Spatiotemporal Data Mining

60

Tao Cheng, James Haworth, Berk Anbaroglu,
Garavig Tanaksaranond, and Jiaqiu Wang

Contents

Abstract

As the number, volume and resolution of spatio-temporal datasets increases, traditional statistical methods for dealing with such data are becoming overwhelmed. Nevertheless, the spatio-temporal data are rich sources of

T. Cheng (✉) • J. Haworth • B. Anbaroglu • G. Tanaksaranond • J. Wang
SpaceTimeLab, Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK
e-mail: tao.cheng@ucl.ac.uk; j.haworth@ucl.ac.uk; b.anbaroglu@ucl.ac.uk; g.tanaksaranond@ucl.ac.uk; w.jiaqiu@ucl.ac.uk

information and knowledge, waiting to be discovered. The field of spatio-temporal data mining (STDM) emerged out of a need to create effective and efficient techniques in order to turn the massive data into meaningful information and knowledge. This chapter reviews the state of the art in STDM research and applications, with emphasis placed on three key areas, including spatio-temporal prediction and forecasting, spatio-temporal clustering and spatio-temporal visualization. The future direction and research challenges of STDM are discussed at the end of this chapter.

60.1   Introduction

With automatic sensor networks and crowd sourcing now being used extensively to monitor a diverse range of phenomena, the amount of data being collected with both spatial and temporal dimensions has increased dramatically. Data collected at two or more locations and times make up space-time series, examples of which include daily temperature series at meteorological stations, monthly crime rates of world capital cities and daily traffic flow on urban roads. These space-time series are massive and continually growing. Spatio-temporal data mining (STDM) is the extraction of unknown and implicit knowledge, structures, relationships, or patterns from these massive datasets. STDM techniques and tasks include spatio-temporal forecasting, spatio-temporal association rule mining, spatio-temporal sequential pattern mining and spatio-temporal clustering and classification, amongst others (Miller and Han 2009). More recently, spatio-temporal visualization has become another hot topic for STDM as we begin to explore new ways of representing spatio-temporal data that go beyond the static map.

Early research efforts on spatio-temporal forecasting focused on adapting existing statistical regression models from the fields of time series analysis, spatial analysis and econometrics to deal with spatio-temporal data. Such models are typically geared towards teasing scarce information from homogenous datasets and have been overwhelmed by the increasing volume and diversity of spatio-temporal data that is now being collected. Increasingly, researchers and practitioners are turning towards less conventional techniques, often with their roots in the machine learning and data mining communities, that are better equipped to deal with the heterogeneous, nonlinear and mutli-scale properties of large scale spatio-temporal datasets. For instance, methods such as artificial neural networks (ANNs) and support vector machines (SVMs) are now being successfully applied to spatio-temporal forecasting problems.

The association (or co-location) rule mining is to infer the presence of spatial features in the neighbourhood of other spatial features (Shekhar et al. 2011). They are spatial extensions of association rules, which were developed by the retail industry to examine the behavior of consumers. A spatio-temporal co-location rule implies a strong association between locations A and B that if the attributes of A take some specific value at a point in time, then with a certain probability, at the same point in time, the attributes of B will take some specific value. A related

STDM task is mixed drove co-occurrence pattern (MDCOP) mining. MDCOPs are subsets of two or more different object types whose instances are often located close to one another in space and time (Shekhar et al. 2011). The drawback of these methods is that only contemporaneous associations are considered so they do not account for the evolution of a spatial process over time.

A logical extension to association mining is to analyze spatio-temporal sequential patterns. This involves finding sequences of events (an ordered list of item sets) that occur frequently in spatio-temporal datasets. Sequential pattern mining algorithms were also first introduced to extract patterns from customer transaction databases. A spatio-temporal sequential pattern means that if at some point in time and space, the attributes in A take some specific value, then with a certain probability at some later point in time, attributes at B will take some specific value. Sequential pattern mining implicitly incorporates the notion of spatio-temporal dependence; that the events at one location at one time can have some causal influence on the events at another location at a subsequent time. A similar concept to sequential patterns are cascading spatio-temporal patterns, which are ordered subsets of events that are located close together and occur in a cascading sequence (Shekhar et al. 2011).

Clustering involves grouping unlabeled objects that share similar characteristics. The goal is to maximize the intraclass similarity and minimize the interclass similarity. Clustering can be used for classification, segmentation and outlier detection, and here clustering is a general term for all these tasks. Widely used spatial clustering techniques e.g., K-means and K-medoids, have been extended to spatio-temporal clustering problems. Designing an effective spatio-temporal clustering algorithm is a difficult task because it must account for the dynamics of a phenomenon in space and time. For instance, when clustering moving objects, a cluster may change its spatial location from one time step to the next but still be the same spatio-temporal cluster. Rules for capturing this type of behavior are difficult to encode in algorithms.

Mining interesting patterns, rules and structures from spatio-temporal data is only part of the task of STDM. The results are not useful if they are not easily understood. For instance, finding a spatio-temporal cluster in a patient register dataset is not useful in itself. On the other hand, confirming this spatio-temporal cluster as a disease outbreak and visualizing it using a platform that epidemiologists and medical professionals can understand is very useful indeed. As a result, space-time visualization has emerged as another important facet of STDM. It explores the patterns hidden in the large data sets by using advanced visualization and animation techniques. This includes conventional 2D maps as well as newly developed 3D space-time cube methods, which can show hotspots and isosurfaces of spatio-temporal phenomena. Integration of data exploration, analysis and visualization in a single platform takes this one step further. The STARS platform (space-time analysis of regional systems, Rey and Janikas 2010) is an excellent example of this that allows exploratory and explanatory analysis and visualization of regional data with spatio-temporal extent. However, despite significant progress, how to visualize large volumes of data in real time and to best make use of the third dimension are problems that are yet to be adequately solved.

This chapter is organized around three main tasks of STDM; space-time modeling and prediction, space-time clustering and space-time visualization. In the following section, we review spatio-temporal autocorrelation and its implications for space-time modeling. Section 60.3 is devoted to space-time modeling and prediction, by either statistical (parametric) approaches or machine learning (non-parametric) approaches. Section 60.4 gives a brief review of space-time clustering and outlier detection, and is followed by an introduction to space-time visualization in Sect. 60.5. The final section summarizes the directions of future research in STDM.

