

RESPONSIBILITY, UNCERTAINTY AND FITNESS-FOR-USE:  
IMPLICATIONS FOR GEO-INFORMATION THEORY.

Allan J. Brimicombe.

Centre of Land & Engineering Surveying,  
Hong Kong Polytechnic,  
Hong Kong.

**ABSTRACT:**

The accuracy and level of uncertainty in GIS products has been a growing concern within the user community. The issue of uncertainty in GIS is explored and current strategies for the measurement, modelling and management of error are reviewed. Current approaches have significant limitations for imprecise spatial data. A general model is proposed as a basis for developing workable solutions. An example of how the model operates is given. Implications are that uncertainty measures should be embedded within GIS data rather than the use of global measures. Users must take responsibility for assessing fitness-for-use of their data for the particular context of an analysis.

**KEY WORDS:** Data uncertainty, Propagation, Theoretical model.

THE ISSUE OF UNCERTAINTY IN GIS

Recognition of the Issue

Geographic Information Systems (GIS), as a tool, is proving successful in apparently reducing practical problems of organizing and integrating spatial data and for carrying out complex spatial analyses. Initial concerns for developing and implementing the technology on the one hand and the pressing need to convert both spatial and non-spatial data from analog to digital formats on the other, has largely resulted in the need to address data and analytical integrity being overlooked. One suspects that in the euphoria of automated overlay analysis, earlier warnings such as those of MacDougall (1975) concerning the integrity of the manual overlay process were either forgotten or deemed no longer relevant. It may even be that operators and managers from non-spatial disciplines were unaware of this dimension to their work, other than the need to address blunders.

As users have become more familiar with the technology and as GIS has been accorded a more sophisticated role such as through incorporation into decision support systems, the quality of results portrayed in the graphical output has been brought increasingly into question. Data quality in GIS began to be discussed in the literature in the early 1980's (e.g. Chrisman, 1982; Mead, 1982; Blakemore, 1984; Newcomer & Szajgin, 1984; Vitek et al., 1984). However it was not until the late 1980's that issue seems to have become of more widespread concern (17 papers in 1987, 2 papers in 1988, 14 papers in 1989). A bibliography has been published (Veregin, 1989) and a book devoted to the subject (Goodchild & Gopal, 1989), but although the issue is now regularly featured in conferences and texts, the number of papers would appear to be in decline. The problem has been identified and defined, but few solutions have emerged. Not surprisingly then, vendors do not seem to be taking much action on the issue.

The Nature of the Issue

Uncertainty can be used as a global term to encompass any facet of the data, its manipulation or its presentation which may raise concern, doubt

or skepticism in the mind of the user as to the validity of the results or intended message. Theoretically this definition would also include mishandling of the data through improper analysis, inappropriate or erroneous use of GIS functions, poor cartographic technique and so on. However, it is not the purpose of this paper to address that aspect of the issue. Mostly it is concerned with the inaccuracies, inexactness or inadequacies inherent in most spatial data sets and how these may propagate to possibly invalidate the informational output of the system. Lack of appropriate visualization techniques to portray data variability is also likely to exacerbate the issue (Beard et al., 1991).

Figure 1 summarizes the possible sources of uncertainty. GIS is a veritable minefield in this regard (see for example Burrough, 1986; Walsh et al., 1987 and Maguire et al., 1991 for general discussions). The end effect results in two broad categories of uncertainty: the position of objects (in 2, 3 or 4 dimensions) and the characterization (description and/or additional dimensionality) of these objects (Table 1). If we also include here the location and correct assignation of any null space, then the above two categories will include concerns over the completeness of the data. Notwithstanding Figure 1, the main cause for concern rests with the ability of GIS to adequately record, manipulate and display the natural variation that exists in most spatial data. Hence Burrough's statement that "many soil scientists and geographers know from field experience that carefully drawn boundaries and contour lines on maps are elegant misrepresentations of changes that are often gradual, vague or fuzzy" (Burrough, 1986).

The Effect

Despite increased awareness on the part of users, canned demonstrations continue to overly impress (Congalton, 1990). Complex manipulations of data, seductive graphics, and implicit faith in the computer easily turns an old adage on its head - garbage in, gospel out! The author believes that this situation has developed due to the absence of an adequate theory or general model for handling uncertainty in GIS.

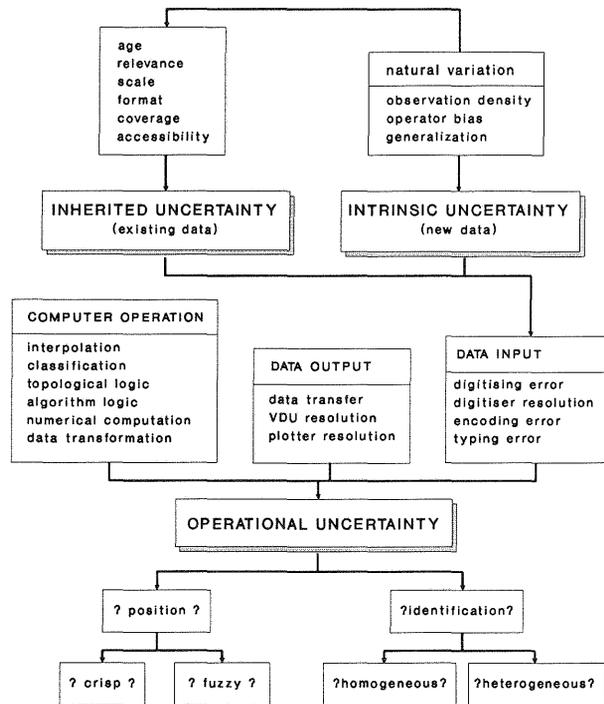


Figure 1: Sources of Uncertainty in GIS

MEASURING, MODELLING AND MANAGING UNCERTAINTY

The general treatment of uncertainty in GIS to date reflects the conceptual closeness of digital maps to their analog roots. Whilst Chrisman may assert that "no map can be picked apart into completely independent pellets of information" (Chrisman, 1991), any graphic output from a GIS can be disaggregated into individually encoded entities. These entities can be subjected to an assessment of their quality, either individually or in sets. Concern with uncertainty in spatial data has concentrated on reducing and modelling error. A recent overview is given by Chrisman (1991) whilst a detailed treatment is provided by Veregin (1989).

