Dealing with Error in Spatial Databases: A Simple Case Study

Gary J. Hunter and Michael F. Goodchild

Abstract
There is now a considerable body of literature on the techniques available for modeling and communicating error in spatial databases. Some error models have solid statistical foundations, while the basis for others is not so strong. In this paper, three basic approaches to the problem are examined. The application investigated is a fundamental one—to determine the position of a given terrain elevation value and to portray the resultant error of the answer. Such a problem can be of critical concern to communities in cases of flood plain mapping, determination of rising sea levels resulting from global warming, or delineation of the full supply level for a proposed reservoir. In this instance, the authors suggest that the application of simple probability theory, when combined with the error estimates supplied by data producers and current computer graphics capabilities, can provide users with more meaningful information concerning the error of their spatial database products. In turn, this information may allow them to better deal with an issue of growing concern.

Introduction
With increased research into error modeling in spatial databases over the past few years, there has been a corresponding growth in the work examining the means by which error can be presented to users. The communication of spatial data error can take many forms, ranging from the use of epsilon bands (Chrisman, 1982) and other descriptors of error such as misclassification matrices, map reliability diagrams, and fuzzy logic (Leung *et al.*, 1992), through to probability surfaces (Lowell, 1992), variability diagrams (MacLean *et al.*, 1992), simulation techniques (Stoms *et al.*, 1990; Fisher, 1991; Goodchild *et al.*, 1992), and advanced computer graphics animation and audio effects (Fisher, 1992).

Clearly, the growing diversity of applications and the number of spatial processes now available mean that there can be neither a single all-embracing error model for spatial data, nor any one optimum method for presenting error. Indeed, Hunter and Goodchild (1993) contend that error management in the future will depend not only on the specific combination of data, systems, and processes employed, but also on the types of users, their respective skill levels, the nature of the applications, the decisions to be made, and the degree of impact that each spatial database has upon the decision-making process.

What is needed is for the research agenda in spatial database error to widen, from its current platform of error modeling and visualization, to embrace the treatment of error from a management (or user-oriented) perspective. While the academic research community continues to have an important role to play in this field (and there remains much that is unknown), alternative techniques and case studies need to be presented to users as a means of educating and assisting them to better deal with error in everyday, operational situations.

In this paper, a common problem (delineation of elevation) is studied and several alternatives for modeling and presenting error are investigated. It is argued that, for the application in question, the use of simple probability theory, when combined with error estimates as supplied by data producers and current computer graphics capabilities, can provide users with much more effective information concerning the error of their spatial database products. In turn, this information can assist them to more effectively deal with this important issue.

The Problem and the Test Site
As mentioned, the application is a common one, that is, to identify the horizontal position of a nominated terrain elevation value. At first glance, it is not a difficult problem to solve but one which can be of critical concern to communities in cases such as flood plain mapping, determining rising sea levels resulting from global warming, or delineating the full supply level for a new reservoir. The problem is that the error of the result must also be presented to users. The significance of the exercise is that error in elevation can have considerable impact upon horizontal delineation of such boundaries, and from a policy-maker's perspective it is important to be aware of the error of spatial database products so that appropriate action may be taken to reduce and absorb the risk associated with any decisions based on that information (see Hunter and Goodchild (1993) for further discussion of error reduction and absorption). In the problem described in this paper, the 350-m elevation is to be delineated for the data set described below. While this elevation was chosen because of its wide variation in horizontal position, its numerical value has no other significance and the treatment given to it is generic in nature.

The data for the exercise come from a 448-row by 334-column subset (86 percent) of the U.S. Geological Survey (USGS) Digital Elevation Model (DEM) for the 7.5-minute,

G.J. Hunter is with the Department of Geomatics, University of Melbourne, Parkville, Victoria 3052, Australia.

M.F. Goodchild is with the National Center for Geographic Information and Analysis, 3510 Phelps Hall, University of California, Santa Barbara, CA 93106.

Photogrammetric Engineering & Remote Sensing,
Vol. 61, No. 5, May 1995, pp. 529–537.

0099-1112/95/6105-529$3.00/0

1:24,000-scale State College (Pennsylvania) mapping quadrangle. The total number of cells in the test file is 149,632 with each one measuring 30 m by 30 m and covering an area of 900 m^2 or 0.09 ha. The DEM was supplied as part of the "Visualization of Spatial Data Quality Challenge," a research contest jointly sponsored by the U.S. National Center for Geographic Information and Analysis (NCGIA), the U.S. Environmental Protection Agency (EPA) Center for Environmental Statistics, the U.S. Department of Agriculture Soil Conservation Service (SCS), and the Statistical Graphics Section of the American Statistical Association (Beard, 1992). Figure 1 shows a contour plot of the test site topography at a contour interval of 20 m and with 100-m contours labeled.

The test site measures approximately 10 km by 13 km and has considerable variation in terrain, with elevation values ranging from 255 m in the north to 743 m in the southeast. The vertical accuracy of the DEM used for the tests is 7 m RMSE, which means that the square root of the average squared difference between the true and observed elevations at selected test points in the quadrangle is 7 m (see USGS (1990a), Weih and Smith (1990), and Acevedo (1991) for a general discussion of USGS DEM construction and accuracy), and it should be noted here that the USGS assumes a Gaussian distribution of random error. The DEM was supplied as an ERDAS (LAN) file following reformatting from the USGS data structure. The file was then converted by the authors to the GRID structure for use in the ARC/INFO system.