60.2   Spatio-Temporal Autocorrelation

An observation from nature is that near things tend to be more similar than distant things both in space and in time. For instance, the weather tomorrow is more likely to be similar to today's weather than the weather a week ago, or a month ago and so on. Similarly the weather 1 mile away is likely to be more similar than the weather 10 miles away or 100 miles away. These phenomena are referred to respectively as temporal and spatial dependence. The presence of dependence in spatial and temporal data violates the stationarity assumption of classic statistical models such as ordinary least squares (OLS) and necessitates the use of specialized modeling and forecasting techniques. Testing for dependence is typically accomplished using an autocorrelation analysis. Autocorrelation is the cross-correlation of a signal with itself and can be measured in temporal data using the temporal autocorrelation function (ACF, Box and Jenkins 1970) or in spatial data using an index such as the familiar Moran coefficient.

These measures are global, implying a degree of fixity in the level of autocorrelation across the space/time such that it can be described by a single parameter. However, this is often unrealistic. Many time series exhibit nonlinear characteristics that make stationarization difficult. Similarly, spatial data often exhibit structural instability over space, which is referred to as heterogeneity. Heterogeneity has two distinct aspects; structural instability as expressed by changing functional forms or varying parameters, and heteroskedasticity that leads to error terms with non-constant variance (Anselin 1988). Ignoring it can have serious consequences including biased parameter estimates, misleading significance levels and poor predictive power. Anselin (1988) provides some methods for testing for heterogeneity. Additionally, a number of local indicators of spatial association (LISA) have been devised. These include a local variant of Moran's I and Getis and Ord's G_i and $G_{\bar{i}}$ statistics, which measure the extent to which high and low values are clustered together.

Although sharing many commonalities in techniques and concepts, the fields of time series analysis and spatial analysis have largely developed separately from one another. The behavior of a variable over space differs from its behavior in time. Time has a clear ordering of past, present and future while space does not and because of this ordering isotropy has no meaning in the space-time context. In time, measurements can only be taken on one side of the axis; hence estimation involves

extrapolation rather than interpolation. Temporal data also has other characteristics, such as periodicity, that are not common in spatial data and scales of measurement also differ between space and time and are not directly comparable.

When a variable Z is observed over time at two or more locations, it is both a spatial series and a time series and can be referred to as a space-time series $z = \{z(s,t)|s \in S, t \in T\}$ in spatial domain S and temporal interval T. A space-time series may exhibit spatio-temporal dependence which describes its evolution over space and time. If the spatio-temporal dependence in a dataset can be modeled then one essentially has predictive information. A number of indices have been devised to this end including space-time (semi) variograms (Heuvelink and Griffith 2010) as well as space-time eigenvector filtering (Griffith 2010). Two indices are described here, the space-time autocorrelation function (ST-ACF), that measures global space-time autocorrelation, and the cross-correlation function (CCF), that measures local space-time autocorrelation between two locations. These indices are extensions of the temporal autocorrelation function and are selected as they are easily interpretable and have a practical application in established space-time modeling frameworks.

60.2.1 The Global Measure

The ST-ACF measures the N^2 cross-covariances between all possible pairs of locations lagged in both time and space (Pfeifer and Deutsch 1980). Given the weighted l^{th} order spatial neighbours of any spatial location at time t and the weighted K^{th} order spatial neighbors of the same spatial location s time lags in the future, the space-time cross-covariance can be given as:

$$\gamma_{lk}(s) = E\left\{\frac{\left[W^{(l)}z(t)\right]'\left[W^{(k)}z(t+s)\right]}{N}\right\} \tag{60.1}$$

Where N is the number of spatial locations, $W^{(l)}$ and $W^{(k)}$ are the $N * N$ spatial weight matrices at spatial orders l and k, $Z(t)$ is the $N * 1$ vector of observations z at time t, $z(t+s)$ is the $N * 1$ vector of observations z at time $(t+s)$ and the symbol $\prime$ denotes matrix transposition. Based on Eq. (60.1), the ST-ACF can be defined as:

$$\rho_{lk}(s) = \frac{\gamma_{lk}(s)}{\left[\gamma_{ll}(0)\gamma_{kk}(0)\right]^{\frac{1}{2}}} \tag{60.2}$$

ST-ACF has been used in STARIMA to calibrate the order of moving average (MA), which define the range of spatial neighbourhoods which contribute to the current location at a specific time lag (Pfeifer and Deutsch 1980). The MA orders are fixed globally both spatially and temporally and a single parameter is estimated for it in practical application such as in Kamarianakis and Prastacos (2005), and Cheng et al. (2011b).

60.2.2 The Local Measure

The cross correlation function (CCF) (see, for example, Box and Jenkins 1970) treats two time series as a bivariate stochastic process and measures the cross covariance coefficients between each series at specified lags. It provides a measure of the similarity between two time series. The CCF is useful if one has reason to believe that the level of autocorrelation in a spatio-temporal dataset is not fixed in time and space. Given two time series X and Y, the CCF at lag k is given as:

$$\rho_{xy}(k) = \frac{E\left[(x_t - \mu_x)(y_{t+k} - \mu_y)\right]}{\sigma_x \sigma_y} \quad k = 0, \pm 1, \pm 2, \pm \cdots \tag{60.3}$$

The CCF measures cross-correlations in both directions, as denoted by subscript k, therefore the temporal lag at which the CCF peaks can be used to determine a transfer function between two series. This is, however, dependent on sufficient spatial and temporal resolution in the data. A peak at lag zero indicates that the current resolution does not capture the direction of influence of one location on another, but the series behave very similarly at the same time (Cheng et al. 2011a). As examples, the global and local measures of road network in central London are shown in Figs. 60.1 and 60.2.

60.3 Space-Time Forecasting and Prediction

Space-time models must account for the combined problems of spatial and temporal data mentioned in the preceding sections. Uptake of space-time models has traditionally been limited by the scarcity of large scale spatio-temporal datasets (Griffith 2010). This is a situation that has been reversed over recent decades and we are now inundated with data and require methods to deal with them quickly and effectively. The models that are currently applied to space-time data can be broadly divided into two categories; statistical (parametric) methods and machine learning (non-parametric) methods. These are described in turn in the following subsections.

