Visualization of Disease Surveillance Data with Geostatistics

Charles M. Croner, Ph.D.
Centers for Disease Control and Prevention, Hyattsville, MD, USA
and
Lee De Cola, Ph.D.
U.S. Geological Survey, Reston, VA, USA

Abstract: Kriging techniques are used to map national public health data routinely reported to the Centers for Disease Control and Prevention (CDC). Although choropleth maps are commonly used to display geographic and temporal disease-related events for political (state and county) jurisdictions, kriged maps may offer opportunities to enhance visual communication of event patterns over time. As a geostatistical modeling technique, kriging takes into account the existing underlying spatial structure of georeferenced information to produce statistically robust estimates of the pattern and standard errors of an underlying phenomenon. This geostatistical technique is applied to model estimates of reported cases by state of Lyme disease.

Keywords: kriging, GIS, disease surveillance, visualization

1. Introduction

1.1 The purpose of this exploratory research is to consider the use of kriging in mapping routinely-reported national public health data. Traditionally, choropleth or area maps are commonly used to display geographic and temporal epidemiologic events. Kriged maps may strengthen our ability to visually communicate event patterns, especially over time. As a geostatistical modeling technique, kriging takes into account the existing underlying spatial structure of georeferenced information (distances among samples or observations). Statistically optimal estimates and their standard errors for locations with missing data may be derived and the actual and estimated data represented as a smoothed surface (Cliff and Ord, 1981; Anselin and Hudak, 1992; Getis and Ord, 1996) using Geographic Information Systems (GIS).

1.2 Kriging is well associated with the earth sciences particularly in studies of minerals, soils and hydrology (Journel and Huijbregts, 1978; Delhomme, 1978; David, 1977; Ver Hoef and Cressie, 1993). The growing availability of computational tools for spatio-temporal analysis (Venables and Ripley, 1994; Kaluzny et al, 1997) has contributed, as well, to the use of kriging in other disciplines. For example, in the pharmaceutical sciences, flow properties of oil-in-water emulsion gels employ kriging to describe summarized response surfaces of aqueous thickeners (Marquardt and Sucker, 1998); in precision farming, kriging generates *k* classes of site membership values from multivariate and remotely sensed data sets to derive well-defined maps for improved crop yield and response (Burrough and Swindell, 1997); in biology, global block kriging is used to estimate vertical and horizontal distributions of zooplankton populations (Kern and Coyle, 2000); and, in ecology, the study of disturbed ecosystems in Russia employed three-dimensional universal kriging based on sequential aerial and space surveys (Vinogradov et al, 2000).

1.3 Increases in applications also are noted in the health sciences. In disease epidemiology and public health, uses of kriging have emerged in the last decade or so including the mapping of influenza-like illness in France (Carrat and Valleron, 1992); the mapping of

peak weeks of rotavirus detection in the United States (Torok et al, 1997); kriging of regression residuals to model local variation of malaria risk (Kleinschmidt et al, 2000); kriging of incidence rates of rare diseases such as childhood cancers in England (Oliver et al, 1994); and, use of punctual kriging of addressed-matched vital statistics in the mapping of spatial patterns of infant mortality and birth defect rates in Iowa (Rushton et al, 1996).

1.4 Kriging applications also are associated with a variety of recent cross-disciplinary studies involving human, or other organismic, health risk and the environment. In the example of human and ecological risk assessment, plutonium contamination of soils around the Rocky Flats Environmental & Technology Site were modeled with the technique of nonparametric indicator kriging to assess risk related to residential occupancy (Litaor et al, 1995); in the study of air quality, suspected association between ambient levels of air pollutants and pediatric asthma employed universal kriging for spatial interpolation of aerometric monitoring station data (Mulholland et al, 1998).

2. Approach

Our approach to employ kriging on routinely reported health data uses the example of Lyme disease by state from the CDC data base. Computations are based on S-Plus software for modeling the spatial structure or spatial correlation of the data. The model is developed through the variogram plot of semivariance by describing how sample data are related to distance and direction. The modeled data results in a kriged national forecast map, displayed with ArcView software in a Geographic Information System (GIS).

