Crime Travel Demand

► CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents

CrimeStat

► CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents

CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents

NED LEVINE
Ned Levine & Associates, Houston, TX, USA

Synonyms

CrimeStat; Spatial statistics program; Crime mapping; Hotspot; Centrographic measures; Interpolation; Space-time interaction; Knox test; Mantel test; Correlated walk; Journey to crime analysis; Geographic profiling; Crime travel demand

Definition

CrimeStat is a spatial statistics and visualization program that interfaces with desktop GIS packages. It is a standalone Windows program for the analysis of crime incident locations and can interface with most desktop GIS programs. Its aim is to provide statistical tools to help law enforcement agencies and criminal justice researchers in their crime mapping efforts. The program has many statistical tools, including centrographic, distance analysis, hot spot analysis, space-time analysis, interpolation, Journey-to-Crime estimation, and crime travel demand modeling routines. The program writes calculated objects to GIS files that can be imported into a GIS program, including shape, MIF/MID, BNA, and ASCII. The National Institute of Justice is the distributor of CrimeStat and makes it available for free to analysts, researchers, educators, and students.[1] The program is distributed along with a manual that describes each of the statistics and gives examples of their use [1].

Historical Background

CrimeStat has been developed by Ned Levine and Associates since the late 1990s under grants from the National Institute of Justice. It is an outgrowth of the Hawaii Pointstat program that was UNIX-based [2]. CrimeStat, on the other hand, is a Windows-based program. It is written in C++ and is multi-threading. To date, there have been three major versions with two updates. The first was in 1999 (version 1.0) with an update in 2000 (version 1.1). The second was in 2002 (CrimeStat II) and the third was in 2004 (CrimeStat III). The current version is 3.1 and was released in March 2007.

[1] The program is available at http://www.icpsr.umich.edu/crimestat.

Scientific Fundamentals

The current version of CrimeStat covers seven main areas of spatial analysis: centrographic; spatial autocorrelation, hot spot analysis, interpolation, space-time analysis, Journey-to-Crime modeling, and crime travel demand modeling.

Centrographic Measures

There are a number of statistics for describing the general properties of a distribution. These include central tendency of the overall spatial pattern, dispersion and directionality. Among the statistics are the mean center, the center of minimum distance, the standard distance deviation, the standard deviational ellipse, the harmonic mean, the geometric mean, and the directional mean [3].

Spatial Autocorrelation

There are several statistics for describing spatial autocorrelation, including Moran's I, Geary's C, and a Moran Correlogram [4,5,3]. There are also several statistics that describe spatial autocorrelation through the properties of distances between incidents including the nearest neighbor statistic [6], the linear nearest neighbor statistic, the K-order nearest neighbor distribution [7], and Ripley's K statistic [8]. The testing of significance for Ripley's K is done through a Monte Carlo simulation that estimates approximate confidence intervals.

Hot Spot Analysis

An extreme form of spatial autocorrelation is a *hot spot*. While there is no absolute definition of a 'hot spot', police are aware that many crime incidents tend to be concentrated in a limited number of locations. The Mapping and Analysis for Public Safety Program at the National Institute of Justice has sponsored several major studies on crime hot spot analysis [9,10,11].

CrimeStat includes seven distinct 'hot spot' analysis routines: the mode, the fuzzy mode, nearest neighbor hierarchical clustering [12], risk-adjusted nearest neighbor hierarchical clustering [13], the Spatial and Temporal Analysis of Crime routine (STAC) [14], K-means clustering, and Anselin's Moran statistic [15].

The *mode* counts the number of incidents at each location. The *fuzzy mode* counts the number of incidents at each location within a specified search circle; it is useful for detecting concentrations of incidents within a short distance of each other (e. g., at multiple parking lots around a stadium; at the shared parking lot of multiple apartment buildings).

The *nearest neighbor hierarchical clustering* routine defines a search circle that is tied to the random nearest neighbor distance. First, the algorithm groups incidents that are closer than the search circle and then searches for a concentration of multiple incidents within those selected. The center of each concentration is identified and all incidents within the search circle of the center of each concentration are assigned to the cluster. Thus, incidents can belong to one-and-only-one cluster, but not all incidents belong to a cluster. The process is repeated until the distribution is stable (first-order clusters). The user can specify a minimum size for the cluster to eliminate very small clusters (e. g., 2 or 3 incidents at the same location). Once clustered, the routine then clusters the first-order clusters to produce second-order clusters. The process is continued until the grouping algorithm fails. The *risk-adjusted nearest neighbor hierarchical clustering* routine follows the same logic but compares the distribution of incidents to a baseline variable. The clustering is done with respect to a baseline variable by calculating a cell-specific grouping distance that would be expected on the basis of the baseline variable, rather than a single grouping distance for all parts of the study area.

