TOWARDS UNCERTAINTY-BASED GEOGRAPHIC INFORMATION SCIENCE (PART B) – THEORIES OF MODELING UNCERTAINTIES IN SPATIAL ANALYSES

Wenzhong Shi

Advanced Research Centre for Spatial Information Technology
Department of Land Surveying and Geo-Informatics
The Hong Kong Polytechnic University
Email: lswzshi@polyu.edu.hk; Tel: +852 - 2766 5975

Commission II

KEY WORDS: uncertain GIS, uncertainty analysis, spatial analysis, multi-data source analysis, overlay, buffer, line simplification, uncertainty-based spatial mining, quality control

ABSTRACT:

Within the framework of uncertainty-based geographic information science, this paper addresses modeling uncertainties in integrating multiple sources of data, modeling uncertainty in overlay analysis, modeling uncertainty in line simplification, uncertainty-based spatial data mining, uncertainty-based spatial queries, theory and methods for controlling the quality of spatial data, modeling uncertain topological relations, and dissemination of uncertainty information of spatial data and analysis.

1. INTRODUCTION

In the first of these two consecutive papers (Shi, 2005a), we raised the point of view that one of the development trends of geographic information science is from determinate geographic information science to uncertainty-based geographic information science. In order to establish uncertainty-based geographic information science, we need to develop a number of corresponding supporting theories. These two consecutive papers describe two major supporting theories that we have developed so far: (a) modeling uncertainties in spatial data, and (b) modeling uncertainties in spatial analysis. Specifically, these include:

- Theories, models, and methods for positional uncertainties in spatial data,
- Theories, models, and methods for thematic uncertainties in spatial data,
- Theories, models, and methods for uncertainties in digital elevation models (DEM),
- Quality improvement techniques and methods for satellite images,
- The modeling of fuzzy topological relationships between spatial objects,
- Theories and methods for uncertainties in spatial analyses and their propagation,
- Uncertainty-based spatial queries, and
- Methods for managing uncertainty information in spatial data and analyses.

The logical relations among the above theories are graphically described in the following diagram:

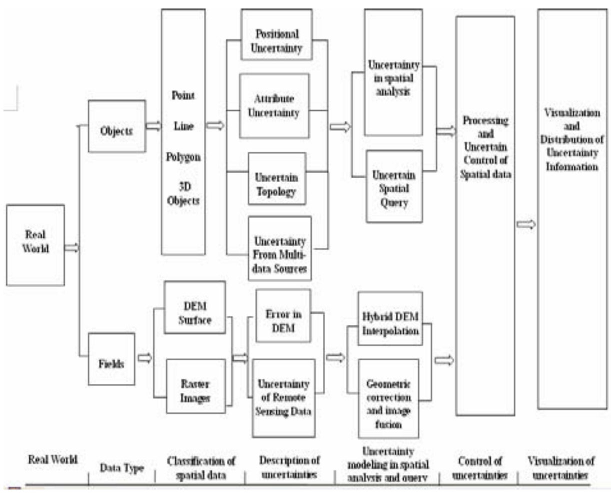

Figure 1. A graphical description of the logical relations among uncertainty theories in geographic information science

The previous paper addressed uncertainty modeling for spatial data, and this paper will concentrate on uncertainty modeling for spatial analyses and others. Specifically, this paper will cover modeling uncertainties in integrating multiple sources of data, modeling uncertainty in overlay analysis, modeling uncertainty in line simplification, uncertainty-based spatial data mining, uncertainty-based spatial queries, the theory and methods for controlling the quality of spatial data, modeling uncertainties in topological relations, and the dissemination of uncertainty information.

2. MODELING UNCERTAINTIES IN INTEGRATING MULTI-DATA SOURCES

2.1 The "S-band" model – integrating attribute and positional uncertainties

Compared with commonly used approaches where positional and thematic uncertainties are modeled separately, the philosophy behind the "S-band" model is that the uncertainty within the fuzzy boundary region of a composed object is related to both positional and thematic uncertainties in the description of the object, as well as the correlation between them. The general "S-band" model can be described by the following formula:

$$\mathbf{U_{PAT}(X)} = F(\mathbf{U_P}(X), \mathbf{U_T}(X), \rho(\mathbf{U_P}(X), \mathbf{U_T}(X))),$$

where $U_{PAT}(X)$ is the PAT uncertainty at X, and $X \in R^2$ is a generic lattice or pattern data location in a two-dimensional Euclidean space. X varies over the index set $D \in R^2$, which is the fuzzy boundary region; $U_P(X)$ is the positional uncertainty at X; $U_T(X)$ is the thematic uncertainty at X; and $\rho(U_P(X), U_T(X))$ is the correlation between $U_P(X)$ and $U_T(X)$ at X.

