

UC Santa Barbara

NCGIA Technical Reports

Title

Taxonomy of Error in Spatial Databases (89-12)

Permalink

<https://escholarship.org/uc/item/87s8f0mg>

Author

Veregin, Howard

Publication Date

1989-12-01

NCGIA

**National Center for
Geographic Information and Analysis**

Taxonomy of Error in Spatial Databases

By:

**Howard Veregin
University of California, Santa Barbara**

Technical Report 89-12

December, 1989

**Simonett Center for Spatial Analysis
University of California**
35 10 Phelps Hall
Santa Barbara, CA 93106-4060
Office (805) 893-8224
Fax (805) 893-8617
ncgia@ncgia.ucsb.edu

State University of New York
301 Wilkeson Quad, Box 610023
Buffalo NY 14261-0001
Office (716) 645-2545
Fax (716) 645-5957
ncgia@ubvms.cc.buffalo.edu

University of Maine
348 Boardman Hall
Orono ME 04469-5711
Office (207) 581-2149
Fax (207) 581-2206
ncgia@spatial.maine.edu

TABLE OF CONTENTS

1. Introduction.....	
2. Measurement of Error in Spatial Databases.....	
2.1. Classification Accuracy.....	
2.1.1. PCC as an Index of Classification Accuracy.....	
2.1.2. Alternatives to PCC.....	
2.1.3. Alternatives to the Classification Error Matrix.....	
2.1.4. Other Issues Pertaining to Classification Accuracy.....	
2.2. Soil Properties.....	
2.2.1. Variations in Soil Properties.....	
2.2.2. Models of Soil Property Variation.....	
2.3. Topographic Maps.....	
2.4. Planimetric Accuracy.....	
2.5. Geographic Base File Standards.....	
2.6. Digital Elevation Models.....	
3. Accuracy of Cartometric Estimates.....	
3.1. Length Estimation.....	
3.1.1. Sources of Error.....	
3.1.2. Fractal Theory.....	
3.2. Area Estimation.....	
3.2.1. Effects of Dot Grid Density.....	
3.2.2. Correcting for Classification Error.....	
4. Errors Introduced during Data Compilation.....	
4.1. Digitizing Error.....	
4.1.1. The Epsilon Band Concept.....	
4.1.2. Digitizing Error as a Serially Dependent Process.....	
4.1.3. The Role of Human Factors.....	

4.2. Generalization Error	
4.3. Choroplethic Mapping	
4.3.1. Data Symbolization.....	
4.3.2. Interactions between Attribute and Positional Error	
4.4. Isometric Mapping	
4.4.1. Factors Affecting Interpolation Accuracy	
4.4.2. Interactions between Interpolation and Measurement Error.....	
4.5. Isoplethic Mapping	
4.6. Cartographic Communication	
5. Propagation of Error through GIS Operations	
5.1. Areal Interpolation	
5.2. Vector to Raster Conversion	
5.2.1. Models of Positional Error.....	
5.2.2. Models of Attribute Error	
5.3. Map Overlay	
5.3.1. Factors Affecting Accuracy.....	
5.3.2. The Spurious Polygon Problem	
5.3.3. Models of Attribute Error	
6. General Issues of Spatial Database Accuracy	
6.1. Modeling Uncertainty.....	
6.2. Geographical Modeling.....	
6.3. Data Quality Issues	
7. Summary	
8. References.....	
9. Author Index	
10. Subject Index	

PREFACE

This report was prepared in conjunction with Research Initiative No. 1 of the National Center for Geographic Information and Analysis. It is being distributed as an aid to research.

The report consists of a review of a body of literature focusing on the issue of spatial database accuracy. The error taxonomy developed in the report organizes different dimensions of accuracy into a practical, user-oriented structure that reflects the interests of those conducting both pure and applied research.

Approximately one-half of the articles reviewed in this report have been published in the professional literature. Conference papers account for the largest share of the remainder, which also includes technical reports, chapters of books and unpublished manuscripts. A less detailed discussion of this same body of literature is provided in NCGIA Technical Paper 89-9, entitled *Accuracy of spatial databases: Annotated bibliography*.

1. INTRODUCTION

The issue of spatial database accuracy is critical to the successful implementation and long-term viability of GIS technology. The value of GIS as a decision-making tool is dependent on the ability of decision-makers to evaluate the reliability of the information on which their decisions are based. Users of GIS technology must therefore be able to assess the nature and degree of error in spatial databases, track this error through GIS operations and estimate accuracy for both tabular and graphic output products.

Evaluation of spatial database accuracy in a GIS environment encompasses a variety of concepts, methods and models. To make matters more complex, the significance of different dimensions of accuracy is a function of data type, application and the sources of error deemed to be important in a particular context. This report consists of a review of a body of literature focusing on various dimensions of spatial database accuracy. The error taxonomy developed in the report organizes these dimensions into a practical, user-oriented structure that reflects the interests of those conducting both pure and applied research. The taxonomy recognizes that different classes of spatial data exhibit different types of error, that error may be introduced during various stages of data compilation and that error may be propagated through spatial operations to appear in modified form in output products.

The exact structure of the taxonomy is revealed in the sections and sub-sections of the report. On the broadest level, the taxonomy identifies five main dimensions of spatial database accuracy — measurement of error in spatial databases (§ 2), accuracy of cartometric estimates (§ 3), errors introduced during data compilation (§ 4), propagation of error through GIS operations (§ 5) and general issues of spatial database accuracy (§ 6). These dimensions are further broken down into secondary and tertiary levels in the report. The relationships between different dimensions in the taxonomy are explicitly identified at appropriate junctures in the text.

The literature review presented in this report is intended as a general overview of the concepts, methods and models developed for evaluating spatial database accuracy in a variety of contexts. Interested readers are referred to the specific articles cited in the text for a more detailed and thorough discussion. Readers will no doubt find that certain dimensions of accuracy have not been dealt with comprehensively in the report. Due to the usual constraints, it has been impossible to compile a thoroughly exhaustive review of all relevant articles and issues. Although incomplete, the report will hopefully serve as a valuable starting point for future research.

2. MEASUREMENT OF ERROR IN SPATIAL DATABASES

This section is concerned with methods for detecting and measuring errors in spatial databases. Various aspects of attribute error are discussed in § 2.1 and § 2.2. The discussion in § 2.1 focuses on classification error in a remote sensing context, or error in databases depicting a categorical attribute derived by classifying remotely sensed imagery. The nature of attribute error in soil maps arising from spatial variations in measured soil properties is discussed in § 2.2. The remaining sub-sections focus primarily on positional error in spatial databases. Data quality standards and methods of evaluating horizontal and vertical error in topographic maps are reviewed in § 2.3. The issue of horizontal accuracy for large-scale planimetric and cadastral maps is discussed in § 2.4. General data quality standards for multi-purpose digital geographic base files are examined in § 2.5. Finally, § 2.6 is concerned with methods for classifying and correcting vertical errors in digital elevation models.

2.1. Classification Accuracy

In remote sensing, classification is often performed by assigning each pixel in an image to one of a set of classes based on the spectral response of the pixel in one or more spectral bands. The resulting map, or classified image, contains a set of polygons defined as groups of contiguous pixels of the same class. Various methods then exist for assessing the accuracy of the classification procedure. For example, one might compare the areas of a sample of polygons on the classified image to their actual areas as determined by ground survey. Alternatively, one could compute the positional error in the boundaries of a sample of polygons with reference to more accurate data. However, the most common approach in assessing classification accuracy is to select a sample of pixels from the image and compare their assigned and actual classes, the latter being determined from ground survey or some data source of higher accuracy. The resulting cross-tabulation is variously referred to as a "classification error matrix," a "confusion matrix" or a "contingency table." The classification error matrix, denoted as C , is a $k \times k$ matrix where k is the number of classes and c_{ij} is the number of sample pixels assigned to class i that actually belong to true class j . The number of pixels in the sample that belong to true class j , t_j , is given by the column sum

$$t_j = \sum_{i=1}^k c_{ij} \quad (2.1)$$

The number of pixels in the sample assigned to class i on the map, m_i , is given by the row sum

