THE ISSUE OF ACCURACY IN GLOBAL DATABASES

Michael F. Goodchild

ABSTRACT

Very few procedures exist for estimating the uncertainty or accuracy of products of spatial databases. Point objects can be treated by direct extension of the theory of measurement errors, but no suitable models exist for more complex spatial objects. The paper reviews the available techniques and describes work in progress. The prospects for error evaluation and reporting will improve if objects are tagged with relevant information, and if the data stored in the system can be obtained earlier in the chain of abstraction and generalization. It is argued that the cartographic representation of data should be regarded as a product rather than an input. The final section of the paper reviews available techniques for global spatial analysis, and identifies areas where further research is needed.

INTRODUCTION

The global databases now being constructed by a number of international agencies take data from a variety of different sources, each with their own characteristics of precision and accuracy. Data from remote sensors will be subject to the errors inherent in the classification and interpretation methods used, in the satellite platform, and in the sensor itself, so that the relationship between the truth as represented in the data which enters the database, and the truth as it appears on the ground, may be complex and obscure. Within the database the remotely sensed data may be combined with data from maps, which is subject to errors in digitizing, in the geodetic base on which the map was based, in the error inherent in the figure of the earth used as the basis for its projection, and again in the nature of the truth which is being represented.

In this paper we review the issues surrounding error and accuracy in global databases. Many of these occur irrespective of scale, but others are specific to the spherical geometry required at global scales. The first section of the paper discusses the state of knowledge about error modeling in spatial databases and spatial analysis, and techniques for dealing with error and estimating the accuracy of spatial database products. This is followed by a section on the more general issues arising from a discussion of error, including implications for methods of spatial data handling. The final section of the paper reviews the field of spatial analysis from a global perspective, and identifies areas where research is needed because existing methods are weak or incomplete.

ERROR MODELING IN SPATIAL DATABASES

Importance of error

The products of spatial databases may be simple responses to queries, such as the attributes of a point, the length of a line or the area of a selected polygon. More sophisticated applications are likely to result in more complex measures, such as calibrated parameters, goodness of fit statistics or simulated patterns of spatial variation. In all cases the accuracy of the results is to some extent determined by errors which will be present in the database. For example, a selected point on a soil map may not have the specified characteristics because the polygon in which it lies, and whose attributes were reported to the user, is not in fact homogeneous, although it was coded as such in the database. Similarly the area of the patch may not be a perfect estimate of the area on the ground having the coded attributes, again because of the lack of true homogeneity.

Because of the unique nature of spatial data, it is very unusual for any spatial data handling system to attempt to generate estimates of measurement error, or confidence limits, for its products. On the other hand numerous experiments have shown that actual levels of error, even for simple products such as point attributes, are far higher than one might expect given the precision with which the system operates, and in many cases unacceptable. Inaccurate products can lead to false inferences, bad decisions and even litigation (for a general discussion of error in spatial databases see Walsh, Lightfoot and Butler, 1987; for an excellent review and technical discussion see Burrough, 1986).

Error problems are likely to be worse in global than in local, planar databases, for several reasons. First, global databases must rely heavily on small scale data since global coverage at large scales is unlikely to be available; but small scale data is highly generalized, with much high- frequency spatial variation removed. Second, although data may be gathered at uniform spatial resolution, it is difficult to devise methods of digital representation and storage with similar uniformity. Spatial resolution varies over many common projections and over most common tesselations of the globe, leading to problems of estimating accuracy. Third, global databases are likely to combine data from numerous, incompatible sources, with problems of completeness and uniformity of coverage, and variations in definition. In fact many of the issues of data integration discussed at this conference impact directly on the question of product accuracy. For example, data from a satellite may be combined with digitized political boundaries; the two sources have entirely different types of error.

Definitions

We define the precision of a product in terms of its representation, as one half of the unit of the least significant digit. In principle one might wish that the product be reported to a precision equal to its accuracy, but in the absence of any clear knowledge of accuracy it is more likely that precision in a spatial data handling system will be determined by the software, particularly the number of digits carried in the internal arithmetic. Of course there will be occasions where for various computational reasons precision falls below accuracy (note that high precision implies a large number of significant digits, so the measure of precision will be numerically small), but it is normal to assume the reverse in most spatial data handling.

