Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 21 (2011) 317–325

Procedia Social and Behavioral Sciences

International Conference: Spatial Thinking and Geographic Information Sciences 2011

Spatio-Temporal Clustering of Road Accidents: GIS Based Analysis and Assessment

V. Prasannakumar[a*], H. Vijith[a], R. Charutha[a], N. Geetha[a]

[a]*Centre for Geoinformation Science and Technology, University of Kerala, Kariavattom, Thiruvananthapuram, Kerala, India - 695 581*

Abstract

Road accident hot spots are evaluated and delineated in a South Indian city where inadequate development of land transport network often leads to traffic congestion and accidents. The patterns of localization and distribution of hotspots are examined with the help of geo-information technology to bring out the influence of spatial and/or temporal factors in their formation. Assessment of spatial clustering of accidents and hotspots spatial densities was carried out following Moran's I method of spatial autocorrelation, Getis-Ord Gi* statistics and point Kernel density functions. The accidents as a whole show a clustered nature while the comparison of spatio-temporal break ups indicates random distribution in certain classes. The Kernel density surface, estimated from the results of hotspot analysis delineates the road stretches as well as isolated zones where the hotspots are concentrated. The results can be effectively used by various agencies for adopting better planning and management strategies for improved traffic conditions as well as accident reduction.

Selection and/or peer-review under responsibility of Yasushi Asami

Keywords: Spatial statistics; auto correlation, Moran's I; Getis-Ord Gi*; accident hotspots; Thiruvananthapuram;

1. Introduction

Road accidents are on the increase, world wide, mainly because the development of transportation infrastructure fails to keep pace with that in other sectors like industry and real estate. Thus, the road traffic accidents are the leading cause of human deaths and/or illness world-wide [1]. In India, every year, nearly 85,000 persons are reported to be killed and 300,000 are injured in accidents on road [2]. Most of these accidents result from human error, mainly carelessness of the drivers or pedestrians. Hence, the probability of accident occurrence, and its severity, can often be reduced by the systematic analysis of the incident scenario and by resorting to appropriate solutions involving the application of proper traffic control devices, suitable roadway design practices and effective traffic police activities. However, the task of devicing effective solutions warrants analysis of spatial and temporal patterns in the zonation of traffic accidents, which can be achieved through the application of geospatial technology [3,4,5,6] . The non-random distribution of accidents, both in time and space, often raises questions about the location and the reasons for that location [7,8,9,10]. Unlike the conventional methods, spatial thinking helps to identify the

*Corresponding author: Tel.: +91-471-2308214
E-mail address: prasannan@cgist.ac.in

1877–0428 © 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of Yasushi Asami

doi:10.1016/j.sbspro.2011.07.020

patterns and suggest reasons for the pattern characteristics. GIS technology has been a popular tool for visualization of accident data and analysis of hot spots and hence is being used by many traffic agencies [11,12,13,14,15,16,17,18].

The Thiruvananthapuram city corporation is one among the fast growing cities in South India, where drastic changes in land use and land cover have taken place. The vehicle density in the city has shown alarming increase in the past few years, but the road conditions remain same resulting in increased incidents of traffic accidents and the situation warrants a detailed scientific study on the spatial and temporal probability of accident occurrence. As conventional techniques of accident analysis had failed in reducing the occurrence of traffic related accidents, the present study was carried out using GIS based spatial statistical techniques to identify the accident hotspots. This paper aims to evaluate and represent the accident hotspots in the Thiruvananthapuram city by modeling real accident location information in conjunction with various spatial attributes using spatial statistics in geographical information technology. The predictive zonation and analysis can help to identify vulnerable locations and zones that require remedial measures. Besides, the model will also help to delineate the safe road segments which in turn can be effectively used as models in the development of safer pathways.