3. Transition from Choropleth Disease Forecast to Spatial Structure

3.1 Annual reported cases of Lyme disease by state for the period 1990-1999 were entered in an Excel spreadsheet. Centroid locations e.g., x, y coordinates of latitude and longitude, also were merged with the respective state data. The data are representative of the 48 conterminous “states” of the U.S. and Washington, D.C. (Alaska and Hawaii are not included). These data are the basis for modeling the temporal trend of Lyme disease.

3.2 A 10-year case time series for each state was used to model incidence forward to establish a stable forecast for 2002 (Fig. 1). The method of autoregressive integrated

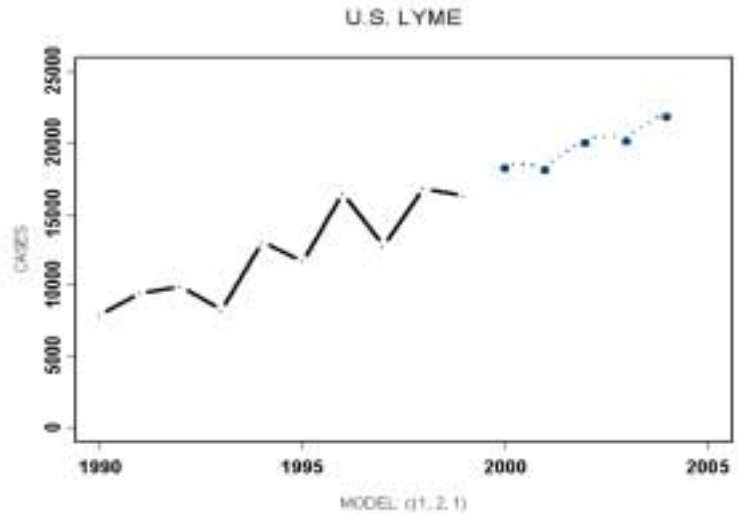

Figure 1: Lyme disease case time series

moving average (ARIMA) was compared with, and selected over, a linear statistical model. Whereas the linear procedure models numbers as independent variables, the ARIMA procedure accounts for temporal considerations and maintains the integrity of the data sequence. The sensitivity of the later was reflected by some cyclical appearance of the disease incidence, both nationally and in the regions (not shown).

3.3 Disease incidence is routinely presented by shaded polygons or a choropleth design (Fig. 2). Each of these state polygons contains a numerical value of Lyme disease incidence represented as an area or shaded value within the national framework. Visual communication of disease risk is oversimplified since all values appear evenly distributed within a polygon. Additionally, while Lyme incidence may be considered a continuous variable, values among contiguous states in a choropleth map can differ abruptly at adjoining borders. Other studied limitations of this design include the visual dominance of larger states over smaller ones, and often the absence of some measure of statistical reliability associated with reported incidence. A logarithmic (log base 10) transformation of reported incidence was done to normalize variability of state data and map disease risk.