We define the accuracy of a product in terms of the magnitude of the difference between the reported value and the true value. The number of sources which can contribute to accuracy depends on how we choose to define truth. If the true value is as shown on the input document, then the errors include those introduced during the digitizing process, and during the manipulation and analysis in the system. However the input document may be an analog map, in which case the truth will be imprecise. Moreover it is unlikely that accuracy with respect to the input document will be considered acceptable, when the input document is itself an approximation to reality; accuracy with respect to ground truth is clearly more appropriate and realistic.

Unfortunately ground truth is often inaccessible or unavailable, and there are many instances of spatial data where no ground truth exists, since the object being represented in the system is itself an abstraction. Examples include soil maps, where a boundary shown on the map may be an abstraction of a transition zone between two generalized soil types.

A spatial database represents reality through objects and associated attributes. Objects may be points, lines, areas or pixels, with complex topological relationships derived from their spatial juxtaposition. Attributes may be nominal, ordinal, interval or ratio. Attributes are often assumed to be spatially invariant or homogeneous over the object; for example a soil map will associate the attributes of a polygon uniformly with every part of the polygon. Alternatively the system may infer systematic spatial variation of attributes over or between objects. For example, a topographic surface may be represented by a TIN, with height varying within each triangle in the form of a plane through the vertices, or according to a polynomial in the locational variables whose coefficients are stored as attributes of the triangle. Surfaces might also be modeled by point samples coupled with some spatial interpolation method such as Kriging, or by contours coupled with a linear interpolation algorithm.

We can partition error in the database into three types: errors in the positioning of objects, leading to distorted lines or point positions; errors in the attributes associated with objects, such as measurement errors at sample points; and errors in the modeling of spatial variation over or between objects, such as errors which result from assuming spatial homogeneity or linear interpolation. The simplest measure of accuracy for a variable would compare the database value at a randomly chosen point with the true value. While this would be an appropriate measure for query operations where the database must return the value at a selected point, it is not necessarily appropriate for more complex operations. For example, by arbitrarily offsetting the origin it is easy to see that one could estimate the areas of polygons accurately even though the database performed poorly on this measure of the accuracy of point attributes.

Measurement errors at points are relatively easy to describe using the Gaussian model, and the theory of errors can be used to predict their effects on derived products provided the analysis deals with each point independently. However it is much more difficult to deal with errors of positioning or modeling, because both are affected by the spatial relationships between objects.

Positioning errors

Consider a point object, such as a weather station. Errors in its position can be modeled as independent distortions of its two coordinates, although the respective standard errors may not be the same. The bivariate normal distribution is used to model distortion, and statistics based on it are used as standards in many mapping organizations. The Circular Map Accuracy Standard (CMAS) assumes a circular model with equal standard errors in the two coordinate directions and no correlation, and defines a circle within which the true point is expected to lie nine times out of ten.

Although the bivariate normal model is adequate for describing errors in individual points, it is likely that errors at neighboring points will show strong correlations, as the error in distances between nearby points is often much less than would be expected based on independent positioning errors. Similarly the errors at the vertices defining a patch are likely correlated, leading to lower than expected errors in patch area.

More difficult problems arise in defining positioning error for more complex objects such as lines or polygons. Most vector data structures represent complex objects as sequences of digitized points; it is assumed that the software will interpret the points as connected by straight line segments.

One solution to the problem of line error, and by extension polygon error, would be to model each point's accuracy using the bivariate, independent model, and to assume that the errors in the line derived entirely from errors in the points. Unfortunately this would be inadequate for several reasons. First, in digitizing a line an operator tends to choose points to be captured fairly carefully, selecting those which capture the form of the line with greatest economy. It would therefore be incorrect to regard the points as randomly sampled from the line, or to regard the errors present in each point's location as somehow typical of the errors which exist between the true line and the digitized representation of it. The latter combine the positioning errors at digitized points with modeling errors arising from differences between the behavior of the true line, and the model implied by the linear interpolation.

Secondly, the errors between the true and digitized line are not independent. If the true line is to the east of the digitized line at some location along the line, then it is very likely that its deviation immediately on either side of this location is also to the east by similar amounts. The relationship between true and digitized lines cannot therefore be modeled as a series of independent errors in point positions. To do so would lead to absurdity; the modeled line would be infinitely long, and would contain numerous loops or self-crossings.

