Chapter 1

WHAT'S SPATIAL ABOUT SPATIAL DATA MINING: THREE CASE STUDIES

Shashi Shekhar, Yan Huang, Weili Wu, C.T. Lu

Abstract Spatial data mining is the process of discovering interesting and previously unknown, but potentially useful, patterns from large spatial datasets. Extracting interesting and useful patterns from spatial datasets is more difficult than extracting the corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation. A popular approach is to apply classical data mining techniques after transforming spatial components into non-spatial components via feature selection. An alternative is to explore new models, new objective functions, and new patterns which are more suitable for spatial data and their unique properties. This chapter investigates techniques in the literature to incorporate spatial components via feature selection, new models, new objective functions, and new patterns.

Keywords: spatial data mining, feature selection, spatial databases, co-location rules, spatial autocorrelation, spatial outliers

1. INTRODUCTION

Widespread use of spatial databases [Gut94, SC01, SCR+99, Wor95] is leading to an increasing interest in mining interesting and useful, but implicit, spatial patterns [Gre00, KAH96, Mar99, RS99, SNM+95]. Spatial data sets and patterns are abundant in many application domains related to NASA, the National Imagery and Mapping Agency(NIMA), the National Cancer Institute(NCI), and the United States Department of Transportation(USDOT). Efficient tools for extracting information from geo-spatial data are crucial to organizations which make decisions based on large spatial data sets. These organizations are spread across many domains including ecology and environment management, pub-

lic safety, transportation, public health, business, travel, and tourism [AM95, HGL93, IES89, Kru95, Hai89, SYH93, SNM+95, YL97].

Extracting interesting and useful patterns from spatial datasets is more difficult than extracting corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation.

There are different representations for spatial data. Because many datasets are collected from satellites, aerial-photographs, and Digital Elevation Maps (DEM), much spatial information is stored as thematic maps which are qualitative and categorical **raster** data, usually maps of land cover classes such as temperature, rainfall, pasture, urban areas, and standing water. Coupled with this **raster** data representation is **vector** data, the other common GIS data model, made up of points, lines, or polygons, with associated attributes. Spatial data mining deals with not only data types such as integers, dates and strings, but also complex data types like points, lines, and polygons. Furthermore, relations between spatial objects add another dimension to the complexity of spatial data mining. Basic spatial relations include metric (e.g. distance), directional (e.g. north of), and topological (e.g. adjacent).

Traditional data mining algorithms [Agr94] often make assumptions (e.g. independent, identical distributions) which violate Tobler's first law of Geography: everything is related to everything else, but nearby things are more related than distant things [Tob79]. In other words, the values of attributes of nearby spatial objects tend to systematically affect each other. In spatial statistics, an area within statistics devoted to the analysis of spatial data is called spatial autocorrelation [Cre93]. Scientists and researchers in several disciplines have created, adapted, and applied statistical techniques to spatial data. For example, in image processing and vision, Markov Random Fields(MRFs) is a popular model to incorporate context for image segmentation and classification. Economists use spatial autoregression models to predict and estimate trends in regional economies. The variogram, a tool to capture spatial information in data, is widely used in geography and remote sensing. In spatial data mining, knowledge discovery techniques which ignore spatial autocorrelation typically perform poorly in the presence of spatial data. The models derived may turn out to be not only biased and inconsistent, but may also be a poor fit to the data set.

In this paper, we provide three case studies in spatial data mining. The first case study examines the classification of spatial datasets. The second case study describes the generalization of association rules to spatial co-location patterns. The third case study focuses on detecting

spatial outliers. In each case study, we bring out unique challenges of spatial data mining relative to classical data mining.

2. SPATIAL CLASSIFICATION

Given a set of data (a training set) with one attribute as the dependent attribute, the classification task is to build a model to predict the unknown dependent attributes of future data based on other attributes as accurately as possible. A set of data (different from the training set) with dependent attributes known in advance is used to validate the model. In spatial classification, the attribute properties of neighboring objects may also have an effect on the membership of objects.

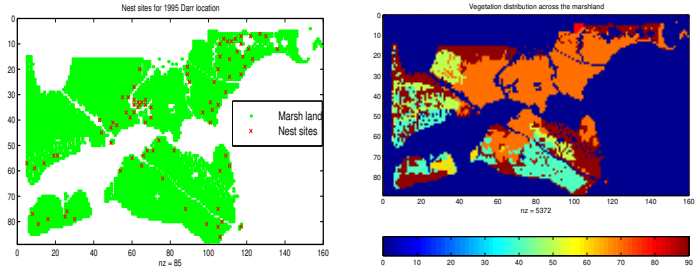
2.1 AN ILLUSTRATIVE APPLICATION DOMAIN

We now introduce an example to illustrate the different concepts in spatial data mining. We are given data about two wetlands, named Darr and Stubble, on the shores of Lake Erie in Ohio USA in order to *predict* the spatial distribution of a marsh-breeding bird, the red-winged blackbird (*Agelaius phoeniceus*). The data was collected from April to June in two successive years, 1995 and 1996.

A uniform grid was imposed on the two wetlands and different types of measurements were recorded at each cell or pixel. In total, values of seven attributes were recorded at each cell. Domain knowledge is crucial in deciding which attributes are important and which are not. For example, *Vegetation Durability* was chosen over *Vegetation Species* because specialized knowledge about the bird-nesting habits of the red-winged blackbird suggested that the choice of nest location is more dependent on plant structure, plant resistance to wind, and wave action than on the plant species.

Our goal is to build a model for predicting the location of bird nests in the wetlands. Typically the model is built using a portion of the data, called the **Learning** or **Training** data, and then tested on the remainder of the data, called the **Testing** data. In the learning data, all the attributes are used to build the model and in the testing data, one value is *hidden*, in our case the location of the nests.

We focus on three independent attributes, namely *Vegetation Durability*, *Distance to Open Water*, and *Water Depth*. The spatial distribution of *Vegetation Durability* and the actual nest locations for the Darr wetland in 1995 are shown in Figure 1.1. These maps illustrate the following two important properties inherent in spatial data.



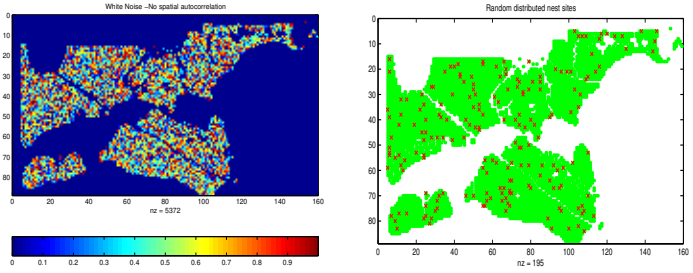
(a) Nest Locations (b) Vegetation Durability

Figure 1.1 (a) Learning dataset: The geometry of the wetland and the locations of the nests, (b) The spatial distribution of *vegetation durability* over the marshland

1. The values of *Vegetation Durability* and the actual nest locations which are referenced by spatial location tend to vary gradually over space (values of *Distance to Open Water* and *Water Depth* are similar). While this may seem obvious, classical data mining techniques, either explicitly or implicitly, assume that the data is *independently* generated. For example, the maps in Figure 1.2 show the spatial distribution of attributes if they were independently generated. Classical data mining techniques like logistic regression [OM97] and neural networks [OO99] were applied to build spatial habitat models. Logistic regression was used because the dependent variable is binary (nest/no-nest) and the logistic function "squashes" the real line onto the unit-interval. The values in the unit-interval can then be interpreted as probabilities. The study concluded that with the use of logistic regression, the nests could be classified at a rate 24% better than random [OO99].

