

Assessing network-based traffic crash risk using prospective space-time scan statistic method

Congcong Miao^a, Xiang Chen^{a,*}, Chuanrong Zhang^{a,b}

^a Department of Geography, University of Connecticut, Storrs, CT, USA

^b Center for Environmental Sciences and Engineering, University of Connecticut, Storrs, CT, USA



ARTICLE INFO

Keywords:

Traffic crash
Risk modeling
Space-time
STSS
Prospective

ABSTRACT

As car ownership and urbanization continue to rise worldwide, traffic crashes have become growing concerns globally. Measuring crash risk provides insight into understanding crash patterns, which can eventually support proactive transport planning and improve road safety. However, traditional spatial analysis methods for crash risk assessment, such as the hotspot detection method, are mainly focused on identifying areas with higher crash frequency. These methods are subject to critical issues in risk analysis due to ignoring crash impacts and background traffic volume information. Aside from the two issues, current crash risk assessment methods, especially those aiming for cluster detection, are subject to the modified temporal unit problem, referring to the temporal effects (i.e., aggregation, segmentation, and boundary) in cluster detection. To alleviate these issues, this paper applies an emerging hot spot detection method, called the prospective space-time scan statistic (STSS) method, for assessing the crash risk at a refined network scale and over multiple years in a case study of Hartford, Connecticut. By identifying the spatial and temporal clusters of the crash risk, the study can provide evidence for tailoring road safety management strategies in neighborhoods characterized by high crash risk.

1. Introduction

As car ownership and urbanization continue to rise worldwide, traffic crashes have become growing concerns globally, incurring substantial deaths, injuries, and economic losses on a daily basis. Although many policy attempts, such as improving vehicle standards, have been made to mitigate the impact of traffic crashes, road traffic deaths and injuries remain a major global development challenge (WHO, 2023). In particular, the United Nations' Sustainable Development Goals (SDG) 3.6, originally set to halve the global deaths and injuries from traffic crashes by 2020, has been extended to 2030 due to the rising scope and frequency of traffic accidents (United Nations, 2020; Mohan et al., 2021; WHO, 2023b).

The occurrence of traffic crashes is closely associated with environmental factors, including land use patterns (Ouyang and Bejleri, 2014), road infrastructure (Papadimitriou et al., 2019), and lighting conditions (Haleem et al., 2015). As these environmental factors vary across space, traffic crashes are subject to a high degree of spatial variations (Ziakopoulos and Yannis, 2020). Therefore, spatial analysis methods, such as summarizing cases delineated by census units or identifying the hot spots of crashes, are frequently employed to identify high-risk areas

(Ziakopoulos and Yannis, 2020; Shahzad, 2020). These methods, however, are subject to three methodological issues in risk analysis. First, risk is a combined measure of both the likelihood for a hazardous event to occur and the potential impact when it occurs (Rausand, 2013; Stoneburner et al., 2002). Past assessments of the crash risk focusing on the historical occurrences (which represent the likelihood) have rarely accounted for the crash impact. For example, locations with fewer crashes but higher mortality rates could be overlooked if the impact component is missing in the assessment. Second, using the total number of cases to represent the crash risk does not account for the variation of traffic volume in space and over time. More specifically, the total cases may not be an appropriate risk indicator, as the traffic volume contributing to the likelihood of crashes is overlooked. This issue shares similarities with the data normalization issue in epidemiology, where the risk for an epidemic outbreak is not only dependent on the absolute disease cases but also the total population susceptible to the disease (Adams et al., 2023). Third, existing crash risk assessment is subject to the Modified Temporal Unit Problem (MTUP) (T. Cheng and Adepeju, 2014). The MTUP arises as the modeling results can be influenced by how the data is organized temporally, including the different temporal durations of data, the different ways of segmenting the temporal data,

* Corresponding author.

E-mail address: xiang.chen@uconn.edu (X. Chen).

and the smallest unit used for such segmentation. While the MTUP exists in all temporal data modeling, it will have considerable influence on long-time series data, such as multiple-year traffic crashes. Traditional risk assessment models (e.g., kernel density) are unable to tackle the MTUP.

In this study, we have modeled traffic crash risk on road networks by considering these issues. Our model identifies the spatial-temporal patterns of traffic crash risk on road networks using the prospective space-time scan statistic (STSS) method. The prospective STSS method has gained wide recognition in epidemiology for its effectiveness in detecting clusters of diseases with a moving space-time window (Kulldorff et al., 1998). This method is particularly helpful for continuous surveillance of hazardous events, as it detects clusters that are active or emerging towards the end of the study period (Kulldorff, 2001). Also, as the STSS can identify the high or low risk clusters without being constrained by a given temporal duration or unit, it can alleviate the issue of MTUP. By using the prospective STSS method, we have derived network-based traffic crash risk using a seven-year crash dataset. To the best of our knowledge, this study is the first to utilize this new method for modeling traffic crash risk and can help support transport planning and policy initiatives that aim to mitigate crash risk in vulnerable neighborhoods.

The rest of the paper is organized as follows: Section 2 provides a review of the spatial and temporal analysis methods for modeling traffic crash risk. Section 3 introduces the prospective STSS method for assessing traffic crash risk. Section 4 presents the result of applying the method to a seven-year traffic crash dataset of Hartford, Connecticut. Section 5 provides a further discussion of the result in terms of the policy implications, advantages, and limitations, followed by the last section that concludes the paper.

2. Literature review

2.1. Spatial modeling of traffic crash risk

Traffic crashes are typically represented as spatial points based on the locations where they occur. Thus, the modeling of traffic crash risk can be guided by two essential properties of spatial point data: spatial heterogeneity and spatial dependence (Mohaymany et al., 2013; Ziakopoulos and Yannis, 2020).

The spatial heterogeneity of traffic crashes relates to the fact that traffic crashes are unevenly distributed over space due to the differences in risk factors within the built environment (Mohaymany et al., 2013). This property can be measured by aggregating crash counts within certain spatial units, such as road intersections (Abdel-Aty and Wang, 2006), road segments (Aguero-Valverde et al., 2016), traffic analysis zones (TAZs) (Ng et al., 2002), and census tracts (Wier et al., 2009). However, the measurement precision could be limited by the scale of the chosen spatial units, and thus, modeling results across different spatial units can be inconsistent (Abdel-Aty et al., 2013; Cai et al., 2017). To complement the lack of consistency, the kernel density estimation (KDE) model was employed for identifying the spatial heterogeneity of traffic crashes (Hashimoto et al., 2016). Specifically, the KDE creates a smooth, continuous density surface, where the hotspots of crashes can be visualized (T. Anderson, 2007; T. K. Anderson, 2009). As traffic crashes happen mostly on road networks, the NetKDE method, which is a network-based KDE, was developed for measuring network-based spatial heterogeneity (Xie and Yan, 2008), and was utilized for identifying high-risk road segments in various case studies (Chen et al., 2018; Loo et al., 2011; Mohaymany et al., 2013). However, a common weakness of all the KDE methods is the relative subjectivity in parameter selection. Specifically, the modeling results can be largely influenced by a smoothing parameter called the bandwidth, and the selection of the bandwidth is relatively difficult (Loo et al., 2011; Xie and Yan, 2008).

The spatial dependence characterizes the interactive effects between nearby crashes (Mohaymany et al., 2013; Ziakopoulos and Yannis,

2020). This property is due to the fact the influence of one crash at a particular location can be related to the occurrence of another crash nearby, as the crashes in near proximity can be induced by similar environmental risk factors. This effect can be measured by Moran's I Index and Geary's C Ratio (Yao et al., 2016; Ziakopoulos and Yannis, 2020; Shahzad, 2020), whereas many analysts prefer Moran's I due to its more favorable distribution characteristics and its greater overall stability and flexibility (Mitra, 2009). When detecting crash-prone locations, the local versions of Moran's I and Geary's C are more suitable, as they can calculate for individual locations and reveal their spatial autocorrelation with nearby locations (Anselin, 1995; Yamada and Thill, 2007, 2010; Tortum and Atalay, 2015). However, the local Moran's I and Geary's C cannot differentiate high-value clusters from low-value clusters (Erdogan, 2009). To solve this issue, another index for local spatial autocorrelation, Getis-Ord Gi*, has been proposed to distinguish between high-value clusters and low-value clusters in terms of the Gi* value (Ord and Getis, 1995; Yamada and Thill, 2010). The Gi* value has been used to identify hotspots and cold spots of different types of traffic crashes at different zonal levels, such as road buffer polygons (Rankavat and Tiwari, 2013), traffic analysis zones (Soltani and Askari, 2017), and administrative regions (Erdogan, 2009). The traditional local Moran's I and Getis-Ord Gi* have also been extended to the network, known as the local indicators of network-constrained clusters (LINCS) method (Yamada and Thill, 2010). The LINCS considers the spatial separation between observations based on the distance along a specific network (Yamada and Thill, 2007, 2010), making it more suitable for analyzing network-constrained scenarios, including traffic crashes (Liu et al., 2019; Nie et al., 2015).

These existing spatial methods aim to identify hotspots of crash locations across space, rather than detecting true clusters based on the nature of crash occurrence. In contrast, STSS searches for significant clustering that exceeds expectations under baseline conditions and can incorporate relevant environmental factors into risk assessment.