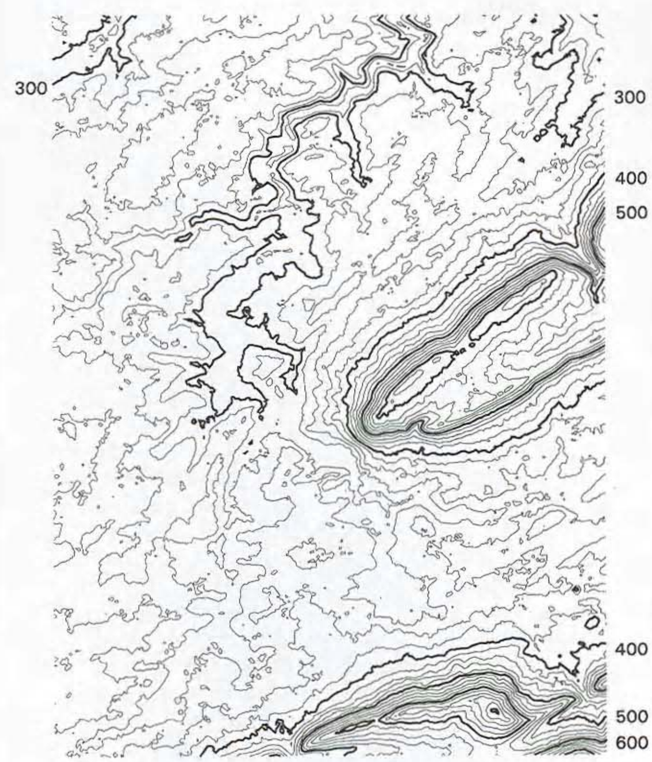


Figure 1. Test site topography (shown at a contour interval of 20m and with 100m contours labeled).

Comparison of Approaches

Error Ignored — the "Do Nothing" Option

The first approach examined (and which tends to be the rule rather than the exception) ignores error completely — in other words, it is the "do nothing" option. A user would simply select the 350-m contour in a vector database derived by interpolation of the DEM (Figure 2a), or else would reclassify and display the grid cells according to whether their elevations are lesser, equal to, or greater than 350 m (Figure 2b). This method makes no attempt to use the accuracy estimate supplied for the DEM by its producer (USGS), and it offers no measure of error of the derived map products. Deservedly, it fails as a solution to the problem.

Epsilon Bands

An advance on the "do nothing" option would be to draw epsilon bands around the interpolated position of the 350-m contour (from the Greek letter epsilon (ε) for error), and deem the true position of the contour to lie somewhere between these two extremes. This philosophy has been used to provide error estimates for digitized features when checking plots of converted linework against source documents, in which a typical quality criteria is that 90 percent of all points tested shall be within 0.5 mm of their source document position. This 0.5 mm, in turn, translates to 0.5 m at a scale of 1:1,000; 5 m at 1:10,000; and 50 m at 1:100,000. Proponents of the epsilon band suggest that this information can then be tagged to digitized features to provide an accompanying horizontal error attribute and the means for automatic generation of epsilon bands (for example, through a buffering routine).

While the general notion of the epsilon band has been discussed elsewhere (for example, by Chrisman (1982), Dunn *et al.* (1990), and Hunter (1992)), Goodchild *et al.* (1992) argue that (except in the case of point features) it is only an error descriptor and does not satisfy the requirements of an error model, because it provides no means of generating a population of distorted versions that satisfy all the requirements of the feature concerned—in this case, contours. For the data used in this case study, no estimate of horizontal positional accuracy is available; the only error estimate provided is a vertical component RMSE of 7 m for the DEM. As it would be unreasonable to suggest that no error has been introduced during the contour interpolation process, the figure of 7 m cannot be automatically transferred from the DEM to the derived contours as the error will most probably have increased. Indeed, the concept of the RMSE and its specific meaning with respect to probability distributions (see next section) cannot be rightfully used in relation to contour error because there is no known statistical justification for doing so. An additional problem with using epsilon bands (although not in this instance) is deciding how to portray them in the case of a contour which possesses both horizontal and vertical error estimates—when four bands (left, right, upper, and lower) could be deemed to exist to describe contour error.

Thus, it is considered that epsilon bands are inappropriate for displaying elevation error in the specific example given in this paper. However, from a pragmatic viewpoint it is recognized that there may be times when an attempt to portray error by epsilon bands is better than nothing at all. Therefore, for the sake of completeness, an epsilon band plot has been produced for the central eastern portion of the test site. The hypothetical scenario is that a user has assessed the error information available for the DEM and, while nothing is