2. Study Area

The Thiruvananthapuram city corporation, the headquarter of the state of Kerala, India, spreads over 141.74 km^2 (Fig.1), bounded in the west by Arabian sea and on the east by the Western Ghats. The region, in general, has undulating topography with low rising ridges as well as lakes and estuaries. Though the city is categorized as fast growing, it is not properly planned and most of the roads are narrow causing heavy traffic congestion. In addition to being the administrative hub of the state, the city has also developed in recent years, as an Information Technology hub with the starting of a number of IT based industries. The fast and speedy development, in turn, has increased the number of vehicles without proportional augmentation in road infrastructure. Since the roads in the area are always under excessive pressure and the likelihood of accidents is more, the possibility of spatio-temporal prediction of accidents is very important to traffic police department as well as transportation planners and engineers.

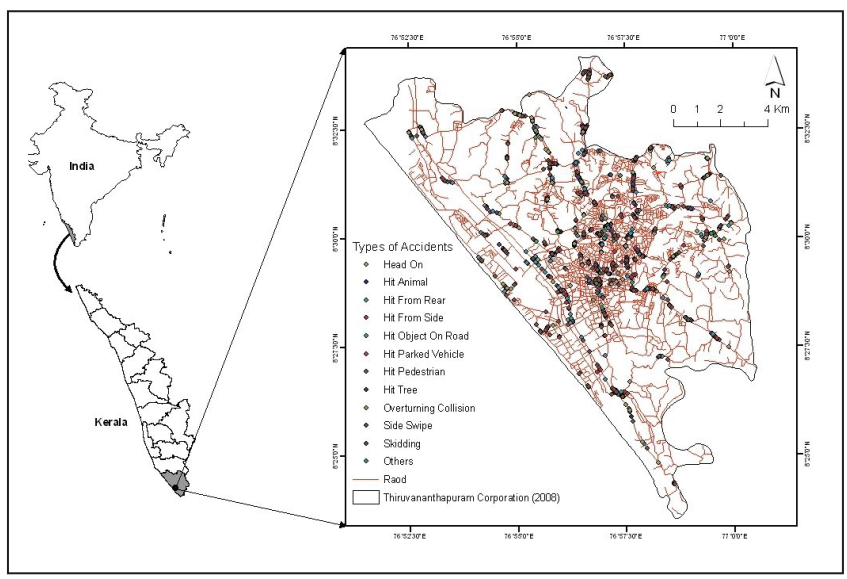

Fig. 1 Map of the study area with accident locations and types.

3. Methodology

The most important purpose of accident hotspots analysis is to identify and generate the information needed to assist the decision makers in adopting suitable measures to prevent and to reduce the accident happenings [5,12,14]. Generally, traffic accidents statistics are considered as assessment indicators to estimate the quantity of the probable incidents on roads in the future. Prediction of spatial pattern of traffic accidents by using actual field data and Geographical Information System (GIS) is relatively recent and the results are well correlatable with the real conditions. The technical details of the datasets and methodology used in the analysis are detailed in the ongoing sections.

3.1. Accident datasets

Road collision (accident) data set used for the present study was obtained from the Traffic Police Headquarters of the Thiruvananthapuram city of Kerala (Fig. 1) and presented in table 1. The datasets represent accident locations for the year 2008, which are 1468 in number and are represented as geocoded x and y coordinates. These accident locations were attributed with detailed information such as place, month, date, day, time, vehicle type, reason, fatality etc. Besides, the corporation area boundary with ward divisions was digitized and was used for the extraction of road networks from the SoI toposheets of scale 1:25,000. After digitization, the road map was updated by GPS field survey.

3.2. Accident data analysis techniques

Spatial statistical mapping is the key to understanding the spatial and temporal occurrence of accidents [3,15,19,20] and spatial statistics comprises a set of techniques for describing and modeling spatial data. Spatial statistical analysis related to road accidents can be preformed on a spatial database incorporating all the desired information and by generating data layers from the available sources updated by field verification. All spatial processings were carried out using ArcGIS 9.3 and its extensions. A Global Moran's I spatial autocorrelation test was carried out for each type of accident incidence in the area. In addition, a hot-spot analysis and Kernel density estimation were also carried out based on the Getis-Ord Gi* statistics and point Kernel density function. Both of these analyses were carried out using ArcGIS's Spatial Statistics tools.