$$m_i = \sum_{j=1}^k c_{ij} \quad (2.2)$$

The total number of pixels in the sample, n , is given by

$$n = \sum_{i=1}^k \sum_{j=1}^k c_{ij} \quad (2.3)$$

Early research on classification accuracy focuses on the proportion of pixels correctly classified (or PCC) as an index of classification accuracy. The sample PCC, usually denoted \hat{p} , is defined as the trace of C divided by the number of pixels in the sample. That is,

$$\hat{p} = \frac{1}{n} \sum_{i=1}^k c_{ii} \quad (2.4)$$

(The parameters defined in equations (1) through (4) are used throughout the remainder of this section.) When the sample is error-free, \hat{p} will equal 1, its maximum value, since in this case C is a diagonal matrix (i.e., the only non-zero elements are on the main diagonal). Many of the earliest studies of classification accuracy focus on the formulation of statistical tests for making inferences about the population PCC based on the sample PCC. Particularly when using small samples, there is a relatively

high probability of finding only a few misclassifications in the sample even if accuracy is quite low for the classified image as a whole. Typically, these tests are based on the binomial distribution or the normal approximation to this distribution, although a number of other tests have also been proposed. These tests are often formulated in terms of simple random sampling, although it is generally agreed that stratified random sampling is a more appropriate sampling methodology. In stratified random sampling, pixels are selected randomly from each of the k classes, such that each row total of C is sufficiently large to permit statistical testing. Further research has identified several problems with PCC as an index of classification accuracy and a number of alternate indices have been proposed. An example is the $\hat{\kappa}$ statistic, which, in contrast to PCC, accounts for the fact that correct pixel assignments can occur by chance.

2.1.1. PCC as an Index of Classification Accuracy

Fitzpatrick-Lins [82] used a classification error matrix to evaluate classification accuracy for a set of 1:100 000 and 1:24 000-scale land use/land cover maps. A sample of points was selected by stratified systematic unaligned sampling, in which points were selected randomly within a set of grid cells. Following the construction of the classification error matrix, the sample PCC was computed. Confidence limits were then computed for the sample PCC using the normal approximation to the binomial distribution,

$$\hat{p} \pm \left[z_{\alpha/2} \left(\frac{\hat{p}(1 - \hat{p})}{n} \right)^{1/2} + \frac{50}{n} \right] \quad (2.5)$$

where: $z_{\alpha/2}$ = the quantile from the standard normal distribution corresponding to a two-tailed confidence level of $1 - \alpha$; and

$50/n$ = a correction factor to account for the discrete nature of the binomial distribution.

The interpretation of the confidence limits is that there is a probability of $1 - \alpha$ that the population PCC falls within the range defined by the upper and lower confidence limits. Equation (2.5) may also be used to compute the minimum sample size required to attain a certain level of accuracy with a certain confidence. Manipulation of the equation (ignoring the correction factor) gives a minimum sample size, n_{\min} , of

$$n_{\min} = p(1 - p) / s^2 \quad (2.6)$$

where: p = the desired population PCC; and

s = the desired standard error for the population PCC.

Fitzpatrick-Lins [83] used a similar approach to evaluate the accuracy of a 1:250 000-scale land use/land cover map. A sample of points was selected by stratified systematic unaligned sampling followed by random sampling within each land use/land cover class to obtain a sufficiently large sample of points for each class. Unlike simple random sampling, application of this sampling scheme implies that the number of points selected from each class is not proportional to the incidence of that class on the map. Hence in calculating the sample PCC, the PCC for each class must be weighted by the relative area of the class as estimated from the map. That is,

$$\hat{p} = \sum_{i=1}^k \hat{p}_i a_i \quad (2.7)$$

where: \hat{p}_i = the sample PCC for class i (i.e., c_{ii}/m_i); and

a_i = the relative area of class i on the map (i.e., the area assigned to class i on the map, expressed as a proportion of total map area).

The confidence limits for \hat{p} are then given as

$$\hat{p} \pm (z_{\alpha/2} v^{1/2} + 50/n) \quad (2.8)$$

where:

$$v = \sum_{i=1}^k \frac{a_i^2 \hat{p}_i (1 - \hat{p}_i)}{m_i} \quad (2.9)$$

Rosenfield *et al* [193] present a similar set of equations.

Thomas & Allcock [206] present a similar approach for calculating confidence limits for the number of correctly classified pixels rather than the sample PCC. The confidence limits are defined as

$$x \pm z_{\alpha/2} \left(\frac{x(n-x)}{n} \right)^{1/2} - e \quad (2.10)$$

where: x = the number of correctly classified pixels in the sample; and

e = a measure of human error in counting the number of sample pixels correctly classified.

The interpretation of the confidence limits is that there is a probability of $1-\alpha$ that the number of correctly classified pixels in the entire image falls within the range defined by the upper and lower confidence limits.

Minimum sample size and confidence limits for the sample PCC may also be derived from the binomial distribution, as shown by van Genderen & Lock [89]. According to the binomial distribution, the probability of observing exactly x correctly classified points in a sample of size n , $P[x]$, is given by

$$P[x] = \frac{n!}{(n-x)! x!} p^x (1-p)^{n-x} \quad (2.11)$$

where: p = the unknown population PCC.

When $x = n$ (i.e., no misclassifications are observed in the sample), equation (2.11) reduces to

$$P[x] = p^n \quad (2.12)$$

Using this equation the authors were able to compile a table showing the value of $P[x]$ for different values of p and n . The table indicates that, particularly for small samples, the probability of observing no misclassifications can be quite high, even when the population PCC is relatively low. Given a desired population PCC, equation (2.12) can be used to calculate the minimum sample size required to achieve a specified confidence level. (The cumulative binomial distribution is actually more appropriate, as discussed below.) This is achieved by substituting the desired population PCC for p and computing the minimum n for which $P[x] \leq 1-\alpha$, where α is the specified confidence level. For example, if the desired PCC is 0.85 and the specified confidence level is 0.05 then the minimum sample size is 19, since $0.85^{19} = 0.046$ while $0.85^{18} = 0.054$.

This method must be extended to account for situations in which one or more misclassifications are observed in the sample (van Genderen *et al* [90]). Having determined the appropriate sample size according to the method described above, it is then possible to calculate the probability of observing exactly x correctly classified points in a sample of size n for a given population PCC of p . Taking the preceding example, if $n = 19$, $p = 0.85$ and 1 misclassification is observed in the sample (i.e., $x = 18$), then by equation (2.11), $P[x] = 0.153$. This forms the basis for a statistical test that employs the cumulative binomial distribution. Consider a situation in which one wishes to infer whether the population PCC is at least as high as some specified value, p_0 (e.g., a minimum accuracy standard). Thus the null hypothesis states that $p \geq p_0$, while the alternate hypothesis states that $p < p_0$. The cumulative probability of observing x or fewer correctly classified points in a sample of size n drawn from a population with an accuracy of p_0 is given by

$$P[r \leq x] = \sum_{r=0}^x \frac{n!}{(n-r)! r!} p_0^r (1-p_0)^{n-r} \quad (2.13)$$

If this probability is higher than some specified level, α , then the test indicates that the null hypothesis that $p \geq p_0$ should not be rejected. Typically, $\alpha = 0.05$, representing a confidence level of 0.95. However, if the probability is less than α , the null hypothesis should be rejected, since there is a very low probability that only as many as x correct classifications will be observed in the sample if the population PCC is actually p_0 . The value of α gives the probability that the null hypothesis will be rejected when it is true. Hence α represents the “producer’s risk,” or the probability that an accurate map will fail the accuracy test and thus be rejected as insufficiently accurate.

In addition to producer’s risk, it is also important to consider “consumer’s risk,” or the probability that a map of insufficient accuracy will pass the accuracy test and be accepted as sufficiently accurate. In this case, the null hypothesis states that $p < p_0$, while the alternate hypothesis states that $p \geq p_0$. The cumulative probability of observing more than x correctly classified points in a sample of size n drawn from a population with an accuracy of p_0 is given by

$$P[r > x] = \sum_{r=x+1}^n \frac{n!}{(n-r)! r!} p_0^r (1-p_0)^{n-r} \quad (2.14)$$

If this probability is larger than some specified value, β , then the test indicates that the null hypothesis that $p < p_0$ should not be rejected. However, if the probability is less than β , the null hypothesis should be rejected, since the test indicates that there is a very low probability that more than x correct classifications will be observed if the population PCC is actually p_0 . The value of β represents the probability that the null hypothesis will be rejected when it is true, and hence β is the consumer’s risk.