60.3.1 Statistical (Parametric) Models

The state of the art in statistical modeling of spatio-temporal processes represents the outcome of several decades of cross-pollination of research between the fields of time series analysis, spatial statistics and econometrics. Some of the methods commonly used in the literature include space-time autoregressive integrated moving average (STARIMA) models (Pfeifer and Deutsch 1980) and variants, multiple ARIMA models, space-time geostatistical models (Heuvelink and Griffith 2010), spatial panel data models (Elhorst 2003), geographically and temporally weighted regression (Huang et al. 2010) and eigenvector spatial filtering (Griffith 2010).
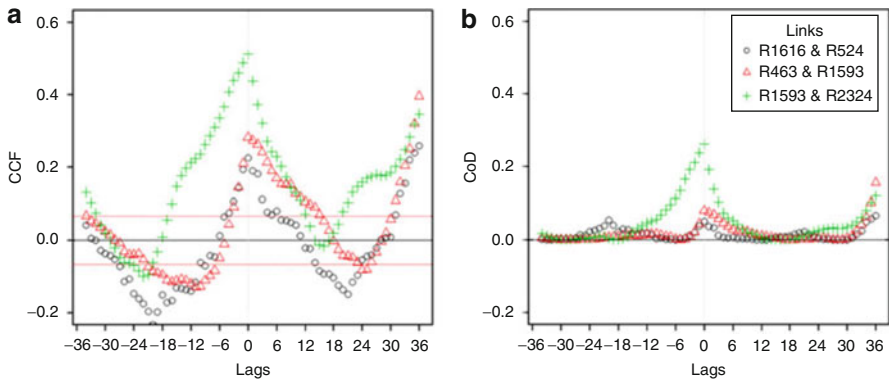
Fig. 60.1 (**a**) CCF and (**b**) coefficient of determination (CCF²) between unit journey times of three pairs of road links in central London in the AM peak period (7–10am) (Cheng et al. 2011a)
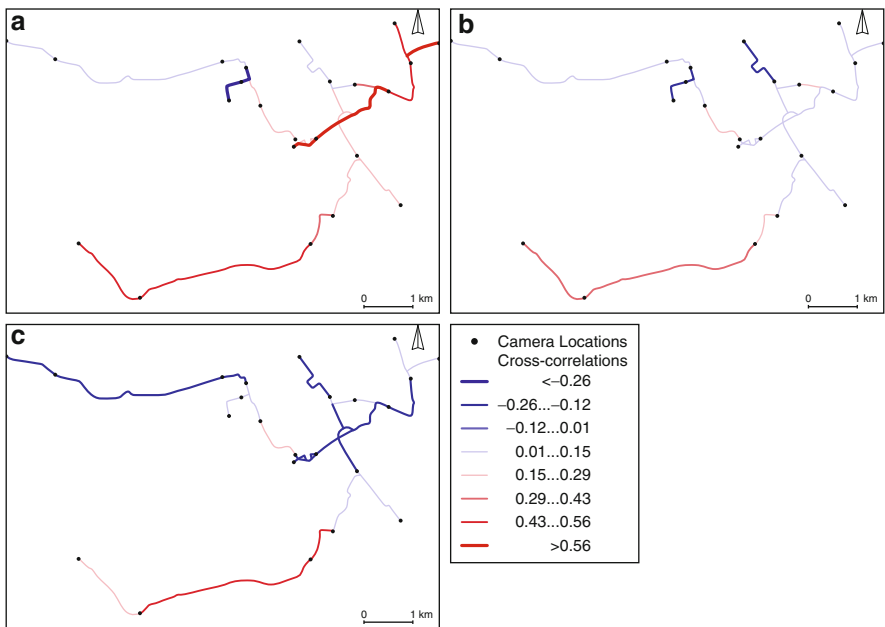

Fig. 60.2 Average CCF between links and their first-order neighbours at temporal lag zero in (**a**) the AM peak; (**b**) interpeak; and (**c**) PM peak (Cheng et al. 2011a)

60.3.1.1 Space-Time Autoregressive Integrated Moving Average

Space-time autoregressive integrated moving average (STARIMA) is a family of models that extend the ARIMA time series model to space-time data (Pfeifer and Deutsch 1980). STARIMA explicitly takes into account the spatial structure in the data through the use of a spatial weight matrix. The general STARIMA model

expresses an observation of a spatial process as a weighted linear combination of past observations and errors lagged in both space and time. A fitted STARIMA model is usually described as a STARIMA (p,d,q) model, where p indicates the autoregressive order, d is the order of differencing and q is the moving average order. The application of STARIMA models has been fairly limited in the literature, with examples existing in traffic prediction (Kamarianakis and Prastacos 2005) and temperature forecasting (Cheng et al. 2011b).

Some important special cases of the STARIMA model should be noted; when $d = 0$ the model reduces to a STARMA model, furthermore, a STARMA model with $q = 0$ is a STAR model and with $p = 0$ is a STMA model. Although the STARIMA model family accounts for spatio-temporal autocorrelation, it has not yet been adequately adapted to deal with spatial heterogeneity and parameter estimates are global. The implication of this is that the space-time process must be stationary (or made stationary through differencing/transformation) for STARIMA modeling to be effective.

60.3.1.2 Spatial Panel Data Models

Panel data is a term used in the econometrics literature for multi-dimensional data. A panel contains observations on multiple phenomena (cross-sections) over multiple time periods. When panel data include a spatial component they are referred to as spatial panel data. Although the term describes the data itself, there are a range of models that have been developed to work with spatial panel data that originate specifically from spatial econometrics that are referred to as spatial panel data models. Methodologically, they are often very similar to those encountered in the spatial statistics literature.

Aspatial panel data models are modified to account for spatial dependence in one of two ways; either with a spatial autoregressive process in the error term; a spatial error model (equivalent to a spatial moving average), or with a spatially autoregressive dependent variable; a spatial lag model (Elhorst 2003). In their standard form, spatial panel data models are global models and do not account for spatial heterogeneity and, as in the spatial statistics literature, this has become a focus of research in recent years. Elhorst (2003) defined a set of spatial panel data models that account for heterogeneity in different ways. The uptake of spatial panel data models has been much more widespread than those mentioned in Sect. 60.3.1 and there have been applications in liquor demand prediction and US state tax competition, amongst many others.

60.3.1.3 Space-Time GWR

Recently, there has been a great deal of interest in extending geographically weighted regression (GWR) to the temporal dimension. In their geographically and temporally weighted regression (GTWR) model, Huang et al. (2010) incorporate both the spatial and temporal dimensions into the weight matrix to account for spatial and temporal nonstationarity. The technique was applied to a case study of residential housing sales in the city of Calgary from 2002 to 2004 and found to outperform GWR and temporally weighted regression (TWR) as well as OLS.

