Article

Analysis of the Spatial Variation of Network-Constrained Phenomena Represented by a Link Attribute Using a Hierarchical Bayesian Model

Zhensheng Wang [1,2,3,4], Yang Yue [1,2,*], Qingquan Li [1,2], Ke Nie [4] and Changbin Yu [5]

[1] Shenzhen Key Laboratory of Spatial Smart Sensing and Services, College of Civil Engineering,
    Shenzhen University, Shenzhen 518060, China; wangzhensheng@szu.edu.cn (Z.W.); liqq@szu.edu.cn (Q.L.)
[2] Key Laboratory for Geo-Environmental Monitoring of Coastal Zone of the National Administration of
    Surveying, Mapping and GeoInformation, Shenzhen University, Shenzhen 518060, China
[3] College of Information Engineering, Shenzhen University, Shenzhen 518060, China
[4] Key Laboratory of Urban land Resources Monitoring and Simulation, Ministry of Land and Resources,
    Shenzhen 518034, China; nieke@whu.edu.cn
[5] School of Management, Ningbo Institute of Technology, Zhejiang University, Ningbo 315010, China;
    yuchangbinboy@gmail.com
* Correspondence: yueyang@szu.edu.cn; Tel.: +86-755-2697-9741

Abstract: The spatial variation of geographical phenomena is a classical problem in spatial data analysis and can provide insight into underlying processes. Traditional exploratory methods mostly depend on the planar distance assumption, but many spatial phenomena are constrained to a subset of Euclidean space. In this study, we apply a method based on a hierarchical Bayesian model to analyse the spatial variation of network-constrained phenomena represented by a link attribute in conjunction with two experiments based on a simplified hypothetical network and a complex road network in Shenzhen that includes 4212 urban facility points of interest (POIs) for leisure activities. Then, the methods named local indicators of network-constrained clusters (LINCS) are applied to explore local spatial patterns in the given network space. The proposed method is designed for phenomena that are represented by attribute values of network links and is capable of removing part of random variability resulting from small-sample estimation. The effects of spatial dependence and the base distribution are also considered in the proposed method, which could be applied in the fields of urban planning and safety research.

1. Introduction

Spatial variation is a classic problem in spatial data analysis and can provide insight into the spatial patterns of geographic phenomena and spatial processes. In traditional analysis methods, it is generally assumed that the spatial events can be located stochastically on a plane, and the spatial association between event locations or sub-areas is analysed using the Euclidean (or planar) distance [1–3], in which the inherent spatial processes are quantified based the assumption of Euclidean geometry [4]. However, this assumption is not appropriate when a spatial phenomenon is apparently constrained to a subset of geographical space, such as a street network. In the real world, there are many events, the existence of which are strongly restricted by networks, such as vehicle crashes on roads, retail services alongside streets, street crimes, and many others. These events are termed network-constrained events or network events for short [4–6]. Euclidean-based methods, which are

designed for events occurring on a continuous plane, are likely to lead to biased conclusions when analysing the network events. Therefore, network spatial analysis has been developed in the past two decades and is widely applied in analysing network-constrained spatial phenomena [7–15].

For spatial phenomena consisting of network events, it is obvious that the Euclidean distance assumption is often violated because the network and Euclidean spaces are not isomorphic and have their own intrinsic properties [16]. For example, vehicle crashes and traffic violations are more likely to occur on the road networks than in other locations within a planar space. In addition, the Euclidean distance might not be an appropriate measure of the spatial separation of events [16] and is inappropriate for analysing point events constrained in a network space [17]. Spatially close point events in the Euclidean space could be far apart from each other in the network space when accounting for connectivity. For instance, bus stations are designed based on a pre-defined road network distance rather than the Euclidean distance. However, in some circumstances, the Euclidean distance assumption may be appropriate for analysing network-constrained phenomena. For instance, the association between traffic-related air pollution exposure and disease mortality at the individual level is estimated using a Euclidean distance-based buffer in epidemiology.

In planar spatial analysis, geo-referenced data can be viewed either as point events (e.g., vehicle crashes, crimes, or residential locations) or as spatial units with attribute values (e.g., population or disease rates) [18]. Analogously, a network-constrained phenomenon can be represented either as a set of points distributed over links of the network or as a set of attribute values assigned to the links [5,16], which drives the need for different analysis methods. These statistical methods have been developed using two commonly used approaches in spatial analysis, namely, the attribute-based approach [5,13,16] and the event-based approach [5,6]. For the former, spatial events are not analysed directly but are assigned to a road network when the exact locations of discrete events are not of interest or available. The aggregated counts of events are treated as attribute values of the links, and this method is also termed the link-attribute-based approach [16]. For the latter, the physical locations of discrete events are analysed directly. This article introduces a Bayesian-based method to analyse spatial variation of network-constrained phenomena that are represented by the link-attribute approach. The motivation for this method is as follows.

An observed spatial pattern of a phenomenon could merely result from spatial variation of the base distribution, and this possibility should be considered in network spatial analysis. For instance, the spatial variation of a disease largely depends on the distribution of the population at risk. Therefore, the procedure of standardization is usually applied to calculate a ratio, such as the relative risk [19], between the observed event counts and the expected event counts, which are determined using the base distribution. In addition, the links are usually divided into shorter segments called basic spatial units (BSUs) in the link-attribute approach [13–15], which enables the detection of spatial patterns at a much finer spatial resolution than that imposed by a given network [16]. However, analyses based on fine spatial units with small base values introduce an extra source of variability into the analysis because of random variation [20]. Typically, links with few (or zero) cases can generate extreme values of the ratios, as the variance of the standardized value is inversely related to the expected case count, and the small size of the base distribution will cause large variability in the estimated results.

The paper introduces a method based on hierarchical Bayesian models to analyse the spatial variability of network-constrained phenomena by considering the random variation resulting from small sample estimation in conjunction with two experiments based on a simplified hypothetical network and a complex road network in Shenzhen that includes 4212 urban facility points of interest (POIs) for leisure activities. For large Chinese cities that are undergoing fast economic growth and rapid urban development, there is an urgent need to study spatial distribution characteristics of urban facilities in order to better understand the urban structure and patterns of human mobility. In a GIS (geographical information system) environment, urban facility POIs, just like many other geographical phenomena including car crashes on a road, crimes, and disease outbreak sites, can be abstracted as points for spatial analysis. Points can be either specific geographical entities or locations of past events.

Based on the spatial location of these point events, spatial analysis has been widely applied to study the characteristics of global or local spatial distributions [16]. The remainder of this paper is organized as follows. A literature review is conducted in Section 2. Section 3 introduces the framework of the proposed method and two experiments. Section 4 reports the details of the application, and the results and comparisons using data on urban facility POIs. The last section concludes the study.

