Spatial data are an important source of scientific information. The development of high capacity and fast desk and laptop computers and the concomitant creation of geographic information systems has made it possible to explore georeferenced or mapped data as never before. This Handbook summarizes, explains, and demonstrates the nature of current models, methods, and techniques particularly designed for the analysis of spatial data. The book is designed to be a desk reference for all researchers just getting into the field of spatial data analysis as well as for seasoned spatial analysts. Relevant references are given whenever possible to direct researchers to the most useful writings on the subject.

Unlike most compendia of this nature, the book starts out by exploring the available software for spatial analysis. We focus on the tools that make analysis possible. The volume then describes briefly but clearly the many techniques embodied in the fields of exploratory spatial data analysis, spatial statistics, geostatistics, and spatial econometrics. In addition, attention is given to the methods used for the analysis of remotely sensed data. Finally, a number of example sections are included that demonstrate the application of spatial analysis in the economic, environmental, and health sciences. The wide range of approaches described helps readers better understand their data and the techniques needed for spatial analysis.

The volume features contributions from the very best scholars in the field. Their explanations are able to communicate the fundamental ideas of their subject area succinctly and accessibly.
A.6 Space-Time Intelligence System Software for the Analysis of Complex Systems

Geoffrey M. Jacquez

A.6.1 Introduction

The representation of geographies (e.g. census units), demographics and populations as unchanging rather than dynamic is due in part to the static world-view of GIS software, which has been criticized as not fully capable of representing temporal change and better suited to ‘snapshots’ of static systems (Goodchild 2000; Hornsby and Egenhofer 2002; Jacquez et al. 2005). This static view hinders the mapping, representation, and analysis of dynamic health, socioeconomic, and environmental information for populations that are dispersed and mobile – a key characteristic of the human condition (Schaerstrom 2003).

Several approaches to modifying GIS to better handle the temporal dimension have been proposed. Yearsley and Worboys (1995) proposed a space time object model that integrates abstract spatial data types with a geometric layer to construct a higher-level topological data model, Raper and Livingstone (1993) used an object oriented approach to represent dynamic spatial processes as spatio-temporal aggregations of point objects, and Peuquet and Duan (1993) formulated an event-based spatio-temporal data model (ESTDM) that maintains spatio-temporal data as a sequence of temporal events associated with a spatial object. See Miller (2005b) for a review of alternative data models.