The general "S-band" model illustrates that the uncertainty at a location X is related to its positional and thematic uncertainty, and the correlation between them. $U_P(X)$ is quantitative in nature, $U_T(X)$ can be either qualitative or quantitative, and the correlation $\rho(U_P(X), U_T(X))$ between them can vary from one case to another. Therefore, we cannot define a function F to cover every case. A concrete form of F can only be given when a practical problem is well defined.

Based on the general theoretical "S-band" model, a practical "S-band" model can be developed. For an example in a GIS-based spatial data analysis, two layers of data are used. On the first layer are geometric features such as administrative boundaries. The uncertainty of this layer resulted from positional errors, e.g., errors from digitizing the boundaries. The second layer is the result of the classification of a remote sensing image; the uncertainty of this layer is caused by an error in classification, i.e., mainly a thematic error. What will be the uncertainty of a new layer generated by the two basic data layers? A practical "S-band" model is designed to solve this problem.

This problem can be defined as:

What is $P(H_{P \cap T})$, given $P(H_P / Z_P)$ and $P(H_T / Z_T)$?

Two methods have been developed to handle the combinational uncertainty resulting from both positional and thematic uncertainties (Shi, 1994). One is based on the probability theory and the other is based on a certainty factor model with a probabilistic interpretation.

2.1.1 Probability Theory-Based Combination Model:

In the probability theory, we have the product rule of probabilities as follows:

$$P(A \cap B) = P(A)P(B/A) = P(B)P(A/B),$$

where $P(A \cap B)$ is the probability that both A and B will occur; P(B/A) is the probability of B given the occurrence of A, i.e., the conditional probability of B given A; P(A/B) is the probability of A given the occurrence of B, i.e., the conditional probability of A given B; P(A) is the probability of A; and P(B) is the probability of B.

2.1.2 Combination Model Based on the Certainty Factor Concept:

When considering the propagation of both certainty and uncertainty, the certainty factor model is used in MYCIN (Shortliffe and Buchanan, 1975) for uncertainty-based reasoning. MYCIN is an expert system used for medical diagnosis. The certainty factor model can overcome the constraint of using a product rule based on the probability theory. Therefore, the certainty factor model can be used as an alternative.

We have then adopted the certainty factor model with the probabilistic interpretation, which was developed by Heckerman (1986) based on the original definition of the certainty factor. A detailed description of the solution can be found in Shi (1994).

2.2 Modeling uncertainties in change detection based on multi-temporal remote sensing data

An approach to determining uncertainties and their propagation in dynamic change detection based on classified remotely sensed images was proposed by Shi and Ehlers (1996). First, the uncertainties of a classified image using the maximum likelihood (ML) classification are determined. The probability vectors, which are generated during the maximum likelihood classification, are used as indicators of uncertainty. Second, the uncertainty propagation of classified multi-data images is described using two mathematical models. One is based on the product rule in the probability theory and the other is based on a certainty factor model with the probabilistic interpretation. Third, a visualization technique, using 3-D and color, is used to visually present uncertainties in the detection of change. This model is different from the “S-band” model. Here instead of the bundles of objects, we model the whole area covered by two raster images.

A case study is carried out, with the results indicating that the propagated change detection uncertainty is much higher than the uncertainty of the individual ML classifications. This increase is due to the propagation of uncertainty and indicates that the assessment of classification uncertainty alone is not sufficient to accurately quantify the uncertainties in the detection of change.

2.3 Rate of disfigurement model for assessing the quality of multi-data source

In the paper (Shi, 2005a), the rate of disfigurement model is proposed to assess the uncertainties of attribute data. In fact, the rate of disfigurement model can also be applied to assess the quality of an analysis based on multiple sources of data, which includes positional, thematic and other types of uncertainties. In this circumstance, there are a number of issues that need to be considered, such as determining the weight of each type of data, determining the sampling size and method(s), and so on.

