Models for Spatial Data

Winter 2022

Xiao Qin, Ph.D., P.E.
Professor, University of Wisconsin-Milwaukee

Introduction

- Spatial analysis of crash data is to study the distribution of crash locations in order to identify the spatial patterns and their **underlying causes**.
- **Spatial association** indicators such as Getis G and Moran's I can measure the clustering of crash attributes of a set of geographic features at a global or a local scale.
- Kernel density estimation, Ripley's k-function and cross-k function analyze crash points by calculating crash **intensity or** the **strength** of correlation between two distinct sets of points.
- **Spatial regression methods** explicitly consider spatial dependency of crash observations and spatial heterogeneity in the relationship between crashes and their contributing factors.

Learning Objectives

- Understand the characteristics of spatial data, data types and data models.
- Use spatial correlation indicators such as Getis G, Moran's I to measure and explain spatial association.
- Use kernel density estimation, K function to perform first-order and second-order spatial analysis.
- Understand the similarities and differences between different spatial econometric models.
- Learn and develop hierarchical Bayesian models to quantify the relationship between crashes and contributing factors.
- Understand the limitations of geographically weighted regression and use the model to its advantages.

Spatial Data and Data Types

- Spatial data identify the geographic location of features, boundaries and other geographic phenomena on the surface of the Earth.
- Spatial data are usually recorded by coordinates, pixels, and typology.
- The main spatial data types are vectors and rasters.
  - Points, lines and polygons are vector data.
  - Other data types such as elevation, temperature, and rainfall precipitation have no distinct shape. Instead, they can be measured for any location and are better represented as **surfaces** than as shapes. The most typical surface is raster, which is a matrix of identically sized square cells.

Spatial Data Models

- In a vector model, the physical representation of the features includes two components:
  - the location, and
  - the characteristics (i.e., attributes) of the feature.
- In a raster model, each cell represents a unit of surface area and contains a measured or estimated value for that location. The raster model stores only the data values, and does not include the location information pertaining to the position of individual grid cells.

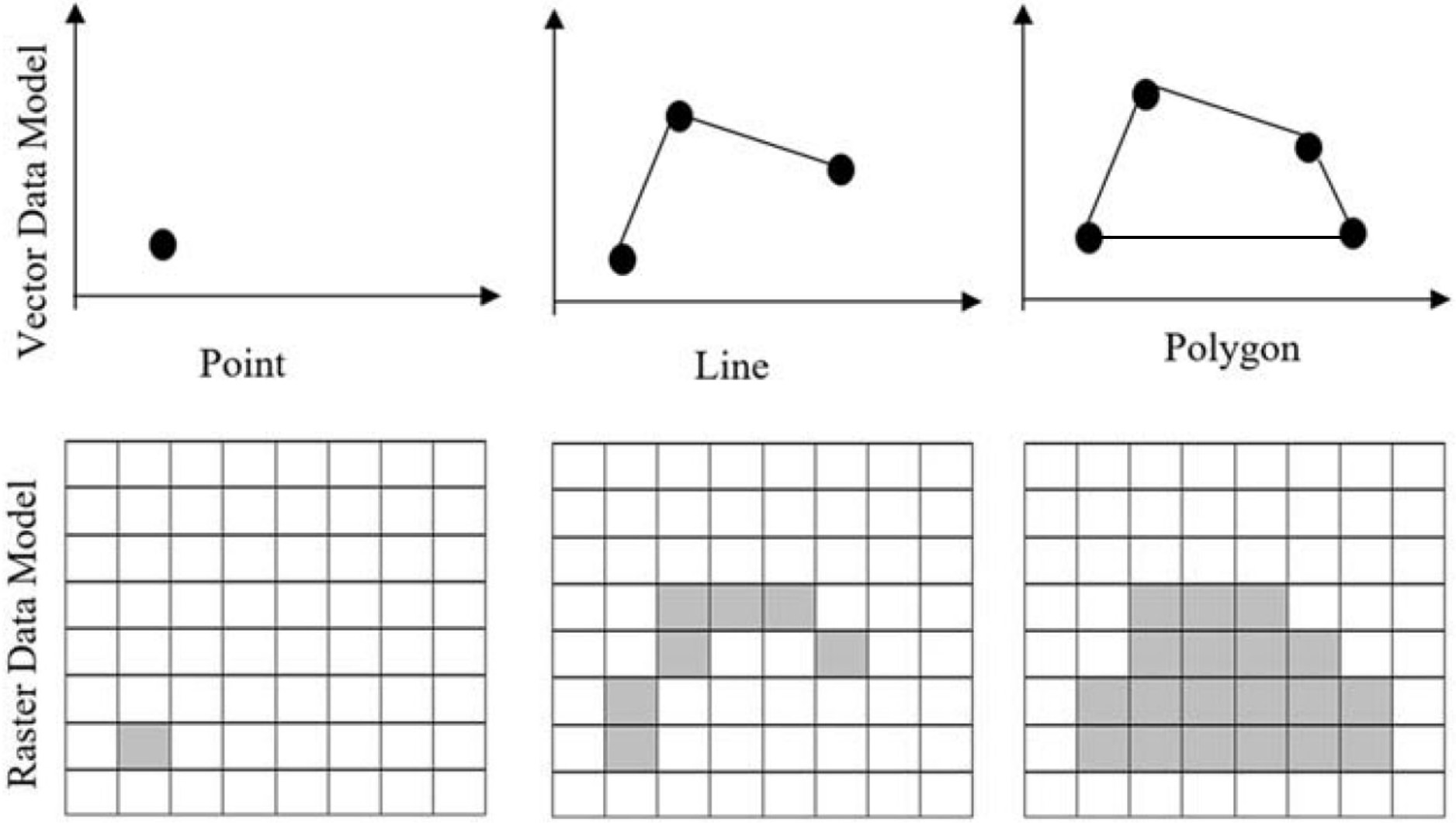


FIGURE 9.1 Vector data model and raster data model.

Measurement of Spatial Association

- Spatial series data possess certain patterns that may be the result of the concentration of weighted points or the areas represented by weighted points.
- Characterize the structure embedded in spatially referenced data and measure the strength of the correlation.
- Measurement can be taken globally or locally.

Global Vs. Local

- A global measure provides the overall trend for the entire region under study.
- Two most popular global statistics for spatial association or specially, spatial autocorrelation:
  - Getis-Ord General G* (d)
  - Moran's I
- Sometimes it is beneficial to examine patterns at a local level, particularly if the pattern generating process is varying over the space.
  - Local $G_i^*(d)$
  - Local Moran's I_i

Getis-Ord $G_i^*(d)$

Getis and Ord (1992) proposed G*(d) as a global statistic to measure the concentration of the high or low values for an entire study area as a function of distance d. For a specific location or subarea i,

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}(d) x_j}{\sum_{j=1}^{n} x_j}, \forall\, j \qquad (9.1a)$$

$$G_i(d) = \frac{\sum_{j=1}^{n} w_{ij}(d) x_j}{\sum_{j=1}^{n} x_j}, \text{ j not equal to i} \qquad (9.1b)$$

Where w_{ij} is a symmetric 0-1 spatial weights matrix with the value of 1 for all subareas defined as being within distance d of a given subarea i; all others are 0, including the i itself. Each subarea i (i = 1, 2, …, n) is identified with its centroid associated with a value x (a weight or attribute) taken from a variable X.

Gi*(d) and Gi (d) measure similar spatial phenomenon and the difference is the x values includes the x at i.

Getis-Ord General G*(d)

- Since $G_i^*(d)$ is a proportion of the sum of all x_j values that are within distance d of i, $G_i^*(d)$ is high if high value x_js are within d of i, and $G_i^*(d)$ is low if low value x_js are within d of i.
- A more general statistic can be defined based on all pairs of values (x_i, x_j) if i and j are within the distance of d of each other.

$$G^*(d) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(d)x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}, \forall j \qquad (9.2)$$

- G*(d) measures the concentration or lack of concentration. If the absolute value of the Z score of G*(d) is greater than a predetermined value, strong spatial association or clustering is present. A **+Z score** means that **high values cluster** together, and a **-Z score** means that **low values cluster** together.

Moran's I

- Moran's I evaluates whether the pattern (of a set of geographic features with attribute values) is spatially **clustered, dispersed, or random** on a **global** scale in a study area.
- Moran (1950) developed Moran's I in Equation (9.3) to measure the correlation of each xi with all neighboring xjs, including itself.

$$I(d) = \frac{\sum_i \sum_j w_{ij}\left(x_i - \overline{x}\right)\left(x_j - \overline{x}\right)}{W \sum_i \left(x_i - \overline{x}\right)^2 / n} \qquad (9.3)$$

where $\overline{x}$ is the mean of x; w_{ij} is a matrix of spatial weights with zeroes on the diagonal (i.e., $w_{ii} = 0$), and W is the sum of all w_{ij}, $W = \sum i \sum j w_{ij}$. Distance d is used to determine the neighbors j.

- As a correlation statistic, values of I(d) range from -1 to +1. A Moran's Index value **near +1** indicates a **clustering** pattern while an index value **near -1** indicates a **dispersed** pattern.

$G_i^*(d)$ Local Statistics

- Local indicators of spatial association (LISA) have been introduced to help detect local clusters.
- LISA assess the significance of local statistics at each location, identify locations of spatial clusters and spatial outliers irrespective of the presence of global spatial association.
- Ord and Getis (1995) developed the local version of G*(d).

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j - \bar{x}\sum_{j=1}^{n} w_{ij}(d)}{s\sqrt{\frac{N\sum_{j=1}^{n} w_{ij}^2(d) - \left(\sum_{j=1}^{n} w_{ij}(d)\right)^2}{N-1}}}, \forall j. \qquad (9.4)$$

$$\text{With } \bar{x} = \frac{\sum_{j=1}^{n} x_j}{n} \text{ and } s = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{x})^2}$$

- G*(d) statistics are often used for **hot spot/cold spot analysis**. The underlying theory is that a feature with a high value is interesting, but it must be surrounded by other features with high values in order to be qualified as statistically significant. Note that local $G_i^*(d)$ is a z-score so no further calculation is needed.

Local Moran's I_i

- Anselin (1995) modified the local Moran's I index as:

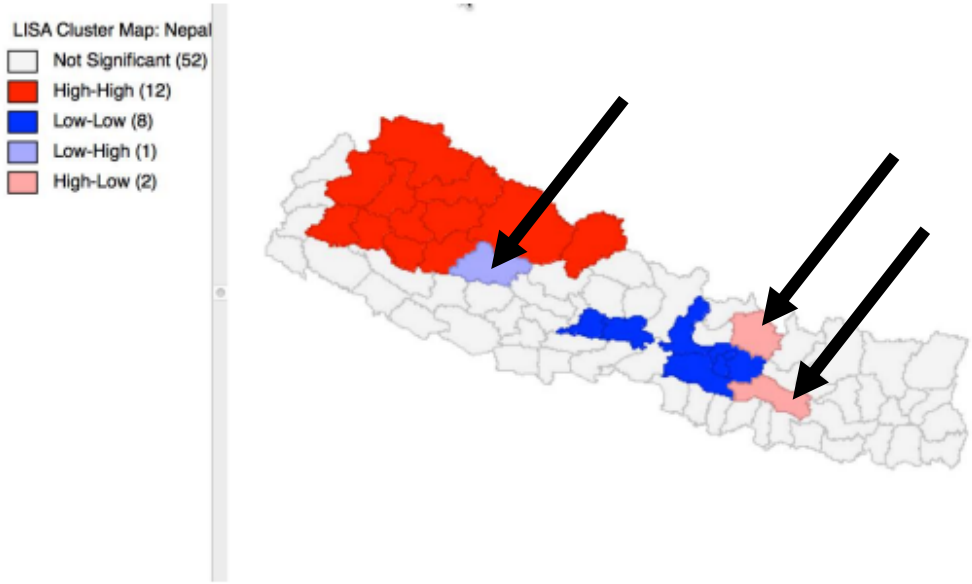
$$I_i = \frac{(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2 / n} \sum_j w_{ij} \left(x_j - \bar{x} \right) \qquad (9.5)$$

- Note that $\sum I_i = NI$. Therefore, global Moran is the average of local Moran statistics.
- A positive value for I indicates a clustering pattern, meaning the feature is surrounded by features with similar values. A negative value for I indicates an outlier, meaning the feature is surrounded by features with dissimilar values.
- Hence, the local Moran's I can help identify the cluster of high values (HH), cluster of low values (LL), an outlier in which a high value is surround primarily by low values (HL), and an outlier in which a low value is surrounded primarily by high values (LH).

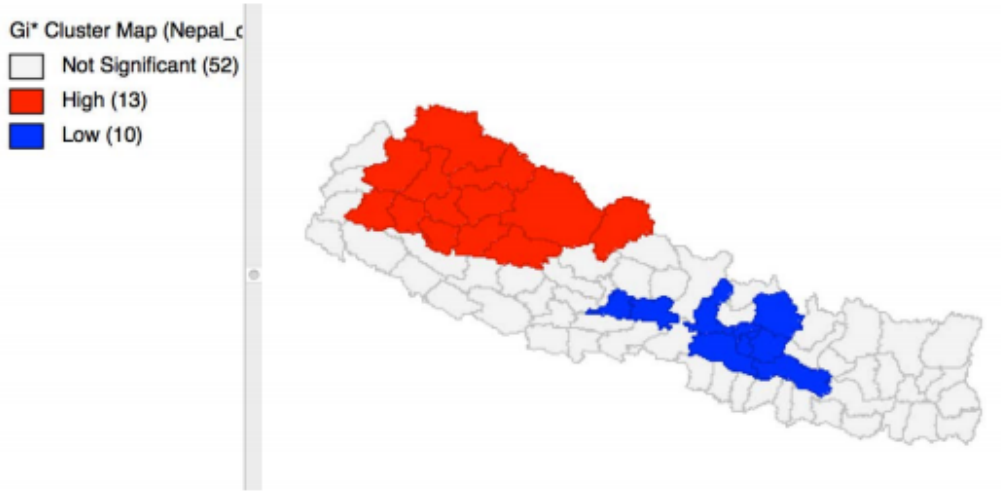
$G_i^*(d)$ and $I_i(d)$

- Both can measure the association among the set of weighted points or areas represented by points, but they are different in formulation.
- $G_i^*(d)$ measures the concentration or lack of concentration of all pairs of (x_i, x_j) such that i and j are within d of each other.
- $I_i(d)$ is used to measure the correlation of each x_i with i all x_js within d.
- This difference means that G statistics are useful when only positive spatial autocorrelation is of interest (i.e., hot spots (clustering of high values) or cold spots (clustering of low values)), whereas Moran's I identifies both spatial clusters and outliers.