Hägerstrand’s (1970) seminal work in time geography has led to geometric and mathematical constructs for quantifying human mobility including geospatial lifelines, space-time prisms, and techniques for propagating location uncertainty through time (Miller 1991; Kwan 2003; Han et al. 2005; Miller 2005a). These, in turn, have provided a quantitative basis for the development of statistics and modeling approaches suited to the analysis of temporally dynamic systems. For example, Sinha and Mark (2005) proposed a Minkowski metric to quantify dissimilarity between geospatial lifelines; Han et al. (2005) present a $K$-function calculated from the spatial pattern of place of residence at specific time slices; $Q$-statistics assess case-control clustering (Jacquez and Meliker 2008) and space-
<table>
<thead>
<tr>
<th>Category</th>
<th>Functionality</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data types</td>
<td>Points</td>
<td>Both static (e.g. space only) and temporally dynamic when points move through time</td>
</tr>
<tr>
<td></td>
<td>Lines</td>
<td>Both static and temporally dynamic</td>
</tr>
<tr>
<td></td>
<td>Polygons</td>
<td>Both static and dynamic such that polygons change shape (e.g. morph) and location</td>
</tr>
<tr>
<td></td>
<td>Mobility histories</td>
<td>For representing geospatial lifelines and activity spaces</td>
</tr>
<tr>
<td></td>
<td>Rasters</td>
<td>Both static and temporally dynamic for representing space-time fields</td>
</tr>
<tr>
<td>Visualization</td>
<td>Linked windows</td>
<td>Cartographic and statistical brushing with time-enabled spatial objects; time synchronization of maps and graphs</td>
</tr>
<tr>
<td>Tables</td>
<td>Tables</td>
<td>Attribute values can change through time</td>
</tr>
<tr>
<td>Maps</td>
<td>Maps</td>
<td>Of point, vector, polygons, mobility histories, raster data, spatial weights</td>
</tr>
<tr>
<td></td>
<td>Cluster maps</td>
<td>Display locations of spatial outliers and clusters of low and high values and how they change through time</td>
</tr>
<tr>
<td></td>
<td>Change maps</td>
<td>Also called difference maps, show absolute and relative change between time periods</td>
</tr>
<tr>
<td></td>
<td>Disparity maps</td>
<td>Show where and when a target population differs significantly from a reference population (e.g. health disparities)</td>
</tr>
<tr>
<td></td>
<td>Animations</td>
<td>Display movies of time-dynamic spatial data</td>
</tr>
<tr>
<td>Statistical graphics</td>
<td>Box plots</td>
<td>All statistical graphics except the time series plot are time enabled, displaying data relationships through time</td>
</tr>
<tr>
<td></td>
<td>Histograms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scattergrams</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Principal coordinate plots</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time series plots</td>
<td>Show how attribute values for space-time objects change through time</td>
</tr>
<tr>
<td></td>
<td>Variogram clouds</td>
<td>Visualize spatial variance at different spatial lags and through time</td>
</tr>
<tr>
<td>Weight sets</td>
<td>Spatial weights</td>
<td>Nearest neighbor, adjacency, distance, and inverse distance weights are time-dynamic when geography (for example, locations of points) changes through time</td>
</tr>
<tr>
<td>Smoothing</td>
<td>Empirical Bayesian</td>
<td>Bayesian smoother using user-specified spatial weights</td>
</tr>
<tr>
<td></td>
<td>Poisson</td>
<td>Poisson smoother for point/rare event data</td>
</tr>
<tr>
<td>Pattern recognition methods</td>
<td>Local Moran</td>
<td>Univariate and bivariate, with or without temporal lags, for points or polygons</td>
</tr>
<tr>
<td></td>
<td>Global Moran</td>
<td>Provided automatically with local Moran</td>
</tr>
<tr>
<td></td>
<td>Local G and G</td>
<td>For points or polygons</td>
</tr>
<tr>
<td></td>
<td>Besag and Newell</td>
<td>For case and population at risk, using points or polygons</td>
</tr>
<tr>
<td></td>
<td>Turnbull</td>
<td>For case and population at risk data, points or polygons.</td>
</tr>
<tr>
<td></td>
<td>Disparity statistics</td>
<td>For reference and target populations using rates and population sizes</td>
</tr>
<tr>
<td></td>
<td>Variogram analysis</td>
<td>Isotropic, anisotropic, automated fitting, time dynamic</td>
</tr>
<tr>
<td>Modeling</td>
<td>Aspatial regression</td>
<td>Linear, logistic and Poisson regression, using full model, best subset, forward or backward stepwise selection. Model is fitted through time.</td>
</tr>
<tr>
<td>(all time-dynamic)</td>
<td>Geographically weighted regression</td>
<td>Linear, logistic and Poisson regression, using user-specified or automatically optimized bandwidth</td>
</tr>
<tr>
<td></td>
<td>Kriging</td>
<td>Using traditional, standardized, residuals, weighted and Poisson estimators; includes simple, ordinary, kriging with a trend and Poisson kriging</td>
</tr>
</tbody>
</table>
Recent technological advances have resulted in Space Time Intelligence Systems (STIS) that implement constructs for representing temporal change (AvRuskin et al. 2004; Greiling et al. 2005; Jacquez et al. 2005; Meliker et al. 2005). The STIS technology has several advantages. First, it is founded on space-time data structures, enabling complex space-time queries not possible in conventional ‘spatial only’ GIS. Second, it incorporates statistical tests for space-time pattern such as univariate and bivariate local indicators of spatial autocorrelation and clustering that are automatically calculated through time, resulting in cluster animations that capture space-time change. Third, it employs dynamic linked windows that enable both cartographic and statistical brushing through time. Fourth, it calculates weight matrices for dynamic systems where points can move through time, and where polygons can morph, merge and divide such that pattern recognition and modeling readily account for dynamic and complex time geographies. Fifth, it constructs spatio-temporal statistical models including linear, Poisson and logistic regression, geographically weighted regression variogram models, and kriging. Finally, it displays animated ‘movies’ for exploring how variables (for example, health outcomes such as maps of incidence, mortality, case counts and expectations, and clusters themselves) change through space and time. Development of the STIS software was funded by grants from the National Institutes of Environmental Health Sciences and the National Cancer Institute. This technology is well suited to the representation, visualization, modeling and simulation of dynamic patterns and processes, and its functionality (see Table A.6.1) is the topic of the balance of this Chapter.

A.6.2 An approach to the analysis of complex systems

Geographic systems typically are large, dynamic and complex. Our approach to analyzing complex systems in STIS consists of three stages: development of cognitive models, exploratory space-time data analysis, and modeling; with each stage informing the others.