Instead of modeling uncertainties in raster-based spatial analyses from previous studies, the uncertainty of modeling in vector-based spatial analyses is another area of research that is more generic and, at the same time, more complicated. Two approaches – analytical and simulating – were applied. Uncertainties and their propagation in overlay, buffer, line simplification, and uncertainty-based spatial data mining will be addressed in the following. In the area of GIS, this is a step further from static error modeling for spatial data to uncertainty modeling for spatial analyses.

3. MODELING UNCERTAINTY IN OVERLAY SPATIAL ANALYSES

Two methods are proposed to estimate the propagation of errors in vector overlays: an analytical error model derived based on the error propagation law, and a simulation error model (Shi, Cheung and Tong, 2004). For each of these two error models, it is proposed that the positional error in the original or derived polygons be assessed by three measures of error: (a) the variance-covariance matrices of the polygon vertices, (b) the radial error interval for all vertices of the original or derived polygons, and (c) the variance of the perimeter and that of the area of the original or derived polygons. Whereas the area and perimeter of the derived polygons have been introduced in previous studies, the x- and y-directional error intervals as well as the radial error interval for the vertices of the derived polygons are newly proposed.

The variance-covariance matrix of the vertices of an original or derived polygon is a relatively comprehensive description of error. However, such a matrix is neither practical nor easy to use as a measure of error and may also be large in size. Therefore, the radial positional error interval has been proposed as more practical for describing the error at the vertices of a polygon.

The results of a case study have demonstrated that the intersecting points of the sides of the original polygon have a higher accuracy than the vertices of the intersecting sides of the original polygon. Furthermore, it is also shown that the number of vertices and the error at the vertices of a polygon are two major factors that may affect the accuracy of parameters of the polygon such as perimeter and area. Increasing (or decreasing) both the number and the error of the vertices of a polygon will similarly influence the error of these polygon parameters. However, there is no consistent trend to the changes in the error of these parameters, if the number of vertices and the error at the vertices are not decreased or increased simultaneously.

4. MODELING UNCERTAINTY IN BUFFER SPATIAL ANALYSES

A method of modeling the propagation of errors in a buffer spatial analysis for a vector-based GIS has been developed (Shi, Cheung and Zhu, 2003). Here, the buffer analysis error is defined as the difference between the expected and measured locations of a buffer. The buffer can be for the GIS point, line-segment, linear, and area features. The sources of errors in a buffer analysis include errors of the composed nodes of point or linear features, and errors of buffer width. These errors are characterized by their pdf.

Four indicators of error and their corresponding mathematical models, in multiple integrals, have been proposed for describing the propagated error in a buffer spatial analysis. These include the error of commission, error of omission, discrepant area, and normalized discrepant area. These indicators are defined for describing various error situations in a buffer analysis. Both the error of commission and the error of omission are suitable for describing a situation where the expected and measured buffers overlap. The definition of these two indicators of error is consistent with the error indicators used for assessing classified images in remote sensing. The discrepant area indicator has been defined taking into consideration the possibility that the expected and measured buffers might not overlap with each other. This error indicator in fact provides a more generic solution for the cases where the measured and expected buffers either overlap with each other or do not. Furthermore, the normalized discrepant area has been proposed considering the mathematical rigors of the definition – to meet the conditions of the definition of distance in algebra.

To apply the proposed mathematical models of the indicators in a practical manner, it is recommended that the Gaussian Quadrature numerical integration method be used to solve the models in multiple integrals. One characteristic of this method is that it is highly accurate, and the authors have also proved in a previous study that it is capable of analyzing errors of spatial features.

5. MODELING UNCERTAINTY IN LINE SIMPLIFICATION

Line simplification is performed in GIS in order to remove redundant points in a line or to reduce the volume of data for representing lines; however such a process leads to positional uncertainties caused by (a) the uncertainty in an initial line and (b) the uncertainty due to the deviation between the initial and simplified lines. In our study (Cheung and Shi, 2004), the uncertainties in a line simplification process are classified as a combination of the propagated uncertainty, the modeling uncertainty, and the overall processing uncertainty. The propagated uncertainty is used to identify the uncertainty effect of the initial line, and the modeling uncertainty represents the uncertainty arising from the line simplification process. The overall uncertainty in the simplified line is modeled by the overall processing uncertainty that integrates both the propagated and the modeling uncertainties in the line simplification process. Three uncertainty indices and corresponding mathematical solutions were proposed for each type of uncertainty by measuring its mean, median, and maximum values. For the propagation uncertainty, we proposed the mean discrepancy, the median discrepancy, and the maximum discrepancy; for the modeling uncertainty, we proposed the mean distortion, the median distortion, and the maximum distortion; for the overall processing uncertainty, we proposed the mean deviation, the median deviation, and the maximum deviation.