One commonly discussed method of dealing with this problem is through the concept of an error band. The observed line is surrounded by a band of width epsilon, known as the Perkal epsilon band (Perkal 1956, 1966; Blakemore 1984; Chrisman 1982). The model has been described in both deterministic and probabilistic forms. In the former, it is proposed that the true line lies within the band with probability 1.0; in the latter, the band is compared to a standard error, or some average deviation from the true line. Blakemore (1984) has used the model to develop confidence limits on estimates of the attributes of a point.

The epsilon band model is too unspecified to be useful in modeling error in spatial databases. It is impossible to use it as a basis for simulation of distortion without some means of characterizing the position of the line within the band, and the autocorrelation of distortion. So although it might conceivably be used as the basis for predicting error in the attributes of a point, it is not sufficiently developed to allow estimates of error in line length, polygon area or more sophisticated forms of analysis.

Further clarification is necessary in dealing with line and polygon errors. In principle, we are concerned with the differences between some observed line, represented by a sequence of points with intervening straight line (or perhaps splined) segments, and a true line. The gross misfit between the two versions can be measured readily from the area contained between them, in other words the sum of the areas of the spurious or sliver polygons. To remove the effects of units of measurement we might divide

mismatch area by the square of the line length, although the length of the digitized line is not an unbiassed estimate of the length of the true line (Mandelbrot 1967; Maling 1968; Hakanson 1978).

To determine the mismatch for a single point is not as simple, however, since there is no obvious basis for selecting a point on the true line as representing the undistorted version of some specific point on the observed line. Most researchers in this field have made a suitable but essentially arbitrary decision, for example that the corresponding point on the true line can be found by drawing a line from the observed point which is perpendicular to the observed line. Using this rule, we can measure the linear displacement error of any selected point, or compute the average displacement along the line.

Consider a randomly chosen point on the line; this may or may not be a digitized point. Under the deterministic form of the Perkal epsilon band, this point estimates a true location which is with certainty within the epsilon band centered on the line. The probability distribution of linear differences is not specified in the model, but we might propose that it should be uniform within the error band, and zero outside. Under the probabilistic form one can calculate the probability that the true point lies within certain distances of the estimated position. Again the probability distribution is not specified. We could propose that it be normal, with mean, median and mode coincident with the estimated position, and with a standard deviation related to, and possibly equal to, epsilon. Honeycutt (1986) has accumulated evidence that the distribution of these distances is not in fact normal but bimodal, so that a finite, nonzero error on either side of the line is more likely than no error at all. Moreover the linear errors are clearly not independently distributed along the line, but strongly autocorrelated.

Studies of database error

The literature contains several discussions of the consequences of inaccurate response to queries about point attributes. In a vector database it is possible to overlay multiple polygon coverages, on the assumption that a point has the attributes of the polygon which contains it; the concatenated attributes of overlaid polygons can then be compared with ground truth. MacDougall (1975), Cook (1983), Newcomer and Szajgin (1984) and Chrisman (1987) have discussed the conceptual basis of error in map overlay and described simple experiments; in general the results show disappointingly high levels of error, even for large-scale databases.

Estimates of area imply queries about the attributes of assemblages of points, and therefore require reference to the spatial autocorrelation which is undoubtedly present in all spatial variables. Despite this there are numerous studies in the literature based on models of independence; for example Greenland and Socher (1985) used an independent norm in their study of mismatch between digital versions of the same objects (see also Rosenfield 1986; Congalton, Oderwald and Mead 1983).

Two forms of error in area estimation have been discussed based on models which incorporate appropriate spatial autocorrelation, either explicitly or implicitly. Mismatch error was defined previously as the total area between an observed and true line, or an observed and true polygon. Switzer (1975) described a model of mismatch error, which he related to the conditional probability that a point has the same (categorical) attributes as another point a given distance away. An empirical test of the model can be found in Muller (1978).

A common use of spatial databases is the estimation of the area of a coterminous patch. This might be done in a vector database by a calculation based on polygon vertices, or in a raster database by counting cells. The error of area estimation is typically much smaller than mismatch error because positive and negative errors tend to cancel eachother. An expression for the relationship between pixel size and the standard error of area estimates was obtained by Frolov and Maling (1969) using a simple statistical model, and generalized by Goodchild (1980) using fractional dimensions (see also Lloyd, 1976).