Measurement

The main focus has been on establishing the extent to which locational and/or attribute errors are present within a dataset and, in the recognition that these can never be entirely eliminated, how to characterize these errors as a metric or statistic.

Adopted wholesale from the mapping sciences has been the testing of geometrical accuracy based on well-defined points having no attribute ambiguity (Bureau of Budget, 1947; ASCE, 1983; ASPRS, 1985). Testing is carried out by reference to a survey of

higher order. Whilst horizontal and vertical accuracy can be treated individually, measures such as Koppe's formula recognize the link between horizontal and vertical accuracy. The logical expectation that well-defined points of no attribute ambiguity are more likely to be accurately surveyed in the first place would certainly bias any outcomes, but nevertheless these accuracy tests are widely accepted. How often these tests are actually carried out is another matter! In the context of GIS, the main consideration is that much spatial data do not contain well-defined points.

Appropriate testing of attribute accuracy depends on the measurement class used. Thus continuous data such as digital elevation models (DEM) can be tested for horizontal and vertical accuracy as above either through interpolating contours or by interpolating to known points. An alternative is statistical analysis of expected and observed values. This type of treatment can be viewed as overlooking certain important issues. For DEMs, consideration should be given to the limitations of source documents (quality of maps or vegetation on aerial photographs), the sampling interval and orientation of sampling (if in lines or on a grid) which will need to be adjusted depending on terrain and proposed use of the data (Theobald, 1988).

Contextual Resolution		Characterization of Spatial Entity	
		Exact	Inexact
Locational Definition of Spatial Entity	Exact	No uncertainty	Uncertain character
	Inexact	Uncertain location	Uncertain character Uncertain location

Table 1: Contingency table for positional and characterization errors (modified from Robinson & Frank, 1985).

For attributes recorded on nominal scales or by discrete classes, the use of misclassification matrices is widely used. Such matrices are of particular importance in testing interpretive data from remote sensing. A number of indices can be derived to summarize the matrix such as percentage pixels correctly classified (PCC) and its variants. Debate has not only centered around the appropriate derivation of indices but also on whether these should reflect accuracy from the producer's or user's point of view (Story & Congalton, 1986). Testing is still reliant on points though chosen through a sampling scheme. Difficulties here arise because classification schemes rarely have the mutual exclusivity of crisp sets, boundaries are often avoided (as in soil sampling) and the position of the sampling point must be correctly located on the ground. Middelkoop (1990) puts forward an alternative approach whereby a confusion matrix, generated by having several experts carry out interpretation of the same test area, is used to study boundary uncertainty.

During the course of data collection and input into a GIS, a number of data accuracy measures become available which could be included in the database for later use in assessing the validity of analyses. For example, if vectors are digitized from a base map then the expected accuracy of the map may be known (e.g.  $\pm 0.5\text{m}$  for planimetric detail), the error in control points for X and Y axes after map registration on the digitizer should also be known (e.g.  $\sigma_x=0.19\text{mm}$ ,  $\sigma_y=0.14\text{mm}$  at map scale) and then tests of human performance using digitizer pucks would indicate accuracies of  $\pm 0.25\text{mm}$  at map scale or half this if a magnifier is used (Rosenberg & Martin, 1988). For attributes, accuracy measures (Pccs or Rmses depending on data class) may result from fieldwork. The author is unaware of any commercial GIS software that automatically records such data and attributes them to entities, even when generated internally by the GIS software (as in map registration or rubber sheeting). Much of what could be used gets left behind along the way.

### Modelling

Modelling in the broadest sense would have to include the choice and nature of metric, statistic or range of verbalizations used to describe error or other uncertainty prior to measurement. More narrowly, this section will consider some currently proposed strategies for handling uncertainty in data transformations within a GIS. Consideration could be given to a very wide range of data transformations (Tobler, 1990). Assuming, from the above section on measurement, something is known about the accuracy of ones data (location and attribute), what is the accuracy of a derivative map compiled by combining data as in overlay analysis?

Map overlay will combine the locational and attribute errors of two or more layers. For vector data, locational errors will result in the spurious polygon or sliver problem. A number of algorithms have been developed and implemented in some GIS software to remove spurious polygons in an equitable way. These employ models based on the epsilon band concept (Blakemore, 1983; Chrisman, 1983; Pullar, 1991), maximum perpendicular deviation (Peucker, 1976) or fuzzy tolerances (Zhang & Tulip, 1990). Slivers are considered undesirable and whilst their removal reduces both database size and processing time and enhances the aesthetic quality of the cartographic product, they are themselves (or their absence) an indication of

quality and their automated removal at each successive stage of a complex analysis would introduce its own uncertainty.