Cognitive and ontological models have to do with the mental representation of the underlying causal mechanisms that drive the relationships observed in a complex system. These usually are based on speculation, an understanding of prior research findings, and by one’s experience with similar systems. They guide exploratory data analysis, and form the basis on which more detailed data-based and process-based models are constructed. They are developed by visualizing and interacting with the data, and are continually refined through data analysis and modeling.
Exploratory Data Analysis (EDA) is founded on exploratory methods for quickly producing and visualizing simple summaries of data sets to reveal relationships and insights that often cause one to refine the cognitive model (Tukey 1977). Exploratory Space-Time Data Analysis (ESTDA) is made possible by software systems that incorporate spatial and temporal data, dynamic linked windows, statistical and cartographic brushing, and can generate hypotheses to be evaluated using clustering, inferential statistics and models. The objective of exploratory techniques is to illuminate and quantify relationships in order to increase the analyst’s knowledge of the complex system, giving rise to testable hypotheses and to relationships that can be modeled.

Models of data include statistical tools such as ANOVA, regression and correlation, and are used to quantify relationships among variables, to test statistical hypotheses, and to identify factors that drive variability in the experimental system. These models require data of sufficient quality to estimate model coefficients (for example, regression intercept), and that the researcher has sufficient knowledge to be able to identify dependent and independent variables, and their relevant parameters. Models of data are often used for interpolation and for prediction but do not necessarily convey information regarding underlying causal mechanisms.

Models of process require a detailed understanding of the mechanics of the system being studied, and incorporate this understanding directly into the model itself. Examples of process models include infection transmission systems in which the population is structured into susceptible, infectious, and immune subgroups, and in which the model parameters describe mechanistic processes such as infection transmission to susceptible individuals (Koopman et al. 2001).

STIS provides a platform for analyzing complex space-time systems, from visualization, the quantification of geographic relationships using weight matrices that change through time, the identification of space-time pattern to generate hypotheses, to models that may be used for estimation and prediction, as described below.

A.6.3 Visualization

The first step is to enter data into STIS and to then create maps, animations, and statistical graphics to explore relationships in the data. Supported data types include points, mobility histories, lines and polygons. STIS reads ESRI shape files, excel, dbf and text files, using time series or time slice formats. A time slice means all objects in the geography change attribute values simultaneously, so that one may assign a time stamp defining an interval that applies to all objects in a data set. An example would be lung cancer mortality rates for white males in U.S. counties from 1950 to 1955. Time series data arise when the values of the attributes change asynchronously among different spatial objects. For example, one location may be sampled at hourly intervals, while another is sampled daily.
After the data are entered one next creates maps, and then animates them to obtain an initial impression of space-time patterns. Time series plots are used to explore how variable values change through time. Linked brushing on the maps, statistical graphics and tables, along with time animation, supports rapid identification of relevant space-time patterns (see Fig. A.6.1).

**A.6.4 Exploratory space-time analysis**

*Dynamic spatial weights*: Cluster analysis, autocorrelation analysis, spatial regression, geostatistics and other techniques in STIS rely on weights to model geographic relationships among the objects. STIS automatically calculates spatial weight matrices needed for cluster analyses, and prompts the user when more detailed weights or kernels are required for methods such as geographically weighted regression and geostatistics. In Fig. A.6.2 the user is exploring the spatial weight connections in counties in the Northeastern United States using cen-
troids with five nearest neighbors (left) and polygon adjacencies (right). The use of centroid locations to represent geographic relationships among area-based data such as counties can produce misleading results since the spatial support (for example, area and configuration of the counties) is ignored (Jacquez and Greiling 2003).

The spatial weights in STIS are dynamic, so that changing geographies are modeled in a realistic fashion. Examples include census geography; zip-code geographies; area-codes; land parcel data; and land use maps, all of which change through time. Dynamic spatial weights are used by cluster analysis and modeling techniques (including geographically weighted regression, variogram models, and kriging) so that temporal change in both geographic relationships and attribute values are fully accounted for.