The *Spatial and Temporal Analysis of Crime* hot spot routine (STAC) is linked to a grid and groups on the basis of a minimum size. It is useful for identifying medium-sized clusters. The *K-means* clustering algorithm divides the points into *K* distinct groupings where *K* is defined by the user. Since the routine will frequently create clusters of vastly unequal size due to the concentration of incidents in the central part of most metropolitan areas, the user can adjust them through a separation factor. Also, the user can define specific starting points (seeds) for the clusters as opposed to allowing the routine to find its own.

Statistical significance of these latter routines is tested with a Monte Carlo simulation. The nearest neighbor hierarchical clustering, the risk-adjusted nearest neighbor hierarchical clustering, and the STAC routines each have a Monte Carlo simulation that allows the estimation of approximate confidence intervals or test thresholds for these statistics.

Finally, unlike the other hot spot routines, *Anselin's Local Moran* statistic is applied to aggregates of incidents in zones. It calculates the similarity and dissimilarity of zones relative to nearby zones by applying the Moran's I statistic to each zone. An approximate significance test can be calculated using an estimated variance.

Interpolation

Interpolation involves extrapolating a density estimate from individual data points. A fine-mesh grid is placed over the study area. For each grid cell, the distance from the center of the cell to each data point is calculated and is converted into a density using a mathematical function (a kernel). The densities are summed over all incidents to produce an estimate for the cell. This process is then repeated for each grid cell [16]. *CrimeStat* allows five different mathematical functions to be used to estimate the density. The particular dispersion of the function is controlled through a bandwidth parameter and the user can select a fixed or an adaptive bandwidth. It is a type of hot spot analysis in that it can illustrate where there are concentrations of incidents. However it lacks the precision of the hot spot routines since it is smoothed. The hot spot routines will show exactly which points are included in a cluster.

CrimeStat has two different kernel function, a single-variable kernel density estimation routine for producing a surface or contour estimate of the density of incidents (e. g., the density of burglaries) and a dual-variable kernel density estimation routine for comparing the density of incidents to the density of an underlying baseline (e. g., the density of burglaries relative to the density of households).