2.2. Temporal modeling of traffic crash risk

In recent years, many scholars have attempted to bring the temporal dimension into the spatial modeling of traffic crash risk. The majority of these studies analyze the spatial and temporal characteristics of traffic crashes separately. Specifically, a common approach is aggregating traffic crashes based on a certain time scale, for which the space-time cube is frequently used as a framework for spatiotemporal aggregation (Z. Cheng et al., 2018; Wu et al., 2022; Yoon and Lee, 2021). Then, traffic crash clusters were detected through cluster detection methods, such as Getis-Ord Gi* (Hazaymeh et al., 2022; Soltani and Askari, 2017) and KDE (Fan et al., 2018; Kazmi et al., 2022; Ouni and Belloumi, 2018; Özcan and Küçükönder, 2020), and these identified clusters were compared over different time periods. For example, Bil et al. (2019) applied a fixed moving window in time and computed spatial clustering within each time window. Kang et al. (2018) utilized space-time kernel density estimation (STKDE) to identify the change in traffic crash concentration over time, whereas the identification results were a comparison of spatial patterns at different points in time (i.e., hourly).

One issue in these temporal modeling studies refers to the MTUP (Cheng and Adepeju, 2014). The MTUP is a problem similar to the classic modifiable areal unit problem (MAUP), whereas it refers to the data uncertainty in the temporal dimension. The MTUP arises as the modeling results can be influenced by how the data is organized temporally, including the different temporal durations of data (i.e., boundary), the different ways of segmenting the temporal data (i.e., segmentation), and the smallest unit used for such segmentation (i.e., aggregation). While the MTUP exists in all temporal modeling, it will have considerable influence on long time series data, such as multiple years of traffic crashes. As such, existing traffic crash risk assessments are generally restricted to revealing high/low clusters in a single temporal unit (e.g., one week) instead of identifying different temporal

periods (e.g., from several weeks to several months) that represent statistically significant high/low clusters.

The MTUP can be alleviated by the prospective STSS method. The STSS method utilizes a cylindrical scanning window, where the base of the window represents the spatial scope and the height of the window represents the time interval (Kulldorff et al., 1998). The scanning process involves adjusting the location, size, and height of the cylindrical window continuously while conducting statistical analyses to detect clusters. As the scanning window is not bounded by a fixed unit on the vertical axis or in time, it can generate temporally unrestricted clusters, eventually overcoming the MTUP.

The STSS method can be generally divided into two groups, the retrospective method and the prospective method. While the retrospective STSS method considers historic clusters for a fixed study period, the prospective STSS method prioritizes the detection of active or emerging clusters. For this reason, the prospective STSS method is more suitable for periodic surveillance, such as disease surveillance (Hohl et al., 2020; Kulldorff, 2001; Takahashi et al., 2008; Vicente Ferreira et al., 2022) and crime prediction (Gao et al., 2013), as it prioritizes clusters that continue to be active until the end of the study period (Kulldorff, 2001). The applications of the STSS method to traffic crash analysis have been relatively limited. We have only identified two case studies. Dai (2012) utilized the STSS method to study clusters of pedestrian injuries, in which the Bernoulli model was employed to evaluate the risk of pedestrian injuries given a traffic crash occurred. Song et al. (2018) employed the STSS method for the detection of traffic crash clusters by employing the space-time permutation model.

However, both of these two studies are retrospective analyses, which are unable to detect active or emerging clusters.

3. Method

3.1. Study area and data

Hartford is the capital city of Connecticut in the United States. As of 2020, Hartford has a population of 121,054 with a land area of 17.4 mile². Along with the exurbanization process and more reliance on private vehicles, Hartford has experienced a substantial surge in automobile usage over the past decades (McCahill and Garrick, 2010a; McCahill and Garrick, 2010b). The road network of Hartford consists of diverse types and varying traffic volumes. In addition to the densely distributed local roads, the city is intersected by two major interstate highways: I-84 and I-91. I-84 primarily traverses this area in the east-west direction, while I-91 serves as a prominent north-south thoroughfare, running through the eastern part of the region.

The traffic crash dataset (Fig. 1) was gathered from the Connecticut Crash Data Repository (CTCDR) maintained by the Connecticut Transportation Safety Research Center (CTSRC). The dataset encompasses seven-year traffic crashes between January 1, 2015, and December 31, 2021. The dataset included comprehensive information for each crash, including date, time, geographical coordinates, crash severity (fatal, injury, or property damage only), and other relevant details. In total, there were 46,515 crashes in the dataset, comprised of 100 fatal crashes, 12,487 nonfatal injury crashes (including serious, minor, or possible



Fig. 1. Three types of traffic crashes recorded from 2015 to 2021 in Hartford, Connecticut. Arterial roads are labeled in red text and neighborhoods are labeled in black text. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

injury), and 33,928 property damage-only (PDO) crashes.

The road network and traffic volume information were collected from the Connecticut Department of Transportation (CTDOT) Open Data portal. The road network includes all public roads and routes as defined by the Federal Highway Administration (FHWA). The traffic volume data was provided as the road-based Annual Average Daily Traffic (AADT). Certain local road segments lack available AADT information, primarily due to their assumed lower traffic volumes and consequent lack of monitoring. We then performed a sensitivity analysis to determine the missing value, identifying that any value below 350 had a minimal impact on the cluster detection result. We then chose AADT = 100 to represent the significantly low traffic on unmonitored roads.

3.2. Crash severity

One contribution of our model is to incorporate the crash impact in terms of crash severity. Severity denotes the intensity level of damage resulting from a crash. While a crash might lead to multiple damages or injuries with varying degrees of severity, the term “crash severity” pertains to the most severe injury caused by the crash (AASHTO, 2010).

In this study, we utilized the Equivalent Property Damage Only (EPDO) method, as documented in the Highway Safety Manual (HSM) (AASHTO, 2010), to represent the severity-weighted number of traffic crashes. For each specific location, its observed EPDO value (c) is calculated from Eq. (1), which combines the severity weight (w) and the observed counts of crashes (f).

$$c = \sum (w \times f) \quad (1)$$

where w is the severity weight corresponding to a severity level, and f represents the number of crashes on this severity level.

The severity weight w was established in consideration of the comprehensive societal costs associated with crashes, which take into account factors such as medical care, emergency services, property damage, and other relevant expenses in the estimation (AASHTO, 2010; Young and Park, 2014). The severity weights, which were derived from the inflation-adjusted value of the comprehensive cost on each crash severity level, are presented in Table 1 (CTSRC, 2020).

The assessment of traffic crash risk is based on specific points of measurement that are subsequently utilized to interpolate information for the entire network (Steenberghen et al., 2010). Thus, we employed road intersections as these measurement points. First, we aggregated point-based traffic crashes onto road intersections by the Thiessen polygons generated for these intersections, as the Thiessen polygon has been widely used as the spatial unit for network-based risk assessment in past studies (Chen et al., 2012; Church and Cova, 2000; Cova and Church, 1997). We also derived the EPDO value of each road intersection by summarizing the weighted traffic crashes within its Thiessen polygon. Second, as the traffic volume is a critical variable used in our STSS method, we derived the traffic volume of each intersection by overlaying its Thiessen polygon with the road network. In cases where one polygon overlaps multiple road segments with varying traffic volumes, the highest value was selected to represent the traffic volume at the intersection.

3.3. Prospective space-time scan statistic

To identify clusters of traffic crashes that are posing threats to road

Table 1
Estimated monetary costs and severity weights by crash severity.

Crash severity	Mean comprehensive cost	Severity weight (w)
Fatal	\$6,415,389	574
Injury	\$123,646	11
PDO	\$11,186	1

safety, we employed the prospective STSS method. The STSS method employs a cylindrical window, where the circular base represents the spatial extent, and the height corresponds to the time interval. The cylindrical window is sequentially positioned at the intersections within the study area and gradually expanded until the maximum spatial and temporal limits are reached. This process generates an unknown large number of overlapping cylinders with varying sizes, each representing a potential cluster.

Compared with the retrospective STSS, the prospective analysis only focuses on significant clusters that remain active or emerging at the ending time of the scan. Specifically, the starting time of the scanning window varies in time, and the ending time of the window is the end of the study period. In mathematical notation, $[T_1, T_2]$ represents the study period for the analysis (where T_1 is the start time and T_2 is the end time), and $[s, e]$ represents the time interval of the cylindrical scanning window (where s is the start time of the window and e is the end time), the prospective analysis examines the cylindrical windows for which $T_1 \leq s \leq e = T_2$. Therefore, the prospective analysis is suitable for the surveillance of recent observations. For example, by conducting the analysis periodically as T_2 increases, we can consistently monitor the concentration of the risk that is induced by recent crashes. The method is illustrated in Fig. 2.

We selected the discrete Poisson model for cluster detection, as the expected number of crashes is assumed to follow a Poisson distribution (Kulldorff, 2022). In the analysis of traffic crashes within each cylindrical scanning window, we used the EPDO value to represent the severity-weighted number of crashes. Under the null hypothesis, the expected EPDO value within each cylinder ($E(c)$) is directly proportional to its traffic volume when no covariates are present. This relationship is shown in Eq. (2):

$$E(c) = a \times \frac{\cup(c)}{\cup(a)} \quad (2)$$

where a is the traffic volume within a cylinder, $\cup(c)$ is the total EPDO values across the study area within the study period, and $\cup(a)$ is the total traffic volume across the study area within the study period.

Under the Poisson assumption, the likelihood ratio $L(Z)/L_0$ for a specific cylindrical scanning window Z is given by Eq. (3):

$$\frac{L(Z)}{L_0} = \left(\frac{c_Z}{E(c_Z)} \right)^{c_Z} \left(\frac{\cup(c) - c_Z}{\cup(c) - E(c_Z)} \right)^{\cup(c) - c_Z} \quad (3)$$

where $L(Z)$ is the likelihood function for cylinder Z , L_0 is the likelihood under the null hypothesis, which is an identical constant for all cylinders (Kulldorff, 2001; Ma et al., 2016). c_Z is the observed EPDO value in cylinder Z , $E(c_Z)$ is the expected EPDO value in cylinder Z under the null hypothesis, and $\cup(c)$ is the total observed EPDO value over the study period for the entire study area. In essence, the likelihood ratio $L(Z)/L_0$ measures the traffic crash risk within a cylinder on the basis of that outside the cylinder (Hohl et al., 2020).

