Predicting and identifying traffic hot spots applying neuro-fuzzy systems in intercity roads

*M. Hadji Hosseinlou; M. Sohrabi

Transportation Departement, Faculty of Civil Engineering, K. N. Toosi University of Technology, Tehran, Iran

ABSTRACT: Providing safety in roads for the purpose of protecting human assets and preventing social and economic losses resulted from road accidents is a significant issue. Identifying the traffic hot spots of the roads provides the possibility of promoting the road safety which is also related to investigate frequency and intensity of occurred accidents. Accidents are multidimensional and complicated events. Identifying the accident factors is based on applying a comprehensive and integrated system for making decisions. Therefore, applying common mathematical and statistical methods in this field can be resulted in some problems. Hence, the new research methods with abilities to infer meaning from complicated and ambiguous data seem useful. Therefore, along with identifying the traffic hot spots, adaptive Neuro-Fuzzy inference system is used to predict traffic hot spots on rural roads. In this process, a fuzzy inference system from Sugeno type is trained applying hybrid optimization routine (back propagation algorithm in combination with a least square type of method) and accident data of Karaj-Chalus road in Tehran province. Then the system was tested by a complete set of data. Finally, the stated system could predict 96.85 % of accident frequencies in the studied blocks. Furthermore, the amount of effective false negative in all cases included only 0.82 % of predictions, which indicated a good approximation of predictions and model credibility.

Keywords: *Accident, safety, optimization, least square, sugeno, hybrid, back propagation*

INTRODUCTION

In this modern world, human beings need more attempts to achieve a proper social, economical, cultural, welfare and political position. They try to use all possibilities to create relation to take advantage of all possible opportunities in each geographical situation. Therefore, roads play significant roles in economic affairs. More than 80 % of transportation is done through road transportation network in Iran, while the casualty statistics resulting from road accidents is very high and has more than 28000 victims per year with their economic and social consequences. The statistics indicate thorough attempt in planning, constructing and safe operating of country roads, while supplying safety isn't possible unless the traffic hot spots are identified within safety inspection plans in existing road networks. Identifying traffic hot spots has been one of the clamorous discussions in transportation and traffic. In the present research, traffic hot spot identification methods are briefly mentioned and then the ability of applied devices, processing and their results are explained. Various methods have been used in identifying the traffic hot spots due to many problems. The simplest way of locating is grading and categorizing them in descending trend according to accident frequencies and recording accidents used in this country. Although this method has clear advantages, efficiency of identifying real traffic hot spots has many capacities for development and promotion (Heydecker and Wu, 2001; Kim *et al.*, 2006; Brown *et al.*, 1992; Norden *et al.,* 1956). In this traditional method, a filed data and historical records for attribution are essentially needed. A main part of researches emphasizes on overcoming this shortage and more development and optimization of identification methods (Norden *et al.,* 1992; Petridou and Moustak, 2000; Tamburri and Smith, 1970). Therefore, many methods based on statistical mathematics, regression to the mean and Bayesian empirical methods have been developed (Ng and Sayed, 2004; Mahal *et al.,* 1982; Wright *et al.,* 1988). Briefly speaking, a broad range of submitted techniques for Traffic hot spot identification (HSID) (including simple categorization of accident rate and frequency, controlling quality rate, identifying the area) have been applied using safety index/number

✉ *Corresponding Author Email: mansour@kntu.ac.ir*
  Tel.: +98 912 337 7434; Fax: +9821 8877 9476