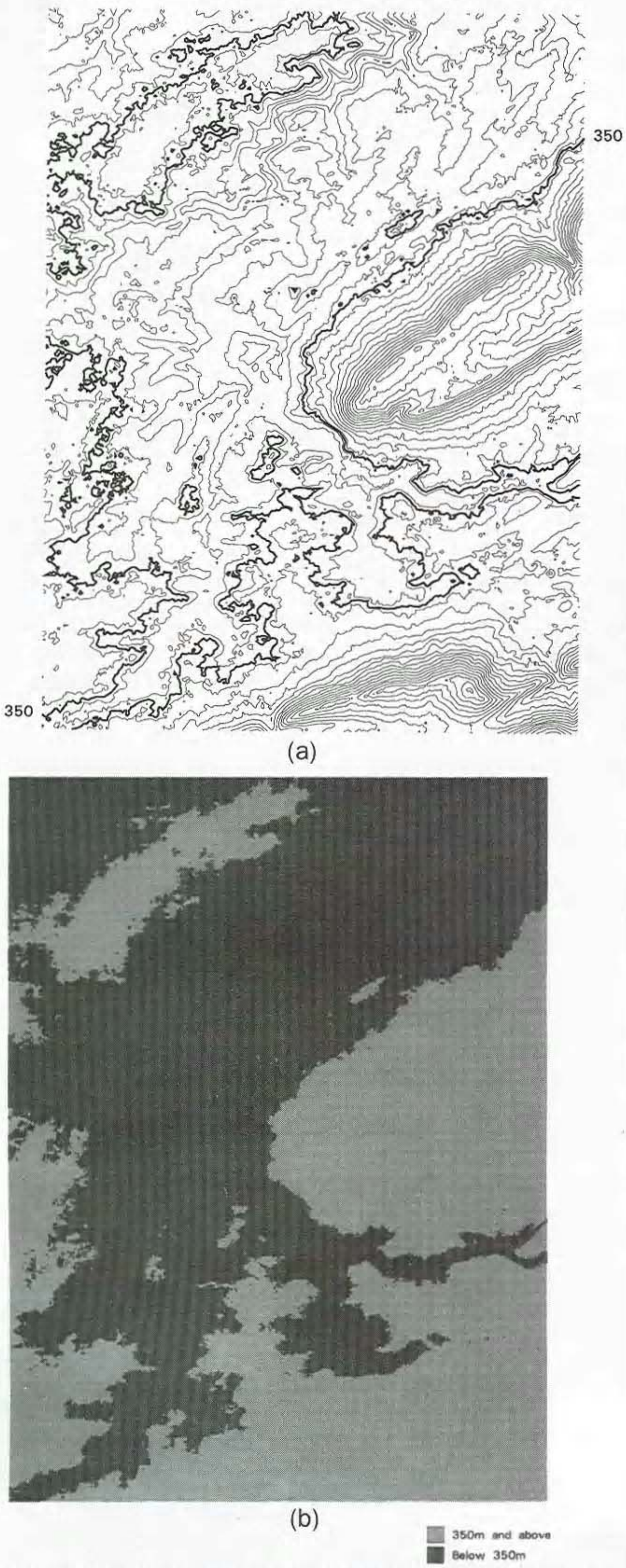


Figure 2. (a) Plot of the 350m Contour (thick line), superimposed over contours at a vertical interval of 20m. (b) Plot of cell elevations, classified according to whether cells are less than, or equal to and greater than 350m.

known about horizontal positional error, has made a personal judgment that the true position of the 350-m contour lies somewhere within half the contour interval of 20 m (that is, 350 m ± 10 m). Hence, in Figure 3, the 350-m contour (thick line) is bounded by the 340-m and 360-m contours to illustrate (in plan view) the bounds of vertical error.

Probability Mapping

The final approach investigated is probability mapping, in which the likelihood of each grid cell's true value exceeding or being exceeded by a nominated threshold is first calculated, and then is used as a means of displaying the error of the data. Assuming a Gaussian or Normal distribution of random error, the concept is illustrated in Figures 4a, 4b, and 4c where cell (p) is shown with an observed elevation value (Z_p) and a vertical error estimate (RMSE). The Gaussian bell curves show the relative likelihoods of various true elevations (Z), given the observed elevation and the estimate of error; as one would expect, true elevations close to the observed elevations are most likely. The probability of each cell's true elevation being greater than some threshold value (Z_t) is evaluated by calculating $P(Z>Z_t)$ and writing the value to a separate file for display. In Figure 4a, where the threshold value Z_t is equal to the observed value Z_p, the characteristics of the normal probability distribution are such that cell (p) has a likelihood of 0.50 (50 percent) that its true elevation will be greater than the threshold value. Alternatively, in Figure 4b where $Z_t - Z_p = 1 \times$ RMSE, the probability is about 0.16 (16 percent), and in Figure 4c the probability is approximately 0.025 (2.5 percent).

Figure 3. Using the Epsilon Band concept to show error in elevation equal to half the interpolated contour interval of 20m (350m contour ±10m).

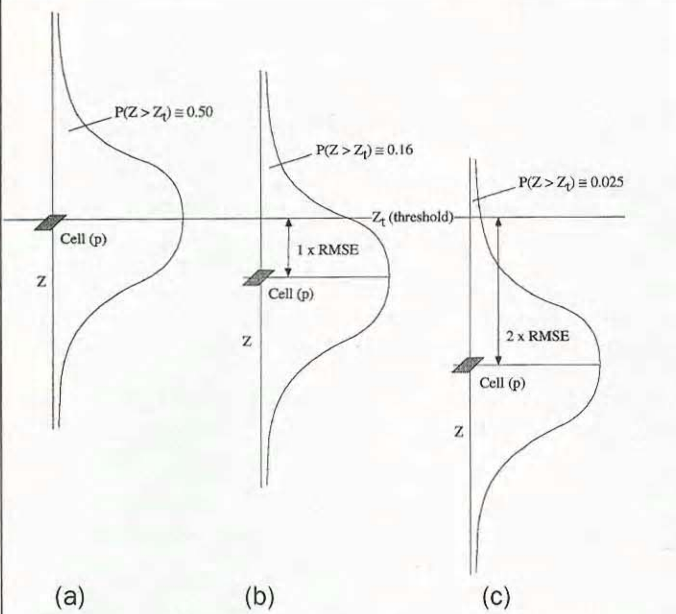


Figure 4. (a) Case 1 where $(Z_t - Z_p) = 0$. (b) Case 2 where $(Z_t - Z_p) = 1 \times$ RMSE. (c) Case 3 where $(Z_t - Z_p) = 2 \times$ RMSE.