Propagation of attribute error is of greater concern. Much of the work in modelling such errors has been carried out using raster data. Newcomer and Szajgin (1984) use conditional probability for overlay assuming a Boolean AND operator. In such cases, the highest accuracy expected is equal to the accuracy of the least accurate layer used. Usually though, accuracy will continue to decrease as more layers are added. Tests by Walsh et al. (1987) seemed to confirm the earlier pessimism that "it is quite possible that map overlays by their very nature are so inaccurate as to be useless and perhaps misleading for planning" (MacDougall, 1975). However, Veregin (1989) demonstrates that a Boolean OR operation for conditional probabilities will result in an accuracy not less than the most accurate layer used. Thus in a complex analysis using a combination of Boolean operators, composite map accuracy at each stage may improve or worsen significantly and hence an ability to track this is desirable. Recording of lineage in GIS operations (Lanter, 1990) seeks to address this requirement. A diagrammatic example of the effects of data reselection, union and intersection using PCC values is provided by Lanter and Veregin (1991).

Alternative approaches have been explored. Evidential reasoning (Shafer, 1976) has been used by Lee et al. (1987) and Garvey (1987) to combine multisource data. Belief functions are assigned to the data which by evidential computation and decision rules result in a measure of the plausibility or support for a particular proposition. Leung (1988) and Wang et al. (1990) have used fuzzy membership functions to assign climatic regions and land suitability classes respectively to conventional datasets. Heuvelink et al. (1989), using mean attribute values for each cell derived from kriging, were able to assess the reliability of their derivative maps by modelling the error propagation as a second-order Taylor expansion.

General, workable solutions have not been demonstrated in the literature. The only study to provide visualization of reliability of composite maps as a continuous surface (rather than global measures) is Heuvelink et al. (1989). Their initial accuracy measures, however, are a product of the kriging process and therefore can only be implemented where interpolation of quantitative point samples by this technique is appropriate.

### Management

If data quality is an important concern to both GIS implementors and users, then management strategies are required for controlling or reducing uncertainty and for ensuring fitness for use of products. Without a general model for handling uncertainty such strategies may be difficult to develop resulting in a series of loosely organized actions that may not achieve the desired goals. Current developments are concerned with consistency checks, metadata and standards.

Logical consistency checks can be carried out both for entities and attributes (Laurini & Thompson, 1992). Topology is most frequently used to check the integrity of vector polygons, tessellations and DEMs. Additional techniques used on DEMs are spatial autocorrelation (Caruso, 1987) and outlier detection using B-splines (Bethel & Mikhail, 1983). Attributes can be assessed for consistency with the

domain, are of the correct data class and fall within predefined ranges. Although these types of tests can detect blunders, their main aim is ensuring logical consistency of the data rather than improving accuracy.

Metadata is 'data about data' (Medyckyj-Scott et al., 1991). Coming under the umbrella of metadata are general descriptions of the data, provenance lineage reports, specifications of source materials and the data dictionary. Its purpose is to allow potential users to assess the suitability of the data for a specific task. Metadata is therefore an integral component of data standards. The proposed standard for Digital Cartographic Data Quality (Moellering, 1988) requires that measures of accuracy (positional and attribute), consistency, completeness and the data's lineage be recorded. Whilst such information is clearly desirable, the true place for quality information is embedded with the data it pertains to. Only in this way can data be proactive about its uncertainty.

#### FURTHER PERSPECTIVES ON POSSIBLE SOLUTIONS

The above short review, though somewhat selective, indicates that many issues are outstanding, the most fundamental of which is a theoretical basis from which to implement the means to record, propagate and visualize uncertainty in GIS. Before presenting a general model proposed to fulfill this function, additional perspectives on the solution to be sought are discussed.

As commented above, much of the concern has been with testing accuracy and eradicating error. This assumes there is 'truth' against which to measure, that is the existence of a binary right or wrong. Most spatial data does not fall into mutually exclusive sets with exact boundaries. Soils and other geomorphic data are typical. More research could be fruitfully expended on how to preserve natural variation and fuzzy boundaries in GIS rather than the current practice of making all data fit a crisp representation in a digital database. Existing techniques of using confusion matrices are inapplicable to aerial photographic interpretive data such as flood susceptibility or slope instability, as the chances of being able to sample flooding events or landsliding may be short-lived or rare and in any case hazardous. Using a panel of experts to review all such interpretive data may actually increase the uncertainty if a consensus cannot be reached!

Recourse to check surveys of higher accuracy may not be the answer either for this type of data. For each inexact phenomenon there is a characteristic resolution (linked to scales of space and time) at which confidence in identification and delineation is at a maximum (Davis et al., 1991). Trying to survey at a different resolution will only increase uncertainty. The object 'village' breaks down into buildings and sub-landuses in one direction making delineation more difficult and gradually reduces to a dot in the other. In a similar way, increasing the number of classes to account for variation may not be helpful at the analytical stage.

Aggregate measures of quality produced by existing testing schemes are not very meaningful as they say nothing about variability of error in space. For large coverages compiled possibly by several individuals, such variability is likely to be an important component in decision making from derivative maps.

Who should hold responsibility for the data? A commonly held view is that users are at the mercy of someone else's data collection. Inadequate metadata or measures of quality leave the user uncertain as to the fitness-for-use of the data. Even with quality measures present, there is no saying how different users may interpret them. The reverse side of the coin is that data collectors have little control over misuse of their data once it is in the GIS (Beard, 1989).

