MANIPULATING UNCERTAINTY-BASED DIGITAL MAPS TO PERFORM LIKE POLYGON-BASED THEMATIC MAPS

Dr. Kim Lowell
La Chaire industrielle en géomatique appliquée à la foresterie
Centre de recherche en géomatique
Université Laval, Pavillon Casault
Ste-Foy, Québec G1K 7P4 CANADA
(418) 656-7998
FAX: (418) 656-7411
Kim.Lowell@for.ulaval.ca

Abstract
Spatial uncertainty has become a topic of considerable interest in recent years. Most research has focused upon descriptions, models, and quantification of spatial uncertainty. Little work has considered what the incorporation of uncertainty into a GIS will mean for end-users of GIS. Using a specific example and spatial operators adapted from fuzzy set theory, this paper describes how users will be able to manipulate cartographic data in an uncertainty-based GIS. It is concluded that the use of uncertainty in GIS will give a user more control over analysis and end products, but will also require that a user take more responsibility for the analysis and the communication of results to others.

1 Introduction

In recent years, there has been a considerable amount of interest in the subject of spatial uncertainty. Some individuals have proposed cartographic error models (Chrisman 1982) while others have distinguished between error analysis and sensitivity analysis [9] . Based on definitions by the latter, error analysis concerns the simplification and input of geographical data into a digital geographic information system (GIS) whereas sensitivity analysis imposes perturbations on geographical phenomena based on one or more underlying assumptions about the nature of spatial error. [8] and [6] are examples of using algorithmic perturbations of deterministic maps based on an underlying assumption concerning the level of spatial autocorrelation among errors. Still others have worked on the quantification of spatial error and/or cartographic simplification. [2] studied the magnitude of digitizing errors -- i.e., converting a paper map to a digital map. [10], [5], and [1] have focused on quantifying real-world-to-map uncertainty -- i.e., cartographic simplification.

In the work that has been conducted on spatial uncertainty, little attention has been paid to how such work will affect end users of GIS. That is, if one supposes that future GIS will be uncertainty-based[1], how will this change user input and the responsibilities of users? How will the incorporation of spatial uncertainty into GIS change the results of digitally based cartographic analysis? How can the results of such analysis be interpreted and communicated to others? The purpose of this paper is to address these questions by using real maps as an example.

2 Data Base

The study area for this work is Montmorency Forest -- the research forest of Laval University -- located 80 km north of Quebec City. For this study, a map of forest types produced by trained aerial photograph interpreters was digitized, as was a map of soils types for the same area; a subset of these maps were employed in this study. Forests and Soils are two phenomena for which it is known that maps are extremely imprecise -- i.e., different cartographers may put type boundary lines in different places or completely omit boundaries found by another cartographer. Thus it is appropriate to represent a given polygon for either of these phenomena as a "fuzzy" field rather than a "solid" polygon having crisp boundaries. On the digitized maps, 20 forest types and 11 soil types were present.

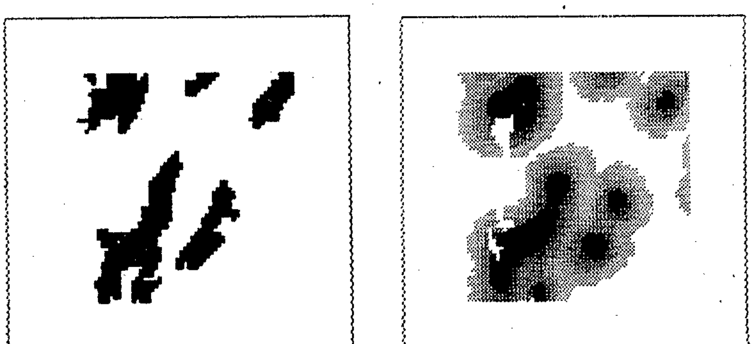
[1] Although many people employ the word "error," the word "uncertainty" is preferred here. It is the belief of the author that "error" implies that some absolute ground-truth exists whereas in the author's domain of forestry, the existence of real ground "truth" is a proposition defended only with great difficulty. "Uncertainty" by contrast does not imply the existence of some attainable truth and reflects the idea that the only "truth" that exists is that which is found by a large number of individuals.