There are several options for putting this theory into practice. One way of automating the process is provided in the IDRISI grid cell-based system (Eastman, 1992) which gives a specific function (PCLASS) to calculate the probability of each cell either exceeding or being exceeded by a user-defined threshold. The RMSE value is added to the image documentation file prior to initiating the PCLASS function. A new file holding cell probability values is then computed and serves as the basis for displaying error employing user-defined or system default look-up tables.

The only deficiency observed with the PCLASS function is that there is no facility for assigning different RMSE values to individual cells within an image. This need can occur in practice when users append several DEMs, each with its own RMSE value, to cover a large project site. One way to overcome this problem is to split the image into separate files on the basis of like-RMSE values, then run PCLASS for each file and re-join the individual probability files.

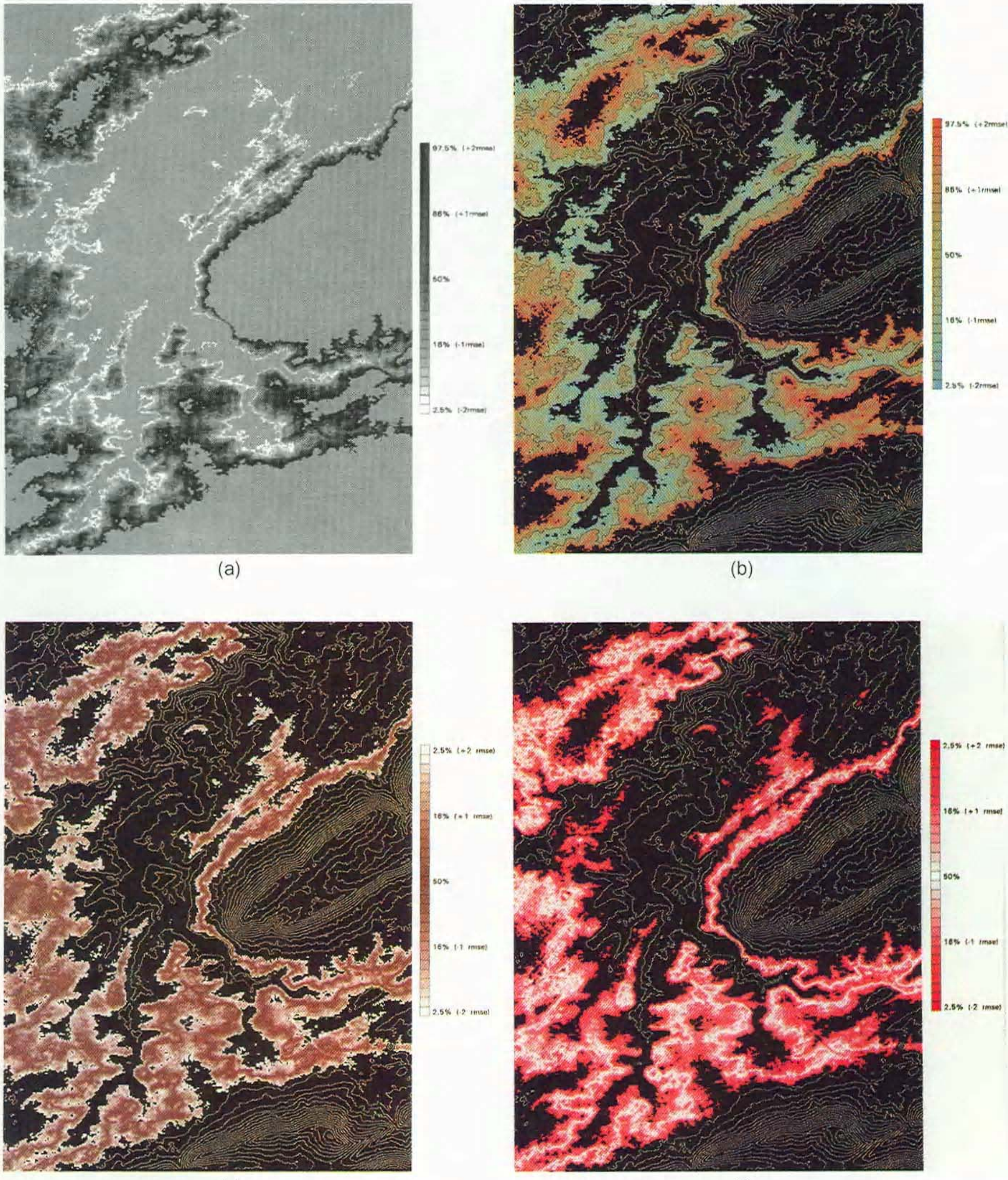
Unfortunately, many of the software packages used in large spatial database applications do not have this capability and consequently, for the purpose of this paper, the GRID software module in the ARC/INFO system (ESRI, 1991) was used. While there is no function in GRID equivalent to IDRISI's PCLASS, it was found that effective results could be obtained just as quickly by other means without having to calculate individual cell probabilities—although this could be still be achieved in GRID with a macro command which integrates the area under the normal curve (see the Appendix for an appropriate formula).