2. The spatial distributions of attributes sometimes have distinct local trends which contradict the global trends. This is seen most vividly in Figure 1.1(b), where the spatial distribution of *Vegetation Durability* is jagged in the western section of the wetland as compared to the overall impression of uniformity across the wetland. This property is called spatial heterogeneity. In section 3.2 we describe a measure which quantifies the notion of spatial autocorrelation.

The fact that classical data mining techniques ignore spatial autocorrelation and spatial heterogeneity in the model-building process is one reason why these techniques do a poor job. A second, more subtle but



(a) pixel property with independent identical distribution

(b) Random nest locations

Figure 1.2 Spatial distribution satisfying random distribution assumptions of classical regression

equally important reason is related to the choice of the objective function to measure classification accuracy. For a two-class problem, the standard way to measure classification accuracy is to calculate the percentage of correctly classified objects. However, this measure may not be the most suitable in a spatial context. *Spatial accuracy*—how far the predictions are from the actuals—is as important in this application domain due to the effects of the discretizations of a continuous wetland into discrete pixels, as shown in Figure 1.3. Figure 1.3(a) shows the actual locations of nests and 1.3(b) shows the pixels with actual nests. Note the loss of information during the discretization of continuous space into pixels. Many nest locations barely fall within the pixels labeled 'A' and are quite close to other blank pixels, which represent 'no-nest'. Now consider two predictions shown in Figure 1.3(c) and 1.3(d). Domain scientists prefer prediction 1.3(d) over 1.3(c), since the predicted nest locations are closer on average to some actual nest locations. The classification accuracy measure cannot distinguish between 1.3(c) and 1.3(d), and a measure of spatial accuracy is needed to capture this preference.

A simple and intuitive measure of spatial accuracy is the Average Distance to Nearest Prediction (ADNP) from the actual nest sites, which can be defined as

$$ADNP(A,P) = \frac{1}{K}\sum_{k=1}^{K} d(A_k, A_k.nearest(P)).$$

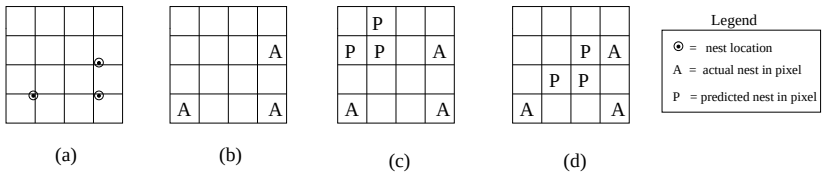


Figure 1.3 (a)The actual locations of nests, (b)Pixels with actual nests, (c)Location predicted by a model, (d)Location predicted by another model. Prediction(d) is spatially more accurate than (c).

Here A_k represents the actual nest locations, P is the map layer of predicted nest locations and $A_k.nearest(P)$ denotes the nearest predicted location to A_k. K is the number of actual nest sites.

2.2 APPROACHES OF MODELING SPATIAL DEPENDENCIES

Decision Tree Approaches. A spatial classification algorithm is proposed in [EKS97]. This method is based on the well-known ID3 [Qui86] algorithm. It employs the "neighborhood" concept and extends the attributes by considering not only properties of the classified objects, but also the attribute values of neighboring objects. Objects are considered neighbors if they satisfy some neighborhood relations such as *overlap, close-to, east, etc.* For instance the economic power of the city can be classified based on the types of neighboring cities. In this example, the dependent attribute is *economic power*. Both the attributes of object city (*e.g. population of city, amount of taxes of city*) and the attributes of neighborhing cities (*e.g. type of neighbor of city, type of neighbor of neighbor of city*) are considered by this spatial classification algorithm. The algorithm "materializes" spatial relationship as attributes, but it does not analyze spatial autocorrelation, which is important property of the spatial data.

Kopersiki, Han, and Stefanovic presented an efficient method for spatial classification [KHS98]. The authors proposed some optimization techniques like the "Two-Step" spatial computation approach to build decision trees. This algorithm includes the spatial relationship as a predicate in decision tree. More accurate and efficient decision trees can be produced via this two-step approach where coarse computation are performed to get a sample of approximate spatial predicates followed by fine computations done only for promising patterns. However, this algorithm does not take into account spatial autocorrelation and does not explicitly factor spatial properties into the model.

Logistic Regression Modeling. Given an $n-$vector $\mathbf{y}$ of observations and an $n \times m$ matrix $\underline{X}$ of explanatory data, classical linear regression models the relationship between y and $\underline{X}$ as

$$\mathbf{y} = \mathbf{X}\beta + \epsilon.$$

Here $\mathbf{X} = [1, \underline{X}]$ and $\beta = (\beta_0, \ldots, \beta_m)^t$. The standard assumption on the error vector ϵ is that each component is generated from an independent, identical and normal distribution, i.e, $\epsilon_i = N(0, \sigma^2)$.

When the dependent variable is binary, as is the case in the "bird-nest" example, the model is transformed via the logistic function and the dependent variable is interpreted as the probability of finding a nest at a given location. Thus, $Prob(y = 1) = \frac{e^{X\beta}}{1+e^{X\beta}}$. This transformed model is referred to as **logistic** regression.

The fundamental limitation of classical regression modeling is that it assumes that the sample observations are independently generated. This may not be true in the case of spatial data. As we have shown in our example application, the explanatory and the independent variables show a moderate to high degree of spatial autocorrelation(see Figure 1.1). The inappropriateness of the independence assumption shows up in the residual errors, the ϵ_i's. When the samples are spatially related, the residual errors reveal a systematic variation over space, i.e., they exhibit high spatial autocorrelation. This is a clear indication that the model was unable to capture the spatial relationships existing in the data. Thus the model is a poor fit to the data. Incidently, the notion of spatial autocorrelation is similar to that of time autocorrelation in time series analysis but is more difficult to model because of the multi-dimensional nature of space.

Spatial Autocorrelation and Examples.. Many measures are available for quantifying spatial autocorrelation. Each has strengths and weaknesses. Here we briefly describe Moran's I measure.

In most cases, Moran's I measure (henceforth MI) ranges between -1 and +1 and thus is similar to the classical measure of correlation. Intuitively, a higher absolute value indicates high spatial autocorrelation. A high positive value implies that like classes tend to cluster together or attract each other. A low negative value indicates that high and low values are interspersed. Thus like classes are de-clustered and tend to repel each other. A value close to zero is an indication that no spatial trend (random distribution) is discernible using the given measure.

All spatial autocorrelation measures are crucially dependent on the choice and design of the contiguity matrix W. The design of the matrix itself reflects the influence of neighborhood. Two common choices are

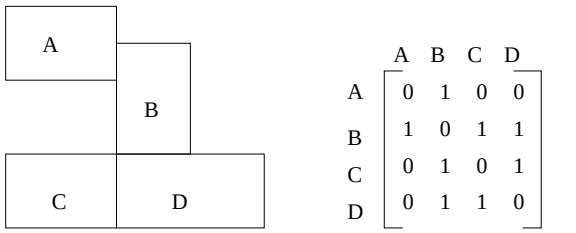


Figure 1.4 A spatial neighborhood and its contiguity matrix

the four and the eight-neighborhood. Thus given a lattice structure and a point S in the lattice, a four-neighborhood assumes that S influences all cells which share an edge with S. In an eight-neighborhood, it is assumed that S influences all cells which either share an edge or a vertex. An eight neighborhood contiguity matrix is shown in Figure 1.4. The contiguity matrix of the uneven lattice (left) is shown on the right hand-side. The contiguity matrix plays a pivotal role in the spatial extension of the regression model.