The mean and median uncertainty indices for each type of uncertainty are different in value although both types measure the central tendency of a particular type of uncertainty. The distributions of all of the types of uncertainty are positively skewed. The mean uncertainty index provides the general value of the type of uncertainty. If users want to minimize the average value of the type of uncertainty, the median uncertainty index will be given. In this study, the relation between the mean uncertainty index for the overall processing uncertainty and the threshold distance in the DP line simplification was studied. It was found that the mean uncertainty index is a monotonic increasing function of the threshold distance. Also, this function is used to determine the threshold distance such that an uncertain simplified line is close to the "true" initial line to a predefined acceptable level of accuracy.

The maximum uncertainty index considers an extreme case in the type of uncertainty. This index for the overall processing uncertainty is important for describing an uncertainty distribution of the simplified line. The maximum deviation is a maximum distance measure for the "true" initial line and the uncertain simplified line. It is considered a buffer width such that the buffer around the mean location of the simplified line contains the "true" initial line. Therefore, points of the simplified line are expressed as a linear combination of their mean locations plus a random number evenly distributed in a square of the maximum deviation. The uncertainty distribution of the simplified line is important for assessing the uncertainty in the spatial results derived from the uncertain simplified line.

According to our experimental study, the uncertainty indices for the propagated and modeling uncertainties were not small when compared with the overall processing uncertainty. Since the propagated and modeling uncertainties assess the effect of the uncertainty in the initial line on the line simplification process, and the distortion between the measured initial line and the corresponding simplified line respectively, it is more rational to measure the uncertainty in the simplified line by considering these two sources of uncertainty. Most previous studies have neglected the uncertainty in the initial line when modeling the uncertainty problem for a line simplification process. The uncertainty indices for the overall processing uncertainty proposed in this study are therefore more comprehensive in order to measure the uncertainty in the simplified line.

6. UNCERTAINTY-BASED SPATIAL DATA MINING

The idea of spatial data mining with uncertainty and a corresponding theoretical framework has been proposed (Shi *et al*, 2003). First, uncertainty in spatial data is presented in terms of its characteristics, perspectives, and role in spatial data mining. Both randomness and fuzziness often appear at the same time. Second, a cloud model, a mathematical model that studies randomness and fuzziness in a unified way, is applied to model uncertainty in spatial data mining. Furthermore, the cloud model may act as an uncertainty transition between a qualitative concept and its quantitative data, which is implemented with the cloud generators. Finally, a case study on landslide-monitoring data mining is conducted. The results show that uncertainty is unavoidable when spatial data mining is carried out, and that the quality of the final discovered knowledge may be improved if the uncertainties are properly modeled.

7. UNCERTAINTY-BASED SPATIAL QUERIES

In a point-in-polygon analysis, a point and a polygon are uncertain due to random errors introduced in the process of capturing data or other processes. Existing research studies model the uncertainties of the point and polygon on the assumption that the uncertainties of all points located inside the polygon follow a circular normal distribution identically and independently. In our study of this field (Cheung, Shi and Zhou, 2004), we have proposed a probability-based uncertainty model for a point-in-polygon analysis in a more generic case regarding existing research problems, including those in which (a) both the point and the polygon are uncertain, (b) the error ellipse of the point, which is more rigorous than the error circle model, should be used to describe the uncertainty of a point, and (c) where the uncertainties of the vertices of a polygon may be correlated and different from each other.

In order to provide the probability of an uncertain point located inside an uncertain polygon, the uncertainties of the point and of the vertices of the polygon are described in terms of probability density functions. The probability of an uncertain point located inside an uncertain polygon is then derived in terms of multiple integrals based on probability and statistical theories. Since this expression involves many variables of integration, we divide the point-in-polygon analysis into different cases depending on the intersection between the polygon and the error ellipse of the point. The mathematical expressions for the probability in the individual cases are also provided.

8. THEORY AND METHODS FOR CONTROLLING THE QUALITY OF SPATIAL DATA

In previous studies, the modeling of uncertainties in either spatial data or spatial analyses has focused mainly on *describing* the quantity of errors or uncertainties. A further question would be whether we can reduce the quantity of these errors or uncertainties. Studies on controlling the quality of spatial data are trying to address this issue.