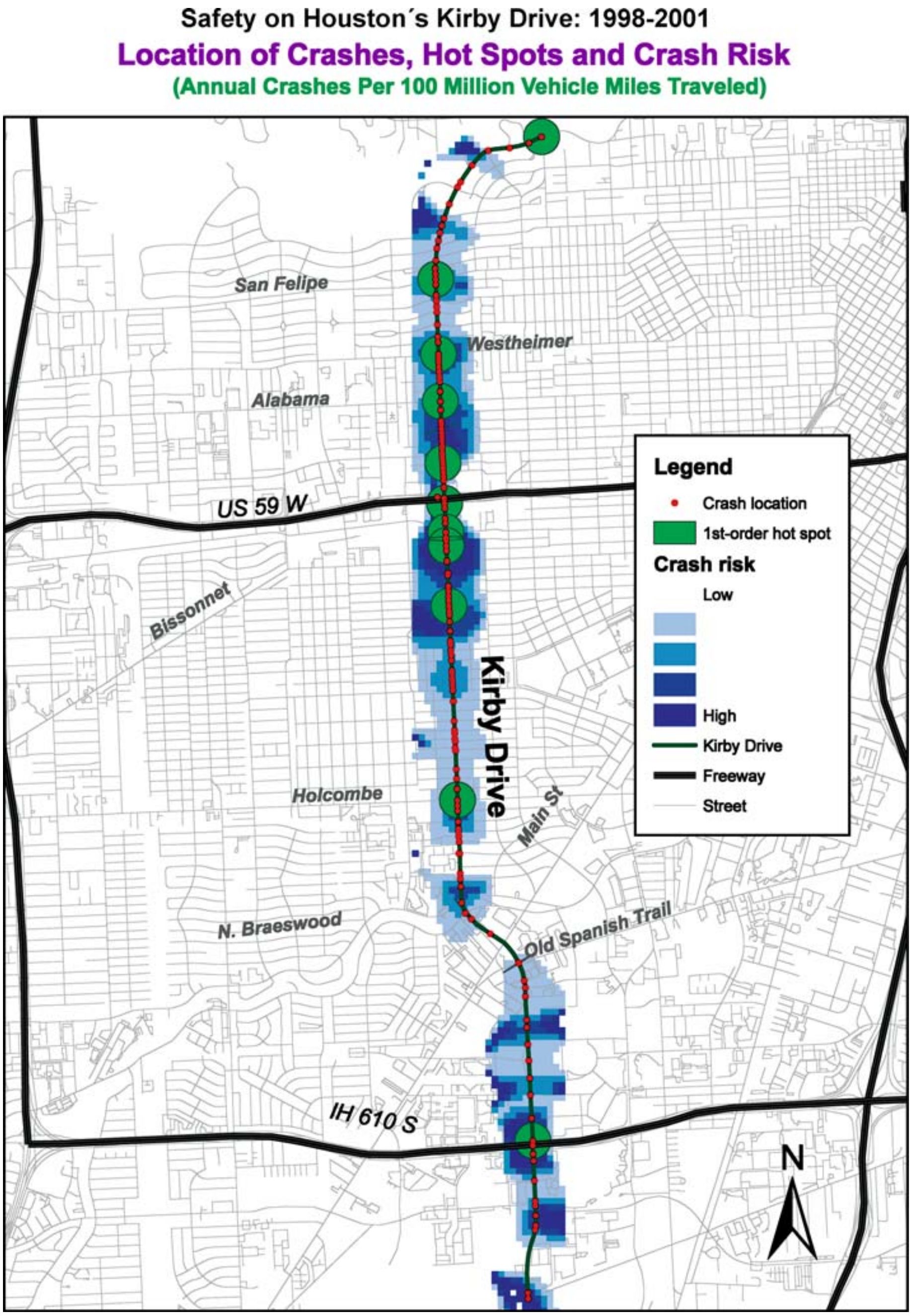
As an example, Fig. 1 shows motor vehicle crash risk along Kirby Drive in Houston for 1999–2001. *Crash risk* is defined as the annual number of motor vehicle crashes per 100 million vehicle miles traveled (VMT) and is a standard measure of motor vehicle safety. The duel-variable kernel density routine was used to estimate the densities with the number of crashes being the incident variable and VMT being the baseline variable. In the map, higher crash risk is shown as darker. As a comparison, hot spots with 15 or more incidents were identified with the nearest neighbor hierarchical clustering routine and are overlaid on the map as are the crash locations.

Space-Time Analysis

There are several routines for analyzing clustering in time and in space. Two are global measures – the *Knox* and *Mantel* indices, which specify whether there is a relationship between time and space. Each has a Monte Carlo simulation to estimate confidence intervals around the calculated statistic.

The third space-time routine is a specific tool for predicting the behavior of a serial offender called the *Correlated*

C



CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents, Figure 1 Safety on Houston's Kirby Drive: 1998–2001

Walk Analysis module. This module analyzes periodicity in the sequence of events committed by the serial offender by distance, direction, and time interval. It does this by analyzing the sequence of lagged incidents. A diagnostic correlogram allows the user to analyze periodicity by different lags. The user can then specify one of several methods for predicting the next incident that the serial offender will commit, by location and by time interval. Error is, of course, quite sizeable with this methodology because serial offenders don't follow strict mathematical rules. But the method can be useful for police because it can indicate whether there are any repeating patterns that the offender is following.

Journey-to-Crime Analysis

A useful tool for police departments seeking to apprehend a serial offender is *Journey-to-crime analysis* (sometimes known as *Geographic Profiling*). This is a method for estimating the likely residence location of a serial offender given the distribution of incidents and a model for travel distance [17,18,19,20]. The method depends on building a typical travel distance function, either based on empirical distances traveled by known offenders or on an *a priori* mathematical function that approximates travel behavior (e. g., a negative exponential function, a negative exponential function with a low use 'buffer zone' around the offender's residence).