Spatial Autoregression Model(SAR). We now show how spatial dependencies are modeled in the framework of regression analysis.

This framework may serve as a template for modeling spatial dependencies in other data mining techniques. In spatial regression, the spatial dependencies of the error term, or, the dependent variable, are directly modeled in the regression equation [Ans88]. Assume that the dependent values y_i' are related to each other, i.e., $y_i = f(y_j)$ $i \neq j$. Then the regression equation can be modified as

$$\mathbf{y} = \rho W\mathbf{y} + \mathbf{X}\beta + \epsilon.$$

Here W is the neighborhood relationship contiguity matrix and ρ is a parameter that reflects the strength of spatial dependencies between the elements of the dependent variable. After the correction term $\rho W\mathbf{y}$ is introduced, the components of the residual error vector ϵ are then assumed to be generated from independent and identical standard normal distributions.

We refer to this equation as the ***Spatial Autoregressive Model (SAR).*** Notice that when $\rho = 0$, this equation collapses to the classical regression model. The benefits of modeling spatial autocorrelation are many: (1) The residual error will have much lower spatial autocorrelation, i.e., systematic variation. With the proper choice of W, the residual error should, at least theoretically, have no systematic variation. (2)

If the spatial autocorrelation coefficient is statistically significant, then SAR will quantify the presence of spatial autocorrelation. It will indicate the extent to which variations in the dependent variable (y) are explained by the average of neighboring observation values. (3) Finally, the model will have a better fit, i.e., a higher R-squared statistic.

As in the case of classical regression, the SAR equation has to be transformed via the logistic function for binary dependent variables. The estimates of ρ and β can be derived using maximum likelihood theory or Bayesian statistics. We have carried out preliminary experiments using the spatial econometrics MATLAB package for making the MATLAB toolbox, which implements a Bayesian approach using sampling-based Markov Chain Monte Carlo (MCMC) methods [LeS97]. The general approach of MCMC methods is that when the joint-probability distribution is too complicated to be computed analytically, then a sufficiently large number of samples from the conditional probability distributions can be used to estimate the *statistics* of the full joint probability distribution. While this approach is very flexible and the workhorse of Bayesian statistics, it is a computationally expensive process with slow convergence properties. Furthermore, at least for non-statisticians, it is a non-trivial task to decide what "priors" to choose and what analytic expressions to use for the conditional probability distributions.

Computing Markov Random Fields with Graph Partitioning Technique. Markov Random Fields(MRFs) generalize Markov Chains to multi-dimensional structures. Since there is no natural order in a multi-dimensional space, the notion of a transition probability matrix is absent in MRFs.

MRFs have found applications in image processing and spatial statistics, where they have been used to estimate spatially varying quantities like intensity and texture for noisy measurements. Typical images are characterized by piece-wise smooth quantities, i.e, they vary smoothly but have sharp jumps(discontinuities) at the boundaries of the homogeneous areas. Because of these discontinuities, the least-squares approach does not provide an adequate framework for the estimation of these quantities. MRFs provide a mathematical framework to model our *priori* belief that spatial quantities consist of smooth patches with occasional jumps.

We will follow the approach suggested in [BVZ99] where it is shown that the maximum a posteriori estimate of a particular configuration of an MRF can be obtained by solving a suitable min-cut multi-way graph partitioning problem.

Example 1: A classification problem with no spatial constraints

Even though MRFs are inherently multi-dimensional, we will use a simple one-dimensional example to illustrate the main points. Consider the graph $G = (V, E)$ shown in Figure 1.5(a). The node-set V itself consists of two disjoint sets X and L. The members of X are $\{x_1, x_2, x_3\}$ and the members of L are $\{l_1, l_2\}$. Typically the $x_i's$ are the pixels and the $l_j's$ are the labels, like *nest* or *no-nest*. There is an edge between each member of the set X and L. Here we will interpret the edge weights as probabilities. For example, $p_1 = Prob(x_1 = l_1) = 0.7$ and $p_2 = Prob(x_1 = l_2) = 0.3$; $p_1 + p_2 = 1$.

Our goal is to provide a *labeling* for the pixel set X. This will be done by partitioning the graph into two disjoint sets(not X and L) by removing certain edges such that:

1. There is a many-to-one mapping from the set X to L. Every element of X must be mapped to one and only element of L.

2. The elements of L cannot belong to the same set. Thus there are no edges between elements of L and therefore the number of partitions is equal to the cardinality of L, and

3. The sum of the weights of the edges removed(the cut-set) is the minimum of all possible cut-sets.

In this example the cut-set is easily determined. For example, of the two edges connecting each element of X and an element of L, remove the edge with the *smaller* weight. Figure 1.5(b) shows the graph with the cut-set removed. Thus we have just shown that when the weights of the edges are interpreted as probabilities, the min-cut graph partition induces a Maximum Apriori Estimate(MAP) estimate for the pixel labels. We prefer to say that the *min-cut induces a Bayesian classification* on the underlying pixel set. This is because we will use the Bayes theorem to calculate the edge weights of the graphs.

Example 2: Adding spatial constraints

In the previous example we did not use any information about the spatial proximity of the pixels relative to each other. We do that now by introducing additional edges in the graph structure.

Consider the graph shown in Figure 1.5(c) in which we have added two extra edges (x_1, x_2) and (x_2, x_3) with a weight λ. In this example we have chosen $\lambda = 0.2$.

Now if we want to retain the same partitions of the graph as in Example 1, then the cut-set has two extra edges, namely (x_1, x_2) and (x_2, x_3).

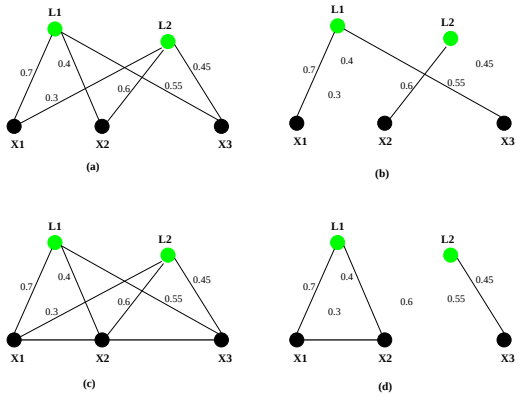


Figure 1.5 (a) Initially each pixel is assigned to both labels with different edge weights. The edge weights correspond to probabilities about assigning each pixel to a different label, (b) A min-cut graph partitioning induces a labeling of the pixel set. Labels which correspond to the maximum probabilities are retained, (c) `Spatial autocorrelation` is modeled by introducing edges between pixel nodes, (d) A min-cut graph partitioning does not necessarily induce a labeling where the labeling with maximum probabilities are retained. If two neighboring pixels are assigned different labels, then the edge connecting the pixels is added to the cut-set.