![Fig. A.6.2.](image)

*Fig. A.6.2.* STIS visualizes spatial weights by outlining the selected location (centroid or polygon) in gold, and the localities to which it is connected in blue. The five nearest neighbors using centroids (left) differ from border adjacencies (right). The spatial weights for queried locations are written to the log view (not shown) for validation

**Pattern recognition:** STIS provides cluster tests for both point data and polygon data, including the local Moran (Anselin 1995), $G$ statistics (Getis and Ord 1992; Ord and Getis 1995), Besag and Newell (1991) and Turnbull’s (Turnbull et al. 1990) tests. Both absolute and relative disparity statistics identify significant differences in outcomes (for example, disease incidence and mortality, tumor staging, health screening utilization) through space and time (Goovaerts 2005). Spa-
Spatial pattern recognition may also be accomplished using variogram analysis, the point of departure for which is the variogram cloud. STIS provides automatic variogram fitting and both the variogram cloud and variogram models are time-dynamic. Basic variogram models include spherical, exponential, cubic, Gaussian and power models (Fig. A.6.3). Automatic variogram fitting selects from among these models to find that model which provides the best fit, along with the corresponding parameter estimates.

**Outlier detection**: An important step in exploratory space-time data analysis is the identification of outliers – observations whose values are unusual when considered in the context of the sample. Outlier analysis methods in STIS include the box plot, anomaly detection using local indicators of spatial autocorrelation, detection of geostatistical outliers via statistical brushing on the variogram cloud, and the exploration of deviations from model predictions using these techniques applied to model residuals.

![Automatic variogram model fitting of soil Cadmium concentrations in the Jura mountains, France.](image)

**Fig. A.6.3.** Automatic variogram model fitting of soil Cadmium concentrations in the Jura mountains, France. Notice the ‘Calculate best fit’ button in the variogram model window (left). Variogram estimators in STIS include traditional, standardized, residuals, weighted and Poisson. Here, an isotropic variogram model modeled a directional spatial pattern, and was then used to predict soil cadmium concentrations using kriging (raster map, right center). Data courtesy Pierre Goovaerts

## A.6.5 Analysis and modeling

STIS provides advanced modeling techniques including aspatial regression, geographically weighted regression, and geostatistics, as summarized below. All of these are time-enabled and automatically model changes in geography (for exam-
ple, morphing polygons and moving points) as well as attributes (for example, how the value associated with a spatial object changes) through time.

Aspatial regression: Exploratory space-time data analysis using visualization and pattern recognition methods often generates hypotheses regarding dependencies and associations among the variables. Before invoking spatial modeling approaches a researcher may first choose to employ aspatial models, and then evaluate pattern in the model residuals to determine whether more detailed space-time models are warranted. The rationale is one of parsimony – if an aspatial model adequately explains the observed variability then a more complex spatial model may not be warranted.

STIS provides linear, logistic and Poisson regression, and for complex models with several variables evaluates the fit of subsets of the independent variables using the full model (all variables), forward stepwise, backward stepwise, and best subset. The criterion for finding the best subset – that combination of independent variables that does the best job of explaining variability in the dependent variable – for linear models include $R^2$, adjusted $R^2$, $C(p)$, and AIC. $R^2$ selects the model with the largest reduction in residual sum of squares, and thus favors complex model with the largest number of terms. The adjusted $R^2$ criterion punishes models with too many terms. The smallest AIC (Akaike information criterion) trades off model fit and model complexity using, for linear regression, the residual sum of squares (RSS) penalized by two times the number of regression term degrees of freedom ($k$ = the number of regression parameters). Finally, the smallest Mallows $C(p)$ is another way of penalizing models with many independent variables. It is the residual sum of squares for the subset model being considered, divided by the error variance for the full model plus twice the number of regression degrees of freedom minus the total number of observations. Similar $C(p)$ values similar to the one for the full model are considered an indication of good candidate models. Appropriate model selection criteria are also provided for Poisson and logistic regression.

Geographically weighted regression (GWR): Most of the functionality and modeling approaches for aspatial regression are available as well in GWR (see Chapter C.5 for more details). Whereas aspatial regression makes strong assumptions regarding stationarity of the regression coefficients, GWR allows the regression coefficients to vary through geographic space and through time, and fits spatially and temporally local regression, with local estimates of model fit (for example, $R^2$, the regression coefficients and model residuals, correlations and other statistics). GWR has been pioneered by A. Stewart Fotheringham and Martin Charlton (currently at the National Center for Geocomputation, National University of Ireland), and Chris Brunsdon (University of Glamorgan, UK). Our implementation of this tool is based primarily upon their book on the topic (Fotheringham et al. 2002), but we have made some changes, which follow from the way in which many in the public health and environmental science fields are likely to use these tools. Our approach to GWR uses an unified framework for including both
Fig. A.6.4. Aspatial regression analysis of breast cancer in the northeastern United States. The user has conducted a linear regression modeling breast cancer mortality in white females as a function of xylene and availability of physicians (MD ratio), with poverty and median age as interaction terms. The regression residuals have been mapped (circles) with the breast cancer mortality in white females (left). A local Moran analysis found significant clusters of high and low residuals (map top center) and a global Moran’s I of 0.18 ($p < 0.001$). The presence of significant spatial autocorrelation in the residuals suggests an important predictor is missing and/or that a more detailed spatial model is needed.