2. Literature Review

Originally, planar spatial methods were directly applied to measure the spatial variation of network events [12,21] and the network autocorrelation of traffic crashes was first examined by using global spatial statistics [22]. In recent years, the primary concern of spatial analysis has been to explore the nonstationarity of spatial processes and to detect specific concentrations using local spatial statistics. In [12], local indicators of spatial association (LISA) [23] were applied to identify traffic hot zones. However, the application of planar methods to network-constrained phenomena can lead to systematic bias and improper pattern inferences [21]. Therefore, many researchers have made significant progress by extending the planar methods to network-based methods. For instance, two popular event-based methods, planar kernel density estimation (KDE) and planar K-function methods have been extended to the network KDE [4,6,14,15] and the network K-function [4,5,11,24] methods, respectively. Using the link-attribute approach, an exploratory methodology named local indicators of network-constrained clusters (LINCS) was introduced to detect the local scale clustering of network events [16]. Two types of LINCS methods, namely ILINCS and GLINCS, are the network extension of the local Moran's I and the local Getis-Ord G statistics, respectively.

With respect to handling background variations in the base distribution, a radical approach is to regenerate the spatial process of event occurrence itself under the null hypothesis of no spatial pattern [4]. There are two alternative assumptions that can be used in this simulation [16]. In the first approach, the probability of each spatial unit in the base distribution being an event is constant over the network and the number of observed events for each link is assumed to follow a Poisson distribution [19]. The second approach assumes the probability of an event occurring for a link is proportional to the link's base value, which can be simulated using a binomial-based null distribution, such as the binomial distribution, the multinomial distribution or the negative binomial distribution [25].

As discussed earlier, the random variability that results from small sample estimation should be considered in network-based spatial analysis. One significant merit of the Bayesian approach is its ability to generate robust estimates in the presence of sparse data or rare events [25], which was first applied in the area of disease mapping [26]. The basic principle of Bayesian methods is that uncertain data can be strengthened by combining them with prior information [25], which makes Bayesian inference an attractive estimation method in many fields. In disease mapping, Bayesian statistics have been applied to remove part of the random component from the map to produce smoothed estimates of relative risk in each area by considering neighbourhood relationships in the data [19,27]. Methods based on Bayesian statistics have also been implemented in other fields that employ small-area spatial analysis, such as crime [28], air pollution [29], and traffic crashes [30]. However, the application of Bayesian approaches to explore the spatial variation of network-constrained phenomena is still limited in terms of the number of existing studies.

It is recognized that Bayesian estimation represents a trade-off between improved precision and the introduction of bias [31]. In the case of empirical Bayesian disease mapping, the posterior estimates of spatially varying disease risk are evaluated based on a weighted combination of two components, namely the local risk and prior information from the surrounding areas [25,27]. The relationship between the two components depends on the population size in the local area. The smoothed risk, which can be estimated using a prior based on the global mean of the neighbours, is more stable and has less uncertainty. In addition, the prior can be arranged as a hierarchical or multilevel structure, which can simplify the estimation of parameter distributions such as the convolution prior. However,

because empirical Bayesian methods tend to oversmooth towards the summarized data from the neighbouring areas, two types of Bayesian methods have been developed to address the limitations of empirical estimates, namely hierarchical Bayesian models [26] and the BYM model introduced by Besag, York and Mollié [32].

The use of hierarchical models estimated in a Bayesian framework to account for different levels of variability of spatial data has been well established in recent years. Hierarchical Bayesian models were originally developed in the field of image analysis, and have subsequently been widely used in disease mapping and ecological studies. In conventional spatial statistics, the results of a regression model explain only a small amount of variance [2,25]. However, in hierarchical Bayesian models, the unexplained "extra variance" [25,26] is represented by either the spatially correlated effects or spatial heterogeneity effects [27,31,33]. Unlike other Bayesian models, spatial dependence is considered in hierarchical Bayesian models. The parameter estimates for a given spatial unit are obtained by "borrowing" strength from neighbouring spatial units [19,31]. In addition, the term "hierarchical" indicates that observed outcomes are modelled conditionally on a set of parameters that are themselves given a probabilistic value in terms of other parameters, which are defined as hyperparameters in Bayesian statistics. In Bayesian inference, posterior estimates can be produced from a weighted combination of the local estimates (also called the likelihood) and the estimates in surrounding spatial units. Empirical Bayesian estimates are inexact and tend to oversmooth towards the global mean [31], whereas in full Bayesian methods, the hyperparameters have hyperprior distributions. As a result, the estimates for each spatial unit better approximate the true value [32].

This study attempts to develop a Bayesian method for analysing the spatial variation in network-constrained phenomena. The high uncertainty and random variation of link-attribute-based data are eliminated by applying a hierarchical Bayesian model to network space. To examine the performance of the Bayesian method proposed here, a simulation study-based hypothetical network and a case study are conducted with the urban facility POI data in Shenzhen, China, respectively. Moreover, we apply two types of LINCS methods to detect the local-scale clusters.

3. Materials and Methods

3.1. Hierarchical Bayesian Models for Network-Constrained Data

In this study, we focus on tangible networks such as road networks and river networks. Links in the network are defined as the edges between two intersections and can be split into BSUs. Let us consider a network that consists of n BSUs, and let y_i be a count value of interest observed at BSU $i(i = 1, \ldots, I)$. As mentioned above, the effects of the base distribution should be considered when measuring the spatial variation of network-constrained phenomena. In this study, we consider a Poisson model for the counts and formulate the Bayesian spatial models within a Poisson framework with a logistic link. Extending this model to the binomial case is straightforward. In the hierarchical Bayesian framework that we consider, the Poisson likelihood of the observed counts is the first level of the model, which is used for modelling the within-segment variability of the event counts conditional on unknown risk parameters. The prior distributions of these parameters are specified at the second level of the model where the spatial dependence is also measured.