Thus the sum of the weights of the edges in the cut-set $WC1$ is

$$WC1 = 0.3 + 0.4 + 0.45 + 2\lambda$$

But now, depending upon λ, the cut-set weight may not be minimal. For example, if $\lambda = 0.2$ then the weight of the cut-set $WC2$ consisting of the edges $\{(x_1, l_2), (x_2, l_1), (x_3, l_1), (x_1, x_2)\}$ is

$$WC2 = 0.3 + 0.4 + 0.55 + 0.2$$

Thus $WC2 < WC1$. What is happening is that if two neighboring pixels are assigned to different labels, then the edge between the two neighbors is added to the cut-set. Thus there is a penalty associated with two neighboring nodes being assigned every time to different labels. Thus we can model `spatial autocorrelation` by adding edges between the pixel nodes of the graph. We can also model `spatial heterogeneity` by assigning different *weights*, the λ's to the pixel edges.

3. SPATIAL CO-LOCATION RULES

Association rule finding [HGN00] is an important data mining technique which has helped retailers interested in finding items frequently bought together to make store arrangements, plan catalogs, and promote products together. In market basket data, a transaction consists

of a collection of item types purchased together by a customer. Association rule mining algorithms [AS94, AS94] assume that a finite set of disjoint transactions are given as input to the algorithms. Algorithms like *apriori* [AS94] can efficiently find the frequent itemsets from all the transactions and association rules can be found from these frequent itemsets. Many spatial datasets consist of instances of a collection of boolean spatial features (e.g. drought, needle leaf vegetation). While boolean spatial features can be thought of as item types, there may not be an explicit finite set of transactions due to the continuity of underlying spaces. We define co-location rules, a generalization of association rules to spatial datasets, in this section.

3.1 ILLUSTRATIVE APPLICATION DOMAINS

Many ecological datasets [LCM+97, NVA+99] consist of raster maps of the Earth at different times. Measurement values for a number of variables (e.g., temperature, pressure, and precipitation) are collected for different locations on Earth. Maps of these variables are available for different time periods ranging from twenty years to one hundred years. Some variables are measured using sensors while others are computed using model predictions.

A set of events, i.e., boolean spatial features, are defined on these spatial variables. Example events include drought, flood, fire, and smoke. Ecologists are interested in a variety of spatio-temporal patterns including co-location rules. Co-location patterns represent frequent co-occurrences of a subset of boolean spatial features. Examples of interesting co-location patterns in ecology are shown in Table 1.1.

The spatial patterns of ecosystem data sets include:

a. **Local co-location patterns** represent relationships among events in the same grid cell, ignoring the temporal aspects of the data. Examples from the ecosystem domain include patterns P1 and P2 of Table 1.1. These patterns can be discovered using algorithms [AS94] for mining classical association rules.

b. **Spatial co-location patterns** represent relationships among events happening in different and possibly nearby grid cells. Examples from the ecosystem domain include patterns P3 and P4 of Table 1.1.

Additional varieties of co-location patterns may exist. Furthermore, temporal natures of the ecosystem data give rise to many other time related patterns. We focus on the above co-location patterns in the following sections.

Table 1.1 Examples of interesting spatio-temporal ecological patterns. Net Primary Production (NPP) is a key variable for understanding the global carbon cycle and the ecological dynamics of the Earth

Pattern #	Variable A	variable B	Examples of interesting patterns
P1	Cropland Area	Vegetation	Higher cropland area alters NPP
P2	Precipitation Drought Index	Vegetation	Low rainfall events lead to lower NPP
P3	Smoke Aerosol Index	Precipitation	Smoke aerosols alter the likelihood of rainfall in a nearby region
P4	Sea Surface Temperature	Land Surface Climate and NPP	Surface ocean heating affects regional terrestrial climate and NPP

3.2 CO-LOCATION RULE APPROACHES

Given the difficulty in creating explicit disjoint transactions from continuous spatial data, this section defines several approaches to module co-location rules. We will use Figure 1.6 as an example spatial dataset to illustrate different models. In Figure 1.6, a uniform grid is imposed on the underlying spatial framework. For each grid l, its neighbors are defined to be the nine adjacent grids (including l). Spatial feature types are labeled beside their instances. We define following basic concepts to facilitate the description of different models.

Definition 1 *A* **co-location** *is a subset of boolean spatial features or spatial events.*

Definition 2 *A* **co-location rule** *is of the form:* $C_1 \rightarrow C_2(p, cp)$ *where* C_1 *and* C_2 *are co-locations,* p *is a number representing prevalence measure and* cp *is a number measuring conditional probability.*

The prevalence measure and the conditional probability measure are called interest measures and defined differently in different models which will be described shortly.

The **reference feature centric model** is relevant to application domains focusing on a specific boolean spatial feature, e.g. cancer. Domain scientists are interested in finding the co-locations of other task relevant

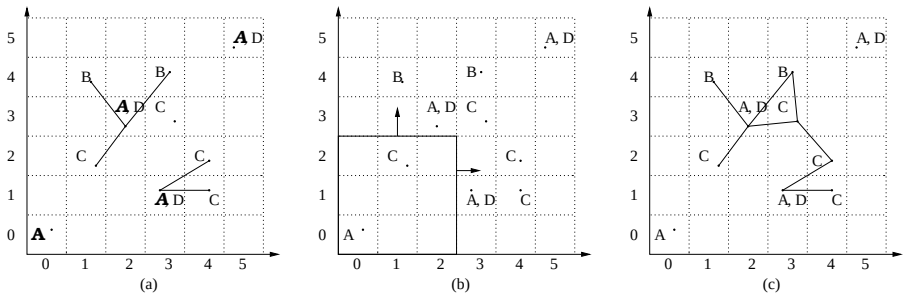


Figure 1.6 Spatial dataset to illustrate different co-location models. Spatial feature types are labeled besides their instances. The 9 adjacent grids of a grid l (including l) are defined to be l's neighbors. a) Reference feature centric model. The instances of A are connected with their neighboring instances of B and C by edges. b) Window centric model. Each 3 X 3 window corresponds to a transaction. c) Event centric model. Neighboring instances are joined by edges.

features (e.g. asbestos, other substances) to the reference feature. This model enumerates neighborhoods to "materialize" a set of transactions around instances of the reference spatial feature. A specific example is provided by the spatial association rule [KH95].

For example, in Figure 1.6 a), let the reference feature be A, the set of task relevant features be B and C, and the set of spatial predicates include one predicate named "*close_to*". Let us define $close_to(a, b)$ to be true if and only if b is a's neighbor. Then for each instance of spatial feature A, a transaction which is a subset of relevant features $\{B, C\}$ is defined. For example, for the instance of A at (2,3), transaction $\{B, C\}$ is defined because the instance of B at (1,4) (and at (3,4)) and instance of C at (1,2) (and at (3,3)) are *close_to* (2,3). The transactions defined around instances of feature A are summarized in Table 1.2.

Table 1.2 Reference feature centric view: transactions are defined around instances of feature A relevant to B and C in figure 1.6 a)

Instance of A	Transaction
(0,0)	ø
(2,3)	$\{B, C\}$
(3,1)	$\{C\}$
(5,5)	ø

With "materialized" transactions, the support and confidence of the traditional association rule problem [AS94] may be used as prevalence and conditional probability measures as summarized in Table 1.3. Since 1 out of 2 non-empty transactions contains instances of both B and C and 1 out of 2 non-empty transactions contain C in Table 1.2, an association rule example is: $is_type(i, A) \wedge \exists j\ is_type(j, B) \wedge close_to(j, i) \rightarrow \exists k\ is_type(k, C) \wedge close_to(k, i)$ with $\frac{1}{1} * 100\% = 100\%$ probability.