A least squares-based method, designed particularly for solving inconsistencies between areas of digitized and registered land parcels, had been proposed for adjusting the boundaries of area objects in a GIS. The principle of this approach is to take the size of the registered area of a land parcel as its true value and to adjust the geometric position of the boundaries of the digitized parcel. First, a generic area adjustment model is derived by incorporating the following two categories of constraints: (a) attribute constraints: the size of the true area of the parcel, and (b) geometric constraints: such as straight lines, right angles, and certain distances. Second, the methods used to adjust the areas of the parcels for different cases are presented. It is demonstrated, via several case studies, that the proposed approach is able to maintain a consistency between the areas of the digitized and registered parcels. This study has solved one of the most critical problems in developing a land/cadastral information system, and this solution has been adopted in the processing of real world cadastral data in Shanghai and other cities in China (Tong, Shi and Liu, 2005).

The processing of cadastral area is further studied due to the assumption that the digitized coordinates of a parcel contain errors, while the registered area of the parcel is regarded as a known value without error, and is less rigorous (Tong and Zhao, 2002). The reason for this is that the registered area of a parcel is also obtained by surveying methods and thus contains errors. We should then treat the registered areas of the parcels as observations with errors. Therefore, it is necessary to solve the inconsistencies between the digitized and registered areas of land parcels under the condition that both the areas and coordinates are treated as observations with errors. The problem then comes to be defined as how to determine the prior knowledge of the weights of these observations since these are two different types of observations, including different sources of data and units of measure. In our study, a least squares adjustment based on the Helmert method is presented for estimating accurate weights between the area and coordinates. At the same time, the insistency between the registered area and digitized area of the parcel is adjusted through the least squares adjustment. The computational results show that the Helmert method can accurately estimate the weights of the observables, and the least squares adjustment based on the variance components of the unit weights can solve the insistency between the registered area and digitized area of the parcel more rigorously. As a result, the accuracy of the parcel area is improved via the surveying adjustment and quality control process.

9. MODELING UNCERTAINTIES IN TOPOLOGICAL RELATIONS

The topological relation between spatial objects is one of the fundamental properties of GIS data, which can be used for spatial analysis, spatial queries and controlling the quality of data. In order to form a theoretical basis for spatial analysis, we extend topological relations to uncertain topological relations based on fuzzy topology and probability theory.

9.1 Extended model of topological relations between crisp GIS objects

The current models for describing topological relations between crisp objects in GIS (Egenhofer, 1993; Cohn and Gotts, 1996; Smith, 1996) are first extended and followed by a new definition of the topological relations between two objects, which is an extension of the traditional definition based on empty and non-empty objects under homeomorphic mapping (Liu and Shi, 2003). Based on this new definition, including the topology of the object itself and several topological properties, we have uncovered a sequence of topological relations between two convex sets.

There are a number of new findings from this study. Among them, two major findings are: (a) that the number of topological relations between the two sets is not as simple as finite; actually, it is infinite and can be approximated by a sequence of matrices, and (b) that the topological relations between two sets are dependent on the shapes of the sets themselves.

9.2 Modeling fuzzy topological relations in GIS

The boundary of an object in GIS can be either crisp or vague/fuzzy. The classical set theory, which is based on a crisp boundary, may not be suitable for handling vague or fuzzy boundary problems. An incorrect modeling of fuzzy/ vague GIS objects may lead not only to the loss of information, but also to incorrect descriptions of the reality in a GIS. For instance, a tide makes it difficult to determine the boundary of a sea, such as the boundary of the Pacific Ocean. Due to the tidal effect, some islands in the Ocean may appear and disappear from time to time. We may also have problems describing the topology of these islands based on a crisp and static GIS. Therefore, the classical set theory may not be a suitable basic tool for describing objects in GIS. Alternatively, the fuzzy set theory provides a useful solution to the description of uncertain objects in GIS. The fuzzy topology can be applied to investigate fuzzy topological relations among objects in GIS.

For modeling fuzzy topological relations among uncertain objects in GIS, the quasi-coincidence and quasi-difference are used (a) to distinguish the topological relations among fuzzy objects and (b) to indicate the effect of one fuzzy object to the other, based on fuzzy topology (Shi and Liu, 2004). Geometrically, features in GIS can be classified as point features, linear features. and polygon or region features. In our study, we first introduced several basic concepts in fuzzy topology, then followed these by definitions of fuzzy points, fuzzy lines, and fuzzy regions for GIS objects. Next, the level of one fuzzy object that affected the other is modeled based on the sum and difference of the membership functions, which are the quasi-coincidence and quasi-difference, respectively.

