Computers, Environment and Urban Systems 41 (2013) 249–261

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

The identification of traffic crash hot zones under the link-attribute and event-based approaches in a network-constrained environment

Becky P.Y. Loo *, Shenjun Yao

Department of Geography, The University of Hong Kong, Pokfulam, Hong Kong

ARTICLE INFO

Article history:
Received 6 April 2011
Received in revised form 4 July 2013
Accepted 4 July 2013
Available online 5 August 2013

Keywords:
Link-attribute
Event-based
Network
Traffic crashes
Hot zone

ABSTRACT

In the spatial analysis of road traffic crashes, a hot zone methodology explicitly uses the network contiguity of more than one road segment as a criterion in identifying crash clusters. In this paper, 603 simulated patterns of traffic crashes in three simplified hypothetical networks and the empirical crash pattern in Hong Kong from 2008 to 2010 (with a total of 30,490 traffic crashes on 1090 km of roads) are analyzed using the link-attribute approach and the network-constrained event-based approach. Procedures for identifying hot zones using statistical thresholds are developed. This paper represents the first systematic comparison of hot zone results using these two different approaches. The results suggest that the link-attribute approach and network-constrained event-based approach are usually consistent but there are major differences between the two approaches.

1. Introduction

Do spatial clusters of traffic crashes exist? If so, then where are they? How can we identify them in an efficient and scientific manner? Traffic crash hot zones (also called black zones) are defined as "a set of contiguous road segments taken together and characterized by a high number" of traffic crashes (Flahaut, Mouchart, Martin, & Thomas, 2003: 992). Traffic crash hot zones are distinct from traffic crash hotspots (also known as blacksites or blackspots) (Cheng & Washington, 2005; Elvik, 1997, 2006; Geurts, Wets, Brijs, & Vanhoof, 2004; Huang, Chin, & Haque, 2009; Miranda-Moreno, Labbe, & Fu, 2007) because of two distinct characteristics. First, hot zones must include more than one road segment (the distinction does not lie in the length of the road segments). Second, those road segments with high crash numbers must be contiguous (Loo, 2009; Moons, Brijs, & Wets, 2009a). With the hotspot methodology, a junction or an individual road segment with a large number of crashes is identified as a hotspot. With the hot zone methodology, more than one individual road segment is taken together to become a single hot zone, which is a single spatial cluster of traffic crashes. In other words, network contiguity of more than one road segment is an essential criterion for identifying hot zones.

This paper addresses the methodological challenges of detecting hot zones using the concept of spatial autocorrelation (Fotheringham, 2009; Getis & Ord, 1992; Goodchild, 1986) and the statistical methods developed under two commonly used approaches in geographic information systems (GIS) analysis. The first approach is the link-attribute approach (Yamada & Thill, 2007, 2010). Spatial events such as traffic crashes are not analyzed directly but are instead assigned to geographic features, such as areas or a road network. For the former, the focus is usually on visualizing and explaining the spatial variability of the crash intensity across areas (polygons), such as traffic zones, census tracts, districts, regions and provinces (Chen, Lin, & Loo, 2012; Erdogan, 2009; Levine,Kim, & Nitz, 1995b). For the latter, traffic crashes are assigned to line and point features, namely roads (links) and junctions (nodes). Links are, in turn, divided into shorter segments called basic spatial units (BSUs) for detailed spatial analysis. Traffic crash numbers or rates are treated as attribute values of these geographic features. The second approach is to consider the physical locations of individual crashes (events) directly. This approach is often termed the event-based approach (Yamada & Thill, 2007). Much event-based research on traffic crashes aims to describe spatial patterns using "point process" tools (Kim & Yamashita, 2007; Levine, Kim, & Nitz, 1995a; Okabe, Satoh, & Sugihara, 2009). To illustrate, the K-means clustering algorithm is used to group crashes into cluster centroids that minimize the sum of thesquared distance from every point to the K centers (Kim & Yamashita, 2007). In other words, all crashes are assigned to one of the spatial clusters. Spatial statistics are used primarily as a tool for data reduction and grouping. When the event-based approach is used to identify local clusters, this goal is accomplished by directly measuring the (physical or network) distance or the degree of concentration among the traffic crashes. With reference to hot zones, the methodology is better developed under the link-attribute approach

* Corresponding author. Tel.: +852 3917 7024; fax: +852 2559 8994.
E-mail address: bpyloo@hku.hk (B.P.Y. Loo).

0198-9715/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compenvurbsys.2013.07.001

but has never been fully developed using the event-based approach (elaborated below). Therefore, this paper provides a new scientific contribution through fulfilling two research objectives. The first objective is to develop and improve the hot zone methodology using both approaches. The second objective is to perform the first systematic comparison of hot zone results under the two approaches. Morespecifically, we ask the following questions: Do network structures matter? Does the relative performance of the two approaches vary with different underlying spatial patterns of traffic crashes? In an experimental setting, the concentrated, dispersed and random spatial patterns of traffic crashes are tested on grids, limited access and organic road structures. In the empirical situation of Hong Kong, road safety implications and recommendations are identified based on hot zones that are identified by the two different approaches.