Clearly some quality data is required, but how much? With complex measures of different dimensions of the data, the database may become disproportionately loaded with quality data. After all, given the nature of most GIS data the level of uncertainty in any uncertainty measures is likely to remain high. Spurious accuracy in quality data should be avoided and a more general approach adopted. An informed user should not be committing large resources from decisions based only on maps. Such maps should provide information on likely sites or appropriate scenarios so that limited funds for detailed studies can be deployed with maximum benefit. Such a user requires guidelines as to where potentially suitable sites carry a high risk of abortive work and where potentially unsuitable sites may be usefully explored. In this situation the data must be proactive about its uncertainty over every part of the map, must be capable of propagation and can be interpreted within the specific context.

#### A GENERAL MODEL FOR HANDLING UNCERTAINTY AND FITNESS-FOR-USE IN GIS

A general model for handling uncertainty in GIS is presented in Figure 2. This model is designed with regard to the issues discussed in the previous sections and will hopefully provide a focus and coordinating principle for future research as discussed in the final section.

The overall structure is based on a communication model (Bedard, 1987). It recognizes that data collection is carried out within a specific context and yet observers will have their own view of the real world (W). Observers will either generate uncertainty measures (if appropriate for the data) or at worst be able to verbalize their uncertainty about a number of dimensions of the data. This information about the uncertainty in the data may be global or pertaining to individual objects or entities. It is further recognized that a different metric (or choice of metrics) may be used internally in the GIS to propagate uncertainty during data analysis. There must then be a mapping from the observers stated uncertainty to the propagation metric. This allows flexibility in collecting uncertainty measures at the observation stage, provided a suitable mapping into the internal metric can be found, whilst remembering that high levels of precision in such transformation may be spurious.

Following propagation the resulting metric and its spatial distribution may not be easily intelligible to the user. A second mapping is therefore required so that any visualization fits the user's real world model in the context of the specific task. The user can then assess fitness for use and take responsibility for the data. Sensitivity analysis is possible if the lineage of analysis is stored and on the basis of the results, issue requirements for improved data. This can relate to a specific portion of a coverage if the distribution of uncertainty is known.

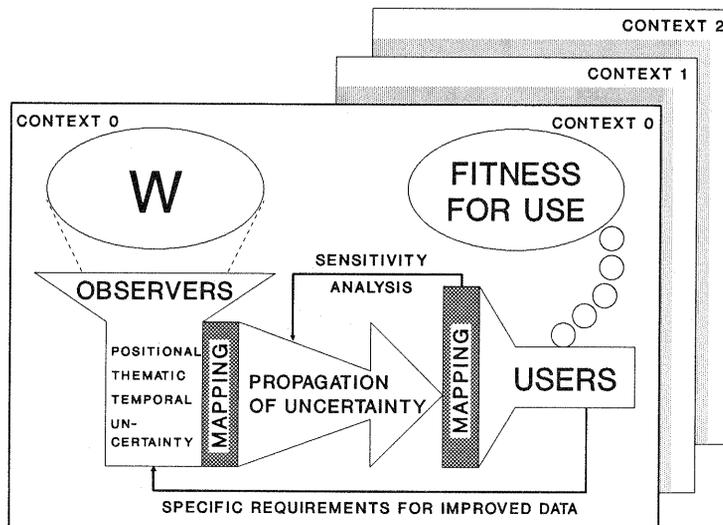


Figure 2: A generalized model for handling uncertainty and fitness for use in GIS.

This model is intended for implementation with existing data structures in GIS software. For vendors it represents evolution rather than revolution.

#### AN EXAMPLE OF USING THE MODEL

Flexibility of the model can best be illustrated through a simple example. A propagation metric is currently being refined that will permit the integration of positional, thematic and temporal uncertainty using fuzzy numbers and fuzzy sets (Zadeh, 1965; Kaufmann & Gupta, 1985). The current example however will be limited to a consideration of thematic or characterization uncertainty.

The propagation metric is summarized in Table 2. These are normal convex fuzzy sets which have been derived through rounding memberships ( $\mu_i | x_i$ ) from triangular fuzzy sets. The simplified fuzzy sets as shown have the important properties that they spread towards the mid-values (i.e. become more fuzzy as one is more uncertain about the data and therefore conform with linguistic use of fuzzy sets) and are a constant orthogonal distance from each other. Thus more complex fuzzy sets resulting from propagation can be unambiguously matched with a simplified fuzzy set. Direct interpretation or constructing representations of verbalizations from fuzzy sets is not always easy. A further feature of these simplified fuzzy sets is that as a series they normalize (i.e. where  $\mu_i=1$ ) from 0 to 1 at an interval of 0.1. Each fuzzy set can therefore be referred to as a number akin to a probability statement (with which there is generally a high level of familiarity) and which to avoid confusion with the literature is called a 'fuzzy expectation' ( $\approx E$ ). Singly or in combination it is possible to map verbalizations about data uncertainty or fitness-for-use in and out of the system.

