

UNCERTAINTY IN GEOGRAPHIC DATA AND GIS-BASED ANALYSES

Objective

As research priority calls for a systematic effort to advance our understanding of uncertainty associated with geographic data and how this uncertainty is propagated through data analyses based on geographic information systems (GIS). Such an understanding is vital to decision-making that relies on geographic data, such as might occur in such areas as environmental management, transportation planning, and national security. Thus, we must develop strategies for identifying, quantifying, tracking, reducing, and reporting (including visually representing) uncertainty in geographic data and GIS-based analyses, and we must develop a standardized means by which uncertainty can be addressed in daily GIS applications.

Background

Geographic data (used interchangeably with "spatial data" in this document) relate information about three attributes of a geographic feature: typology (the type of geographic feature), location, and spatial dependence (the spatial relationships with other features). For example, data about a forest can indicate the forest type and species combination (as typological attributes), the location and size of the forest (the locational attributes), and its proximity to surrounding landscape features (attributes related to spatial dependence). Because such attributes change over time, geographic data are very complex and difficult to manage.

Geographic data are essentially observations about geographic features or phenomena, referred to as "geographic reality". Geographic reality often cannot be measured exhaustively because it is nearly impossible to obtain measurements for every point across an entire landscape. Accurate measurements are also difficult to obtain because of continuous (slow or rapid) variation of the landscape over time and because of the limitations of instruments, financial budgets, and human capacity. Thus, when geographic data are developed, they are merely approximations of geographic reality. Therefore, a fundamental discrepancy exists between geographic data and the reality that they are intended to represent. This discrepancy, or uncertainty, is propagated through, and may be further amplified by, data management and analyses in a GIS environment. The basic GIS schemes (Coulter 1992) for representing geographic data are not dynamic but record only a static, invariable view of the world. They do not depict complex objects that consist of interacting parts, nor do they display variation at many levels of detail over space and over time. Thus, uncertainty must be recognized as a basic element in all GIS results.

Uncertainty analysis assesses the discrepancy between geographic data in GIS, and the geographic reality that the data are intended to represent. Uncertainty information associated with a geographic data set should be perceived as a map depicting varying degrees of uncertainty associated with each of the features or phenomena represented in the data set. Uncertainty and error differ in that uncertainty is a relative measure of the discrepancy while error tends to measure the value of the discrepancy (Goodchild et al. 1994, p.142). Because the true value for every geographic feature or phenomenon is rarely determinable and the exact value of this discrepancy cannot be obtained in most cases, uncertainty rather than error should be used to describe the quality of geographic data and GIS products. Uncertainty exists in all three components of

geographic data: in the typological attributes, the locational attributes, and attributes related to spatial dependence. Recognition of these forms of uncertainty is important because policymakers increasingly use geographic data and GIS techniques to support their policy decisions. Geographic data are often used under the assumption that they are free of errors. The beguiling attractiveness, the high aesthetic quality of cartographic products from GIS, and the analytical capability of GIS further contribute to an undue credibility, at times, of these products (Abler 1987, p. 305). Acceptance of the accuracy of these data is often not warranted (Goodchild and Gopal 1989 pp. xii-xiii). Policymakers who use error-laden data without consideration of their intrinsic uncertainty are likely to reach inappropriate decisions.

The UCGIS Approach

Uncertainty exists in every phase of the life cycle of geographic data (data collection, data representation, data analyses, and final results), and it transcends the boundaries of disciplines and organizations. Our proposed research involves an inter-institutional (UCGIS) research team consisting of domain experts, GIS experts (including spatial statisticians), application users (including decision makers), data producers, and GIS software vendors to plan and execute the following three research tasks:

1. To study and understand the mechanics of how uncertainty arises in geographic data and how it is propagated through GIS-based data analyses.
1. To develop techniques for reducing, quantifying, and visually representing uncertainty in geographic data and for analyzing and predicting the propagation of this uncertainty through GIS-based data analyses.
1. To participate in promoting and advising a testing institute for
 - o testing new methods of managing uncertainty in geographic data and GIS analyses;
 - o implementing strategies and methodologies for reducing, quantifying, tracking, and reporting uncertainty in GIS implementation, in geographic data collection and generation, and in spatial data standards and decision-making processes; and
 - o improving communications about findings on uncertainty throughout the entire GIS life cycle, from data collection and analysis through the decision-making processes.

Importance to National Research Needs

Increasingly, policy decision-making involves the use of geographic data and GIS techniques. For example, making detailed policy decisions on how to preserve the Florida Everglades calls for detailed environmental analyses using state-of-the-art GIS technology and geographic data about the Everglades. In locating and allocating urban resources (such as transportation planning, fire station location and fire truck routing, school zoning, etc.), decision makers often employ GIS techniques and geographic data. The reliability of the resulting decisions very much depends on the quality of the geographic data used in evaluating alternatives.

Concern about uncertainty in spatial data and analyses is not new, but systematic efforts to study the problem are much more recent. Data errors and uncertainties in GISs were identified in the research agenda of the National Center for Geographic Information and Analysis (NCGIA) as one

of the most important impediments to the successful implementation of GISs (NCGIA 1989). The center devoted its first research activity to improving the accuracy of spatial databases. Progress has been made in this field since then (Goodchild and Gopal 1989; Goodchild 1992; Mowrer et al. 1996). Despite recent progress, most research findings are applicable only to artificial or exhaustively well-known data sets, and much remains unknown. The current state of GIS technology in dealing with uncertainty falls short of the goals described by Goodchild (1993, p. 98): (1) each object in a GIS database would carry information describing its accuracy; (2) every operation or process within a GIS would track and report error; and (3) accuracy measures would be a standard feature of every product generated by a GIS. The research proposed here calls for a major systematic effort to attack this deficiency in GIS research.

