A Methodology for Reporting Uncertainty in Spatial Database Products

Gary J. Hunter, Mario Caetano and Michael F. Goodchild

The term 'uncertainty' is used to refer to differences between the information provided by a spatial database, and the corresponding information that would be available to someone able to observe and measure the real world directly. It includes the effects of errors made during creation of the database, as well as those of the information loss that occurs during generalization. A general model of spatial data uncertainty is presented, and examples of its applications are described. The model forms the foundation for a general approach to handling uncertainty in the application of spatial databases and GIS. The use of the model is illustrated with a simple example of the analysis of a wildfire in a remotely sensed image. The approach allows uncertainty to be modeled and visualized, and its effects on the results of analysis to be simulated and evaluated.

We experience difficulties in articulating the quality of information represented in a database principally because we don't understand how to analyze data based on information about its qualities. Even if a database were to provide us with a feature's quality attributes along with its shape and topology, our prevailing tools are usually incapable of modeling most of its properties, and those procedures we do have are pathetically crude . . .

(Dutton 1984, p. 276).

Dr. Gary Hunter is an assistant professor with the Department of Geomatics at the University of Melbourne, Australia. He teaches undergraduate and postgraduate courses in land law and development, cadastral surveying, spatial data algorithms, and geographic information systems. His paper on spatial database accuracy, co-authored with Michael Goodchild, won the Horwood Critique Prize at URISA '93 in Atlanta.

Mario Caetano is a graduate student researcher at the Centro Nacional de Informacao Geografica of Portugal. He recently completed his M.A. degree in geography at the University of California, Santa Barbara, and is continuing his Ph.D. work on burned vegetation mapping in Mediterranean landscapes with remote sensing data at CNIG in Lisbon.

Dr. Michael Goodchild is a professor of geography at the University of California, Santa Barbara, and director of the National Center for Geographic Information and Analysis. He was editor of *Geographical Analysis* for several years and currently serves on the board of six journals and book series. He is well known for his interest in quality and accuracy issues in GIS, and is editor of the book *Accuracy of Spatial Databases*.

While researchers have invested considerable resources examining the modeling and communication of uncertainty in spatial database products, the results of their labor will not be recognized until the user community can apply the techniques in everyday, operational situations. Thus, the debate about uncertainty has now reached the stage where there is a critical need for tools to be developed to assist users in better understanding the outputs derived from their systems, as acknowledged in the quotation given above by Dutton (1984). Through such knowledge they will be placed in an improved position to make data quality assessments, by comparing the quality of their products against the requirements of the tasks for which the products are to be used.

Before discussing the various options available for dealing with this problem, some explanatory remarks are required regarding the use of the term "uncertainty." In general terms, uncertainty denotes a lack of sureness or definite knowledge about an outcome or result, and synonyms include "doubt" (a lack of certainty witnessed by the inability to make a decision), "dubiosity" (a vagueness or conceptual confusion), "skepticism" (implying a lack of faith or trust in the reliability of something), and "mistrust" (a genuine belief based upon suspicion).

In the context of spatial databases, the authors suggest there is a clear distinction to be made between "error" and "uncertainty," since the former implies that some degree of knowledge has been attained about differences (and the reasons for their occurrence) between the results or observations and the truth to which they pertain. On the other hand, "uncertainty" conveys the fact that it is the lack of such knowledge which is responsible for hesitancy in accepting those same results or observations without caution, and often the term "error" is used when it would be more appropriate to use "uncertainty."

It is well known that there are many potential sources of error in spatial databases (Hunter and Beard 1992), but because so little is understood about the way in which those errors (either singly or in conjunction with each other) affect the outcome of the final products (be they displays, maps, graphs or reports), there is a resultant uncertainty concerning the level of trust which should be placed in them. In some ways, the distinction between error and uncertainty is analogous to the legal belief that a person is "innocent until proven guilty," since in many cases conceptual models of spatial database error simply do not exist. It is suggested that until that situation improves, "uncertainty" offers a more appropriate means of describing such lack of proof. This does not mean that uncertainty should always be substituted for error, as there already exist several well-established and accepted error models for given spatial operations. They are properly described as such, however in situations where there is little knowledge of the actual errors involved—as in the case study described—it is uncertainty which will be referred to by the authors.

One of the largest sources of uncertainty in spatial databases is a byproduct of the process of cartographic generalization. For example, a map may show an area as having a uniform land cover class, even though it is known that the land-cover class in the area is not in fact uniform. This leaves the user of the database uncertain as to the actual land-cover class to be found at a specific point within the area. A very naive user who is unaware of the nature of cartographic generalization might see such a difference between database contents and ground truth as error.

At this time, there are three options available (Goodchild, Lin, and Leung 1993) for dealing with uncertainty in spatial databases, and communicating such information to users, *viz.*:

1. omit all reference to it,
2. attach some form of descriptor to the output,
3. show samples from the range of possible maps.