3.2.1. Spatial Autocorrelation: Moran's I method

The Spatial autocorrelation (Moran’s I method), works not only on feature locations or attribute values alone but on both feature locations and feature values simultaneously. Given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed, or random. Moran’s I is one of the oldest indicators of global spatial autocorrelation and is still used for determining spatial autocorrelation [15,20,21,22]. It compares the value of the variable at any one location with the value at all other locations and can be represented as:

Table 1 Accident statistics of Thiruvananthapuram Corporation for the year 2008*.

Sl.No.	Causes	Accidents	No. of injured
1	Drivers faults	1341	3222
2	Mechanical defect	09	21
3	Fault of pedestrian	17	32
4	Others	78	162
5	Unknown causes	23	49

* Those registered as police cases

$$I = \frac{N\sum_i \sum_j Wi.j\,(X_i-\overline{X})(X_j-\overline{X})}{(\sum_i \sum_j Wi.j)\ \sum_i (X_i-\overline{X})(X_j-\overline{X})^2} \qquad (1)$$

where N is the number of cases, Xi is the variable value at a particular location, Xj is the variable value at another location, X is the mean of the variable, and Wij is a weight applied to the comparison between location i and location j. Wij is a distance-based weight matrix which is the inverse distance between locations i and j (1/dij).

The spatial pattern analysis tool in the ArcGIS calculates the Moran's I Index value and a Z score, which indicate statistical significance. In general, a Moran's Index value near +1.0 indicates clustering while an index value near -1.0 indicates dispersion. In the case of the Spatial Autocorrelation tool, the null hypothesis states that "there is no spatial clustering of the values". When the Z score is large (or small) enough and falls outside the desired significance, the null hypothesis can be rejected. When the null hypothesis is rejected, the next step is to inspect the value of the Moran's I Index. If the value is greater than 0, the set of features exhibits a clustered pattern and the pattern is dispersed if the value is less than 0[21,23,24].

3.2.2. Hot Spot Analysis

A hot spot is a location or a small area within an identifiable boundary showing concentration of incidents [21]. The three major processes involved in the estimation of desired hotspots of accident incidents are collection of events, mapping of clusters using Getis-Ord GI* function and density estimation using Kernel density tool.

Collect-event function available with the spatial statistic tool was used for performing the function, which in turn will yield a new weighted point feature class with a field ICount that indicates the sum of all the accidents happened in a unique geographic location. This weighted point feature was used as the input for running the hotspot function (Getis-Ord GI*) to identify whether features with high values or features with low values tend to cluster in the study area. This tool works by looking at each feature within the context of neighbouring features. If a feature's value is high, and the values for all of it's neighbouring features is also high, it is a part of a hot spot. The local sum for a feature and its neighbours is compared proportionally to the sum of all features; when the local sum is much different than the expected local sum, and that difference is too large to be the result of random chance, a statistically significant Z score is the result [21,25,26]. The statistical equation for calculating Gi and Gi* can be written as,

$$Gi^*(d) = \frac{\sum j\ W_{ij}(d)\,x_j - W^*_i\,\overline{x}^*}{s^*\left\{\left[(nS^*1i) - W_i^{*2}\right]/(n-1)\right\}^{1/2}} \qquad (2)$$

where ‘W_{ij} (d)’ is a spatial weight vector with values for all cells ‘j’ within distance d of target cell i, W^*_i is the sum of weights, S^*_{1i} is the sum of squared weights and s^* is the standard deviation of the data in the cells.

The Gi* statistics is actually a Z score. For statistically significant positive Z scores, the larger the Z score, the more intense the clustering of high values. For statistically significant negative Z scores, the smaller the Z score, the more intense the clustering of low values. Finally the Kernel density hotspots with the populated field as GiZScore were performed with the point density calculator function available with the spatial analyst tool. It calculates the magnitude per unit area from each hot spot features using the populated GiZScore field. The output of the Kernel density function is a raster file displaying the areas of high and low clusters of accident occurrence [27,28].