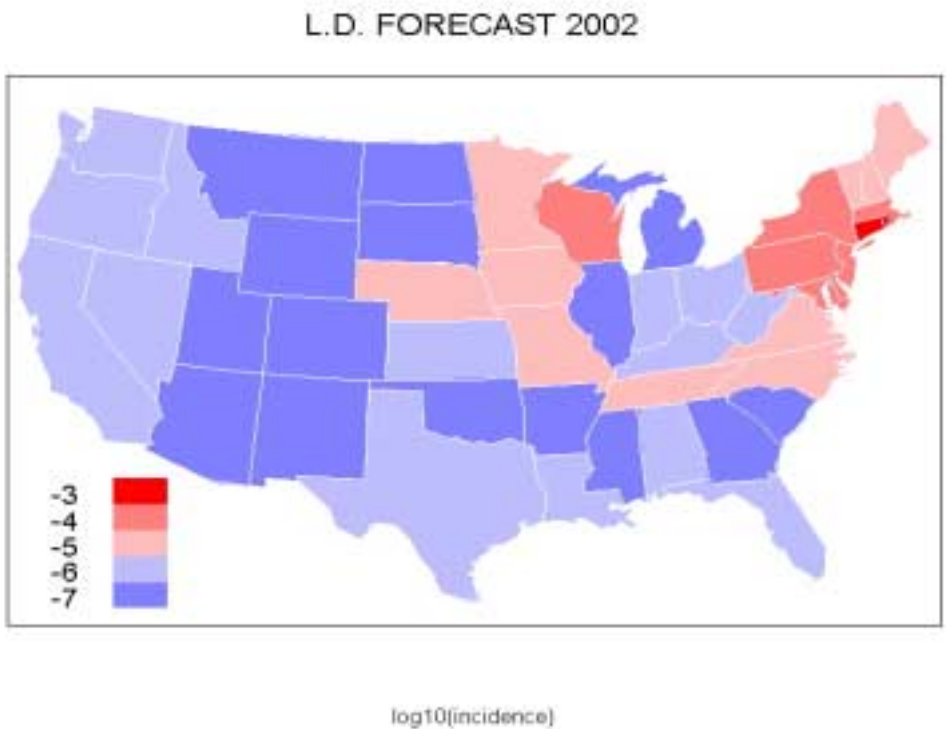

Figure 2: Choropleth design and log incidence

3.4 As a transition from a choropleth design to a “surface” more representative of existing correlation between pairs of points, the centroids of the 2002 forecast were statistically modeled. The resulting “best fit” model of global or large-scale variation is a 2-degree generalized least squares polynomial and predicted surface (Fig. 3). This polynomial provides the “best fit” in terms of least distortion of pattern in the spatial structure of the data. Here an expected elevation towards the northeastern U.S. is observed.