Models of database error

Although we can discuss positional, attribute and modeling errors as conceptually different sources of inaccuracy in spatial databases, it is clear that strong relationships exist between them. For example, on a soil map the boundaries represent transitions between attributes, and are therefore controlled by the same processes which determine the homogeneity of attributes within polygons, and the assignment of attributes themselves. Similarly positional, attribute and modeling errors in topographic

data are controlled by the same underlying processes controlling the topography itself.

A reasonable goal of research in error modeling might be a stochastic process capable of simulating intuitively acceptable errors. If this model could be defined and parametrized, then it could be used as the basis for confidence limits on products. The model might take one of two forms. First, the process might distort an existing map from its true appearance to a simulation of its observed form. Second, the process might have no relationship to any true map, but simulate a family of versions of the same random dataset; with appropriate parameters the dataset could be made to resemble a range of real coverages. We refer to these as the "truth plus distortion" and "artificial truth" approaches, respectively. The number of required parameters is clearly greater in the second case.

Goodchild and Dubuc (1987) have described an artificial truth process which begins with random surfaces or fields. One random surface will generate an isopleth map with contour or spot height features; two or more will generate a choropleth map or polygon coverage with associated attributes. The surfaces must be continuous but need not be differentiable; the degree of spatial autocorrelation is controlled through the form of the variogram function. The surfaces are then passed through a classifier of an appropriate number of dimensions. The results have the properties of choropleth maps and it is possible to produce a range of conditions by varying the parameters of the generating process. However the number of parameters is large, and it is unlikely that any useful calibration will be possible by analysis of real datasets.

It is possible to implement a truth plus distortion process when the data source is remotely sensed imagery. Suppose the product of the interpretation and classification process is a vector of probabilities for each pixel, consisting of the probabilities that the cell is a member of each of a set of prescribed classes. Several standard classification methods are capable of yielding such vectors, including discriminant and likelihood procedures. Simple independent trials in each pixel will yield distorted versions of the same map. However the result will be only weakly spatially autocorrelated, and interpolated boundaries will tend to be unreasonably complex. Instead, the probability vector for each pixel must be modified to take account of outcomes in neighboring pixels.

Given suitable vectors of probabilities and a spatially autocorrelated realization process, it may be possible to develop appropriate techniques for attaching confidence limits to spatial database products. These will clearly lead to more useful products than those based on assignment of each pixel to a single class.

IMPLICATIONS FOR SPATIAL DATA HANDLING

All of the methods reviewed in the previous section have a common characteristic; they attempt to estimate error based entirely on the information available in the database. In this section we consider the implications of this general strategy for spatial data handling.

The accuracy of an estimate from a soil map is dependent to a large extent on the validity of the assumption of homogeneity within each of the polygons in the database. However in most cases the database contains no explicit information on homogeneity. We can assume that homogeneity is more valid in the middle of each patch than at the edges, but there have been no systematic and extensive studies of the truth of this assumption. Indirect information on the homogeneity of patches is available in the configuration of the boundary, since it is likely that more heterogeneous patches will have more complex boundaries, but again there have been no objective tests of this relationship. Moreover the processes which control the configuration of the boundary are complex, and include the predilection of cartographers for smooth curves. In summary, a polygon database likely contains very imprecise and indirect information on the accuracy of the homogeneity assumption, and more generally, it is unlikely that models of distortion can be successfully calibrated from the contents of spatial databases.

It follows that if confidence limits are to be generated for spatial database products, then additional information must be included on which to base estimates. Two general approaches seem appropriate. The more straightforward argues for tagging of objects and coverages with relevant statistics; the second, for a more radical reexamination of the contents of many databases.

Accuracy tagging

The most readily accessible indicator of accuracy of a coverage is the scale of the manuscript, in the case of

digitized map documents, or the pixel size in the case of remote sensing and scanning. Unfortunately there is a tendency to see spatial data handling as scale-free; despite the ease with which products can be generated at any scale from the same database, the scale of the input document is a major determinant of the accuracy of those products. For example, the accuracy of many map products is published in the form of a linear CMAS, which translates into different distances on the ground depending on the scale of the map.

However although scale may be a reasonable indicator of positional error, it is largely unrelated to homogeneity or attribute error, or to the error in line and polygon features, all of which tend to be much larger than the scale would suggest. The accuracy of each object depends on the cartographic processes which generated it, in the form of abstraction and generalization, and these are sensitive both to scale and also to the nature of the object. For example, the processes determining the accuracy of road and river features on a topographic map are quite different, despite a possibly common scale of depiction. The curvature of a river feature is a direct function of its discharge, and therefore related indirectly to its width, whereas the curvature of a road feature is related to its class and position in the hierarchy of the road network, as well as to the cultural milieu.