Figure 3 shows sample categorical data sets, Layer 1 and Layer 2. Each has four categories of data. The observer who collected the data (say, by aerial photographic interpretation) has identified four levels of uncertainty and has assigned each polygon to one of them. Uncertainty that a polygon contains category A may arise either because there is insufficient evidence in the photograph or because the unit contains a level of heterogeneity. The

four linguistic classes of uncertainty have been mapped as  $\approx E$  by the observer (see Table 3). A linguistic class may be cover a range of  $\approx E$  combined with a Boolean OR, or overlap with the adjacent class. Within the GIS data structure, pointers can be made to a lookup table of  $\approx E$  which is a small overhead compared with the 3,628,000 possible fuzzy sets between 0 and 1. In Figure 4 the two layers have been overlaid to form map L1L2 (the numbers in some of the polygons are for reference below). The following Boolean selection is carried out with the conventional binary output from a GIS given in Figure 5:

$$((( A \text{ OR } W ) \text{ AND } \# C ) \text{ AND } \# Y )$$

where # indicates NOT. The uncertainties can be propagated using the same operators. In order to see the results of this, two user's assessment schemes are used (see Table 3) where Scheme B is more stringent in terms of its requirements for fitness-for-use. Figure 6 shows the distribution of

$X_i$

	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1	
f e	0	1										
u x	.1	.9	1	.9								
z p	.2	.5	.9	1	.9	.5						
z e	.3		.4	.8	1	.8	.4					
y c	.4			.3	.7	1	.7	.3				
t	.5				.3	.7	1	.7	.3			
a	.6					.3	.7	1	.7	.3		
t	.7						.4	.8	1	.8	.4	
i	.8							.5	.9	1	.9	.5
o	.9									.9	1	.9
n	1											1

$\mu$

Table 2: Fuzzy expectation,  $\approx E$ , as underlying simplified fuzzy sets.

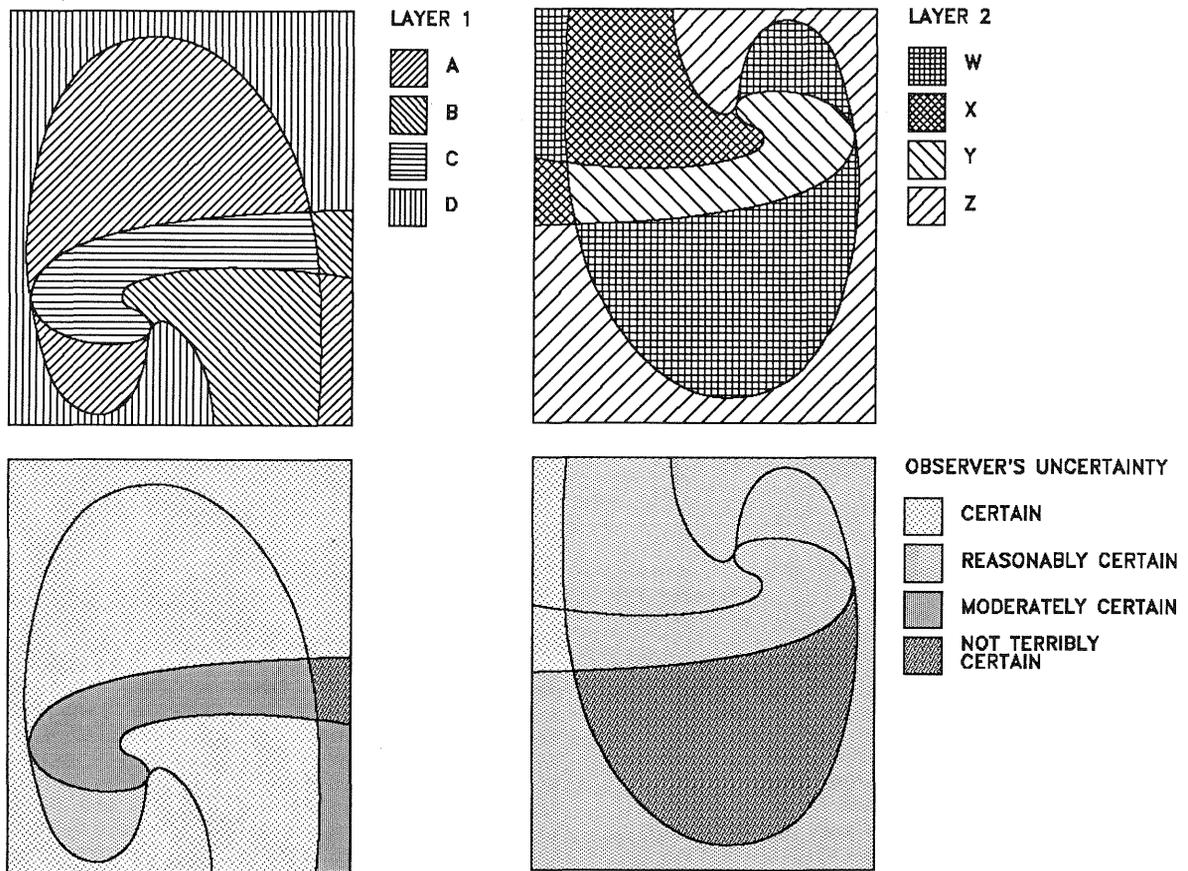


Figure 3: Categorical maps, Layer 1 & Layer 2 with their respective uncertainty displays according to the observer's terminology.