CrimeStat has a Journey-to-Crime routine that uses the travel distance function and a Bayesian Journey-to-Crime routine that utilizes additional information about the likely origins of offenders who committed crimes in the same locations. With both types – the traditional distance-based and the Bayesian, there are both calibration and estimation routines. In the calibration routine for the Journey-to-Crime routine, the user can create an empirical travel distance function based on the records of known offenders where both the crime location and the residence location were known (typically from arrest records). This function can then be applied in estimating the likely location of a single serial offender for whom his or her residence location is not known.

The Bayesian Journey-to-Crime routine utilizes information about the origins of other offenders who committed crimes in the same locations as a single serial offender. Again, based on a large set of records of known offenders, the routine estimates the distribution of origins of these offenders. This information can then be combined with the travel distance function to make estimates of the likely location of a serial offender where the residence location is not known. Early tests of this method suggest that it is 10–15% more accurate than the traditional travel distance only method in terms of estimating the distance between the highest probability location and the location where the offender lived.

As an example, Fig. 2 shows a Bayesian probability model of the likely residence location of a serial offender who committed five incidents between 1993 and 1997 in Baltimore County, Maryland (two burglaries and three larceny thefts). The grid cell with the highest probability is outlined. The location of the incidents is indicated as is the actual residence location of the offender when arrested. As seen, the predicted highest probability location is very close to the actual location (0.14 of a mile error).

Crime Travel Demand Modeling

CrimeStat has several routines that examine travel patterns by offenders. There is a module for modeling crime travel behavior over a metropolitan area called *Crime Travel Demand* modeling. It is an application of travel demand modeling that is widely used in transportation planning [21]. There are four separate stages to the model. First, predictive models of crimes occurring in a series of zones (crime destinations) and originating in a series of zones (crime origins) are estimated using a non-linear (Poisson) regression model with a correction for over-dispersion [22]. Second, the predicted origins and destinations are linked to yield a model of crime *trips* from each origin zone to each destination zone using a gravity-type spatial interaction model. To estimate the coefficients, the calibrated model is compared with an actual distribution of crime trips.

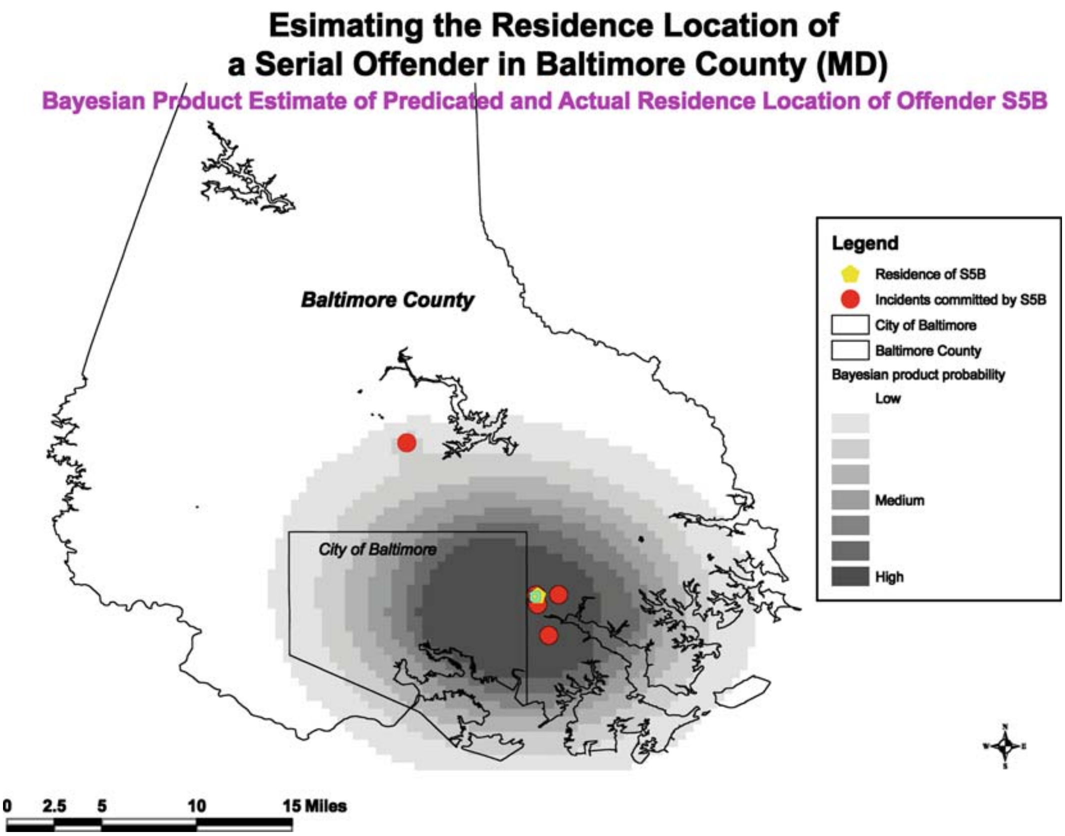
In the third stage, the predicted crime trips are separated into different travel modes using an approximate multinomial utility function [23]. The aim is to examine possible strategies used by offenders in targeting their victims. Finally, the predicted crime trips by travel mode are assigned to particular routes, either on a street network or a transit network. The cost of travel along the network can be estimated using distance, travel time, or a generalized cost using the A* shortest path algorithm [24].