60.3.1.4 Space-Time Geostatistics

Space-time geostatistics is concerned with deriving space-time covariance structures and semivariograms for the purpose of space-time interpolation and forecasting. The aim is to build a process that mimics some patterns of the observed spatiotemporal variability, without necessarily following the underlying governing equations (Kyriakidis and Journel 1999). The first step usually involves separating the deterministic component $m(u, t)$ of space time coordinates u and t. Following this, a covariance structure is fitted to the residuals. The simplest approach is to separate space and time and consider the space-time covariance to be either a sum (zonal anisotropy model) or product (separable model) of separate spatial and temporal covariance functions. Although simple to implement, these models have the disadvantage that they do not consider space-time interaction. They assume a fixed temporal pattern across locations and a fixed spatial pattern across time. Additionally, it is not straightforward to separate the component structures from the experimental covariances. For example, an experimental spatial covariance will be influenced by temporal variability resulting from the time instant at which the data was measured.

The second approach is to model a joint space-time covariance structure. This approach is generally accepted to be more appropriate. Combinations of the two approaches have also been described in the literature (Heuvelink and Griffith 2010). Once an appropriate space-time covariance structure has been defined, one can use standard Kriging techniques for interpolation and prediction; Space-time geostatistical techniques are best applied to stationary space-time processes. Highly nonstationary spatio-temporal relationships require a very complicated space-time covariance structure to be modelled for accurate prediction to be possible. Despite being spatio-temporal in nature, the main function of space-time geostatistical models is space-time interpolation and they encounter problems in forecasting scenarios where extrapolation is required (Heuvelink and Griffith 2010).

60.3.2 Machine Learning (Non-parametric) Approaches

In parallel to the development of statistical space-time models, there was a multidisciplinary explosion of interest in non-parametric machine learning methods, and many of these have been successfully adapted to work with spatio-temporal data due to their innate ability to model complex nonlinear relationships. There is a wide range of machine learning algorithms available, in this section we focus on two of the most popular; the artificial neural network and the support vector machine.

60.3.2.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) are a family of non-parametric methods for function approximation that have been shown to be very powerful tools in many application domains (see Fischer 2006 for example), often dealing with complex real world sensor data. They were initially inspired by the observation that biological learning is governed by a complex set of interconnected neurons. The key

concept is that, although individual neurons may be simple in structure, their interconnections allow them to perform complex tasks such as pattern recognition and classification.

Since its inception, the term ANN has become an umbrella term for a broad class of flexible non-linear models for regression and classification with a range of different architectures. ANNs have been widely applied in spatial and temporal analysis. Kanevski et al. (2009) have applied various types of ANN to spatial and environmental modeling problems including radial basis function neural networks (RBFNN), general regression neural networks (GRNN), probabilistic neural networks (PNN) and neural network residual Kriging (NNRK) models and have gained excellent results. The authors note that the strength of ANNs is that they learn from empirical data and can be used in cases where the modeled phenomena are hidden, non-evident or not very well described. This makes them particularly useful in modeling the complex dependency structures present in space-time data that cannot be described theoretically. Hsieh (2009) also provides a good review of ANN methods applied to spatial problems.

60.3.2.2 Support Vector Machines

Another widely used machine learning technique is the support vector machine (SVM, SVR in the regression case). SVMs are a set of supervised learning methods originally devised for classification tasks that are based on the principles of statistical learning theory (Vapnik 1999). SVMs make use of a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory. The key to their strong performance is that the learning task is formulated as a convex optimization problem meaning that, for a given set of parameters, the solution is globally optimal provided one can be found. Therefore, SVMs avoid the problem of getting stuck in local minima which are traditionally associated with ANNs. This has led to SVMs outperforming most other systems in a wide variety of applications within a few years of their introduction.

SVMs have been successfully used to model time series in a number of application areas including financial time series and traffic flow prediction. Compared to time series analysis, the uptake of SVM in the spatial sciences was initially slow but has seen a rapid increase in popularity in the past 5 years or so. The book "Machine Learning for Spatial Environmental Data" (Kanevski et al. 2009) provides a good introduction to some of the machine learning methods currently being used to model spatial data. Recently, SVMs have been applied to spatio-temporal avalanche forecasting (Pozdnoukhov et al. 2011). The approach involves incorporating the outputs of simple physics based and statistical approaches to interpolate meteorological and snowpack related data over a digital elevation model of the region. The decision boundary is used to discriminate between safe and dangerous conditions.

60.3.2.3 Other Methods

ANNs and SVMs are two methods that are widespread in temporal and spatial analysis, however, the field of machine learning is huge and a comprehensive

review is beyond the scope of this chapter. Various other methods have been used including nearest neighbour regression, kernel (ridge) regression, Gaussian processes, self-organizing maps (SOM), principal components analysis (PCA) and regression trees, which are introduced in Hsieh (2009). This list is non-exhaustive and readers are also referred to the text of Kanevski et al. (2009) for detailed introduction in the context of spatial data.

60.3.3 Summary

In this section, the complex, nonlinear, non-stationary properties of spatio-temporal data and their implications for space-time models were outlined. The question is which model should one choose for a given spatio-temporal dataset? The answer to this depends on the data. In the literature, space-time analysis is typically applied to data with low spatial and/or temporal resolution which is acquired after the event. In the tradition of spatial analysis, the practical use of such data is to elicit causal relationships between variables that can give some valuable insights into the underlying processes. In this case, the use of parametric statistical models may be preferable because of their explanatory power and interpretability.

However, these days, more and more data sources are becoming available in (near) real time at high spatial and temporal resolutions. Extracting meaningful relationships from such data is a task that is secondary to forecasting and it is likely that machine learning approaches, with their greater flexibility, will play an ever increasing role. Generally, machine learning methods have a wider field of application than traditional geostatistics due to their ability to deal with multi-dimensional nonlinear data. They are also well suited to dealing with large databases and long periods of observation. In particular, the SVM approach is favorable because it avoids the curse of dimensionality faced by other methods. One of the future research directions in this area lies in improving the interpretability of the structure and output of machine learning algorithms. Another way is to use a hybrid framework with both statistical and machine learning approaches (Cheng et al. 2011b).

60.4   Space-Time Clustering

60.4.1 Introduction

Another very important task of STDM is to extract meaningful patterns and relationships from massive spatio-temporal data that are not necessarily explicit. In this situation, we may wish to search for structure in the dataset without an apriori hypothesis. Hypotheses can be then be formed and refined aposteriori from the results. This is known as unsupervised learning. One of the most important unsupervised learning tasks in STDM is clustering. This involves grouping space-time series into clusters, where the similarity of data within a cluster and the dissimilarity between the clusters are high. Clustering can also be used to detect outliers.