If scale is a determinant of accuracy then it is sufficient to tag each coverage tile in the database by the scale of its source. But the arguments just presented suggest that each object should be tagged, either by direct measures of accuracy, or by indirect indicators, such as key attributes, which might be used as predictors of accuracy once a more comprehensive theory can be developed.

Rethinking the contents of spatial databases

Much of the input to spatial databases consists of digital versions of standard cartographic products, such as soil maps. We have seen how the lack of accuracy information in such products leads to difficulties in spatial analysis, as there is no alternative but to take such data at face value, despite the obvious errors inherent in assuming homogeneity.

The cartographic view of spatial data is a highly abstracted representation, designed to display data in the most informative manner possible under the constraints of pen and paper technology. For example, the topographic map is conventionally displayed using contours, because these have been found to give informative and readily comprehended height information, and can be generated by hand using pen and paper with reasonable efficiency. Alternatives which might be more informative, such as perspective views, have not been part of the cartographic tradition because of the difficulty of generating them in a manual environment.

Spatial data handling technology is not as constrained as cartography; it is easy to generate perspective views, for example, from digital representations of terrain such as TINs or DEMs. It would make little sense to digitize the cartographic view of contours if a DEM were available, since the accuracy of heights interpolated from contours is uneven, and deteriorates rapidly as one moves away from the contours themselves. Instead much more accurate products could be generated if the DEM were input directly, allowing the cartographic view to be generated if and when desired.

In general, and from the perspective of product accuracy being taken in this paper, a spatial database should represent reality, using data which is as raw and unabstracted as possible, rather than the cartographic representation of reality, which is often highly abstracted. If the cartographic view is necessary, it can be generated from the database using appropriate rules. It is much easier to assign indices of accuracy to raw data than to abstractions and interpretations.

SPATIAL ANALYSIS ON THE GLOBE

Most of the discussion thus far has concerned general databases, independently of the specific context of the globe. We have assumed, for example, that in the case of raster databases we are dealing with a uniform tesselation and a constant pixel size. In a global database it is likely that pixels will vary in size, or that the tesselation will be nonuniform, or both (see for example Tobler and Chen 1986; Mark and Lauzon 1986).

There has been relatively little explicit consideration of the unique problems of handling global data. In some fields there exists an almost complete literature on the adaptation of standard problems in planar spatial analysis to the sphere (problems of non-sphericity will be ignored in this discussion) while in other fields there is a clear

need for the development of appropriate methods. We define a method as appropriate if its results are invariant under rotation of the spherical referencing system.

One field which has received significant attention is spatial interpolation, which has obvious applications in spatial data handling, particularly in generalization from point samples to continuous surfaces. Willmott, Rowe and Philpot (1985) and Legates and Willmott (1986) have adapted distance-based methods of spatial interpolation to the sphere in methods which ensure appropriate continuity across the poles. Spline interpolation has been developed for the sphere by Wahba (1981), Freeden (1981), Lawson (1984), Renka (1984) and Dierckx (1984); Jupp and Kent (1987) have considered the special problem of fitting a smooth path to time-dependent events on the sphere. Finally spectral methods are particularly appropriate on the sphere and have been used for spatial interpolation by Balmino, Lambeck and Kaula (1973), Swarztrauber (1979) and Dierckx (1986).

Statistical hypothesis tests for spatial distributions on the sphere can be developed from spherical distributions. Mardia (1972) and Batschelet (1981) review a number of stochastic processes yielding unimodal, bimodal and linear distributions on the sphere and additional tests are given by Costanzo and Gale (1984). While these can be used as the basis for models of error in point positions on the sphere, as with planar data there are no obvious ways of dealing with more complex objects.

The problem of measuring central tendency and dispersion on the sphere was first discussed by Fisher (1953). A more recent and extensive literature deals with the problems of finding one or more locations on the sphere which minimize functions of distance, in other words spherical extensions of planar locational analysis (see for example Drezner and Wesolowsky 1978; Aly, Kay and Litwhiler 1979; Katz and Cooper 1980; Drezner 1981, 1983, 1985; Wesolowsky 1983; Drezner and Wesolowsky 1983). One of the more interesting motivations for this research concerns the optimal paths of earth observing satellites.