Producer’s and consumer’s risk may be calculated for the entire sample if simple random sampling is employed. In the case of stratified random sampling, these probabilities must be weighted to account for the disproportionate sample size of certain classes. Aronoff [8, 9, 10], Ginevan [94] and Rosenfield *et al* [193] describe hypothesis tests for producer’s and consumer’s risk and illustrate how these tests may be formulated in terms of a critical value of x . (Note that nomenclature varies considerably among the authors.)

Tests of producer’s and consumer’s risk may be combined to minimize the probability of accepting an inaccurate map and maximize the probability of accepting an accurate map (Aronoff [8] and Ginevan [94]). First, one selects the lowest level of acceptable accuracy, p_L , the consumer’s risk, β , and an initial sample size, n . Equation (2.14) is then evaluated using the selected values of n and p_L (in place of p_0) to find the smallest value of x such that $P[r > x] \leq \beta$. Next, a selected high level of accuracy, p_H , is inserted into equation (2.13) in place of p_0 . This equation is then evaluated to yield the producer’s risk, $\alpha = P[r \leq x]$. The value of n can be manipulated such that the values of α and β reflect the consequences of potential errors. Note that when $p_L = p_H$, $\alpha + \beta = 1$.

2.1.2. Alternatives to PCC

While PCC remains a standard index of classification accuracy, a number of authors have proposed alternative indices that circumvent some of the limitations of PCC. One such limitation is that PCC does not differentiate between “errors of omission” (i.e., the omission of a point from its true class) and “errors of commission” (i.e., the assignment of a point to an incorrect class). A second limitation is that PCC does not account for correct classifications that occur by chance alone. Hence it confuses the accuracy of a classification procedure with the accuracy of its outcome. PCC is also highly sensitive to the row and column totals of the classification error matrix, such that a high PCC may be obtained if the incorrectly classified classes represent only a small percentage of the total sample.

Errors of omission and commission can be differentiated by computing the PCC for each row and column of the classification error matrix (Story & Congalton [202]). “Producer’s accuracy” (not to be confused with producer’s risk) for a given class is defined as the number of correctly classified points for that class divided by the corresponding column total. That is, the producer’s accuracy for class j is defined as c_{jj}/t_j . This value represents the probability that a point actually belonging to class j has been correctly assigned to this class. A value less than 1 is indicative of omission error (i.e., the omission of one or more points from their true classes). “User’s accuracy” for a given class is defined as the number of correctly classified points for that class divided by the corresponding row total. That is, user’s accuracy for a given class i is defined as c_{ii}/m_i . This value represents the probability that a point assigned to class i actually belongs to that class. A value less than 1 is indicative of commission error (i.e., the assignment of one or more points to an incorrect class). These indices have been employed in an operational setting by Robinson *et al* [185] and Kenk *et al* [134].

Consumer’s risk may also be extended to individual classes (Aronoff [9, 11]). The “minimum accuracy value” is defined as the highest value of p_0 in equation (2.14) for which the inequality $P[c > x] \leq \beta$ still holds for an observed number of correct classifications x in a sample of size n . This may be computed for the sample as a whole or for individual classes. In the latter case, the result is referred to as the “minimum class accuracy.” The minimum class accuracy may also be used to calculate a “loss function” for each class as

$$L_i = (1 - q_i) w_i m_i \quad (2.15)$$

where: L_i = the maximum expected loss for class i ;
 q_i = the minimum class accuracy for class i ; and
 w_i = a weight describing the cost of misclassification for class i .

The maximum expected loss for the whole map is equal to the sum of the maximum expected losses for each class. The maximum expected loss can be used to compare the relative costs of different classification procedures. Aronoff [10] illustrates how the costs of different misclassifications can be computed and describes a classification algorithm that minimizes the loss function for each class.

Another alternative to PCC is the $\hat{\kappa}$ (or KHAT) statistic. Unlike PCC, $\hat{\kappa}$ accounts for correct classifications that occur by chance alone. The formulation of $\hat{\kappa}$ is given by Hudson & Ramm [125] as

$$\hat{\kappa} = \frac{\hat{p} - \Theta}{1 - \Theta} \quad (2.16)$$

where:

$$\Theta = \frac{1}{n^2} \sum_{i=1}^k m_i t_i \quad (2.17)$$

The $\hat{\kappa}$ statistic thus represents the ratio of beyond-random agreement to expected disagreement in a random case. The value of $\hat{\kappa}$ is equal to 0 for random agreement and 1 for perfect agreement. It is negative when agreement is less than that expected in a random case. Experiments performed by Rosenfield & Fitzpatrick-Lins [192] show that, relative to $\hat{\kappa}$, PCC gives an inflated estimate of classification accuracy by ignoring chance agreement. Congalton *et al* [55] and Hudson & Ramm [125] give the approximate large sample variance of $\hat{\kappa}$. (Note that there is a small typographical error in the formula given by Hudson & Ramm.)

The $\hat{\kappa}$ statistic may also be computed for individual classes in the classification error matrix. For a given class i , $\hat{\kappa}_i$ is defined as

$$\hat{\kappa}_i = \frac{n c_{ii} - m_i t_i}{n m_i - m_i t_i} \quad (2.18)$$

Experimental results obtained by Rosenfield & Fitzpatrick-Lins [192] indicate that, relative to $\hat{\kappa}_i$, the user's accuracy gives an inflated estimate of accuracy. A number of authors have employed the $\hat{\kappa}$ statistic in conjunction with other indices of classification accuracy described above (e.g., Robinson *et al* [185] and Dicks & Lo [71]).

Congalton & Mead [54] and Congalton *et al* [55] describe how $\hat{\kappa}$ may be used to compare alternate classification methods. (Note that the formulation of $\hat{\kappa}$ given by the authors is incorrect.) Values of $\hat{\kappa}$ are computed from the classification error matrix associated with each method. The confidence interval for each value is then calculated based on the variance of $\hat{\kappa}$. If any two confidence intervals overlap, the difference between the accuracies of the two classification methods is assumed to be statistically insignificant.

Carstensen [40] demonstrates how the $\hat{\kappa}$ statistic may be used to quantify the level of agreement between different maps depicting nominal attributes (e.g., soil classes). The approach consists in superimposing two maps and constructing a classification error matrix based on the areas of agreement and disagreement. The value of $\hat{\kappa}$ is then computed from the classification error matrix. In the case of ordinal attributes the approach must be modified by calculating a weighted $\hat{\kappa}$ statistic. Various weighting schemes are described by Greenland *et al* [105]. For example, the elements of the two minor diagonals adjacent to the main diagonal of the classification error matrix can be assigned weights of 0.5. In this case, a disagreement of one ordinal class may be envisaged as half an error. Alternatively, the elements of each minor diagonal in the classification error matrix can be assigned weights inversely proportional to the distance between the minor diagonal and the main diagonal. Hence, the greater the disagreement in the number of ordinal classes, the greater the significance of the error.

The "GT index" (i.e., the ground truth index) developed by Turk [212] is an index of classification accuracy similar to the $\hat{\kappa}$ statistic. Like $\hat{\kappa}$, the GT index is the ratio of beyond-chance agreement to expected agreement. For a given true class j in the classification error matrix, the GT index, g_j , is defined as

$$g_j = \frac{c_{jj}/t_j - \mu_j}{1 - \mu_j} \quad (2.19)$$

The value of μ_j represents random agreement and is derived by an iterative transformation of the classification error matrix. An experiment conducted by Turk shows that, relative to g_j , the producer's accuracy for a given class j gives an inflated estimate of classification accuracy.

Other transformations of the classification error matrix may also be used in classification accuracy assessment. Card [39] demonstrates how knowledge of the mapped areas of each class can be used to improve estimates of classification accuracy. The classification error matrix is transformed to produce a matrix Θ in which

$$\theta_{ij} = a_i c_{ij} / m_i \quad (2.20)$$

where: a_i = the relative area of class i (i.e., the area assigned to class i on the map, expressed as a proportion of total map area).