The first option ("do nothing" approach) treats the problem by ignoring it; undoubtedly the easiest solution to adopt, but one which potentially places at risk the reputations of decision-makers (and their agencies) who have to act on the basis of such information. The second option would see the use of descriptors such as epsilon bands, misclassification matrices, reliability diagrams, and root mean-square error estimates. In effect, these are a caveat to users and while they give warnings about product uncertainty, they provide little assistance in showing how the resultant output might vary spatially. Although, with further development they can be more usefully interpreted, as Hunter and Goodchild (1995a) have shown in the case of the root mean-square error estimate for digital elevation models (DEMs). Finally, different versions of the same map might be presented to users to illustrate the uncertainty to which their products are subject due to the particular combination of data, error estimates, algorithms and process models which have been chosen for the task.

This latter approach is the one preferred by the authors, since it would appear to have the greatest potential benefit in both communicating uncertainty and at the same time educating the user community in the significance of this issue. Accordingly, this paper presents a methodology which permits uncertainty reporting for certain types of spatial database products. By presenting the level of uncertainty which resides in an output, such a methodology might assist agencies in determining the degree of uncertainty they are willing to tolerate before it either changes the decisions made on the basis of that information, or else (in the worst case) causes the benefits of spatial database usage to be lost. In the reverse role, the methodology could provide advance testing of different combinations of data, error estimates, algorithms and models to assess which ones are most likely to suit a user's needs.

At this stage, the methodology is restricted to the study of grid-cell data and, specifically, the outputs derived from the use of DEMs; however, even in this limited role it has considerable relevance to natural resource and environmental applications where the raster data model has greater suitability for representing inherently continuous variation. In addition, the raster model more easily accommodates simulation techniques such as those used in this research. The paper discusses:

1. the underlying model of uncertainty employed,
2. its potential applications,
3. how the model can be integrated into an overall methodology to handle uncertainty, and finally
4. a case study of its use relating to the adjustment of remotely sensed data for topographic effects, as applied to the detection of burnt land in mountainous areas of central Portugal.

The Underlying Model of Uncertainty

The basis of the approach is a model of the uncertainty present in a spatial database. While the database may indicate that a point has some characteristic value, such as its elevation or land cover class, in general it is clear that the value recorded in the database may not be the true value. The amount of uncertainty is sometimes known—for example, producers of DEMs often publish estimates of root mean square error, and producers of land cover maps may provide information on misclassification probabilities. In such cases it is possible to model uncertainty and its effects by introducing random variation. Where the amount of uncertainty is unknown, it is possible to introduce different amounts of random variation to explore their effects. However, the spatial nature of the data requires that the random variation be spatially autocorrelated, and thus demands the use of specialized techniques.

An error model can be defined as a mechanism for introducing random variation in order to represent error or uncertainty. Any one execution of the mechanism creates one pattern of distortion, and thus one sample from a population of possible patterns of distortion. The term for such a single execution of the mechanism is a 'realization' of the error model. One realization might represent the errors introduced by one person's effort to digitize a map, or the uncertainty generated by one cartographer's generalization. A sample of realizations might represent the variability due to different foresters' interpretations of the same aerial photograph. When several realizations are displayed rather than a single map, the effect is to convey a sense of uncertainty. In addition, realizations can be used to investigate the effects of uncertainty on the results of GIS analysis due to error propagation. One error model for categorical spatial data has been described by Goodchild, Sun and Yang (1992).

The traditional Gaussian model (where the standard deviation is a measure of error or uncertainty) is useful for modeling variation in single measurements, but cannot be used to deal with the spatial case where errors display strongly correlated patterns. The error at one point in a DEM, for example, is likely to be strongly correlated with errors at neighboring points. In a previous paper, Hunter and Goodchild (1995b) argued that while it is possible to perturb a data set according to an error descriptor (such as an RMSE value for a DEM) without consideration of spatial autocorrelation between point sample elevations, the process may be stochastic but nevertheless lacks 'truthfulness'—since adjacent elevations can be severely distorted creating large pits and peaks which do not intuitively occur at the resolution of a 30m $\times$ 30m grid. This approach produces what are known as "random maps."

On the other hand, assumption of complete spatial dependence between neighboring points produces realizations of the DEM which are "truthful" but not stochastic, since elevations are constrained to maintain their relative differences to each other and the introduction of a noise value has the effect of moving all DEM elevations up or down by a constant amount. Hence, there is a need to find the appropriate pattern of spatial dependence for any particular application. For the model used in this paper, and by Goodchild, Sun and Yang (1992), spatial dependence is described by a single parameter r in the range $0 < \rho < 0.25$, which meets the dual requirements of being stochastic as well as "truthful." The limit of 0.25 arises because a raster is used, and spatial dependence is defined by the relationship between a cell's value and those of the four neighbors with which it shares a common edge (Cliff and Ord 1981, p. 147). This is perhaps the simplest possible model of spatial dependence. If additional information were available about the actual spatial dependence between errors in a given application, this simple model could be replaced by a more accurate version.

By producing distorted versions of the DEM for different ρ values, and by studying the change in differences between the realized data and the original DEM, it is possible to make reasonable deductions as to what the appropriate ρ values may be, or at least to place limits on the range, and to gain insight into the effects of various levels of spatial dependence on the outcomes of GIS analysis. Hunter and Goodchild (1995b) derived separate realizations of slope gradient and aspect values from a DEM with the latter, in particular, showing a marked change in response at approximately $\rho = 0.24$, while slope gradient only started to vary significantly from $\rho = 0.20$ onwards.