From these maps, fuzzy surface representations were developed using a process developed by [10]. Briefly, one skeletonizes a polygonal map so that a surface is populated by features -- points or lines -- which are considered to have "100% certainty" of being a given map type (points) or definitely having two different types on either side (lines). This skeletonizing is done subjectively -- i.e., a user decides which features have absolute certainty -- but it can be automated by techniques such as inverse buffering of polygons to retain their cores and/or by preserving boundaries of types such as clearcuts or deep lakes known to have precise boundaries. Once the map skeleton is identified, spatial interpolation is conducted for each type. That is, if Type A is considered to be present at a point with a certainty of 1.00, this means that the possibility of having Type B (and all other types) at this point is 0.00. Thus for Type A (and B and C...) the surface is populated by features having values of either 1.00 or 0.00. Interpolation for Type A will fill the surface with intermediate values. In this study, linear interpolation was employed, but there is no reason that other functional forms -- e.g., gaussian -- could not be employed.

It is noted that fuzzy surfaces such as those described can be obtained by other methods. [1] and [5] have worked on the quantification of photo-interpretation errors as well as the form of the distribution of such errors across type boundaries; these errors were found to be distributed in a gaussian fashion. Thus one could generate error distributions around a mean line using well-defined statistical parameters; these in turn can be converted to probabilities using cumulative frequency distributions. Another way to produce such surfaces involves employing a simulation technique such as described in [8] or [6] to "perturb" a map although such perturbations require that assumptions about the value of non-intuitive parameters be made -- e.g., the statistic *rho* which is a measure of spatial autocorrelation.

In this paper, the relative merits of alternative ways of obtaining fuzzy surfaces are not at issue. What is important is understanding that for the purposes of this paper, one will have a fuzzy surface representation available represented in the raster data structure. Each raster cell contains a set of k values between 0 and 1 each representing the certainty of having one of the k types present at that location. The certainty values sum to 1.0 over all k types. Perhaps an easier way to visualize this -- rather than a set of cells each having k values -- is to think of it as a set of k surfaces (Fig. 1). There is one surface for each of the k categories and a given surface shows the certainty (not necessarily the probability) of having a particular type at a particular location. In this study, these surfaces were derived from an interpolation process from features having high certainty and are intended to represent real-world-to-map uncertainty. However, these surfaces could also be derived from perturbing processes, remote sensing preclassification information, or other techniques. Moreover, they could be representative of digitizing errors, positioning errors, and/or other types of cartographic uncertainty. The point of this paper is not to select the best method for generating such surfaces nor to quantify and/or describe all possible types of cartographic error. Rather the purpose of this paper is to show, given a digital uncertainty representation, how these can be manipulated in a GIS, what is required of a user to do so, and how the results can be treated.

Figure 1. Example of deterministic and fuzzy map type.



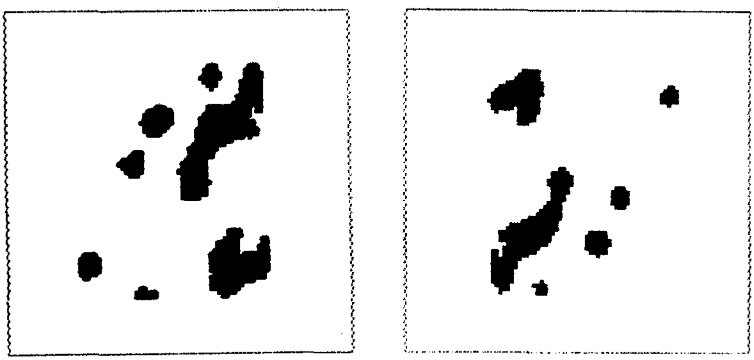
3 GIS Single Map Manipulations

3.1 What (categorical) attribute is "here?"

One of the most employed GIS queries for a single map is "Where will I find Type *k*?" In conventional GIS, this query is resolved by searching the data base for all the polygons that have an attribute of *k* recorded for the variable Type, and coloring the polygons with a particular colour. In the case of an uncertainty-based GIS, one does not have a set of polygons and one cannot say something is definitely Type *k*. (The sole exception is when the certainty value for *k* is 1.0 and the certainty value for all other types is 0.0; this special case is not considered here.) Thus in an uncertainty-based GIS, it is necessary to find another way of determining where Type *k* is present.