The **window centric model** is relevant to applications like mining, surveying and geology, which focus on land-parcels. A goal is to predict sets of spatial features likely to be discovered in a land parcel given that some other features have been found there. The window centric model enumerates all possible windows as transactions. In a space discretized by a uniform grid, windows of size kXk can be enumerated and materialized, ignoring the boundary effect. Each transaction contains a subset of spatial features of which at least one instance occurs in the corresponding window. The support and confidence of the traditional association rule problem may again be used as prevalence and conditional probability measures as summarized in Table 1.3. There are 16 3X3 windows corresponding to 16 transactions in Figure 1.6 b). All of them contain A and 15 of them contain both A and B. An example of an association rule of this model is: *an instance of type A in a window* $\rightarrow$ *an instance of type B in this window* with $\frac{15}{16} = 93.75\%$ probability. A special case of the window centric model relates to the case when windows are spatially disjoint and form a partition of space. This case is relevant when analyzing spatial datasets related to the units of political or administrative boundaries (e.g. country, state, zip-code). In some sense this is a local model since we treat each arbitrary partition as a transaction to derive co-location patterns without considering any patterns cross partition boundaries. The window centric model "materializes" transactions in a different way from the reference feature centric model.

The **event centric model** is relevant to applications like ecology where there are many types of boolean spatial features. Ecologists are interested in finding subsets of spatial features likely to occur in a neighborhood around instances of given subsets of event types. For example, let us determine the probability of finding at least one instance of feature type B in the neighborhood of an instance of feature type A in Figure 1.6 c). There are four instances of type A and two of them have some instance(s) of type B in their 9-neighbor adjacent neighborhoods. The conditional probability for the co-location rule is: *spatial feature A at location l* $\rightarrow$ *spatial feature type B in* $9-$ *neighbor neighborhood is* 50%.

Neighborhood is an important concept in the event centric model. Given a reflexive and symmetric neighbor relation R, we can define neighborhoods of a location l which satisfies the definition of neighborhood in Topology [Wor95] as follows:

Definition 3 *A* **neighborhood** *of l is a set of locations $L = \{l_1, \ldots, l_k\}$ such that l_i is a neighbor of l i.e. $(l, l_i) \in R(\forall i \in 1 \ldots k)$.*

We generalize the neighborhood definition to a collection of locations.

Definition 4 *For a subset of locations L' if L' is a neighborhood of every location in $L = \{l_1, \ldots, l_k\}$ then L' is a* **neighborhood** *of L.*

In another word, if every l_1 in L' is a neighbor of every l_2 in L, then L' is a neighborhood of L.

The definition of neighbor relation R is an input and is based on the semantics of application domains. It may be defined using topological relationships (e.g. connected, adjacent), metric relationships (e.g. Euclidean distance) or a combination (e.g. shortest-path distance in a graph such as roadmap). In general there are infinite neighborhoods over continuous space and it may not be possible to materialize all of them. But we are only interested in the locations where instances of spatial feature types (events) occurs. Even confined to these locations, enumerating all the neighborhoods incurs substantial computational cost because support based pruning cannot be carried out before the enumeration of all the neighborhoods is completed and the total number of neighborhoods is obtained. Furthermore, this support-based prevalence measure definition may not be meaningful because the value of the prevalences may be extremely small due to the fact that many neighborhoods are contained in bigger neighborhoods and counted multiple times. Thus the participation index is proposed to be a prevalence measure as defined below.

Definition 5 *For a co-location $C = \{f_1, \ldots, f_k\}$ and a set of locations $I = \{i_1, \ldots, i_k\}$ where i_j is an instance of feature $f_j (\forall j \in 1, \ldots, k)$ if I is a neighborhood of I itself then I is an* **instance** *of C.*

In other words, if elements of I are neighbors to each other, then I is an instance of C. For example, {(3,1),(4,1)} is an instance of co-location $\{A, C\}$ in Figure 1.6 c) using a 9-neighbor adjacent neighbor definition.

Definition 6 *The* **participation ratio** *$pr(C, f_i)$ for feature type f_i of a co-location $C = \{f_1, f_2, \ldots, f_k\}$ is the fraction of instances of f_i which participate in the co-location C. It can be formally defined as $\frac{|distinct(\pi_{f_i}(all\ instances\ of\ co-location\ C))|}{|instances\ of\ \{f_i\}|}$ where π is a projection operation.*

For example, in Figure 1.6 c), instances of co-location $\{A, B\}$ are {(2,3), (1,4)} and {(2,3)), (3,4)}. Only one instance (2,3) of spatial feature A out of four participates in co-location $\{A, B\}$. So $pr(\{A, B\}, A) = \frac{1}{4} = .25$.

Definition 7 *The* **participation index** *of a co-location* $C = \{f_1, f_2, \ldots, f_k\}$ *is* $\prod_{i=1}^{k} pr(C, f_i)$.

In Figure 1.6 c), participation ratio pr($\{A, B\}, A$) of feature A in co-location $\{A, B\}$ is .25 as calculated above. Similarly $pr(\{A, B\}, B)$ is 1.0. The participation index for co-location $\{A, B\}$ is .25 X 1.0 = .25.

The conditional probability of a co-location rule $C_1 \rightarrow C_2$ in the event centric model is the probability of finding C_2 in a neighborhood of C_1 or it can be formally defined as:

Definition 8 *The conditional probability of a co-location rule* $C_1 \rightarrow C_2$ *is* $\frac{|distinct(\pi_{C_1}(all\ instances\ of\ co-location\ C_1 \cup C_2))|}{|instances\ of\ C_1|}$ *where* π *is a projection operation.*

For details of algorithms which mine co-location rules in event centric model refer to [SH01].

4. SPATIAL OUTLIERS

Outliers have been informally defined as observations which appear to be inconsistent with the remainder of that set of data [BL94], or which deviate so much from other observations as to arouse suspicions that they were generated by a different mechanism [Haw80]. Spatial outliers are observations that are inconsistent with those in their neighbhorhood, even though they may not be inconsistent with the overall population. The identification of spatial outliers can lead to the discovery of unexpected knowledge and has a number of practical applications in areas related to transportation, epidemiology, precision agriculture, natural resource management, voting irregularity, and weather prediction.

Many outlier detection algorithms have been recently proposed; however, spatial outlier detection remains challenging for some reasons. First, the choice of a neighborhood is non-trivial. Second, the design of statistical tests for the spatial outliers needs to account for the distribution of the attribute values at various locations as well as the distribution of aggregate function of attribute values over the neighborhoods. In addition, the computational cost of determining parameters for a neighborhood-based test can be high due to the possibility of join computations