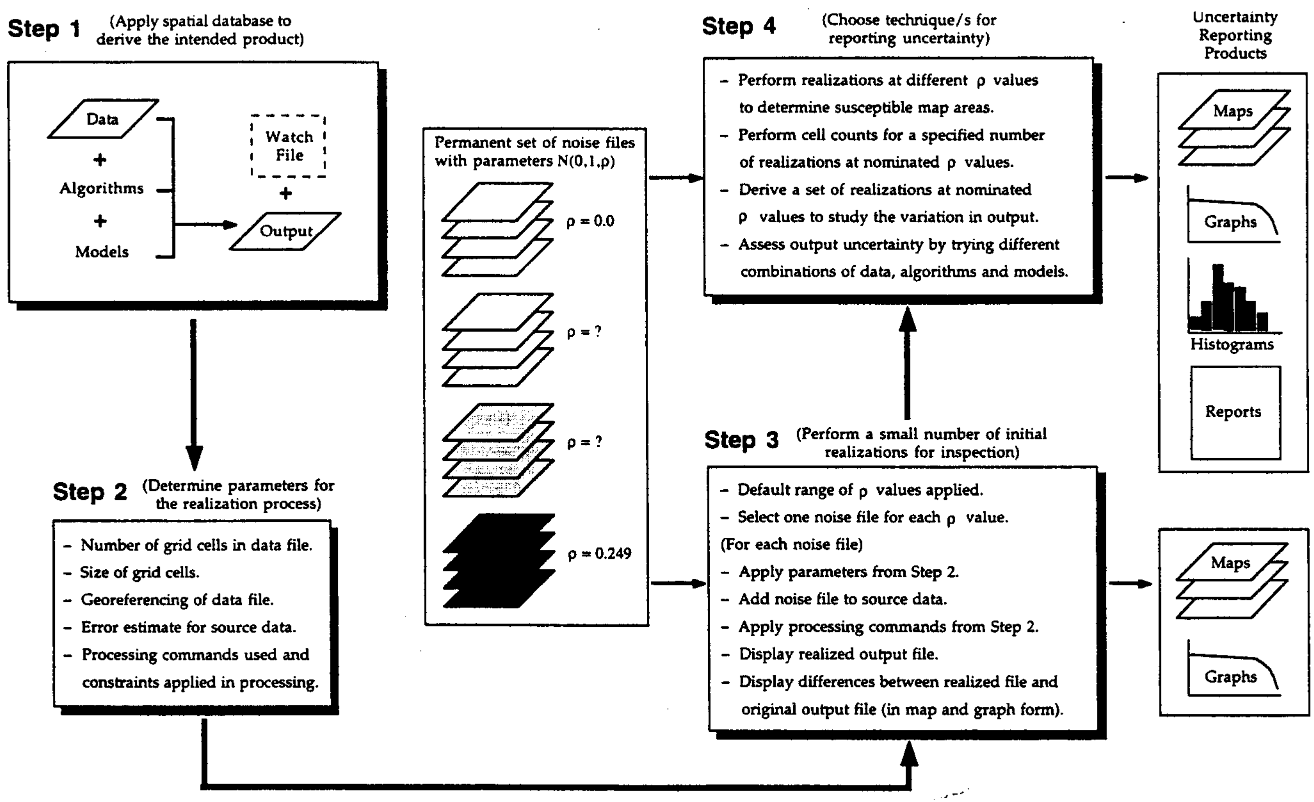
Of course, the realization process need not stop there, as the different slope gradient and aspect maps can be input to, say, hydrologic models to produce alternative realizations of drainage basin parameters, which in turn can be used to derive realized runoff characteristics. At any stage, the differences between the realized maps and the original (as produced from the source data without any consideration of uncertainty) can be analyzed to assess the resultant effects. The attractiveness of this approach is that even though we do not know how error is being propagated, its effects are nevertheless displayed.

A Methodology to Handle Uncertainty

The methodology to handle uncertainty embodies the model previously described (Figure 1). It consists of four steps, with the first one requiring the user to combine whatever data, processes and models are needed to generate the desired output-in other words, applying the spatial database as would normally be done without any consideration of uncertainty. From the beginning of the procedure, a log or watch file is kept of the commands used which will later be applied in producing the realizations.

In the second step, the parameters necessary for the realization process are determined. By reading system variables associated with the source data file, the num-

FIGURE 1. A Methodology for Reporting Uncertainty in Spatial Database Products.



ber of rows and columns in the data file, the cell size, and the geo-referencing details of the data can be ascertained. These will be required later when the noise files are to be transformed to agree with the actual data sets used. An error estimate for the source data will also need to be identified, and this can take the form of either a global value for the file, or else a separate field in the database which may be subject to spatial variation.

At this stage, the watch file of commands may need to be edited by the user to include only those which were finally applied in the procedure. Any constraints applied during processing will also be embedded in this file, such as in a viewshed computation where cells immediately surrounding the viewing point are usually masked out or held fixed (and therefore assumed to be always seen) so that their elevations are not perturbed, thereby possibly obscuring large areas of the viewshed.