The definition of the scan statistic S is the maximum likelihood ratio over all possible cylinders, given by Eq. (4) (Kulldorff, 2001):

$$S = \max \left\{ \frac{L(Z)}{L_0} \right\} \quad (4)$$

When the likelihood ratio is maximized over all locations and sizes of the cylindrical window, the window with the maximum likelihood ratio S is identified as the most likely space-time cluster. To determine the statistical significance of the space-time clusters, we obtained the p -value through the Monte Carlo hypothesis testing with 999 rounds of simulations (Kulldorff, 2022).

3.4. Relative risk of traffic crash

The relative risk (RR) of the traffic crash for each cluster is calculated

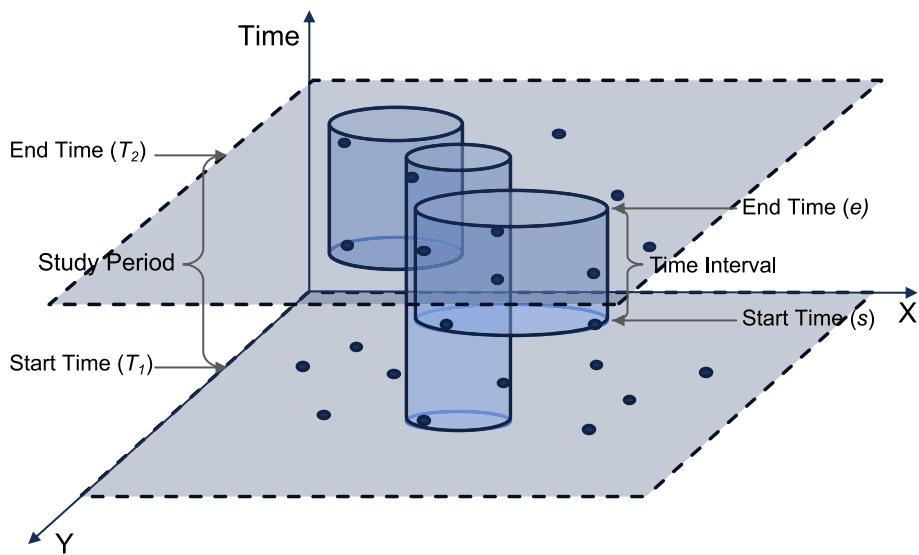


Fig. 2. An illustration of the prospective STSS.

by Eq. (5):

$$RR = \frac{c/E(c)}{(\cup(c) - c)/(\cup(c) - E(c))} \quad (5)$$

where c is the observed EPDO value within the cluster, $E(c)$ is the expected EPDO value for the same cluster as derived by Eq. (2), and $\cup(c)$ is the total observed EPDO value of the entire study area within the study period. In essence, RR indicates the estimated risk within the cluster in comparison to the estimated risk outside of the cluster (Desjardins et al., 2020).

Based on the identified clusters, we further calculated the network-based relative risk. This calculation was based on the intersections of the road network. The intersections within the same cluster may not have a uniform value of relative risk, as they may vary in terms of traffic volume and EPDO values. For each intersection within the identified clusters, the relative risk (RR_i) was calculated by Eq. (6):

$$RR_i = \frac{c_i/E(c_i)}{(\cup(c) - c_i)/(\cup(c) - E(c_i))} \quad (6)$$

where c_i is the observed EPDO value of intersection i , $E(c_i)$ is the expected EPDO value for intersection i , and $\cup(c)$ is the total sum of observed EPDO values of the entire study area during the study period.

The calculation of $E(c_i)$ is similar to Eq. (2) but uses the traffic volume of the intersection instead of within the entire cylindrical cluster. For those intersections located outside the identified clusters, we consider their RR_i values to be 1, implying their traffic crash risk values are not significantly different from the rest of the study area.

The relative risk of the traffic crash along the road network can be derived from road segments. Based on a past study on risk modeling (Chen et al., 2012), the relative risk of a road segment ($RR_{s,t}$) can be determined by averaging the RR values of the two endpoints (s and t) of that segment, as shown in Eq. (7):

$$RR_{s,t} = \frac{RR_s + RR_t}{2} \quad (7)$$

where RR_s and RR_t are the RR values of the two endpoints of the road segment (s, t).

3.5. Coefficient of variation

In this study, we performed risk assessments for a seven-year traffic crash dataset using the prospective STSS method. In essence, we derived

the spatial distribution of the network-based crash risk values annually. This analysis can be quantified on the temporal scale by analyzing the temporal variation of risk, which is a crucial aspect of traffic safety analysis and can reflect both the likelihood and the resulting severity of future traffic crashes (Mannering, 2018). For example, roads consistently exhibiting a high risk (meaning low-variation, high-risk roads) should be prioritized for proactive transportation planning, such as investigating the underlying environmental factors that can lead to increased risk levels.

In this regard, we applied the coefficient of variation (COV) as a variation index for the RR value. The COV is defined as the standard deviation of the target variable divided by its mean value. In our study, the value of COV reflects the temporal variation of traffic crash risk. A higher COV value indicates greater variation in the RR value, while a lower COV value indicates less variation. For any given road segment, the calculation of its COV is given in Eq. (8).

$$COV = \sqrt{\frac{\sum_{t=1}^n (RR_t - \bar{RR})^2}{n-1}} / \bar{RR} \quad (8)$$

where RR_t is the RR value of a road segment at time point t , \bar{RR} is the average RR value of this road segment across all time points, and n represents the number of times the analysis is conducted (i.e., $n = 7$ in our case).

3.6. Covariate adjustment

The STSS method allows for adjustments based on specific covariates. The objective is to identify clusters that cannot be explained by the covariates (Kullendorff, 2022). When selecting a covariate for adjustment, two conditions should be met: 1) the covariate is related to the phenomenon under study and 2) the covariate is not randomly distributed geographically. Previous research has established significant associations between land use characteristics and traffic crash frequency (Huang et al., 2018; Pulugurtha et al., 2013; Wier et al., 2009). For this reason, we attempted to adjust for land use characteristics. The land use characteristics of the study area were collected from Hartford zoning regulations (City of Hartford, 2022). There were nine different land use types in this study area, as further described in the Appendix.

To detect risk clusters that cannot be explained by the covariates, we need to first estimate the expected EPDO value for each location based on the covariates. Then, this estimated expected EPDO value will be used in the scanning process to determine whether the observed EPDO

value is significantly higher or lower than expected. In practice, we overlaid the Thiessen polygons generated from the road intersections with the zoning map of land use to calculate the area of each land use type in square kilometers within each polygon. Considering the Thiessen polygon provides coverage for each intersection, we also included the length of roads in kilometers within each polygon as a covariate. We then developed a regression model to estimate the covariate-adjusted expected EPDO values of road intersections, considering the AADT, the road length, and the area of each land use as the independent variables. We used the negative binomial model with log-link for the regression as it is suitable for area-level estimation of traffic crashes (Pulugurtha et al., 2013; Schneider et al., 2010). The covariate adjustment in greater detail is included in the Appendix.

4. Results

4.1. Prospective analysis of traffic crash risk

We implemented the prospective STSS method in the SaTScan 10.1 (Kulldorff, 2022). We conducted a periodical analysis of the traffic crash data from 2015 to 2021. Let $[T_1, T_2]$ represent the entire study period, T_1 remained the same on January 1, 2015, while T_2 advanced from December 31, 2015, to December 31, 2021 (Table 2). This setting served the purpose of periodical surveillance—as T_2 extended further in time, a larger amount of data and more recent data were included in the assessment. For this study, we set the maximum spatial cluster size as a radius of 1.0 km to prevent extremely large clusters. Also, we set the upper temporal boundary of the window as 50% of the study period. This setting aligned with other studies using the STSS method (Kulldorff et al., 1998; Kulldorff, 2001; Song et al., 2018; Xu and Beard, 2021).

Fig. 3 illustrates the identified statistically significant clusters and their RR values calculated by Eq. (5). The seven maps reveal the active clusters by the end of each year from 2015 to 2021. The areas with RR values <1 indicate a significantly lower risk compared to other areas. The areas with RR values >1 are considered high risk, which are further divided into four levels. Areas not covered by any cluster mean that their traffic crash risks are neither significantly higher nor significantly lower than other areas. The figure shows that a substantial portion of the low-risk clusters are identified near the two interstates, I-84 and I-91, which is especially evident by the end of year 2018. Conversely, areas with primarily local roads are on the lower end of the high risk, with RR values ranging from 1 to 10.

Table 3 presents a summary of the 27 statistically significant active clusters (20 high-risk clusters and 7 low-risk clusters) identified by the end of 2021 (Fig. 3, 2021/12/31). The recurrence interval serves as an alternative to the p -value for assessing statistical significance in the prospective analysis by reflecting how often the observed value could occur under the null hypothesis (Kleinman, 2004; Kulldorff, 2022). One advantage of the recurrence interval is that a larger value indicates a higher level of statistical significance, contrasting with the counterintuitive interpretation of smaller p -values as more alarming.