Next, estimates of the relative area of each true class j , π_j , are calculated as

$$\pi_j = \sum_{i=1}^k \theta_{ij} \quad (2.21)$$

Estimates of the proportion of correct classifications for each true class j , ϕ_j , are computed as

$$\phi_j = \theta_{jj} / \pi_j \quad (2.22)$$

Estimates of the proportion of correct classifications for each assigned class i , λ_i , are computed from the untransformed classification error matrix as

$$\lambda_i = c_{ii} / m_i \quad (2.23)$$

An overall estimate of the proportion of correct classifications, η , is computed as

$$\eta = \sum_{i=1}^k \theta_{ii} \quad (2.24)$$

Approximate confidence limits for η can be computed as

$$\eta \pm z_{\alpha/2} [V(\eta)]^{1/2} \quad (2.25)$$

where: $z_{\alpha/2}$ = the quantile from the standard normal distribution corresponding to a confidence level of $1-\alpha$; and

$V(\eta)$ = the variance of η .

Similar confidence limits can be computed for θ_{ij} , π_j , ϕ_j and λ_i . The asymptotic variances of these estimators under simple random and stratified random sampling are presented by the author.

Accuracy assessments based on the classification error matrix may be adversely affected by the presence of zeros in the matrix (Maxim & Harrington [157]). Use of a sampling zero in place of the correct non-zero population probability can affect statistical testing. The presence of zeros may also produce biased estimates of the conditional and unconditional probabilities derived from the classification error matrix. The conditional probability ζ_{ij} is defined as

$$\zeta_{ij} = c_{ij} / t_j \quad (2.26)$$

The unconditional probability ξ_{ij} is defined as

$$\xi_{ij} = c_{ij} / n \quad (2.27)$$

To circumvent the problems posed by sampling zeros, smoothed estimates of the conditional and unconditional probabilities can be calculated using a pseudo-Bayesian approach. First, the prior unconditional probability is computed for each element of the classification error matrix. The prior unconditional probability ψ_{ij} may be estimated in terms of the number of classes as

$$\psi_{ij} = 1 / k^2 \quad (2.28)$$

In this case, the estimate of ψ_{ij} is the same for all elements of the matrix. However, a unique estimate for each element may be obtained if some level of prior knowledge exists. The estimates given by equation (2.28) are consistent with the notion of a high level of entropy in the prior unconditional probabilities relative to the unconditional probabilities ξ_{ij} . The entropy of the prior unconditional probabilities, E_ψ , is defined as

$$E_\psi = - \sum_{i=1}^k \sum_{j=1}^k \psi_{ij} \ln(\psi_{ij}) \quad (2.29)$$

The entropy of the unconditional probabilities, E_ξ , is defined in an analogous manner.

A weighting factor is next defined as the pseudo-sample size associated with the prior unconditional probabilities. Some *ad hoc* estimates for this weighting factor, w , include $k^2/2$ and $1/n^2$. A smoothed classification error matrix, C^* , is then constructed such that

$$c_{ij}^* = \frac{n}{n + w} (c_{ij} + w\psi_{ij}) \quad (2.30)$$

must equal 1. A value greater than 1 indicates that the two classes exhibit beyond-random locational agreement. Hence the value of r is directly proportional to the level of classification accuracy. This approach accounts for agreement attributable to chance alone in a manner analogous to the $\hat{\kappa}$ statistic.

Two maps of the same area can also be compared using “equivalent isopleths,” or isopleths defined such that the area between successive isopleths is identical (Court [58]). (As described in § 4.5, isopleths are non-intersecting lines joining points of equal value for some phenomena distributed over space.) By superimposing the two maps containing the same number of equivalent isopleths, it is possible to construct a “resemblance matrix” showing the area of agreement and disagreement between the maps. The “coefficient of quantile correlation” can be computed from this matrix as an index of agreement between the two maps.

2.1.4. Other Issues Pertaining to Classification Accuracy

Some of the implications of spatial autocorrelation for classification accuracy have been examined by Congalton [52]. In the context of remotely sensed data, spatial autocorrelation in classification errors can be estimated from a “difference image”, computed by superimposing a classified image and an image in which each pixel has been assigned to its true class. Pixels that agree on the two images are assigned a value of 0 on the difference image, while pixels that disagree are assigned a value of 1. The autocorrelation on the difference image, ρ , is computed as the number of 0-1 joins. That is,

$$\rho = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \delta_{ij} (z_i - z_j)^2 \quad (2.39)$$

where: $\delta_{ij} = 1$ if pixels i and j are neighbors and 0 otherwise;

$z_i =$ the value (1 or 0) of pixel i on the difference image;

$z_j =$ the value (1 or 0) of pixel j on the difference image; and

$n =$ the total number of pixels.

In an empirical test, Congalton computed the spatial autocorrelation for three images of varying complexity, using spatial lags of between 1 and 30 pixels. (The lag is determined by the definition of δ_{ij} .) Positive spatial autocorrelation was observed for all three images, but a decrease in positive spatial autocorrelation was observed as the lag increased.

This approach can be used to assess the effects of different sampling schemes on estimates of classification accuracy (Congalton [53]). For each of the three difference images described above, the proportion of misclassified pixels (i.e., the number of pixels assigned a value of 1, divided by the total number of pixels in the image) and the variance of this proportion were calculated. These served as the population parameters of interest. Sample estimates of these parameters were obtained by sampling each difference image according to five different schemes (simple random, stratified random, cluster, systematic and stratified systematic unaligned sampling). Simple and stratified random sampling provided consistently accurate sample estimates of the population parameters. For less complex difference images, systematic and stratified systematic unaligned sampling greatly over-estimated the population parameters. Cluster sampling performed adequately as long as clusters were relatively small. These results indicate that different sampling schemes vary in terms of the bias and precision of sample estimates of PCC as a function of the level of spatial autocorrelation in misclassifications.

Sampling methodology is also of importance in devising standards for classification accuracy assessment procedures. Mead & Szajgin [159] argue that procedures and reporting formats should be standardized to permit comparison across studies. Sampled points should be well-distributed over space and between assigned classes. The appropriateness of a given sampling scheme depends on sample size, the number of assigned classes, the confidence level desired and the relative significance of different types of misclassification. Loelkes [146] provides some guidelines for assessing classification accuracy by ground survey. In assessing classification accuracy, it must be recognized that many apparent

discrepancies between ground and mapped data are attributable to the effects of cartographic generalization (see § 4.2).

Curran & Williamson [66] also focus on the accuracy of ground survey data. Empirical results indicate that the errors associated with remote sensing (i.e., sensor calibration, signal digitizing error and variability associated with the sensor and the scene) may be outweighed by the errors in ground survey data. Errors in ground data are associated primarily with the spatial variability of the scene, as well as sample processing error and variations in ground personnel performance. In order to ensure that accuracy assessments are meaningful, the ground survey should be designed to attain high levels of accuracy for a relatively small number of ground observations, rather than attempting to obtain a large number of observations.

Havens *et al* [112] describe two methods for assessing classification accuracy for situations in which ground survey data are unavailable. The first method estimates the probability of error analytically, using an *a posteriori* density function. The second method, called the *majority-rule* method, is based on estimation of the dominant class in an area.

Star [199] identifies a number of other sources of classification error in a remote sensing context. Classification procedures in remote sensing are often performed without regard to the underlying frequency distribution of pixel values. The assumption of a normal distribution is typically not verified empirically and it is unclear how deviations from normality affect classification accuracy. Simulation studies are needed that examine the effects of deviations from normality, random noise in the input data, training field selection criteria and alternate classification algorithms.

Estes [78] maintains that for some remote sensing products, especially those that have been transformed by some GIS operation, the degree of accuracy is unestimable using standard statistical techniques. While modeling the effects of error propagation may be useful, it may be more cost-effective to evaluate the accuracy of the output products directly. The efficiency of any method of accuracy assessment must be balanced against its cost and ease of application.

2.2. Soil Properties

Soil maps exemplify a type of spatial data known as a “categorical coverage” (Chrisman [49]). The delineation of mapping units on soil maps is typically achieved by transforming a set of soil properties, measured for a sample of profiles, into a set of mutually exclusive classes. Hence the boundaries between mapping units are not defined *a priori*, as in the case of enumeration districts for collecting census data, but *a posteriori* as a function of the assignment of soil classes. Goodchild & Dubuc [100] (see § 4.3.2) provide a succinct description of categorical coverages based on a “phase-space” model. If m continuous variables (i.e., soil properties), z_1 through z_m , are distributed in two-dimensional space, then the m -dimensional space defined by these variables is denoted as the phase-space. Categorical coverages are constructed by partitioning the phase-space into k domains, each of which corresponds to a particular class. These domains therefore define the transformation of a set of m continuous variables into a set of k classes. The phase-space model clearly illustrates the interdependence between the taxonomic definitions that prescribe the partitioning of phase-space and the structure of the resulting map.