Table 1.3 Interest measures for different models

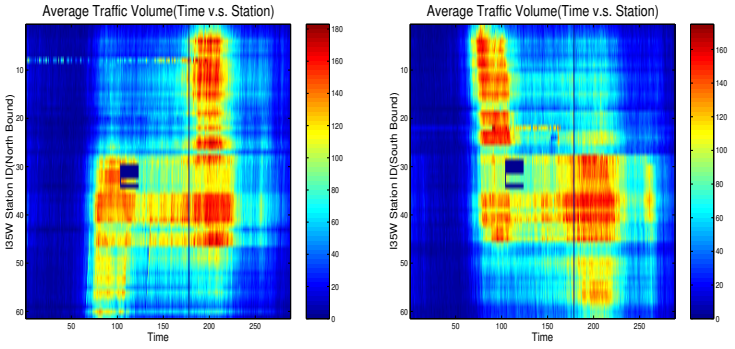
Model	Items	transactions defined by	Interest measures for $C_1 \rightarrow C_2$	
			Prevalence	Conditional probability
local	boolean feature types	partitions of space	fraction of partitions with $C_1 \cup C_2$	$Pr(C_2$ in a partition given C_1 in the partition)
reference feature centric	predicates on reference and relevant features	instances of reference feature C_1 and C_2 involved with	fraction of instance of reference feature with $C_1 \cup C_2$	$Pr(C_2$ is true for an instance of reference features given C_1 is true for that instance of reference feature)
window centric	boolean feature types	possibly infinite set of distinct overlapping windows	fraction of windows with $C_1 \cup C_2$	$Pr(C_2$ in a window given C_1 in that window)
event centric	boolean feature types	neighborhoods of instances of feature types	participation index of $C_1 \cup C_2$	$Pr(C_2$ in a neighborhood of $C_1)$

4.1 AN ILLUSTRATIVE APPLICATION DOMAIN: TRAFFIC DATA SET

In 1995, the University of Minnesota and the Traffic Management Center(TMC) Freeway Operations group started the development of a database to archive sensor network measurements from the freeway system in the Twin Cities. The sensor network includes about nine hundred stations, each of which contains one to four loop detectors, depending on the number of lanes. Sensors embedded in the freeways and interstate monitor the occupancy and volume of traffic on the road. At regular intervals, this information is sent to the Traffic Management Center for operational purposes, e.g., ramp meter control, as well as research on traffic modeling and experiments.

In this application, we are interested in discovering the location of stations whose measurements are inconsistent with those of their spatial neighbors and the time periods when those abnormalities arise. The outlier detection tasks are: a) Build a statistical model for a spatial dataset; b) Check whether a specific station is an outlier; c) Check whether stations on a route are outliers.

Figure 1.7 shows an example of traffic flow outliers. Figure 1.7(a) and (b) are the traffic volume maps for I-35W North Bound and South Bound, respectively, on January 21 1997. The X-axis is the 5-minute time slot for the whole day and the Y-axis is the label of the stations installed on the highway, starting from 1 in the north end to 61 in the south end. The abnormal dark line at time slot 177 and the dark rectangle during time slot 100 to 120 on the X-axis and between station 29 to 34 on the Y-axis can be easily observed from both (a) and (b). This dark line at time slot 177 is an instance of temporal outliers, where the dark rectangle is a spatial-temporal outlier. Moreover, station 9 in Figure 1.7(a) exhibits inconsistent traffic flow compared with its neighboring stations, and was marked as a spatial outlier.



(a) I-35W North Bound (b) I-35W South Bound

Figure 1.7 An example of an outlier

4.2 SPATIAL OUTLIER DETECTION APPROACHES

Outliers in a spatial data set can be classified into three categories, namely, set-based outliers, multi-dimensional space-based outliers, and graph-based outliers. A set-based outlier is a data object whose attributes are inconsistent with attribute values of other objects in a given

data set regardless of spatial relationships. Both multi-dimensional space-based outliers and graph-based outliers are called spatial outliers, that is, data objects that are significantly different in the attribute space from the collection of data objects among spatial neighborhoods. However, multi-dimensional space-based outliers and graph-based outliers are based on different spatial neighborhood definitions. In multi-dimensional space-based outlier detection, the definition of spatial neighborhood is based on Euclidean distance, while in graph-based spatial outlier detection, the definition is based on graph connectivity.

Many outlier detection algorithms [ABKS99, BL94, BKNS99, KN97, KN98, PS88, RR96, YSZ99] have been recently proposed, as shown in Figure 1.8. The set-based outlier detection algorithms [BL94, Joh92] consider the statistical distribution of attribute values, ignoring the spatial relationships among data objects. Numerous distribution-based outlier detection tests, known as discordancy tests [BL94, Joh92], have been developed for different circumstances, depending on the data distribution, the number of expected outliers, and the types of expected outliers. The main idea is to fit the data set to a known standard distribution, and develop a test based on distribution properties.

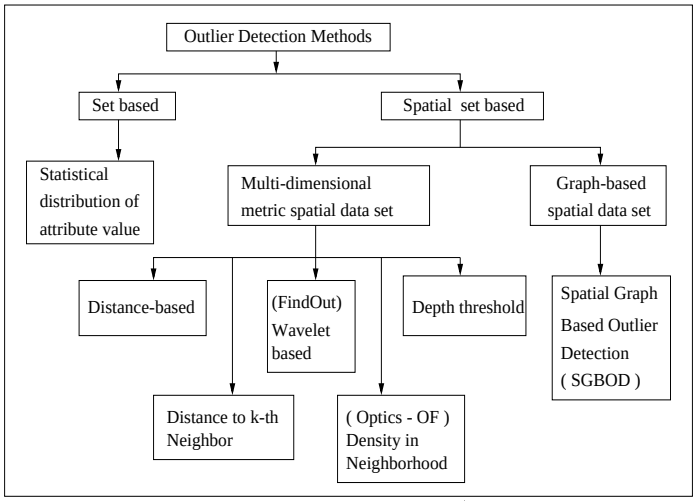


Figure 1.8 Classification of outlier detection methods

The multi-dimensional space-based methods model data sets as a collection of points in a multidimensional space, and provide tests based on concepts such as distance, density, and convex-hull depth. Knorr

and Ng presented the notion of distance-based outliers [KN97, KN98]. For a k dimensional data set T with N objects, an object O in T is a $DB(p, D)$-outlier if at least a fraction p of the objects in T lies greater than distance D from O. Ramaswamy et al. [RRS] proposed a formulation for distance-based outliers based on the distance of a point from its k^{th} nearest neighbor. After ranking points by the distance of each point to its k^{th} nearest neighbor, the top n points are declared as outliers. Breunig et al. [BKNS99] introduced the notion of a "local" outlier where the outlier-degree of an object is determined by taking into account the clustering structure in a bounded neighborhood of the object, e.g., k nearest neighbors. They formally defined the *outlier factor* to capture this relative degree of isolation or outlierness. Their notions of outliers are based on the same theoretical foundation as density-based cluster analysis [ABKS99]. In computational geometry, some depth-based approaches [RR96, PS88] organize data objects in convex hull layers in data space according to peeling depth [PS88], and outliers are expected to be found from data objects with shallow depth value. Conceptually, depth-based outlier detection methods are capable of processing multi-dimensional datasets. However, with the best case computational complexity $\Omega(N^{\lceil k/2 \rceil})$ for computing a convex hull, where N is the number of objects and k is the dimensionality of the dataset, depth-based outlier detection methods may not be applicable for high dimensional data sets. Yu et al. [YSZ99] introduced an outlier detection approach, called *FindOut*, which identifies outliers by removing clusters from the original data. Its key idea is to apply signal processing techniques to transform the space and find the dense regions in the transformed space. The remaining objects in the non-dense regions are labeled as outliers.

Multi-dimensional Euclidean spatial based methods detect outliers in multidimensional data space. These approaches have some limitations. First, the multi-dimensional approaches assume that the data items are embedded in isometric metric space and do not capture the spatial graph structure. Consider the application domain of traffic data analysis. A multi-dimensional method may put a detector station in the neighborhood of another detector even if the detectors were on opposite sides of the highway (e.g., I-35W north bound at exit 230, and I-35W south bound at exit 230), leading to the potentially incorrect identification of a bad detector. Secondly, these methods do not exploit apriori information about the statistical distribution of attribute data. Finally, they seldom provide the confidence measure of the discovered outliers.

