the venue for interaction a virtual online data super-store. Additionally, the accessibility to data is no longer tied to ownership or control of the acquiring technology, subject to considerable government restriction, or restricted to only those entities that can invest in considerable computer storage. In recent years, several private companies have built the requisite infrastructure to make data more accessible to a broader audience near-real-time via the internet. Advances in associated standards, such as ISO JPEG 2000 (JP2) and Geography Markup Language (GML), are integral to these market advances, as well as the value-added tools and technologies that are now being designed to facilitate true interoperability.
Clearly, there are geopolitical issues helping to drive the current business of earth observation. International tensions, especially since the 9/11 terrorist attacks on the World Trade Center, have certainly kept defense and security high on the priority list of most nations. Combine defense with natural disasters, such as the 2004 Sumatran Tsunami, and increased attention on global warming and there is an increased demand for four-dimensional high-resolution remote sensing of our environment. Furthermore, in addition to the traditional "remote" platforms – satellite and aerial – individual citizens have become another earth observing-platform generating real-time views of what exists at any given moment in any given place. So, the final trend in earth observation is this ubiquity of data; the push of timely, high-resolution data to any device, anytime, anywhere.

Cross References

► Bayesian Spatial Regression for Multi-source Predictive Mapping
► Intelligence, Geospatial
► Intergraph: Real Time Operational Geospatial Applications
► Photogrammetric Applications
► Standards, Critical Evaluation of Remote Sensing
► Temporal GIS and Applications

Recommended Reading

Avery, T.E., Berlin, G.L.: Interpretation of Aerial Photographs. 4th edn. Burgess Publishing, Minneapolis, MN (1985)
Baker, J.C., O'Connell, K.M., Williamson, R.A. (eds.): Commercial Observation Satellites: At the Leading Edge of Global Transparency. RAND-ASPRS, Santa Monica, CA (2001)
Jensen, J.R.: Remote Sensing of the Environment: An Earth Resource Perspective. Prentice-Hall, Upper Saddle River (2000)
Stoney, W.E.: Markets and Opportunities. Earth Imag. J. **2**(1), 10–14 (2005)

Evolution of GIS and LBS

► Computer Environments for GIS and CAD

Evolutionary Algorithms

► Geographic Dynamics, Visualization And Modeling

Evolving Spatial Patterns

► Patterns in Spatio-temporal Data

Exchange, Data

► OGC's Open Standards for Geospatial Interoperability

Exchange Format

► Geography Markup Language (GML)

Exploratory Cartography

► Exploratory Visualization

Exploratory Data Analysis

► Movement Patterns in Spatio-temporal Data

Exploratory Spatial Analysis

► Geographic Knowledge Discovery

Exploratory Spatial Analysis in Disease Ecology

LANCE A. WALLER
Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, GA, USA

Synonyms

Conservation medicine; Epidemiology, spatial; Epidemiology, landscape

Definition

Disease ecology represents an intersection of population ecology, community ecology, epidemiology and environmental health focused on the study of (typically infectious) disease within a population of individuals. Analysis often focuses on the interactions of at-risk individuals with each other and with elements of their environment. Accordingly, spatial factors often play key roles in disease ecology due to the influence of heterogeneous geographic distributions (e. g., host distributions, vector abundance, and landscape barriers to contact) on the spread of an infectious disease.

The phrase "exploratory spatial analysis" encompasses a wide range of goals and techniques for investigating patterns and process within georeferenced data. The phrase intersects other phrases in the literature and merits brief discussion. Bailey and Gatrell (1995) distinguish between "spatial analysis" and "spatial data analysis" with the former referring to the exploration of spatial association through basic geographic information system operations such as layering, buffering, and linking databases via spatial location, and the latter referring to statistical analysis of spatially referenced observations. In addition, Haining (2003, Chapter 5) provides a detailed conceptual framework for "exploratory spatial data analysis" (ESDA), the primary goals of which are to summarize spatial properties and patterns and to formulate hypotheses from geographically referenced data using methods making minimal data assumptions and which are numerically and graphically resistant to the impact of isolated outlying observations. Such subtle distinctions have their place in the categorization of analytic techniques within geographic information science, but in the sections below, all of these components are stressed in the setting of disease ecology. More specifically, modern analysis of disease ecology often involves a synergistic interaction between geographic information systems, mathematical models of population dynamics, and statistical analysis of reported public health data. As a result, visual and quantitative summaries of complex spatial and spatio-temporal patterns play a key role in the study of disease ecology.