Example Local Moran's I Compared to Local Gi*



local Moran



local Gi*

(source "Geospatial Analysis)

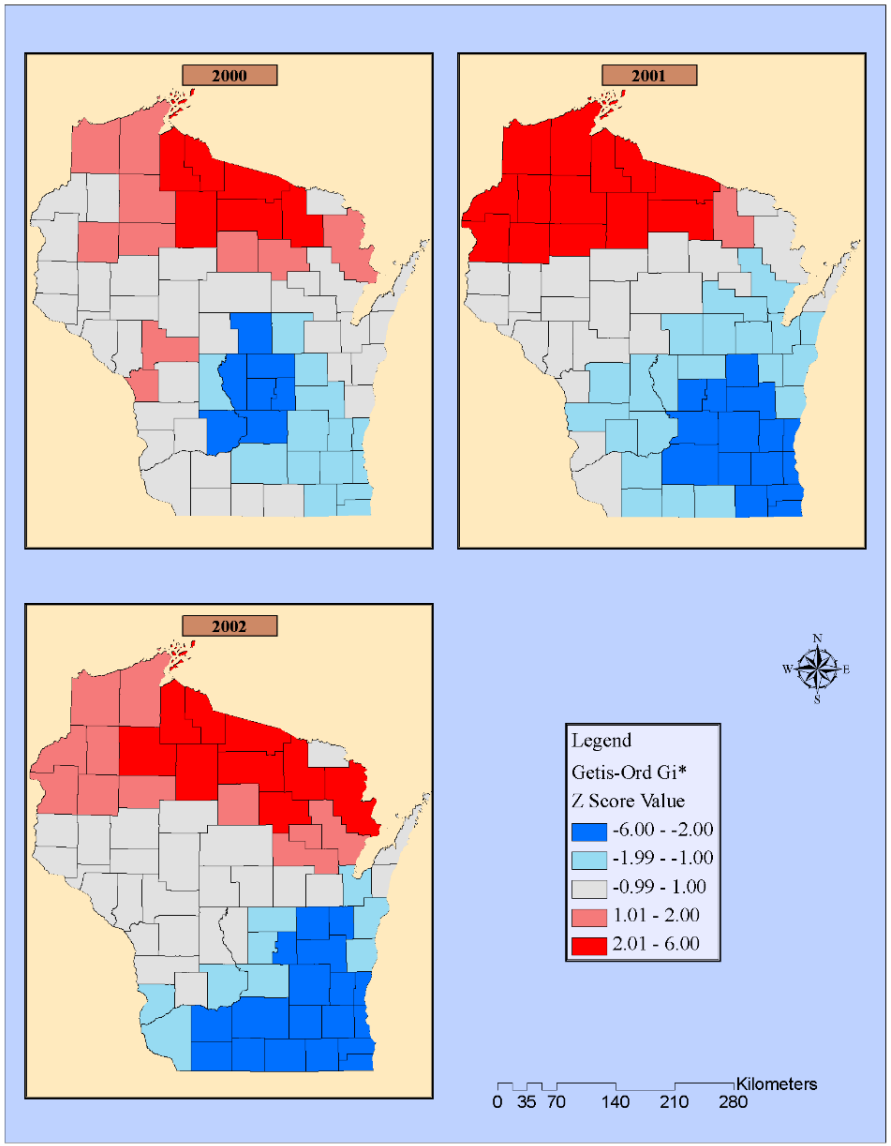
TABLE 9.1 Standard normal variates for G(d) and $I(d)$.

Situation	Z(G)	Z(I)
HH	++	++
HM	+	+
MM	0	0
Random	0	0
HL	−	− −
ML	−#	−
LL	− −	++

#, tends to be more negative than HL; +, moderate positive association; ++, strong positive association (high positive Z scores); *0*, no association; *HH*, pattern of high values of *x*s within *d* of other high *x* values; *L*, low values; *M*, moderate values; *H*, high values; *Random*, no discernible pattern of *x*s.

Table is adapted from Getis, A., Ord, J.K. 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24 (3), 189–206.

Example 9.1 Clusters of Snow-Related Crashes in Wisconsin Calculated by G*i(d)



- The reason for spatial clusters of snow-related crashes in the northern region is likely due to the fact that northern counties in Wisconsin experience more snowfall and snowstorm events.
- Counties with a Z-score between +2 and –2 represent locations that may have a high or low relative crash rate value, but are not part of a statistically significant spatial pattern or cluster.

(Source: Khan et al., 2008)

Spatial Weights

- The spatial weights matrix (or spatial weighting matrix, weighting factor) is the proximity measure that determines the influence of site j on site i where $i \neq j$.
- The measurement can be either adjacency-based or distance-based.
  - In some textbooks, adjacency-based is also referred to as contiguity-based.
- The rule of thumb is that the adjacency-based measure is more common for area or zonal variables and the distance-based measure is more common for point data.
- When the rule is relaxed, the concept of adjacency can be extended for point data based on the distance d_{ij} against a predetermined value; or the concept of distance can be applied to zonal variables in which the distance is measured from zonal centroid to centroid.

Spatial Weights Matrix

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}$$

The weights express the neighbor structure between the observations as a n × n matrix W where the elements wij of the matrix are the spatial weights.

Distance Decay Models

- *Contiguity weights:* The most common neighboring relation is contiguity, which means the two spatial features share a common border of non-zero length.
- *Distance-band weights*: adjacency relation can be constructed from distance based on a predetermined cutoff value.
- The adjacency-based measure might cause the issue of discontinuity and abrupt change along the border. Rather than expressing spatial influence as a binary value based on adjacency, the spatial weight is often expressed as a continuous value using **a distance decay function**.

- Inverse distance weighting:

$$w(d_{ij}) = \frac{1}{d_{ij}^k} \tag{9.7}$$

A generalized powered exponential family:

$$w(d_{ij}) = \exp[-(\phi d_{ij})^k],\ k \in [0, 2]; \phi > 0, \tag{9.8}$$

where ϕ is the principal decay parameter; *k* is a smoothing factor. When *k* = 2, this is a Gaussian distance decay function. A variant of Equation (9.8) is

$w(d_{ij}) = \exp[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2]$, where *h* is the bandwidth.

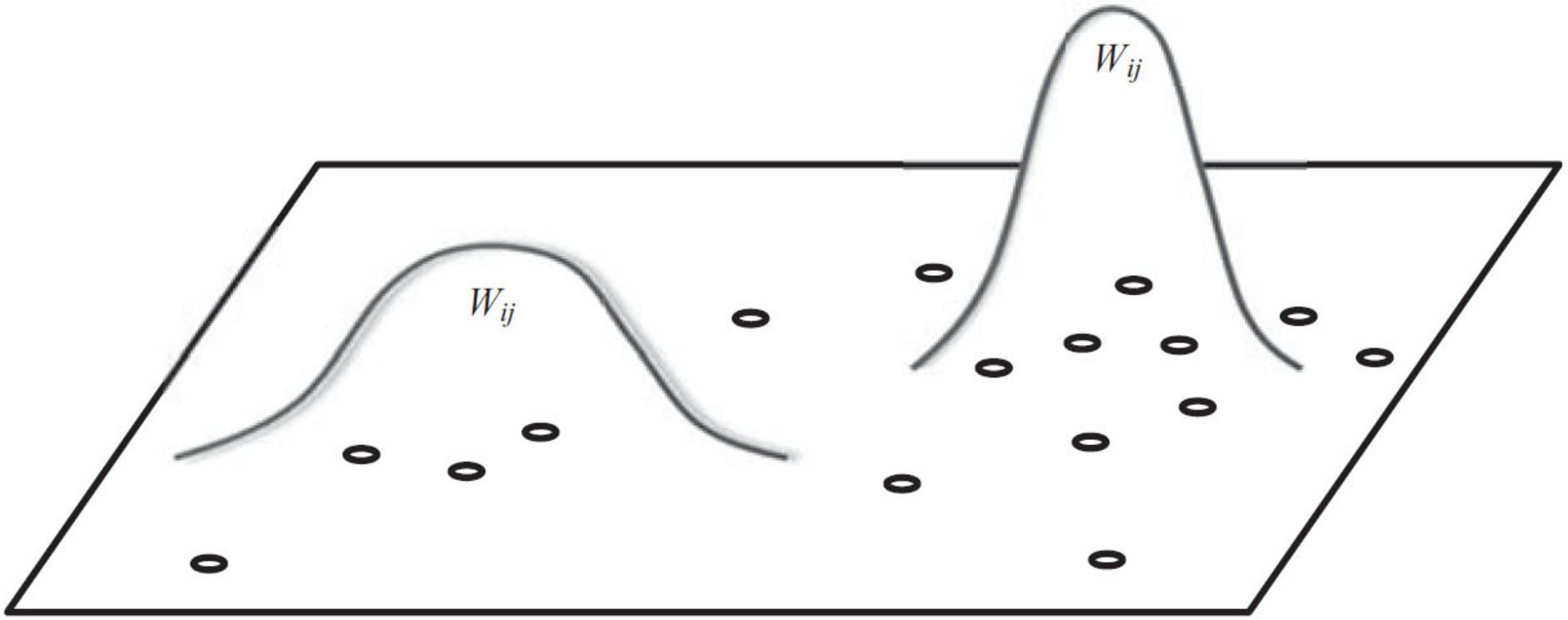


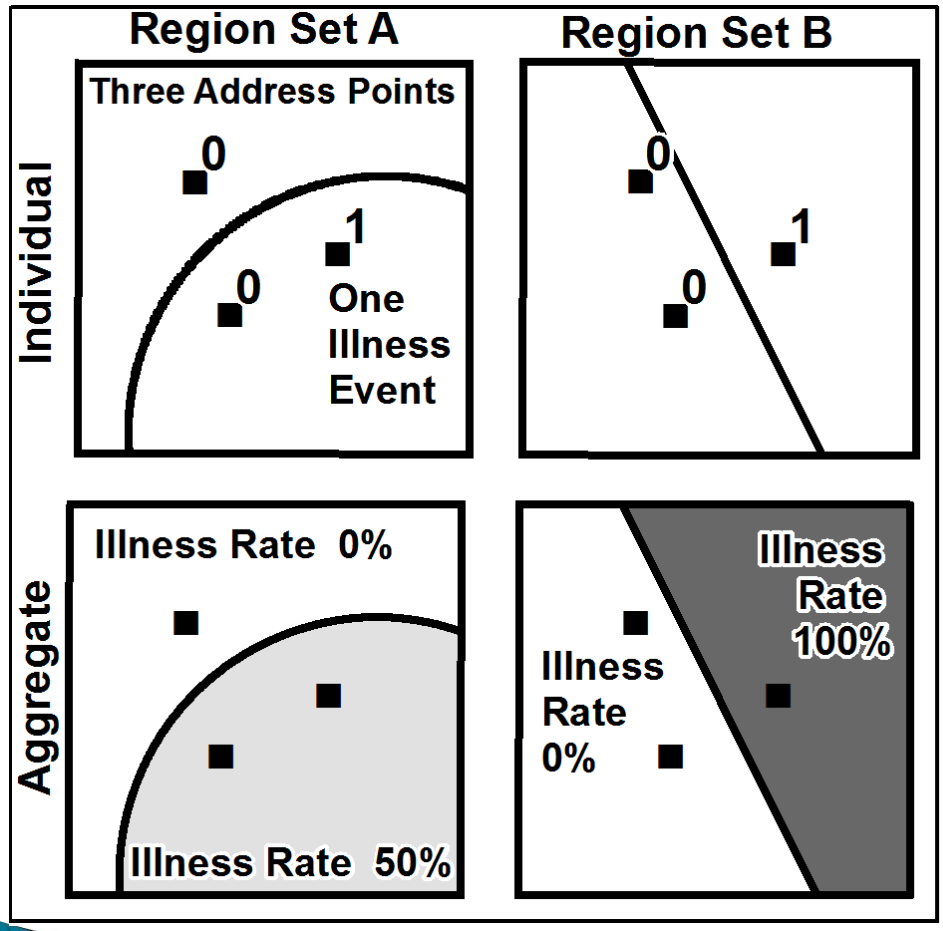
FIGURE 9.3 Adaptive spatial weighting function (◯ data point).

- The spatial weighting function can either be universal (i.e., applied equally at each point) or adaptive, depending on the location of a point as shown in Fig. 9.3.

Point Data Analysis

- Point data analysis studies the distribution of the location of point data in the hope that the spatial patterns observed will provide information about the underlying process that generates the points.
- In safety analysis, researchers often aggregate crashes by location based on pre-established boundaries and scales.
- But the patterns observed can vary by the choice of scales and boundaries. Other techniques such as kernel density estimation allow researchers to analyze crash point patterns directly.
- Researchers may also be interested in studying the location association between two sets of data (i.e., co-location), such as the association between a run-off road crash location and a horizontal curve location.

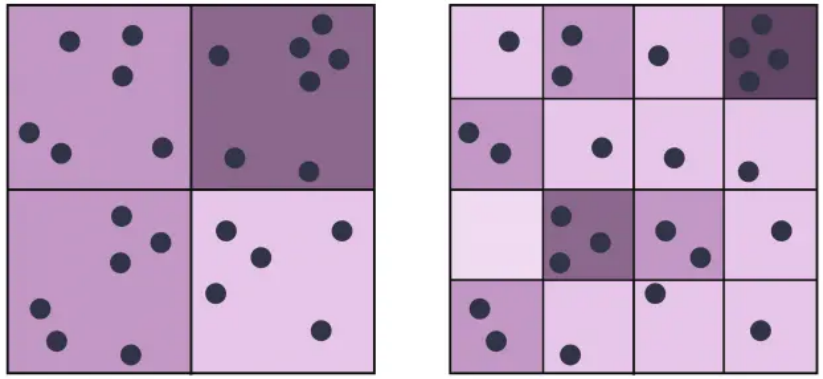
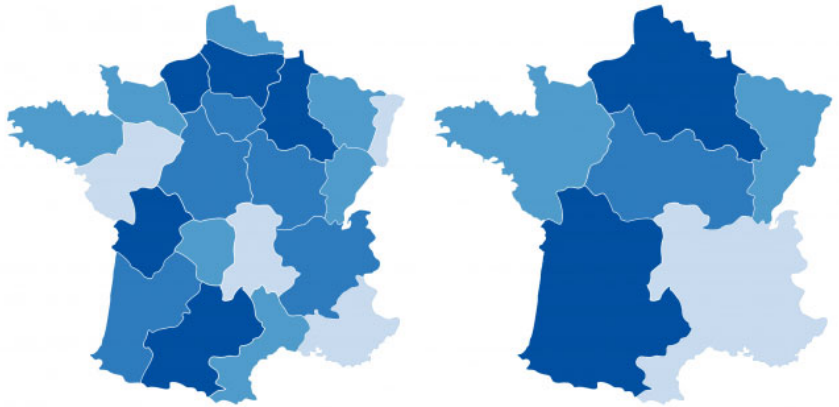
Modifiable Areal Unit Problem (MAUP)



- The modifiable areal unit problem (MAUP) is a source of statistical bias that occurs when you aggregate point data such as the scale and zonal effect.

MAUP Effects

There are two types of biases for the MAUP: Scale effect and Zonal effect? So, do we have a solution??

The scale effect occurs when maps show different results at different levels of aggregation.	The zonal effect occurs when you group data by various artificial boundaries.
	