The most obvious way is to check if Type *k* has the highest certainty value for each cell. The GIS can then highlight all those cells for which this condition is met (Fig. 2a). The danger in this process is that one may have a cell with values of 0.34, 0.33, and 0.33 for Types *k*, *l*, and *m*, respectively. A maximum-likelihood assignment such as that described would label such a cell as Type *k* even though there is considerable confusion about what type is dominant on this cell. An alternative technique for determining "*k*-ness" would be to specify some threshold for certainty values for being Type *k* -- an obvious choice would be 0.50 (Fig. 2b). Still another alternative might be to calculate the entropy (Phipps 1981) of each cell (a measure of variability) and ask to know which cells have a maximum-likelihood of being *k* and also high entropy. (High entropy indicates high variability.) This would distinguish between cells having values of [0.21, 0.19, 0.20, 0.20, 0.20] (low entropy) and [0.40, 0.15, 0.15, 0.15, 0.15] (high entropy). In neither case is the maximum certainty value greater than 0.50, but in the second case there is considerably more likelihood that the cell is actually the first type on the list. Finally, another way to determine *k*-ness would be to assign cells to a given type based both on certainty values and a user-specified amount of spatial autocorrelation. This would be useful if one is working with satellite images, for example. In such cases, maximum-likelihood classifications and/or even entropy evaluation would probably lead to a considerable amount of salt-and-peppering on the final classified image. Specification of the amount of spatial autocorrelation would cause pixels identified as Type *k* to be more grouped as low certainty values could still lead to a categorization of *k* if surrounding pixels were definitely *k*. This notion is in a accord with the concept prevalent in cartography of the presence of polygons on maps and not isolated occurrences of a given type.

Figure 2. Thresholding of fuzzy surfaces.

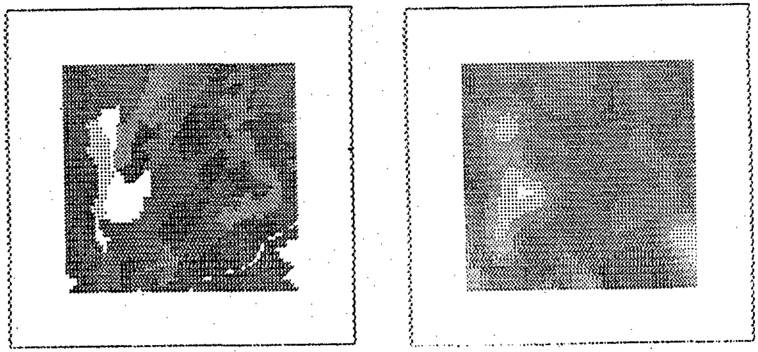


a. Maximum likelihood of having *k*. b. > 50% likelihood of being *k*.

3.2 *"What (continuous) attribute is "here?"*

Another common query concerns the spatial distribution of a continuous variable/attribute. For example, one may want to know which are the counties that have a population larger than a certain threshold. In the case of conventional GIS, such a query is handled in exactly the same way as a categorical request: one searches a data base for all the polygons which respond to the threshold specified for the attribute of interest (Fig. 3a). And, indeed, in uncertainty-based GIS, one may adopt the same approach. One may classify a cell into a category using any of the methods described in the previous section, and then search the attributes for each of these types to determine which types (and therefore cells) respond to a given criterion.

Figure 3. Basal area maps.



a. Deterministic. b. Continuous

An alternative way would be to calculate the value for a cell using the certainty values as indicative of a mixture of the types present at a location. Thus given certainty values [0.5, 0.3, 0.2] for types having attributes of [10, 20, 30] one would estimate a value of $(.5*10 + .3*20 + .2*30 =)$ 17 instead of the GIS classifying the cell as the first type and then responding with a value of 10. As before, one then determines which cells respond to the specified criterion. While such an approach allows a continuum of values (Fig. 3b) it rests on the assumption that the certainty values -- however obtained -- are somehow indicative of sub-cell resolution. [7] and [11] found only limited relationship between pre-classification probabilities and sub-pixel resolution with remotely sensed images. And what little work has been done on this subject for categorical maps has shown limited success with such an approach. [10] realized some success with forest species composition with the methods employed here for generating fuzzy surfaces, but almost none for estimating basal area. The author is aware of no studies in which polygon perturbation was used to generate fuzzy surfaces for which resulting surfaces were checked against ground-truth for either full cell or sub-cell resolution.