While not a direct step in the realization procedure as such, the noise files to be employed would usually be previously computed and then permanently stored in the system for future use. The way in which they are generated has already been described by Hunter and Goodchild (1995b). To date, it has been considered sufficient for most applications tested for about ten files to be held for each ρ value, although users would have the option of creating a greater number of noise files for specific tasks in the final module of the methodology. The default ρ values chosen for the noise files are 0.0, 0.05, 0.10, 0.15, 0.20, 0.21, 0.22, 0.23, 0.24, 0.245, and 0.249.

Again, the user has the option of producing noise files with other ρ levels in the final module. As for the maximum value of ρ offered (0.249), experience has shown there is little to be gained from using ρ values higher than this since the realization process becomes so constrained that there is no discernible difference between the realized maps and the original product. In Step 3 of the methodology, it is expected that users will want to see a small number of initial trial realizations and the default range of ρ values listed above is applied. A single realization for each value is performed by first applying the parameters derived from Step 2 to georeference and transform the coordinates of the noise grid. Next, the error estimate (usually an RMSE for DEMs) is applied to map the noise values from a Normal distribution of N(0,1) to N(0,RMSE). This adjusted noise file is added to the source data to produce a realization to which the commands employed to create the original database product are applied. The realized maps and the differences between the realizations and the original outputs can be displayed in map or graph form.

Finally, in Step 4 of the procedure the user may choose one or more approaches for more detailed investigation of product uncertainty, as discussed in the previous section, and with a greater variety of reporting output products available.

A Case Study in Assessing Uncertainty

The case study to be discussed relates to mapping areas burnt by forest fires through the use of remotely sensed Thematic Mapper (TM) imagery. In rugged areas in particular, many researchers have reported the problem of confusion between shadowed and burnt regions as they both appear the same in most of the bands. Similarly, if terrain corrections are not applied, it is difficult to discriminate between differing degrees of fire severity since a given area may seem darker not because the fire was more intense, but because it lies on a slope that receives less light by area unit.

Traditionally, DEMs have not been used as part of the assessment for fire severity mapping, but research at the University of California, Santa Barbara, is underway to determine how consideration of topographic effects on the satellite signal may be used to counter this problem, which requires a radiance model to be applied to correct (or normalize) the radiance data for terrain differences (Caetano 1993, personal communication). Normalized radiance values are already commonly used in other applications of remote sensing in mountainous areas, and once derived may be used with traditional image-analysis techniques such as supervised and unsupervised image classification, and density slicing.

The test site lies in central Portugal near Pampilhosa da Serra, and a DEM with a cell size of 30m x 30m was used as the basis for normalization of the TM imagery. Figure 2a shows a hill-shaded view of the DEM covering the test site, while the same area is also delineated on the unclassified TM Band 4 scene in Figure 2b in which the effects of fire clearly show in the middle of the image as regions of dark gray/black color. The portion of the DEM used for this research measures 353 rows by 272 columns (or about 10.6 km by 8.2 km), with elevations ranging from 287m to 1020m. Unlike DEM data supplied by the U.S. Geological Survey, an estimate of the elevation error for the DEM is not available, and so on the advice of researchers familiar with the Portuguese digital mapping program the standard deviation for elevation error has been estimated to be 10m.

The traditional procedure used by image analysts is to calculate the slope gradient and aspect for each cell in the DEM and then combine them with the sun's zenith and azimuth angles at the time of image capture (taken

FIGURE 2a. A hill-shaded view of the test site DEM.

FIGURE 2b. Illustration of the darker burnt area in the unclassified TM scene.

from the file header or else calculated for the time of day and the latitude and longitude of the site). This information is used to compute the cosine of the incidence angle, which has values in the range –1.0 to +1.0. Thus, a cell with an aspect equal and opposite to the sun's azimuth (in other words, facing the sun), and a gradient equal to the sun's zenith angle (that is, perpendicular to the sun's rays), will receive the maximum amount of radiance and have an incidence angle of 0° with a cosine of +1.0. The formula for the cosine of the incidence angle (i) is given by Equation 1:

$$\cos(i) = \cos(\text{sun zenith}) * \cos(\text{cell gradient}) + \sin(\text{sun zenith}) * \sin(\text{cell gradient}) * \cos(\text{sun azimuth–cell aspect}) \quad (1)$$

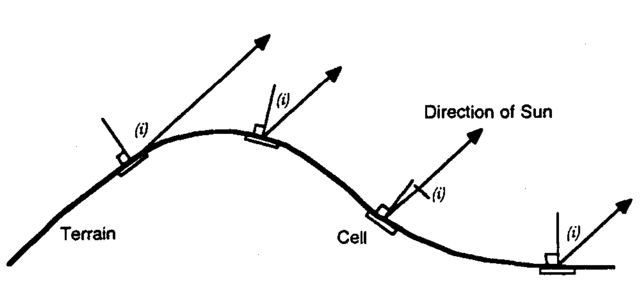
Cells which have an incidence angle cosine equal to or less than zero either lie in a plane parallel to the direction of the sun's rays or else are on reverse hill slopes (Figure 3). These cells are deemed to be in 'self shadow' and are not operated on in the traditional research procedure due to the difficulty of working with diffused light. There is a further process which identifies cells that are in 'cast shadow' from larger features which obscure them from direct sunlight, and these cells also are usually excluded from further calculations.

