ELSEVIER

Fuzzy Sets and Systems 113 (2000) 3–5

FUZZY
sets and systems

www.elsevier.com/locate/fss

Introduction: special issue on 'Uncertainty in geographic information systems'

Michael F. Goodchild *

National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA

Geographic information systems (GIS) manipulate digital representations of phenomena on the Earth's surface. Use has grown dramatically over the past two decades, along with the associated software industry, in response to the needs of utility companies to manage and maintain very large and complex networks of underground pipes, transformers, valves, poles, and other facilities. GIS are also used in environmental management, to inventory and manage productive forests, oil and gas fields, or agriculture. Thus a large industry has grown up around the production of software and data, and its use to solve a host of practical problems. A recent state-of-the-art review is provided by Longley et al. [6].

Information is geographic if it links instances of features, classes, measurements, events, transactions, or any other generalizable *thing* to some location at or near the surface of the Earth, and to some time. Symbolically, the primitive element of geographic information is the tuple $\langle x, z \rangle$, where x is a location in space–time and z is some thing. Examples include a measurement of air temperature at a weather station at a given time, or the presence of a particular type of vegetation at some location. It seems straightforward to imagine building a database of such tuples, such as occurs in the management of weather records. Problems arise, however, because the space–time frame is continuous, and any *complete* representation must therefore imply an infinite number of tuples.

Many clever methods have been devised for getting around the problem, all of them based in one way or another on the observation that virtually all geographic phenomena exhibit some degree of *spatial dependence* or *spatial autocorrelation*, meaning that the things present at one location can be guessed or imputed with some level of accuracy from knowledge of the same phenomena at nearby locations. For example, the field of atmospheric temperature exhibits strong spatial autocorrelation, allowing reasonable estimates of temperature anywhere to be made from observations at a surprisingly small number of weather stations. Spatial autocorrelation typically declines with distance, so the best estimates are made near weather stations. *Cross-correlation* is also common, allowing the presence of one thing to be used to infer the presence of others at the same location.

The various ways of exploiting auto- and cross-correlations provide the basis for a range of GIS *data models*. For example:

- Measurements are recorded at sample locations spaced distance S apart in a regular grid. This model (the *grid* model) is commonly used to represent the Earth's topographic surface in digital elevation models (DEMs).
- Measurements are recorded at irregularly-spaced sample points. This model is commonly used for weather data.

* Tel.: +1-805-893-8049; fax: +1-805-893-3146.
E-mail address: good@ncgia.ucsb.edu (M.F. Goodchild)

0165-0114/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0165-0114(99)00008-1

- The surface of the Earth is divided into square cells of side S. In each cell of this *raster* model an average value is recorded (remote sensing), or a class is identified (classified imagery).
- The surface of the Earth is partitioned into regions having approximately uniform value of the relevant class or measurement. This model (*polygon* or *area class* model) is commonly used for mapping vegetation, soils, or land ownership.
- The surface of the Earth is divided into a mesh of irregular triangles (the *triangulated irregular network* or TIN model). Each triangle models the land surface as a plane, and continuity of value is imposed across triangle edges.
- A set of points is identified as having the same *thing*. The set might be disjoint (point objects) or connected into lines or regions (line or area objects). Nothing is recorded for other points not in the set.

Some degree of approximation or generalization is involved in all of these cases. For the two cases associated with a well-defined value of S, or spatial resolution, there is a clear parameter associated with the approximation, representing the level of detail of the representation. In other cases, however, there is no clear parametrization of level of detail. In the TIN model, for example, the dimension of the smallest triangle merely provides an upper bound on S. S is roughly analogous to the *representative fraction* or scale used to define level of detail in the creation of paper maps, but the rigorous definition of this parameter for paper maps has no equivalent for digital data.

While this approach has obvious advantages in allowing an infinitely variable world to be represented in a digital store of finite size, accounting for the enormous interest in GIS, it inevitably raises problems of *uncertainty*, since the digital representation cannot be more than an approximation to the real geographical phenomenon. Only in a very few instances is it possible to remove uncertainty entirely. For example, by definition the set of points contained within the boundary of a land parcel belongs uniformly to the owner; and the lines connecting the corner points of such a parcel are by definition straight. Uncertainty arises in geographic data because it is impossible to measure location x without error given uncertainty in the definition of the Earth's spatial frame and the limitations of measuring instruments; because it is impossible to define many *things* in ways that are perfectly agreed and replicable among observers; because many spatial databases ignore the temporal dimension, and thus fail to address change; and because the *definitions* inherent in the data may not be passed successfully from the creator to the user.