The procedure adopted was first to select all cells which contained elevation values within given error distribution limits centered around the elevation threshold (that is, 350 m ± 2 RMSE, or 350 m ± 14 m). Referring back to Figures 4a, 4b, and 4c, it can be deduced that cells selected within this range have between 2.5 percent and 97.5 percent probability of possessing a true value which exceeds 350 m. Alternatively, limits of 350 m ± 3 RMSE would see cells chosen with probabilities lying between 0.1 percent and 99.9 percent of exceeding 350 m. At this stage, it should be noted that the choice of RMSE limits is context-dependent and is best left to users to determine.

The cells are then displayed by means of color shading which is applied so as to gradually vary between nominated color extremes in user defined increments. A command (SHADECOLORRAMP) exists in the ARC/INFO software to achieve this; however, the same result could be easily achieved in other systems by employing color look-up tables. In Plate 1a, cells outside the limits of 350 m ± 2 RMSE have been masked in a neutral gray color, and the key legend at the right side of the image depicts the color ramp chosen which varies from white (350 m − 2 RMSE) through to black (350 m + 2 RMSE). In this case, the color ramp represents an answer to the query "Show the probability of each cell exceeding the threshold value of 350 m." The color ramp was constructed in 29 equal one-metre increments, taking values from 336 m through to 364 m (29 values are required and not 28 because DEM elevations are supplied as integers to the nearest metre). For reference purposes, the nominal position of the 350-m contour derived through interpolation of the DEM is shown as a white line.

While the black-and-white image in Plate 1a helps to show the error in elevation, an obvious improvement can be seen in Plate 1b where color has been employed for the first time, with supporting background contours depicted at an interval of 20 m outside the 350-m ± 2 RMSE region. As in Plate 1a, the interpolated position of the 350-m contour is shown for reference purposes only. While the key legend still denotes the probability of a cell's true elevation exceeding 350 m, if the hypothetical situation is that the valley will be flooded to the 350-m elevation, then, by grading the color ramp from blue to brown, a user can more effectively see which cells will most probably be inundated (maximum blue color) and which cells will most likely remain above the water level (maximum brown color).

Alternatively, if a user is more interested in the most likely position of the 350-m contour, then color shading such as in Plates 1c and 1d may be more appropriate. The color ramps in these two figures have their maximum or minimum color, respectively, centered around the 350-m value. The key legends in Plates 1c and 1d show that cells with values at the upper limit of 350 m + 2 RMSE have approximately a 2.5 percent probability of being exceeded by a value of 350 m, while those cells at the lower limit of 350 m − 2 RMSE have a similarly small likelihood of exceeding 350 m. In areas where the slope is steep, cells with equal (or almost equal) probability of exceeding or being exceeded by a value of 350 m show as either a thin brown or thin white line, denoting small positional error (see the eastern half of each image), as opposed to the flatter terrain in the western half of the two images where there is considerable error in the position of the 350-m elevation. While there is no conceptual difference between Plates 1c and 1d, they represent alternative views of the same data set which may appeal in different ways to different users. Clearly, the assessment of the two figures is subjective, although informal discussions with other spatial database users have shown that most prefer the color shading in Plate 1d because of its clarity in showing the most likely position of the contour. However, at this time no cognitive studies have been conducted to either verify or



(a) (b)

(c) (d)

Plate 1. (a) Using a graduated white/black color ramp to show the probability of cell values exceeding 350m. (b) Using a blue/brown color ramp to show the probability of cell values exceeding 350m. (c) Using a white/brown/white color ramp to show the probability of cell values exceeding or being exceeded by a value of 350 m. (d) Using a red/white/red color ramp to show the probability of cell values exceeding or being exceeded by a value of 350 m.

refute this finding, and it clearly requires greater research into the application of visual variables and visually ordered scales.

Managing Error

Having visualized the error in the position of the 350-m elevation, the question remains as to how the information can be applied in practice. From a management perspective, users must choose between either reducing the error in their data or else absorbing (accepting) it. For instance, if a critical site is located near the western edge of the image where there are large areas of flat terrain with elevations very close to the 350-m threshold, the usual form of error reduction would be to recollect and reprocess elevation data with a higher accuracy for the area of interest. This will create regions in the image with different RMSE values, in which case it becomes necessary to calculate individual cell probabilities based on separate RMSEs, so that variation in probability can be displayed as opposed to changes in elevation shown in Plates 1a, 1b, 1c, and 1d.

Once the error in delineating the 350-m elevation is satisfactorily reduced, the user must then absorb the remaining error. Generally, this occurs by simply accepting the fact that the data will never be perfect and that there will always be some likelihood of error because of the way DEMs model reality. In other cases, additional factors may be built in to the absorption process. For example, with reservoir construction, a buffer zone is applied around the full supply level contour to be certain that all land proposed to be inundated is actually purchased from landowners.

Alternatively, there may need to be a change in the way elevations are perceived by thinking of them in terms of confidence levels. For instance, a user may want the 350-m elevation depicted with a confidence level of 90 percent, in which case cells with at least a 0.90 probability of exceeding 350 m would be selected (that is, 350 m + 1.282 RMSE, or 359 m). This concept already occurs in flood plain mapping where it is standard practice to compute flood levels which, for instance, have a 1 percent probability of being met or exceeded in any year (termed a 100-year event). Elsewhere in civil engineering, drainage design calculations are commonly based on rainfall runoff statistics for 10-, 20-, and 50-year storms (that is, storms with 0.10, 0.05, and 0.02 probabilities, respectively, of occurring in any one year). The point being made is that users of this type of information are quite capable of understanding that, while a 100-year flood may not occur within the next 150 years, it might also happen twice in the next 6 months. In other words, they are able to live with the error which occurs in natural phenomena, and it is suggested that users of spatial databases need to develop similar attitudes to the data they use and to the products they develop. In the context of this paper, such an attitude would see a user decide beforehand what level of error may be tolerated in the position of the 350-m contour, rather than simply selecting it with a single command and acting on its position without consideration for the error which may be present.