4. Result and Discussion

The analysis of temporal and spatial pattern of accidents in the study area, focuses on the total number of accidents as well as accidents happened during monsoon and non-monsoon time (temporal) and

accidents near to religious places and educational institutions (spatial). The monsoon and non monsoon periods are considered as temporal variables since these are the two major seasons in the state that could affect the traffic and may influence the chances of accidents. Similarly two major institutions viz, religious places and educational institutions are considered separately under spatial variables due to the association of large number of people with these point features. Evaluation of basic characteristics of the accident data sets indicates that among the total number of accidents (1468), there exists a temporal partition as monsoon (702) and non-monsoon (766); and a spatial association to educational institution (820) and religious places (333).

Much of traffic accident mapping is devoted to detecting high accident density areas known as hot spots. There are many ways to define and analyze hotspots for a given set of incident locations. As a first step, spatial pattern was assessed by spatial autocorrelation method (Moran's I). To perform this task, spatial autocorrelation coefficient using Moran's I index and associated Z-score were computed for all the datasets (total accidents, temporal and spatial partitions). While deriving the Moran's I index, it was noted that, a clustered pattern exists for datasets of total accidents, non-monsoon and accidents near to educational institutions. Other two datasets, monsoon and those associated with religious places, are classified as random in nature. The unpartitioned total accident dataset showed a high positive Z score (2.25 standard deviation), statistically significant at 0.05 level with Moran's I index of 0.01 with critical value of 1.96 (Fig. 2). In the case of accident datasets of non-monsoon, Z score was 1.98 standard deviation (statistically significant at 0.05) with Moran's I index of 0.03 with critical value of 1.96 (Fig. 3). While running the Moran's I index for accident locations near educational institutions, it was categorized as somewhat clustered due to random chance (Fig. 4) with a Z score of 1.18 standard deviation and Moran's I index of 0.02. The Moran's classification of pattern of accidents in monsoon season and that near to religious places is neither clustered nor dispersed with Z score less than 1 and Moran's I index of <0.03 (Fig. 5).

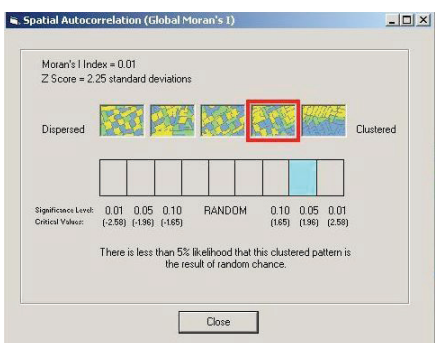

Fig. 2 Moran's I classification of entire accident data.

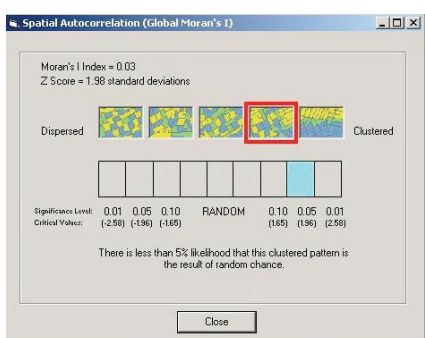

Fig.3 Moran's I classification of accidents during non-monsoon period.

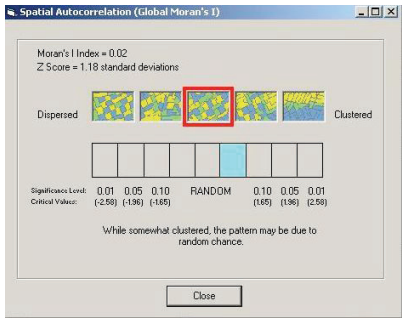

Fig. 4 Moran's I classification of accidents near educational institutions.

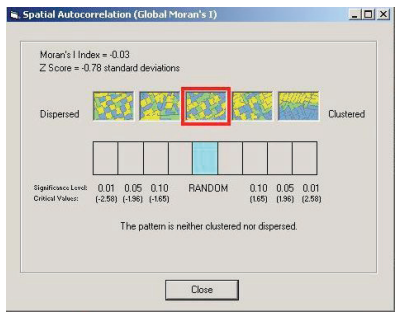

Fig. 5 Moran's I classification of accidents near religious places.