Like soil maps, maps derived by classifying remotely sensed imagery are also categorical coverages. However, many of the models of error assessment described in § 2.1 are of limited utility in the context of soil mapping. More so than many other phenomena distributed over space, soil properties exhibit relatively high-frequency spatial variations, or large fluctuations in values over relatively short distances. Even at relatively large map scales, it is generally not feasible to delineate mapping units within which soil properties are strictly homogeneous. The degree of internal or within-unit variation depends on the minimum mapping unit size, which is constrained by the limits on cartographic fidelity

Smoothed conditional probabilities, ζ_{ij}^* are computed as

$$\zeta_{ij}^* = c_{ij}^* / \sum_{i=1}^k c_{ij}^* \quad (2.31)$$

Smoothed unconditional probabilities, ξ_{ij}^* , are computed as

$$\xi_{ij}^* = c_{ij}^* / n \quad (2.32)$$

These estimators may be useful when the classification error matrix is sparse.

A further alternative to PCC involves “matrix normalization,” as described by Congalton *et al* [55] and Kenk *et al* [134]. Matrix normalization is an iterative procedure in which the rows and columns of the classification error matrix are balanced until each row and column sums to 1. The PCC for the normalized matrix is computed as in equation (2.4). The purpose of matrix normalization is to eliminate the effects of sample size and account for both errors of omission and commission. Hence the PCCs for any two normalized matrices can be directly compared.

The classification error matrix may also be employed to correct estimates of polygon area derived by counts of pixels on a classified image. This approach generally involves pre-multiplication of the vector of area estimates for each class by the inverse of a transformed classification error matrix (Hay [114]). Bauer *et al* [16] and Hixson [116] describe an alternate method based on the *relative* area of each class. Maxim *et al* [158] describe a second variant that accounts for “detection error,” or error in identifying objects of interest. Chrisman [44] and Prisley & Smith [180] describe another variant in which matrix inversion is not required. These techniques are discussed in greater detail in § 3.2.2.

Analysis of variance (ANOVA) may also be used as an alternative to PCC in evaluating classification accuracy (Fitzpatrick-Lins [81], Rosenfield [190] and Rosenfield & Melley [194]). The analysis is based on an arcsin or logit transformation of producer’s accuracy for each class in the classification error matrix. These transformations are designed to produce a more normal probability distribution. Hypothesis testing is based on the F-ratio calculated for each factor in the ANOVA model, or on multiple-range tests, which permit within-factor comparisons. This approach has been used to compare accuracies across classes and between maps of different scales.

Differences in classification accuracy across classes and between maps has also been assessed with log-linear analysis (Rosenfield [191]). In the simplest case, the logit function is computed for the user’s accuracy associated with each class on each map and an additive model is developed that assumes no interactions between class and map effects. More complex models have been developed that consider multi-factor interactions between classes, classification algorithms and image enhancement techniques in the context of remotely sensed imagery (Congalton *et al* [55]). In such models, the relative importance of different factors and factor interactions is evaluated by systematically searching all possible combinations of factors and interactions. The optimal model combines a simple combination with a good fit to the observed data.

Differences in classification accuracy between maps has also been evaluated with a standard t-test (Rosenfield & Melley [194]). The test is based on the deviations in producer’s accuracy for each class between pairs of maps. These deviations are summed and divided by the number of classes to obtain the mean deviation. When divided by the standard error of the mean, the mean deviation is assumed to follow a t-distribution. A t-test may then be applied to determine whether the maps exhibit evidence of a significant difference in accuracy.

It is also possible to test whether the misclassifications along any row of the classification error matrix are distributed uniformly across all classes (Hay [113]). The cumulative binomial probability distribution (equation (2.14)) is employed in this test. For a given row of the classification error matrix, the values of n , x and p_0 in equation (2.14) are defined as follows:

$$n = \sum_{j=1}^k c_{ij} \quad i \neq j \quad (2.33)$$

$$x = \max(c_{ij}) \quad i \neq j \quad (2.34)$$

$$p_0 = 1 / (k - 1) \quad (2.35)$$

Thus n is the number of misclassifications in row i of the matrix, x is the maximum number of misclassifications for a given class in row i and p_0 is the probability of selecting a given incorrect class in row i . Equation (2.14) is evaluated using these values and the result, $P[r > x]$, is compared to some probability threshold (e.g., 0.05). If $P[r > x]$ is less than this threshold, then one may conclude that the misclassifications along row i are not distributed uniformly across all classes.

2.1.3. Alternatives to the Classification Error Matrix

Other indices of classification accuracy have been developed that are not based on the classification error matrix. For example, classification accuracy may be calculated in terms of the disagreement between the mapped areas of all classes and their actual areas as derived from a survey of higher accuracy (Fitzpatrick-Lins [81]). The total area of disagreement is given by

$$d = \frac{1}{2} \sum_{i=1}^k |a_i - \hat{a}_i| \quad (2.36)$$

where: a_i = the actual area of class i ; and

\hat{a}_i = the area of class i estimated from the map.

(Alternatively, a_i and \hat{a}_i may represent the area of class i on two different maps.) A low value of d indicates that the estimated and actual areas are similar. However, the value of d may approach 0 even if the map is highly inaccurate, since the index depends only on the total area of each class and not the locational agreement between each class.

Locational agreement may be defined for any pair of classes on two different maps (Adejuwon [1]). In the context of classification accuracy, one of these maps might represent the accuracy standard. In a random case, the probability of any two classes occurring at the same location, q_{ij} , is defined as

$$q_{ij} = a_i a_j / A^2 \quad (2.37)$$

where: a_i = the area of class i on map 1;

a_j = the area of class j on map 2; and

A = the total map area.

The observed area of locational agreement between classes i and j , a_{ij} , can be computed planimetrically by superimposing the two maps. If randomness holds, then the ratio

$$r = \frac{a_{ij}}{A q_{ij}} \quad (2.38)$$

as a function of map scale. Maps of larger scale are capable of depicting higher-frequency variations in soil properties, but must balance this increase in map complexity against an inevitable decrease in map interpretability (see Jenks & Caspall [131], § 4.3.1). Internal variation in soil properties is therefore normally accepted as an inevitable feature of soil maps. The adequacy of a given set of mapping units is defined in relative terms, such as the ratio of within-unit variance to the between-unit variance. As internal variation cannot be wholly eliminated, the crucial issue is the manner in which variation is documented, such that the map user is cognizant of the limitations of the map.

As a further consequence of internal variations in soil properties, the boundaries between mapping units on soil maps are inherently indeterminate. As Burrough [34] notes, such boundaries are statistical constructs rather than actual physical features of the environment. Boundaries have no absolute or precise location, but rather reflect the manner in which soil properties vary over space and the degree to which such variations may be resolved as a function of map scale. In certain cases, boundaries may coincide with abrupt changes in soil properties, but these boundaries may prove to be less distinct when examined at a finer spatial scale. In regions of continuous variation, boundaries are often masked by high-frequency variations superimposed upon the regional trend. Boundaries may also be artifactual when soil properties are sampled at a distance greater than the wavelength of the high-frequency component of variation. Hence boundaries between mapping units should be viewed in the sense of “edges,” or locations where the average rate of change in soil properties is at a maximum.

As a result of the presence of high-frequency variations in soil properties, the reliability of soil maps cannot be modeled solely in terms of classification accuracy (see § 2.1). Reliability is only partially defined by the number of correctly and incorrectly classified points within each mapping unit (i.e., mapping unit “purity”). It also depends on the variation in soil properties within mapping units (i.e., mapping unit “homogeneity” or “uniformity”). It is necessary to develop models that account for the existence of variation at different spatial scales. Some models of this type are described below, following a brief review of studies describing the nature of variation in soil properties.