Fig. 4 shows the RR values of the intersections within the active clusters in Fig. 3. This figure was derived by applying Eq. (6) to the 1555

intersections in the study area. Further, by applying Eq. (7) to the 2754 road segments, we derived the network-based RR, as shown in Fig. 5. It should be noted that this figure combines the low-risk segments (RR value <1) and no statistically different risk segments (RR value = 1), while the high-risk segments (RR value >1) are divided into four levels. Fig. 4 as a point-based visualization allows for a closer examination of the spatial heterogeneity of the risk level, while Fig. 5 as a line-based visualization is helpful for policy interventions to target high-risk roads. The detected risk distributions in Figs. 3 through 5 are summarized in Table 2.

4.2. The temporal variation of traffic crash risk

The RR values at the end of 2021 and the COV of RR from 2015 to 2021 across the road network are illustrated by the bivariate map in Fig. 6, where the red gradient represents the value of RR, and the blue gradient represents the COV. This figure visualizes the current spatial patterns and temporal variation of the crash risks. Based on the two dimensions of measurement, we can roughly categorize the road segments into four risk types: (a) low-risk, low-variation, (b) high-risk, low-variation, (c) high-risk, high-variation, and (d) low-risk, high-variation.

4.3. The land use adjusted traffic crash risk

Using the negative binomial regression model, we estimated the expected EPDO values for each road intersection on a monthly basis from 2015 to 2021. Subsequently, we conducted a prospective STSS analysis based on these model-estimated expected EPDO values and the corresponding observed EPDO values from 2015 to 2021. The statistically significant clusters identified in this analysis are the active clusters by the end of 2021 that cannot be explained by AADT, road length, and land use variables. We also derived the RR values for the intersections within these clusters and the network-based RR for the road network, as shown in Fig. 7.

Compared to the clusters revealed without covariate adjustment (Fig. 3, 2021/12/31), there is a noticeable reduction in the coverage area of high-risk clusters identified by the covariate-adjusted analysis (Fig. 7a). Only a few intersections have aggregations of high-risk clusters. Some areas exhibit significantly high traffic crash risks regardless of whether the covariates are adjusted for, such as the Upper Albany neighborhood section along US-44 (Fig. 7b and c). This result indicates that beyond the land use, traffic volume, and road length, there are other environmental factors contributing to the elevated traffic crash risks in these areas. These consistently high-risk areas need to be prioritized in transport safety management.

5. Discussion

We would like to further discuss how these risk types manifest in the study area (Fig. 6). First, it is observed that two interstate highways (I-91 and I-84) are all categorized under low-risk low-variation (light blue segments, Fig. 6a). This pattern shows that considering the notably higher traffic volume along the highways, both the relative risk and the

Table 2
Summary of the detected high/low-risk distribution in seven periods.

Analysis ID	Time span of data	High-risk clusters		Low-risk clusters		High-risk intersections		Low-risk intersections		High-risk road length		Low-risk road length	
		#	#	#	%	#	%	#	%	km	%	km	%
1	2015/1/1–2015/12/31	19	10	367	23.96	595	38.84	162.34	37.61	133.41	30.91		
2	2015/1/1–2016/12/31	15	6	515	33.62	471	30.74	192.40	44.58	94.11	21.8		
3	2015/1/1–2017/12/31	20	10	522	34.07	460	30.02	200	46.34	110.28	25.55		
4	2015/1/1–2018/12/31	20	13	442	28.85	430	28.07	170.55	39.52	117.93	27.33		
5	2015/1/1–2019/12/31	21	11	477	31.14	378	24.67	177.54	41.14	100.04	23.18		
6	2015/1/1–2020/12/31	17	10	546	35.64	423	27.61	200.06	46.36	98.99	22.93		
7	2015/1/1–2021/12/31	20	7	558	36.42	417	27.22	204.13	47.3	87.97	20.38		

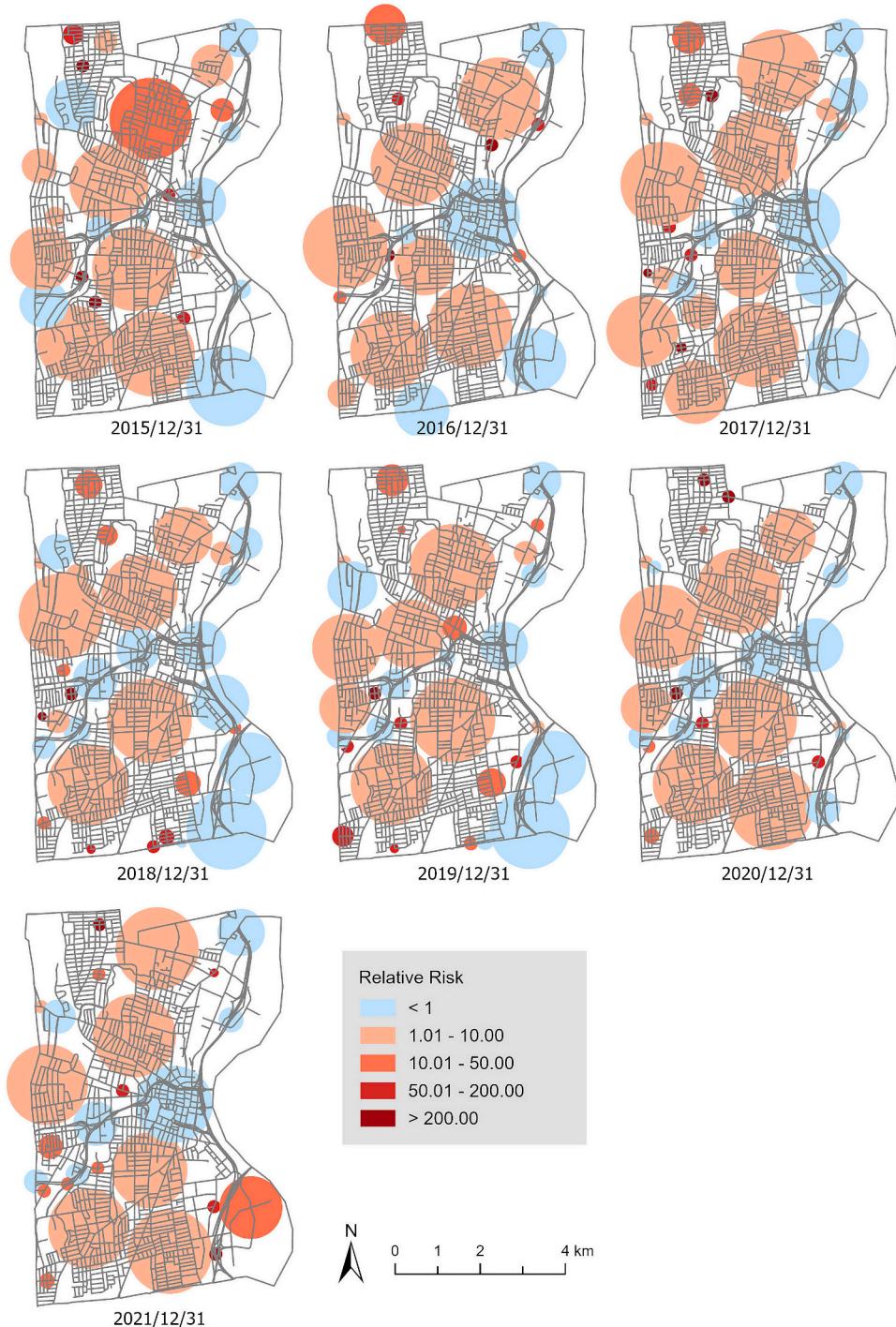


Fig. 3. High-risk clusters (four levels of red circles) and low-risk clusters (blue circles) calculated by the prospective STSS method in seven periods, where each panel starts from 2015/1/1 and ends on the given date. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

risk variation in these areas are significantly low. As the relative risk is a space-time weighted measure of the EPDO value, it suggests that the most severe crashes did not occur along the highways and are more likely to happen on the local roads. Second, the high-risk low-variation areas (dark blue segments) mean that their crash risks have been consistently high (e.g., Fig. 6b). These areas could be regarded as the most vulnerable neighborhoods that need further investigations on the environmental factors inducing the risk. For example, Fig. 6b depicts a section of the Upper Albany neighborhood, with US-44 acting as a central spine with densely populated residential streets on both sides. A

further investigation shows that this neighborhood has experienced a shift in modal split from mass transit to private cars (Patel et al., 2005), which might have made the local built environment unable to accommodate the increasing traffic volume. US-44 within this area serves not only as a major commuting route for residents but also as a commercial corridor for local businesses (Patel et al., 2005). This mixed road function could have further intensified traffic complexity, contributing to the consistently high risk levels in this area.

We would like to highlight four advantages of applying the prospective STSS method for proactive transportation planning. First, the

Table 3

Attributes of prospective space-time clusters (2015/1/1–2021/12/31).