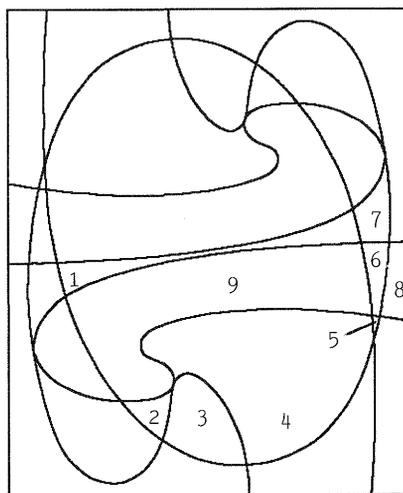


Figure 4: Overlay map L1L2

MAPPING IN OBSERVER'S UNCERTAINTY		MAPPING OUT USER'S ASSESSMENT	
linguistic	≈E	linguistic	≈E
certain	1 OR 0.9	good acceptable unacceptable	scheme A
reasonably certain	0.8		scheme B
moderately certain	0.7 OR 0.6		1 OR 0.9 OR 0.8
not terribly certain	0.6 OR 0.5		0.7 OR 0.6
certain			1 OR 0.9 0.8 OR 0.7 (below)

Table 3: Mappings of observer's uncertainty and user's assessment

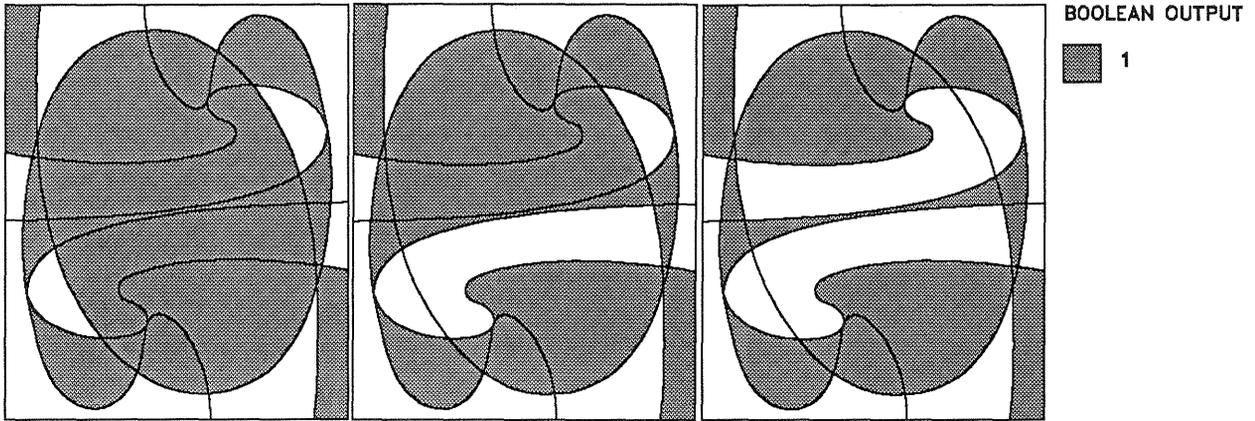
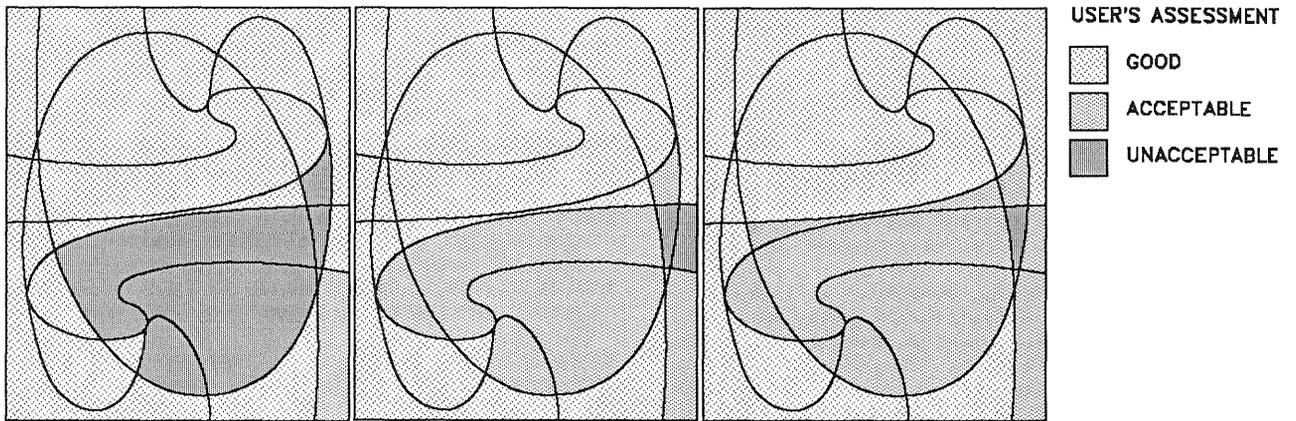
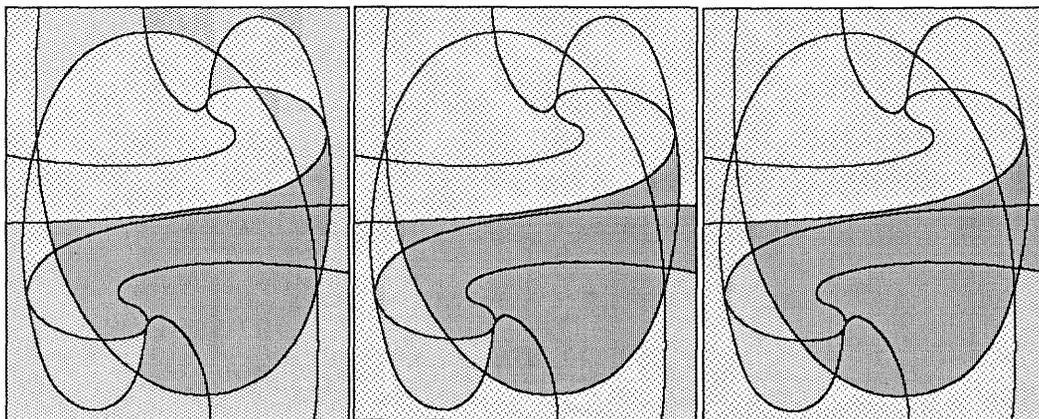


Figure 5: Boolean analysis to identify ((( A OR W ) AND #C ) AND #Y )



USER'S ASSESSMENT SCHEME A



USER'S ASSESSMENT SCHEME B

Figure 6: Uncertainty propagation through the Boolean analysis in the context of the two assessment schemes.

uncertainty in both Schemes for each of the stages in Figure 5. Note how there is an assessment for all polygons regardless of the Boolean outcome in Figure 5. Thus in Scheme A, doubt can be cast on the acceptance of unit 6 and in Scheme B doubt can be cast on both the acceptance of units 1-7 and the rejection of units 8 and 9 in the traditional binary form of the analysis.