3.3 *Summary of Single Map Requests*

It is apparent that if uncertainty is included in GIS data bases, users will have to take more responsibility for the results of spatial analysis. In particular, users will have the responsibility of determining how single maps are to be summarized. To do this, a number of relatively simple algorithms for the assignment of values has been presented, and it is certain that as uncertainty becomes an increasingly integral part of GIS, others will be developed. Users will also have to accept the responsibility of determining what is a "reasonable certainty" of having a particular type or attribute value at a given location. The advantage of such responsibility is that the user will have more control over the system and the analysis. Presently, for natural resource maps which are produced from interpretations of phenomena which have no absolute ground truth, GIS only gives users information about what is the most likely map type or attribute value that one will have at a given location. The incorporation of uncertainty in a GIS will provide users with the ability to determine what level of certainty is necessary before one is willing to accept that a given type or attribute value does or does not occur at a particular location.

4 GIS Multiple Map Manipulations

4.1 Fuzzy Operators

One of the principle reasons that GIS has gained popularity in recent years is its ability to integrate geographic data from different sources. This means that maps of various phenomena constructed at different scales and using different map projections can be readily overlain and summarized. Map combination such as this is most often thought of in terms of "polygon overlay." Thus if one has a cartographic representation which is not based on polygons such as an uncertainty-based GIS, one must question what sort of overlay is possible.

Uncertainty-based GIS can be considered a variant of fuzzy set theory [13] in which an object (a location) is not assigned in a Boolean fashion to one single map type. Instead a location has a likelihood of belonging to more than one type. As such, one may employ operators for fuzzy sets to conduct spatial overlay. In this paper the fuzzy operators of Hard and Soft ANDs and ORs [13] (Table 1) are demonstrated although other fuzzy operators have been described [3]. ANDs are the intersection of two sets (a location is underlain by Forest A AND Soil 1) whereas ORs give the union of sets (a location is underlain by Forest A OR Soil 1).

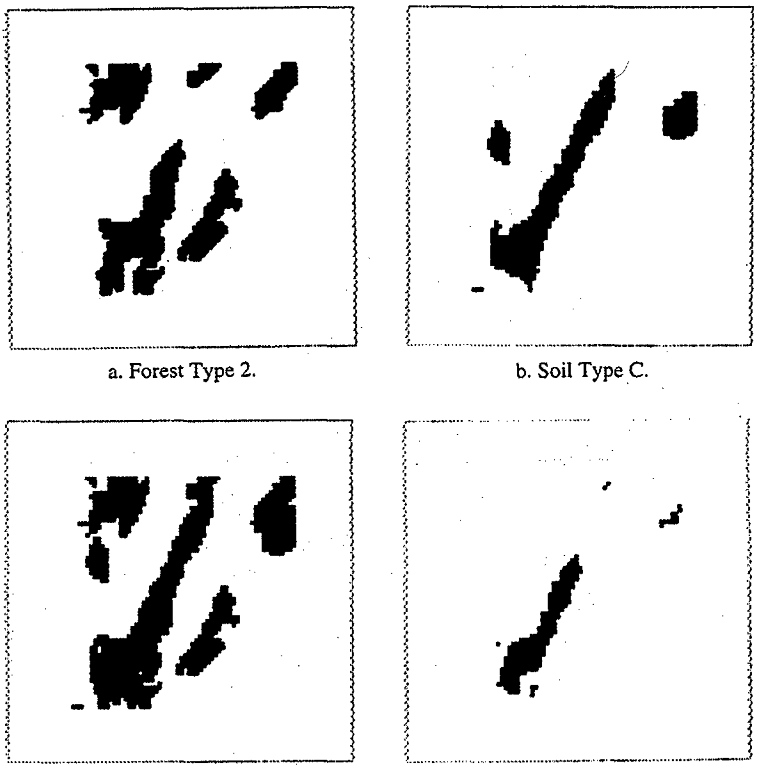
Table 1. Summary of fuzzy spatial operators.