Once calibrated, the model can be used to examine possible interventions or policy scenarios. For example, one study examined the travel behavior of individuals who were involved in Driving-while-Intoxicated (DWI) motor vehicle crashes in Baltimore County. Neighborhoods where a higher proportion of DWI drivers involved in crashes were identified as were locations where many DWI crashes had occurred. Interventions in both high DWI driver neighborhoods and the high DWI crash locations were simulated using the model to estimate the likely reduction in DWI crashes that would be expected to occur if the interventions were actually implemented.

Key Applications

CrimeStat is oriented mostly toward the law enforcement and criminal justice fields, but it has been used widely by researchers in other fields including geography, traffic safety, urban planning, sociology, and even fields like botany and forestry. The tools reflect a range of applications that criminal justice researchers and crime analysts might find useful, some describing the spatial distribution and others being targeted to particular offenders.

For example, hot spot analysis is particularly useful for police departments. Police officers, crime analysts and researchers are very familiar with the concentration of crime or other incidents that occur in small areas. Further they are aware that many offenders live in certain neighborhoods that are particularly poor and lacking in social amenities. There is a large literature on high crime



CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents, Figure 2 Estimating the residence location of a serial offender in Baltimore County (MD)

areas so that the phenomenon is very well known (e. g., see [25,26]). The hot spot tools can be useful to help police systematically identify the high crime areas as well as the areas where there are concentrations of offenders (which are not necessarily the same as the high crime locations). For example, the hot spot tools were used to identify locations with many red light running crashes in Houston as a prelude for introducing photo-enforcement. The Massachusetts State Police used the neighbor nearest hierarchical clustering algorithm to compare heroin and marijuana arrest locations with drug seizures in one small city [27].

Another criminal justice application is the desire to catch serial offenders, particularly high visibility ones. The Journey-to-Crime and Bayesian Journey-to-Crime routines can be useful for police departments in that it can narrow the search that police have to make to identify likely suspects. Police will routinely search through their database of known offenders; the spatial narrowing can reduce that search substantially. The CrimeStat manual has several examples of the Journey-to-Crime tool being used to identify a serial offender. As an example, the Glendale (Arizona) Police Department used the Journey-to-Crime routine to catch a felon who had committed many auto thefts [28].

Many of the other tools are more relevant for applied researchers such as the tools for describing the overall spatial distribution or for calculating risk in incidents (police typically are interested in the volume of incidents) or for modeling the travel behavior of offenders. Two examples from the CrimeStat manual are given. First, the spatial distribution of *"Man With A Gun"* calls for service during Hurricane Hugo in Charlotte, North Carolina was compared with a typical weekend [29]. Second, the single-variable kernel density routine was used to model urbanization changes in the Amazon between 1996 and 2000 [30].

Future Directions

Version 4 of CrimeStat is currently being developed (CrimeStat IV). The new version will have a complete restructuring to modernize it consistent with trends in computer science. First, there will be a new GUI interface that will be more Windows Vista-oriented. Second, the code is being revised to be consistent with the .NET framework and selected routines will be compiled as objects in a library that will be available for programmers and third-party applications. Third, additional statistics relevant for crime prediction are being developed. These include a spatial regression module using Markov Chain Monte Carlo methods and an incident detection module for identifying emerging crime hot spot spots early in their sequence. Version 4 is expected to be released early in 2009.

