Measurement 81 (2016) 123–162

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

A practical taxonomy of methods and literature for managing uncertain spatial data in geographic information systems

Madjid Tavana [a,b,*], Weiru Liu [c], Paul Elmore [d], Frederick E. Petry [d], Brian S. Bourgeois [d]

[a] Business Systems and Analytics Department, Distinguished Chair of Business Analytics, La Salle University, Philadelphia, PA 19141, USA
[b] Business Information Systems Department, Faculty of Business Administration and Economics, University of Paderborn, D-33098 Paderborn, Germany
[c] School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, United Kingdom
[d] Geospatial Sciences and Technology Branch, Naval Research Laboratory, Stennis Space Center, USA

ARTICLE INFO

Article history:
Received 3 November 2015
Received in revised form 14 November 2015
Accepted 4 December 2015
Available online 14 December 2015

Keywords:
Uncertainty
Spatial data
Geographic information systems
Taxonomy
Literature review

ABSTRACT

Perfect information is seldom available to man or machines due to uncertainties inherent in real world problems. Uncertainties in geographic information systems (GIS) stem from either vague/ambiguous or imprecise/inaccurate/incomplete information and it is necessary for GIS to develop tools and techniques to manage these uncertainties. There is a widespread agreement in the GIS community that although GIS has the potential to support a wide range of spatial data analysis problems, this potential is often hindered by the lack of consistency and uniformity. Uncertainties come in many shapes and forms, and processing uncertain spatial data requires a practical taxonomy to aid decision makers in choosing the most suitable data modeling and analysis method. In this paper, we: (1) review important developments in handling uncertainties when working with spatial data and GIS applications; (2) propose a taxonomy of models for dealing with uncertainties in GIS; and (3) identify current challenges and future research directions in spatial data analysis and GIS for managing uncertainties.

1. Introduction

The modern geospatial revolution enhanced by geographic information systems (GIS) has greatly increased the understanding of our physical environment. The basic components of GIS include [26]: (1) a data input component for collecting and processing spatial data; (2) a data storage and retrieval component for organizing spatial data; (3) a data manipulation and analysis component for changing spatial data; and (4) a data reporting component for displaying spatial data. Spatial data are not always precise and uncertainty in geographical data is widely accepted due to the way the world is perceived, measured, and represented [51]. Varsi [40,41] has observed that vagueness is a major factor in geographical information representation since concepts such a river's length or a mountain's height in a specific area are uncertain as the specification of a river or peak are vague concepts. Baofu [2, p. 297] states "all geographical data are inherently inaccurate, and these inaccuracies will propagate through GIS operations in ways that are difficult to predict." Couclelis [10] further describes uncertainty as an inherent property of complex geospatial knowledge that must be managed effectively. Many of the problems associated with the accurate measurement of spatial databases and GIS are also prevalent in all types of database systems. Uncertainty

* Corresponding author at: Business Systems and Analytics Department, Distinguished Chair of Business Analytics, La Salle University, Philadelphia, PA 19141, USA. Tel.: +1 215 951 1129; fax: +1 267 295 2854.

E-mail addresses: tavana@lasalle.edu (M. Tavana), w.liu@qub.ac.uk (W. Liu), paul.elmore@nrlssc.navy.mil (P. Elmore), fred.petry@nrlssc.navy.mil (F.E. Petry), bsb2@nrlssc.navy.mil (B.S. Bourgeois).

URL: http://tavana.us/ (M. Tavana).

http://dx.doi.org/10.1016/j.measurement.2015.12.007
0263-2241/

in many of these systems is not simply an error or flaw to be reduced or eliminated but an important component of the system that must be taken into consideration. Therefore, uncertainty plays a critical role in the analysis of spatial data and GIS which contain descriptive as well as positional data. The uncertainty can be represented by a wide range of values that may include the actual measurement of the object as only one point. Fig. 1 illustrates the complexity that can be observed in a real-world example. This figure is an image of the Louisiana gulf coastal region in the area of the Atchafalaya Bay and illustrates the difficulty of specifying the characteristics of the spatial features. The boundary between the coastline and the Gulf of Mexico, the relationship of the various waterways and their characterization are difficult to specify as they exhibit both spatial and temporal uncertainty.

The remainder of this paper is organized as follows. In Section 2, we present a review of the statistical and non-statistical methods used for managing uncertain spatial data in GIS. More specifically, we review fuzzy set/possibility theory and rough set theory used for managing vague/ambiguous data and probability theory and Dempster–Shafer (D–S) theory for managing imprecise/inaccurate/incomplete spatial data. In Section 3, we discuss our study and results and in Section 4, we draw our conclusions and outline future research directions.

2. Managing uncertainties in spatial data

In this section, we examine some practical approaches used to represent various aspects of geospatial data. Uncertainty can refer to vagueness, ambiguity, imprecision, inaccuracy, incompleteness, or anything that is undetermined. In this study, we refer to "vagueness" as the inability to clearly understand the meaning of a word or phrase; "ambiguity" as multiple meanings in a word or phrase; "imprecision" as the level of variation associated with a set of measurements; "inaccuracy" as a situation where the assessment fails to give the true measurement; and "incompleteness" as the lack of relevant measurement.

A wide range of statistical and non-statistical methods have been proposed in the literature to model uncertainties in spatial data. In this study, we present a practical taxonomy of these methods by grouping them into two general categories: statistical and non-statistical methods. As shown in Fig. 2, statistical methods are often used to model imprecise, inaccurate, or incomplete spatial data while non-statistical methods are used to handle vague or ambiguous spatial data. Probability theory and D–S theory are the most widely used statistical methods for modeling uncertain spatial data while fuzzy set/possibility theory and rough set theory are the most commonly used non-statistical methods for managing uncertainties in spatial data modeling.

2.1. Statistical approaches

In this study, we identified 42 papers which applied D–S theory in a GIS environment. Malpica et al. [25] present a survey of (D–S) theory in GIS. Here we discuss how probability and D–S theory have been used to represent geospatial data with uncertainty.

The D–S theory of evidence (also referred to as the *belief function theory* or *evidential reasoning theory*) is general framework formalized by Shafer [35] for representing and reasoning with uncertain, imprecise, or incomplete information. Shafer's seminal book was based on Dempster's original idea [13] on the modeling of uncertainty in terms of upper- and lower-probabilities induced by a multivalued mapping [22]. The key concept in D–S theory is that an amount of probability mass (a value in $[0,1]$) can be assigned to a subset of a set of solutions to a question (such as all the possible values of size of a particular space) rather than just a singleton set, as in the case of probability

Fig. 1. Gulf of Mexico coastal region: Atchafalaya Bay area.