In the first level, the likelihood model assumes that the observed event counts y_i for each BSU follow a Poisson distribution centred on π_i to capture the within-segment variability of the counts:

$$y_i \sim Poisson(\pi_i) \tag{1}$$

Therefore, π_i is an estimate of the true number of events in BSU i, which could be calculated by:

$$\pi_i = E_i \times R_i \tag{2}$$

where R_i is the ratio between the observed event counts y_i and the expected event counts E_i for BSU i, which could be termed the relative risk. E_i is given by:

$$E_i = \frac{\sum_{j=1}^{I} y_j}{\sum_{j=1}^{I} n_j} \times n_i, i = 1, \ldots, I \tag{3}$$

$$R_i = \frac{y_i}{E_i} R_i = \frac{y_i}{E_i} \tag{4}$$

where n_i is the total number of events for BSU i. It is important to identify a reasonable base distribution which is largely dependent on the phenomena of interest. For instance, when shopping facilities or traffic crashes are the research subjects, the number of all types of POIs or traffic volumes can be viewed as the corresponding base distribution. Within a Bayesian regression context, the relative risk in BSU i can be parameterized as a function of a series of explanatory variables [31]. At the second level of the model, we split the ratio R_i on a logarithmic scale into an overall intercept α and main spatial effects S_i, which is assumed to be approximately normally distributed:

$$\log(R_i) = \alpha + S_i \tag{5}$$

Therefore, the estimate of the true number of events for BSU i is given by

$$\log(\pi_i) = \log(E_i) + \alpha + S_i \tag{6}$$

A noninformative prior distribution (flat distribution) is given for the intercept term α. The spatial dependence is represented by means of a spatial weights matrix that defines a set of spatial neighbours δ_i for each unit i. A weight matrix $\mathbf{W} = (w_{ij})$ is then defined to measure the proximity between segments in the given network. In the simplest case, $w_{ij} = 1$ if segment i and j share a common node, and is 0 otherwise, which is denoted in this paper by $i \sim j$; it might also be defined by the network distance between the midpoints of segments. Although the weight matrix in network space can be complex in some situations, we consider the above two types of spatial neighbourhoods because of their much broader applications in the field. The spatial dependence is modelled using a conditional autoregressive process (CAR) [19]. Given a matrix $\mathbf{W}$, the conditional distribution of a set of parameters μ_i is specified by:

$$p(\mu_i | \mu_j, j \neq i) \sim N(\overline{\mu_i}, \sigma_\varepsilon^2 / k_i) \tag{7}$$

$$\overline{\mu_i} = \frac{1}{k_i} \sum_{i \sim j} \mu_j \tag{8}$$

where σ_ε^2 is an unknown variance parameter, k_i is the number of neighbours of segment i, $\boldsymbol{\mu} = (\mu_1, .., \mu_I)$ denotes the random effects in Bayesian spatial models, and $\overline{\mu_i}$ is the conditional expectation of μ_i. Thus, the value of a parameter associated with segment i is affected by the average value of its neighbours with additional variability. The variance parameter σ_ε^2 controls the amount of variation between the random effects. This variance structure recognizes the fact that in the presence of strong spatial correlation, the more neighbors a unit has the more information there is in the data about the value of its random effect. However, a strong spatial dependence is defined by the CAR process, which has only one free parameter linked to the conditional variance σ_ε^2. The main spatial effect can be divided into two parts: a spatial unstructured effect and a spatial structured effect, which indicate the spatial heterogeneity and spatial dependence, respectively [19]. Thus, to increase the method's flexibility, we apply a convolution BYM model to combine the CAR process (spatial structured effect) with an unstructured exchangeable normal component (spatial unstructured effect). The resulting model can be written as:

$$S_i \sim N(\mu_i, \sigma_v^2), i = 1, \ldots, I \tag{9}$$

$$\mu_i \sim CAR(\mathbf{W}, \sigma_\varepsilon^2) \tag{10}$$

where σ_ν^2 is variance of the unstructured component. We further define a third level of the model so that the variation parameters that are involved in the second level (Equations (9) and (10)) are themselves treated as unknown and given hyperprior distributions. For the hierarchical standard deviation, we specify a uniform distribution on the interval (0,100) because this range is wide enough to cover any realistic value in log-transformed modelling [19].

Bayesian inference is based on the joint distribution of all parameters, which was considered difficult and intractable in the past. In this study, the posterior means of all parameters were estimated using Markov chain Monte Carlo (MCMC) algorithms. We used the free software WinBUGS, which is based on MCMC algorithms, to implement the model [34]. The deviance information criterion (DIC) is applied to evaluate the goodness of fit of the model; if the differences in DIC are greater than 5, the model with the lowest DIC is selected as best model [35]. The DIC, which is a hierarchical modelling generalization of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), is particularly helpful in Bayesian model selection problems where the posterior distributions of the models are acquired based on MCMC simulation [19].

### 3.2. The ILINCS and GLINCS Approaches

Based on the planar Moran's I statistic, the network autocorrelation analysis modifies the spatial weight matrix to reflect the network connectivity between links. Like the local I statistic, the local G statistic can be applied to analyse network-constrained phenomena by modifying the weight matrix [13,15]. The local I statistic aims to determine the autocorrelation between a region and its neighbours; however, the local G statistic measures the concentration of attributes of a variable around a region [36]. In [37], the authors revised the local G statistics to allow the variable to be nonpositive and the weight matrix to be nonbinary. There are two versions of the local G statistic, namely $G_i(d)$ and $G_i^*(d)$. The only difference between the two statistics is whether the unit i is itself included or not. In [16], the authors recommended performing analyses with the local I and G* statistics simultaneously because the two statistics are complementary to each other.

The Moran's I and Getis-Ord G statistics follow an asymptotic normal distribution when either the normality assumption or the randomization assumption holds [16,20]. However, in the context of local-scale cluster detection in a network space, the randomization assumption is preferred over the normality assumption because each link is usually connected to a relatively small number of other links [23]. To measure the null distributions of the local statistics, statistical inference based on Monte Carlo simulation was recommended in previous studies [5,13,23]. In this study, we applied the ILINCS and GLINCS methods to explore local spatial patterns on a network, which incorporates a Monte Carlo simulation to assess the statistical significance of detected clusters.

3.3. POI Data and Analysis Design

To illustrate the validity of the proposed method, two experiments are carried out in this paper. The first experiment involves a hypothetical network, and the second experiment uses data from Shenzhen city. In the first experiment, the methodology is used to analyse the variations of repeated simulated link-attribute phenomena in a simplified hypothetical network. The second experiment uses a road network system and urban facility POI data for 2013 from Futian District, an advanced region of Shenzhen city (Figure 1). Compared to other large cities in China, Shenzhen is a young immigrant city with rapid economic growth because of its high-tech industries and finance. However, because it is a thriving city with a high population density, it is important to study the distribution of urban facilities throughout the road network, and this need has drawn the attention of policymakers and scholars [38]. We chose a type of urban facility POIs which are most likely to be located on road networks, such as hotels, restaurants, pubs, cinemas and art galleries, which cover numerous leisure