Cluster ID	Start date	End date	Duration (month)	Radius (km)	Observed EPDO	Expected EPDO	Relative risk	p-value	Recurrence interval	# of intersections
1	2018/7/1	2021/12/31	42	0.97	14,701	2585.94	6.01	<0.0001	>100 years	114
2	2018/7/1	2021/12/31	42	0.89	14,194	2543.04	5.88	<0.0001	>100 years	88
3	2018/7/1	2021/12/31	42	0.90	16,802	41,171.80	0.36	<0.0001	>100 years	277
4	2018/7/1	2021/12/31	42	0.99	10,420	2831.26	3.81	<0.0001	>100 years	133
5	2018/7/1	2021/12/31	42	0.52	2297	11,752.13	0.19	<0.0001	>100 years	57
6	2021/11/1	2021/12/31	2	0	601	0.032	18,858.86	<0.0001	>100 years	1
7	2021/7/1	2021/12/31	6	0.1	1187	11.85	100.65	<0.0001	>100 years	2
8	2020/4/1	2021/12/31	21	0	574	0.17	3433.06	<0.0001	>100 years	1
9	2019/8/1	2021/12/31	29	0	1175	16.12	73.25	<0.0001	>100 years	1
10	2021/5/1	2021/12/31	8	0.29	1262	32.98	38.47	<0.0001	>100 years	14
11	2018/7/1	2021/12/31	42	0.94	6266	1810.47	3.53	<0.0001	>100 years	112
12	2018/7/1	2021/12/31	42	0.97	3109	633.40	4.96	<0.0001	>100 years	40
13	2021/2/1	2021/12/31	11	0.74	756	15.31	49.55	<0.0001	>100 years	12
14	2018/7/1	2021/12/31	42	0.56	441	3477.01	0.13	<0.0001	>100 years	20
15	2021/3/1	2021/12/31	10	0	598	9.54	62.87	<0.0001	>100 years	1
16	2019/6/1	2021/12/31	31	0	693	20.63	33.69	<0.0001	>100 years	1
17	2018/7/1	2021/12/31	42	0.30	711	3222.68	0.22	<0.0001	>100 years	9
18	2019/11/1	2021/12/31	26	0	585	22.67	25.87	<0.0001	>100 years	1
19	2018/7/1	2021/12/31	42	0.29	301	2160.99	0.14	<0.0001	>100 years	6
20	2019/8/1	2021/12/31	29	0	635	38.21	16.66	<0.0001	>100 years	1
21	2018/7/1	2021/12/31	42	0.28	626	2459.61	0.25	<0.0001	>100 years	12
22	2021/8/1	2021/12/31	5	0	593	46.76	12.71	<0.0001	>100 years	1
23	2018/7/1	2021/12/31	42	0.95	2297	1386.65	1.66	<0.0001	>100 years	54
24	2018/7/1	2021/12/31	42	0	274	65.46	4.19	<0.0001	>100 years	1
25	2018/8/1	2021/12/31	41	0.18	62	1.96	31.60	<0.0001	>100 years	6
26	2018/10/1	2021/12/31	39	0.40	135	429.72	0.31	<0.0001	>100 years	8
27	2018/7/1	2021/12/31	42	0.035	275	118.95	2.31	<0.0001	>100 years	2

method incorporates the crash impact, which is a largely missing yet critical risk component, in crash risk assessment. Modeling the crash impact in terms of different severities is a necessary step to quantify different degrees of societal losses, as it extends the risk analysis from a pure physical assessment into a socioeconomically weighted evaluation. Second, the method offers a solution to alleviate the MTUP effect in crash risk assessment by handling the time scale, as the STSS method are not constrained by a fixed temporal unit. This relaxation in the temporal component is essentially useful for studying traffic crashes over a long term, as high-risk clusters could appear not within a predefined time unit (e.g., a week) but can last for different time periods (e.g., from several days to several months). Traditional risk assessment models (e.g., KDE), however, are not able to unveil these different time periods with statistical significance. Third, the method can alleviate the regression-to-the-mean (RTM) effect. The RTM is a statistical

phenomenon where short-term observations fluctuate around the data average due to inherent data randomness ([Sharma and Datta, 2007](#)). The STSS method runs the Monte Carlo tests on cumulative long-term historical data. This simulation-based method applied to the longitudinal data can identify clusters that are not likely to be affected by data randomness ([Kulldorff, 2022](#)), and thus this identification is less subject to the RTM effect. Fourth, the prospective STSS method is suitable for continuous risk monitoring, as it focuses on the detection of currently active clusters. This feature can be of importance for traffic safety analysis since active clusters might be linked to certain risk factors that require urgent attention. As time progresses, more recent data can be incorporated into recurrent analyses, eventually allowing for time series analysis or even near-real-time monitoring of risk fluctuation.

Three issues regarding applying the prospective STSS method for crash risk modeling need further attention. First, as the spatial size of

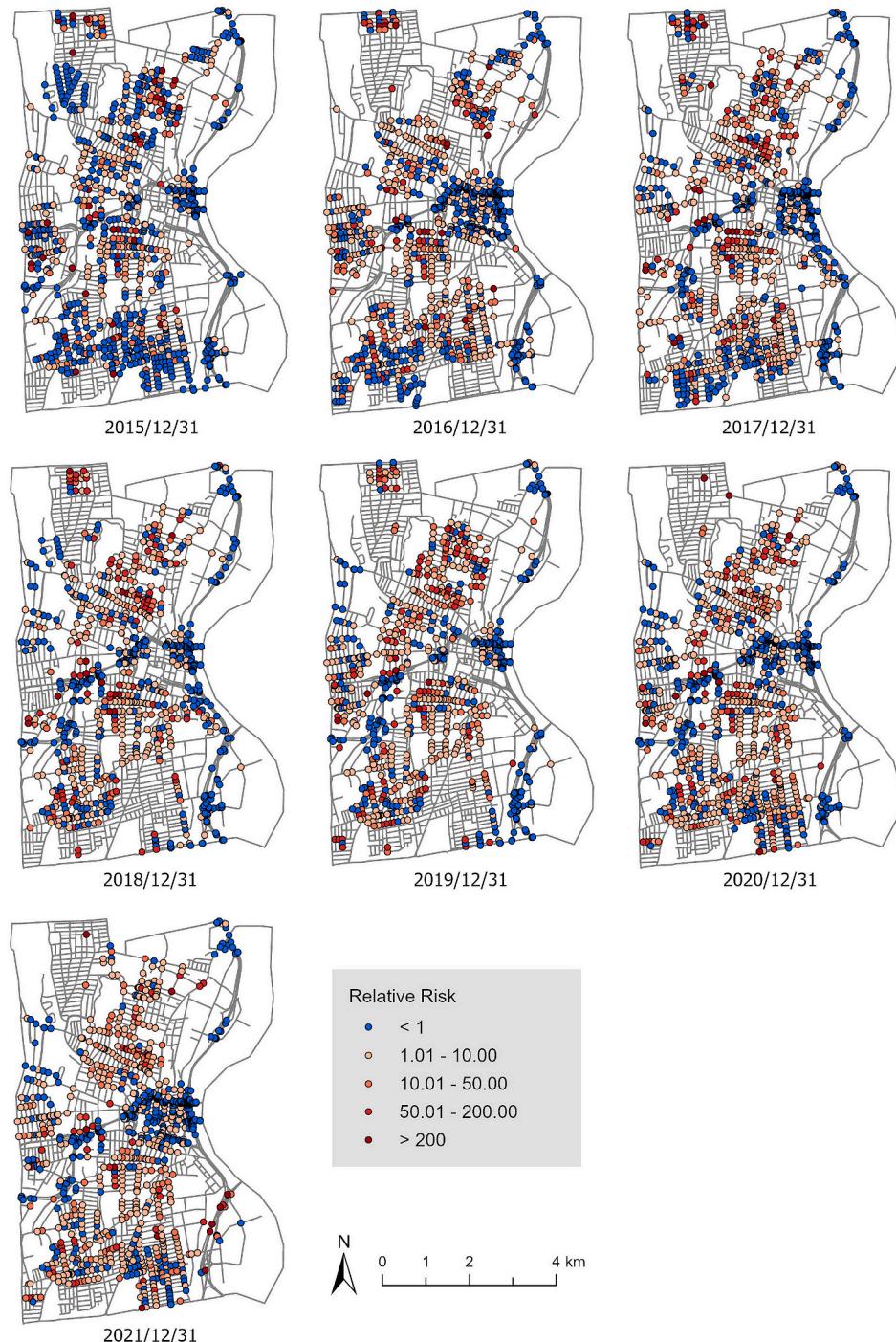


Fig. 4. High-risk intersections (four levels of red dots) and low-risk intersections (blue dots) in seven periods, where each scan starts from 2015/1/1 and ends on the given date. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

emerging clusters expands, the effectiveness of the method becomes less reliable. For example, we observed that high-risk clusters may contain low-risk locations and vice versa. Thus, a careful selection of the upper bound of the spatial scanning window is key to the utilization of the method. In this study, we set this upper bound as 1.0 km. This variable, which represents the maximum search radius of the algorithm, needs to be appropriately evaluated by sensitivity analysis as it may not apply to another study area where the distribution patterns of crashes are different. For example, in areas with fewer crashes (e.g., rural areas), the upper bound of the spatial scanning window must be appropriately increased. Second, we used the Thiessen polygon as our spatial unit.

Polygons' shape and size are subject to road density. While Thiessen polygons have less variation in urban areas than in rural areas, this variation can lead to uncertainties in the risk distribution. Third, crash risk modeling is highly data-dependent. Data on the long-term observation of the traffic volume over the road network are not always available. In our case, road segments lacking the traffic volume were assigned a small AADT value, as these road segments were assumed to carry a low traffic volume. While this substitution is necessary for model implementation, in the future, more accurate data interpolation methods are needed to solve the missing data issue.