Other Issues Arising from the Research

DEM Interpolation Error

Thus far, the paper has concentrated on visual interpretations based on the knowledge that this class of DEM is subject to an RMSE of 7 m. This knowledge comes from the data quality statement that accompanies each USGS DEM product, which in turn results from a program of random point sampling (USGS, 1990b). In reality, however, the structure of errors present in DEM products is much more complex, and dependent on the process used to construct each DEM.

One known error structure is commonly termed the "Firth Effect." It is found in DEMs produced by photogrammetry, when the DEM is created row by row as a series of transects in alternating directions (usually along the north-south axis). Because operators tend to underestimate elevation when moving up slope, and overestimate when moving down slope, the DEM shows a pattern of strong positive correlation of errors between adjacent points in the same column, and strong negative correlation of errors between adjacent points in the same row. When contoured, the surface shows a characteristic herringbone pattern.

During the research, examination of the hill shaded portion of the image in the northeast and southeast corners revealed a series of abrupt, approximately north-south and east-west steps in cell elevations. As these steps do not correspond with the position of artificial features such as roads, it was suspected that elevation error had been induced during DEM production by treating the image as separate patches and then joining them to assemble the complete DEM.

Figure 5a shows a hill shaded portion of the DEM extracted from the northwest corner of the test file (consisting of 169 columns by 114 rows, or 19,266 cells representing an area of 5070 m by 3420 m). In this figure, several regularly spaced east-west steps in elevation appear as dark lines on the image. To further examine them, contours at a 2-metre interval were interpolated from the DEM as in Figure 5b. While the apparent error does not show as a complete grid pattern in the latter figure, there is sufficient regularity to suggest that the DEM has been constructed in approximately 500-m by 500-m patches, with elevation steps at the patch edges which vary in magnitude from zero to about 10 m. This view was subsequently confirmed in the detailed description of DEM construction given by Allam (1978) and by other technical advice received.

The explanation is that this error occurs in DEMs produced with the Gestalt Photo Mapper II (GPM 2) automated photogrammetric system. This system performs image correlation in patches (measuring 9 mm by 8 mm at photography scale) with minimal edge matching between patch boundaries. In effect, the interpolation error is bowl-shaped across each patch, with overestimation of elevation at the edges and underestimation at the patch center. As for the divergence in patch bearing from grid north (about 3°), this is due to the readjustment of the GPM 2 machine coordinates to the Universal Transverse Mercator coordinate system.

From an error management perspective, users should realize that about 50 percent of the DEMs constructed by the USGS have been produced with the GPM 2 system primarily for making orthophotos—in which case the error occurring at patch edges is considered negligible. Given that these DEMs are now being used for purposes other than those for which they were originally intended (that is, the production of orthophoto maps), it is suggested that, while the likelihood of these errors being detected during accuracy testing is minimal, there is increased need for users to be aware of the error in this form of elevation data and the products derived from it. In addition, the existence of errors such as these highlights the need for improved data quality reporting to provide more information to users to allow them to better determine the appropriateness of DEM data for their applications. When such information becomes available, it will be possible to develop much more sophisticated methods of vi-

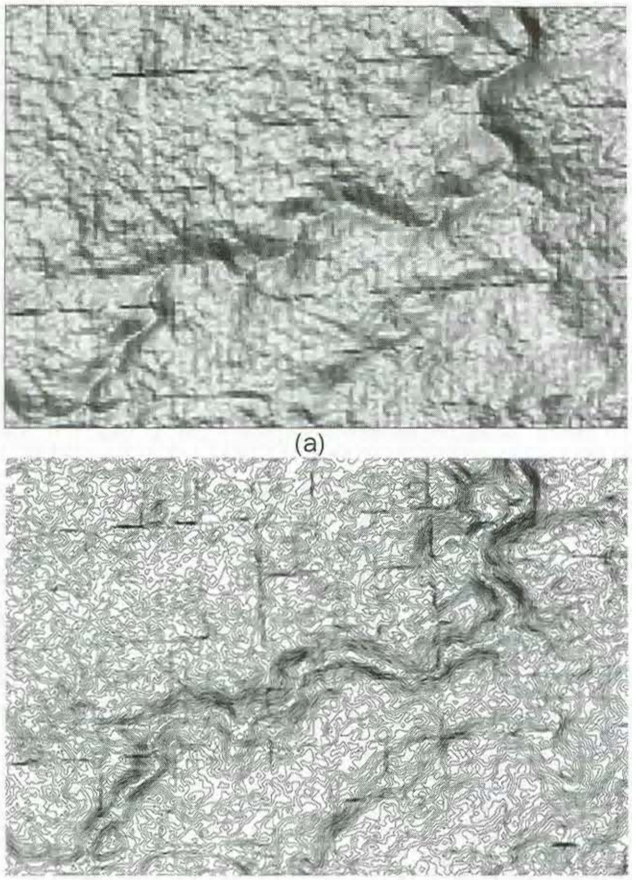


Figure 5. (a) A hill shaded portion of the DEM showing several regularly spaced, east-west steps in elevation caused by interpolation errors during production. (b) Two 2-meter interval contours for the image in Figure 5a, showing the steps in elevation caused by producing the DEM in 500-m by 500-m patches with minimal edge matching.

sualization that reflect the specific patterns of error known to exist in the data, because the assumption that DEM errors are random and independent is obviously not correct. To help deal with this issue, a forthcoming paper by the authors (Hunter and Goodchild, in review) presents a simulation-based approach which takes spatial autocorrelation into account.