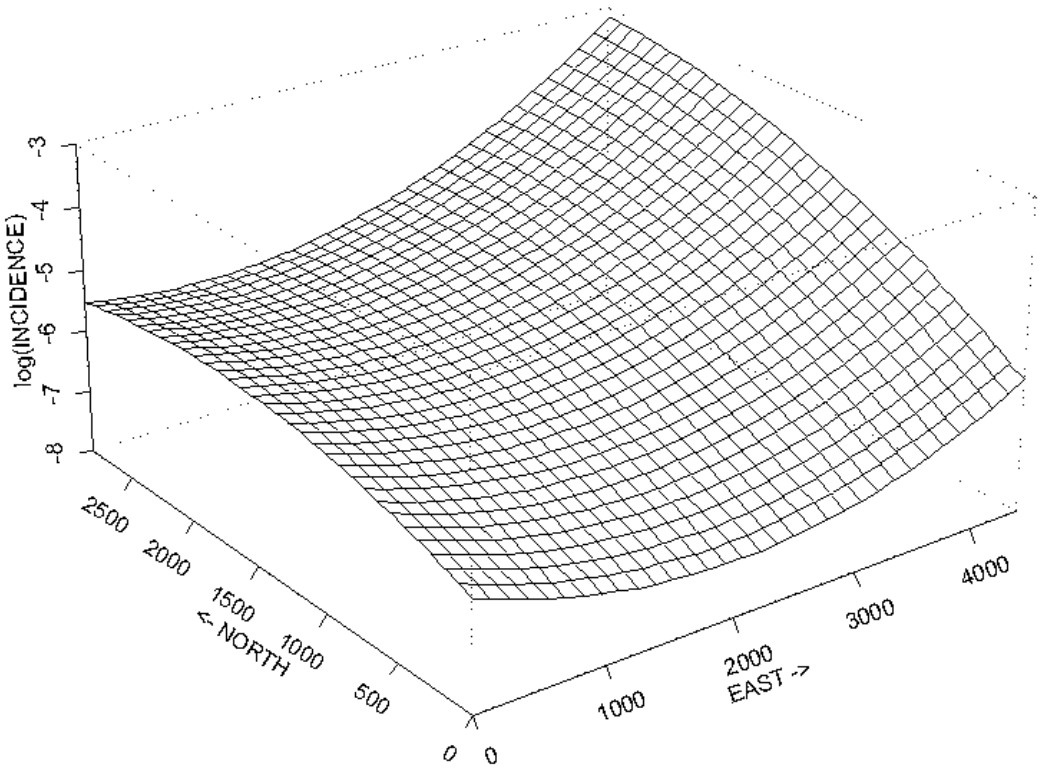

Figure 3: Polynomial trend surface

4. Kriged Spatial Structure and Mapped Risk

4.1 In the next step, state-to-state or local variation is modeled to produce the kriged surface (Mathematica 3.0 software). In the kriging process, correlations between states were derived at certain bands of distance through a correlogram and the correlation structure was used to make choices for kriging the 49 data points to a surface. This procedure looks at correlations in values between states and provides a covariance function to determine the distance and shape (spherical) of the spatial autocorrelation. For Lyme disease, those states closest in distance to each other were found to have the highest correlated values (reflective of Tobler's first law of geography: Tobler 1970). Conversely, those states least correlated in value tend to be furthest apart. This modeled covariation is taken into account to make kriging choices at varying spatial scales. For this covariance model, a distance parameter of about 500 kilometers around each data point was selected for kriging. Values were then generated at each node on a regularly spaced grid superimposed over the U.S. For example, a 32x32 grid (or approximately 1,000 grid cells) results in a more definitive surface (Fig. 4) than the previous trend polynomial while retaining the high values in the northeast.

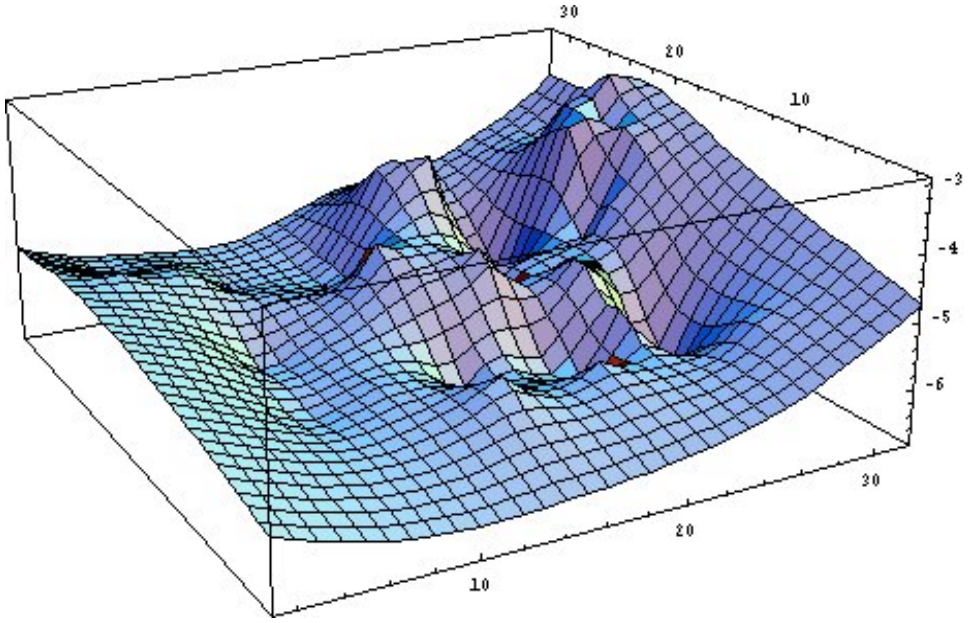

Figure 4: Kriged 32x32 grid surface

4.2 The spatial resolution of the grid set at 256x256 (or approximately 65,000 cells) results in a highly definitive surface (not shown) expressed as an isarithmic map of Lyme disease risk (Fig. 5).This smoothed or contour map corresponds to the original state data