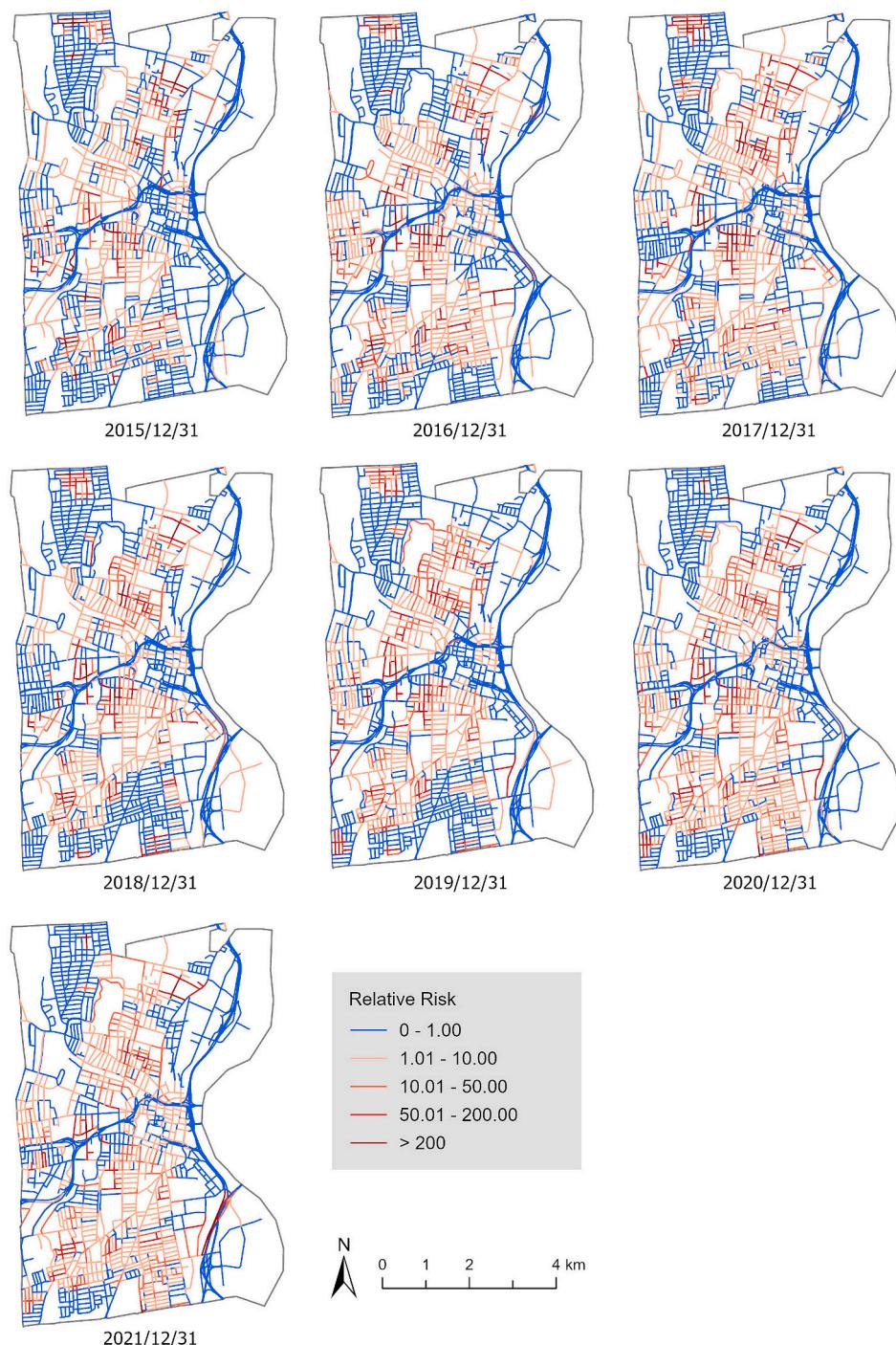


Fig. 5. Network-based high-risk roads (four levels of red segments), low-risk, and no statistically different risk roads (blue segments) in seven periods, where each scan starts from 2015/1/1 and ends on the given date. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6. Conclusions

In this paper, we have developed a new crash risk modeling method based on the prospective STSS method. Unlike traditional hot spot detection methods that require a pre-defined bandwidth for smoothing or a weight matrix representing the proximity between locations, the STSS method operates with less dependency on these arbitrary parameters. With more refined spatial and temporal traffic volume data, this method can be applied to periodic monitoring of traffic crash risk across the road network. Such applications can reveal the space-time patterns

of crash risk and can further support proactive traffic safety management.

Funding

Dr. Zhang is partially supported by USA NSF grants No. 2022036, 2118102 and 2303575. Authors have the sole responsibility for all of the viewpoints presented in the paper.

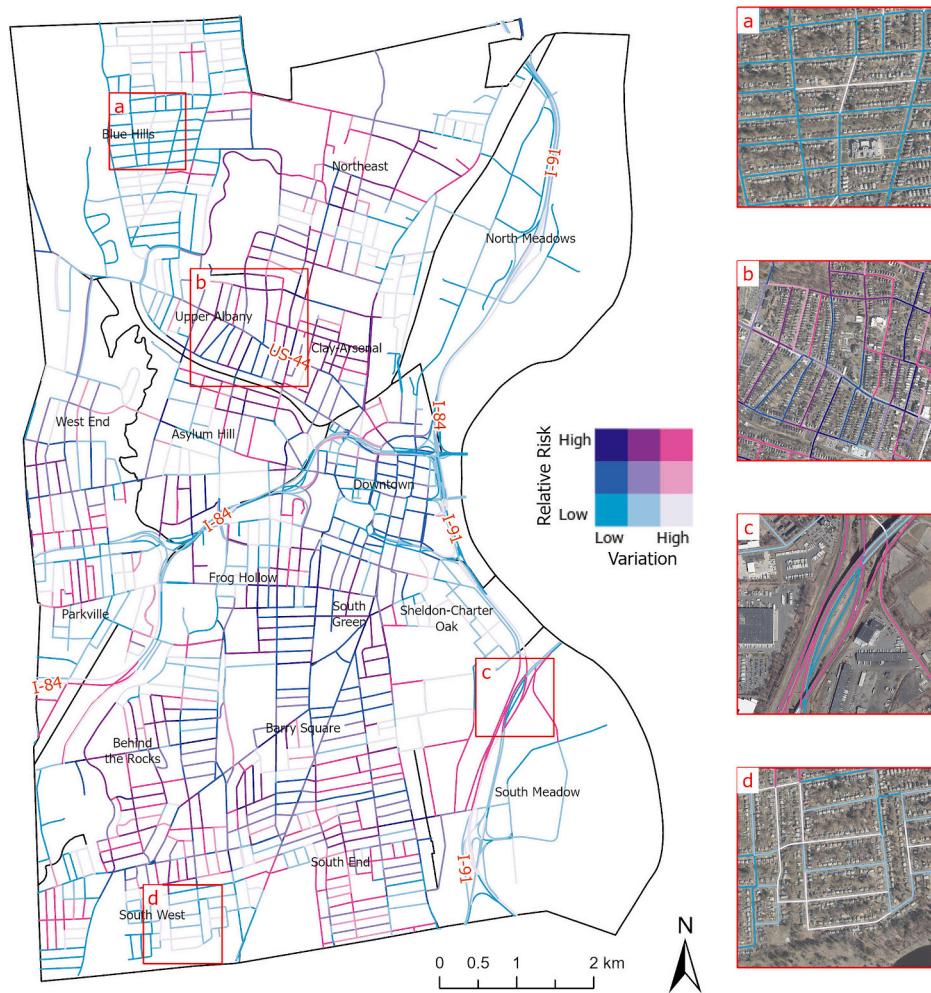


Fig. 6. Traffic crash relative risk by the end of 2021 and the coefficient of variation of relative risk from 2015 to 2021 with four featured neighborhoods: (a) low-risk, low-variation, (b) high-risk, low-variation, (c) high-risk, high-variation, (d) low-risk, high-variation.

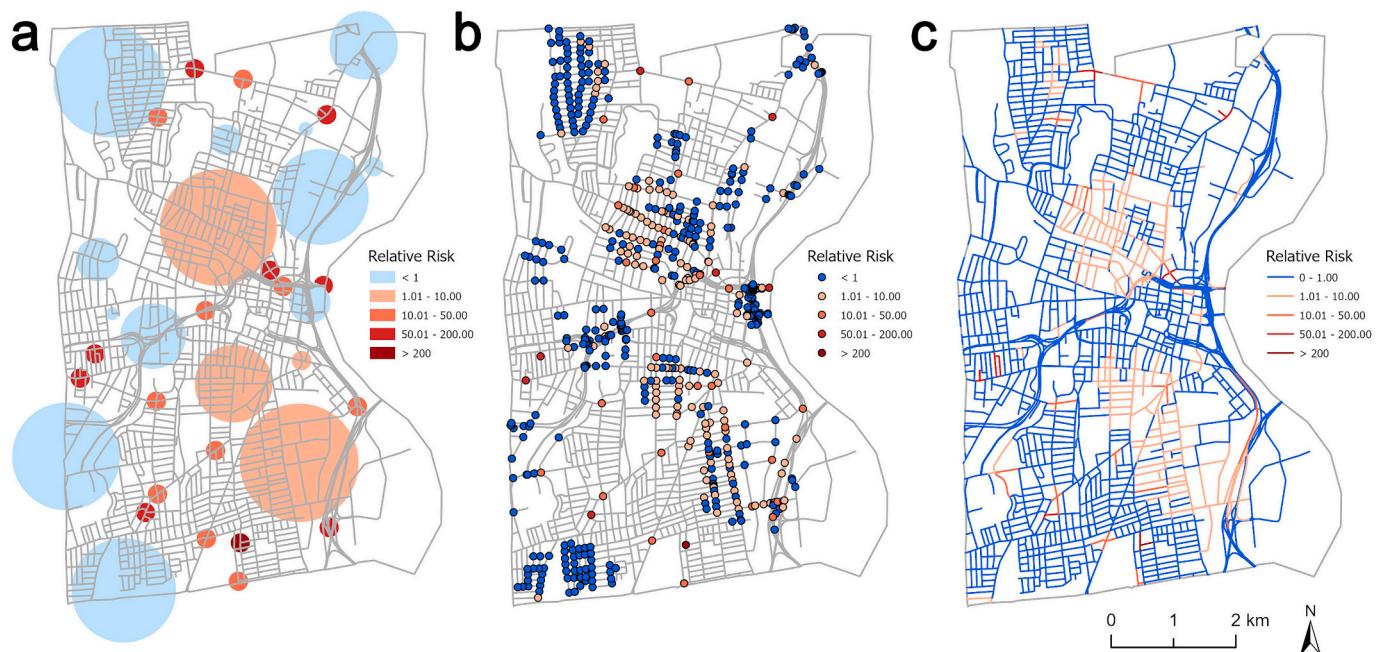


Fig. 7. Covariate-adjusted traffic crash risk in terms of (a) clusters, (b) intersections, and (c) networks.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

CRediT authorship contribution statement

Congcong Miao: Conceptualization, Methodology, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Xiang Chen:** Conceptualization, Writing – original draft, Writing – review & editing, Supervision. **Chuanrong Zhang:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors have no conflicts of interest to declare.