The hotspots and Kernel density surfaces were derived for total accidents as a whole and particularly for accidents during monsoon and non-monsoon times and incidents near to educational institution and religious places. The hotspots for each classified types of accidents were calculated using the Getis-Ord Gi* function followed by the event calculation. The Getis-Ord Gi* statistics identifies spatial clusters of high values (hot spots) and of low values (cold spots). The output of hotspot analysis tool is GiZScore and GiPValue for each feature. These values represent the statistical significance of the spatial clustering of values, given the conceptualization of spatial relationships and the scale of analysis. A high GiZScore and small GiPValue (probability) for a feature indicates a spatial clustering of high values where as a low negative GiZScore and small GiPValue indicates a spatial clustering of low values[21,27]. The higher the GiZScore, the more intense is the clustering. A Z score near zero indicates no apparent spatial clustering. The GiZScore for unclassified total accident locations varies between -2.093 to 7.781 and the GiPValues from 0 to 0.998. From this, it is inferred that statistically significant positive GiZScore (high values) indicates accident hotspots, while statistically significant negative GiZScore (low values) indicates coldspots. In the case of accidents during monsoon and non-monsoon seasons, the GiZScore and GiPValues range from 5.556 to -1.250 and 4.286 to -2.171 and 0 to 0.211 and 0.029 respectively. Similar hotspots Gi* statistics were observed for accidents near educational institutions and religious places. The GiZScore ranges from 5.753 to -1.552 and 5.748 to -1.384 with GiPValues 0 to 0.120 and 0 to 0.166 respectively for incidents near educational institutions and religious places.

The Kernel density estimated from the derived hotspots (total accidents, monsoon, non-monsoon, near educational institutions and religious places), suggests spatial variability in the distribution pattern of accident hotspots and coldspots in the Thiruvananthapuram corporation. The advantages of such two dimensional hotspot surface representations, particularly of road accidents, can provide a more realistic continuous model of accident hotspot patterns, over space and time. The hotspots spatial density estimated for the unclassified total accidents (Fig.6) and the spatial distribution of accidents during the monsoon and non-monsoon seasons (Fig. 7a&b) shows high variability in the hotspot and coldspot

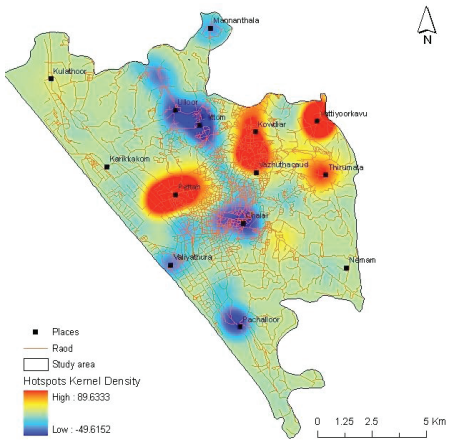

Fig. 6 Spatial characteristics of accident hotspots (unpartitioned data)

clusters. The same changes are observed in the spatial variability of coldspots from monsoon to non-monsoon period. While analyzing spatial pattern of accidents near to educational institutions, the hotspots are concentrated in Kowdiar –Vazhuthacaud stretch, Vattiyoorkavu and Thirumala regions. At the same time the accidents near to religious places show isolated high concentration near to Pettah and Thirumala areas (Fig. 8 a&b). Both the images show isolated coldspots, which are well distributed in the study area. Thus, the spatial pattern of accidents analysed in the present study enables the analyst to quickly and aesthetically locate statistically satisfactory accident hotspots in the Thiruvananthapuram city corporation area.