Construction of Graduated Color Ramps to Depict Error

While the graduated color ramps used throughout this paper to display probabilities were generated automatically by basic system commands, it should be noted that the changes in color did not vary exactly in linear proportion to the probabilities being represented. For example, in Plate 1a, while the variation in grey-scale from white to black is proportional to the change in elevation, it is not proportional to the change in probability due to the nature of the probability function.

In Figure 6, color variation is shown on the horizontal axis and the probability of a cell exceeding an elevation of 350 m is charted vertically. In this graph, the straight (thick) line indicates a linear relationship between change in color and probability, while the curved (thin) line shows the color variation actually generated. Again referring to Plate 1a, at the 16 percent probability level there is 25 percent grey shading, and at 84 percent probability only 75 percent grey shading has been applied. This difference will increase if shading is scaled to a wider interval, for example, between ±3 RMSE rather than ±2 RMSE.

Although a plot of RMSE difference from the 350-m elevation versus grey-scale variation is correctly generated using the method described in this paper, it is accepted that, if it is probability which is intended to be displayed rather than RMSE difference, then users will need to create customized shade look-up tables which give the correct linear result. Alternatively, if individual cell probabilities are calculated and displayed as such (as in the IDRISI PCLASS function), then the correct relationship will be achieved.

Conclusion

This paper has investigated the common problem of determining the position of a given elevation from a DEM under terms of error. The problem has applications in flood plain management, in engineering projects, and in assessing the effects of new phenomena such as rising sea levels resulting from global warming. Several methods were used to depict a given elevation and, more importantly, to display the error of the answer. The authors suggest that the use of fundamental probability theory, in conjunction with the error statistics supplied by data producers and computer graphics, gives a simple yet effective visualization of the result which can be easily implemented. Clearly, the paper has raised questions for further research, such as "What error displays are most effective for users ?" and "What are the implications for users by thinking of spatial data in terms of confidence lev-

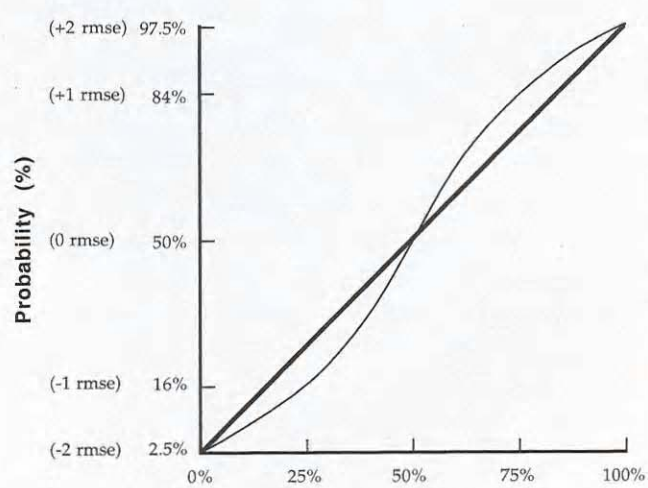


Figure 6. A plot of variation in color versus cell probability of exceeding 350 m, showing that color ramps constructed by the default ARC/INFO system command do not vary linearly with change in probability as desired.

els?" In addition, the research has shown that further attention should be paid to the need for greater information on DEM error beyond the simple RMSE, because it is known to vary markedly between different DEM creation methods.

Acknowledgments

The National Center for Geographic Information and Analysis is supported by the National Science Foundation, grant SES 88-10917. This research constitutes part of the Center's Research Initiative 7 on Visualization of the Quality of Spatial Data. The advice of Dr. Mark Kumler, of the University of Colorado, is gratefully acknowledged in helping to explain the DEM error discussed in the fifth section of the paper.

References

Allam, M.M., 1978. DTM Application in Topographic Mapping, *Photogrammetric Engineering & Remote Sensing*, 44(12):1513-1520.

Avecedo, W., 1991. First Assessment of U.S. Geological Survey 30-Minute DEMs: A Great Improvement over Existing 1-Degree Data, *Proceedings of the 1991 ACSM/ASPRS Annual Conference*, Baltimore, Maryland, 2:1-12.

Beard, M.K., 1992. *Call for Participation: Visualization of Spatial Data Quality Challenge*, National Center for Geographic Information and Analysis, University of California, Santa Barbara, 4 p.

Chrisman, N.R., 1982. *Methods of Spatial Analysis Based on Error in Categorical Maps*, PhD Dissertation, University of Bristol, United Kingdom.

Dunn, R., A.R. Harrison, and J.C. White, 1990. Positional Accuracy and Measurement Error in Digital Databases of Land Use: An Empirical Study, *International Journal of Geographical Information Systems*, 4(4):385-398.

Eastman, R.J., 1992. *IDRISI: User's Guide (Version 4.0)*, Graduate School of Geography, Clark University, Worcester, Massachusetts, 178 p.

ESRI, 1991. *ARC /INFO User's Guide: Cell-Based Modeling with GRID (Version 6.0)*, Environmental Systems Research Institute, Redlands, California, 277 p.