activities for urban residents. The number of these facilities is 4212 and the total number of all types of POIs in the study region is regarded as the base distribution.
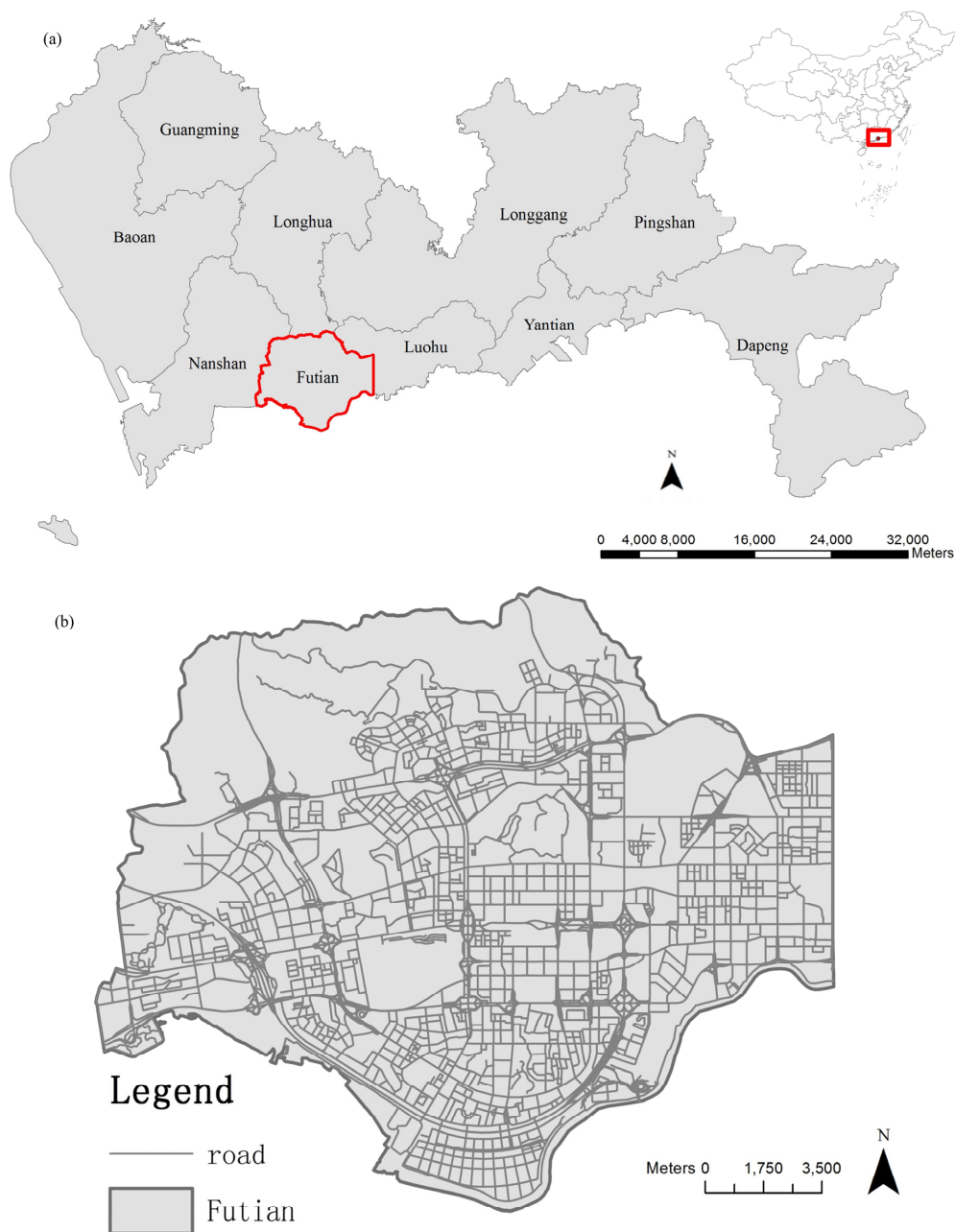
Figure 1. Shenzhen city and the study region: (**a**) the position of the study region (red box); (**b**) the road network in the study area.

To apply the LINCS and the proposed methods, it is necessary to convert the event-based POIs into link-attribute-based data. We apply the following procedure in this study. Routes are first split into shorter segments at reference points using a network segmentation algorithm [13,16]. Taking 1 km as the standard BSU length, the above splitting process results in 4372 BSUs for the 651 km of road network in the study area (Figure 2b). Secondly, we count the number of urban facilities along network segments and assign the resulting values to the BSUs as new attributes. To avoid edge effects, a facility that is located at the end nodes of more than one BSU is randomly assigned to one of the BSUs. Finally, in the LINCS methods, the allocated number of facility POIs for each BSU is standardized

according to the base distribution, which is approximated by the total number of POIs in the case study. By applying the Monte Carlo method, the local statistics are simulated to identify the spatial pattern in the given network [13,16]. The proposed method, which is based on a hierarchical Bayesian model, is implemented in WinBUGS. For each model in the hypothetical study, three parallel MCMC chains that each contain 10,000 MCMC iterations are simulated and visualized with time series plots and Gelman–Rubin statistics [39]. The posterior distributions of the unknown parameters are acquired after a burn-in of 1000 iterations. We apply a longer run for the case study (50,000 iterations after 10,000 iterations of burn-in), and posterior estimates are used as the inputs for calculating LINCS without adjustment for the base distribution. Figure 2 describes the spatial variation of the base distribution over the study region.
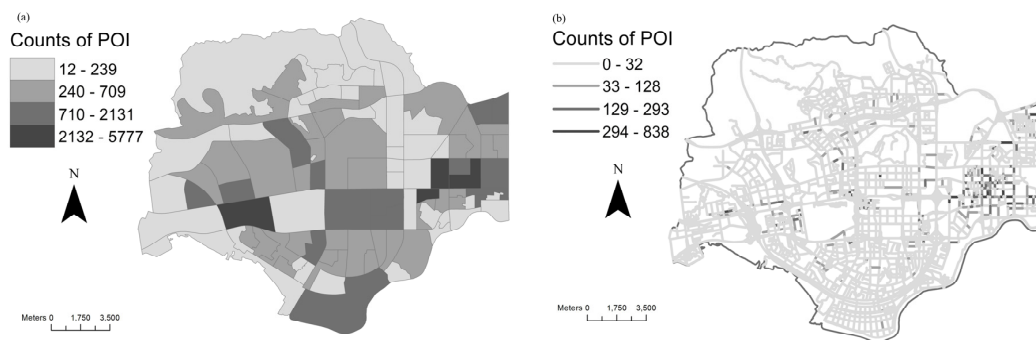

Figure 2. The base distribution in the study region: (**a**) the facility points of interest (POI) counts in 98 communities; (**b**) the POI counts for the 4372 basic spatial units (BSUs).