Cross References

- ► Autocorrelation, Spatial
- ► Crime Mapping and Analysis
- ► Data Analysis, Spatial
- ► Emergency Evacuation, Dynamic Transportation Models
- ► Hotspot Detection, Prioritization, and Security
- ► Movement Patterns in Spatio-temporal Data
- ► Nearest Neighbors Problem
- ► Patterns in Spatio-temporal Data
- ► Public Health and Spatial Modeling
- ► Routing Vehicles, Algorithms
- ► Sequential Patterns, Spatio-temporal
- ► Statistical Descriptions of Spatial Patterns

Recommended Reading

1. Levine, N.: CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 3.1). Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. March (2007)
2. Levine, N.: Spatial statistics and GIS: software tools to quantify spatial patterns. Journal of the American Planning Association **62**(3), 381–392 (1996)
3. Ebdon, D.: Statistics in Geography (2nd edn. with corrections). Blackwell, Oxford (1988)
4. Moran, P.A.P.: The interpretation of statistical maps. J. Roy. Stat. Soc. B **10**, 243–251 (1948)
5. Geary, R.: The contiguity ratio and statistical mapping. The Incorporated Statistician **5**, 115–145 (1954)
6. Clark, P.J., Evans, F.C.: Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology **35**, 445–453 (1954)
7. Cressie, N.: Statistics for Spatial Data. John Wiley & Sons, Inc, New York (1991)
8. Ripley, B.D.: Spatial Statistics. John Wiley & Sons, New York (1981)
9. Harries, K.: Mapping Crime: Principle and Practice. NCJ 178919, National Institute of Justice, US Department of Justice: Washington, DC. Available at http://www.ncjrs.org/html/nij/mapping/pdf.html (1999)
10. LaVigne, N., Wartell, J.: Crime Mapping Case Studies: Success in the Field (vol. 1). Police Executive Research Forum and National Institute of Justice, US Department of Justice: Washington, DC (1998)
11. Eck, J., Chainey, S., Cameron, J., Leitner, M., Wilson, R.E.: Mapping Crime: Understanding Hot Spots. Mapping and Analysis for Public Safety, National Institute of Justice (2005)
12. Everitt, B.S., Landau, S., Leese, M.: Cluster Analysis, 4th edn. Oxford University Press, New York (2001)
13. Levine, N.: Risk-adjusted nearest neighbor hierarchical clustering, In: Levine, N. (ed.) CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 3.0), Chapter 6. Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. November (2004)
14. Block, C.R.: STAC hot spot areas: a statistical tool for law enforcement decisions. In: Proceedings of the Workshop on Crime Analysis Through Computer Mapping. Criminal Justice Information Authority, Chicago, IL (1994)
15. Anselin, L.: Local indicators of spatial association – LISA. Geogr. Anal. **27**(2), 93–115 (1995)
16. Bailey, T.C., Gatrell, A.C.: Interactive Spatial Data Analysis. Longman Scientific & Technical, Burnt Mill, Essex, England (1995)
17. Brantingham, P.L., Brantingham, P.J.: Notes on the geometry of crime. In: Brantingham, P.J., Brantingham, P.L. (eds.) Environmental Criminology, pp. 27–54. Waveland Press, Inc., Prospect Heights, IL (1981)
18. Canter, D., Gregory, A.: Identifying the residential location of rapists, J. Forens. Sci. Soc. **34**(3), 169–175 (1994)
19. Rossmo, D.K.: Overview: multivariate spatial profiles as a tool in crime investigation. In: Block, C.R., Dabdoub, M., Fregly, S. (eds.) Crime Analysis Through Computer Mapping, pp. 65–97. Police Executive Research Forum: Washington, DC (1995)
20. Levine, N.: Bayesian Journey to Crime estimation (update chapter). In: Levine, N. (ed.) CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 3.1). Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. March 2007. Available at http://www.icpsr.umich.edu/crimestat (2007)
21. Ortuzar, J.D., Willumsen, L.G.: Modeling Transport, 3rd edn. John Wiley & Sons, New York (2001)
22. Cameron, A.C., Trivedi, P.K.: Regression Analysis of Count Data. Cambridge University Press, Cambridge, UK (1998)
23. Domencich, T., McFadden, D.L.: Urban Travel Demand: A Behavioral Analysis. North-Holland Publishing Co., 1975. Reprinted 1996. Available at: http://emlab.berkeley.edu/users/mcfadden/travel.html (1975)
24. Sedgewick, R.: Algorithms in C++: Part 5 Graph Algorithms, 3rd edn. Addison-Wesley, Boston (2002)
25. Cohen, L.E., Felson, M.: Social change and crime rate trends: a routine activity approach, Am. Soc. Rev. **44**, 588–608 (1979)
26. Wilson, J.Q., Kelling, G.: Broken Windows: The Police and Neighborhood Safety. Atlantic Monthly **29**(3), 29–38 (1982)
27. Bibel, B.: Arrest Locations as a Means for Directing Resources. In: Levine, N. (ed.) CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 3.0), Chapter 6. Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. November (2004)
28. Hill, B.: Catching the Bad Guy. In: Levine, N. (ed.) CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 3.0), Chapter 10. Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. November (2004)
29. LeBeau, J.L.: Distance Analysis:Man With A Gun Calls For Servicein Charlotte, N.C., 1989. In: Levine, N. (ed.) CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 3.0), Chapter 4. Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. November (2004)