2.2.1. Variations in Soil Properties

Bascomb & Jarvis [14] examined the variations in soil properties for profiles within a set of mapping units depicting a single soil series. Mapping unit purity was calculated as the percentage of profiles meeting all of the criteria of the soil series. Purity was observed to be on the order of 60 percent, a value in agreement with previous research. Low mapping unit purity was observed to result from the rigidity of taxonomic definitions. This suggests that an increase in purity might be achieved by allowing for flexibility in the assignment of classes. Within-unit variation was also observed to differ widely for different soil properties. In a decision-making context, mapping units are intended to delineate areas within which similar management practices may be applied. Hence internal variations in soil properties are important in assessing the confidence with which different management practices may safely be applied.

Beckett & Burrough [20, 21] compared the utility of a number of single-property and general-purpose soil maps produced by free and grid survey at scales ranging from 1:20 000 to 1:70 000. On single-property maps, soil classes were defined along specific ranges of the property. (In the terminology of Goodchild & Dubuc [100], the phase-space is one-dimensional.) Map utility was defined as the capacity of the map to predict soil conditions at any point, which depends on the variability of the property within mapping units. Mapping units on general-purpose soil maps were defined in terms of a set of consistent profile classes, or a group of profiles with a particular set of properties. For general-purpose soil maps, utility can be defined variously as the completeness and stability of the profile classes, the ability of the map to predict the correct mapping unit or profile class at any point, the uniformity of soil properties within each mapping unit or profile class, or the mean or modal values of the soil properties within each mapping unit or profile class. The uniformity of soil properties within all mapping units can be measured by the ratio of the within- or between-unit variance to the total map variance. For individual mapping units, the coefficient of variation gives the within-unit variation in a property as

a proportion of the unit mean. Empirical results indicate that coefficients of variation tend to be higher for mapping units than profile classes, since it is not possible to reduce the variance for mapping units below that of the profile classes from which the mapping units are derived. Reducing within-unit variance is more difficult for general-purpose than single-property maps. Substantial reduction is possible only by producing single-property maps at a relatively large scale.

Webster & Beckett [223] examined soil map quality in the context of variations in soil properties within and between mapping units. The adequacy of a given set of mapping units is defined by the intraclass correlation coefficient, or the ratio of the variance between mapping units to the total map variance. If a given soil property is absolutely uniform within each mapping unit, the coefficient takes a maximum value of 1. The within-unit variance measures the uniformity of soil properties within each mapping unit and is therefore more useful than the intraclass correlation when soil maps are employed for decision-making purposes. The ability of a given set of mapping units to minimize within-unit variance depends on the coefficient of variation for each mapping unit. For soil properties with high coefficients of variation, only a small reduction in within-unit variance is possible and the intraclass correlation coefficient will be relatively low. Although within-unit variation can be reduced by increasing map scale, the minimum size of the mapping unit should reflect the minimum size of the area normally considered in a decision-making context. It is necessary to document the degree and character of within-unit variation, since this variation provides information about the applicability of different management practices in mapping units.

Nichols [174] examined the reliability of a set of computerized soil maps constructed by mapping the dominant soil class within each of a set of grid cells. Reliability was estimated as the percentage agreement in the area of each soil class relative to the original map. An inverse relationship was observed between accuracy and cell size and between accuracy and the number of classes within each grid cell. These relationships arise from sampling effects associated with grid cell size and are discussed in greater detail in § 3.2.1 and § 5.2.1.

Empirical results suggest that it is often difficult to derive meaningful mapping units on the basis of taxonomic definitions alone (Crosson & Protz [65]). Observed differences in soil properties between mapping units indicate that many properties do not exhibit significant differences in modal values between units. Hence soil properties do not always provide an unequivocal set of criteria for defining the boundaries between adjacent mapping units. Spatial variables, such as landscape position, may be needed to delineate more meaningful mapping units.

The effects of taxonomic definitions on the reliability of mapping units have been examined in greater detail by Webster [222]. Uncertainty in class assignments is inherent in any soil taxonomy, due to error in measuring soil properties and vagueness in class definitions. Attempts to eliminate this uncertainty (e.g., the 7th *Approximation*, a US Department of Agriculture soil taxonomy) demand an unattainable level of precision that results in internal inconsistency in class assignments. For single-purpose maps, a modicum of vagueness must be tolerated in class assignments to accommodate uncertainty and the effects of local conditions. Mapping units on general-purpose soil maps are often derived by hierarchical subdivision of profiles to reach the required level of generalization. This approach is problematic because it is based on the assumption that soil properties exhibit a nested structure. That is, at any level in the hierarchy, the variations within groups are assumed to be less than the variations between groups. To the degree that this structure does not exist, profiles with similar properties will be separated into different mapping units. This approach assumes that soil classes may be defined monothetically. A more appropriate model is a polythetic one in which no single property is essential or sufficient and membership depends on the possession of a set of common attributes. Soil taxonomies also need to consider the limitations of pedogenesis as a criterion for defining classes. The identification of genetically significant properties as the basis for soil classification is problematic because significance can only be assessed after classification has been performed.

The relationship between the taxonomic definitions of soil classes and the delineation of mapping units has been explored further by Campbell & Edmonds [38]. The authors compared observed soil variations with the types of variations that can be described with the *Soil Taxonomy* (devised by the US Soil Conservation Service). The *Soil Taxonomy* defines three kinds of pedological units. “Taxonomic units” are specific soil classes (e.g., orders, suborders, series, etc.) with no constraint as to their spatial manifestation. “Genetic units” are separate areas of the earth’s surface subjected to uniform pedogenesis and exist wholly in the spatial domain. “Mapping units” are legend items on a soil map and, like genetic units, exist in the spatial domain. Within mapping units, soil properties exhibit continuous variation associated with natural variations and discontinuous or abrupt changes associated with non-soil formations. Theoretically, a mapping unit is equivalent to a taxonomic or genetic unit, but this equivalence cannot be achieved in practice due to limits on cartographic fidelity. Hence the user is often unaware of the amount, character and pattern of variation on soil maps. These variations may be reduced by increasing map scale, but a limit of map complexity is reached beyond which an additional increase in scale is unjustified. Hence most mapping units do not conform to the most basic notions of uniformity and the critical issue is the manner in which within-unit variations are communicated to the map user. Communication may be achieved through verbal descriptions of mapping unit variability, a method applied in older (pre-1935) US soil surveys and modern surveys in some European countries.

Alternate mapping units can also be defined that account for complex pedological patterns and provide a means of depicting spatial variability. For example, “elementary soil areals” delineate soils belonging to a single class of the lowest rank and are bordered by other areals or non-soil formations. This creates a soil cover that describes the continuous and abrupt variations between neighboring units. Other mapping units, including the “genon” and “pedotop,” delineate areas with similar pedological attributes resulting from the same pedogenesis. That is, they define areas of uniform ecology and structure rather than areas of uniform composition. The communicative power of these mapping units lies in their ability to depict spatial variations without requiring additional cartographic detail. In contrast, the *Soil Taxonomy* is biased towards taxonomic distinctions between soil classes, such that the map user receives little information about variations within mapping units.

2.2.2. Models of Soil Property Variation

Variations in soil properties may be modeled as a set of superimposed processes operating at different spatial scales that give rise to definitive soil patterns (Burrough [32]). Variation is commonly viewed as a composite of functional or systematic variation that can be explained and random variation or noise that is unresolvable. However, this distinction is scale-dependent, since an increase in the scale of analysis often reveals structure in random variation. This scale-dependence can be accounted for using the geostatistical concept of a regionalized variable. A regionalized variable is a function that takes a value at any point in two-dimensional space and incorporates both a general structural component and a local random component of variation. The observed values of a regionalized variable are viewed as a realization of a spatial random function with a certain probability distribution. An estimate of the semivariance of the regionalized variable, $\hat{\gamma}(\mathbf{h})$, is given by

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2n} \sum_{i=1}^n [z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})]^2 \quad (2.40)$$

where: $z(\mathbf{x}_i)$ = the observed value at point i ;

\mathbf{h} = the spatial lag; and

n = the number of points separated by the spatial lag.

The plot of $\hat{\gamma}(\mathbf{h})$ against \mathbf{h} is referred to as the semivariogram. Semivariograms of soil tend to exhibit a non-zero value of $\hat{\gamma}(\mathbf{h})$ as the lag approaches 0. This value is referred to as the “nugget variance” and corresponds to random variation. This random component is the result of high-frequency variation associated with processes acting at spatial scales finer than that resolvable by the minimum lag defined in the semivariogram.