4. Results and Discussion

4.1. A Simplified Hypothetical Network

In this section, the proposed method is first tested by analysing a simulated network-constrained phenomenon represented by link attributes in a simplified hypothetical network. The length of the simplified road network is 10 km; using the standard BSU length of 200 m, the simplified network is split into 80 BSUs. Then, we randomly distribute 5000 events over the network to represent the base distribution. Figure 3 describes the simplified hypothetical network and the spatial distribution of simulated network events.
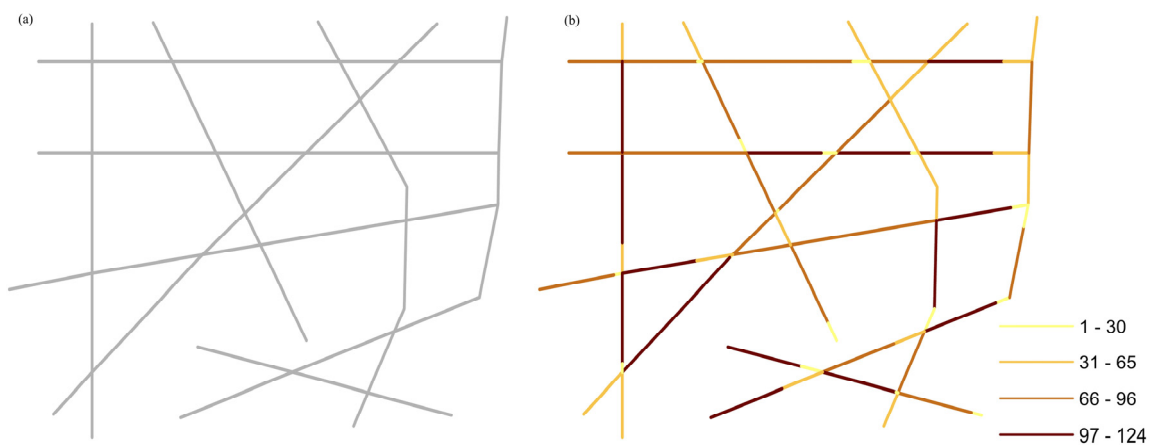
Figure 3. The spatial distribution of the hypothetical network and simulated network events: (**a**) a 10-km hypothetical network; (**b**) the number of randomly distributed 5000 network events along 80 BSUs.

Since the spatial patterns might depend on the BSU length, we also apply a shorter standard BSU length, 100 m, to implement the network segmentation. Two kinds of spatial weight matrices, node-based and distance-based matrices, are also used. Moreover, we simulate a set of randomly distributed point events to represent the observed network phenomenon and the spatial distributions of the events are identical in models with same BSU length and event counts. Table 1 shows the summary statistics of the proposed method. The results indicate that the model, which has 50 network events with a standard BSU length of 200, shows a higher performance with a lower DIC, and there is no considerable difference in the smoothing properties of the CAR model using two types of spatial weight matrices because the differences in DIC are less than 5. Figure 4a shows crude estimates of the relative risk. Based on the results of M_2, the spatial distribution of posterior estimates is depicted in Figure 4b. A comparison of Figure 4a,b reveals that the extreme values of the crude risk can be smoothed by the proposed method, which is based on Bayesian statistics, for network-constrained phenomena. This could imply that the spatial variation of risk estimates in a network space is attributable to the spatial dependence and extra variability relative to planar space. LINCS methods are then applied to detect spatial clusters within the simulated phenomenon based on the results of M_2 (Figure 5).

Table 1. The results of the proposed method under different scenarios.

Weight Matrix	Model	BSU Length	Event Counts	Dbar [1]	Dhat [2]	pD [3]	DIC [4]
Node-based	M_1	100	50	199.457	192.344	7.113	206.570
	M_2	**200**	**50**	**156.022**	**150.979**	**5.044**	**161.066**
	M_3	100	100	276.031	266.234	9.796	285.827
	M_4	200	100	200.296	192.094	8.202	208.497
	M_5	100	200	368.890	360.617	8.273	377.162
	M_6	200	200	254.844	249.845	4.999	259.843
Distance-based	M_7	100	50	199.146	192.457	6.690	205.836
	M_8	**200**	**50**	**155.513**	**149.565**	**5.948**	**161.461**
	M_9	100	100	275.655	265.605	10.050	285.705
	M_{10}	200	100	203.896	196.413	7.483	211.379
	M_{11}	100	200	368.364	360.303	8.062	376.426
	M_{12}	200	200	254.877	249.359	5.518	260.395

[1] **Dbar** denotes the posterior expected deviance; [2] **Dhat** denotes the deviance evaluated at the posterior expectations; [3] **pD** denotes the effective number of model parameters, [4] **DIC:** deviance information criterion.
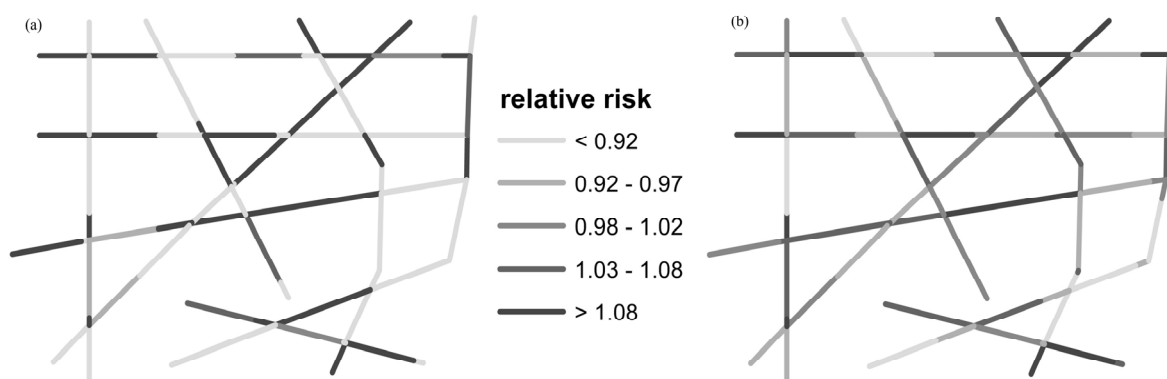
Figure 4. Distributions of relative risk: (**a**) crude risk; (**b**) smoothed risk.