and rate methods/accident patterns identification methods and various applications of Bayesian categorization in techniques with ability to improve identification methods of black spots according to the history of accidents and the ones expected through comparing with similar zones (Hauer *et al.,* 2002; Milton and Mannering, 1998; Li *et al.,* 1994; Tarko *et al.,* 1996). The performed studies in this respect showed that applying simple methods of categorization with Bayesian methods can remarkably increase the capacity of accident data set. Therefore, collecting comprehensive information about accident history is required within research limitation and reference population (Davis and Yang, 2001; Cafiso et al., 1999; Roger *et al.,* 1980). Shortages of this method are as follows: 1) They needed a macro reference population, 2) Selecting reference population was completely optional, 3) Reference population always includes required elements for identifying unsafe situations (Hauer and Persaud, 1984; Schluter *et al.*,1997;Washington and Cheng, 2005). It was found in studies that there may be two types of errors in identifying traffic hot spots which are false positive and false negative. False positive includes errors which may occur in identifying a point of research span. There is a safe condition area which is used in identifying of removing the faults. Since the error is for the purpose of assurance it is called false positive (Washington and Cheng, 2005). But false negative is an error because of which an originally traffic hot spot area is identified as a safe one. Since this error causes omission of some areas of research span which are traffic hot spots, it is called false negative (Washington and Cheng, 2005). To remove these shortages, lots of attempts have been done which lead to superiority of imperial methods resulting in Bayesian pattern (Hauer, 1996; Cheng and Washington, 2005; Sayed and Abdelwahab,1998). On the other hand, studying methods based on new mathematical science such as Neuro-Fuzzy system indicates that such tools with their unique abilities to infer meanings from ambiguous and deficient data can be useful in many transportation problems such as identifying traffic hot spots (Akiyama and Shao, 1993; Busch *et al.,* 1995; Teodorovic, 1999; Huang and Ren, 1999; Lee *et al.*, 2003; Pappis and Mamdani, 1977; Sayed *et al.*, 1995). The present research includes introduction, methodology, data collections and analysis, results and discussions, and conclusion. This research was performed in K. N. Toosi University of technology, (2007).

MATERIALS AND METHODS

According to what has been previously stated and concerning defective and confused records of accidents in Iran, it is tried to find successful results in identifying traffic hot spots by collecting existing data in adaptive Neuro-Fuzzy inference system and taking advantage of the capabilities of such system (Akiyama and Shao, 1993; Busch *et al.,* 1995; Teodorovic, 1999; Pappis and Mamdani, 1977). Fuzzy and Neuro systems are the best approximate reasoning and optimization tools in soft computing. Soft computing means a series of new calculative methods in computer, artificial intelligence and other applied sciences. Complicated research, modeling and analysis are required for all these fields for which the accurate scientific methods in past weren't successful with their easily, analyzing and full solutions.

1. In data research by means of soft computing, in contrast to hard computing, there is no need to presupposition. Data research algorithm by soft computing search the relation between data in 2 ways, with supervision and without supervision (non automatic). Moreover, data research tools not only can use various data, but also the accurately classified number ones (Hong Xing, 2001).

2. Soft computing is based on human mind and proceeds with accepting inaccuracy. The conducting principle here is taking advantage of inaccuracy to control the problem and reduce the solution expenses (Lee, 2005).

3. Contrary to hard computing methods in which accuracy and full modeling of a fact is concentrated; the soft method takes into account minor facts, insufficiency and distrust. The hard computing needs tough and hard conditions for implying and supporting the reply (Hauer *et al.,* 2002).

4. Professor Zadeh believes that despite traditional method of hard computing, there is a tolerance in soft computing for any ambiguity, distrust and relative facts in order to achieve cheaper solutions and better relation between facts (Li, 2001).

5. Neuro systems besides fuzzy logic are one of the tools of soft computing in inferring meanings of complicated and ambiguous data and extracting patterns understanding of which is important for human and other computerized complicated techniques. An adaptive Neuro-Fuzzy system can be considered as an expert in informative category for analysis. This expert can be applied for estimating new desirable situations and replies to "what if …" Questions (Kwang Hyung, 2005).

Now, regarding the complicatedness and ambiguity of a series of factors involving in accidents, modeling such process by common methods of mathematics (which have been broadly applied during the last 50 years) is very difficult by existing accident information throughout Iran. Therefore, developing Neuro-Fuzzy systems is more acceptable in such situations (Roess *et al.,* 2004). Consequently by using MATLAB software and self organized algorithms in Adaptive Neuro-Fuzzy Inference System (ANFIS), Sugeno type of fuzzy inference system was used in present research by applying optimization routine back propagating algorithm in combination with a least square type of method.

