recommended class interval, MAX means the maximum value,

MIN represents the minimum value and the NC shows the number of class.

Using Eq.9 and the values determined as MAX=0.61, MIN=0.24 and NC=5, RCI=0.08 value is calculated. The safety levels and definition of the intervals are shown in Table 7.

Table 5 Sample Entropy Calculations for Black Spot Centers and Sections around

Black Spot Center	Accident Location	GP	TV (veh/hrs)	AS (km/h)	AR	Entropy Value
Üçgen Intersection	U1	30.7	1500	80	0.10	0.29
	U5	31.52	1677	80	0.25	0.27
Karayolları Intersection	K3	29.89	1111	78	0.50	0.35
	K7	30.17	1100	80	0.40	0.36
Çınar Square	Ç2	25	700	50	0.35	0.38
	Ç6	32.33	748	50	0.09	0.39
Kiremitçi Mah. Intersection	Ki2	30.5	652	50	0.44	0.42
	Ki4	32	500	65	0.73	0.55
Yeni Adliye Intersection	Ya1	23.5	500	75	0.87	0.55
	Ya3	36	525	72	0.50	0.57
İstasyon Intersection	İ4	24	1265	75	0.09	0.51
	İ7	34.57	1000	80	0.16	0.39
Sevindik Intersection	S3	27.33	800	101	0.11	0.47
	S5	29.58	1400	78	0.41	0.30
Emniyet Müd. Intersection	E4	31.6	975	85	0.21	0.40
	E6	30.95	800	80	0.32	0.45
Ulus Caddesi Intersection	U1	25.5	500	60	0.80	0.51
	U4	24	532	70	0.46	0.51
Devlet Hastanesi Intersection	H3	35	700	55	0.20	0.43
	H6	29.5	688	50	0.125	0.40
25. Cadde Intersection	25c2	30.68	711	70	0.52	0.46
	25c3	26.33	955	80	0.14	0.38

3. Validation of Entropy Calculations

To generalize entropy based classification approach in traffic safety researches, validation studies are employed. Truth value calculations and chi-square tests are used as validation studies. Safety levels (classes) of black spots are considered in truth value calculations. The safety levels of each black spot are determined regarding entropy value of the black spots. This value is compared to the safety level of the black spot centers specified by average entropy values. The truth value is assigned as 1 for the same result otherwise it is assigned as 0. Based on these calculations, 18 of 73 data provided different results for training data set and 9 of 47 data provided for testing data set. Therefore the truth value rates are calculated as 0.75 for training and 0.81 for testing data set.

Table 7 Safety Levels and safety Classes determined by Entropy Values

Entropy Interval	Definition	Safety Level (Class)
0,22<E<0.30	Definitely Unsafe	1
0.31<E<0.39	Unsafe	2
0.40<E<0.48	Approximately Safe	3
0.49<E<0.58	Safe	4
0.58<E<0.67	Definitely Safe	5

Chi-square test is used as the second validation study. In this test, the calculated entropy value of each black spot is compared to the expected entropy value. The expected entropy value is specified by calculating average entropy value of black spots around the center and it is assigned as the value for the corresponding center. Chi-square value for training data set is calculated as 0.67 and corresponding critical value is determined as 84 from chi-square table. Therefore, it is understood that calculated entropy values are compatible with the expected entropy values for training data set. Similar results are also achieved for testing data set. The calculated chi-square value is identified as 0.14 and the critical chi-square value is obtained as 62 for testing data. Results for sample testing data are given in Table 8. As a result, validation test results showed that Shannon Entropy approach can be used for safety level determination of black spots.

Table 8 Chi-Square Test Results for Sample Test Data Set

Black Spot Center	Accident Location	GP	TV (veh/hrs)	AS (km/h)	AR	Calculated Entropy Value (C)	Expected Entropy Value (E)	Chi-Square (C-E)^2/E
Üçgen	ut1	19.22	1100	87	0.09	0.34	0.32	0.0019
	ut2	8	925	85	0.06	0.33	0.32	0.0010
	ut3	18	1111	80	0.05	0.32	0.32	0.0001
Karayolları	kt1	21.14	1045	75	0.21	0.34	0.34	0.0000
	kt2	27.5	1100	77	0.33	0.35	0.34	0.0002
Çınar	ct1	16	800	55	0.14	0.33	0.33	0.0001
	ct2	28.5	775	47	0.07	0.36	0.33	0.0020
İstasyon	it1	26.29	1625	75	0.28	0.26	0.31	0.0076
	it2	24.35	1200	70	0.18	0.30	0.31	0.0000
Emniyet	et1	14.5	575	70	0.11	0.44	0.39	0.0080
	et2	18	1090	72	0.17	0.31	0.39	0.0158
Ulus	ut1	32.25	450	60	0.1	0.57	0.53	0.0020
	ut2	22.78	511	70	0.25	0.52	0.53	0.0005