The spatial lag defined in equation (2.40) closely resembles the notion of “step-size” used in computing the fractional Hausdorff-Besicovitch dimension, D (see equation (3.12)). The value of D can be estimated from a double-log plot of the semivariogram using regression techniques. For a one-dimensional transect, the value of D is estimated from the slope of the regression line, b , by the equation

$$D = (4 - b) / 2 \quad (2.41)$$

In contrast to other geographical and geological phenomena, soil properties tend to have a relatively high value of D . This is attributable to high-frequency variations in soil properties, since the value of D tends to decline as lower-frequency processes begin to dominate.

Variations in soil properties can be approximated as a stochastic fractal process in which the value of D is an index of the relative balance between high- and low-frequency variations. This stochastic process may be represented by a random walk or Brownian motion function. When $D = 1.5$, this function has the property that for any series of equally spaced lags, the function $[z(x_i) - z(x_i + h)]$ is normally distributed with a mean of 0 and a variance proportional to h . In accordance with the notion of self-similarity (see § 3.1.2), a change in the lag results in a rescaled function that has a probability distribution identical to the original. That is, the Brownian functions at all scales are statistically equivalent. A family of self-similar scaled functions can be defined by using different values of D as the rescaling parameter.

Application of this model to soil property data indicates that these properties do not behave as *ideal* Brownian functions. Although finer scales of analysis clearly reveal finer levels of detail, soil properties exhibit self-similarity over only limited ranges of the spatial lag. The semivariance does not always increase monotonically with the spatial lag, but rises in a series of steps associated with abrupt changes as finer scales of variation are encountered. Moreover the distribution of $[z(x_i) - z(x_i + h)]$ for soil properties is different from that of a Brownian function, in that a greater proportion of values occur at the lower end of the range of the function. The Brownian function is approximated only for relatively large spatial lags, where the structure of variation is no longer resolvable and is perceived as random noise.

An alternate, non-Brownian nested model of superimposed processes operating at different spatial scales is described by Burrough [33, 35]. The model is an extension of the notion that the value of a property at a point is a combination of systematic and random components of variation. The value at point x , $z(x)$, is defined as

$$z(x) = \sum_{i=1}^m f_i(x) + \epsilon \quad (2.42)$$

where: $f_i(x)$ = an independent spatial random function;

m = the number of functions; and

ϵ = the random noise component.

The scales and weights associated with each function can be derived from the semivariogram. Empirical results suggest that, for one-dimensional transects, variations in soil properties are consistent with this model, particularly when abrupt changes occur.

Several models of mapping unit boundaries are described by Campbell [37]. These models define the form of variation in soil properties at the interface between two mapping units. The random model assumes that no coherent structure exists in soil property variation. For a given property, the value at any point is defined as the mean value of the property plus a random component. The gradual transition model defines variations in soil properties as a n^{th} -order polynomial trend surface in the x and y dimensions. The abrupt change model assumes that a sharp break in soil properties occurs at mapping unit boundaries. For a given property, the value at any point in a particular mapping unit is defined as

the mean value of the property in the mapping unit plus a random component. Empirical results indicate that pH and silt content are most adequately represented by a 3rd- and 4th-order trend surface, respectively. For sand content, the abrupt boundary model is more appropriate. These results indicate that different models of spatial variation may be required for different properties and, in accordance with Burrough's findings, suggest that variations in soil properties occur at different spatial scales.

Fisher [80] presents a technique analogous to the dasymetric mapping method to account for the effects of variations in soil properties on mapping unit homogeneity. In dasymetric mapping, a variable of interest is available for a set of areal units, but is assumed to be distributed non-uniformly within these units. Knowledge of "limiting variables," or variables that affect the spatial distribution of the variable of interest, are used to derive a map that more accurately depicts the distribution of the variable. (Flowerdew [84], § 5.1, describes an analogous application of the dasymetric method in the context of areal interpolation.) In the case of soil maps, knowledge of the relationships between soil classes and other environmental variables (e.g., slope) may be used to assess the accuracy of a given set of mapping units and suggest how mapping units might more reliably be delineated. Such relationships can be elicited through knowledge engineering to define a set of rules for mapping the distribution of soil classes. These rules will often be specific to a particular set of classes, although more general rules might also be developed. Reliability will be enhanced by increasing the number of limiting variables examined, particularly if these variables represent independent sources of evidence (see Tikunov [209]).

2.3. Topographic Maps

This section reviews topographic map accuracy standards developed by different mapping agencies, as well as various techniques that have been proposed as the basis for developing alternate standards. In the United States, the definitive accuracy standard for topographic maps is the National Map Accuracy Standard (NMAS), adopted by the Bureau of the Budget in 1947 and currently applied to the US Geological Survey (USGS) topographic map series. This standard is based on compliance with a horizontal and a vertical accuracy standard which define the limit of acceptable error in the horizontal and vertical map dimensions. Compliance testing is based on a comparison of at least 20 well-defined map points relative to a survey of higher accuracy. Horizontal error is defined in terms of the horizontal distance between each map and survey point (i.e., the discrepancy between the x- and y-coordinate values). Vertical error is defined in terms of the discrepancy between the z-coordinates of the map and survey points. The horizontal accuracy standard states that at most 10 percent of the map points may have a horizontal error greater than 1/30 in. for map scales greater than 1:20 000, or 1/50 in. for scales of 1:20 000 or less. The vertical accuracy standard states that at most 10 percent of the map points may have a vertical error greater than one-half of the contour interval of the map. The discrepancies in the vertical dimension may be reduced by shifting the location of points by an amount equal to the allowable horizontal error.

Since NMAS is simply a statement of compliance with an accuracy standard, it provides no accuracy information and ignores the magnitude of horizontal and vertical error at points. A number of alternatives to NMAS have been proposed that address these shortcomings by defining statistical expressions of topographic map accuracy. Defenders of NMAS (e.g., Blakney [29]) argue that such alternatives add to the cost of accuracy assessment without substantially enhancing the reliability of the assessment. Other authors maintain that quantitative expressions of accuracy are preferable, but caution that no single statistical method of assessing accuracy exists that is universally applicable (Thompson [207, 208]). All statistical methods rely on a set of assumptions about the reliability of the data obtained for accuracy testing and the statistical distribution of error. Quantitative expressions of map accuracy also ignore factual or qualitative map errors.

The American Society of Civil Engineers [3] has proposed the Engineering Map Accuracy Standard (EMAS) as an alternative to NMAS for large-scale maps. EMAS is considered to be more flexible than NMAS, in that it is applicable to maps prepared for a variety of applications. The level of acceptable error is permitted to vary as a function of this application. EMAS is a statistical expression of map accuracy based on errors in the x-, y- and z-coordinates of at least 20 well-defined and well-distributed sample points. The error in the x-coordinate of sample point i, δ_x , is defined as

$$\delta_x = \hat{x}_i - x_i \quad (2.43)$$

where: \hat{x}_i = the x-coordinate at sample point i on the map; and

x_i = the x-coordinate at sample point i as determined from a survey of higher accuracy.

The mean error in x, $\bar{\delta}_x$, is defined as

$$\bar{\delta}_x = \frac{1}{n} \sum_{i=1}^n \delta_x \quad (2.44)$$

where: n = the number of sample points.

The standard error in the x dimension, s_x , is defined as

$$s_x = \left[\frac{1}{n-1} \sum_{i=1}^n (\delta_x - \bar{\delta}_x)^2 \right]^{1/2} \quad (2.45)$$

Analogous calculations are also performed in the y and z dimensions. All variables are expressed in terms of ground distance. The standard error as defined in equation (2.45) is similar, but not identical, to the root mean squared error (RMSE) as defined in equation (2.52). The standard error will be equal to the RMSE only when the mean error (equation (2.44)) is equal to 0. Thus in contrast to the RMSE, systematic bias associated with the mean error has been removed in the calculation of the standard error.

Compliance testing for EMAS is performed by comparing the computed mean errors and standard errors in the x, y and z dimensions to their respective maximum acceptable limits (referred to as the "limiting" errors). A test for bias in the x dimension is performed as

$$t_x = \frac{(|\bar{\delta}_x| - \bar{\delta}_{x_0}) n^{1/2}}{s_x} \quad (2.46)$$

where: $\bar{\delta}_{x_0}$ = the limiting mean error in x.