Data collection and analysis

To study the function of adaptive Neuro-Fuzzy system in identifying traffic hot spots, achievable data about occurred accidents for Karaj-Chalus two-lane major road is selected which is one of the main roads connecting the capital to the north of Iran. This road is among beautiful roads throughout the world and has a high traffic volume; meanwhile it has potential for accidents due to its mountainous situation. However, since 70 years from its construction, it has not been adapted with new traffic standards. Accident frequency and intensity were selected for identifying traffic hot spots of this road. In this regard, the data related to accidents during 3 months was collected by considering one of the crowded months. That was related to 122 points of this road. Finally, 11 items were selected for identifying the process. Table 1 shows the collected data together with used labels in introducing models. These data include 7 related ones to geometric design future and place status. They were adopted from topographical and geometrical drawings of the road including the situation of horizontal plan, height, vertical slope, horizontal sight push, road width, bridge and tunnel effect and application of studied area. The next 3 data also indicate the amount of traffic volume per day, the amount of hourly traffic volume in the day of accident and the time of accident in the road. Item 11 shows accident frequency of the road which is used as objective function. The collection of these data presents a reliable interpretation of road situations in the time of accident occurrence for all recorded accidents and their occurrence blocks. It is obvious that traffic accident data is too much and complicated to obtain all data. It is worth noting that the data shown in Table 1 is based on existing reliable information in the road. Therefore, only the information represented in Table 1 should be selected. After

collecting data, they were categorized into 2 training and testing groups. The former are latter groups produced for adapting the model with road situations and for studying the credibility of prediction results from ANFIS in estimating expected frequencies, respectively.

As stated before, Neuro-Fuzzy system adopts the ambiguous and invisible concepts and through which the required parameters and components can be adjusted for predicting other data not being confronted with yet. Then the adapted system can predict the frequency of accidents related to the situations of input data introduced in the beginning of this section.

1. At first step, the relation between data and accident frequency is found by using the means of identifying proper input data in Neuro-Fuzzy system. According to direct correlation with target function, the data area is as follows: representative of horizontal plan, representative of sight distance, bridge and tunnel effect, super elevation difference, zone user, longitudinal slope, representative of width difference, passenger car equivalent volume per hour and day, representative of accident time situation.

2. At next step, it was found through studying a broad range of components that the positions (123456) as mentioned above, lead to the best result among all components with more than 6 inputs.

3. Then it was specified that data such as the amount of commuting car equivalent per hour (input 9) and time representative (input 10) have not proper overall situations due to insufficient frequency and high scatteredness comparing with other inputs.

4. As per above results, three combination of input data which have proper cover on road and contain optimized components from previous stages were studied. The results showed that the combination (1234568) is the best combination for predicting the amount of accident frequency of Karaj-Chalus road. The components of this combination include the horizontal plan, road width,

Table 1: The collected data together with used labels in introducing models

Collected data	Used labels
Horizontal Align co.	1
Width difference	2
Super elevation difference	3
Slop co.	4
Tunnel and bridge co.	5
Sight distance difference	6
Residential co.	7
Total Vehicle / Day	8
Total Vehicle / h	9
Day and h co.	10
Accident amplitude	11

311

height difference, longitudinal slope, bridge and tunnel effect, sight push and the amount of car equivalent per day. This combination could produce the accident frequency in 96 % with a very high approximation while the other combination could estimate target function, but the above combination provides the best result.

RESULTS AND DISCUSSION

To clarify the quality of predictions, it is necessary to make a statistical review on the amount and type of occurred errors. The summary of results from applying adaptive Neuro-Fuzzy inference system in predicting accident frequency in Karaj-Chalus road is shown in Figs. 1 to 3, using the (1234568) data combination. Figs. 1 and 2 indicate a very proper approximation of accident frequencies in ANFIS system by putting the expected and predicted amount of accident frequency beside or over each other. Table 2 shows the data distribution in various error amounts. According to Table 2, nearly 83 % of data lead to accurate prediction. Moreover , 7.8 % of predictions have less than 5 % error rate and this error in 76 % of cases produce less than one unit of error. Therefore, in 90.4 % of predictions, the occurred error is very low. In 4.1 % of cases, there is an error of 5 to 15 % which has been occurred due to the system that is acceptable. In 5.6% of remaining cases, the occurred error is more than 15 %. This 5.6 % (including 33 data items) needs more discussion in view of type and reason of produced error is less than unit and only because of very small frequency they are subject to percent mutation. This group of expansion (including 42 %) is ignorable in practice. If the prediction of second position in view of error type is considered, 15 cases of 33 errors were based on the type of false negative (without overlooking predictions with less than unit error and confusion leading to negative frequency). Of course 8 predictions

among these cases are about frequency one unit. 3 cases of these 8 predictions and generally 5 cases of 15 errors are included in less than unit error and in all 33 cases are more than 15 % that equals 14 numbers of total data. Finally adding up to the predictions and omitting less than unit error in more than 15 % of effective errors constitutes 3.15 % of total predictions. It means the adapted system has had acceptable predictions in more than 96.85 % of total ones. The results of Table 2 is depicted in Fig. 3. In other words, root means square error is 6.58 % in this model. In fact, each of the predictions is as mean with 6.58 % error which is very low and acceptable.