Data availability

Not applicable.

Acknowledgments

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jtrangeo.2024.103958>.

References

- AASHTO, 2010. Highway Safety Manual. American Association of State Highway and Transportation Officials. <https://books.google.com/books?id=M4fQiyfVRr8C>.
- Abdel-Aty, M., Wang, X., 2006. Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. *Transp. Res.* 40 (1), 98–111.
- Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. *Transp. Res. A Policy Pract.* 49, 62–75. <https://doi.org/10.1016/j.tra.2013.01.030>.
- Adams, A.M., Chen, X., Li, W., Zhang, C., 2023. Normalizing the pandemic: exploring the cartographic issues in state government COVID-19 dashboards. *J. Maps* 19 (1), 1–9. <https://doi.org/10.1080/17445647.2023.2235385>.
- Agüero-Valverde, J., Wu, K.-F., Ken, Donnell, E.T., 2016. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. *Accid. Anal. Prev.* 87, 8–16. <https://doi.org/10.1016/j.aap.2015.11.006>.
- Anderson, T., 2007. Comparison of spatial methods for measuring road accident ‘hotspots’: a case study of London. *J. Maps* 3 (1), 55–63. <https://doi.org/10.1080/jom.2007.9710827>.
- Anderson, T.K., 2009. Kernel density estimation and K-means clustering to profile road accident hotspots. *Accid. Anal. Prev.* 41 (3), 359–364. <https://doi.org/10.1016/j.aap.2008.12.014>.
- Anselin, L., 1995. Local indicators of spatial association-LISA. *Geogr. Anal.* 27 (2), 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>.
- Bil, M., Andrášik, R., Sedoník, J., 2019. A detailed spatiotemporal analysis of traffic crash hotspots. *Appl. Geogr.* 107, 82–90. <https://doi.org/10.1016/j.apgeog.2019.04.008>.
- Cai, Q., Abdel-Aty, M., Lee, J., Eluru, N., 2017. Comparative analysis of zonal systems for macro-level crash modeling. *J. Saf. Res.* 61, 157–166. <https://doi.org/10.1016/j.jsr.2017.02.018>.
- Chen, X., Kwan, M.-P., Li, Q., Chen, J., 2012. A model for evacuation risk assessment with consideration of pre- and post-disaster factors. *Comput. Environ. Urban Syst.* 36 (3), 207–217. <https://doi.org/10.1016/j.compenvurbsys.2011.11.002>.
- Chen, X., Huang, L., Dai, D., Zhu, M., Jin, K., 2018. Hotspots of road traffic crashes in a redeveloping area of Shanghai. *Int. J. Inj. Control Saf. Promot.* 25 (3), 293–302. <https://doi.org/10.1080/17457300.2018.1431938>.
- Cheng, T., Adepeju, M., 2014. Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PLoS One* 9 (6), e100465. <https://doi.org/10.1371/journal.pone.0100465>.
- Cheng, Z., Zu, Z., Lu, J., 2018. Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification of urban road intersections. *Sustainability* 11 (1), 160. <https://doi.org/10.3390/su11010160>.
- Church, R.L., Cova, T.J., 2000. Mapping evacuation risk on transportation networks using a spatial optimization model. *Transp. Res. Part C: Emerg. Technol.* 8 (1–6), 321–336. [https://doi.org/10.1016/S0968-090X\(00\)00019-X](https://doi.org/10.1016/S0968-090X(00)00019-X).
- City of Hartford, 2022. Zoning Regulations | Hartford, CT | Municode Library. https://library.municode.com/ct/hartford/codes/zoning_regulations.
- Cova, T.J., Church, R.L., 1997. Modelling community evacuation vulnerability using GIS. *Int. J. Geogr. Inf. Sci.* 11 (8), 763–784. <https://doi.org/10.1080/136588197242077>.
- CTSRC, 2020. Connecticut Roadway Safety Management System (CRSMS) Analytical Tool Suite User Manual.
- Dai, D., 2012. Identifying clusters and risk factors of injuries in pedestrian–vehicle crashes in a GIS environment. *J. Transp. Geogr.* 24, 206–214. <https://doi.org/10.1016/j.jtrangeo.2012.02.005>.
- Desjardins, M.R., Hohl, A., Delmelle, E.M., 2020. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. *Appl. Geogr.* 118, 102202. <https://doi.org/10.1016/j.apgeog.2020.102202>.
- Erdogan, S., 2009. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. *J. Saf. Res.* 40 (5), 341–351. <https://doi.org/10.1016/j.jsr.2009.07.006>.
- Fan, Y., Zhu, X., She, B., Guo, W., Guo, T., 2018. Network-constrained spatio-temporal clustering analysis of traffic collisions in Jianghan District of Wuhan, China. *PLoS One* 13 (4), e0195093. <https://doi.org/10.1371/journal.pone.0195093>.
- Gao, P., Guo, D., Liao, K., Webb, J.J., Cutler, S.L., 2013. Early detection of terrorism outbreaks using prospective space-time scan statistics. *Prof. Geogr.* 65 (4), 676–691. <https://doi.org/10.1080/00330124.2012.724348>.
- Haleem, K., Alluri, P., Gan, A., 2015. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accid. Anal. Prev.* 81, 14–23. <https://doi.org/10.1016/j.aap.2015.04.025>.
- Hashimoto, S., Yoshiki, S., Saeki, R., Mimura, Y., Ando, R., Nanba, S., 2016. Development and application of traffic accident density estimation models using kernel density estimation. *J. Traffic Transp. Eng. (English Edition)* 3 (3), 262–270. <https://doi.org/10.1016/j.jtte.2016.01.005>.
- Hazaymeh, K., Almagbile, A., Alomari, A.H., 2022. Spatiotemporal analysis of traffic accidents hotspots based on geospatial techniques. *ISPRS Int. J. Geo Inf.* 11 (4), 260. <https://doi.org/10.3390/ijgi11040260>.
- Hohl, A., Delmelle, E.M., Desjardins, M.R., Lan, Y., 2020. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spatial Spatio-Temp. Epidemiol.* 34, 100354. <https://doi.org/10.1016/j.sste.2020.100354>.
- Huang, Y., Wang, X., Patton, D., 2018. Examining spatial relationships between crashes and the built environment: a geographically weighted regression approach. *J. Transp. Geogr.* 69, 221–233. <https://doi.org/10.1016/j.jtrangeo.2018.04.027>.
- Kang, Y., Cho, N., Son, S., 2018. Spatiotemporal characteristics of elderly population’s traffic accidents in Seoul using space-time cube and space-time kernel density estimation. *PLoS One* 13 (5), e0196845. <https://doi.org/10.1371/journal.pone.0196845>.
- Kazmi, S.S.A., Ahmed, M., Mumtaz, R., Anwar, Z., 2022. Spatiotemporal clustering and analysis of road accident hotspots by exploiting GIS technology and kernel density estimation. *Comput. J.* 65 (2), 155–176. <https://doi.org/10.1093/comjnl/bxz158>.
- Kleinman, K., 2004. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *Am. J. Epidemiol.* 159 (3), 217–224. <https://doi.org/10.1093/aje/kwh029>.
- Kulldorff, M., 2001. Prospective time periodic geographical disease surveillance using a scan statistic. *J. R. Stat. Soc. A. Stat. Soc.* 164 (1), 61–72. <https://doi.org/10.1111/1467-985X.00186>.
- Kulldorff, M., 2022. SatScan Users Guide (10.1) [Computer Software]. <https://www.satscan.org/>.
- Kulldorff, M., Athas, W.F., Feurer, E.J., Miller, B.A., Key, C.R., 1998. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *Am. J. Public Health* 88 (9), 1377–1380. <https://doi.org/10.2105/AJPH.88.9.1377>.
- Liu, Y., Wang, S., Fu, X., Xie, B., 2019. A network-constrained spatial identification of high-risk roads for hit-parked-vehicle collisions in Brisbane, Australia. *Environ. Plan. A: Econ. Space* 51 (2), 279–282. <https://doi.org/10.1177/0308518X18810531>.
- Loo, B.P.Y., Yao, S., Wu, J., 2011. Spatial point analysis of road crashes in Shanghai: A GIS-based network kernel density method. In: 2011 19th International Conference on Geoinformatics, pp. 1–6. <https://doi.org/10.1109/Geoinformatics.2011.5980938>.
- Ma, Y., Yin, F., Zhang, T., Zhou, X.A., Li, X., 2016. Selection of the maximum spatial cluster size of the spatial scan statistic by using the maximum clustering set-proportion statistic. *PLoS One* 11 (1), e0147918. <https://doi.org/10.1371/journal.pone.0147918>.
- Mannering, F., 2018. Temporal instability and the analysis of highway accident data. *Analyt. Methods Accident Res.* 17, 1–13. <https://doi.org/10.1016/j.amar.2017.10.002>.
- McCahill, C., Garrick, N.W., 2010a. Losing Hartford: transportation policy and the decline of an American city. In: 18th Annual Meeting of the Congress for the New Urbanism.
- McCahill, C.T., Garrick, N.W., 2010b. Influence of parking policy on built environment and travel behavior in two New England cities, 1960 to 2007. *Transp. Res. Record J. Transp. Res. Board* 2187 (1), 123–130. <https://doi.org/10.3141/2187-16>.