5. Conclusion

The present study which investigates and compares different kinds of traffic accidents in terms of spatial and temporal aspects is the first attempt of its kind in the Thiruvananthapuram city corporation. The results of the spatial statistics and cluster analysis elicit the spatial and temporal variations of accident highs (hotspots) and lows (coldspots) in the area. The assessment of spatial characteristics of the accident data by Moran's I method, derived Z score, Moran's I index and critical value indicates that

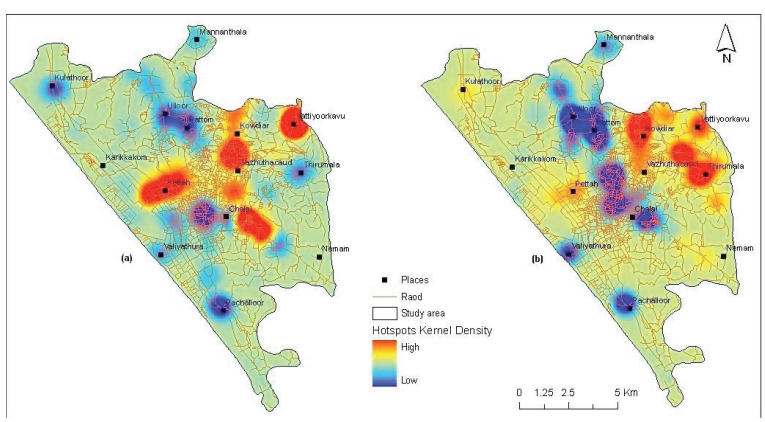

Fig. 7 Spatial characteristics of accident hotspots (a). monsoon and b). non-monsoon period).

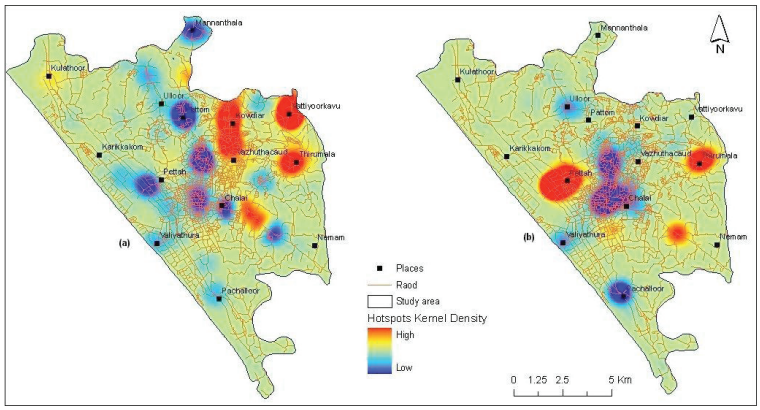

Fig. 8 Spatial characteristics of accident hotspots associated with (a). educational institutions (b). religious places

accident datasets, as a whole, are categorized as clustered in nature, followed by accidents during non-monsoon and those associated with educational institutions. Among the analyzed data, monsoon time spatial densities, estimated based on the Getis-Ord Gi* statistics and point Kernel density function, for much of the datasets, indicate that the accident hotspots as well as cold spots are clustered around specific sectors with isolated highs and lows and show spatial variation among the datasets. While the results can be effectively used for the successful management of traffic and reduction of accidents, micro level studies can be attempted with additional information on geographical boundary related to population, vehicle details and road characteristics.

Acknowledgements

The paper is a contribution of the Centre for Geoinformation Science and Technology, a joint venture of the University of Kerala, DST and KSCSTE. Financial support from DST and KSCSTE is gratefully acknowledged

References

[1] ICMR. Development of a feasibility module for road traffic injuries surveillance. *Ind Council Med Res Bul* 2009;**39(10-12)**: 42-50.

[2] Mohan D. Road accidents in India. *Inter Assoc Traffic and safety Scie Res* 2009; **33(1)**: 75 – 79.

[3] Levine N, Kim KE, Nitz LH. Spatial analysis of Honolulu motor vehicle crashes: I. spatial patterns. *Acc Anal and Prev* 1995; **27(5)**: 663-674.

[4] Harkey DL. Evaluation of truck crashes using a GIS-based crash referencing and analysis system transportation research record. *J Trans Res Board* 2007;**1686(1999)**: 13-21.

[5] Li L, Zhu L, Sui DZ. A GIS-based bayesian approach for analyzing spatial–temporal patterns of intra-city motor vehicle crashes. *J Trans Geog* 2007;**15(4)**: 274-285.

[6] Cheng W, Washington S. New criteria for evaluating methods of identifying hotspots. Transportation Research Record: Journal of the Transportation Research Board, No. 2083, Transportation Research Board of the National Academies, Washington, D.C, 2008, pages 76–85.