CONCLUSION

From this research, the following results will be explicitly obtained:

a) In this identification method, it was attempted to present a method for predicting accident frequency by applying statistics of accidents in a certain road. Hence, the model was adapted for predicting the series of tests. Therefore, the model could write the training data as the similar accident frequency. The result showed that the model can estimate the accident frequency in more than 96 % of cases with good quality and proper coverage over peripheral conditions. In fact, the ability of the model to produce real data with average error of 6.58 % evidences the assurance of processing result. While considering average accident frequency (equals 13.4), each data items was identified and 0.88 % error was averagely found.

b) The credibility of results was studied according to predictable usages for the model. It should be taken into consideration that such identifications may be used for future trends of accidents in the roads without written record. Therefore, the results shall be free of any false
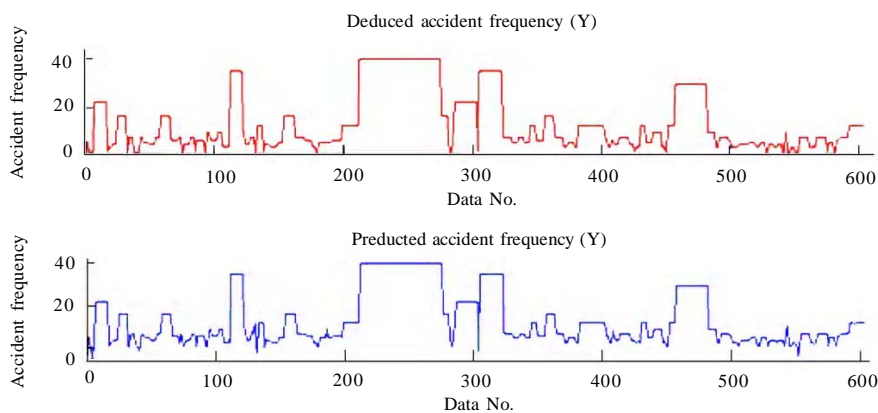

Fig. 1: Expected and predicted amount of accident frequency by ANFIS system

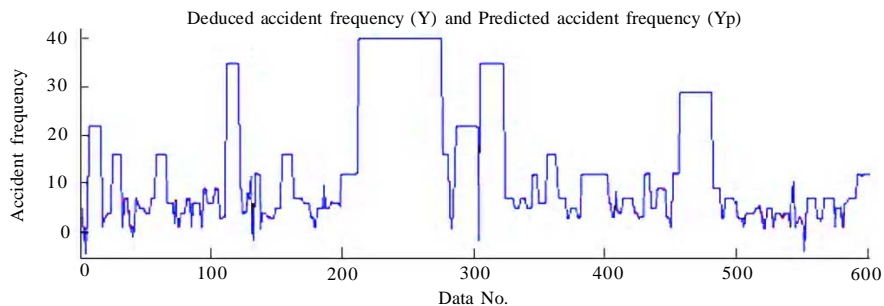Deduced accident frequency (Y) and Predicted accident frequency (Yp)

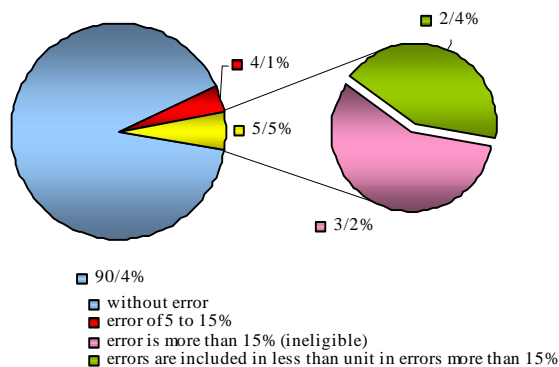Fig. 2: Comparison of expected and predicted amount of accident frequency by ANFIS system

Fig. 3: Summary of results from applying adaptive Neuro-Fuzzy inference system in predicting accident frequency in Karaj-Chalus road

4/1%
2/4%
5/5%
3/2%

90/4%
□ without error
□ error of 5 to 15%
□ error is more than 15% (ineligible)
□ errors are included in less than unit in errors more than 15%