geographical weighting, and an extra non-geographical weight dataset that allows for user-supplied knowledge of the ratio variances at each source point. One example of this type of weight is the use of population data as a weight set for mortality rates, which has the effect of assigning higher ‘confidence’ to mortality rates derived from areas with higher populations. Our goal is to treat this type of weighting together with geographic weighting within a unified framework. As a result, STIS uses a maximum weighted likelihood approach to calculate the regression parameters, parameter variances, parameter $R$-square, expected $y$-values, residuals and $y$-standard errors as well as the ‘local model’ $R$-square. This approach boils down to treating geographically weighted regression as a local extension of weighted aspatial regression. As a consequence GWR can be straightforwardly extended to non-linear regression procedures such as logistic and Poisson regression with parameter values and parameter variances calculated from a weighted log-likelihood formulation.

A key question when using GWR is the construction of the local kernel used to identify those observations to use when fitting a local regression. For kernels of fixed size STIS uses either a number of nearest neighbors or a range (distance) from the central observation, and allows weights to be assigned to the observations based on proximity to the center (for example, using Gaussian and bi-square
decay functions). Researchers may also choose to use adaptive kernels that determine the kernel bandwidth by an iterative estimation procedure that minimizes the sum of the differences between the observed value of the dependent variable and the model’s estimate of that value. This effectively results in a bandwidth specification that results in the best model ‘fit’ over the range of bandwidths specified by the researcher. We have found GWR to be particularly useful when concerned with prediction, since it typically results in mean local $R^2$ values that exceed the $R^2$ from the corresponding aspatial regression. In the course of an analysis it is important to first derive a reasonable regression model using aspatial techniques before proceeding to GWR.

**Geostatistics**: Geostatistics provides powerful techniques for prediction, interpolation and simulation (see Chapter A.7 for information on geostatistical software). As noted earlier, STIS provides automated variogram estimation methods for modeling spatial relationships through time (see Chapter B.6 for more details on the variogram and kriging). The variogram may then be used in kriging to develop models of how variable values change through space and time. The current release of STIS provides kriging of continuous attributes with or without secondary information. It supports simple kriging, ordinary kriging, kriging with a trend, factorial kriging and Poisson kriging. Underlying variogram models may account for directional components, and the search strategies for fitting the local kriging equations may be anisotropic as well. When the data are time-dynamic one can estimate the variogram model through time, or alternatively can specify one variogram model and then apply it over the entire time interval.

### A.6.6 Concluding remarks

This chapter has provided a quick overview of some of the methods and functionality that are now available in the Space-Time Intelligence System software. The development of this software has been motivated by a desire to break the bonds of what has been called ‘technological determinism’. This arises when tools and methods dictate the approaches that are used to solve problems, as summarized in the aphorism ‘When one has a hammer everything starts to look like a nail’. In spatial analysis two factors lead to technological determinism. First, there still is a strong tradition of using static data models as the basis for developing statistical approaches for spatial data. One still often sees observations subscripted to identify their location, but we less often see a subscript denoting time – when that observation was observed. This is an oversimplification when data in reality are time-dynamic, and results in the application of statistical methods that assume static data to systems that are in fact highly time-dynamic. Second, many of the software tools, such as Geographical Information Systems, were originally founded on a ‘static world view’ that may be appropriate for geology and other fields where system change is slow, but is less appropriate in economic geography, medical ge-
ography and other fields where the systems under scrutiny are highly dynamic. It is unusual, for example, to find software in which the underlying assumption is that the location, extent and attributes associated with an object may change through time. The STIS software is a solution to this problem, and the assumption of dynamic objects leads naturally to time-enabled data views, tables, maps, statistical graphics, and analysis methods, including clustering, regression, geographically weighted regression, variogram analysis and kriging. In the near future we expect to include Q-statistics – methods for the analysis of case-control data that account for residential mobility, covariates and risk factors (Jacquez et al. 2006; Jacquez and Meliker 2008). STIS was created by BioMedware, and is being distributed by TerraSeer, Inc. Details on the methods are available on the TerraSeer website, www.Terraseer.com.

References


Tukey JW (1977) Exploratory data analysis. Addison-Wesley, Reading [MA]