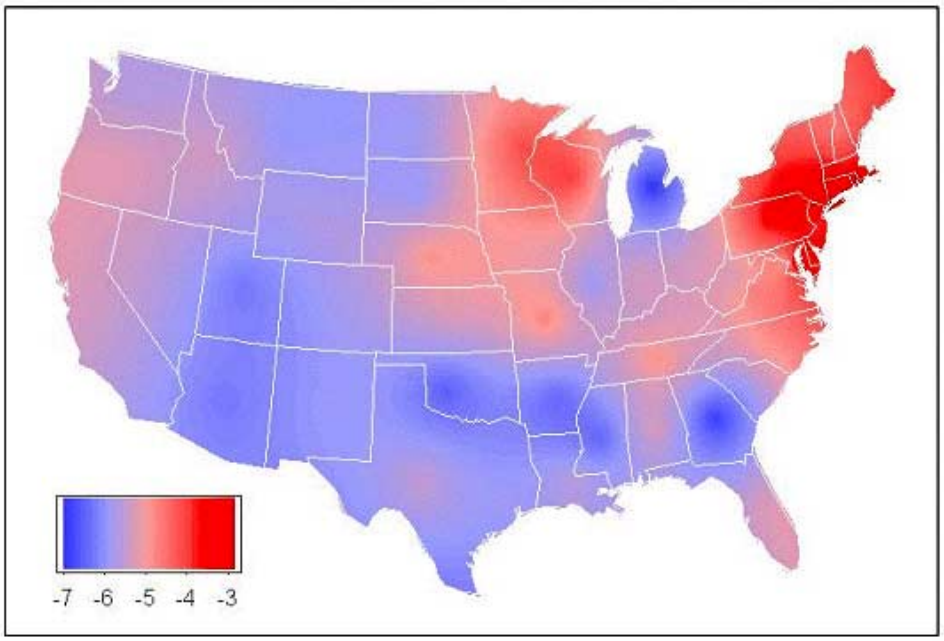

Figure 5: Kriged and smoothed Lyme disease risk map

and reflects unbiased estimates with least error taking into account existing spatial autocorrelation.

5. Discussion

5.1 A variety of discussion points are raised in this first attempt to model and forecast the underlying spatial structure of reported disease incidence, through the geostatistical procedure of kriging. These include the selection of spatial scale and resolution, estimates of error variances, kriging other potential covariates of disease risk, comparison with other methods of smoothing, and kriging as a potential visual or cognitive tool for communicating spatiotemporal change.

5.2 It should be noted while kriging provides best unbiased estimates mentioned earlier, the use of state data provides more of a test of procedures, but not the best choice for analysis. Centroids of variable and large size states can mask localized clusters of disease occurrence and therefore bias true distance relationships of disease occurrence among states. The next step is to use county-level data which will result in larger scale spatial resolution and improved measurement of disease distance relationships.

5.3 A statistical strength of kriging in the analysis of the spatial structure of data is the procedure yields estimates of standard errors or error variances for expected cases. The plan is to develop the error surface much like that shown in the kriged incidence map, for both state and county reported Lyme disease cases. One promising development has been recent work to incorporate both kriged incidence and uncertainty of Lyme disease into a single map display (De Cola, 2001).

5.4 The plan also is to move beyond kriging the spatial pattern of a single dependent variable and explore multivariate analyses (Hayes et al, 1999) of other possible predictor relationships with disease outcome. Georeferenced data including temperature and rainfall; remotely sensed soil moisture, land cover, and vegetation; and available vector distribution, abundance, tick infection prevalence and human exposure data (Fish and Howard, 1999) are potential associative covariates for study. Whether administrative, irregular grid, or point measurements, georeferenced data from these and other sources can be converted into a regular raster data structure (Burroughs and McDonnell, 1998) and kriged.

5.5 Kriging is one of several approaches to spatiotemporal analysis based on random function theory (Atkinson and Tate, 2000). Theoretically it appears to be the best grid generation method for estimation of a mapped surface. Among other spatiotemporal methods of analysis, Bayesian smoothing techniques may be the more robust (Bernardinelli and Montomoli, 1992). Random function theory recently has been combined with Bayesian maximum entropy (BME) methods (Christakos and Serre, 2000). A comparison of these methods should be incorporated into future research.