Historical Background

The end goal of most geographic analysis in disease ecology is the identification of the locations where disease cases are likely to occur and the factors driving these patterns. Meeting this goal requires assessment of the modes of transmission of the disease, as well as the interactions between various hosts, vectors, and the landscape which they inhabit. Most ecologists take an a priori approach, seeking mathematical models of the underlying dynamics that produce patterns mirroring those observed in data (Hudson et al. 2001, Collinge and Ray 2006). In contrast, many public health researchers take an a posteriori approach using statistical techniques to quantify patterns and associations within observational data sets (Halloran 1998). The two epistemological approaches have coexisted at least since the ground-breaking work in both paradigms of Sir Ronald Ross regarding infectious disease in general and malaria in particular in the early 1900's (Ross 1905). However, there has been relatively little cross-fertilization between the two until relatively recently (Hilborn and Mangel 1997, Halloran 1998, Burnham and Anderson 2002).

The growing availability of georeferenced health and environmental data has led to an increased use of geographic information science within disease ecology and often provides a meeting ground between the a priori and a posteriori points of view, further facilitating a more holistic strategy in quantitative methods for disease ecology and related fields such as conservation medicine (Aguirre 2002) and spatial epidemiology (Ostfeld et al. 2005).

In particular, geographic information systems and science provide multiple mechanisms for exploratory spatial analysis in disease ecology. First, basic layering and buffering operations provide insight into the construction of mathematical models by revealing geographic associations between disease incidence, land forms, populations, and local habitats. Second, geographic display of model-based predictions links local observations with model output for assessing fit. Finally, and perhaps most importantly, the geographic display of model output allows statistical a posteriori assessment of model-driven a priori associations by matching spatial patterns arising within the modeled and observed outcomes.

Scientific Fundamentals

The basic components available for spatial analysis in disease ecology involve quantitatively linking the elements of dynamic models of disease transmission and associated measurable data. Beginning with dynamic models of disease transmission, disease ecology typically involves a combination of models of infectious disease transmissions and population dynamics. Halloran (1998) provides a thorough description of what typically are referred to as "SIR" or "SEIR" models where individuals within the population are considered to move between several states: "S" defines those individuals who are not yet infected but are susceptible to infection, "I" defines individuals who are infected (and perhaps, but not always, infectious), and "R" defines individuals who are "recovered" or "removed" through death or immunity. SEIR models incorporate an additional state ("E") for individuals exposed to the pathogen but not yet infected or infectious. One next defines a model, typically based on differential equations, defining the rates of transition from state to state. Such models have seen wide popularity and are used to model a variety of diseases in a variety of settings.

Such models provide theoretical insight into the progression of disease outbreaks and offer testbeds for various intervention scenarios (e. g., does vaccinating children or older adults have a greater impact on the final number of individuals infected during an influenza outbreak?). However, many models are parameterized based on quantities of biological interest (e. g., the basic reproductive number defined as the average number of cases a single infectious individual would infect in a totally susceptible population), or simplifying assumptions (e. g., randomly mixing populations of individuals) that limit connection to data observed in local settings. Halloran (1998) nicely illustrates how the elegant mathematical structure of such models often clashes with the realities of observed data introducing complications in statistical analysis. Moving to a geographical setting raises even more complications such as spatial heterogeneities in population density, interaction rates between "S"s and "I"s, and the influence of local environmental factors.