4. Conclusions and Discussions

Black spots and safety levels are considered in this study. The safety levels of black spots and center of black spots are determined and classified using Shannon Entropy approach. Geometrical and physical conditions, traffic volume, average speed and accident rate are considered as effective parameters on safety level determination. Using these parameters, entropy values of black spots are calculated and average entropy value is specified for the corresponding centers. 5 levels (classes) are assigned based on calculated entropy value intervals. 11 black spots' centers are taken into consideration and it is resulted that 5 of 11 are in the 2^{nd} safety level, 3of 11 are in the 3^{rd} safety level and 3 of 11 are in the 4^{th} safety level. Validation search of the Shannon Entropy approach is made using truth value method and chi-square test. Both of the tests proved that Shannon Entropy approach can be used for safety level determination purposes. Similar results can be obtained for different cities or locations if the systematic approach defined in this research is used.

The results obtained can be used for investment priorities assignment in practical manner. For example, it should urgently be concentrated to the Ucgen, Karayollari, Cinar, İstasyon and Emniyet black spots' centers that are in the 2nd safety level (unsafe). The types of intersections can be changed considering geometrical conditions and traffic volumes. On the other hand, comparing to the centers which are in the 2nd safety level, there is no urgency for the Sevindik, Hastane and 25. Cadde black spots' centers that are in the 3rd safety level (approximately safe). It is concluded that the Kiremitci, Yeni Adliye and Ulus black spots' centers are in the 4th safety level (safe) and investments about traffic safety for these locations can be planned for a long period of time regarding transportation master plan and safety planning issues. These results can be useful for decision makers who are trying to find optimum investment assignment.

Acknowledgement

The data used in this study are obtained from Local Police Dept. of Denizli city for a research project (105G090) supported by TUBITAK. These supports are appreciated.

References

Abdel-Aty M.and Pange, A. (2007) Crash data analysis: Collective vs. individual crash level approach, *Journal of Safety Research* 38, 581–587

Bayazıt, M. and Oguz B. (1985) *Statistics for Engineers*, Birsen Publishing , Istanbul, 187 p.

Cheng,W. and Washington, S.P. (2005). Experimental evaluation of hotspots identification methods *Accident Analysis and Prevention*, 37, 870-881.

FHWA, (1982). *Safety Effectiveness of Highway Design Features*.

FHWA, (1991). *Effective Highway Accident Counter-measures*.

Flahaut, B., Mouchart, M., Martin, E.S. 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 item sets, *Accident Analysis and Prevention*, 37, 787-799.

Hawas Y. (2007). A fuzzy-based system for incident detection in urban street networks, *Transportation Research Part C*, 15, 69–95.

ITE (1993). *The Traffic Safety Toolbox-A primer on Traffic Safety*, Second Edition.

Meuleners, L.B., Hendric, D., Lee, A.H. and Legge, M. (2008). Effectiveness of the Black Spot Programs in Western Australia. *Accident Analysis and Prevention*, 40, 1211-1216.

Moreno, L.M., Labbe, A. and Fu, L. (2007). Bayesian multiple testing procedures for hotspot identification. *Accident Analysis and Prevention*, 39, 1192-1201.

Murat, Y.Ş., Fırat, M. and Altun, S. (2008). Analysis of Traffic Accidents using Fuzzy Clustering and Geographical Information Systems, *Proceedings of the 10th International Conference on Applications of Advanced Technologies in Transportation*, 27-31 May 2008, Athens, Greece.

Retting, R.A., Weinstein, H.B., Williams, A.F. and Preusser, D.F. (2001). A Simple method for identifying and correcting crash problems on urban arterial streets, *Accident Analysis and Prevention*, 33, 723-734.

Saplıoğlu, M., ve Karaşahin, M. (2006). *Urban Traffic Accident Analysis by using Geographic Information System*, Pamukkale Üniversitesi Mühendislik Fakültesi, Mühendislik Bilimleri Dergisi sayı,12, 3, 321-332.

Vorko, A and Jovic, F. (2000). Multiple attribute entropy classification of school-age injuries, *Accident Analysis and Prevention* 32, 445–454.

Wang, Y. and Nihan, N. (2004). Estimating the risk of collisions between bicycles and motor vehicles at signalized intersections, *Accident Analysis and Prevention* 36, 313–321.