The incidence angle cosines for all cells in the DEM are then used as a means of normalizing the radiance values of pixels in the TM image, given that radiance is affected by the nature of the terrain to which it applies. At this point it should be noted that corrections will have already been made to ensure that both the DEM and the TM images have the same georeferencing and cell/pixel size. The radiance values, being the raw signals from the image in the range 0 to 255, are then normalized by computing the value they would have if each pixel was horizontal, as in Equation 2:

$$L_H = L/\cos(i) \quad (2)$$

where L denotes the radiance value, and L_H is the normalized radiance value.

FIGURE 3. Variation in the incidence angle (i) with cell postion.



Having discussed the analyst's traditional procedures for deriving normalized radiance values for each pixel, it is clear there is considerable potential for applying the realization methodology to assess the uncertainty present in the final output, which would include any effects arising from the DEM elevation error, the algorithms used to calculate slope gradient and aspect, and the formulae applied to determine the incidence angle cosines and the normalized radiance value. In this example we focus on dealing with the uncertainty associated with normalization, since that was the primary concern of the researcher conducting the major study; however, there is no reason why the process outlined here could not continue through to the next stage of analysis of the effects of uncertainty on fire-severity modeling.

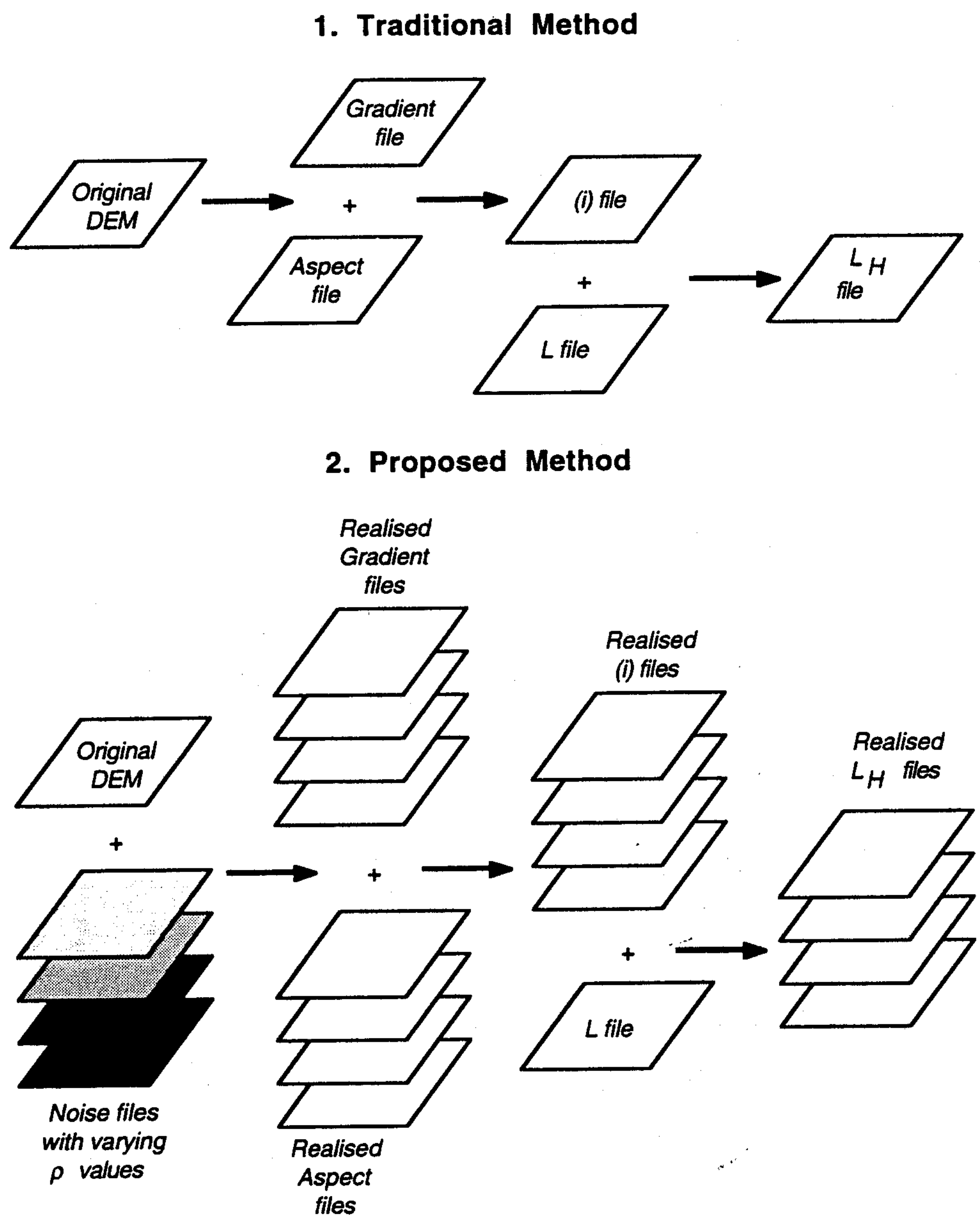
The difference between the traditional approach to calculating the terrain-corrected L_H values and the proposed approach permits their uncertainty to be assessed (Figure 4). The latter technique applies elevation noise files, with varying levels of spatial autocorrelation, to the original DEM to establish corresponding sets of slope gradient and aspect files for the test site. Pairs of gradient and aspect realization files (for each given ρ value) are then taken in turn and used to calculate the corresponding incidence angle cosine file, which is applied to the original TM radiance file (L) used for the analysis. The process results in the creation of a family of realized L_H files whose outputs can then be analyzed. The entire process was automated by using a macro command script and applied using the ARC GRID software (ESRI 1992).