Table 2: Data distribution in various error amounts

Error amounts percent	Number of data in error amounts	Percentage of data in error amounts
100%<	1/7 %	10
[70 %-100 %)	0/8 %	5
[60 %-70 %)	0/5 %	3
[40 %-60 %)	0/3 %	2
[30 %-40 %)	0/7 %	4
[20 %-30 %)	0/7 %	4
[15 %-20 %)	0/8 %	5
[10 %-15 %)	0/8 %	5
[5 %-10 %)	3/3 %	20
[2 %-5 %)	5/6 %	34
[1 %-2 %)	2/2 %	13
<1 %	82/6 %	498

negative as much as possible. However, as previously stated, there was only 5 false negative with more than 1 or 1 error unit. This amount of false negative equals 0.82 % of predicted cases. Such a result also shows the credibility of applied method.

c) As expected before, in identifying traffic hot spots regarding Karaj-Chalus road, a reliable success was obtained in view of producing amounts similar to reality in accident frequency. These results confirm the application of such systems in analyzing complicated traffic problems.

The first interpretation of the above 3 results is that according to abilities of such models, it is expected that such systems shall be used in various inspections of road safety. It should be added that the system can be trained by extracted data from the road with completer accident history and used for the roads with similar conditions. This matter can be very useful in studying the behavior of roads which haven't been utilized yet and it is predicted to operate under various traffic volumes. Other application of such processing is the ability of this system to study the amount of factors effectiveness in accident occurrence. Therefore, using

the results of Neuro-Fuzzy inference system, the hidden relation among the data can be found. Such result can be applied for giving priority in cases where a limited budget is available and one situation with several hazardous reasons is under investigation.

REFERENCES

Akiyama, T.; Shao, C. F., (1993). Fuzzy mathematical programming for traffic safety planning on an urban expressway. Transport. Plan. Tech., 17 (2), 179-190 **(12 pages)**.

Brown, B.; Farley, C.; Forgues, M., (1992). Identification of dangerous highway locations: Results of Community Health Department study in Quebec. In transportation research record, Transportation Research Board (TRB), National Research Council, Washington D.C., 1376, 78-85.

Busch, F.; Cremer, M.; Ghio, A.; Henninger, T., (1995). A muti-model approach with fuzzy reasoning for traffic state estimation and incident detection on motorways. In: ERTICO (Ed.), Proceedings of the 1[st.] world congress on applications of transport telematics and intelligent vehicle-highway systems. Towards an intelligent transport systems. Artech House, Boston and London, 145-1252.

Cafiso, S.; La Cava, G.; Cutello, V., (1999). A fuzzy model for road accidents analysis. Fuzzy Information Processing Society, NAFIPS, 18th International Conference of the North American, 139-143(5), ISBN:0-7803-5211-4.

Cheng, W.; Washington, S. P., (2005). Experimental evaluation

of hotspot identification methods. Accident Anal. Prev., 37 (5), 870–881 **(12 pages)**.

Davis, G. A.; Yang, S., (2001). Bayesian identification of high-risk intersections for older drivers via gibbs sampling. Transportation Research Board (TRB), National Research Council, Transportation Research Record, J. Transport. Res. Board, 1746, 84-89.

Hauer, E.; Persaud, B. N., (1984). Problem of identifying hazardous locations using accident data. In transportation research record Transportation Research Board (TRB), National Research Council, Washington D.C., 975, 36-43.

Hauer, E., (1992). Empirical bayes approach to the estimation of unsafe: The multivariate regression approach. Accident Anal. Prev., 24 (5), 457-477 **(21 pages)**.

Hauer, E.; Harwood, D. W.; Council, F. M.; Griffith, M. S., (2002). Estimating safety by the empirical bayes method: A tutorial. In transportation research record. Transportation Research Board (TRB), National Research Council. J. Transport. Res. Board, 1784, 126-131.

Heydecker, B. G.; Wu, J., (2001). Identification of sites for accident remedial work by Bayesian statistical methods: An example of uncertain inference. Adv. Eng. Softw., 32 (10-11), 859–869 **(11 pages)**.

Huang, S.; Ren, W., (1999). Use of neural fuzzy networks with mixed genetic/gradient algorithmin automated vehicle control. IEEE T. Ind. Electron., 46 (6), 1090-1102 **(13 pages)**.