In this following subsection, we describe a general framework for detecting spatial outliers in a spatial data set with an underlying graph structure. The detailed work can be found in [SLZ01].

Choice of Spatial Statistic. For spatial statistics, several parameters should be pre-determined before running the spatial outlier test. First, the neighborhood must be selected, based on a fixed cardinality or a fixed graph distance or a fixed Euclidean distance. Second, the aggregate neighborhood function must be chosen, e.g., mean, variance, and auto-correlation. The third parameter that should be pre-determined is the values used for comparing a location with its neighbors, either just one attribute or a vector of attribute values. Finally, the statistic must be chosen.

The statistic used in the graph-based outlier detection method is $S(x) = [f(x) - E_{y \in N(x)}(f(y))]$, where $f(x)$ is the attribute value for a data record x, $N(x)$ is the fixed cardinality set of neighbors of x, and $E_{y \in N(x)}(f(y))$ is the average attribute value for neighbors of x. Statistic $S(x)$ denotes the difference of attribute value of each data object x and the average attribute value of $x's$ neighbors.

Characterizing the Distribution of the Statistic.

Lemma 1 *Spatial Statistic* $S(x) = [f(x) - E_{y \in N(x)}(f(y))]$ *is normally distributed if attribute value* $f(x)$ *is normally distributed.*

Proof:

Given the definition of neighborhood, for each data record x, the average attribute values $E_{y \in N(x)}(f(y))$ of $x's$ k neighbors can be calculated. Since attribute values $f(x)$ are normally distributed and an average of normal variables is also normally distributed, the average attribute values $E_{y \in N(x)}(f(y))$ over neighborers is also a normal distribution for a fixed k cardinality neighborhood.

Since the attribute value and the average attribute value over neighbors are two normal variables, the distribution of difference $S(x)$ of each data object x and the average attribute value of $x's$ neighbors is also normally distributed. ■

Test for Outlier Detection. The test for detecting an outlier in a graph structure can be described as follows. $|\frac{S(x)-\mu_s}{\sigma_s}| > \theta$. For each data object x with an attribute value $f(x)$, the $S(x)$ is the difference of the attribute value of data object x and the average attribute value of its neighbors. μ_s is the mean value of all $S(x)$, and σ_s is the standard deviation of all $S(x)$. The choice of θ depends on the specified confidence interval. For example, a confidence interval of 95 percent will lead to $\theta \approx 2$.

4.3 COMPUTATION OF TEST PARAMETERS

We now describe the Test Parameters Computation(TPC) algorithm to calculate the test parameters, e.g., mean and standard deviation for the statistics. The computed mean and standard deviation can then be used to validate the outlier of the incoming data set.

Given an attribute data set V and the connectivity graph $G = (V, E)$, the TPC algorithm first retrieves the neighbor nodes from G for each data object x, then it computes the difference of the attribute value of x to the average of the attribute values of $x's$ neighbor nodes. These different values are then stored as a set called the AvgDist_Set. Finally, the AvgDist_Set is used to get the distribution value μ_s and σ_s.

4.4 COMPUTATION OF TEST RESULTS

The neighborhood aggregate statistics value, e.g., mean and standard deviation, computed in the TPC algorithm can be used to verify the outlier of an incoming data set. The two verification procedures are Route Outlier Detection(ROD) and Random Node Verification(RNV). The ROD procedure detects the spatial outliers from a user specified route with the graph, e.g., all stations along a highway. The RNV procedure check the outlierness from a set of randomly generated nodes. The steps to detect outliers in both ROD and RNV are similar

Given a route RN in the data set V with graph structure $G = (V, E)$, the ROD algorithm first retrieves the neighboring nodes from G for each data object x in the route RN, then it computes the difference $S(x)$ between the attribute value of x and the average of attribute values of $x's$ neighboring nodes. Each $S(x)$ can then be tested using the spatial outlier detection test $|\frac{S(x)-\mu_s}{\sigma_s}| > \theta$. The θ is pre-determined by the given confidence interval. The data objects with a statistic greater than θ are then declared as outliers.

5. CONCLUSION

In this chapter, we provide a new viewpoint for understanding the spatial data mining literature. Two major approaches are identified: the classical data mining method after feature selection and new spatial data mining approaches. We mainly focus on the second approach. We introduced the spatial autoregression model and Markov random fields with graph partitioning to remove trends resulting from spatial dependency in classical location prediction models. A new pattern, co-location rules mining, which is associated with classical association rules mining,

but different due to the intrinsic implicit transactions in spatial data, is introduced along with new prevalence measures and conditional probability definitions. Spatial outlier detection models spatial dependence via neighborhood graphs. Finally, there are other interesting spatial patterns including hotspot analysis, aggregate proximity [KN96, KR96], and boundary shape matching [KNS97]. We briefly describe one of these in following paragraph and plan to explore some of these in future work.

Acknowledgments
We thank Jiawei Han (Simon Fraser University, CA), James Lesage, and Sucharita Gopal(Boston University) for valuable insights during the discussion on spatial data mining at the Scientific Data Mining workshop 2000 held at the Army High Performance Computing Research Center. We also thank Ugyar Ozesmi (Ericyes University, Kayseri, Turkey), and Pusheng Zhang (University of Minnesota) for contributions of dataset and graphs and Xiaobin Ma and Hui Xiong for their valuable feedback on early versions of this paper. We would like to thank Steve Klooster et. al. from NASA for the examples in Table 1.1, and James Lesage (http://www.spatial-econometrics.com/) for making the MATLAB toolbox available on the web.

References

[ABKS99] M. Ankerst, M.M. Breunig, H.P. Kriegel, and J. Sander. Optics: Ordering points to identify the clustering structure. In *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, Philadelphia, Pennsylvania, USA*, pages 49–60, 1999.

[Agr94] R. Agrawal. Tutorial on database mining. In *Thirteenth ACM Symposium on Principles of Databases Systems*, pages 75–76, Minneapolis, MN, 1994.

[AM95] P.S. Albert and L.M. McShane. A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data. *Biometrics(Publisher: Washington, Biometric Society, Etc.), 1:627-638*, 1995.

[Ans88] L Anselin. *Spatial Econometrics: methods and models.* Kluwer, Dordrecht, Netherlands, 1988.

[AS94] R. Agrawal and R. Srikant. Fast algorithms for Mining Association Rules. In *Proc. of Very Large Databases*, may 1994.

[BKNS99] M.M. Breunig, H.P. Kriegel, R. T. Ng, and J. Sander. Optics-of: Identifying local outliers. In *Proc. of PKDD '99, Prague, Czech Republic, Lecture Notes in Computer Science (LNAI 1704), pp. 262-270, Springer Verlag*, 1999.

[BL94] V. Barnett and T. Lewis. *Outliers in Statistical Data*. John Wiley, New York, 3rd edition, 1994.

[BVZ99] Y. Boykov, O. Veksler, and R. Zabih. Fast Approximate Energy Minimization via Graph Cuts . *Proc. of International Conference on Computer Vision*, September 1999.

[Cre93] N.A. Cressie. *Statistics for Spatial Data (Revised Edition)*. Wiley, New York, 1993.

[EKS97] M. Ester, H.-P. Kriegel, and J. Sander. Spatial Data Mining: A Database Approach. In *Proc. Fifth Symposium on Rules in Geographic Information Databases*, 1997.