A spatial outlier is a spatially referenced object whose thematic attribute values are significantly different from those of other spatially referenced objects in its spatial neighborhood. It represents an object that is significantly different from its neighborhoods even though it may not be significantly different from the entire population. A spatial-temporal outlier is a spatial-temporal object whose thematic attribute values are significantly different from those of other spatially and temporally referenced objects in its spatial or/and temporal neighborhoods. Identification of ST-outliers can lead to the discovery of unexpected, interesting, and implicit knowledge, such as local instability or deformation (Cheng and Li 2006). Nowadays spatial and spatio-temporal clustering has been widely used to understand the spatial patterns hidden in spatial databases with applications in epidemic studies, crime hotspot analysis and social networks.

The three domains of space-time series data can be used to define the similarity between observations for clustering. The first is the thematic domain, where the attributes define the characteristics of the object. The second is the spatial domain, which is used to describe the location of the object. Finally, the temporal domain is used to store the timing of the object. These domains are used to answer the questions *what*, *where* and *when* respectively. Initial research on clustering focused on the thematic domain, with methods such as k-means, k-medoids and their variants being popular. Research into clustering using the spatial domain gained popularity in the early twenty-first century. Initial research on spatial clustering has focused on point data. Popular algorithms such as DBSCAN and BIRCH are the outputs of this research area. The spatial distance or the spatial density derived from the spatial locations of the points is considered for clustering. Clustering has also been conducted by combining spatial adjacency with thematic domains or by combining the spatial distance with the thematic distance. Temporal-thematic clustering is mainly applied to group time series data, in order to know whether customers are changing over time, or to determine if credit card fraud transactions change over time.

Very few algorithms consider the spatial, temporal and thematic attributes seamlessly and simultaneously in the clustering. Capturing the dynamicity in the data is the most difficult challenge in spatio-temporal clustering, which is the reason that traditional clustering algorithms, in which the clustering is carried out on a cross-section of the phenomenon, cannot be directly applied to spatio-temporal phenomena. The arbitrarily chosen temporal intervals may not capture the real dynamics of the phenomena since they only consider the thematic values at the same time, which cannot capture the influence of flow (i.e., time-lag phenomena). It is only recently that this has been attempted. We pay particular attention to spatio-temporal scan statistics, a method that has shown promising performance in a range of STDM tasks such as health, crime and transport studies.

60.4.2 Spatio-Temporal Scan Statistics

Spatio-temporal scan statistics (STSS) is a clustering technique that was originally devised to detect disease outbreaks (Neill 2008). The goal is to automatically detect

regions of space that are "anomalous," "unexpected," or otherwise "interesting." Spatial and temporal proximities are exploited by scanning the entire study area via overlapping space-time regions (STRs). Each STR represents a possible disease outbreak with a geometrical shape which is either a cylinder or rectangular prism. The base corresponds to the spatial dimension and the height corresponds to the temporal dimension. The dimensions of the STR are adjustable parameters. For instance, the maximum spatial dimension (e.g., the circular base of a cylindrical STR) can represent the maximum possible boundary of an outbreak, and the height of the STR could be the maximum allowable time to detect the outbreak. The dimensions of the STR are allowed to vary in order to detect outbreaks of varying sizes.

The initial proposition of STSS is based on the comparison of the disease rates inside and outside of a STR. If the disease rate inside the STR is significantly higher than outside the STR, then a possible disease outbreak is detected. However, this does not take into account the temporal variations (e.g., seasonal trends), which are inherent in epidemiological data. More recently, an expectation based approach was proposed to accommodate the temporal trend, where the observed value of an STR is compared with its expected value based upon historical data. Based on the statistical distribution that the data is assumed to follow, comparison is made via a likelihood ratio score function. If a STR has a likelihood ratio score bigger than 1, the STR is a potential disease outbreak. To reduce the false-alarm rate (reporting a disease outbreak where in reality there is no outbreak), the significance of the potential STRs is further tested via Monte Carlo simulation. If the STR is found to be significant at this stage, then a disease outbreak is recorded (Neill 2008)

STSS has the significant drawback that the entire study region has to be scanned, which is computationally intensive and limits the method's scalability. Although previous research has shown that this problem can be tackled via efficient spatial-indexing methods. The assumption that a disease outbreak is a regular geometrical shape is also not realistic (e.g., disease might have spread via the river, thus affecting the people near the river bed) and remains as a limitation of the method. This problem might be tackled by generating irregularly shaped STRs.

60.5   Space-Time Visualization

Representing a phenomenon that evolves over space and time has emerged as a contentious issue within the GIS community. The contentious issue comes from the fact that most geographic phenomena change over time; for example, forest fires, storms, water contamination and also traffic congestion, but representing time on a map is still difficult. It is because GIS has its roots in mapping, which originally was designed to represent static phenomena, not dynamic process. Geographic visualization enhances traditional cartography by providing dynamic and interactive maps. Many new techniques on visualizing time on maps have been proposed. These techniques can be divided into three broad types: (static) 2D and 3D maps, and animation.

60.5.1 2D maps

There are various ways to represent time on static 2D maps, either as a single static map or multiple snap shots. Since all time steps are shown at the same time, the map-readers don't need to retain events temporarily in their minds thus preventing lapses of certain critical information. However, this technique can only present a few time steps at a time due to the limitation of the available map media (computer screen, paper, etc.).

This section will discuss some interesting static map techniques. The techniques are divided according to the type of data to be presented: geometric change of spatial objects (movement, size, shape, etc.), attribute change of spatial objects, and travel time.

60.5.1.1 Representing Geometric Change of Spatial Objects

Monmonier (1990) presents movement of spatial objects by drawing movement paths or pinpoints of objects on a 2D plane. Arrows are added to represent directions of movements. This technique is called a "dance map" since it is similar to a diagram of foot paths in a ballroom dance. Dance maps can display both discrete and continuous movement. When data are captured at fixed time intervals, a dance map can display the rate of movement (or rate of change) very well. Color or variety in sizes of objects can be added to the map, but the number of objects is limited by occlusion.

Another visualization technique presented by Monmonier (1990) is the *chess map* (map series). Each map contains a snap shot representing a time slice. A series of maps are laid out continuously in the manner of a chess board for users to compare events between time slices, allowing the comparison of many different time slices at a single sitting. The disadvantage of chess maps is that a large space is required to present multiple maps at the same time. In addition, the users must determine by themselves as to how the changes occurred, and at which time slices.