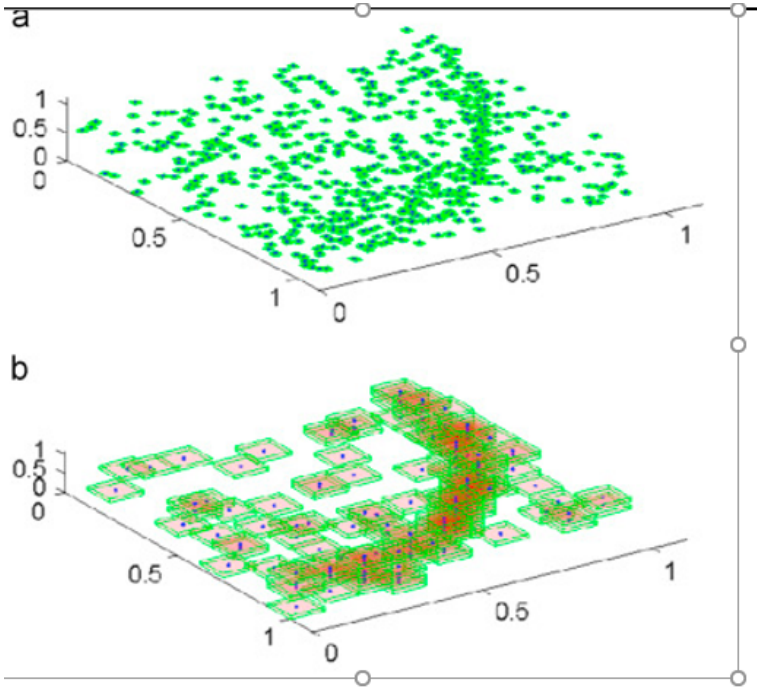
First-order Process

- Point patterns can be defined by running the statistical test of a point pattern against the point pattern generated through a random process, also known as complete spatial randomness (CSR).
- In the first-order property of point process, a point pattern is a data set X, consisting of a series of points $\{x_1,\cdots,x_n\}$. All points are contained in a study area **R**.
- The intensity at the point x, denoted by $\lambda(x)$ is:

$$\lambda(x) = \lim_{r \to 0} \frac{E\left(N\left(C(x,r), \boldsymbol{X}\right)\right)}{\pi r^2} \qquad (9.12)$$

Data Patterns at First-order Intensity

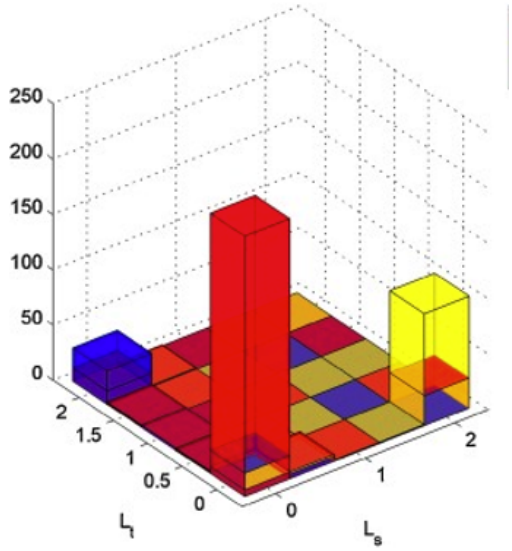
The most straightforward way to analyze data patterns at first-order intensity is to divide the area into grid squares, count the number of events in each square, divide the counts by the area of the squares, and then formulate a density function over space.



c Grid counts for crashes

?	?	?	?
?	?	?	?
?	?	?	?
?	?	?	?

d bar chart



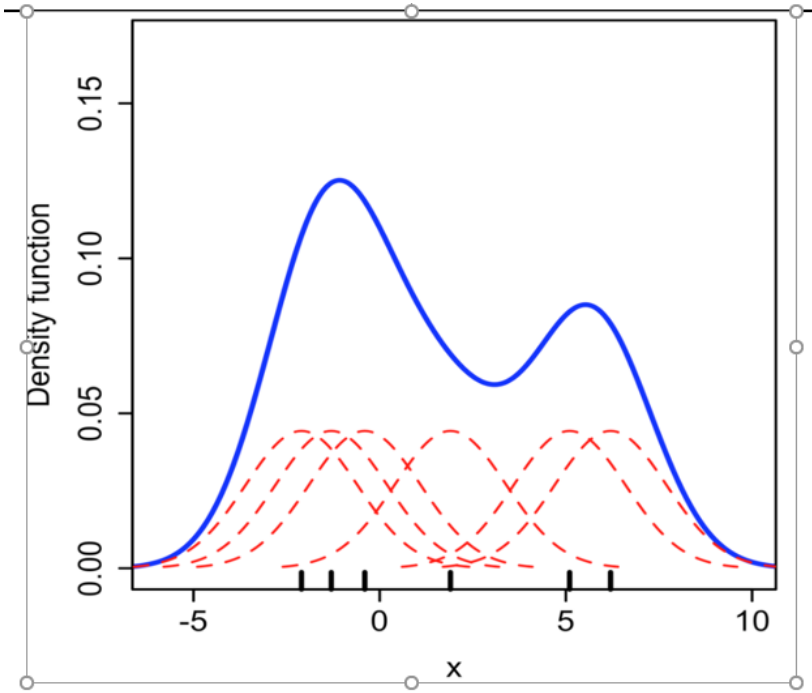
Kernel Density Function

- A better alternative is the kernel density estimation (KDE). KDE is a non-parametric method used to estimate the probability density function of a random variable.
- Let $P = \{x_1, \cdots, x_n\}$ be a univariate independent and identically distributed sample drawn from some distribution with an unknown density function. The shape of this unknown function can be estimated through its kernel density estimates as follows:

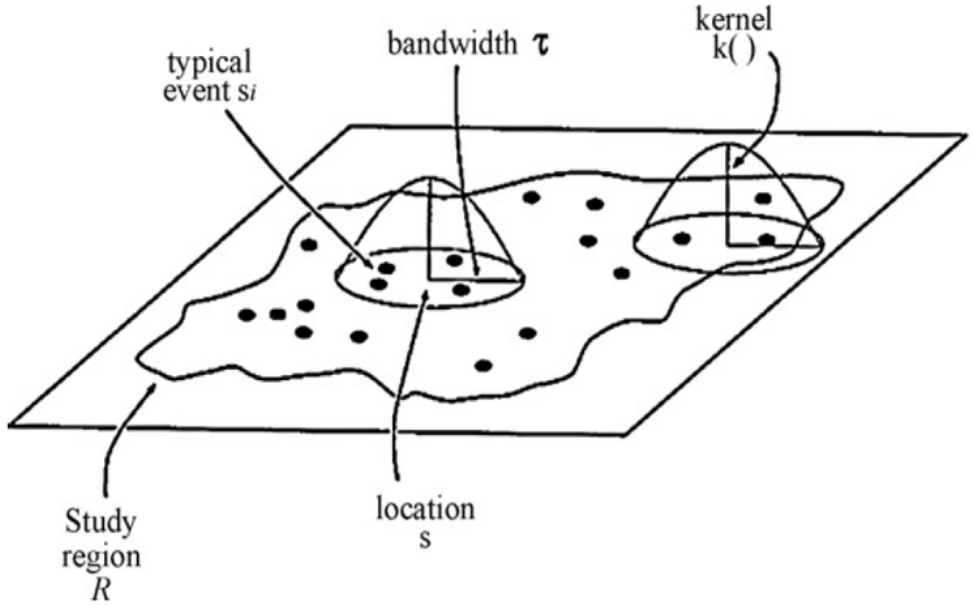
$$\hat{f}_k(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \qquad (9.15)$$

where *x* is the variable of interest; h is the bandwidth that controls the amount of smoothing; d is the dimension (e.g., *d* = 1 is one-dimensional kernel like a roadway link; *d* = 2 is two-dimensional kernel like an area); and, *K*(.) is the kernel function.

One- and Two-dimensional Kernel Density Function



d=1



d=2

Kernel Function

Kernel functions can take on different forms: uniform, normal, exponential and spherical.

Uniform: $K = \begin{cases} 1, & d_{ij} \le h \\ 0, & d_{ij} > h \end{cases}$ (9.16a)

Normal (or Gaussian): $K = \exp\left(-\frac{\left(\frac{d_{ij}}{h}\right)^2}{2}\right)$ (9.16b)

Exponential: $K = exp\left(-\frac{d_{ij}}{h}\right)$ (9.16c)

Spherical: $K = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h}\right)^2\right]^2, & d_{ij} \le h \\ 0, & d_{ij} > h \end{cases}$ (9.16d)

Notes about Kernel Functions

- The kernel function determines the shape of the hump.
- Parameter h (bandwidth) controls the spread of the hump.
- A small h results in a very rapid distance decay, whereas a larger value will result in a smoother decay.
- Also, a large h value tends to smooth out local effects, while a smaller h value tends to produce rugged surfaces with many spikes.
- The kernel function can be either universal (i.e., applied equally at each point) or adaptive to the location of a point. The location-specific kernel is presumed to improve accuracy in prediction, but it can be computationally challenging.

Example 9.2 Use local spatial autocorrelation and the kernel method for identifying crash hot spots

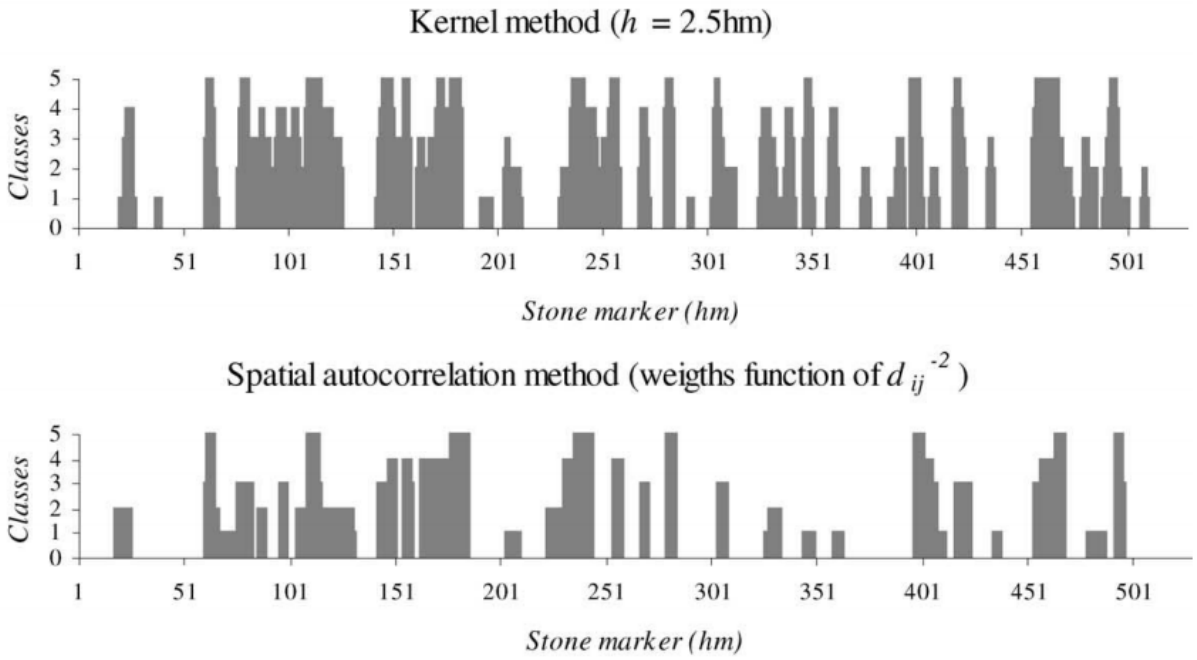


Fig. 9.5 Comparison of the dangerousness of the N29 determined by kernel and the spatial correlation methods (Source: Flahaut et al., 2003)

- Local Moran's I are calculated for the n units based on the number of crashes per hectometer and a spatial contiguity matrix (0-1 matrix) is considered.
- The kernel function for KDE is assumed to be a normal distribution. The reference bandwidth (href = 59.6 hm) and the optimal bandwidth (hopt = 26.5 hm) - are compared.

Second-order Process

- The second-order property of point process examines the correlations (i.e., interactions) between two distinct points in **R**.
- When considering the interaction between a pair of points in the study area **R**, the second-order intensity function can be formulated as:

$$\gamma(x_1, x_2) = \lim_{r_1 \to 0, r_2 \to 0} \frac{E\left(N(C(x_1, r_1), X) N(C(x_2, r_2), X)\right)}{\pi {r_1}^2 \pi {r_2}^2} \quad (9.14)$$

- The distance $d=x_1-x_2$, between a pair of points can also be called displacement or pairwise distance. When only the distance between the two points is of interest, the expression of γ(d) referred to as a stationary process is more appropriate. Ripley's K function is probably the most known second-order measure for summarizing a point pattern.

Ripley's (Planar) K-function

- When the second-order property of data is characterized as an isotropic and stationary process, the second-order intensity function depends only on the distance between the two points
- Ripley's K is the best-known second-order statistic in spatial statistics.
- Ripley (1976) introduces the planar K-function analysis to depict the spatial distribution of point events in a given data set P.
- Ripley's K tests point patterns on various spatial scales, handles all event to event distances, and does not aggregate points into areas.

Ripley's K-function

- An estimate of $K^{pl}(d)$ (pl stands for planar) is:

$$\widehat{K}^{pl}(d) = \frac{1}{\hat{\lambda} n} \sum_{i \neq j} \sum I_h(d_{ij}) = \frac{|A|}{n^2} \sum_{i \neq j} \sum I_d(d_{ij}) \qquad (9.21)$$

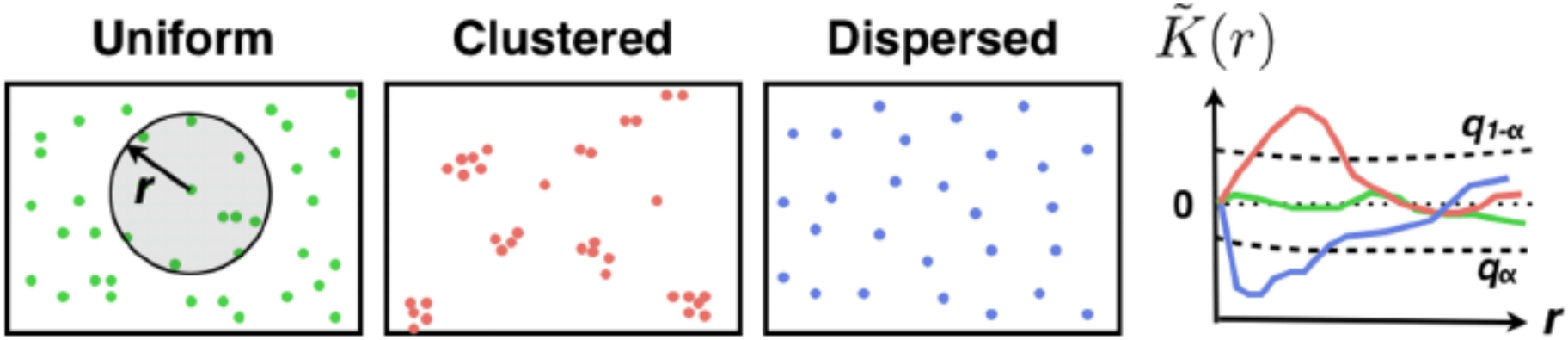
Where n is the number of points in the dataset and |A| is the size of study area. $I_d(d_{ii})$ is an indicator function defined as:

$$I_d(d_{ij}) = \begin{cases} 1 & if\ d_{ij} \leq d \\ 0 & otherwise \end{cases}.$$

- So, the K function for a distance d is the average number of points found in a circle of radius d centered on an event, divided by the intensity of points which is the number of events divided by the study area.