Benefits

To society

A thorough understanding of uncertainty in spatial data and greater availability of tools for quantifying and visualizing this uncertainty will benefit society in several ways. Researchers will be able to judge the quality and proper uses of large amounts of existing spatial data. The tremendous amount of energy and money invested in collecting these data will then become more cost-effective.

As information about uncertainty in spatial data is improved and becomes more readily available, decision makers will be able to better evaluate risks associated with various policy alternatives.

To the development of geographic information science

The future well-being of geographic information science in society largely depends on the quality and the proper use of geographic data as a fuel. The proposed research will directly contribute to the understanding and the documentation of the quality and the proper use of geographic data and will therefore substantially enrich the general knowledge pool in geographic information science. ·) An inter-institutional and interdisciplinary approach (the UCGIS approach) will facilitate communication among the various parties dealing with spatial data. New GIS needs and demands can be identified, and advancements in the field can be quickly transferred into daily GIS operations. This interaction will certainly advance the healthy development of geographic information science as a field.

To the field of uncertainty research

A systematic effort addressing uncertainty in spatial data will help streamline existing research efforts, will enrich our knowledge about uncertainty associated with spatial data, and will invent and perfect the strategies for managing uncertainty in GIS analyses and in decision-making processes.

As existing research efforts are streamlined, isolated research findings can be transferred from laboratory settings into daily GIS operations, making individual investments more cost-effective.

Priority Areas for Research

Short term

- Develop extensions to existing models of spatial data representation to allow the representation of the spatial variation of uncertainty within a geographic data set. Under the new extensions, the spatial variation of uncertainty about geographic features or phenomena represented in a data set is a reported element of that data set.
- Develop and test techniques for measuring or estimating uncertainty associated with each geographic feature or phenomenon recorded in a data set.
- Study the effects of data processing on the uncertainty in a single geographic data set. We will develop methods to quantify these effects and integrate this uncertainty information in the data sets.
- Study and develop effective tools for presenting uncertainty in spatial data, particularly visualization techniques including visual and audio methods. Effective means of visual representation must be able to show the spatial variation in all three uncertainty components (typological, locational, spatial dependence).

Medium term

- Study and model the uncertainty that results from incompatibilities of spatial data from different scales. We will quantify and report this uncertainty as part of the uncertainty that is propagated through computer analyses.
- Conduct case studies to determine how best to link research findings on uncertainty from various stages of the GIS life cycle. We will streamline research findings and demonstrate how they can be used in daily GIS applications.
- Study how the uncertainty information can be properly used in decision making, particularly in terms of risk analyses. We will thoroughly examine the impact that communicating uncertainty has on decision-making.

Long term

- Develop schemes for representing geographic reality as a complex, dynamic whole. Current representation depicts merely a slice of geographic reality with respect to time, space, and geographic attributes. This simple, static representation is limited and is a major source of error in spatial data. We will develop more complex representation that will ultimately allow us to minimize uncertainty in spatial data.
- Establish a clearer understanding of the interaction between forms of uncertainty generated at each stage in the life cycle of GIS data. The most important aspect of the proposed research is to understand how uncertainty is generated by differences in data sources and by differences in data scales and how it is then propagated and amplified through spatial analyses. We plan to clarify the relationships between the various origins of uncertainty so that the uncertainty in the final GIS results can be estimated and reported more accurately.

References

- Abler, R. F., 1987. The National Science Foundation National Center for Geographic Information and Analysis. *International Journal of Geographical Information Systems*. 1:303-326.
- Couclelis, H., 1992. People manipulate objects (but cultivate fields): Beyond the raster-vector debate In: U. Frank, I. Campari and U. Formentini, editors, *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space (Lecture Notes in Computer Science Vol. 639)*. Berlin-Heidelberg:Springer-Verlag, pp. 65-77.
- Goodchild, M. F., 1992. Research initiative 1: Accuracy of spatial databases. *Closing Report*, Santa Barbara, CA: National Center for Geographical Information and Analysis, University of California.
- Goodchild, M. F., 1993. Data models and data quality: Problems and prospects. In: M. F. Goodchild, B. O. Parks, and L. T. Steyaert, editors, *Environmental Modeling with GIS*. Oxford University Press: New York, pp. 94-103.
- Goodchild, M. F., and S. Gopal, editors, 1989. *Accuracy of Spatial Databases*. New York: Taylor and Francis.
- Goodchild, M. F., B. Buttenfield, and J. Wood, 1994. Introduction to visualizing data quality, In: Hilary M. Hearshaw and David J. Unwin, editors, *Visualization in Geographical Information Systems*. New York: John Wiley and Sons, pp. 141-149.
- Mowrer, H. Todd, R. L. Czaplewski, and R. H. Hamre, editors, 1996. Spatial accuracy assessment in natural resources and environmental sciences: Second International Symposium, U.S. Forest Service. Rocky Mountain and Range Experiment Station, Ft. Collins , CO. General Technical Report RM-GTR-277. 728 p.
- NCGIA (National Center for Geographic Information and Analysis), 1989. The research plan for the National Center for Geographic Information and Analysis. *International Journal of Geographical Information Systems*. 3(2):117-136.