60.5.1.2 Representing Thematic Attribute Change of Spatial Objects

A *change map* shows changes or differences against a reference time period, as an absolute value or percentage, such as population increase every 10 years compared with 1990 (Monmonier 1990). The change map is good for representing quantitative attributes. Readers do not have to calculate the amount of change by themselves.

Another way to show the change is to add "small charts on maps" to visualize time series data on maps. The advantage of small charts on maps is that map readers are informed of the locations of the data on the maps as well as how their attributes change over time. However, when plotting many charts simultaneously, the base map can become overcrowded. Moreover, the charts can be easily overlapped when the data locations are very close to one another. An example of small charts on maps is given in Andrienko and Andrienko (2007).

Visual variables (colors, sizes, texture) can be applied to represent variation in attributes at different locations. The classic example of this technique is Minard's map showing Napoleon's doomed campaign to Moscow in 1869. Time was displayed as an axis on the map (parallel to the axis of the geographical position), and the number of remaining soldiers was shown by the thickness of the lines. Another good example is spatial treemaps (Slingsby et al. 2010) that represent traffic variables (traffic speed and traffic volume) of areas of London. Each grid cell on the map represents a borough. The level of brightness of each cell on spatial treemaps is used to represent the value of a traffic variable (speed and volume). Time is also mapped onto small cells within each area. This technique allows the visualization of a large number of time points, since it exploits every pixel on the map to represent data.

Rank Clock has recently been used to visualize the dynamics of city size changes (http://www.bartlett.ucl.ac.uk/casa/pdf/paper152.pdf), where the time is arranged as a clock, the thematic attribute (the size of the city) shows as a dot along the time line. By linking all the dots of a spatial unit over time, the trajectory of rank change is shown.

2D Space-time coloured pixels is widely used to study patterns of traffic congestion in space-time. It was used to display data from loop detectors. The space-time coloured pixels consists of two axes: a device position axis and a time axis. Each pixel represents the magnitude of traffic parameter, in colours, measured from a monitoring device at a particular time. Any anomaly of the detectors can also be shown easily by this approach.

60.5.1.3 Travel Time

The previous two subsections use time as a reference for other types of data (changes in geometry and in attributes of spatial objects). Here we pay special attention for travel time representation since time itself is the data to be represented and special techniques are developed for this purpose. There are two techniques that are used to present travel time on maps.

A *cartogram* is a map that distorts geographic space on maps to represent attributes of spatial data. For example, the tube map of London arrange all the tube lines in six zones in order to show the distance to the centre of London (Zone 1), which is not the exact physical (geometric) locations of the tube lines and stations. Using this technique, travel time on transportation networks can be represented using distance on a map, an example of which is the travel time tube map that distorts real geographic layout of tube lines in London in order to show travel time between stations.

The "isochrone" is another technique that is employed to represent travel time. Isochrones are similar to contour lines on a map, but an isochrone line connects points of equal travel time from a given origin (Brunsdon et al. 2007). The isochrone is a great alternative to the cartogram as it does not distort the underlying map.

60.5.2 3D Visualization

60.5.2.1 3D Space-Time Cube

The 3D space-time cube (or, alternatively, space-time aquarium) was proposed by Hagerstand (1970). A 3D space-time cube consists of two dimensions of geographic locations on a horizontal plane, and a time dimension in the vertical plane (or axis). The space-time cube is normally applied to represent trajectories of objects in 3D space-time dimension, or "space-time paths." Trajectories are normally from GPS data nowadays, and they are represented as lines in the 3D space-time.

3D space-time cube has two main limitations. Firstly, the 3D display makes it difficult to refer space-time paths to geo-locations and time. Secondly, the space-time cube has difficulty in displaying large amounts of data. However, interactive techniques can be used to reduce cluttering when displaying a large amount of data. With interactive functions, users can decide which data to display and can zoom and rotate the cube on its axes. Data aggregation (such as generalized space-time path) can also improve visualization on 3D space-time cube.

60.5.2.2 3D Isosurface

An isosurface is a three-dimensional analog of an isoline. It is a surface that represents points of a constant value (e.g., pressure, temperature, velocity, density) within a volume of space. Isosurface has been employed in various applications such as medical imaging, fluid dynamics, astrophysics, chemistry and quantum mechanics. Isosurfaces are popularly used to visualize volumetric datasets, which consist of a 3D location with one scalar or vector attribute. The data sets are structured as (x,y,z,v), where (x,y,z) are the spatial coordinates and v is an attribute. The 3D isosurface has also been applied to visualize incident data, which are structured as (x,y,t), where (x,y) are two dimensional spatial coordinates and t is the time when the incident occurred (Brunsdon et al. 2007). Isosurfaces have great potential to show the development of space-time processes such as congestion on the traffic network.

60.5.2.3 3D Wall Map

The 3D wall map is a 2D road map with an additional time dimension to display change. Each layer represents the situation at a time. Cheng et al. (2010) employed the technique to represent travel delay during the morning peak in central London at four consecutive Mondays in October 2009. The layout of the link map represents the real geographical layout of the road network. The colours between layers represent the unit journey time (minutes per kilometre), with yellow and red colours showing the highly congested areas (travel time more than 5 min per kilometre) (Fig. 60.3).

60.5.3 Animated Maps

The first computer based animation map was created by Tobler in 1970 (Tobler 1970). He used 3D animated maps to display simulated urban growth data
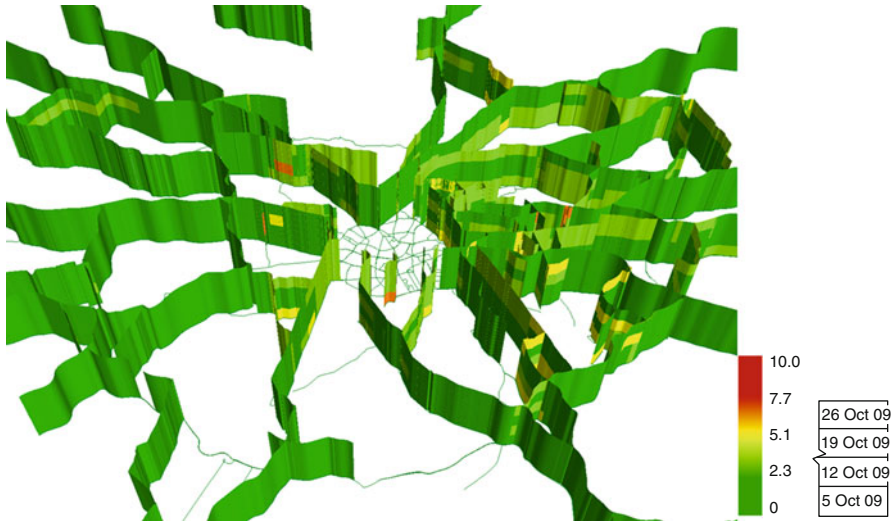
Fig. 60.3 Wall map of travel delay (mins/km) of outbound roads during the morning peak on 5,12,19,26 October 2009 (Cheng et al. 2011a)