Many of the standard methods of planar spatial analysis have yet to be adapted to the context of the sphere. For example there is as yet no extension of the Douglas and Poiker (1973) line generalization algorithm, and only limited literature on the generation of Thiessen polygons and polygon skeletons. There is no spherical version of point pattern analysis, and no literature on spatially autocorrelated processes. It is clear that much research needs to be done in developing a complete set of spatial analytic techniques for the spherical case.

CONCLUDING REMARKS

In this paper we have considered a number of outstanding issues in spatial analysis and spatial statistics which affect the development and use of global databases. Some of these are common to the planar case; there is a similar lack of methods for attaching confidence limits to the products of spatial data handling in both cases. Others are unique, and the previous section has identified some of the areas in which there is a need for further development of analytic methods.

It would be naive to suppose that suitable models of error can be devised, particularly in the short term. There are good reasons to believe that appropriate models will be difficult to calibrate and deal with because of the large numbers of parameters involved. Effective solutions are likely to require much better understanding of the processes which create spatial variation than we currently possess. However the incentive to develop such understanding is substantially higher now than it has been, because of the development of spatial data handling technology. And irrespective of the number of parameters, there are valid uses of simulation models in studying the effects of controlled data parameters on the performance of spatial data handling systems.

Perhaps the greatest improvement in the understanding of error in spatial databases will come when the database representation is seen as the source rather than the product of the cartographic view of the data. If objects can be tagged with accuracy information; if the raw data from which the cartographic view was derived can be included in the database; and if the products of remotely sensed imagery can be tagged with vectors of class probabilities, then far more information will be available on which to base estimates of reliability.

REFERENCES

Aly, A.A., D.C. Kay and D.W. Litwhiler, Jr., 1979. "Location dominance on spherical surfaces". Operations Research 27 972-81.

Balmino, G., K. Lambeck and W.M. Kaula, 1973. "A spherical harmonic analysis of the earth's topography". Journal of Geophysical Research 78 478-81.

Batschelet, E., 1981, Circular Statistics in Biology. Academic Press, London and New York.

Blakemore, M., 1984. "Generalization and error in spatial databases". Cartographica 21 131-9.

Burrough, P.A., 1986. Principles of Geographic Information Systems for Land Resources Assessment. Monographs on Soil and Resources Survey No 12. Oxford University Press.

Chrisman, N.R., 1982. "Methods of spatial analysis based on errors on categorical maps". Unpublished Ph.D. thesis, University of Bristol.

Chrisman, N.R., 1987. "The accuracy of map overlays: a reassessment". Landscape and Urban Planning 14 427-39.

Congalton, R.G., R.G. Oderwald and R.A. Mead, 1983. "Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques". Photogrammetric Engineering and Remote Sensing 49 1671-8.

Cook, B.G., 1983. "Geographic overlay and data reliability". In Proceedings of the US/Australia Workshop on Design and Implementation of Computer-Based Geographical Information Systems ed. D. Peuquet and J. O'Callaghan. IGU, Amherst, New York, 64-9.

Costanzo, C.M. and N. Gale, 1984. "Evaluating the similarity of directional data". Professional Geographer 36 182-7.

Dierckx, P., 1984. "Algorithms for smoothing data on the sphere with tensor product splines". Computing 32 319-42.

Dierckx, P., 1986. "Spectral approximation on the sphere". SIAM Journal on Scientific and Statistical Computing 7 611-23.

Drezner, Z., 1981. "On location dominance on spherical surfaces". Operations Research 29 1218-9.

Drezner, Z., 1983. "Constrained location problems in the plane and on a sphere". IIE Transactions 15 300-4.

Drezner, Z., 1985. "A solution to the Weber location problem on the sphere". Journal of the Operational Research Society 36 333-4.

Drezner, Z. and G.O. Wesolowsky, 1978. "Facility location on a sphere". Journal of the Operational Research Society 29 997-1004.

Drezner, Z. and G.O. Wesolowsky, 1983. "Minimax and maximin facility location problems on a sphere". Naval Research Logistic Quarterly 30 305-12.

Douglas, D.H. and T.K. Peucker, 1973. "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature". Canadian Cartographer 10 112-22.