Name	Function
Hard OR (Union)	Maximum$(f_1,f_2,...f_n)$
Soft OR (Union)	Minimum$(1, f_1+f_2+...+f_n)$
Hard AND (Intersection)	Minimum$(f_1,f_2,...f_n)$
Soft AND (Intersection)	$(f_1 \mathrm{X} f_2 \mathrm{X}...\mathrm{X} f_n)$

For the purposes of this paper. Forest Type 3 (FT3) and Soil Type C (STC) are used. Fig. 4 shows the deterministic representation of each as well as the result of a Boolean AND and Boolean OR. Figure 5 shows the fuzzy surface representation of each as well as a Hard AND and Hard OR. As can be seen, the Boolean overlays -- particularly the AND -- generally yield small and disjoint areas. Depending on the motivation for an overlay, a Boolean map may provide little useful information to a GIS user. For example, if one is searching for not only the FT1/STC combination, but also a minimum contiguous amount of area then the Boolean map has identified too few candidate areas. That is, one has no information about what map type combinations may be found around each of the areas identified even though the phenomena Forests and Soils have boundaries that are not precise. The maps resulting from the fuzzy operators, conversely, provide a user with more information in that they show not only where there is a great likelihood of finding the type combination sought, but also how this likelihood degrades across the surface. Moreover, if this map does not provide enough information or candidate areas, one may employ a Soft AND or OR to have a more liberal interpretation of where a given type combination may exist.

Once an uncertainty-based overlay has been conducted, it remains to communicate the results to users of the analysis. In Boolean map overlay, this is not a problem since polygon maps resulting from deterministic overlays are readily understood. The most direct way to do this with fuzzy maps is to employ thresholding to convert fuzzy to deterministic maps. Note that this gives the user still more flexibility in the analytical process. One may choose a relatively high or low threshold for acceptance of the existence of a given combination depending on what one wishes to communicate. For example, Figure 6 shows the Hard AND of Fig. 5d thresholded at 0.25 and 0.75. Depending on the criticality of having this combination at a given location, one can accept any of these (or other) thresholds.

Figure 4. Deterministic overlay.



a. Forest Type 2. b. Soil Type C.

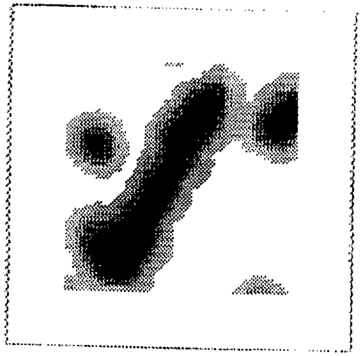
c. OR (Union). d. AND (Intersection).

4.2 *Summary of Multiple Map Operations*

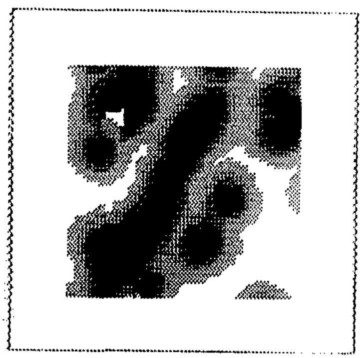
The use of an uncertainty-based GIS provides users more control over the overlay process than does Boolean overlay as there are a number of fuzzy operators which can be used to conduct overlay. This flexibility, of course, also requires that a user assume more responsibility for the analysis and its results. It must also be recognized that this flexibility will increase the potential for abuse by map makers. One can very easily envision a situation in which two groups on opposite sides of a social/political issue wish to show that a particular natural habitat is growing or shrinking. One group can employ a certain fuzzy operator and threshold value while the other employs a different one to "prove" a particular conclusion. Map consumers may thus be left with a baffling array of choices as to which is "correct." While this reflects a more realistic way of interpreting maps -- i.e., accepting that maps are but a representation of reality requiring human interpretation and not reality itself -- it puts more onus on map consumers to understand the map analysis process and its results.

Figure 5. Fuzzy overlay.