in the Detroit region of the US. However, animated maps were not widely used for many years due to the difficulty in distributing and playing back large data files. However, with improvements in computing power and internet technology over the past three decades, animation maps have become a very active area of research and are now distributed widely on the internet. Weather maps and traffic maps are two of the many examples.

An animated map has two outstanding advantages. The first one is that an animation map can be used as an alternative to a static map. It can be employed to emphasize key attributes by using, for example, blinking symbols "to attract attention to a certain location on the map" (Kraak and Klomp 1995). The second advantage is that it provides additional visual variables called "dynamic variables" such as "duration," "rate of change," "order of change," "frequency," "display time," and synchronization (MacEachren et al. 2004).

60.5.4 Visual Analytics: The Current Visualization Trend

Visual analytics is an outgrowth of the field of scientific and information visualization. It refers to "the science of analytical reasoning facilitated by interactive visual interfaces" (Thomas and Cook 2005). The emergence of visual analytics has been driven by the fact that we have no proper tools to leverage large amounts of data. Visual analytics is an iterative process that involves information gathering, data pre-processing, knowledge representation, and decision making. Normally, unknown data are visualized in order to give a basic view of that data, then users will use their perception (intuition) to gain further insights from the images

produced by visualization. The insights generated by this human perception are then transformed into knowledge. After users have gained certain knowledge, they can generate hypotheses that will be used to carry out further analysis using available data analysis and exploration techniques. The results from analytical process will be visualized for presentation, and further gain in knowledge.

Visual analytics is much more than simple visualization. It can rather be seen as an integral approach to combining visualization with human factors, and data analysis (Keim et al. 2008). Visual analytics is becoming more important to many disciplines including scientific research, business enterprise, and other areas that face problems of overwhelming avalanche of data. GIS, also, is now facing this massive data problem. The concept of visual analytics was introduced into GIS, namely "Geovisual Analytics". Geovisual analytics has its specific focus on space and time; posing different specific research problems, and demands special approaches in solving generic research problems of Visual Analytics.

60.6 Conclusions

Since the concept of knowledge discovery from databases (KDD) was proposed in 1988, tremendous progress has been made in data mining and spatial data mining (Miller and Han 2009; Shekhar et al. 2011). STDM is only possible based upon the progress in those areas, along with GIS and geocomputation. This chapter introduces the fundamentals of STDM, which consists of space-time prediction, clustering and visualization.

As for space-time prediction, we have discussed the statistical (parametric) models, including families of STARIMA models, space-time geostatistical models, spatial panel data models, and space-time GWR. The challenge in statistical models lies in the non-stationary and non-linearity of space-time data. How to calibrate the spatio-temporal autocorrelations in the models is the bottleneck of statistical approaches. For low spatio/temporal resolution data, use of parametric statistical models may be preferable because of their explanatory power and interpretability. Due to their ability to deal with multi-dimensional nonlinear data machine learning methods are becoming more popular for large datasets. We have briefly introduced artificial neural networks (ANNs), support vector machines (SVMs), and other methods (Kernel-based approach and self-organized maps) for space-time analysis. However, the interpretability of machine learning is low, and a hybrid framework with both statistical and machine learning approaches might be helpful for this.

Space-time clustering can be used to extract meaningful patterns (clusters) in the data. It can also be used to detect outliers or emerging phenomena (epidemic outbreak or traffic congestion). Considering the spatial, temporal and thematic attributes seamlessly and simultaneously, and the dynamicity in the data is the most difficult challenge in spatio-temporal clustering. Spatio-temporal scan statistics (STSS) sheds lights on this aspect, though efforts are needed to improve computation efficiency and to reduce the false alarm rate.

Space-time visualization explores the patterns hidden in the large data sets by using advanced visualization and animation techniques. This includes conventional static 2D maps as well as newly developed 3D wall maps and isosurface, which shows the hotspots in space-time. Recently "Visual Analytics" and "Geovisual Analytics" have emerged as an iterative process (or tools) that involves information gathering, data pre-processing, knowledge representation/visualization, and decision making. Still, real-time visualization of dynamic processes is still very challenging due to large volume and high dimensions of the data. For examples, methods are needed to show the evolution and dissipation of crime or traffic congestion in space and time simultaneously.

However, the field of STDM is far from mature, and further research is needed in the following areas:

a. New methods and theory are needed for mining crowd sources such as data contributed by citizens and volunteers. These are often extremely noisy, biased, and nonstationary. One example of such data is the trajectory data obtained from smart phones or other sensors. This area is relevant to the recent development of citizen sciences and VGI in particular.

b. Theory and methods need to be developed to extract meaningful patterns from those individual sensors and put them under the framework of networks and network complexity such as transport and social-networks made up of those individual. Under network, the interaction and dynamic flows should be considered in mining spatio-temporal patterns. This aspect is relevant to complexity theory and network dynamics in particular.

c. STDM for emergency and tipping points detection, leading to the generation of, actionable knowledge, i.e., finding the emergent patterns and tipping points of economic crises and disease epidemics. It is important to find outliers, but more important is finding the critical points before the system breaks down so that mitigating action can be taken to avoid the worst scenarios such as traffic congestion and epidemic transmission.

d. Another challenge of STDM is how to calibrate, explain and validate the knowledge extracted. A good example of this is the calibration of spatial (or spatio-temporal) autocorrelation. Higher order spatial autocorrelation models have been developed, but the pitfalls have also been found (LeSage and Pace 2011). Nonstationarity and autocorrelation is fundamental to our observation (or our empirical test) of reality, it is hard to prove that the higher order autocorrelation comes from the first to the second, then to the third; or from the first to the third directly, which makes the explanation unconvincing. Furthermore, validation is difficult – so far Monte Carlo simulation is the main tool for simulation, which is also based upon a statistical distribution, which is hardly provable. This makes machining learning more promising in future STDM.

e. Technically, grid computation and cloud computation allow data mining to be implemented at multiple computer sources. Even so, when the data volume is increased, the capacity of software and hardware is still limited. How to scale the algorithm to larger networks will always be a challenge for data mining given the

increase of data volume is far quicker than the improvement in the performance of data processors.