The software used in this small example is Genamap which provides lineage tables through the use and recording of active IDs. Table 4 shows the analysis lineage. This type of record would permit sensitivity analysis to quickly recompute the  $\approx E$  scores for changing levels of initial uncertainty and hence identify what decrease in uncertainty is required to improve the end result. Thus for unit 6 in Scheme A, the observer should collect additional data on category B in Layer 1, only in the area of unit 6, so that it can be improved to 'moderately certain' or better from the observer's point of view. For Scheme B, category W in Layer 2 must also be resurveyed so that it is mapped in as 'reasonably certain' or better.

#### IMPLICATIONS FOR GEO-INFORMATION THEORY

The model has a number of implications for geo-information theory. There is a distinct need to improve the ability of GISs to record natural variation and uncertainty in data, not at the generalised level of the coverage but at the level of the individual object or entity. Precise, crisp boundaries and totally homogeneous areal units are in fact a special case in our mosaic of natural environment. Why are we using a tool more suited to these special cases for all our spatial analysis? The point, line and polygon may be inappropriate primitives to record imprecisely bounded objects.

To have proactive data, observers will need to exercise greater responsibility in recording levels of uncertainty and will have to develop strategies in order to do so. Yet the user must take responsibility for assessing fitness-for-use since it is the user who is most familiar with the prevailing context. Sensitivity analysis of data should become a standard routine. Data structures and database design will need to be modified in order to facilitate this. There is the potential for involving expert systems. Current relationships between user and data will undoubtedly change.

Probably the greatest impact of the model is likely to be in the area of research. Ways of combining uncertainty measures will have to be sought. New measures may be developed that can be better embedded within the data. There are a number of candidate metrics that could be used for propagating uncertainty. These will need to be investigated, as will the mapping from source uncertainty measures. The propagation will need to be modelled for a wide range of data transformations. The mapping out process and visualization of uncertainty distributions is already a major concern (Beard et al. 1991). Most of all, the model provides a framework around which whole or partial solutions to the problem of uncertainty in GIS can be built and tested.

#### ACKNOWLEDGEMENTS

This research has been supported by the Hong Kong Polytechnic.

Genamap is a registered trademark of GENASYS II Inc. Fort Collins, Colorado.

#### REFERENCES

- American Society of Civil Engineers, 1983. Map Uses, Scales and Accuracies for Engineering and Associated Purposes. ASCE, New York, USA.
- American Society of Photogrammetry and Remote Sensing, 1985. Accuracy specifications for large scale maps. Photogrammetric Engineering and Remote Sensing 51:195-199
- Beard, K., 1989. Use error: the neglected error component. Proceedings AutoCarto 9, Maryland :808-817
- Beard, K.; Buttenfield, B.P. & Clapham, S.B., 1991. Visualization of spatial data quality. NCGIA Technical Paper 91-26, NCGIA, Santa Barbara, USA.
- Bedard, Y., 1987. Uncertainties in land information systems. Proceedings AutoCarto 8, Maryland :175-184
- Bethel, J.S. & Mikhail, E.M., 1983. On-line quality assessment in DTM. ASPRS/ACSM Annual Convention :576-584
- Blakemore, M., 1983. Generalisation and error in spatial data bases. Cartographica 21 :181-139
- Bureau of Budget, 1947. National Map Accuracy Standards, US Government Printing Office, Washington D.C., USA.
- Burrough, P.A., 1986. Principles of Geographical

ID	Entries	Selection/Criteria			
		Map/ID	Selection	Table	Criteria
1	12	L1L2	ATTRIBUTE	TAGV1	BOUNDED A
5	12	L1L2	ATTRIBUTE	TAGV2	BOUNDED W
10	25	L1L2	ATTRIBUTE	TAGV1	BOUNDED #C
11	25	L1L2	ATTRIBUTE	TAGV2	BOUNDED #Y
12	19	1, 5	SETS	-	OR
13	18	10, 12	SETS	-	AND
14	17	11, 13	SETS	-	AND

Table 4: Simplified extract of the Genamap ACTIVE ID list showing lineage of analysis depicted in Figure 5.