Fisher, P.F., 1991. Simulation of the Error of a Viewshed, *Proceedings of the Tenth International Symposium on Computer-Assisted Cartography, AUTO-CARTO 10*, Baltimore, Maryland, March, pp. 205-218.

———, 1992. Real-Time Randomization for the Visualization of Uncertain Spatial Information, *Proceedings of the 5th International Symposium on Spatial Data Handling*, IGU Commission on GIS, Charleston, South Carolina, 2:491-494.

Goodchild, M.F., S. Guoqing, and Y. Shiren, 1992. Development and Test of an Error Model for Categorical Data, *International Journal of Geographical Information Systems*, 6(2):87-104.

Hunter, G.J., 1992. Techniques for Assessing the Effect of Processing Errors in Spatial Databases, *Proceedings of the URISA '92 Annual Conference*, Washington, D.C., 1:24-33.

Hunter, G.J., and M.F. Goodchild, M.F., 1993. Managing Error in Spatial Databases: Putting Theory into Practice, *Proceedings of the URISA '93 Annual Conference*, Atlanta, Georgia, 14 p.

———, (in review). Modeling the Uncertainty of Slope Gradient and Aspect Estimates Derived from Spatial Databases, *Geographical Analysis (journal)*, 25 p.

Leung, Y., M.F. Goodchild, and C-C. Lin, 1992. Visualization of Fuzzy Scenes and Probability Fields, *Proceedings of the 5th International Symposium on Spatial Data Handling*, IGU Commission on GIS, Charleston, South Carolina, 2:480-490.

Lowell, K.E., 1992. On the Incorporation of Error into Spatial Data Systems, *Proceedings of the GIS/LIS '92 Annual Conference*, San Jose, California, 1:485-493.

MacLean, A.L., 1992. The Use of Variability Diagrams in a GIS to Improve the Interpretation of Digital Soil Maps, *Proceedings of the 1992 ACSM/ASPRS Annual Conference*, Washington D.C., 3: 66-76.

Rohlf, F.J., and R.R. Sokal, 1981. *Statistical Tables, Second Ed.*, W.H. Freeman and Company, New York.

Stoms, D.M., F.W. Davis, C.B. Cogan, M.O. Painho, B.W. Duncan, and J. Scepan, 1990. Sensitivity of Habitat Models to Uncertainties in GIS Data: A California Condor Case Study, *Proceedings of the GIS/LIS '90 Conference*, Anaheim, California, 1:69-78.

USGS, 1990a. *Digital Elevation Models*, Earth Science Information Center, U.S. Geological Survey, Reston, Virginia, 2 p.

———, 1990b. *Digital Elevation Models: Data Users Guide 5*, Earth Science Information Center, U.S. Geological Survey, Reston, Virginia, 51 p.

Weih, R.C., and J.L. Smith, 1990. Characteristics and Limitations of USGS Digital Elevation Models, *Proceedings of the Conference on Application of GIS, Simulation Models, and Knowledge-Based Systems for Landuse Management*, Department of Agricultural Engineering, Virginia Polytechnic, pp.139-148.

(Received 4 June 1993; revised and accepted 21 December 1993)

Appendix

Calculating the Area Under the Normal Distribution Curve

In calculating cell probabilities, it is necessary to calculate the area under the normal curve by evaluating the integral of the normal probability function. This requires a numerical approximation solution to the problem. One method is shown below (Rohlf and Sokal, 1981, p. 81) and approximates the integral with a maximum error of 7.5×10^{-8}. The left hand side of Equation 1 represents the value given in distribution tables for the argument X, which is the difference between the criticial point and the mean: i.e.,

$$P(X) - \frac{1}{2} = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} e^{\frac{-X^2}{2}} t\,(b_1 + t\,\{b_2 + t\,[b_3 + t\,(b_4 + t\,b_5)]\}) \quad (1)$$

where

$$p = 0.2316419$$
$$t = \frac{1}{(1 + pX)}$$
$$b_1 = 0.319381530$$
$$b_2 = -0.356563782$$
$$b_3 = 1.781477937$$
$$b_4 = -1.821255978$$
$$b_5 = 1.330274429$$

Gary J. Hunter

Gary Hunter is an Assistant Professor in the Department of Geomatics, and Deputy-Director of the Center for GIS and Modelling, at The University of Melbourne, Australia. He gained his PhD from that university on the subject of handling uncertainty in spatial databases. He teaches undergraduate and postgraduate students cadastral surveying, land law and administration, land development, and GIS. Before joining the university, he worked for the Australian Army on mapping surveys throughout South East Asia, and then on a variety of engineering and cadastral surveys for local and state government agencies in Victoria. He is a Licensed Surveyor in Victoria and was a national councillor of AURISA in 1990-92. His work has been published in the *International Journal of Geographical Information Systems, Photogrammetric Engineering*

& Remote Sensing, and *Cartography and Geographic Information Systems*. In 1993, he won the Horwood Prize with Michael Goodchild for the outstanding paper at the URISA '93 conference in Atlanta.

Michael F. Goodchild
Michael F. Goodchild is Professor of Geography at the University of California, Santa Barbara, and Director, National Center for Geographic Information and Analysis. He received his BA degree from Cambridge University in Physics in 1965 and his PhD in Geography from McMaster University in 1969. His research interests are in spatial analysis, the modeling of uncertainty in spatial data, and geographic data models. He is editor of three books on GIS, most recently *Environmental Modeling with GIS* (Oxford University Press).