Please notice that the content of this chapter is mainly around spatial data in point, line and lattices, but not on image data, which is another broad area of research. Also, due to the limit of length, we do not include the progress on space-time simulation, which includes agent-based modeling (ABM) and cellular automata (CA). ABM has been used across many disciplines to demonstrate the impact of individual decisions and choices on the nature of a system (Gilbert 2007). Such examples include the individual behavior of birds in flocks, ants in colonies and people in crowds – all entities are acting independently yet contribute to a larger body. There is great potential within ABM to replicate and predict system changes over space and time. In (Manley et al. 2011), the agent-based simulation has demonstrated the link between individual choice and behavior in abnormal conditions with the formation and movement of urban road congestion. CA is a discrete model studied in computability theory, mathematics, physics, complexity science, theoretical biology and microstructure modeling. It consists of a regular grid of cells, each in one of a finite number of states, such as "On" and "Off." It has been widely used in urban planning and landuse change modeling.

References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht

Box G, Jenkins G (1970) Time series analysis: forecasting and control. Holden-Day, San Francisco

Brunsdon C, Corcoran J, Higgs G (2007) Visualising space and time in crime patterns: a comparison of methods. Comput Environ Urban Syst 31(1):52–75

Cheng T, Li Z (2006) A multi-scale approach for spatial-temporal outlier detection. Trans GIS 10(2):253–263

Cheng T, Tanaksaranond G, Emmonds A, Sonoiki D (2010) Multi-scale visualization of inbound and outbound traffic delays in London. Cartogr J 47:323–329

Cheng T, Haworth J, Wang J (2011a) Spatio-temporal autocorrelation of road network data. J Geograph Syst. http://www.springerlink.com/content/4l84v7072737621p/ Accessed 12 Oct 2011

Cheng T, Wang J, Li X (2011b) A hybrid framework for space–time modeling of environmental data. Geogr Anal 43(2):188–210

Elhorst JP (2003) Specification and estimation of spatial panel data models. Int Reg Sci Rev 26(3):244–268

Fischer MM (2006) Spatial analysis and geocomputation. Springer, Berlin/ Heidelberg

Gilbert N (2007) Agent-based models. Sage, London

Griffith DA (2010) Modeling spatio-temporal relationships: retrospect and prospect. J Geogr Syst 12(2):111–123

Hägerstrand T (1970) What about people in regional science? Papers Reg Sci 24(1):1–12

Heuvelink GBM, Griffith DA (2010) Space-time geostatistics for geography: a case study of radiation monitoring across parts of germany. Geogr Anal 42(2):161–179

Hsieh WW (2009) Machine learning methods in the environmental sciences: neural networks and
    Kernels, 1st edn. Cambridge University Press, Cambridge
Huang B, Wu B, Barry M (2010) Geographically and temporally weighted regression for modeling
    spatio-temporal variation in house prices. Int J Geogr Inf Sci 24(3):383–401
Kamarianakis Y, Prastacos P (2005) Space-time modeling of traffic flow. Comput Geosci
    31(2):119–133
Kanevski M, Timonin V, Pozdnukhov A (2009) Machine learning for spatial environmental data:
    theory, applications, and software. CRC Press, Boca Raton
Keim D, Andrienko G, Fekete JD, Görg C, Kohlhammer J, Melançon G (2008) Visual analytics:
    definition, process, and challenges. Inf Visual 4950:154–175
Kraak MJ, Klomp A (1995) A classification of cartographic animations: towards a tool for the
    design of dynamic maps in a gis environment. In: Proceedings of the seminar on teching
    animated cartography. Madrid, Spain, pp 29–35
Kyriakidis PC, Journel AG (1999) Geostatistical space–time models: a review. Math Geol
    31(6):651–684
LeSage JP, Pace RK (2011) Pitfalls in higher order model extensions of basic spatial regression
    methodology. http://www.be.wvu.edu/econ_seminar/documents/11-12/lesage.pdf. Accessed
    on 15 Nov 2011
MacEachren A, Gahegan M, Pike W, Brewer I, Cai G, Lengerich E, Hardisty F (2004) Geovisua-
    lization for knowledge construction and decision-support. IEEE Comput Graph Appl 24:13–17
Manley E, Cheng T, Emmonds A (2011) Understanding route choice by using agent-based
    simulation. In: Proceedings of 11th international conference of geocomputation, London,
    20–22 July 2011, pp 54–58
Miller HJ, Han J (2009) Geographic data mining and knowledge discovery: an overview.
    In: Miller H, Han J (eds) Geographic data mining and knowledge discovery, 2nd edn. Taylor
    and Francis, Boca Raton
Monmonier M (1990) Strategies for the visualization of geographic time-series data.
    Cartographica 27(1):30–45
Neill DB (2008) Expectation-based scan statistics for monitoring spatial time series data. Int
    J Forecast 25(3):498–517
Pfeifer PE, Deutsch SJ (1980) A three-stage iterative procedure for space-time modeling.
    Technometrics 22(1):35–47
Pozdnoukhov A, Matasci G, Kanevski M, Purves RS (2011) Spatio-temporal avalanche forecast-
    ing with support vector machines. Nat Hazards Earth Syst Sci 11:367–382
Rey SJ, Janikas MV (2010) STARS: Space-time analysis of regional systems. In: Fischer MM,
    Getis A (eds) Handbook of applied spatial analysis: software tools, methods and applications.
    Springer, Berlin/Heidelberg, pp 91–112
Shekhar S, Evans MR, Kang JM, Mohan P (2011) Identifying patterns in spatial information:
    a survey of methods. Wiley Interdiscip Rev Data Min Knowl Discov 1(3):193–214
Slingsby A, Wood J, Dykes J (2010) Treemap cartography for showing spatial and temporal traffic
    patterns. J Maps 2010:135–146
Thomas JJ, Cook KA (2005) Illuminating the path: the research and development agenda for visual
    analytics. IEEE, Los Alamitos
Tobler W (1970) A computer movie simulating urban growth in the detroit region. Econ Geogr
    46:234–240
Vapnik V (1999) The nature of statistical learning theory, 2nd edn. Springer, London