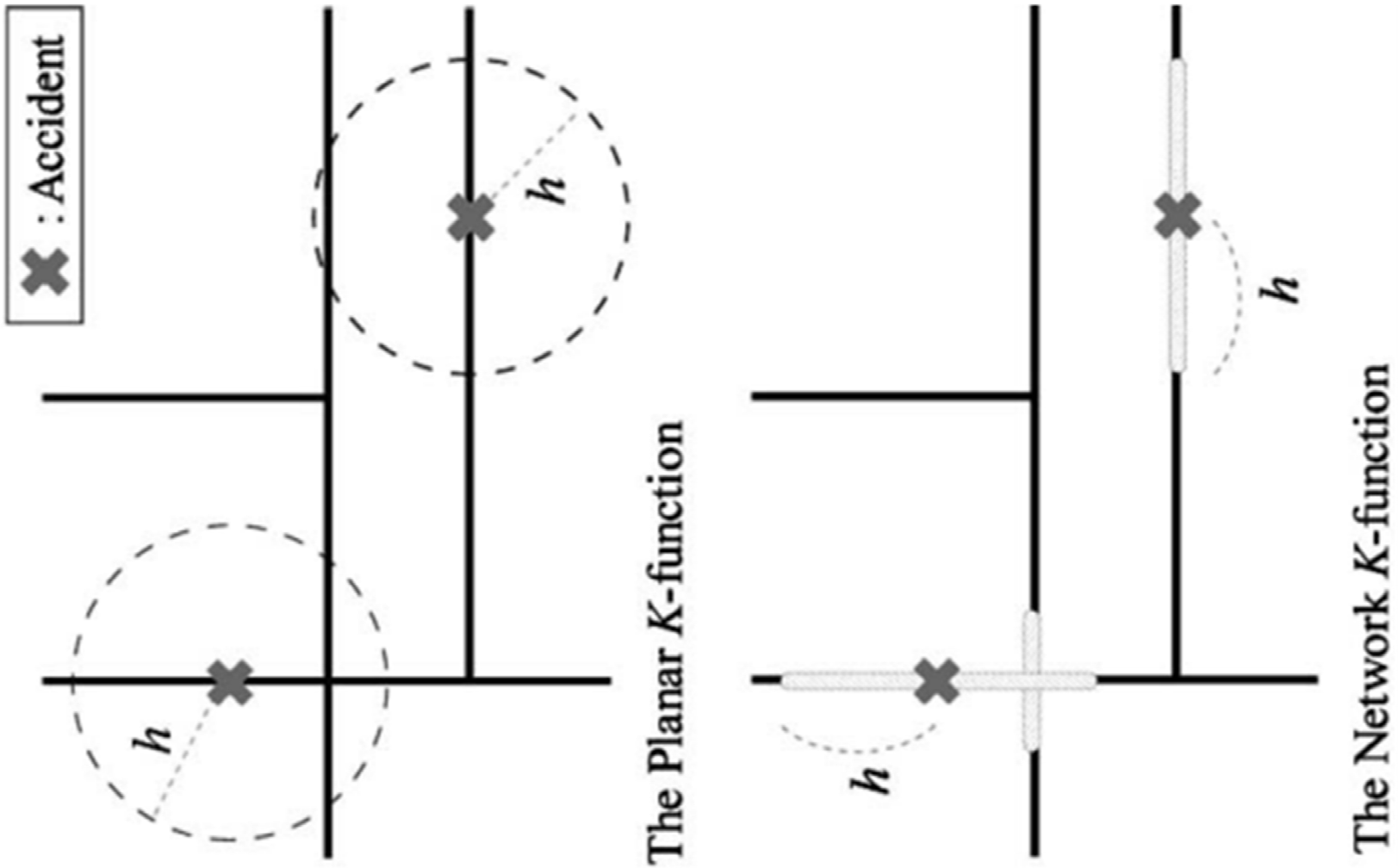
Ripley's K-function

- When the point pattern is CSR following the Poisson point process and where events are independently and uniformly distributed over the study area, the theoretical value of $K^{pl}(d)$ is given by πd^2.
- In practice, the Monte Carlo simulation is often used to calculate pseudo-significance levels by repeated randomization. This technique determines the expected values of $\widehat{K}^{pl}(d)$, the upper and lower significance envelopes under the null hypothesis of CSR. Cressie (1993) had a more detailed discussion of the planar K-function analysis.



Source:
(doi:10.1371/journal.pone.0080914.g001)

Basic concepts of the planar and network K-functions



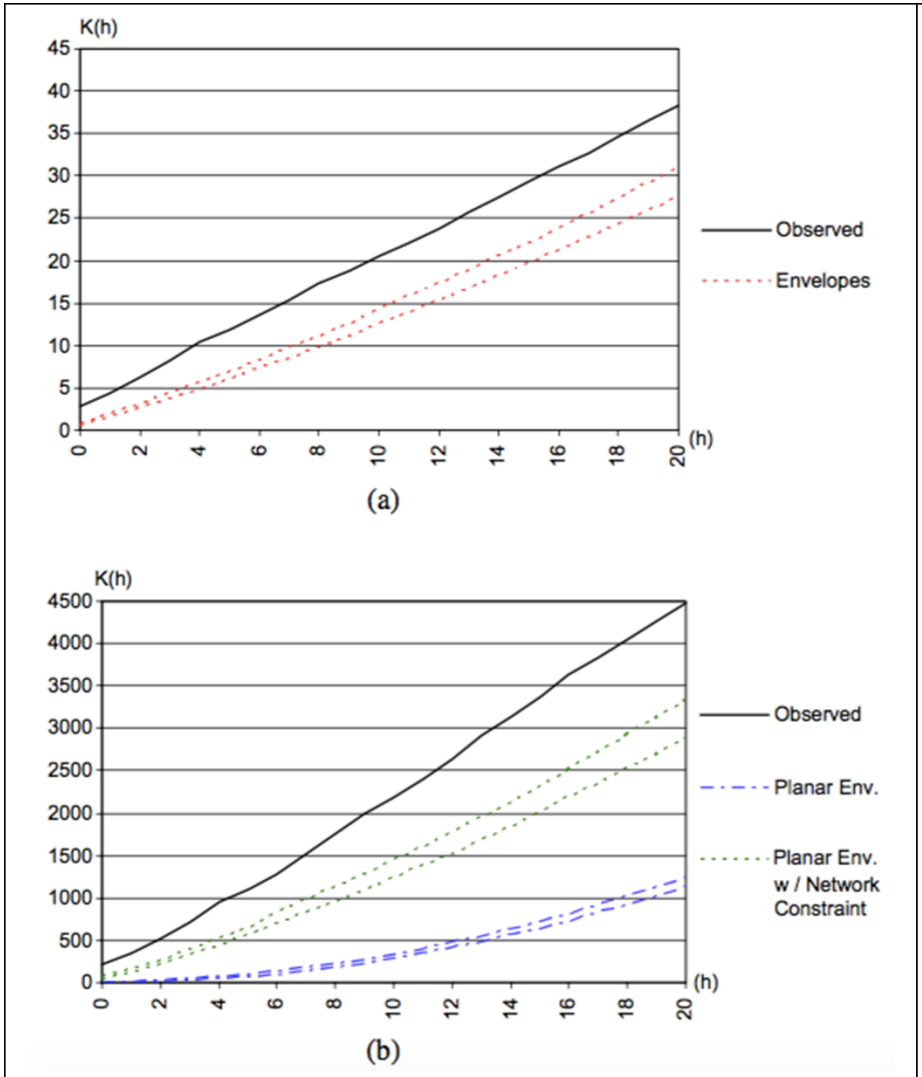
Source: Okabe and Yamada (2001)

Network K-function

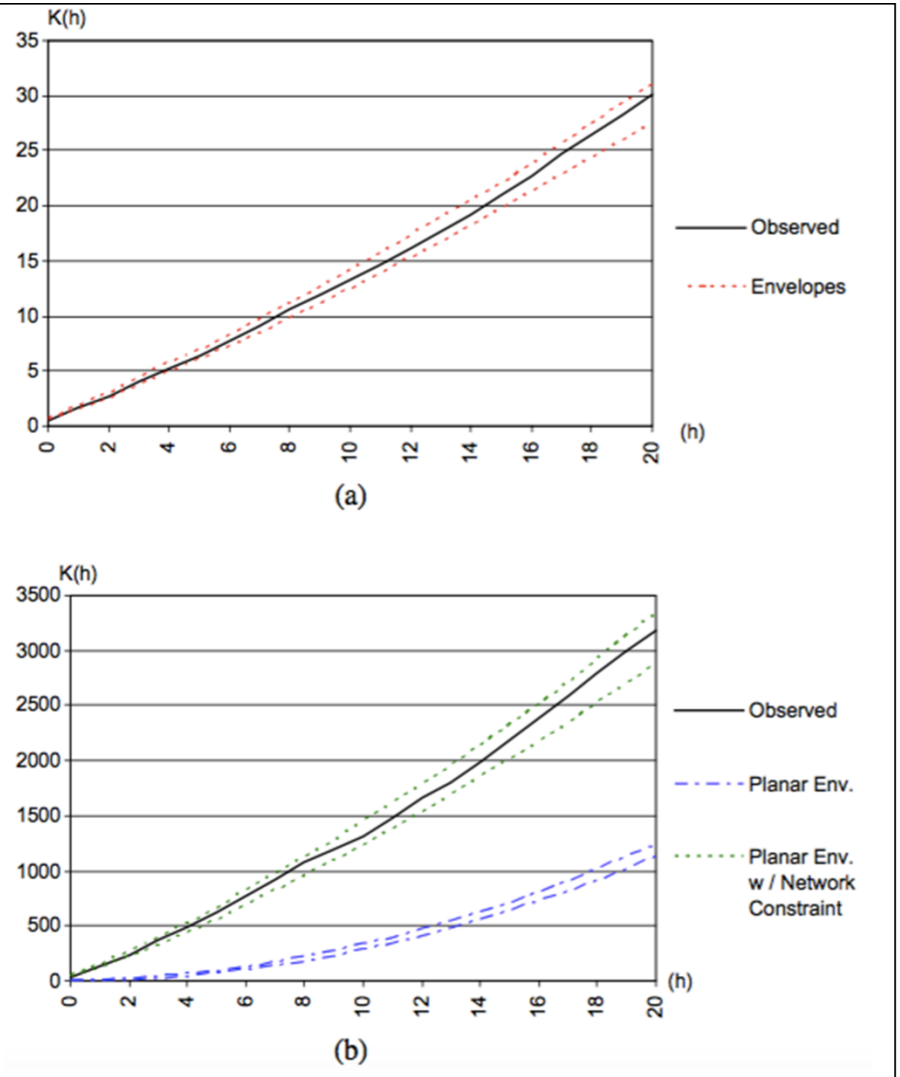
- Okabe and Yamada (2001) and Yamada and Thill (2004) extended from a Euclidean space to a network space. For An observed point pattern, the estimator of the network K-function $\widehat{K}^{net}(d)$ (net is for network) is given by:

$$\widehat{\boldsymbol{K}}^{\boldsymbol{net}}(\boldsymbol{d}) = \frac{|\boldsymbol{L_T}|}{\boldsymbol{n(n-1)}} \sum_{\boldsymbol{i \neq j}} \sum \boldsymbol{I_d(d_{ij})} \qquad (9.24)$$

where d_{ij} is the network distance between points x_i and x_j, and $I(\ )$ is the indicator function defined previously. Note that an unbiased estimator of the density of points, $\frac{n-1}{L_T}$, is used here rather than, $\frac{n}{L_T}$. See Okabe and Yamada (2001) for more details on the method derivation.



a) K-function values for the 450 accidents: the network K-function (upper) and the planar K-functions (below)



b) K-function values for the simulated random pattern: the network K-function (upper) and the planar K-functions (below)

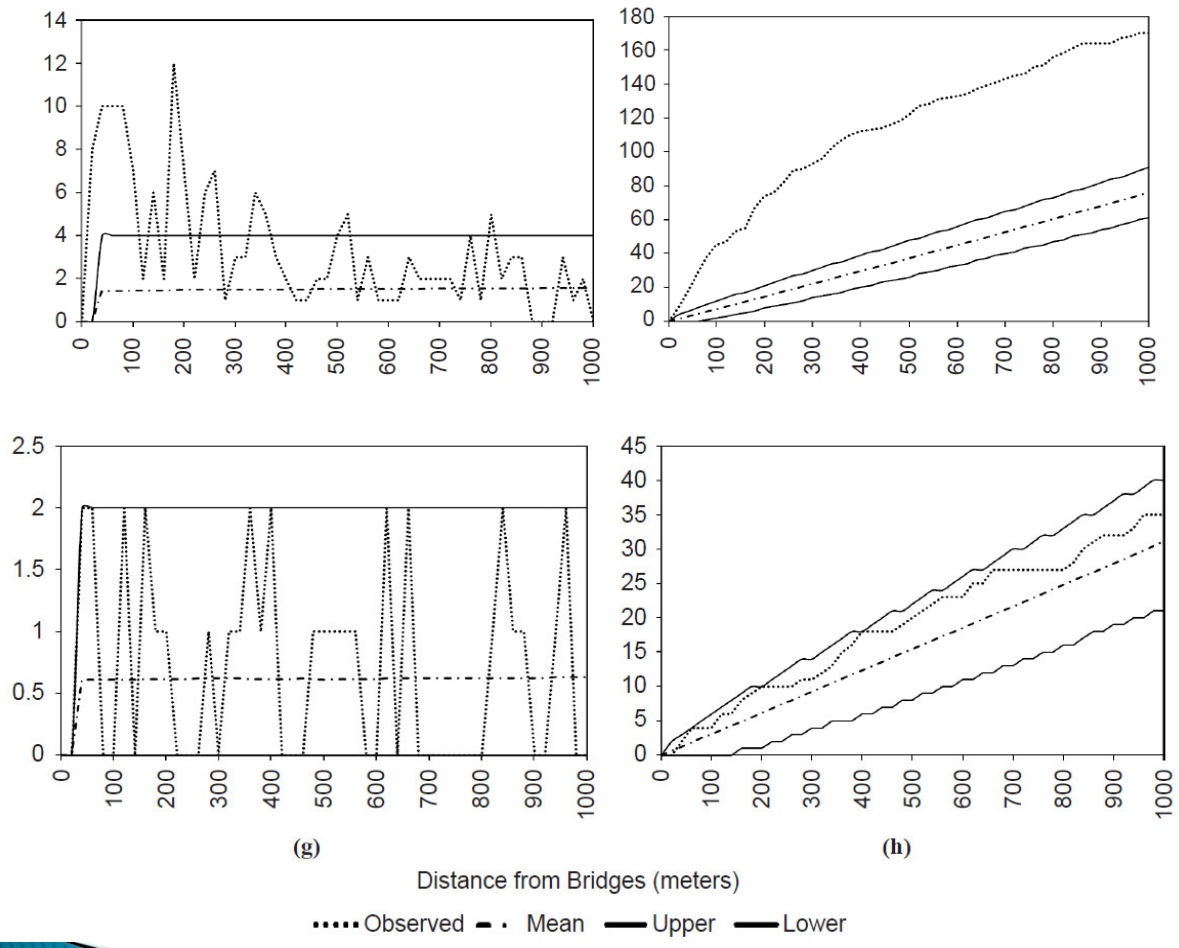
Monte Carlo simulation is a conventional approach to generate the values for the expected value of K^{net} (h) and its upper and lower envelops for a given network .