The statistical analysis of disease ecology data is complicated by the often non-linear models driving disease dynamics and the correlations between disease events induced by the infectious nature of the disease in question. Such correlations arise during the course of an epidemic as the disease passes from individual to individual in space and time, perhaps through intermediate hosts such as vectors or reservoir species.

Recent developments in statistical methodology include the development of hierarchical models allowing the analyst to incorporate uncertainty both in the observations and in the underlying model driving observed dynamics. The general conceptual structure of the hierarchical approach involves three primary stages of the underlying stochastic model: the data model, the process model, and the parameter model. To begin, the data model is a probabilistic description of the data given the underlying process and a set of parameters, i. e., if the underlying dynamic model was known, how would observations arise? Next, it is necessary to define the process model, that is, the set of possible underlying dynamic processes of interest given a set of model parameters (e. g., local transition rates between disease states and the basic reproductive number associated with the disease of interest). Finally, in the Bayesian framework, it is necessary to define prior distributions for each of the data and process parameters. The overall analytic frameworks explores the posterior distribution of these parameters which is proportional to the product of the three hierarchical stages, i. e., [data | process, data parameters]* [process | process parameters]*[data parameters, process parameters] using the bracket notation to denote any general probability distribution and | to denote conditional probabilities. The hierarchical structure not only links process and data models, but also provides mechanisms for incorporating spatial and spatio-temporal correlations within the prior distributions for data and process parameters. Wikle (2003) provides an excellent introductory review of such models and their applications in ecology and climatology.

The hierarchical framework has some conceptual appeal for its linkage of mathematical and statistical modeling, but this appeal can come at a fairly high computational cost. Implementation typically relies on computationally-intensive Markov chain Monte Carlo sampling of the posterior distribution through iterative updates of model parameters. While increasingly common, such approaches are not yet a standard component in statistical or geographic information system software, thereby requiring expertise spanning geographic information science, mathematical modeling, and statistical computing, often best met through interdisciplinary teamwork.

Key Applications

To date, key applications in exploratory spatial analysis of disease ecology data fall primarily into one of the three areas mentioned above. At the risk of oversimplification, it is possible to categorize general trends in disease ecology applications from each of the three classes of analysis. Most geographic information system-based strategies rely on remote sensing data such as precipitation or vegetation cover to define potential habitats for disease vectors or reservoir species, then layer human residences or activity patterns in order to define "risk maps" of areas with high and low propensity for vector-to-human transmission. In contrast, most approaches based on statistical analysis of public health reporting data focus on reported human cases and seek to quantify associations with local factors, some perhaps involving layered remote sensing data. Finally, most spatial ecology approaches rely on stochastic or deterministic models of dynamic spread to predict case locations, then compare these predicted patterns to those observed in the public health data. The analytic stereotypes listed here reveal both the overlap in goals and data, but also highlight the different emphasis placed on different data and modeling pieces by the three different disciplinary approaches.

More recent key efforts begin to lean on more than one framework and brief examples are provided here. Brownstein et al. (2003) expand the risk map format above in an investigation of the spatial distribution of Lyme disease vectors in the United States. Their approach incorporates local climate information based on statistical prediction from fixed sampling stations and links the predictions to observed vector abundance through the use of logistic regression models incorporating spatial correlation. This approach links state-of-the-art geographic information science with sophisticated spatial statistics but requires linking data between a geographic information system and multiple statistical packages in order to provide all of the computation necessary.

A series of papers regarding the spatio-temporal spread of raccoon rabies in the northeastern United States also provides examples of linking geographic information science, mathematical modeling, and spatial statistics. Smith et al. (2002) proposed a cellular automata model of the spread of raccoon rabies across the state of Connecticut. In follow-up work, Waller et al. (2003) use spatial statistics to assess the fit of the mathematical model a posteriori, not only comparing predicted and observed outcomes, but also comparing numerical summaries of the spatial patterns generated by various competing models (e. g., including or not including a delay in spread associated with crossing a river). Finally, Russell et al. (2004) utilized the model of Smith et al. (2002) to provide a priori predictions of the spread of the outbreak through the state of New York.