[7] Maher MJ, Mountain LJ. The identification of accident blackspots: a comparison of current methods. *Acc Anal and Prev* 1988; **20(2)**: 143–151.

[8] Liang LY, Hua LT, Ma'some DM 2005. Traffic accident application using Geographic Information System. *J Eastern Asia Soc Trans Studies* 2005; **6**: 3574 – 3589.
[9] Randall TA, Churchill CJ, Baetz BW. 2005. Geographic information system (GIS) based decision support for neighbourhood traffic calming. *Can J Civil Eng* 2005; **32**: 86-98.
[10] Schuurman N., Cinnamon J, Crooks VA, Hameed SM. Pedestrian injury and the built environment: an environmental scan of hotspots. *BMC Public Health* 2009; 9(233): (doi:10.1186/1471-2458-9-233).
[11] Miller HJ. Potential contribution of spatial analysis to geographic information systems for transportation (GIS-T). *Geog Anal* 1999; **31(4)**: 373–399.
[12] Cook WD, Kazakov A, Persaud BN. 2001. Prioritizing highway accident sites: a data envelopment analysis model. *J Oper Res Soc* 2001; **52**: 303-309.
[13] Ghosh SK, Parida M, Uraon JK. Traffic accident analysis for Dehradun city using GIS. *ITPI Journal* 2004;**1(3)**: 40-54.
[14] Eck JE, Chainey S, Cameron LG, Leitner M, Wilson RE. Mapping crime: Understanding hot spots. Special report. National Institute of Justice, U.S. Department of Justice, USA. www.ojp.usdoj.gov/nij accessed on 29/12/2011.
[15] Mitchell A. *The ESRI guide to GIS analysis. Volume 2: Spatial measurement and statistics.* ESRI Press, Redlands, California;2005.
[16] Anderson T. Comparison of spatial methods for measuring road accident hotspots: a case study of London. *J Maps* 2006: 55-63.
[17] Erdogan S, Yilmaz I, Baybura T, Gullu M. 2008. Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. *Acc Anal and Prev* 2008; **40**: 174–181.
[19] Burrough PA, McDonnell R. *Principles of Geographical Information Systems*. Oxford Press. 1998.
[21] ESRI. ESRI online help for spatial statistics and analysis. http://resources.esri.com/help/9.3/arcgisdesktop/com/. Accessed on 7/12/2010.
[23] Bailey TC, Gatrell AC. *Interactive spatial data analysis*. Addison Wesley Longman Limited, Harlow, England.1995.
[18] Deepthi JK, Ganeshkumar B. 2010. Identification of accident hot spots: A GIS based implementation for Kannur District, Kerala. *Inter J Geomatics and Geosc* 2010;1(1): 51-59.
[20] Haining R. *Spatial Data Analysis: Theory* and *Practice*. New York: Cambridge University Press.2003.
[22] Jing N, Cai W. Analysis on the spatial distribution of logistics industry in the developed East Coast Area in China. *Ann Regional Scie* 2010; **45**: 331–350.
[24] Griffith DA. *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. Springer, Berlin, Germany. 2003.
[25] Ord JK, Getis A. Local spatial autocorrelation statistics: distribution issues and an application. *Geog. Anal* 1995; **27 (4)**: 286-306.
[26] Getis A, Aldstadt J. Constructing the spatial weights matrix using a local statistic. *Geog Anal* 2004;**36(2)**: 90-104.
[27] DeGroote JP, Sugumaran R, Brend SM, Tucker BJ, Bartholomay LC. 2008. Landscape, demographic, entomological, and climatic associations with human disease incidence of West Nile virus in the state of Iowa, USA. *Inter J Health Geog* 2008; **7(19)**: doi:10.1186/1476-072X-7-19.
[28] Zhixiao, X., Jun, Y., 2008. Kernel density estimation of traffic accidents in a network space. Geography/Geology Faculty Publications. Paper 3. http://digitalcommons.wku.edu/geg_fac_pub/3 accessed on2/3/2011.