Cross-K Function

- Cross-K function is for bivariate point patterns as compared to univariate point patterns.
- The network cross-K function describes the relationship between the patterns of two sets of points (e.g., **A** = {a_1, a_2, . . . , a_{na}} and **B** = {b_1, b_2, . . . , b_{nb}}) placed on a finite planar network, and shows whether the set of points **B** affects the location of the set of points **A**.
- The effect can be examined with the following null hypothesis: the set of points **A** is randomly distributed, regardless of the location of the set of points **B**.
- The cross-K function is defined as follows:
- $$K^{ba}(d) = \frac{1}{\rho_a} E(\textit{number of points of A within network distance d of a point } b_i \textit{ in B})$$

(9.25)

Example 9.3: Apply Network K-Functions to Study Ice-Related Crashes and Bridge locations



(Source: Khan et al., 2009)

- Two graphs showing the results of incremental (left) and cumulative (right) cross K-function values up to 1 km from either side of a bridge.
- Y is crash count (incremental or cumulative); x is the distance away from a bridge location)

Above: Barron County; Below: Bayfield County

The results of the K-function analysis show for the northwest counties, suggesting that bridge locations in Barron County are more prone to ice-related crashes than are the locations in Rusk.

Spatial Regression Analysis

- Crash data such as frequency or severity within close proximity in space or time can be correlated due to shared weather, land use characteristics, driver behavior, policies and enforcement practices (Levine et al, 1995).
- In spatial regression, the **residuals are not independent** of each other, but are spatially structured and correlated.
- Applying spatial regression analysis is to explicitly consider the **spatial dependence** of crash observations in the regression model.
- Ignoring the spatial dependence of crash data will result in inefficient, biased and inconsistent parameter estimates and statistical inferences.
- In the highway safety analysis, spatial regression analysis has been used to identify and rank areas with potential improvements for safety, estimate the varying effects of crash contributing factors over the space, and recognize spatial patterns and scopes.

Spatial Regression Techniques

- In general, the two techniques in the spatial regression analysis are:
  - Spatial econometrics models for *continuous spatial data* and
  - Bayesian hierarchical models for *non-negative random count* spatial data.

Spatial Econometrics Methods

- Spatial econometrics is a sub discipline of econometrics that handles spatial autocorrelation and spatial heterogeneity in regression models. Herein, spatial heterogeneity refers to structural instability, either in the form of non-constant error variances (heteroskedasticity) or in the form of regression coefficients.
- They are particularly appealing in modeling crashes because crash events are regarded as the consequence of complicated interactions between many factors: driver, vehicle, roadway, and environment.
- Models used to estimate such geographic phenomena require the specification of spatial effects in econometric models such as
  - the spatial autoregressive model (SAR)
  - the spatial error model (SEM).

Spatial Autoregressive Model

- Autocorrelation means a correlation exists between the values of the same variable.
- The spatial autoregressive model (SAR) (also known as the lagged response model or spatial lag model) deals with spatial autocorrelation by adding an explanatory variable in the form of a spatially lagged dependent variable to a multivariable regression model, as formulated below:

$$\boldsymbol{y} = \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{\rho}\boldsymbol{W}\boldsymbol{y} + \boldsymbol{\varepsilon} \tag{9.26}$$

where $\boldsymbol{y}$ is an n × 1 vector of observations for all locations, *i*; $\boldsymbol{x}$ is an n × k matrix of observations on the explanatory variables, $\boldsymbol{\beta}$ is a k × 1 vector of regression coefficients, ρ is the coefficient of the spatial lag, $\boldsymbol{W}\boldsymbol{y}$ is the spatially lagged dependent variable, and $\boldsymbol{W}$ is the spatial matrix that specifies the spatial dependence between observations; $\boldsymbol{\varepsilon}$ is an n × 1 vector of normally distributed random error terms, with zero mean and constant variance σ^2.

Spatial Autoregressive Model (cont'd)

- Crash count within close proximity in space or time can be correlated due to shared factors, observed or unobserved (e.g., weather, land use characteristics, driver characteristics and behavior, policies and enforcement practices).
- The SAR model expresses the notion that the value of a variable (e.g. crash count) at a given location is related to the values of the same variable measured at nearby locations through the spatial weights matrix $\boldsymbol{W}$.
- The SAR model rearranges and multiplies both sides by $(\boldsymbol{I} - \rho\boldsymbol{W})$, and is reformulated as:

$$\boldsymbol{y} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{x}'\boldsymbol{\beta} + (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{\varepsilon} \qquad (9.27)$$

Spatial Error Model

- In the SEM, the error term - not the dependent variable - is considered as autoregressive. The SEM is formulated as:

$$\boldsymbol{y} = \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ where } \boldsymbol{\varepsilon} = \lambda \boldsymbol{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu} \qquad (9.28)$$

- Note that the error term is made up of a spatially weighted error vector $\lambda\mathbf{W}\boldsymbol{\varepsilon}$ where λ is the spatial autoregressive coefficient and a vector of *i.i.d* random errors **u** with zero mean and variance σ^2.
- We can re-arrange the expression for $\boldsymbol{\varepsilon}$ to obtain: $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{x}'\boldsymbol{\beta}$; and, the spatial error model can be arranged as:

$$\boldsymbol{y} = \boldsymbol{x}'\boldsymbol{\beta} + \lambda \boldsymbol{W}(\boldsymbol{y} - \boldsymbol{x}'\boldsymbol{\beta}) + \boldsymbol{\mu} = \boldsymbol{x}'\boldsymbol{\beta} + \lambda \boldsymbol{W}\boldsymbol{y} - \lambda \boldsymbol{W}\boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{\mu} \qquad (9.29)$$

Spatial Error Model (cont'd)

- In the SEM model, $\boldsymbol{y = x'\beta + \lambda W y - \lambda W x\beta + \mu}$

the dependent variable **y** is a combination of

1. a general (global) linear trend component $\boldsymbol{x'\beta}$,
2. plus a pure spatial autocorrelation component $\lambda\mathbf{W}\mathbf{y}$,
3. plus a (negative) weighted average of predicted neighboring value $\boldsymbol{\lambda W x\beta}$,
4. plus an *i.i.d* random error term $\boldsymbol{u}$.

- So, SEM can be viewed as a form of the mixed SAR with the additional spatial component of the neighboring trend $\boldsymbol{\lambda WX\beta}$.

Example 9.4 Model area-wide crash count with spatial correlation and heterogeneity (Quddus, M.A., 2008)

Zonal-level crash counts (2000-2002) are modeled by traffic characteristics, roadway characteristics, and socio-demographic factors. The spatial autoregressive model (SAR) and the spatial error model (SEM) are formulated as:

SAR: $ln\left(\frac{y_i}{EV_i}\right) = \rho \boldsymbol{W}\, ln\left(\frac{y_i}{EV_i}\right) + \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma^2 \boldsymbol{I})$

SEM: $ln\left(\frac{y_i}{EV_i}\right) = \boldsymbol{x}_i'\boldsymbol{\beta} + \mu_i;\ \mu_i = \lambda \boldsymbol{W}\mu_i + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma^2 \boldsymbol{I})$

where y_i are number of crashes at ward i; EV_i is the exposure variable; $\boldsymbol{x}$ are covariates and $\boldsymbol{\beta}$ is the estimable coefficients. **W** is the spatial weights matrix.

Estimation results for SEM models

Spatial error models (SEM)	Total serious		Total slight		Motorised slight	
	Coef	*t*-stat	Coef	*t*-stat	Coef	*t*-stat
Traffic characteristics						
ln(Traffic flow (PCU km/h))	0.5713	3.08	0.5882	3.57	0.3499	1.56
Average speed (km/h)	−0.0215	−2.77	−0.0163	−1.34	−0.0101	−1.12
Road characteristics						
Length of motorway (km)	0.1025	2.89	0.1228	3.89	0.0730	1.70
Length of A-road (km)	0.0507	2.07	0.0358	1.64	0.0873	2.93
Length of B-road (km)	0.0091	0.40	−0.0071	−0.35	0.0174	0.64
Length of minor road (km)	0.0191	2.57	0.0166	2.51	0.0207	2.29
ln(Road curvature)	–	–	–	–	−1.9405	−0.29
Socio-demographic factors						
ln(Resident population aged less than 60)	0.1209	0.83	0.0562	0.43	0.1656	0.92
ln(Resident population aged 60 or over)	−0.2588	−2.67	−0.2295	−2.64	−0.2493	−2.14
ln(# of employees)	0.1204	3.25	0.1269	3.85	0.0382	0.85
ln(# of households with no cars)	0.3160	3.26	0.3344	3.86	0.1843	1.56
Constant	−16.95	−6.10	−15.14	−6.12	−11.21	−3.33
Spatial autoregressive coef (lamda)	0.8246	16.02	0.8478	17.71	0.7778	13.23
Observations	633		633		633	
Tests for lamda = 0						
Wald test	Reject H0: of no spatial dependence		Reject H0: of no spatial dependence		Reject H0: of no spatial dependence	
Likelihood ratio test	Reject H0: of no spatial dependence		Reject H0: of no spatial dependence		Reject H0: of no spatial dependence	
Lagrange multiplier test	Reject H0: of no spatial dependence		Reject H0: of no spatial dependence		Reject H0: of no spatial dependence	

- Modeling crash rate rather than directly crash count is a compromising strategy because in spatial econometric models, the error term is normally distributed.
- Another issue emerges due to the natural logarithm of the dependent variable crash rate which is a non-negative real. Hence, data with zero counts cannot be modeled.
- The autoregressive coefficient ρ in the SAR model was statistically insignificant, suggesting SAR may not be an appropriate model specification.
- A SEM model with autoregressive coefficient λ was statistically significant.

Generalized Linear Model with Spatial Correlation

- Generalized Linear Mixed Model
- Hierarchical Bayesian Model.

Generalized Linear Mixed Model (GLMM)

- The generalized linear mixed model (GLMM) can accommodate spatial random effects and provide a flexible means of modeling spatially correlated counts.
- The spatial Poisson GLMM model can be specified as:

$$\lambda_i = exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \phi_i\right) \quad (9.31)$$

Where ϕ_i is assumed to have conditional autoregressive (CAR) (Besag 1974 and 1975). A subclass of CAR called the intrinsic conditional autoregressive (ICAR) is used here. The conditional distribution of the random effects is: $\phi_i|\phi_{-i} \sim N\left(\mu_{\phi_i}, \sigma^2_{\phi_i}\right)$

- A more general CAR model is called the Besag York Mollié (BYM) (1991) model. It is a lognormal Poisson model which includes both an ICAR component for spatial smoothing (i.e., spatial autocorrelation) and an ordinary random-effects component for non-spatial heterogeneity.

Hierarchical Bayesian Model

- A more popular approach to modeling crash count with spatial correlation in safety literature is the **hierarchical Bayesian model (HBM) with spatial random effects**.
- HBM are flexible in configuring complicated models by specifying a variety of components and prior distributions in a hierarchical structure.
- HBM is capable of accounting for high data variance due to unobserved heterogeneity, discovering geographic patterns and trends, and increasing accuracy of model estimates through “borrowing strengths” from neighboring sites.

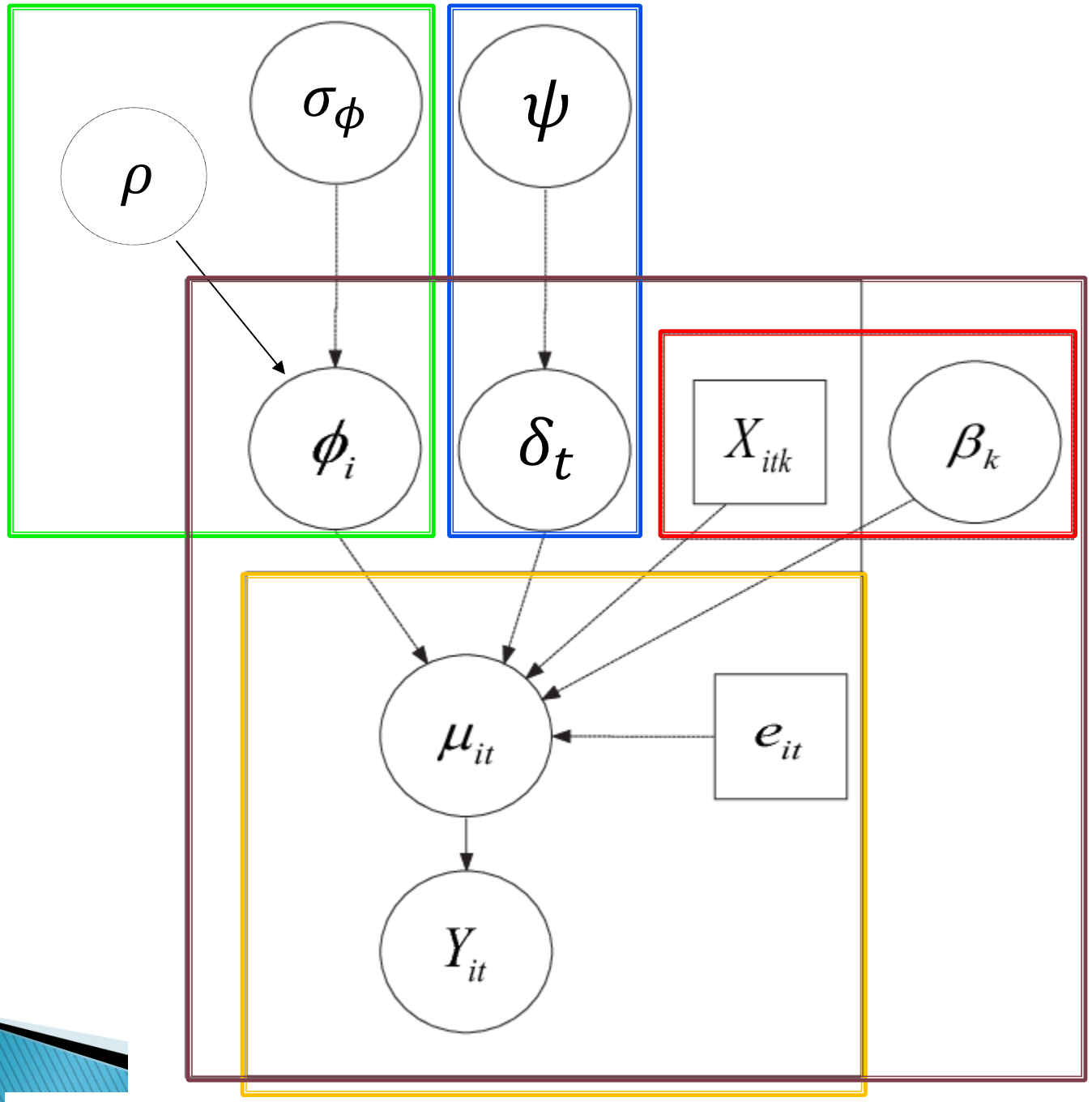
The Rationale Behind…

Safety literature tells us about three mechanisms of spatial effects:

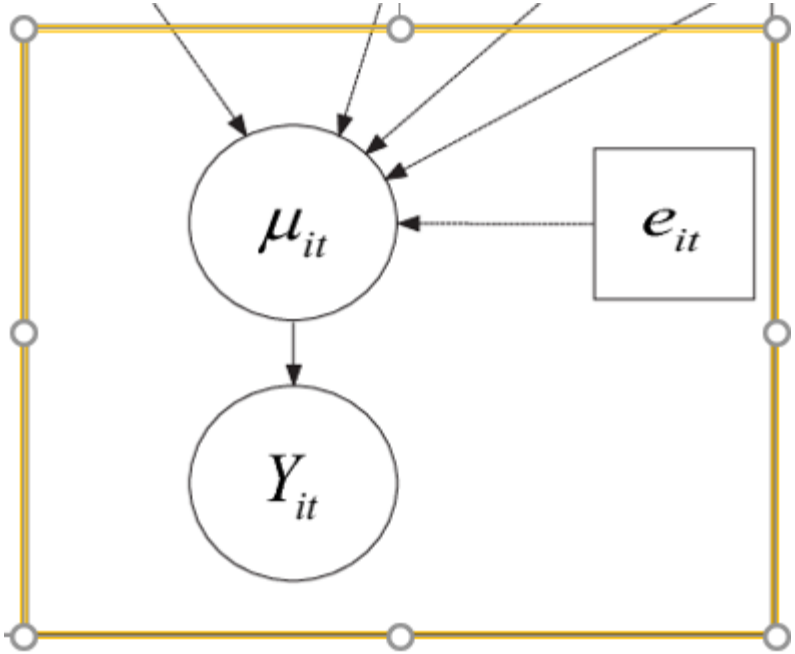
1. interaction, clustering, spillover, externality, diffusion, attraction, and copycat effects via some manners through which actions or phenomena at a given location affect those of other locations;
   - An interesting example of the spillover effect in road safety is the traffic crash "migration effect": the crash rate rises at sites that are untreated but that are "neighbors" to treated sites.
2. varying degrees of measurement errors over space;
3. mis-specification of functional forms for the mean of the response in the model.

However, "*a thorough understanding of the mechanisms of spread remains elusive due to limitations both with the data and development in spatial models*".

Miaou & Song (2005) "Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence", *Accident Analysis and Prevention* 37 (2005) 699–720



Structure of the hierarchical Bayesian Spatial-temporal Model



At the first level of model hierarchy, the number of crashes (Y_{it}) at site *i* and the time period *t* is assumed to be a mutually independent and Poisson distributed random variable and is defined as:

$$y_{it}|\mu_{it} \sim Poisson(\mu_{it})$$

for i = 1,2,…,I and t = 1,2,…,t. (9.33)

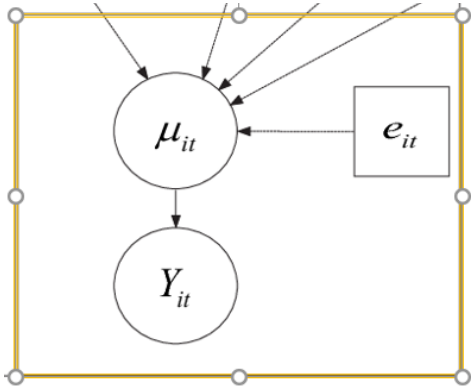
The mean (μ_{it}) of Poisson is modeled as:

$$\mu_{it} = \nu_{it}\lambda_{it} \quad (9.34)$$

The rate λ_{it}, as a non-negative real, can be specified as:

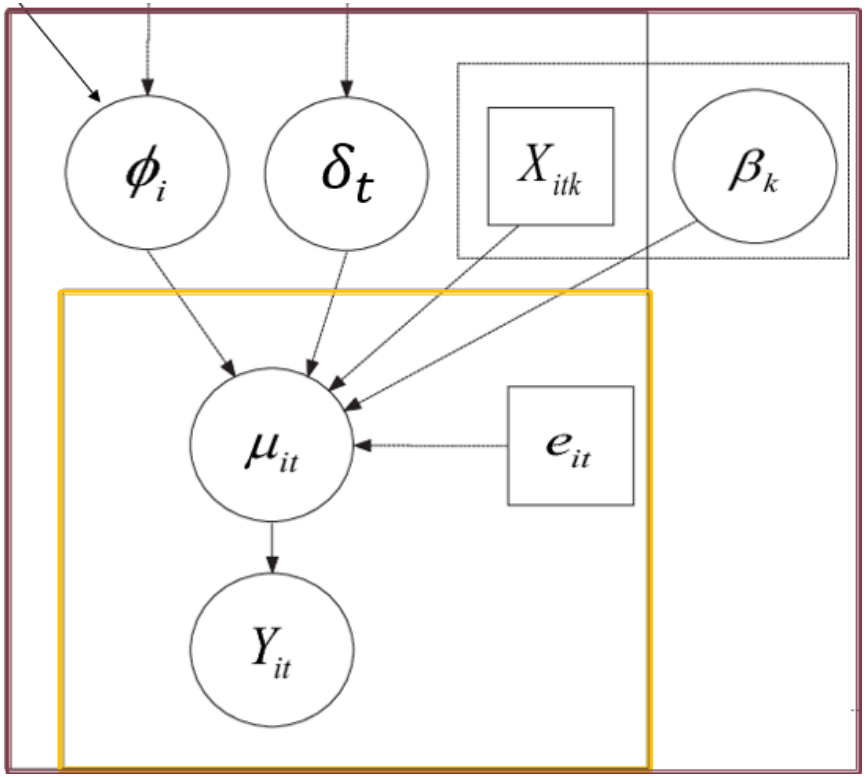
$$\lambda_{it} = exp(\mu_{it} + e_{it}) \quad (9.35)$$

where e_{it} is an exchangeable, unstructured random error.



Configuration And Assumptions of ε_{it}

- ε_{it} can take different forms.
- When the exchangeable random error e_{it} is assumed to be **normal**, $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$, the hierarchical model is the **Poisson-lognormal** model.
- When $exp(\varepsilon_{it})$is assumed to be **gamma**, $exp(\varepsilon_{it}) \sim Gamma(\theta, \theta)$, then the hierarchical model is the **Poisson-gamma** model with the dispersion parameter θ.
- According to Miaou et al. (2003), the gamma distribution is preferred over the lognormal distribution. When θ increases, the amount of overdispersion (due to spatial effects) decreases.
- Non-informative inverse gamma and gamma distributions are assumed for *hyperprior* σ_e^2 and θ, respectively, such as $Ga(0.5, 0.0005)$.
- Gelman has extensively discussed the choice of prior distributions for variance parameters in hierarchical models (Gelman, 2006).



The mean (μ_{it}) of Poisson is modeled as:

$$\mu_{it} = \nu_{it}\lambda_{it} \qquad (9.34)$$

The rate λ_{it}, as a non-negative real, can be specified as:

$$\lambda_{it} = \exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \delta_t + \phi_i + \varepsilon_{it}\right) \qquad (9.35)$$

Where log is the natural logarithm,

X_{itk} are kth explanatory covariate,

βs are the regression coefficients to be estimated from the data,

δ_t represents the time effect,

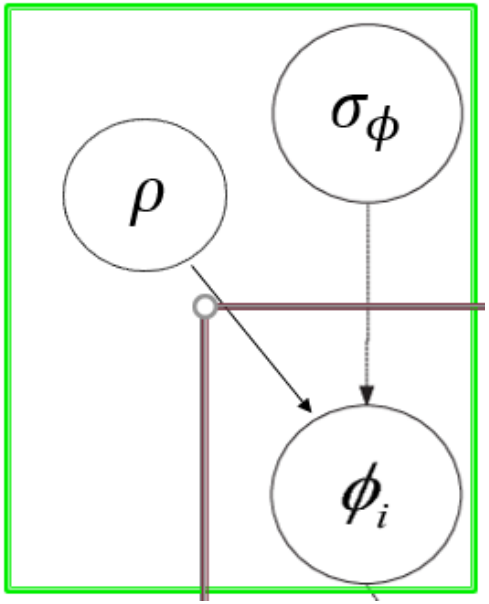
ϕ_i represents the random spatial effect; e_{it} is an exchangeable, unstructured random error.

All ϕ_i, δ_t and ε_{it} are mutually independent.

Hierarchical Bayesian Model

- **At the second level of hierarchy**, the prior distribution of
  - spatial effect, ϕ_i,
  - temporal, δ_t, and
  - unknown coefficients β s

  These are the components that represent the knowledge of the analyst's.
- (**Third level of hierarchy**) Their prior distributions may contain new unknown parameters called *hyper-parameters* whose distributions, also called hyperpriors, constitute the third level of the hierarchy.

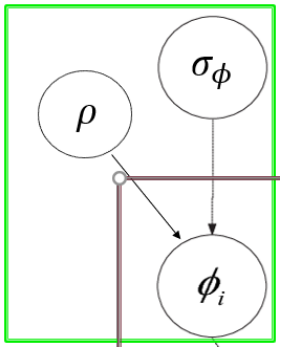


Modeling Spatial Effect ϕ_i

- ϕ_i represents the random spatial effect.
- Besag's Gaussian conditional autoregressive (CAR) model (Besag 1974 and 1975) and its variants are probably the most popular spatial models and the CAR model is formulated as:

$$P(\phi_i|\phi_{-i}) \propto \frac{1}{\sigma_\phi} exp\left\{-\frac{1}{2\sigma_\phi^2}\sum_{j\in C_i} w_{ij}\left(\phi_i - \rho\phi_j\right)^2\right\} \qquad (9.37)$$

Where $P(\phi_i|\phi_{-i})$ is the conditional probability of ϕ_i given ϕ_{-i} and ϕ_{-i} represents all ϕ except ϕ_i.
$\propto$ means "proportional to",
ρ is a parameter that determines the sign and magnitude of the overall spatial dependence.
σ_ϕ^2 controls variability of ϕ and is a fixed effect parameter across all sites.
C_i is a set of sites that are neighbors of site i or have influence on site i and w_{ij} is a spatial weighting factor associated with the pair (i, j),



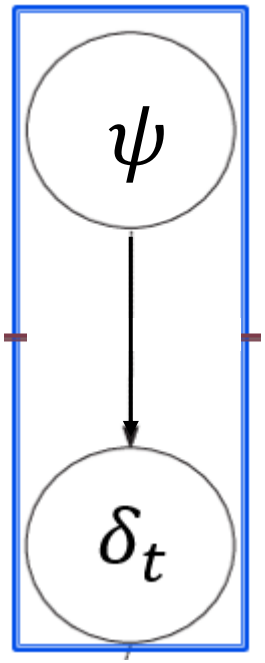
Modeling Spatial Effect ϕ_i (cont'd)

- Alternatively, Gaussian conditional autoregressive (CAR) can also be formulated as

$$p(\phi_i|\phi_{-i}) \propto \exp\left(-\frac{c_{i+}}{2\sigma_\phi^2}\left[\phi_i - \rho\sum_{j\neq i}\frac{c_{ij}}{c_{i+}}\phi_j\right]^2\right)$$

$$i = 1, 2, \ldots, n$$

- Where $C_{i+} = \sum_{j\neq i} C_{ij}$; then, $\phi_i|\phi_{i-1} \sim N\left(\rho\sum_{j\neq i}\left(\frac{c_{ij}}{c_{i+}}\right)\phi_j, \frac{\sigma_\phi^2}{c_{i+}}\right)$
- When ρ is assumed to be 1, Equation 9.37 can be simplified as $\phi_i|\phi_{-i} \sim N\left(\mu_{\phi_i}, \sigma_{\phi_i}^2\right)$. σ_ϕ^2 controls the variability of ϕ, and its inverse is precision (precision=1/variance).
- Most safety literature (Aguero-Valverde and Jovanis, 2006, 2008, 2010; Quddus, 2008) adopts the suggestion in Wakefield et al. (2000) to diffuse the hyperprior density of the precision (or $\frac{1}{\sigma_\phi^2}$): $Ga(0.5, 0.0005)$.



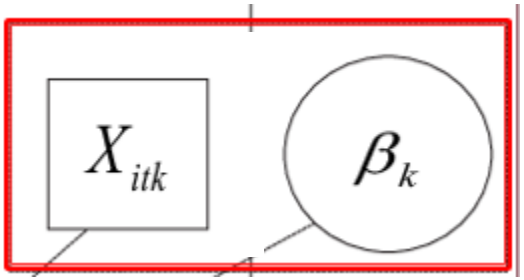
Modeling Temporal Effect: δ_t

- δ_t represents the time effect which is assumed to be fixed with noninformative normal priors.
- Miaou et al.(2003) suggested two types of fixed effects for δ_t:
  - fixed effects that vary by time *t* (e.g., If the data are described in yearly crash count, then use a year-wise fixed effect model with δ_1=0 for the first year as the baseline); and

$$\delta_t = \psi t$$

  - order-one autoregressive (AR(1)) with the same coefficient for all *t* s.

$$\delta_t = \psi \delta_{t-1} + \omega_t$$



Configuration And Assumptions of β

- Covariates x_{itk} can be treated either as fixed effects or random effects. The model that includes both fixed and random effects of covariates is called a mixed effect model.
  - **For fixed effects**, a non-informative prior is assumed (e.g., $\beta \sim N(0, 1000)$ (Aguero-Valverde and Jovanis, 2006, 2008 and 2010), $\beta \sim N(0, 10000)$ (Quddus, 2008)).
  - **For random effects**, the prior for random effects follows a probabilistic structure with *hyperpriors* whose probabilistic distribution needs to be specified. For example, $\beta_k \sim N\left(\mu_{\beta_k}, \sigma^2_{\beta_k}\right)$ and μ_{β_k} and $\sigma^2_{\beta_k}$ are assumed to follow a non-informative normal and inverse gamma distribution, respectively (Miaou et al. (2003), Miaou & Song (2005)).
- Most of the safety literature using hierarchical Bayesian models with random spatial effects treat βs as fixed effects.