9.3 Computable fuzzy topology in GIS

The existing topological models can define conceptual definitions of interiors, boundaries, and exteriors. However, in many cases, we need to compute the level of an interior, boundary and exterior. Therefore, we may need to propose new formulae to compute levels of interiors, boundaries, and exteriors. Based on fuzzy topology, we developed a computable fuzzy topology for calculating the interior, boundary, and exterior of a spatial object once the membership function is known.

Before we could determine a computable fuzzy topology, we first defined two new operators: interior and closure operators, which are used to further define a computable fuzzy topology. We know that each interior operator corresponds to one fuzzy topology and that each closure operator corresponds to one fuzzy topology (Liu and Luo, 1997). In general, if we define two operators, interior and closure, separately, they will define two fuzzy topologies, respectively. These two topologies may not cohere to each other. That is, the open set defined by the interior operator may not be the complement of the closed set, which is defined by the closure operator. In order to define a computable fuzzy topology, we want these two operators to be able to cohere with each other so that they can define the same fuzzy topology, in which the open sets are the image of the interior operator and the closed sets are the image of the closure operator. At the same time, we set the complement of an open set to be a closed set. The interiors, boundaries, and closures of fuzzy spatial objects in GIS can thus be computed (Liu and Shi, 2005).

9.4 Qualitative and quantitative fuzzy topological relations under several invariant properties

The above fuzzy topological relations are elementary relations in the study of topological relations between spatial objects in GIS. Many researchers had developed their model in this field based on these relations (Clementini and Di Felice, 1996; Shi and Guo, 1999; Tang and Kainz, 2001).

The properties of topological spaces that are preserved under homeomorphic mappings are called the topological invariants of the spaces. To study the topological relations, we first need to investigate the properties of a fuzzy mapping, especially a homeomorphic mapping. Based on the newly developed computational fuzzy topology, methods for computing the fuzzy topological relations of spatial objects have been proposed in this study. Specifically, the following areas are covered: (a) the homeomorphic invariants of the fuzzy topology have been proposed; (b) the connectivity based on the newly fuzzy topology has been defined; (c) for modeling the topological relations between spatial objects, the concepts of a bound on the intersection of the boundary and interior, and the boundary and exterior have been defined based on the computational fuzzy topology, which are $(A_\alpha \wedge \partial A)(x) < 1-\alpha$ and $((A^c)_\alpha \wedge \partial A)(x) < 1-\alpha$. With these, we can guarantee the properties that remain unchanged in a GIS transformation, such as the maintenance of topological consistency when transferring a map from one system to another. Moreover, among these topological relations, we can extract useful topological relations, as this activity commonly occurs in GIS (Shi and Liu, 2005a).

9.5 Shape-dependant topological relations

Topology is normally considered to be independent of the shapes of spatial objects. This may not necessarily be true when describing relations between spatial objects in GIS. We have proven (Shi and Liu, 2005b) that the topological relations between spatial objects are dependent on the shapes of spatial objects. That is, that the topological relations of non-convex sets cannot be deformed to the topological relations of convex sets. The significant theoretical value of this research is in its finding that the topologies of spatial objects are shape-dependent. This indicates a necessary consideration of both topologies and shapes of objects when describing topological relations among spatial objects in GIS. As a result, some of the spatial data modeling, queries, and analyses based on the existing understanding of the topologies of spatial objects may be needed to be re-assessed.

10. DISSEMINATION OF UNCERTAINTY INFORMATION OF SPATIAL DATA AND ANALYSIS

Two technologies have been developed for disseminating uncertainty-related information for GIS: (a) to manage error metadata in a database, and (b) to distribute uncertain information by web-service technology.

10.1 Object-oriented error metadata database

An object-oriented error metadata database – EMMS has been designed and developed for managing metadata at the object level rather than at the map level or feature level (Gan and Shi, 2002). It captures quality information and temporal information, information about sources of data, processing steps, and cartographic generalizations on the dataset. Such information relating to data sources and data validation will be captured automatically during the process of being manual inputted by operators. In the example of the Hong Kong 1:20,000 Digital Topographic Map Series, the EMMS is used to display metadata, especially when data are conflated. Metadata for other datasets such as 1:10,000 and 1:50,000 maps in Hong Kong or elsewhere can also adopt the EMMS technology.