Figure 5a,c present the results from the LINCS methods adjusted for the base distribution given 999 conditional permutations and a significance level of 0.01; one BSU is identified as having significant negative autocorrelation (Figure 5a) and there are significant spatial concentrations in the simulated phenomenon (Figure 5c). Using the posterior estimates as the attribute values of the BSUs for

computing the LINCS and using a significance level of 0.01, two BSUs are identified as having significant network autocorrelation (Figure 5b), and Figure 5d shows the distribution of significant concentrations of values. A comparison of Figure 5a,b reveals that there are significant differences in results between ILINCS and the proposed method; the links with significant concentrations of values identified by GLINCS (Figure 5c) do not always correspond to links with significant concentrations identified by the proposed method (Figure 5d). This implies that spatial dependence could be a major factor influencing the spatial pattern inferences and simulation methods, such as MCMC and permutation methods, should be applied in this field.
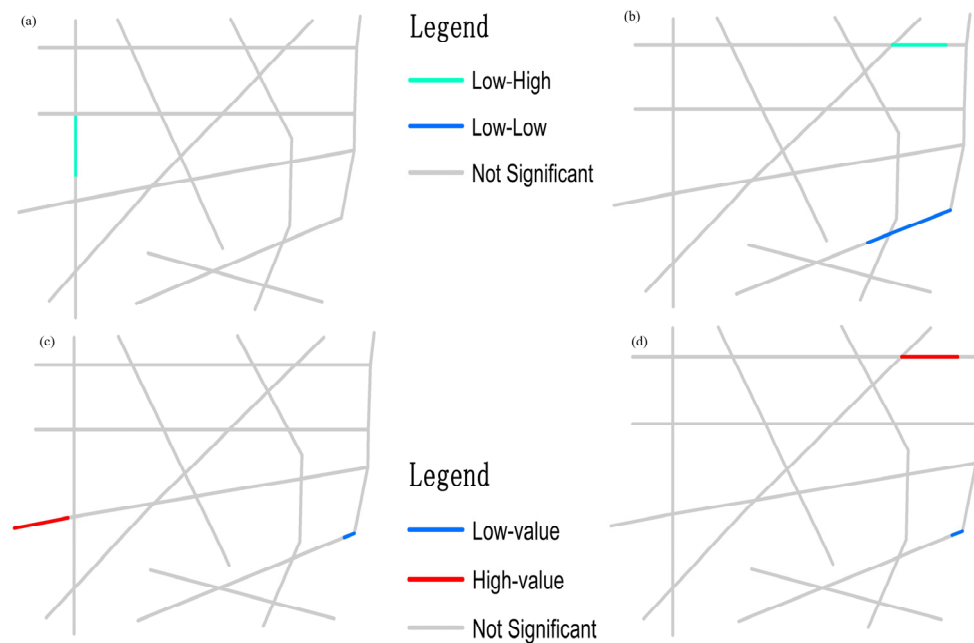
Figure 5. Distributions of spatial clusters, using: (**a**) ILINCS adjusted for the base distribution; (**b**) ILINCS using posterior risk and not adjusted for the base distribution; (**c**) GLINCS adjusted for the base distribution; (**d**) GLINCS using posterior risk and not adjusted for the base distribution.

4.2. Spatial Patterns of Urban Facilities in Futian

In this section, the proposed method is used to analyse the spatial variation of urban facility POIs for leisure activities in a complex road network. In practical research treating subjects such as spatial analysis of crimes and traffic crashes, the identification of high-high autocorrelation and high-value concentration patterns attract more attention than others. Therefore, in the case study, the LINCS methods are applied to detect hotspots of urban facilities for leisure activities distributed along 651 km of road network in Futian, Shenzhen. The structure of the road network is complex; it contains expressways, primary roads, secondary roads and branch roads. The proposed method is implemented on one personal computer with an Intel® Core™ 2 Duo CPU and 4 GB RAM (Lenovo T430) running a 64-bit OS (Windows 7 Professional). The experimental results are stored as shapefiles and visualized in ArcGIS 10.0 software. The LINCS are calculated based on 999 iterations of Monte Carlo simulations using a significance level of 0.001 and a binary connectivity matrix. We apply two kinds of spatial weight matrices, node-based and distance-based matrices, for the proposed method. Table 2 shows the summary statistics of the proposed method. The results indicate that there is no significant difference in the performances of the CAR models using two types of spatial matrices. The posterior distribution of the parameters of hierarchical Bayesian models using two types of spatial matrix is summarized in Table 3.

Table 2. The results of the proposed method using two kinds of spatial matrices.

Weight Matrix	Dbar [1]	Dhat [2]	pD [3]	DIC
Node-based	1012.100	893.768	118.332	1130.430
Distance-based	1011.310	891.668	119.647	1130.960

[1] **Dbar** denotes the posterior expected deviance; [2] **Dhat** denotes the deviance evaluated at the posterior expectations; [3] **pD** denotes the effective number of model parameters.

Table 3. The results of the posterior distribution of the parameters.

Weight Matrix	Node	Mean	sd [1]	MC Error [2]	Median	Credible Level	
						2.5%	97.5%
Node-based	α	0.05996	0.06838	2.781×10^{-4}	0.06007	-0.07689	0.1948
	σ_v^2	7.762	40.57	1.06	4.491	2.342	20.6
	σ_ε^2	41.56	564.3	16.36	4.245	1.223	158.9
Distance-based	α	0.08824	0.03024	1.332×10^{-4}	0.0884	0.02829	0.1473
	σ_v^2	6.395	48.99	1.982	1.666	0.9203	19.41
	σ_ε^2	36.74	489.1	21.28	5.871	1.814	264.5

[1] **sd** denotes standard deviation; [2] **MC Error** denotes Markov Chain errors.

Figure 6a,c shows the distributions of the local statistics for the standardized POI counts adjusted for the base distribution. For the proposed method, the posterior distribution is acquired based on three parallel MCMC chains that each contains 50,000 iterations after a 10,000 burn-in. Then, the posterior estimates based on node-based CAR model are applied as inputs for calculating LINCS without adjustment for the base distribution (Figure 6b,d). The results of the local-scale analyses are summarized in Table 4.

In Figure 6a, there is no significant high-high network autocorrelation in the study region; while a number of BSUs are identified as having significant high-high clustering patterns based on the posterior risk (Figure 6b). This implies that the proposed method based on hierarchical Bayesian models is helpful for detecting hotspots within network-constrained phenomena. Figure 6d presents the results of the local statistics using posterior risk without adjustment for the base distribution, where a smaller number of links are identified as having significant high value concentrations than in the results from GLINCS (Figure 6c). A comparison of Figure 6c,d reveals that the BSUs with significant high-value concentrations using posterior risk without adjustment for the base distribution always indicate significant clusters of high value using GLINCS. This implies that the selected POIs tend to form a clustered pattern with adjustment for the base distribution and that posterior risk can be used to identify clusters of urban facilities.