5.6 The kriging method, compared to traditional map design, may offer a means to enhance or supplement the ability to visually communicate temporal disease patterns. Unlike choropleth maps, kriging is reflective of tracking an underlying disease process. The kriging design, possibly through a growing suite of sophisticated computer animation tools, may offer improved readability of this process.

Although the choropleth map tends to spatially complicate the measurement of an underlying disease process, in its defense the design is “easy” or straightforward to construct and comprehend. This design remains the method of choice to communicate estimated spatial density of reported disease incidence. The use of geostatistics requires more complex computational choices such as solving for a large set of simultaneous equations at every grid node, prior study of the data to test for stationarity, determine the form of the variogram, set neighborhood size and other considerations. The next step in this research is to explore differences in choropleth and kriged county maps of reported Lyme disease to determine if there is, in fact, a need to model the spatial structure of this disease.

References

-Getis A, Ord JK 1996 Local spatial statistics: an overview. In Longley P, Batty M (eds) Spatial analysis: modelling in a GIS environment. Cambridge (UK), GeoInformation International: 269-85.
-Anselin L, Hudak S 1992 Spatial econometrics in practice: a review of software options. Regional Science and Urban Economics 22:509-536.
-Cliff AD, Ord JK 1981 Spatial process: models and applications. London, Pion.
-Journel AG, Huijbregts CJ 1978 Mining geostatistics. London, Academic Press.
-Delhomme JP 1978 Kriging in the hydrosciences. *Advances in Water Resources* 1(5):251-266.
-David M 1977 Geostatistical ore reserve estimation. Amsterdam, Elsevier.
-Marquardt D, Sucker H 1998 Oil-in water-emulsion gels: determination and mathematical treatment of flow properties. European Journal of Pharmaceutics and Biopharmaceutics 46(1):115-24.
-Burrough PA, Swindell J 1997 Optimal mapping of site-specific multivariate soil properties. Ciba Foundation Symposium. Utrecht, The Netherlands: 219-20.
-Kern JW, Coyle KO 2000 Global block kriging to estimate biomass from acoustic surveysfor zooplankton in the western Aleutian Island. Canadian Journal of Fisheries and Aquatic Sciences 57(10):212-21.
-Andre B, Trochu F, Dansereau J 1996 Approach for the smoothing of three-dimensional reconstructions of the human spine using dual kriging interpolation. Medical and Biological Engineering and Computing 34(3):185-91.
-Carrat F, Valleron AJ 1992 Epidemiologic mapping using the “kriging” method: application to an influenza-like illness epidemic in France. American Journal of Epidemiology 135(11):1293-1300.
-Torok TJ, Kilgore PE, Clarke MJ, Holman RC, Bresee JS, Glass 1997 Visualizing geographic and temporal trends in rotavirus activity in the United States, 1991 to 1996. Pediatric Infectious Disease Journal 16(10):941-46.
-Kleinschmidt I, Bagayoko M, Clarke GPY, Craig M, Le Sueur D 2000 A spatial statistical approach to malaria mapping. International Journal of Epidemiology 29(2):355-361.
-Oliver WR, Muir KR, Mann JR 1998 Binomial cokriging for estimating and mapping the risk of childhood cancer. IMA Journal of Mathematics Applied in Medicine and Biology 15(3):279-97.
-Rushton G, Krishnamurthy R, Krishnamurti D, Lolonis P, Song H The spatial relationship between infant mortality and birth defect rates in a US city. Statistics in Medicine 15(17-18):1907-19.