The adopted procedure resulted in a set of 10 realized L_H files for each of the ρ values 0.0, 0.05, 0.10, 0.15, 0.20, 0.21, 0.22, 0.23, 0.24 and 0.245. For the purpose of analysis, the 96,016 grid cells in every realized file were subtracted from their corresponding cells in the original L_H file to provide a "difference" file. This difference represents the amount by which the final L_H value might be expected to vary under terms of uncertainty due to variation in the elevations of the original DEM and the subsequent series of spatial operations that were performed on the data.

The mean and standard deviations of each set of 10 difference files were then calculated for the range of ρ values applied. The results are shown in Figures 5a and 5b, with a gradual increase noticeable in the mean and standard deviation of the differences as ρ varies from 0 to 0.20, followed by sudden decreases as ρ approaches 0.25. Figures 5a and 5b therefore represent the results of 96,016 x 10 x 10 (or 9,601,600) individual calculations.

At this stage, analysis of the results shows that the average greatest difference that might be expected in L_H values is about 2.5 units with a standard deviation of

FIGURE 4. **Comparing the traditional technique of creating the L_H file with the proposed method which provides for assessment of L_H uncertainty.**



approximately 13 units. These extremes occur around $\rho = 0.20$. However while such global statistics are useful in their own right, they say nothing about the spatial variation of the differences and, accordingly, further analysis was made of the realizations made at $\rho = 0.20$.

Taking the 10 realized L_H difference files at $\rho = 0.20$, a composite file was calculated and displayed such that cells with an L_H difference within ±2 standard deviations of the overall mean for the file were shaded as gray color, while cells with an L_H difference greater than ±2 standard deviations were shaded as white and black color respectively. The result is shown in Figure 6a where it can be seen that the white and black cells, representing outlying values or those most susceptible to the spatial operations applied, tend to occur on west-facing slopes of north-south ridgelines when compared with a hill-shaded view of the test site DEM with contours overlaid at an interval of 100m (Figure 6c).

FIGURE 5a. Showing the mean difference in L_H for each set of 10 realized files plotted against variation in the spatial auto correlation parameter (ρ).

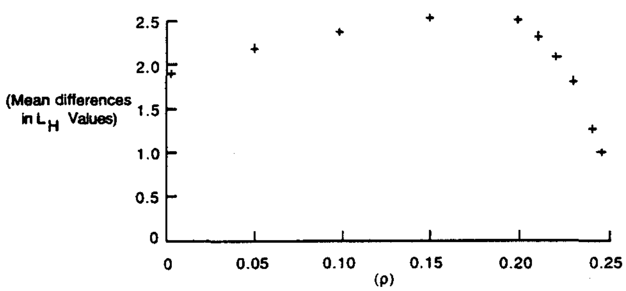
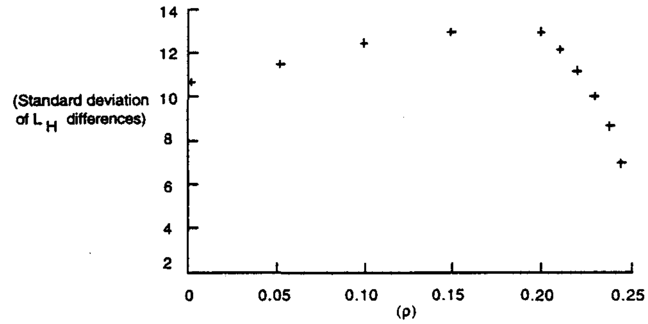


FIGURE 5b. Showing the standard deviation of the differences for the same sets of 10 files.



The file used in Figure 6a was then hill-shaded from the northeast to communicate both the size and spatial variation of the differences, and cell values beyond the ±2 standard deviation threshold show as a highly disturbed pattern while cells with differences within the threshold display as relatively smooth gray color (Figure 6b). One site in particular, in the top northeast corner of the image contains a significant L_H difference witnessed by its long shadow extending to the southwest. Given that this file represents the mean difference value occurring after 10 independent realizations, there is the suggestion of either an anomalous DEM elevation or TM radiance value present which warrants closer inspection of the original data.

Having illustrated the spatial variation in the uncertainty of the L_H values, further explanation was sought

FIGURE 6a. Showing the mean L_H difference file after 10 realizations at ρ = 0.20, with cells below and above the ± 2 standard deviation threshold shown as black and white color respectively.