Table 4. Number of BSUs with significant high-high autocorrelation (local I_i) and high value concentration (local G_i^*) (0.001 significance level).

Statistic Used	Pattern	Data Type	
		Raw POI Counts Adjusted for Base Distribution	Posterior Risk without Adjustment
ILINCS (local I_i statistic)	High-high network autocorrelation	0	121
GLINCS (local G_i^* statistic)	Cluster of high values	273	140

In a network space, spatial dependence still has a significant impact on spatial pattern detection. The method based on hierarchical Bayesian models can remove the part of random variability resulting from small-sample estimation, which is valuable for exploring spatial patterns in network-constrained phenomena. Moreover, it is convenient to combine explanatory variables in the proposed method, which can be used to better understand the determinants of unrevealed spatial process. One notable

limitation is that the weight matrix **W** was set to be simple spatial weight matrices in the case study; the road links were split into 1-km BSUs. Although the results confirm the findings that there is no considerable difference in the smoothing properties of the CAR model using two types of spatial weight matrices, it is often worthwhile to search for the most effective scale of clustering by examining multiple values of standard BSU length with multiple weight matrices. At the areal level, we created several spatial neighbourhoods (e.g., adjacency-based, distance-based, and similarity-based matrices) and applied them to a large dataset of hypertension admissions in Shenzhen. The results indicated that spatial weight matrices had limited impact on the performance of CAR models. Although defining network neighbours is a complex task, it could reduce potential bias resulting from an assumed cluster size and might also mitigate the modifiable areal unit problem associated with the aggregation of events into link segments, which should be investigated in further research.
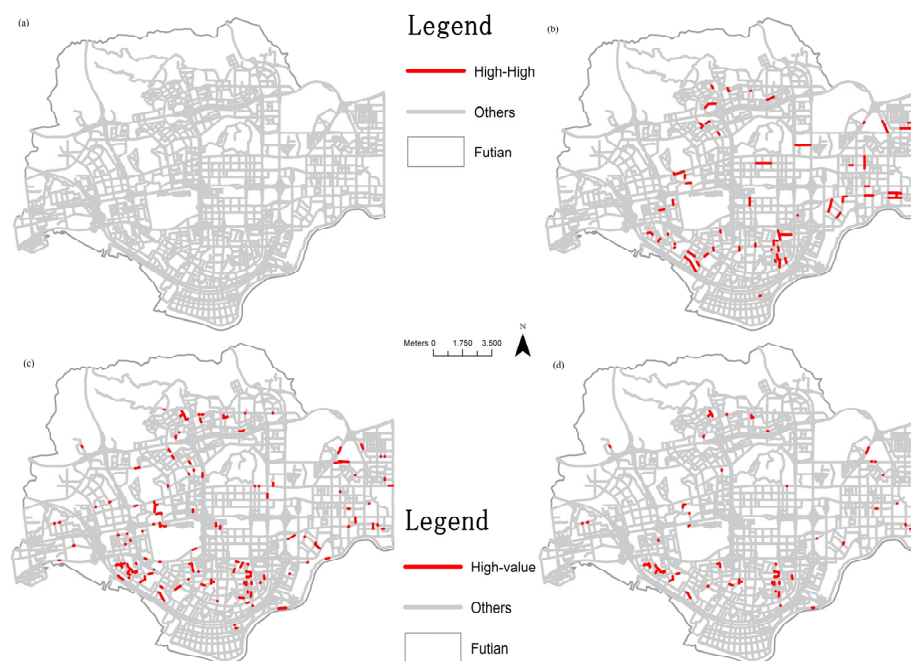
Figure 6. The results of local statistics: (**a**) ILINCS adjusted for the base distribution; (**b**) ILINCS using posterior risk and without adjustment for the base distribution; (**c**) GLINCS adjusted for the base distribution; (**d**) GLINCS using posterior risk without adjustment for the base distribution.

5. Conclusions

Given the ongoing progress in data collection and statistical inference, research interests in spatial analysis have shifted from the meso-scale to the micro-scale to measure the spatial patterns of complex spatial processes [23]. For instance, in road crash analysis, a traffic crash blackspot is defined as an individual road segment with a large number of crashes. A set of continuous road segments with elevated crash counts describes a traffic crash hot zone. In this study, we applied a method based on hierarchical Bayesian models to explore the spatial variation of network-constrained phenomena represented by a link attribute in conjunction with two experiments based on a simplified hypothetical network and a complex road network in Shenzhen with 4212 urban facility POIs for leisure activities. Statistically precise estimates were generated by incorporating prior information from adjacent links to remove the random variation resulting from small numbers of observations. The proposed method incorporated a spatial weights matrix to account for spatial dependence and considered background variations in the base distribution.

In the case study, the proposed method was applied to explore the spatial variation of urban facility POIs for leisure activities in Shenzhen, China and LINCS methods were further applied to

detect the local-scale clusters. In modern cities, urban facilities such as community public service facilities, medical institutions, and retail stores, are significant components that are necessary to the daily lives of urban residents. Moreover, urban facilities are generally expected to form spatial clusters in geographical space because of the inherent association between various types of urban facilities. For example, financial facilities, commercial and consulting facilities, and retail shops are often clustered in central business districts (CBD), which is an important issue for urban planning [40]. However, in the case study, we did not account for the types of individual urban facility POIs (e.g., banks, retail stores and cinemas) nor the relationship between different types of POIs (e.g., schools and residential areas). Incorporation of these characteristics and information on human mobility would significantly increase the practical usefulness of the proposed method, which could be applied in the fields of urban planning and safety research.