[Gre00] G. Greenman. Turning a map into a cake layer of information. *New York Times, http://www.nytimes.com/library/tech /00/01/circuits/articles/20giss.html*, Feb 12 2000.

[Gut94] R.H. Guting. An Introduction to Spatial Database Systems. In *Very Large Data Bases Journal(Publisher: Springer Verlag)*, October 1994.

[Hai89] R.J. Haining. Spatial Data Analysis in the Social and Environmental Sciences. In *Cambridge University Press, Cambridge, U.K*, 1989.

[Haw80] D. Hawkins. *Identification of Outliers*. Chapman and Hall, 1980.

[HGL93] M. Hohn, L. Gribki, and A.E. Liebhold. A Geostatistical Model for Forecasting the Spatial Dynamics of Defoliation Caused by the Beypsy MothLymantria dispar (Lepidoptera: Lymantriidae). *Environmental Entomology (Publisher: Entomological Society of America, 22:1066-1075*, 1993.

[HGN00] J. Hipp, U. Guntzer, and G. Nakaeizadeh. Algorithms for Association Rule Mining - A General Survey and Comparison. In *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2000.

[IES89] Issaks, Edward, and M. Svivastava. Applied Geostatistics. In *Oxford University Press, Oxford*, 1989.

[Joh92] R. Johnson. *Applied Multivariate Statistical Analysis*. Prentice Hall, 1992.

[KAH96] K. Koperski, J. Adhikary, and J. Han. Knowledge Discovery in Spatial Databases: Progress and Challenges. In *Proc. ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Montreal, Canada. 55-70*, 1996.

[KH95] K. Koperski and J. Han. Discovery of Spatial Association Rules in Geographic Information Databases. In *Proc. Fourth International Symposium on Large Spatial Databases, Maine. 47-66*, 1995.

[KHS98] K. Koperski, J. Han, and N. Stefanovic. An Efficient Two-Step Method for Classification of Spatial Data. 1998.

[KN96] E.M. Knorr and R.T. Ng. Extraction of Spatial Proximity Patterns by Concept Generalization. In *Proc. Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon. AAAI Press. 347-350*, 1996.

[KN97] E. Knorr and R. Ng. A unified notion of outliers: Properties and computation. In *Proc. of the International Conference on Knowledge Discovery and Data Mining*, pages 219–222, 1997.

[KN98] E. Knorr and R. Ng. Algorithms for mining distance-based outliers in large datasets. In *Proc. 24th VLDB Conference*, 1998.

[KNS97] E.M. Knorr, R.T. Ng, and D.L. Shilvock. Finding boundary shape matching relationships in spatial data. In *Proc. 5th International Symposium, SSD 97. Springer-Verlag, Berlin, 29-46*, 1997.

[KR96] E.M. Knorr and Ng R.T. Finding aggregate proximity relationships and commonalities in spatial data mining. *IEEE Trans. Knowl. and Data Eng. 8(6):884-897*, 1996.

[Kru95] P. Krugman. Development, Geography, and Economic theory. In *MIT Press, Cambridge, MA*, 1995.

[LCM+97] Z. Li, J. Cihlar, L. Moreau, F. Huang, and B. Lee. Monitoring Fire Activities in the Boreal Ecosystem. *Journal Geophys. Res,. 102(29):611-629*, 1997.

[LeS97] J.P. LeSage. Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, (20):113–129, 1997.

[Mar99] D. Mark. Geographical Information Science: Critical Issues in an Emerging Cross-disciplinary Research Domain. In *NSF Workshop*, February 1999.

[NVA+99] D.C. Nepstad, A. Verissimo, A. Alencar, C. Nobre, E. Lima, P. Lefebvre, P. Schlesinger, C. Potter, P. Moutinho, E. Mendoza, M. Cochrane, and V. Brooks. Large-scale Impoverishment of Amazonian Forests by Logging and Fire. *Nature, 398:505-508*, 1999.

[OM97] U. Ozesmi and W. Mitsch. A spatial habitat model for the Marsh-breeding red-winged black-bird(agelaius phoeniceus l.) In coastal lake Erie wetlands. *Ecological Modelling (Publisher: Elsevier Science B. V.)*, (101):139–152, 1997.

[OO99] S. Ozesmi and U. Ozesmi. An Artificial neural network approach to spatial habitat modeling with interspecific interaction. *Ecological Modelling (Publisher: Elsevier Science B. V.)*, (116):15–31, 1999.

[PS88] F. Preparata and M. Shamos. *Computatinal Geometry: An Introduction*. Springer Verlag, 1988.

[Qui86] J.R. Quinlan. Induction of Decision Trees. *Machine Learning*, 1986.

[RR96] I. Ruts and P. Rousseeuw. Computing depth contours of bivariate point clouds. In *Computational Statistics and Data Analysis, 23:153–168*, 1996.

[RRS] S. Ramaswamy, R. Rastongi, and K. Shim. Efficient algorithms for mining outliers from large data sets. In *Bell Laboratories, Murray Hill, NJ.*

[RS99] J.-F. Roddick and M. Spiliopoulou. A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research. *SIGKDD Explorations 1(1): 34-38 (1999)*, 1999.

[SC01] S. Shekhar and S. Chawla. Spatial Databases: Issues, Implementation and Trends. *Prentice Hall (under contract)*, 2001.

[SCR+99] S. Shekhar, S. Chawla, S. Ravada, A. Fetterer, X. Liu, and C.-T. Lu. Spatial Databases - Accomplishments and Research Needs. *Trans. on Knowledge and Data Engineering 11(1): 45-55 (1999)*, 1999.

[SH01] S. Shekhar and Y. Huang. Co-location Rules Mining: A Summary of Results. *Proc. Spatio-temporal Symposium on Databases*, 2001.

[SLZ01] S. Shekhar, C.T. Lu, and P Zhang. Detecting Graph-based Spatial Outliers: Algorithms and Applications . In *Department of Computer Science Techinical Report TR 01-014,*

University of Minnesota: http://tiberius.cs.umn.edu/tech-reports/listing/, 2001.

[SNM+95] P. Stolorz, H. Nakamura, E. Mesrobian, R.R. Muntz, E.C. Shek, J.R. Santos, J. Yi, K. Ng, S.Y. Chien, R. Mechoso, and J.D. Farrara. Fast Spatio-Temporal Data Mining of Large Geophysical Datasets. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining, AAAI Press, 300-305*, 1995.

[SYH93] S. Shekhar, T.A. Yang, and P. Hancock. An Intelligent Vehicle Highway Information Management System. *Intl Jr. on Microcomputers in Civil Engineering (Publisher: Blackwell Publishers), 8(3)*, 1993.

[Tob79] W.R. Tobler. *Cellular Geography, Philosophy in Geography.* Gale and Olsson, Eds., Dordrecht, Reidel, 1979.

[Wor95] M.F. Worboys. *GIS A Computing Perspective.* Taylor and Francis, 1995.

[YL97] Y. Yasui and S.R. Lele. A Regression Method for Spatial Disease Rates: An Estimating Function Approach. *Journal of the American Statistical Association, 94:21-32*, 1997.

[YSZ99] D. Yu, G. Sheikholeslami, and A. Zhang. Findout: Finding outliers in very large datasets. In *Department of Computer Science and Engineering State University of New York at Buffalo, Technical report 99-03, http://www.cse.buffalo.edu/tech-reports/*, 1999.