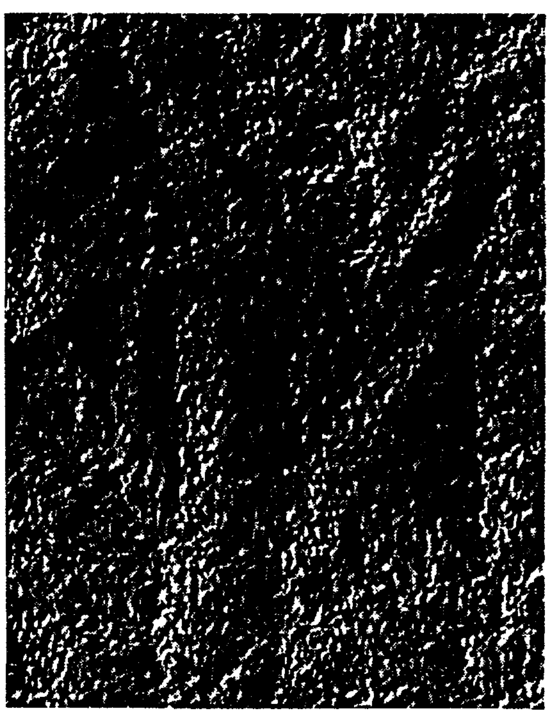


FIGURE 6b. Showing a hill-shaded view of the L_H difference file used in Figure 6a, with cells outside the ± 2 standard deviation threshold showing as disturbed areas (note the anomaly in the upper right corner).

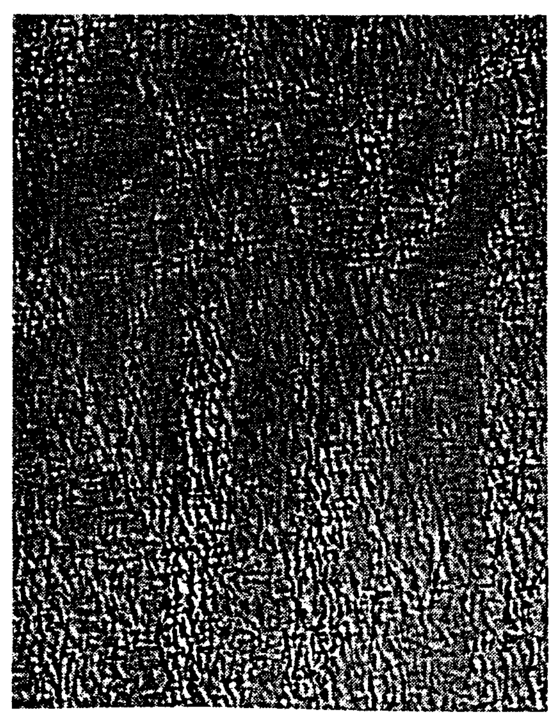


FIGURE 6c. Showing a hill-shaded view of the DEM by comparison with 100m contours overlaid.

as to the reason for the apparent correlation between significant differences in L_H and west-facing slopes. This can be explained by the location of the sun at the time of the TM image capture, which was at an azimuth of 117° and a zenith angle of 36° (during the middle of the northern hemisphere summer). From this position, the pixels in shadow are clearly affected the most, which confirms the problems encountered when working with pixels in diffused light. It was for this reason that cells found to be in shadow (and thus having an incidence angle cosine < 0) were not removed from the realization process, but instead deliberately retained to demonstrate any likely susceptibility to variation. Thus, the masking of such pixels during traditional analysis may be considered a valid approach to the problem.

At the same time, it was seen that for the remainder of the image the greatest mean difference in L_H that might be expected is about 2.5 units with a standard deviation of some 13 units. It is left to the user of the data to decide whether this variation is acceptable for the task at hand, and this assessment of product quality (or fitness for use) forms part of the management approach which needs to be adopted in such cases. If the variation is acceptable, then the methodology proposed has confirmed that the particular combination of DEM and TM imagery; the algorithms for gradient, aspect and incidence angle cosine; and the model for terrain correction of the L_H value are suitable for the purpose intended.

On the other hand, if these differences are unacceptable then uncertainty reduction methods will need to be employed, such as:

- choosing more accurate DEM data;
- selecting alternative algorithms and models;
- taking certain areas shown to be most susceptible to the effects of uncertainty out of the analysis; or
- employing TM imagery from other epochs.

To this end, the realization process may be repeated using different combinations of data, algorithms, and models to determine which one produces the least uncertainty in the final product. At the time of writing, work is already underway using external funding to develop a simple software toolbox which will embody the procedures described, in order that users may more easily automate the modeling and analysis procedures applied in this research.

Potential Applications of the Model

The potential applications of the model lie in four areas. First, the realization process may be used to highlight areas of a map that are susceptible to changes in parameter values. For instance, Hunter and Goodchild (1995b) demonstrated that the calculation of slope aspect from a DEM was particularly susceptible to variation in terrain elevation in relatively flat regions, while large hillside slopes remained relatively stable. While such a conclusion is already fairly well established, this may not always be necessarily so; where complex process models are applied, their effects may still be largely unknown. In other applications, the observed differences might be used as input to subsequent sensitivity analysis to understand how changes in parameters impact upon the decision-making process, such as in land use suitability and capability studies.

Secondly, the technique can be useful in cases where differences *per se* are not as important as assessing the likelihood of a cell's membership of a particular class. An example of this can be found in viewshed computations where cells are computed as being either seen or not seen from a viewing point, and similar requirements may be made in calculating the extent of drainage basins. Sets of realizations taken at different ρ values can be added to compute a "score" for each cell (together with a mean and standard deviation), which in turn may be used to calculate the probability of a cell satisfying the criteria associated with the operation, thereby overcoming the "in or out" Boolean responses normally associated with spatial databases. Users can thus nominate a confidence level to be met when assessing the results of the process (for example, 'cells must have a 90 percent probability of being seen').