Kim, D. G.; Washington, S.; Oh, J., (2006). Modeling crash types: New insights into the effects of covariates on crashes at rural intersections. J. Transp. Eng., 132 (4), 282-292 **(11 pages)**.

Lee, K. H., (2005). First course on fuzzy theory and applications, Springer-Verlag Inc. 335.

Lee, C.; Hellinga, B.; Saccomanno, F., (2003). Proactive freeway crash prevention using real-time traffic control. Can. J. Civil Eng., 30 (6), 1034–1041 **(8 pages)**.

Li, H. X., (2001). Fuzzy neural intelligent systems: Mathematical foundation and the applications in engineering, CRC Press Inc. 357.

Li, J.; Abdelwahab, W.; Brown, G., (1994). Joint effects of access and geometry on two-lane rural highway safety in British Columbia. Can. J. Civil Eng., 21 (6), 1012–1024 **(13 pages)**.

Mahal, D.; Hakkert, A. S.; Phrasker, J. N., (1982). A system for the allocation of safety resources on a road network. Accident Analysis and Prevention, 14 (1), 45-56 **(2 pages)**.

Milton, J.; Mannering, F., (1998). The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation, 25 (4), 395-

413 **(19 pages)**.

Ng, J. C. W.; Sayed, T., (2004). Effect of geometric design consistency on road safety. Can. J. Civil Eng., 31 (2), 218–227 **(10 pages)**.

Norden, N.; Orlansky, J.; Jacobs, H., (1956). Application of statistical quality-control techniques to analysis of highway accident data. Highway Research Board (HRB), National Council Washington D.C., 117, 17-31.

Pappis, C.; Mamdani, E., (1977). A fuzzy controller for a traffic junction. Institute of Electrical and Electronics Engineers (IEEE), Transactions on systems, Man and cybernetics, 7 (17), 707-717.

Petridou, E.; Moustak, M., (2000). Human factors in the causation of road traffic crashes. Eur. J. Epidemiol., 16 (9), 819-826 **(8 pages)**.

Roger, P. R.; Prassas, E. S.; William, R.; MCSHane-pearson, (2004). High risk crash analysis, final report 558-Prepared by: Traffic Engineering 3th. (Ed.) Prentice Hall.

Sayed, T.; Abdelwahab, W.; Navin, F., (1995). Identifying Accident-Prone Locations Using Fuzzy Pattern Recognition. J. Transp. Eng., 121 (4), 352-358 **(7 pages)**.

Sayed, T.; Abdelwahab, W., (1998). Comparison of fuzzy and neural classifiers for road accidents analysis. J. Comput. Civil Eng., 12 (1), 42-47 **(6 pages)**.

Schluter, P. J.; Deely, J. J.; Nicholson, A. J., (1997). Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. The Statistician, 46 (3), 293–316 **(24 pages)**.

Tarko, A. P.; Sinha, K. C.; Farooq, O., (1996). Methodology for identifying highway safety problem areas. In Transportation Research Record, Transportation Research Board (TRB), National Research Council, Washington D.C., 1542, 49-53.

Tamburri, T. N.; Smith, R. N., (1970). The safety Index: Method of evaluating and rating safety benefits. In highway research record, Highway Research Board (HRB), National Research Council, Washington D.C., 332, 28-43.

Teodorovic, D., (1999). Fuzzy logic systems for transportation engineering, Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Yugoslavia. Transport. Res. Part A, 33 (5). 337-364 **(28 pages)**.

Washington, S.; Cheng, W., (2005). High risk crash analysis, final report 558. Department of Civil Engineering. University of Arizona. Tucson, AZ 85721.

Wright, C. C.; Abbess, C. R.; Jarrett, D. F., (1988). Estimating the regression to mean effect associated with road accident black spot treatment: Towards a more realistic approach. Accident Anal. Prevent., 20 (3), 199-214 **(16 pages)**.

AUTHOR (S) BIOSKETCHES
Hadji Hosseinlou, M., Assistant professor, Transportation Department, Faculty of Civil Engineering, Khaje Nasir Toosi University of Technology, Tehran, Iran. Email: *mansour@kntu.ac.ir*

Sohrabi, M., Graduated student of transportation engineering, Transportation Department, Faculty of Civil Engineering, Khaje Nasir Toosi University of Technology, Tehran, Iran. Email: *m_sohrabi@sina.kntu.ac.ir*