Fisher, R.A., 1953. "Dispersion on a sphere". Proceedings of the Royal Society of London, Series A 217 295-305.

Freeden, W., 1981. "On spherical spline interpolation and approximation". Mathematical Methods for the Applied Sciences 3 551-75.

Frolov, Y.S. and D.H. Maling, 1969. "The accuracy of area measurements by point counting techniques". Cartographic Journal 6 21-35.

Goodchild, M.F., 1980. "A fractal approach to the accuracy of geographical measures". Mathematical Geology 12 85-98.

Goodchild, M.F. and O. Dubuc, 1987. "A model of error for choropleth maps with applications to geographic information systems". Proceedings, AutoCarto 8. ASPRS/ACSM, Falls Church, Virginia, 165-74.

Greenland, A. and R.M. Socher, 1985. "Statistical evaluation of accuracy for digital cartographic data bases". Proceedings, AutoCarto 7. ASPRS/ACSM, Falls Church, Virginia, 212-21.

Hakanson, L, 1978. "The length of closed geomorphic lines". Mathematical Geology 10 141-67.

Honeycutt, D.M., 1986. "Epsilon, generalization and probability in spatial data bases". Unpublished manuscript.

Jupp, P.E. and J.T. Kent, 1987. "Fitting smooth paths to spherical data". Applied Statistics 36 34-46.

Katz, I.N. and L. Cooper, 1980. "Optimal location on the sphere". Computers and Mathematics with Applications 6 175-96.

Lawson, C.L., 1984. "C1 surface interpolation for scattered data on a sphere". Rocky Mountain Journal of Mathematics 14 177-202.

Legates, D.R. and C.J. Willmott, 1986. "Interpolation of point values from isoline maps". American Cartographer 13 308-23.

Lloyd, P.R., 1976. "Quantization error in area measurement". Cartographic Journal 13 22-6.

MacDougall, E.B., 1975. "The accuracy of map overlays". Landscape Planning 2 23-30.

Maling, D.M., 1968. "How long is a piece of string?". Cartographic Journal 5 147-56.

Mandelbrot, B.B., 1967. "How long is the coast of Britain? Statistical self-similarity and fractional dimension". Science 156 636-8.

Mardia, K.V., 1972. Statistics of Directional Data. Academic Press, London and New York.

Mark, D.M. and J.P. Lauzon, 1986. "Approaches for quadtree-based geographic information systems at continental and global scales". Proceedings, AutoCarto 7. ASPRS/ACSM, Falls Church, Virginia, 355-64.

Muller, J.-C., 1978. "Map gridding and cartographic errors: a recurrent argument". Canadian Cartographer 14 152-67.

Newcomer, J.A. and J. Szajgin. "Accumulation of thematic map error in digital overlay analysis". American Cartographer 11 58-62.

Perkal, J., 1956. "On epsilon length". Bulletin de l'Academie Polonaise des Sciences 4 399-403.

Perkal, J., 1966. "On the length of empirical curves". Discussion Paper No. 10, Michigan Inter-University Community of Mathematical Geographers, Ann Arbor.

Renka, R.J., 1984. "Interpolation of data on the surface of a sphere". ACM Transactions on Mathematical Software 10 417-36.

Rosenfield, C.H., 1986. "Analysis of thematic map classification error matrices". Photogrammetric Engineering and Remote Sensing 52 68-6.

Swarztrauber, P.N., 1979. "On the spectral approximation of discrete scalar and vector functions on the sphere". SIAM Journal of Numerical Analysis 16 934-49.

Tobler, W.R. and Zi-tan Chen, 1986. "A quadtree for global information storage". Geographical Analysis 18 360-71.

Wahba, G., 1981. "Spline interpolation and smoothing on the sphere". SIAM Journal of Scientific and Statistical Computing 2 5-16.

Walsh, S.J., D.R. Lightfoot and D.R. Butler, 1987. "Recognition and assessment of error in geographic information systems". Photogrammetric Engineering and Remote Sensing 53 1423-30.

Wesolowsky, G.O., 1983. "Location problems on a sphere". Regional Science and Urban Economics 12 495-508.

Willmott, C.J., C.M. Rowe and W.D. Philpot, 1985. "Small-scale climate maps: a sensitivity analysis of some common assumption associated with grid-point interpolation and contouring". American Cartographer 12 5-16.