The study of uncertainty in geographic data has ancient roots, but the stimulus provided by the rapid growth of GIS has produced an explosion of interest in recent years. Maling [7] and Goodchild and Gopal [2] provide early reviews of the issues. Research focuses on problems of defining, describing, and modeling uncertainty, its impact on the results of analysis [4], and its visualization and communication [3]. In addition, uncertainty is introduced into GIS through the imperfect specification of user queries, particularly when these are expressed in plain language using poorly-defined terms. Also, models of the physical processes affecting the Earth's surface, which are often used in conjunction with GIS to predict future landscapes, are themselves subject to uncertainty. Finally, uncertainty is introduced because of the varying interpretations attached to geographic data by its users, especially when these differ between participants in the same research project.

Early work in uncertainty in GIS relied heavily on traditional probabilistic frameworks, and made use of the foundations provided by geometric probability (e.g., [5]). But it became clear that uncertainty was so pervasive, and so problematic, that more general frameworks were needed, including fuzzy sets, Bayesian statistics, and subjective probability. Today, researchers in GIS uncertainty make use of all of these frameworks, despite their apparent incompatibilities, selecting the framework that best fits the problem at hand.

This special issue of *Fuzzy Sets and Systems* is devoted to a selection of cutting-edge papers on uncertainty in GIS. The papers range in style from conceptual discussion of the philosophical problems posed by vagueness, to fuzziness in reasoning with geographic data, to fuzziness in the assignment of locations to classes, to fuzzy definitions of object boundaries (and see [1]). Collectively, they illustrate the profound implications of the inherent uncertainty in geographic data, a feature that distinguishes geographic data from most of the information commonly processed today in digital systems.

At the same time, they raise substantial issues. One of the attractions of GIS is its simplicity; with GIS, everyone is empowered to manipulate geographic data, and to make good-looking maps, whether or not they possess qualifications in cartography or related fields. Geographic data are inherently uncertain, yet as these papers demonstrate, the analysis and manipulation of geographic data under uncertainty requires a high level of understanding of the properties of fuzzy sets, probabilities, and related theoretical frameworks. If research on uncertainty is to be successful, it must speak to the vast majority of GIS users who have no sophisticated quantitative or set-theoretic understanding.

A basic contention of fuzzy sets is that while it is not possible to assign a location x to a class z, it is nevertheless possible to define a membership level $p(z)$ in that class. In GIS, it is usually assumed that this is done because the class is not well-defined. Paradoxically, therefore, this approach requires us to believe that an observer who does not know exactly what is meant by a class z can nevertheless assign a numerical membership function to it, and that this membership function *is also meaningful to another observer, who also cannot define* z.

A second and possibly more serious problem arises when we consider typically applications of such data. For example, it is often useful to measure the total area assigned to class z. For crisp data, this is done by summing the area associated with the class. For fuzzy data, it might be done by summing the memberships. But in reality, the existence of class z at x is not independent of the existence of class z at a location some small distance away, say $x + \delta x$. Spatial dependence is clearly relevant in determining how much area is z, and associating levels of confidence with these estimates. Yet a simple assignment of membership to x falls short of characterizing the necessary dependencies.

In summary, uncertainty is endemic in geographic data. Fuzzy sets and systems provide a comprehensive framework for dealing with its impacts in analysis, reasoning, and modeling, as the papers in this special issue illustrate. But several very challenging problems remain, and much remains to be done before we have a comprehensive approach to uncertainty that is meaningful to the vast majority of GIS users. I look forward to much progress in the future.

References

[1] P.A. Burrough, A.U. Frank (Eds.), Geographic Objects with Indeterminate Boundaries, Taylor and Francis, London, 1996.

[2] M.F. Goodchild, S. Gopal (Eds.), Accuracy of Spatial Databases, Taylor and Francis, London, 1989.

[3] H.M. Hearnshaw, D.J. Unwin (Eds.), Visualization in Geographic Information Systems, Wiley, New York, 1994.

[4] G.B.M. Heuvelink, Error Propagation in Environmental Modelling with GIS, Taylor and Francis, London, 1998.

[5] M.G. Kendall, P.A.P. Moran, Geometric Probability, Hafner, New York, 1963.

[6] P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind (Eds.), Geographical Information Systems: Principles, Techniques, Management and Applications, Wiley, New York, 1998.

[7] D.H. Maling, Measurement from Maps: Principles and Methods of Cartometry, Pergamon, New York, 1989.