The remainder of this paper is structured as follows. First, a literature review is conducted. Then, various methodological problems and solutions for applying the link-attribute and event-based approaches to identify hot zones are reported. With the suggested procedures, the hot zone methodology is applied to 603 simulated crash patterns in three simplified hypothetical networks and the empirical crash pattern in Hong Kong from 2008 to 2010 (a total of 30,490 traffic crashes on 1090 km of roads) using the two different approaches. Based on the results of the hot zone analysis, some relative strengths and weaknesses of the two approaches are identified. Moreover, directions for further research are suggested.

2. Literature review

Using the link-attribute approach, Black and Thomas (1998) illustrate that a positive network autocorrelation (in other words, spatial co-variance or spatial co-occurrence) of road crashes can lead to the spatial clustering of more than one traffic hotspot to form hot zones. The Black and Thomas paper was the first major theoretical work that clearly distinguishes traffic crash hotspots from hot zones. However, the analysis focused on network autocorrelations on a global level through establishing whether hot zones were present or not within an entire dataset by using Moran's I and the associated z score tests. Global spatial statistics summarize the overall pattern of a spatial phenomenon within the whole study area, and local spatial statistics detect and describe specific concentrations. The first paper that documented methods for identifying local traffic hot zones (with a binary variable of 0 and 1) for individual road segments using the link-attribute approach came much later, with the work of Flahaut (2004). Local indices of spatial autocorrelation (LISA) were introduced to examine the crash patterns of 2363 crashes on 460 km (4604 BSUs) of numbered road networks in a Belgian province from 1998 to 2000. This work also explained the statistical advantages of hot zones and further developed logistic regression models to explain traffic crash hot zones with road characteristics and local environmental conditions.

Following the link-attribute approach, Loo (2009) proposed a three-stage hot zone methodology and presented the first large-scale empirical study of hot zones. By comparing the hot zone results with the official "blacksite" methodology in Hong Kong, her work shows that the hot zone methodology was superior and more flexible in many aspects, especially in the detection of hazardous road locations on expressways and in rural areas, where intersections were fewer. To compare the hot zone results with the official "blacksite" methodology, simple numerical threshold definitions and only one year's crash data were used. Through applying Loo's three-stage hot zone methodology, Moons et al. (2009a) systematically identified and compared characteristics of hotspots and hot zones on highways in a Belgium province. Their conclusion was that "the hot zone methodology supplements the hot spot methodology, and it is superior and more flexible in some ways" (Moons et al., 2009a: 298).

In contrast to the link-attribute approach, none of the event-based papers explicitly distinguishes hotspots and hot zones. Hence, the papers reviewed here mainly provide the necessary background for introducing the network-constrained event-based approach. Yamada and Thill (2004) were among the first to argue convincingly that spatial statistics of network-constrained events must be different from spatial analysis methods that were designed for planar space. Network-constrained events (such as traffic crashes) are defined as those in which there are both location and movement constraints imposed by the network space. To provide a clear distinction (because different procedures are involved), the latter is described as a network-constrained, rather than a planar, event-based approach. Using the empirical data from New York State, Yamada and Thill (2004)further explained and compared the methods and results of the planar and network-constrained K-function methods. Their findings "clearly indicate that the planar K-function analysis is problematic since it entails a significant chance of over-detecting cluster patterns" (157). More recently, Yamada and Thill (2007) introduced a method called local indicators of network-constrained clusters (LINCS) to identify hotspots by using the event-based approach. Theoretically, traffic crashes can occur at every possible location over the entire road network. However, it is neither practical nor feasible to examine the clustering tendency at every possible point for cluster identification using the network-constrained event-based approach. Hence, they suggested using reference points (RPs) along the network with an equal interval *Int*. The network-constrained local K-function was then used to analyze 1628 vehicle crasheson 145 miles of highway in Buffalo, New York. It should be noted that the degree of clustering of the K-function among neighboring RPs (network contiguity) was not explicitly considered. Using RPs, Xie and Yan (2008) used a network-constrained KDE approach to examine 3,226 traffic crashes in the Kentucky area. They concluded that the network-constrained KDE was more appropriate than the planar KDE for traffic crash analysis. The methodological challenges of defining the *lixel* length (a concept similar to the BSU length under the link-attribute approach) and search bandwidth were addressed. However, the detection of hotspots was through visual inspection. In contrast to RPs, Steenberghen, Aerts, and Thomas (2010) used "focus points" for measuring a Dangerousness Index based on the influence distance (50 m) along the shortest-path distancebetween the "focus points" and the locations of the traffic crashes. A Monte Carlo method was used to determine the statistical significance of the Dangerousness Index. A map showing the Dangerousness Index in the central business district of Brussels (162 km^2) was presented. Once again, only visual inspection of crash clusters was conducted, and individual road segments were not further classified as hotspots or hot zones. Arguably, if the results of the network-constrained event-based approach are examined carefully, then hot zones can still be identified visually. However, the procedures and related methodological issues (such as the rules and thresholds for hot zone identification) have never been systematically examined.

Last, there are a few papers that address both the link-attribute and event-based approaches in identifying local clusters of traffic crashes. A pioneer comparative study was performed by Flahaut et al. (2003). These authors compared the link-attribute approach using LISA with the planar event-based approach to identify hot zones in Belgium. In comparing the results, the highest correlation was achieved at a spatial weighting structure (d_{ij}^{-2} weights) for Moran's I (link-attribute approach) and a specific window width (2.5 hectometer) for KDE (planar event-based approach). The comparison, however, was performed on the highly simplified scenario