- Mitra, S., 2009. Spatial autocorrelation and Bayesian spatial statistical method for analyzing intersections prone to injury crashes. *Transp. Res. Record J. Transp. Res. Board* 2136 (1), 92–100. <https://doi.org/10.3141/2136-11>.
- Mohan, D., Jha, A., Chauhan, S.S., 2021. Future of road safety and SDG 3.6 goals in six Indian cities. *IATSS Res.* 45 (1), 12–18. <https://doi.org/10.1016/j.iatssr.2021.01.004>.
- Mohaymany, A.S., Shahri, M., Mirbagheri, B., 2013. GIS-based method for detecting high-crash-risk road segments using network kernel density estimation. *Geo-spat. Inf. Sci.* 16 (2), 113–119. <https://doi.org/10.1080/10095020.2013.766396>.
- Ng, K., Hung, W., Wong, W., 2002. An algorithm for assessing the risk of traffic accident. *J. Saf. Res.* 33 (3), 387–410. [https://doi.org/10.1016/S0022-4375\(02\)00033-6](https://doi.org/10.1016/S0022-4375(02)00033-6).
- Nie, K., Wang, Z., Du, Q., Ren, F., Tian, Q., 2015. A network-constrained integrated method for detecting spatial cluster and risk location of traffic crash: a case study from Wuhan, China. *Sustainability* 7 (3), 2662–2677. <https://doi.org/10.3390/su7032662>.
- Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geogr. Anal.* 27 (4), 286–306. <https://doi.org/10.1111/j.1538-4632.1995.tb00912.x>.
- Ouni, F., Belloumi, M., 2018. Spatio-temporal pattern of vulnerable road user's collisions hot spots and related risk factors for injury severity in Tunisia. *Transport. Res. F: Traffic Psychol. Behav.* 56, 477–495. <https://doi.org/10.1016/j.trf.2018.05.003>.
- Ouyang, Y., Bejleri, I., 2014. Geographic information system-based community-level method to evaluate the influence of built environment on traffic crashes. *Transp. Res. Record J. Transp. Res. Board* 2432 (1), 124–132. <https://doi.org/10.3141/2432-15>.
- Özcan, M., Küçükönder, M., 2020. Investigation of spatiotemporal changes in the incidence of traffic accidents in Kahramanmaraş, Turkey, using GIS-based density analysis. *J. Indian Soc. Remote Sens.* 48 (7), 1045–1056. <https://doi.org/10.1007/s12524-020-01137-0>.
- Papadimitriou, E., Filtness, A., Theofilatos, A., Ziakopoulos, A., Quigley, C., Yannis, G., 2019. Review and ranking of crash risk factors related to the road infrastructure. *Accid. Anal. Prev.* 125, 85–97. <https://doi.org/10.1016/j.aap.2019.01.002>.
- Patel, B.N., McGrane, J., Burnham, K.J., Habesch, N.O., Rimiller, J.H., Allu, P.K., Bertoli, J.L., Hath, G., Correia, C., Barrett, J., Patel, M., 2005. *The Development of Hartford's Neighborhood Traffic Calming Master Plan*. Department of Public Works.
- Pulugurtha, S.S., Duddu, V.R., Kotagiri, Y., 2013. Traffic analysis zone level crash estimation models based on land use characteristics. *Accid. Anal. Prev.* 50, 678–687. <https://doi.org/10.1016/j.aap.2012.06.016>.
- Rankavat, S., Tiwari, G., 2013. *Pedestrian Accident Analysis in Delhi Using GIS*.
- Rausand, M., 2013. *Risk Assessment: Theory, Methods, and Applications*, vol. 115. John Wiley & Sons.
- Schneider, R.J., Diogenes, M.C., Arnold, L.S., Attaset, V., Griswold, J., Ragland, D.R., 2010. Association between roadway intersection characteristics and pedestrian crash risk in Alameda County, California. *Transp. Res. Record J. Transp. Res. Board* 2198 (1), 41–51. <https://doi.org/10.3141/2198-06>.
- Shahzad, M., 2020. Review of road accident analysis using GIS technique. *Int. J. Inj. Control Saf. Promot.* 27 (4), 472–481. <https://doi.org/10.1080/17457300.2020.1811732>.
- Sharma, S.L., Datta, T.K., 2007. Investigation of regression-to-mean effect in traffic safety evaluation methodologies. *Transp. Res. Record J. Transp. Res. Board* 2019 (1), 32–39. <https://doi.org/10.3141/2019-05>.
- Soltani, A., Askari, S., 2017. Exploring spatial autocorrelation of traffic crashes based on severity. *Injury* 48 (3), 637–647. <https://doi.org/10.1016/j.injury.2017.01.032>.
- Song, J., Wen, R., Yan, W., 2018. Identification of traffic accident clusters using Kulldorff's space-time scan statistics. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 3162–3167. <https://doi.org/10.1109/BigData.2018.8622226>.
- Steenberghen, T., Aerts, K., Thomas, I., 2010. Spatial clustering of events on a network. *J. Transp. Geogr.* 18 (3), 411–418. <https://doi.org/10.1016/j.jtrangeo.2009.08.005>.
- Stoneburner, G., Goguen, A., Feringa, A., 2002. Risk management guide for information technology systems: recommendations of the National Institute of Standards and Technology (NIST SP 800-30; 0 ed., p. NIST SP 800-30). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.800-30>.
- Takahashi, K., Kulldorff, M., Tango, T., Yih, K., 2008. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *Int. J. Health Geogr.* 7 (1), 14. <https://doi.org/10.1186/1476-072X-7-14>.
- Tortum, A., Atalay, A., 2015. Spatial analysis of road mortality rates in Turkey. *Proc. Inst. Civ. Eng. Transp.* 168 (6), 532–542. <https://doi.org/10.1680/jtran.14.00029>.
- United Nations, 2020. The Sustainable Development Goals Report 2020. United Nations. <https://www.un-ilibrary.org/content/books/9789210049603>.
- Vicente Ferreira, R., Roberto Martines, M., Hartung Toppa, R., Maria de Assunção, L., Richard Desjardins, M., Delmelle, E., 2022. Utilizing prospective space-time scan statistics to discover the dynamics of coronavirus disease 2019 clusters in the state of São Paulo, Brazil. *Rev. Soc. Bras. Med. Trop.* <https://doi.org/10.1590/0037-8682-0607-2021>.
- WHO, 2023. Global Status Report on Road Safety 2023. World Health Organization. <https://iris.who.int/bitstream/handle/10665/375016/9789240086517-eng.pdf?sequence=1>.
- WHO, 2023b. *World Health Statistics 2023 – Monitoring Health for the SDGs*.
- Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accid. Anal. Prev.* 41 (1), 137–145. <https://doi.org/10.1016/j.aap.2008.10.001>.
- Wu, P., Meng, X., Song, L., 2022. Identification and spatiotemporal evolution analysis of high-risk crash spots in urban roads at the microzone-level: using the space-time cube method. *J. Transp. Saf. Secur.* 14 (9), 1510–1530. <https://doi.org/10.1080/19439962.2021.1938323>.
- Xie, Z., Yan, J., 2008. Kernel density estimation of traffic accidents in a network space. *Comput. Environ. Urban. Syst.* 32 (5), 396–406. <https://doi.org/10.1016/j.compenvurbsys.2008.05.001>.
- Xu, F., Beard, K., 2021. A comparison of prospective space-time scan statistics and spatiotemporal event sequence based clustering for COVID-19 surveillance. *PLoS One* 16 (6), e0252990. <https://doi.org/10.1371/journal.pone.0252990>.
- Yamada, I., Thill, J.-C., 2007. Local indicators of network-constrained clusters in spatial point patterns. *Geogr. Anal.* 39 (3), 268–292. <https://doi.org/10.1111/j.1538-4632.2007.00704.x>.
- Yamada, I., Thill, J.-C., 2010. Local indicators of network-constrained clusters in spatial patterns represented by a link attribute. *Ann. Assoc. Am. Geogr.* 100 (2), 269–285. <https://doi.org/10.1080/00045600903550337>.
- Yao, S., Loo, B.P.Y., Yang, B.Z., 2016. Traffic collisions in space: four decades of advancement in applied GIS. *Ann. GIS* 22 (1), 1–14. <https://doi.org/10.1080/19475683.2015.1085440>.
- Yoon, J., Lee, S., 2021. Spatio-temporal patterns in pedestrian crashes and their determining factors: application of a space-time cube analysis model. *Accid. Anal. Prev.* 161, 106291. <https://doi.org/10.1016/j.aap.2021.106291>.
- Young, J., Park, P.Y., 2014. Hotzone identification with GIS-based post-network screening analysis. *J. Transp. Geogr.* 34, 106–120. <https://doi.org/10.1016/j.jtrangeo.2013.11.007>.
- Ziakopoulos, A., Yannis, G., 2020. A review of spatial approaches in road safety. *Accid. Anal. Prev.* 135, 105323. <https://doi.org/10.1016/j.aap.2019.105323>.