- Information Systems for Land Resources Assessment. Clarendon Press, Oxford, UK.
- Caruso, V.M., 1987. Standards for digital elevation models. ASPRS/ACSM Annual Convention, Vol 4:159-166
- Chrisman, N.R., 1982. A theory of cartographic error and its measurement in digital databases. *Proceedings AutoCarto5* :159-168
- Chrisman, N.R., 1983. The role of quality information in the long-term functioning of a geographic information system. *Cartographica* 21 :79-87
- Chrisman, N.R., 1991. The error component of spatial data, in *Geographical Information Systems Principals and Applications* (eds Maguire et al.), Longman, London, UK. Vol 1:165-174
- Congalton, R.G., 1990. Beware the black box. *Proceedings GIS'90*, Anaheim, USA., Vol 2:851-853
- Davis et al., 1991. Environmental analysis using integrated GIS and remote sensing data: some research needs and priorities. *Photogrammetric Engineering & Remote Sensing* 57:689-697
- Garvey, T.D., 1987. Evidential reasoning for geographic evaluation for helicopter route planning. *IEEE Transactions on Geoscience and Remote Sensing* GE-25:294-304
- Goodchild, M.F. & Gopal, S., 1989. *The Accuracy of Spatial Databases*. Taylor & Francis, London, UK.
- Heuvelink, G.B.M.; Burrough, P.A. & Stein, A., 1979. Propagation of errors in spatial modelling with GIS. *International Journal of Geographic Information Systems* 3:303-322
- Kaufmann, A. & Gupta, M.M., 1985. *Introduction to Fuzzy Arithmetic*. Van Nostrand Reinhold, New York, USA.
- Lanter, D.P., 1990. Lineage in GIS: the problem and a solution. NCGIA Technical Paper 90-6, NCGIA, Santa Barbara, USA.
- Lanter, D.P. & Veregin, H., 1991. A lineage information program for exploring error propagation in GIS applications. *Proceedings 15th International Cartographic Association Conference*, Bournemouth, Vol 1:468-472
- Laurini, R. & Thompson, D., 1992. *Fundamentals of Spatial Information Systems*, Academic Press, London, UK
- Lee, T.; Richards, J.A. & Swain, P.H., 1987. Probabilistic and evidential approaches for multisource data analysis. *IEEE Transactions on Geoscience and Remote Sensing* GE-25:283-292
- Leung, Y., 1988. *Spatial Analysis and Planning Under Imprecision*. North-Holland, Amsterdam, Netherlands.
- MacDougall, E.B., 1975. The accuracy of map overlays. *Landscape Planner* 2:23-30
- Maguire, D.J.; Goodchild, M.F. & Rhind, D.W., 1991. *Geographical Information Systems Principals and Applications*. Longman, London, UK.
- Mead, D.A., 1982. Assessing data quality in geographic information systems, in *Remote Sensing for Resource Management*, Soil Conservation Society of America, Iowa :51-62
- Medyckyj-Scott, D., Newman, I., Ruggles, C. & Walker, D., 1991. *Metadata in Geosciences*. Group D Publications, Loughborough, U.K.
- Middelkoop, H., 1990. Uncertainty in GIS: a test of quantifying interpretation input. *ITC Journal* 1990(3): 225-232
- Miller, R.G.; Karimi, H.A. & Feuchtwanger, M., 1989. Uncertainty and its management in geographic information systems. *Proceedings CISM National Conference on GIS*, Ottawa :252-260
- Moellering, H., 1988. The proposed standard for digital cartographic data. *The American Cartographer*, 15:9-140
- Newcomer, J.A. & Szajgin, J., 1984. Accumulation of thematic map errors in digital overlay analysis. *The American Cartographer* 11:58-62
- Peucker, T.K., 1976. A theory of the cartographic line. *International Yearbook of Cartography*, 16:134-143
- Pullar, D., 1991. Spatial overlay with inexact numerical data. *Proceedings AutoCarto 10*, Baltimore :313-329
- Rosenberg, D.J., & Martin, G. 1988. Human performance evaluation of digitizer pucks for computer input of spatial information. *Human Factors*, 30(2):231-235
- Shafer, G., 1976. *A Mathematical Theory of Evidence*. Princeton University Press, USA.
- Stoms, D.M., 1987. Reasoning with uncertainty in intelligent geographic information systems. *Proceedings GIS 87*, San Francisco, Vol. II:693-700
- Story, M. & Congalton, R.G., 1986. Accuracy assessment: a user's perspective. *Photogrammetric Engineering & Remote Sensing*, 52:397-399
- Theobald, D.M., 1988. Accuracy and bias in surface representation. *First Special Meeting on the Accuracy of Spatial Databases*, NCGIA, Santa Barbara, USA.
- Tobler, W., 1990. GIS transformations. *Proceedings GIS'90*, Vol 1:163-166
- Veregin, H., 1989a. Accuracy of spatial databases: annotated bibliography. NCGIA Technical Paper 89-9, NCGIA, Santa Barbara, USA.
- Veregin, H., 1989b. A taxonomy of error in spatial databases. NCGIA Technical Paper 89-12, NCGIA, Santa Barbara, USA.
- Veregin, H., 1989c. Error modelling for the map overlay operation, in *Accuracy of Spatial Databases* (eds Goodchild & Gopal), Taylor & Francis, London, UK. :3-18
- Vitek, J.D.; Walsh, S.J. & Gregory, M.S., 1984. Accuracy in geographic information systems: an assessment of inherent and operational errors. *Proceedings PECORA IX Symposium* :296-302
- Walsh, S.J.; Lightfoot, D.R. & Butter, D.R., 1987. Recognition and assessment of error in geographic information systems. *Photogrammetric Engineering & Remote Sensing* 53:1423-1430
- Wang, F.; Hall, G.B. & Subaryono., 1990. Fuzzy information representation and processing in conventional GIS software, database design and implications. *International Journal of Geographic Information Systems* 4:261-283
- Zadeh, L.A., 1965. Fuzzy sets. *Information & Control* 8(3):338-53
- Zhang, G. & Tulip, J., 1990. An algorithm for the avoidance of sliver polygons in spatial overlay. *4th International Symposium on Spatial Data Handling*, Zurich, Vol 1:141-150