Another example occurs in soil classification which is often dependent in part upon slope-gradient estimates, in which realizations of a soil map can be produced based upon previous realizations of the slope-gradient map, enabling users to select cells on the basis of having a given probability of belonging to a defined gradient range. At the same time, cell class counts (and therefore area estimates) may be made for a set of realized maps with the mean and standard deviation of the area being reported to users.

Thirdly, a user might want to display several realizations of a map to understand the degree of variation associated with the processes involved. For example, instead of interpolating contours from a DEM just once, several realizations might be made to assess not only the impact of elevation uncertainty on the process, but also the variation due to the interpolation procedure itself. This could also be applied to other raster-to-vector conversion procedures when class polygons or linear features such as stream patterns are required, thus producing a family of possible boundaries or linear features.

Finally, simulations can be undertaken to study the effect on map products where competing data sets, error estimates, algorithms, and process models are available. This 'reverse engineering' approach might also be applied by users who, having already studied several possible realizations of a desired map, and having identified areas exhibiting levels of uncertainty considered unacceptable, wish to see how different uncertainty reduction options (for example, recollecting data at a higher accuracy) would affect the final outcome—before returning to the field site or purchasing alternative data sets.

Conclusions

In this paper the authors have presented a methodology that allows uncertainty to be reported for certain types of spatial database products. The work recognizes the critical need for tools to be developed to assist users in improving their understanding of the quality of the outputs from their systems. The methodology has been applied to communicate the uncertainty associated with using digital elevation models to correct remotely sensed imagery for terrain effects when assessing mountainous areas burnt by fire in central Portugal. The results of the study show that given the particular combination of DEM data, algorithms, models and imagery employed, the normalized radiance values of pixels in some locations are highly susceptible to variation in the input parameters. As such, the procedures that have been applied permit users to study the uncertainty associated with the analysis, identify where its effects are most severe, study its impacts on the final spatial database product derived, and offer the opportunity for alternative data and algorithms to be tested to determine which combinations yield uncertainty levels that are acceptable for the task at hand. We believe the result will be an analysis that is better informed with respect to the effects of uncertainty. On the other hand, we cannot avoid the need to involve the user in any ultimate decision about whether the analysis meets its objectives—in that sense the technique remains partially subjective.

While the basis of the approach lies in statistical error modeling, the results can be visualized readily within a GIS, and the paper has demonstrated techniques that can be employed by users lacking any great depth of statistical knowledge or training. In this sense, the visual, intuitive nature of GIS analysis and modeling holds the key to novel, practical approaches to the management of uncertainty in spatial data. The methodology outlined in Figure 1 can be implemented in many existing GISs using scripting or macro languages, hiding most of the computation from the user, and exploiting visual techniques for eliciting the key information needed to implement the model.

On the other hand, the basis of the technique is complex and it will probably never be possible to package them in a form which is fully understandable by all GIS users, given the diversity of backgrounds common in this field. For that reason we suggest that the toolbox containing these techniques be operated by the analysts in an organization who can in turn feed the results to users. Readers interested in obtaining the toolbox and an associated tutorial manual should contact the first author.

Acknowledgments

The authors gratefully acknowledge the support received for this research from the Australian Key Center for Land Information Studies (AKCLIS), and the National Center for Geographic Information and Analysis (NCGIA). The NCGIA is supported by the National Science Foundation, Grant SBR 88-10917 and this research constitutes part of the Center's Research Initiative 7 on Visualization of the Quality of Spatial Data.

References

Cliff, A.D. and Ord, J.K. 1981. *Spatial Processes: Models and Applications*, Pion, London.

Dutton, G. 1984. "Truth and its Consequences in Digital Cartography", In *Proceedings of the ACSM 44th Annual Meeting*, Washington, DC, pp. 273–283.

ESRI. 1992. *ARC/INFO User's Guide: GRID Command References (version 6.1)*, Environmental Systems Research Institute Inc., Redlands, California.

Goodchild, M.F., G. Sun and S. Yang. 1992. "Development and Test of an Error Model for Categorical Data", *International Journal of Geographical Information Systems*, Vol. 6 (2):87–104.

Goodchild, M.F., C. Lin and Y. Leung. 1993. "Visualizing Fuzzy Maps," In *Visualization in Geographic Information Systems*, pp. 158-167, H. Hearnshaw and D. Unwin (eds.), Wiley, Chichester.

Hunter, G.J. and K. Beard. 1992. "Understanding Error in Spatial Databases", In *The Australian Surveyor*, Vol. 37 (2):108–119.

Hunter, G.J. and M.F. Goodchild. 1995a. "Dealing with Uncertainty in Spatial Databases: A Simple Case Study", *Photogrammetric Engineering and Remote Sensing*, Vol. 61 (5):529–537.

Hunter, G.J and M.F. Goodchild. 1994b. "Modeling the Uncertainty of Slope Gradient and Aspect Estimates Derived from Spatial Databases," submitted for publication to *Geographical Analysis*.