Future Directions

Spatial analysis in disease ecology provides a meeting ground for individuals from multiple disciplinary backgrounds and, as the illustrations cited above reveal, research in the area increasingly borrows across disciplinary borders. Such interactions provide broader insight and a larger set of analytic tools to address the underlying questions of interest but at the same time require methodological and computational flexibility to make best use of the available methods, models, and data to meet this goal.

Much future research remains in order to provide a comprehensive set of analytic tools providing quantitative links between landscape patterns and disease incidence and prevalence in space and time. Current methods often rely primarily on geographical information systems-based overlays of landscape and health data, mathematical models of dynamics in space and time, or statistical summaries from regression-type models of association. In order to best address the issues raised above, new approaches are required which utilize elements of all three paradigms in order to fully understand the forces and influences driving observed patterns in emerging diseases as well as to suggest and evaluate potential public health intervention strategies.

Cross References

► Data Analysis, Spatial
► Exploratory Visualization
► Geographic Dynamics, Visualization And Modeling
► Hierarchical Spatial Models
► Patterns in Spatio-temporal Data
► Processes and Events
► Statistical Descriptions of Spatial Patterns

Recommended Reading

Aguirre, A.A., Ostfeld, R.S., Tabor, G.M., House, C., Pearl, M.C.: Conservation Medicine: Ecological Health in Practice. Oxford University Press, Oxford (2002)

Bailey, T.C., Gatrell, A.C.: Interactive Spatial Data Analysis. Longman Scientfic and Technical, Essex (1995)

Brownstein, J.S., Holford, T.R., Fish, D.: A climate-based model predicts the spatial distribution of the Lyme disease vector *Ixodes scapularis* in the United States. Envrion. Health Persp. **111**, 1152–1157 (2003)

Burnham, K.P., Anderson, D.R.: Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd edn. Springer, New York (2002)

Collinge, S.K., Ray, C. (eds.): Disease Ecology: Community Structure and Pathogen Dynamics. Oxford University Press, Oxford (2006)

Haining, R.: Spatial Data Analysis. Cambridge University Press, New York (2003)

Halloran, M.E.: Concepts in infectious disease epidemiology. In: Rothman, K.J., Greenland, S. (eds.) Modern Epidemiology, 2nd edn. Lippincott-Raven, Philadelphia (1998)

Hilborn, R., Mangel, M.: The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton (1997)

Hudson, P.J., Rizzoli, A., Grenfell, B.T., Heesterbeek, H., Dobson, A.P. (eds.): The Ecology of Wildlife Diseases. Oxford University Press, Oxford (2001)

Ostfeld, R., Glass, G., Keesing, F.: Spatial epidemiology: an emerging (or re-emerging) discipline. Trends Ecol. Evol. **20**, 328–336 (2005)

Ross, R.: A logical basis of the sanitary policy of mosquito reduction. Science **22**, 689–699 (1905)

Russell, C.A., Smith, D.A., Waller, L.A., Childs, J.F., Real, L.A.: *A priori* prediction of disease invasion dynamics in a novel environment. Proceedings of the Royal Society: Biol. Sci. **271**:1534, 21–25 (2004)

Smith, D., Lucey, B., Waller, L.A., Childs, J.E., Real, L.A.: Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. Proceedings of the National Academy of Sciences, USA. **99**, 3668–3672 (2002)

Waller, L.A., Smith, D., Real, L.A., Childs, J.: Monte Carlo assessments of goodness of fit for ecological simulation models. Ecol. Model. **164**, 49–63 (2003)

Wikle, C.: Hierarchical models in environmental science. Int. Stat. Rev. **71**, 181–199 (2003)