30. Amaral, S., Monteiro, A.M.V., Câmara, G., Quintanilha, J.A.: Evolution of the Urbanization Process in the Brazilian Amazonia In: Levine, N. (ed.) CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 3.0), Chapter 8. Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC. November 2004 (2004)

Cross-Covariance Models

► Hurricane Wind Fields, Multivariate Modeling

CSCW

► Geocollaboration

Customization

► Mobile Usage and Adaptive Visualization

Cyberinfrastructure for Spatial Data Integration

ILYA ZASLAVSKY
San Diego Supercomputer Center, University of California San Diego, San Diego, CA, USA

Synonyms

E-Science; Heterogeneity; Virtualization, Resource; Standards

Definition

The term cyberinfrastructure (CI) refers to a new research environment that supports integration of geographically distributed computing and information processing services to enable a new level of data-intensive collaborative science enterprise. It includes high-performance data management and storage hardware and software, combined with secure data access and advanced information- and knowledge-management technologies and a variety of search, analysis, visualization, modeling and collaboration tools linked over high-speed networks, to create an enabling end-to-end framework for scientific discovery. CI applications in earth sciences span such disciplines as earthquake modeling and prediction, ecology, atmospheric sciences, hydrology and oceanography.

Historical Background

The term was first articulated at a press briefing on the Presidential Decision Directive (PDD-63) in 1998, in reference to information systems as the major component of the nation's critical infrastructures in need of protection (http://www.fas.org/irp/offdocs/pdd/pdd-63.htm). In 2003, the National Science Foundation's (NSF's) blue ribbon panel used the term in outlining the need to efficiently connect high-performance computing resources, information resources, and researchers, to support scientific discovery. Several large information technology projects were funded by the NSF, focused on CI development in earth science and other domains. In June 2005, an Office of Cyberinfrastructure was created at the NSF (http://www.nsf.gov/dir/index.jsp?org=OCI). At the same time, the National Institutes of Health supported advanced CI projects in biomedical sciences, and a range of infrastructure projects were developed in industry. These developments were accompanied by the emergence of relevant information exchange standards, more importantly *web services*, and *service-oriented architecture* (SOA), which now form the backbone of the large CI projects.

Several of these projects have been using spatial information technologies for monitoring distributed resources, searching for resources and extracting data fragments based on their spatial properties, integrating spatial data of different types, creating composite maps, and serving spatial data in standard formats, etc.

Scientific Fundamentals

The challenges of integrating spatial information from physically distributed spatial data sources derive from:

- Extreme heterogeneity in how web-accessible information is collected, represented, described, interpreted and queried
- Exponentially increasing volumes of available data, numbers of users and applications
- Volatility of the web, with autonomously managed data sources and services
- The need to transform data into standard agreed-upon forms suitable for spatial representation and analysis

The mechanisms proposed within CI projects to address these challenges follow the idea of *resource virtualization*, that is, decoupling information sources and services from their specific implementation, geographic location, or physical configuration. Such resources can be pooled together to address computation- or data-intensive problems. For spatial data, this typically involves:

- Standardization of spatial metadata, specifically following FGDC Content Standard for Digital Geospatial Metadata, ISO 19115 and 19139, and standards-compliant registration of different types of spatial data and services to catalogs that can be accessed and queries in a standard fashion