Example 9.5 Spatial analysis of fatal and injury crashes in Pennsylvania (Aguero-Valverde et al., 2006)

- Aguero-Valverde et al. developed spatial models of road crash frequency for the State of Pennsylvania at the county level.
- The models include socioeconomic, transportation-related, and environmental factors.
- The results from full Bayesian (FB) hierarchical spatial models are compared with the traditional negative binomial (NB) model.
- Conclusions:
  - no evidence of spatial correlation is found in fatal crashes but statistically significant spatial correlation in injury crashes.
  - the effects of the covariates on fatal and injury crash risk are found to be mostly consistent between the negative binomial and full Bayes models.
  - Variables just marginally significant in the NB models are generally not significant in the FB models.

$$\log(\theta_{ij}) = \alpha + \sum_k \beta_k x_{ijk} + v_i + u_i + (\varphi + \delta_i)t_j \quad (7)$$

Full Bayes model of injury crashes with spatial correlation (u_i), time trend (φ), and space × time interactions (δ_i)

Variable	Estimate		Credible set 95%	
	Mean	S.D.	2.50%	97.50%
Intercept	3.110	0.764	1.593	4.582
DVMT	−0.092	0.013	−0.118	−0.067
Infrastructure mileage	0.323	0.051	0.227	0.425
Mileage density	0.131	0.022	0.086	0.174
Mileage of federal aid roads (%)	0.026	0.007	0.013	0.040
Persons 0–14 (%)	0.047	0.020	0.008	0.088
Persons 15–24 (%)	0.015	0.012	−0.008	0.039
Persons 65 and over (%)	0.013	0.016	−0.018	0.044
Total precipitation	5.86E−05	2.27E−04	−3.85E−04	5.01E−04
σ_u^2	0.176	0.038	0.115	0.263
φ	−0.015	0.004	−0.022	−0.008
σ_δ^2	6.66E−04	2.15E−04	3.31E−04	0.001164

$\bar{D}$, 3261.0; $D(\bar{\theta})$, 3159.0; DIC, 3363.0; and p_D, 102.1.

Modeling Local Relationships in Crash Data

- The geographically weighted regression (GWR) model is a more sophisticated modeling technique to help understand local variations of the data (Fotheringham et al., 2002).
- *Site-specific parameters* are useful in revealing the influence of unobserved data heterogeneity and are more accurate in prediction compared with a global model in which the parameters do not vary across the space.

Geographically Weighted Regression (GWR)

GWR can be formulated as:

$$y_i = \beta_0(\mu_i, \upsilon_i) + \Sigma_k x'_{ik}\beta_k(\mu_i, \upsilon_i) + \varepsilon_i \qquad (9.40)$$

where $\beta_k(u_i, v_i)$ (k = 0, m) are a set of specific coefficients to site (u_i, v_i) of point i; $\boldsymbol{x}_i$ are explanatory variables; y_i are dependent variables; and, ε_i is the error term.

Geographically Weighted Poisson Regression (GWPR)

- The geographically weighted Poisson regression (GWPR) and the geographically weighted negative binomial regression (GWNBR) model are appropriate and consistent with the state of practice in crash count modeling.
- In GWPR or GWNBR, the log transformation of μ_i is:

$$\mu_i = exp\left(\beta_0(\mu_i, \upsilon_i) + \beta_1(\mu_i, \upsilon_i)\upsilon_i + \beta_2(\mu_i, \upsilon_i)x_{i2} + \cdots + \beta_k(\mu_i, \upsilon_i)x_{ik}\right) \quad (9.41)$$

- In GWPR or GWNBR, βs can be formulated as:

$$\boldsymbol{\beta}(\mu_i, \upsilon_i) = \left(\boldsymbol{x}'\boldsymbol{W}(\mu_i, \upsilon_i)\boldsymbol{x}\right)^{-1}\boldsymbol{x}'\boldsymbol{W}(\mu_i, \upsilon_i)\boldsymbol{y}$$

Notes of GWPR

- When calibrating GWR, it is assumed that the observations that are closer to point i are more influential in the estimation of βs than the observations that are farther away (the first law of geography: everything is related to everything else, but near things are more related than distant things" (Tobler 1970)).
- This weight determines how many points will be used to calibrate the coefficients for point i and their contributions in the model calibration.
- The degree of influence is commonly specified through a distance-decayed weight function, W.
- The Gaussian function and bi-square function (adaptive kernel) are commonly used.
- The selection of bandwidth is important, even more important than the choice of kernel density.
- In most of the safety literature on GWRP, either an adaptive kernel or a trial-and-error procedure is adopted to determine the optimal bandwidth (Hadayeghia, et al. 2010; Li et al. 2013; Xu et al., 2018).

Example 9.6 Develop planning level models using Geographically Weighted Poisson Regression (Hadayeghi, A., et al., 2010)

- The purpose is to investigate the spatial variations in the relationship between the number of zonal crashes and the transportation planning predictors using the Geographically Weighted Poisson Regression (GWPR) modeling technique.
- This study was based on 481 traffic analysis zones in 2001, Toronto, CA.
- The dependent variable of each developed model is the number of zonal crashes per year.
- Other data include one-day travel survey data, traffic volume, and street network and land use data.
- The model parameters were estimated using the maximum likelihood method for GWPR in the “GWRx3.0” package.

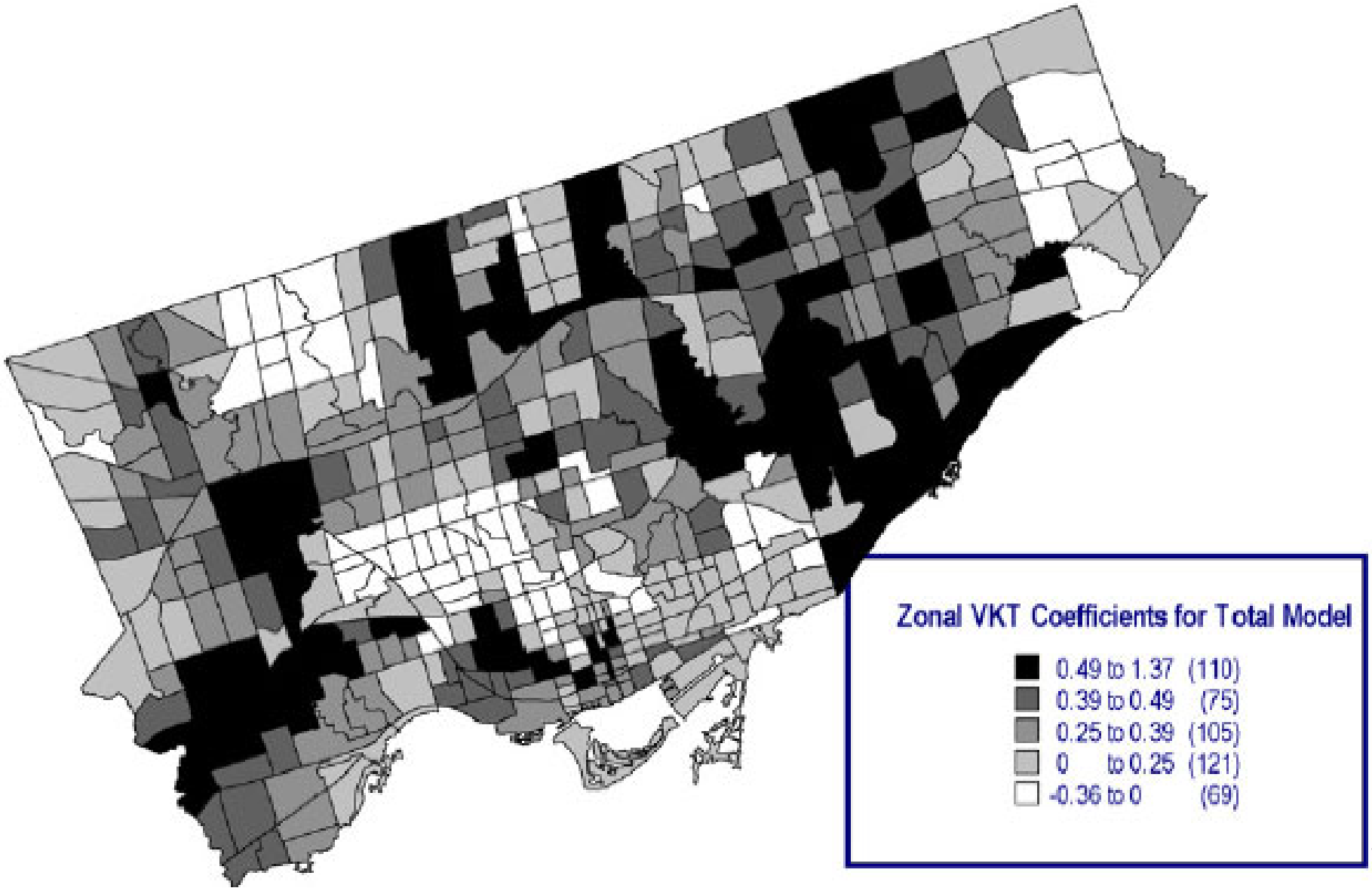
Macro-level GWPR collision prediction models based on traffic and road network variables, total collisions.

Parameters	GWPR model number							
	#1	#2	#3	#4	#5	#6	#7	#8
ln(*A*)	−6.4, 8.10 (1.09, 2.4,8.1)	1.06, 4.60 (3.31,3.84,4.14)	−3.23, 6.92 (1.2,2.3,3.3)	−6.42, 7.41 (1.32,2.53,3.74)	−4.58, 7.31 (1.55,2.67,3.74)	−3.81, 5.88 (1.50,2.56,3.47)	−3.97, 6.23 (1.41,2.42,3.47)	−2.839, 5.526 (1.49,2.50,3.40)
ln(VKT)	−0.36, 1.37 (0.12,0.33,0.47)	−0.009, 0.38 (0.03,0.07,0.13)	−0.33, 0.89 (0.14,0.27,0.38)	−0.32, 1.298 (0.06,0.23,0.38)	−0.26,1.17 (0.07,0.22,0.35)	−0.224, 0.979 (0.08,0.21,0.34)	−0.22, 0.99 (0.09,0.21,0.34)	−0.165, 0.857 (0.10,0.21,0.34)
Total arterial road kilometers		0.07, 0.33 (0.17,0.22,0.26)						
Total expressway kilometers		−0.15, 0.23 (0.05,0.08,0.12)						
Total collector kilometers		−0.02, 0.17 (0.05,0.09,0.12)						
Total laneway kilometers		−0.023, 0.48 (0.04,0.06,0.10)						
Total local road kilometers		−0.048, 0.016 (−0.02,−0.009,0.002)						
Total ramp kilometers		−0.29, 0.26 (0.01,0.07,0.13)						
Total road kilometers			−0.04, 0.16 (0.03,0.04,0.06)			−0.060, 0.175 (0.01,0.03,0.05)	−0.043, 0.156 (0.01,0.03,0.05)	−0.038, 0.118 (0.01,0.03,0.05)
Number of 4-legged signalized intersections				−0.165, 0.487 (0.08,0.14,0.21)		−0.083, 0.372 (0.05,0.10,0.17)	−0.098, 0.350 (0.05,0.10,0.16)	−0.069, 0.347 (0.05,0.09,0.15)
Number of 3-legged signalized intersections				−0.514, 0.564 (0.04,0.15,0.23)		−0.195, 0.543 (0.06,0.13,0.19)	−0.23, 0.474 (0.06,0.12,0.18)	−0.269, 0.419 (0.06,0.13,0.19)
Total number of signalized intersections					−0.067, 0.348 (0.09,0.13,0.18)			
Total rail kilometers								−0.56, 1.12 (−0.11,0.01,0.15)
Number of schools							−0.26, 0.24 (−0.06,0.00,0.05)	
GWPR AICc	16,304	13,404	14,602	8,137	9,910	8,220	7,629	8,032
Global AICc	27,410	20,431	24,373	20,041	20,161	18,805	18,634	18,779

Minimum, Maximum (Lower Quartile, Median, Upper Quartile).

- It is clear from the table that the signs of Traffic Exposure (e.g. VKT) coefficients for each TAZ are not always the same (underlined in red).
- This is counterintuitive because traffic exposure is expected to have a positive effect on the number of crashes; therefore, its coefficient should be positive.

The figure depicts the local parameters of VKT for the total crash. The parameters clearly demonstrate spatial variations across the city.



Example: Modeling crash spatial heterogeneity: Random parameter versus geographically weighting (Xu, P., et al., 2015)

- Modeling location-referenced crash data may need to consider:
  - Spatial dependence between crash observations; and
  - Spatial heterogeneity in the relationships
- This study developed four types of models and quantitatively investigated spatial heterogeneity in regional safety modeling using two approaches:
  - negative binomial model (NB)
  - Bayesian negative binomial model with conditional autoregressive prior (NB-CAR)
  - random parameter negative binomial model (RPNB)
  - semi-parametric geographically weighted Poisson regression model (S-GWPR).

Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting Accid. Anal. Prev. 16–25

- The two methods have intrinsic differences:
  - the local regression coefficients in RPNB are drawn independently from some univariate distributions, and not necessarily referred to specific locations,
  - the local coefficients in GWPR are assumed to be a function of the coordinates in geographical space.
- Based on a 3-year data set from the county of Hillsborough, Florida.

Models with total crashes frequency as the dependent variable.

Total crashes	NB	CAR	RPNB						S-GWPR					
			Mean	Min	Lwr	Med	Upr	Max	Mean	Min	Lwr	Med	Upr	Max
Intercept	**4.15**	**4.10**	**4.10**						4.11	2.08	3.89	4.18	4.40	6.03
LnDVMT	**0.62**	**0.63**	**0.68**	0.40	0.67	0.69	0.71	0.85	0.68	0.15	0.56	0.65	0.76	1.38
s.d._LnDVMT			**0.18**											
Inter_density	***0.05***	***0.07***	**0.06**						0.26	−2.13	−0.02	0.23	0.57	2.70
P_seglen25	−0.05	−0.03	−0.03	−0.21	−0.04	−0.03	−0.02	0.11	0.02	−0.55	−0.10	0.01	0.17	0.49
s.d._P_seglen25			**0.14**											
P_seglen45	**0.06**	**0.07**	***0.05***						0.06					
P_seglen55_65	−0.01	−0.02	−0.03						−0.08	−4.70	−0.17	−0.04	0.13	1.06
POP_density	**0.26**	**0.20**	**0.26**						0.12	−0.81	0.04	0.15	0.25	1.09
MHINC	**−0.11**	**−0.08**	**−0.15**	−0.36	−0.17	−0.15	−0.13	0.15	−0.16	−0.76	−0.30	−0.16	−0.01	0.34
s.d._MHINC			**0.19**											
Over-dispersion	**0.40**	**0.30**	**0.32**											
CAR effects		**0.35**												

Note: the italicized bold numbers mean statistically significant at 90% significance level in the NB, CAR and RPNB; while the bold ones mean statistical significant at 95% significance level; min, lwr, med, upr and max refer to the minimum, lower quartile, median, upper quartile and maximum of values in the local parameters, and all other abbreviations are defined as in Table 1.

Measures of model goodness of fit.