Metadata for individual objects may not yield instant benefits to data producers. However, in the long run, the error metadata can support users in making decisions regarding the appropriateness-for-use of a dataset. It also facilitates the determination of positional errors of GIS objects in the geometric form of point, line, and polygon. Quality control and quality assurance can also be implemented through the installation of the EMMS for data providers. Data users can ultimately select appropriate GIS data for a particular application based on the metadata in the EMMS.

10.2 Web service-based uncertainty information distribution techniques

Based on web service technology, an internet-based distributed computing environment has been built to provide a solution to the web service-based uncertainty information service and the metadata distribution (Shi, 2005b). Specifically, it can be used to design the web service-based data quality information system, to provide the technical framework for the web service-based data quality information system, and to provide a practical solution to the distribution of uncertainty information. Figure 2 is an example of the user interface of a web service-based information system for disseminating the error information of points, where the error of each point feature is illustrated by the error of its ellipse.

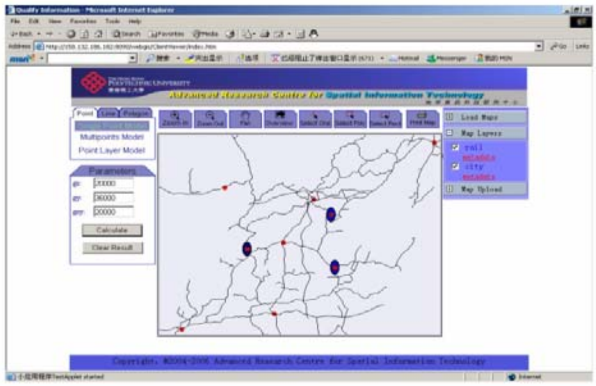

Figure 2. An interface of a Web service-based information system for disseminating uncertainty-related information for users.

11. CONCLUDING REMARKS

This paper addresses a number theories or solutions in modeling uncertainties in spatial analyses, which is one the two major supporting theories for uncertain geographic information science. At the same time, it also takes research on modeling uncertainties in spatial data a step further. Progress has been made in the following areas:

- from the modeling of static uncertain data to dynamic uncertain spatial analysis;
- from modeling errors for single types of data to multiples sources of data;
- from *measuring* the quality of data to *controlling* the quality of data, and possibly *reducing* overall errors in GIS data via least square or other possible solutions, such as image fusion and geometric corrections to remote sensing images;
- from traditional positional and attribute error modeling to spatial data mining, uncertain topological relations; and
- to disseminating uncertain information using the latest information technology – web service.

This and the previous paper (Shi 2005a) only addressed two major supporting theories of uncertainty-based geographic information science. This is just a beginning in the development of uncertainty-based GIS. We expect more developments, which will enrich uncertainty-based GIS in the near future.

ACKNOWLEDGMENTS

The work described in this paper was supported by grants from the Hong Kong Polytechnic University (PolyU 5071/01E), and Hong Kong Research Grants Council (Project No. PolyU 5167/03E).

REFERENCES

Cheung, C.K, W Z Shi, 2004, Estimation of the positional uncertainty in line simplification in GIS, *The Cartography Journal*, Vol. 41, No.1, pp.37-45.

Cheung, C.K., W.Z. Shi, and X. Zhou, 2004, A probability-based uncertainty model for point-in-polygon analysis. *GeoInformatica*, Vol. 8, No. 1, pp. 71-98.

Clementini, E., and D.P. Felice, 1996, An algebraic model for spatial objects with indeterminate boundaries. *Geographic Objects with Indeterminate Boundaries*, edited by Burrough, P.A., and A.U. Frank, Taylor & Francis, London. Pp. 155-169.

Cohn, A.G., and N.M. Gotts, 1996, The "egg-yolk" representation of regions with indeterminate boundaries. *Geographic Objects with Indeterminate Boundaries* edited by Burrough, P.A. and A.U. Frank, Taylor & Francis, London. Pp. 171-187.

Egenhofer, M., 1993, A model for detailed binary topological relations. *Geomatica*, 47 (3) & (4), pp. 261-273.

Gan, E. and W. Z. Shi, 2002, Error Metadata Management System. In Shi, W.Z., P.F. Fisher and M. F. Goodchild (ed.) *Spatial Data Quality*, Taylor and Frances. Pp. 251-266.