Author Contributions: Zhensheng Wang designed the approach and conceived the experiments; Yang Yue and Qingquan Li contributed analyses and discussions; Zhensheng Wang and Ke Nie wrote most of the manuscript; Changbin Yu implemented the experiment.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gatrell, A.C. *Distance and Space: A Geographical Perspective*; Oxford University: Oxford, UK, 1983.
2. Haining, R.P. *Spatial Data Analysis: Theory and Practice*; Cambridge University Press: Cambridge, UK, 2003.
3. Illian, J.; Penttinen, A.; Stoyan, H.; Stoyan, D. *Statistical Analysis and Modelling of Spatial Point Patterns*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
4. Okabe, A.; Okunuki, K.; Shiode, S. SANET: A toolbox for spatial analysis on a network. *Geogr. Anal.* **2006**, *38*, 57–66. [CrossRef]
5. Yamada, I.; Thill, J.C. Local indicators of network-constrained clusters in spatial point patterns. *Geogr. Anal.* **2007**, *39*, 268–292. [CrossRef]
6. Okabe, A.; Satoh, T.; Sugihara, K. A kernel density estimation method for networks, its computational method and a GIS-based tool. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 7–32. [CrossRef]
7. Black, W.R. Network autocorrelation in transport network and flow systems. *Geogr. Anal.* **1992**, *24*, 207–222. [CrossRef]
8. Okabe, A.; Yomono, H.; Kitamura, M. Statistical analysis of the distribution of points on a network. *Geogr. Anal.* **1995**, *27*, 152–175. [CrossRef]
9. Okabe, A.; Kitamura, M. A computational method for market area analysis on a network. *Geogr. Anal.* **1996**, *28*, 330–349. [CrossRef]
10. Okabe, A.; Okunuki, K. A computational method for estimating the demand of retail stores on a street network and its implementation in GIS. *Trans. GIS* **2001**, *5*, 209–220. [CrossRef]
11. Yamada, I.; Thill, J.C. Comparison of planar and network K-functions in traffic accident analysis. *J. Transp. Geogr.* **2004**, *12*, 149–158. [CrossRef]
12. Flahaut, B.; Mouchart, M.; Martin, E.S.; Thomas, I. The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach. *Accid. Anal. Prev.* **2003**, *35*, 991–1004. [CrossRef]
13. Loo, B.P.Y.; Yao, S. The identification of traffic crash hot zones under the link-attribute and event-based approaches in a network-constrained environment. *Comput. Environ. Urban Syst.* **2013**, *41*, 249–261. [CrossRef]
14. Xie, Z.; Yan, J. Kernel density estimation of traffic accidents in a network space. *Comput. Environ. Urban.* **2008**, *32*, 396–406. [CrossRef]
15. Nie, K.; Wang, Z.; Du, Q.; Ren, F.; Tian, Q. A network-constrained integrated method for detecting spatial cluster and risk location of traffic crash: A case study from Wuhan, China. *Sustainability* **2015**, *7*, 2662–2677. [CrossRef]
16. Yamada, I.; Thill, J.C. Local indicators of network-constrained clusters in spatial patterns represented by a link attribute. *Ann. Assoc. Am. Geogr.* **2010**, *100*, 269–285. [CrossRef]

17. Miller, H.J. Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geogr. Anal.* **1999**, *31*, 373–399. [CrossRef]

18. Miller, H.J.; Wentz, E.A. Representation and spatial analysis in geographic information systems. *Ann. Assoc. Am. Geogr.* **2003**, *93*, 574–594. [CrossRef]

19. Wang, Z.; Du, Q.; Liang, S.; Nie, K.; Lin, D.; Chen, Y.; Li, J. Analysis of the spatial variation of hospitalization admissions for hypertension disease in Shenzhen, China. *Int. J. Environ. Res. Public Health* **2014**, *11*, 713–733. [CrossRef] [PubMed]

20. Gelfand, A.E.; Diggle, P.; Guttorp, P.; Fuentes, M. *Handbook of Spatial Statistics*; CRC Press: Boca Raton, FL, USA, 2010.

21. Okabe, A.; Sugihara, K. *Spatial Analysis along Networks: Statistical and Computational Methods*; John Wiley & Sons: Hoboken, NJ, USA, 2012.

22. Black, W.R.; Thomas, I. Accidents on Belgium's motorways: A network autocorrelation analysis. *J. Transp. Geogr.* **1998**, *6*, 23–31. [CrossRef]

23. Anselin, L. Local indicators of spatial association-LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [CrossRef]

24. Okabe, A.; Yamada, I. The K-function method on a network and its computational implementation. *Geogr. Anal.* **2001**, *33*, 271–290. [CrossRef]

25. Congdon, P. *Applied Bayesian Modelling*; John Wiley & Sons: Hoboken, NJ, USA, 2014.

26. Lawson, A.B. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*; CRC Press: Boca Raton, FL, USA, 2013.

27. Banerjee, S.; Carlin, B.P.; Gelfand, A.E. *Hierarchical Modeling and Analysis for Spatial Data*; CRC Press: Boca Raton, FL, USA, 2014.

28. Li, G.; Haining, R.; Richardson, S.; Best, N. Space–time variability in burglary risk: A Bayesian spatio-temporal modelling approach. *Spat. Stat.* **2014**, *9*, 180–191. [CrossRef]

29. Pirani, M.; Gulliver, J.; Fuller, G.W.; Blangiardo, M. Bayesian spatiotemporal modelling for the assessment of short-term exposure to particle pollution in urban areas. *J. Expo. Sci. Environ. Epidemiol.* **2014**, *24*, 319–327. [CrossRef] [PubMed]

30. MacNab, Y.C. Bayesian spatial and ecological models for small-area accident and injury analysis. *Accid. Anal. Prev.* **2004**, *36*, 1019–1028. [CrossRef] [PubMed]

31. Best, N.; Richardson, S.; Thomson, A. A comparison of Bayesian spatial models for disease mapping. *Stat. Methods Med. Res.* **2005**, *14*, 35–59. [CrossRef] [PubMed]

32. Besag, J.; York, J.; Mollié, A. Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Stat. Math.* **1991**, *43*, 1–20. [CrossRef]

33. Bernardinelli, L.; Clayton, D.; Montomoli, C. Bayesian estimates of disease maps: How important are priors? *Stat. Med.* **1995**, *14*, 2411–2431. [CrossRef] [PubMed]

34. Lunn, D.J.; Thomas, A.; Best, N.; Spiegelhalter, D.J. WinBUGS-a Bayesian modelling framework: Concepts, structure, and extensibility. *Stat. Comput.* **2000**, *10*, 325–337. [CrossRef]

35. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Linde, A.V.D. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2002**, *64*, 583–639. [CrossRef]

36. Getis, A.; Ord, J.K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **1992**, *24*, 189–206. [CrossRef]

37. Ord, J.K.; Getis, A. Local spatial autocorrelation statistics: Distributional issues and an application. *Geogr. Anal.* **1995**, *27*, 286–306. [CrossRef]

38. Myint, S.W. An exploration of spatial dispersion, pattern, and association of socio-economic functional units in an urban system. *Appl. Geogr.* **2008**, *28*, 168–188. [CrossRef]

39. Gelman, A.; Rubin, D.B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **1992**, *7*, 457–472. [CrossRef]

40. She, B.; Zhu, X.; Ye, X.; Guo, W.; Su, K.; Lee, J. Weighted network Voronoi Diagrams for local spatial analysis. *Comput. Environ. Urban Syst.* **2015**, *52*, 70–80. [CrossRef]