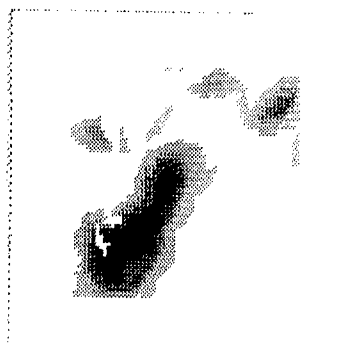
a. Forest Type 2.



b. Soil Type C.

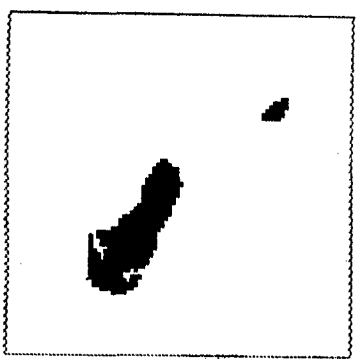


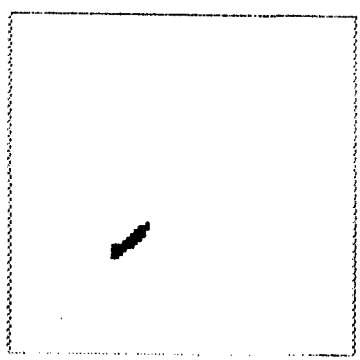
c. Hard OR (Union).



d. Hard AND (Intersection).

Figure 6. Thresholding of Hard AND (Fig. 5d).





5 Summary and Conclusions

Employing an uncertainty-based GIS will require considerably more input from increasingly sophisticated users. The advantage of this is that users will have more control over the analytical process and its results. There are, of course, a number of disadvantages. The first is that the information needed to fuel such GIS is not available and there appear to be few studies being conducted to obtain information on most types of cartographic uncertainty. Of the processes that are available, some require a certain amount of user subjectivity -- e.g., in skeletonizing a map -- while others require non-intuitive information that may never be widely available -- e.g., knowledge of the amount of spatial autocorrelation of errors for a particular phenomenon and/or map.

6 Literature Cited

[1] Aubert, E., 1995. Quantification de l'incertitude spatiale en photointerprétation forestière à l'aide d'un SIG pour le suivi spatio-temporel des peuplements. M.Sc. Thesis. Université, Laval, Québec, Canada. 92 pp.

[2] Bolstad, P., Gessler, P., Lillesand, T., 1990. A variance components analysis of manually digitized map data. Surveying and Land Information Systems 50:201-207.

[3] Burrough, P., 1989. Fuzzy mathematical methods for soil survey and land evaluation. Journal of Soil Science 40:477-492.

[4] Chrisman, N., 1982. Methods of Spatial Analysis Based on Error in Categorical Maps. Ph.D. dissertation. University of Bristol, England. 261 pp.

[5] Edwards, G., Lowell, K., 1995. Modeling uncertainty in photointerpreted boundaries. Photogrammetric Engineering and Remote Sensing (in press).

[6] Fisher, P., 1992. First expeiments in viewshed uncertainty: simulating fuzzy viewsheds. Photogrammetric Engineering and Remote Sensing 58:345-352.

[7] Fisher, P., Pathirina, S., 1990. The evaluation of fuzzy membership of land cover classes in the suburban zone. Remote Sensing of the Environment 34:121-132.

[8] Goodchild, M., Guoqing, S., Shiren, Y., 1992. Development and test of an error model for categorical data. International Journal of GIS 46:87-104.

[9] Lodwick, W., Monson, W., Svoboda, L., 1990. Attribute error and sensitivity analysis of map operations in geographical information systems: suitability analysis. Int'l J. of GIS 4:413-428.

[10] Lowell, K., 1994. A fuzzy surface cartographic representation for forestry based on Voronoi diagram area stealing. Canadian Journal of Forest Research 24:1970-1980.

[11] Marsh, S., Switzer, P., Kowalik, W., Lyon, R., 1980. Resolving the percentage of component terrains within single resolution elements. Photo. Engineering and Remote Sensing 46:1079-1086.

[12] Phipps, M., 1981. Entropy and community patterns analysis. J. of Theoretical Bio. 93:253-273.

[13] Zadeh, L.A., 1965. Fuzzy Sets. Information and Control 8:338-353.