	Total crashes models			Severe crashes models		
	R_d^2	MAD	AIC	R_d^2	MAD	AIC
NB	0.59	35.14	18923	0.52	3.86	2259
CAR	0.75	28.42	–	0.77	2.59	–
RPNB	0.76	27.86	12385	0.69	3.08	1755
S-GWPR	**0.80**	**25.23**	**10237**	**0.81**	**2.36**	**1428**

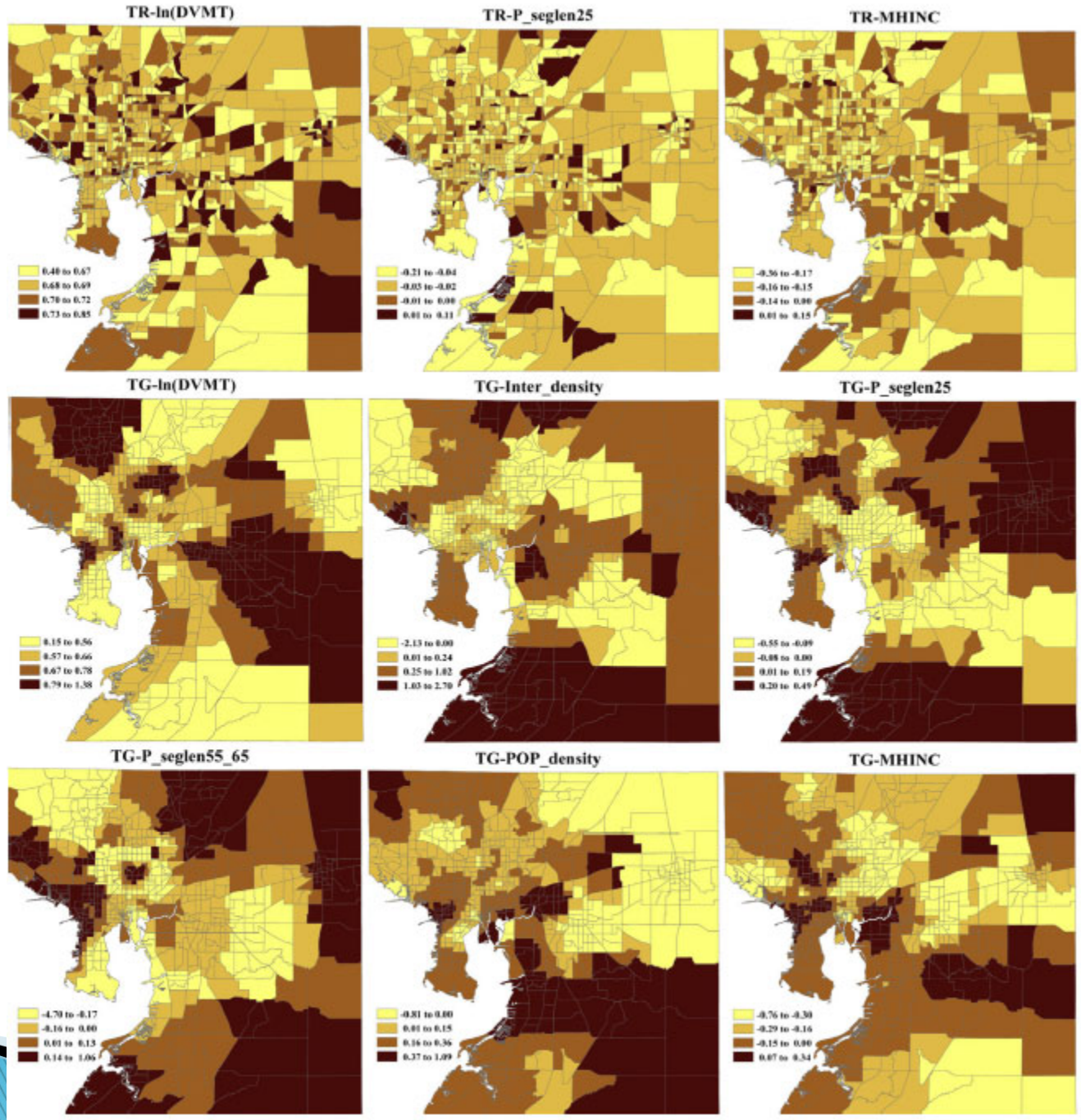


Fig. 1. Local estimates of predicting variables in RPNB and S-GWPR with total crashes frequency as the dependent variable.

The authors concluded:

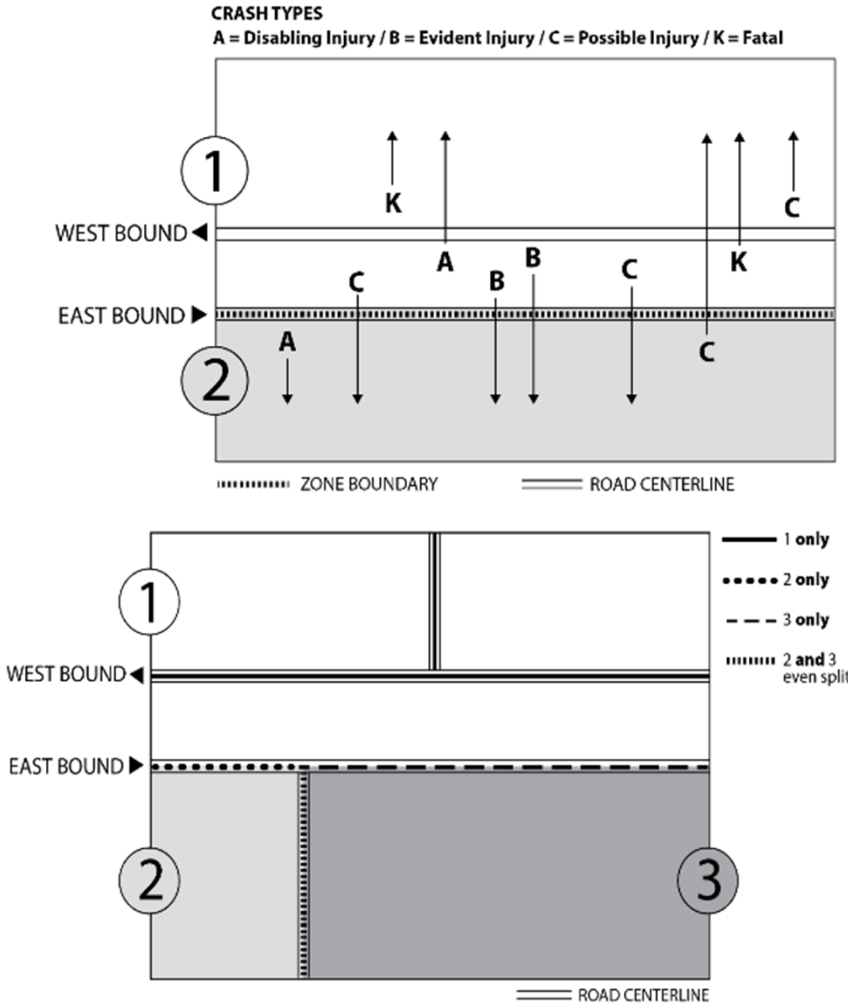
- Both RPNB and S-GWPR successfully capture the spatially varying relationship, but the two methods yield notably different sets of results.
- S-GWPR performs best with the highest value of R^2 as well as the lowest mean absolute deviance and Akaike information criterion measures. Whereas the RPNB is comparable to the CAR, in some cases, it provides less accurate predictions.
- A moderately significant spatial correlation is found in the residuals of RPNB and NB, implying their inadequacy in accounting for the spatial correlation.
- Despite S-GWPR's superior performance, the models calibrated in the study are not spatially transferable because they produce a set of local parameters for a specific geographic region.

Thoughts about Future Research

- Transport applications of big data, cloud computing and connected & autonomous vehicle technologies that can be used to provide a more **integrated** spatial environment.
- Standardize **crash data and built environmental** data collection and processing such as crash data assignment, area type classifications and socioeconomics, land use, extent of multimodal transportation infrastructure, area-wide operational characteristics and strategies; and develop better understanding of their safety and equity implications.

Thoughts about Future Research (cont'd)

- **Crash assignment** acknowledges that boundary roads are a product of both adjoining zones.
  - Crashes are assigned *randomly, evenly* to each zone.
  - Road data (i.e., vehicle miles traveled - VMT) are split between boundary zones.
- Avoiding duplicate data lends analysis to target setting.
  - Allows users to aggregate zones to larger geographies (e.g., counties).
- Integration produces a trivial amount of "error."
  - No more than 1 or 2% of baseline conditions.



Source: Macro-Level Safety Planning Analysis Chapter for the Highway Safety Manual (NCHRP 17-81)

Thoughts about Future Research (cont'd)

Transferability of results of spatial analysis is the creation of common frameworks for the two famous problems (boundary and MAUP).

- How and why were boundaries handled the way they were (i.e., Modifiable Area Unit Problem – MAUP – etc.)?
  - Keep it simple, repeatable while acknowledging reality on the ground.
  - Preliminary investigation showed minimal differences between schemes.
  - Consistent results across agencies in final models.
  - Subsequent analysis shows similar results using different zone types.
- Why choose a specific areal unit (e.g., census block) for this effort?
  - Consistent (and nested) geographic definitions.
  - Consistent data definitions.
  - Publicly available and easily accessible.
  - Existing component of transportation planning practice (i.e., Census informs agencies' travel demand model process).

Thoughts about Future Research (cont'd)

- **Spatial dependence and spatial heterogeneity** affect the coefficient in different directions: can we distinguish them; is it necessary to separate their effects; and how can we separate them?
- The determination of the **spatial impacts** of implemented road safety measures would be thoroughly studied.
- More **thoughtful study** design:
  - Higher vehicle speed is an indicator of safer driving condition or a contributing factor to crashes?
  - Crashes involving pedestrian was revealed to be significantly affected by more location-related factors, while pedestrian origin was revealed to be significantly affected by more demographic related factors.

References

1. Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accid. Anal. Prev. 38 (3), 618–625.
2. Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. Transport. Res. Rec. 2061 (1), 55–63.
3. Aguero-Valverde, J., Jovanis, P.P., 2010. Spatial correlation in multilevel crash frequency models: effects of different neighboring structures. Transport. Res. Rec. 2165 (1), 21–32.
4. Anselin, L., 1995. Local indicators of spatial association – LISA. Geogr. Anal. 27, 93–115.
5. Banergjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Modeling and Analysis for Spatial Data. CRC Press.
6. Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. Roy. Stat. Soc. B 36 (2), 192–225.
7. Besag, J., 1975. Statistical analysis of non-lattice data. J. Roy. Stat. Soc. Series D (The Stat.) 24 (3), 179–195.
8. Besag, J., York, J., Molli´e, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Stat. Math. 43 (1), 1–20.
9. Bowman, A.W., 1984. An alternative method of cross-validation for the smoothing of density estimates. Biometrika 71, 353–360.
10. Charlton, M., Fotheringham, A.S., Brunsdon, C., 2003. Software for Geographically Weighted Regression. University of Newcastle, Newcastle, United Kingdom.
11. Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York.
12. da Silva, A.R., Rodrigues, T.C.V., 2014. Geographically weighted negative binomial regression—incorporating overdispersion. Stat. Comput. 24 (5), 769–783.
13. Flahaut, B., Mouchart, M., Martin, E.S., Thomas, I., 2003. The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. Accid. Anal. Prev. 35 (6), 991–1004.
14. Fotheringham, A.S., Brunsdon, C., Charlton, M.E., 2000. Quantitative Geography, Perspectives on Spatial Data Analysis. SAGE.
15. Fotheringham, A.S., Brunsdon, C., Charlton, M.E., 2002. Geographically Weighted Regression: the Analysis of Spatially Varying Relationship. Wiley, Chichester.
16. Gelfand, A.E., Vounatsou, P., 2003. Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics 4 (1), 11–15.
17. Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal. 1 (3), 515–534.
18. Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24 (No. 3), 189–206.
19. Gomes, M.J.T.L., Cunto, F., da Silva, A.R., 2017. Geographically weighted negative binomial regression applied to zonal level safety performance models. Accid. Anal. Prev. 106, 254–261.
20. Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. Accid. Anal. Prev. 42 (2), 676–688.
21. Khan, G., Qin, X., Noyce, D.A., 2008. Spatial analysis of weather crash patterns. J. Transport. Eng. 134 (5), 191–202.
22. Khan, G., Santiago-Chaparro, K.R., Qin, X., Noyce, D.A., 2009. Application and integration of lattice data analysis, network K-functions, and geographic information system software to study ice-related crashes. Transport. Res. Rec. 2136 (1), 67–76.
23. Levine, N., Kim, K.E., Nitz, L.H., 1995. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. Accid. Anal. Prev. 27 (5), 675–685.
24. Li, Z., Wang, W., Pan, L., Bigham, J.M., Ragland, D.R., 2013. Using geographically weighted Poisson regression for county-level crash modeling in California. Saf. Sci. 58, 89–97.
25. Miaou, S.P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accid. Anal. Prev. 37, 699–720.
26. Miaou, S.P., Song, J.J., Mallick, B.K., 2003. Roadway traffic crash mapping: a space-time modeling approach. J. Transport. Stat. 6 (1), 33–57 1094-8848.
27. Moran, P.A.P., 1950. Notes on continuous Stochastic phenomena. Biometrika 37 (1), 17–23 10.2307/2332142. JSTOR 2332142.
28. Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically weighted Poisson regression for disease association mapping. Stat. Med. 24 (17), 2695–2717.
29. Okabe, A., Yamada, I., 2001. The K-function method on a network and its computational implementation. Geogr. Anal. 33 (3), 271–290.
30. Okabe, A., Okunuki, K.I., Shiode, S., 2006. SANET: a toolbox for spatial analysis on a network. Geogr. Anal. 38 (1), 57–66.
31. Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geogr. Anal. 27, 286–306.
32. Qin, X., Han, J., Zhu, J., 2009. Spatial analysis of road weather safety data using a Bayesian hierarchical modeling approach. Adv. Transport. Stud. 18, 69–78.
33. Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accid. Anal. Prev. 40 (4), 1486–1497.
34. Ripley, B.D., 1976. The second-order analysis of stationary point process. J. Appl. Probab. 13, 255–266.
35. Silva, A.R., Rodrigues, T.C.V., 2016. A SAS. Available at: http://support.sas.com/resources/papers/proceedings16/8000-2016.pdf.
36. Spiegelhalter, D., Best, N., Carlin, B.P., Linde, A., 2002. Bayesian measures of model complexity and fit. J. Roy. Stat. Soc. B 64 (4), 583–639.
37. Tobler, W., 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. 46, 234–240 Supplement.
38. Wakefield, J.C., Best, N.G., Waller, L., 2000. Bayesian Approaches to Disease Mapping, on Spatial Epidemiology: Methods and Applications. Oxford University Press.
39. Xu, P., Huang, H., 2015. Mo[illegible] crash spatial heterogeneity: random parameter versus geographically weighting Accid. Anal. Prev. 16–25.
40. Yamada, I., Thill, J.C., 2004. Comparison o[illegible] and network K-functions in traffic accident analysis. J. Transport Geogr. 12 (2), 149–158.