-Venables, WN, Ripley BD 1994 Modern applied statistics with S-Plus. New York, Springer-Verlag. -Kaluzny SP, Vega SC, Cardoso TP, Shelly AA 1997 S+SpatialStats: users manual for Windows and UNIX. MathSoft, Inc. New York, Springer-Verlag.
-Georgopoulos PG, Purushothaman V, Chiou R 1977 Comparative evaluation of methods for estimating potential human exposure to ozone: photochemical modeling and ambient monitoring. Journal of Exposure Analysis and Environmental Epidemiology 7(2):191-215.
-Vinogradov BV, Koshel SM, Kulik KN, Musin OR, Fedotov PB 2000 Prognosis of the spatiotemporal dynamics of disturbed ecosystems using three-dimensional universal kriging based on sequential aerial and space surveys. Doklady Biological Sciences 371(1-6):147-51.
-Kawaoka PY, Yamashiro SM 1996 Separation of cardiac conduction system and atrial activities by spatial regression. Medical Engineering & Physics 18(1):45-50.
-Trochu F, Brailovski V, Meunier MA, Terriault P, Qian YY 1996 Thermomechanical analysis of shape memory devices. Bio-Medical Materials and Engineering 6(6):389-403.
-Stytz MR, Parrott RW 1993 Using kriging for 3D medical imaging. Computerized Medical Imaging and Graphics 17(6):421-42.
-Litaor MI, Ellerbroek D, Allen L, Dovala E 1995 Comprehensive appraisal of 239+240Pu in soils around Rocky Flats, Colorado. Health Physics 69(6):923-35.
-Shinn NJ, Bing-Canar J, Cailas M, Peneff N, Binns HJ 2000 Determination of spatial continuity of soil lead levels in an urban residential neighborhood. Environmental Research 82(1):46-52.
-Mulholland JA, Butler AJ, Wilkinson JG, Russell AG, Tolbert PE 1998 Temporal and spatial distributions of ozone in Atlanta: regulatory and epidemiologic implications. Journal of the Air & Waste Management Association 48(5):418-26.
-Tobler WR 1970 A computer movie simulating urban growth in the Detroit region. Economic Geography 46:234-40.
-Carlon C, Critto A, Marcomini A, Nathanail P 2001 Risk based characterisation of contaminated industrial site using multivariate and geostatistical tools. Environmental Pollution 111(3):417-427.
-White DL, Bushek D, Porter DE, Edwards D 1998 Geographic information systems (GIS) and kriging: analysis of the spatial and temporal distributions of the oyster pathogen Perkinsus marinus in a developed and an undeveloped estuary. Journal of Shellfish Research 17(5): 1473-1476.
-Gronskei KE, Bartonova A, Brechler J, Walker SE, Larssen S 2000 Air pollution exposure monitoring and estimation VII. Estimation of population exposure in a central European airshed. Journal of Environmental Monitoring 2(4): 344-50.
-Stein A, Turkman KF, Bermudez P, van Heerd R, de Bruijn P 1999 In search of spatial extremes. Novartis Foundation Symposium. Wageningen Agricultural University, The Netherlands:2208-26;discussion 26-32, 52-8.
-Kinney PL, Aggarwal M, Nikiforov SV, Nadas A 1998 Methods development for epidemiologic investigations of the health effects of prolonged ozone exposure. Part III. An approach to retrospective estimation of lifetime ozone exposure using a questionnaire and ambient monitoring data (U.S. sites). Research Report (Health Effects Institute) (81):79-108.
-Ver Hoef JMV, Cressie N 1993 Multivariable spatial prediction. Mathematical Geology 25: (2) 219-240.
-Hayes EB, Fish D, Howard CA, Orloski KA, Dennis DT, Pieman J Geographical distribution of Lyme disease risk in the United States, VIII International Conference on Lyme Borreolosis and Tick-Borne Diseases, Munich, June 1999.

-Fish D, Howard CA 1999 Appendix: methods used for creating a national Lyme disease map. Centers for Disease control and Prevention. Morbidity and Mortality Weekly Report (June 4) 48(RR-7):19-25.
-Burroughs P, McDonnell R 1998 Principles of geographic information systems. Oxford University Press, New York.
-Atkinson PM, Tate NJ 2000 Spatial scale problems and geostatistical solutions: a review. The Professional Geographer 52(4):607-623.
-Bernardinelli L and Montomoli C 1992 Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. Statistics in Medicine 11:983-1007.
-Christakos G, Serre ML 2000 Spatiotemporal analysis of environmental exposure-health effect associations. Journal of Exposure Analysis and Environmental Epidemiology 10(2):168-197.