Heckerman, D., 1986, Probabilistic interpretation of MYCIN's certainty factors. *Uncertainty in Artificial Intelligence* eds. by Kanal, L.N., and J.F. Lemmer, pp. 167-196.

Liu, K.F., and Shi, W.Z., 2005, A fuzzy topology for computing the interior, boundary, and exterior of spatial objects quantitatively in GIS, *Photogrammetric Engineering and Remote Sensing* (submitted).

Liu, K., and W.Z. Shi, 2003, Analysis of topological relationships between two sets. *In: Shi, W.Z., and M. F. Goodchild and P.F. Fisher (ed) Proceedings of the Second International Symposium on Spatial Data Quality*, March 19th – 20th, Hong Kong, pp. 61-71.

Liu Y.M. and Luo M.K., 1997, *Fuzzy Topology*. World Scientific.

Shi, W.Z., 1994, *Modeling Positional and Thematic Uncertainties in Integration of Remote Sensing and Geographic Information Systems*. ITC Publication, Enschede, ISBN 90 6164 099 7, 147 pages.

Shi, W.Z., and W. Guo, 1999, Topological relations between uncertainty spatial objects in three-dimensional space. *Proceedings of International Symposium on Spatial Data Quality*, July in Hong Kong, pp. 487-495.

Shi, W.Z., X.H. Tong, and D.J. Liu, 2000, An Approach for Modeling Error of Generic Curve Features in GIS. *ACTA Geodaetica et Cartographica Sinica*. Vol. 29, No. 1, pp. 52-58.

Shi, W.Z., C.K. Cheung and C.Q. Zhu, 2003, Modeling error propagation in vector-based buffer analysis, *International Journal of Geographic Information Science* Volume 17, Number 3, pp. 251-271.

Shi, W. Z, S.L. Wang, D.R. Li, and X.Z. Wang, 2003, Uncertainty-Based Spatial Data Mining, *Asia GIS 2003 Conference*. October 16 – 18, 2003, Wuhan, China. Session 1, paper 3, 20 pages.

Shi, W.Z., C.K. Cheung and X.H. Tong, 2004, Modeling error propagation in vector-based overlay spatial analysis, *ISPRS Journal of Photogrammetric and Remote Sensing*, Vol. 59, Issues 1-2, pp.47-59.

Shi, W.Z. and K.F. Liu, 2004, Modeling fuzzy topological relations between uncertain objects in GIS, *Photogrammetric Engineering and Remote Sensing*, Vol. 70, No. 8, pp. 921-930.

Shi, W.Z. and K.F. Liu, 2005a, *Computing the fuzzy topological relations of spatial objects based on induced fuzzy topology*, *International Journal of Geographic Information Science* (submitted).

Shi, W.Z. and K.F. Liu, 2005b, Are topological relations dependent on the shape of spatial objects? *Progress in Nature Science* (Accepted)

Shi, W. Z., 2005a, Towards uncertain geographic information science (part A) -- modeling uncertainties in spatial data. *Proceedings of International Symposium on Spatial Data Quality*, August 25-25, 2005, Beijing.

Shi, W.Z., 2005b, *Principle of modeling uncertainties in spatial data and analysis*, Science Press, Beijing, 408 pages.

Shortliffe, E.H., and B.G. Buchanan, 1975, A model of inexact reasoning in medicine. *Mathematical Biosciences 23*, p 351-379.

Smith, B., 1996, Mereotopology: A theory of parts and boundaries. *Data & Knowledge Engineering* 20, p 287-303.

Tang, X.M., and W. Kainz, 2001b, Analysis of topological relations between fuzzy regions in general fuzzy topological space. *Proceedings of the SDH Conference 02*, Ottawa, Canada.

Tang, X.M., and Kainz, W., 2001, Analysis of Topological Relations between Fuzzy Regions in General Fuzzy Topological Space. *Proceedings of SDH conference 02'*, Ottawa, Canada.

Tong, X.H., W.Z. Shi and D.J. Liu, 1999, Uncertainty Model of Circular Curve Features in GIS. For *ACTA Geodaetica et Cartographica.* Vol. 28, No. 4, pp. 325-329.

Tong X.H., Shi, W.Z., Liu, D.J., 2005, A least squares-based method for adjusting the boundaries of area objects. *Photogrammetric Engineering and Remote Sensing*, Vol. 71 No.2: pp. 189-195.