The degree of bias is considered to be statistically insignificant if

$$|t_x| < t_{n-1, \alpha} \quad (2.47)$$

where: $t_{n-1, \alpha}$ = the percentile from the t-distribution with n-1 degrees of freedom and corresponding to a 1- α , one-tailed confidence limit.

A test for precision in the x dimension is performed as

$$\chi_x^2 = \frac{s_x^2 (n-1)}{s_{x_0}^2} \quad (2.48)$$

where: s_{x_0} = the limiting standard error in the x dimension.

The degree of imprecision is assumed to be statistically insignificant if

$$\chi_x^2 < \chi_{n-1, \alpha}^2 \quad (2.49)$$

where: $\chi_{n-1, \alpha}^2$ = the percentile from the χ^2 distribution with $n-1$ degrees of freedom and corresponding to a $1-\alpha$, one-tailed confidence limit.

An analogous set of tests are performed in the y and z dimensions.

Specific values for δ_{x_0} and s_{x_0} are not provided in the EMAS statement. As EMAS is intended to facilitate accuracy testing for a variety of special-purpose maps, these values are assumed to be application-specific. However, the limiting horizontal and vertical standard errors can be computed from the horizontal and vertical map accuracy standards of NMAS. The limiting vertical standard error is defined by

$$s_{x_0} = 0.608 \text{ VMAS} \quad (2.50)$$

where: VMAS = the vertical map accuracy standard (i.e., one-half the contour interval).

The limiting horizontal standard error is defined by

$$s_{x_0} = s_{y_0} = 0.466 \text{ CMAS} \quad (2.51)$$

where: CMAS = the circular (or horizontal) accuracy standard expressed at full scale (i.e., the appropriate fraction, 1/30 or 1/50, weighted by the denominator of the scale representative fraction).

The constants in equations (2.50) and (2.51) are derived from the 90th percentile of a univariate and bivariate normal distribution, respectively (Rosenfield [189]).

These relationships are exploited to define three classes of map accuracy by the American Society of Photogrammetry [4], whose proposed accuracy standard for large-scale maps is similar to EMAS. † Merchant [160] gives a more detailed discussion of the American Society of Photogrammetry standard. The limiting vertical standard error for class 1 maps is defined in accordance with the NMAS vertical accuracy standard (i.e., one-half the contour interval). The limiting horizontal standard error is more stringent than the NMAS horizontal accuracy standard, and corresponds to 1/47 in. expressed at full scale. The limiting standard errors for class 2 and 3 maps are defined by multiplying the class 1 limiting standard errors by a factor equal to the accuracy class. Hypothesis testing for precision (equations (2.48) and (2.49)) is performed for a given accuracy class using the limiting horizontal standard error in place of s_{x_0} and s_{y_0} , and the limiting vertical error in place of s_{z_0} . As in the case of NMAS, points may be shifted by an amount equal to the limiting horizontal standard error for the corresponding accuracy class when assessing vertical error. Points exhibiting errors in excess of three times the limiting standard error in the x , y or z dimensions for the corresponding accuracy class are interpreted as gross errors and are subject to correction. Hypothesis testing for bias (equations (2.46) and (2.47)) is performed using a value of 0 for the limiting mean errors, δ_{x_0} , δ_{y_0} and δ_{z_0} .

Merchant [161] presents a revised version of the American Society of Photogrammetry standard, which is referred to as the American Society of Photogrammetry and Remote Sensing spatial accuracy specification for large scale topographic maps. The revised standard expresses accuracy in terms of a limiting RMSE, rather than a limiting standard error. Hence the test for bias (equations (2.46) and (2.47)) is discarded and systematic bias associated with the mean error is not removed in computing map accuracy. The RMSE in the x dimension is defined as

$$\text{RMSE}_x = \left[\frac{1}{n} \sum_{i=1}^n \delta_{x_i}^2 \right]^{1/2} \quad (2.52)$$

† As noted in the American Society of Civil Engineers [3], EMAS was conceived by a joint committee of the American Society of Civil Engineers, the American Congress on Surveying and Mapping and the American Society of Photogrammetry.

This is identical to equation (2.45) when the mean error, $\bar{\delta}_x$, is equal to 0. RMSEs in the y and z dimensions are defined in analogous manner. Hypothesis testing for precision is performed by comparing the calculated RMSEs to the limiting horizontal and vertical RMSEs. The limiting horizontal RMSE is computed in the same fashion as the limiting standard errors in the American Society of Photogrammetry standard. However, the limiting vertical RMSE is somewhat more stringent, and is defined as one-third of the contour interval of the map. Spot heights depicted on the map have a limiting RMSE of one-sixth of the contour interval. Merchant also stipulates the requirements for the selection of sample points against which the coordinate values of the mapped points are compared. The sampling method is intended to ensure a dispersed distribution of sample points. However, Kellie & Bryan [133] have demonstrated empirically that sample points selected randomly or along profiles provide equally as efficient estimates of accuracy. Random methods may be preferred due to their simplicity and relatively low cost.

Another alternative to NMAS is provided by Koppe's formula, which accounts for the effects of terrain slope on mean vertical error. Based on empirical observations, Koppe observed a high correlation between mean vertical error, $\bar{\delta}_v$, and terrain slope, α , according to an equation of the form

$$\bar{\delta}_v = \pm (A + B \tan \alpha) \quad (2.53)$$

Coefficients A and B are empirically-derived constants for a particular map, where A represents the vertical error when the terrain slope is zero and B is related to the horizontal error at a given point. The equation shows that a given degree of horizontal error will produce a greater degree of vertical error as the terrain slope rises. Hence any increase in the terrain slope is associated with an increase in vertical error (Imhof [126]).

Gustafson & Loon [109] show that the NMAS horizontal and vertical map accuracy standards may be combined into a single expression using Koppe's formula. Coefficient A is defined as the constant vertical error and is equivalent to s_{z_0} as given in equation (2.50). Coefficient B is the allowable horizontal shift in a point when assessing vertical accuracy, expressed at full scale. It is related to s_{x_0} or s_{y_0} in equation (2.51), but assumes a univariate rather than a bivariate distribution. Regression analysis can also be used to estimate coefficients A and B based on the observed vertical error and terrain slope associated with a sample of points. Regression results for a set of USGS 1:24 000-scale topographic maps show that coefficients A and B are lower than the NMAS requirements, indicating compliance with the accuracy standard. However, coefficient B is higher for US maps than European maps of comparable scale, as a consequence of more stringent horizontal accuracy standards in the latter case.

Koppe's formula affords a number of advantages over NMAS as a statement of map accuracy. In contrast to Koppe's formula, NMAS is simply a statement of compliance with an accuracy test rather than a statistical expression of accuracy. Hence NMAS does not facilitate comparison of map accuracy with other countries, provides no accuracy information and ignores the magnitude of error at sample points. More importantly, NMAS ignores the effects of terrain slope on vertical accuracy. Indeed, the effect of the allowable horizontal shift in NMAS is to produce a vertical accuracy standard that is less stringent as terrain slope increases (Gustafson [108]).

The mean horizontal error, $\bar{\delta}_h$, can also be computed by transposing the coefficients of Koppe's formula and reversing the slope function. That is,

$$\bar{\delta}_h = \pm (B + A \cot \alpha) \quad (2.54)$$

Hence, once the coefficients in Koppe's formula have been estimated for a particular map, the horizontal errors in points or contour lines can easily be determined. This equation indicates that horizontal error is inversely related to terrain slope.

Equation (2.54) may be used to derive bands of horizontal error around digitally-encoded contour lines (Yoeli [229]). Coefficients A and B are estimated empirically from a sample of points drawn from the map. The value of δ_h is calculated for each digitized point on the contour lines based on the terrain slope at that point. A new point is then located on either side of each digitized point as a function of the value of δ_h and the direction of the terrain slope at the point. Finally, the error bands are created by joining these new points by straight line segments. The resulting error bands each represent a zone of uncertainty within which the true contour line is located. The technique provides a convenient tool for representing and evaluating topographic map accuracy. It is analogous to the epsilon band concept (see § 4.1.1), in which the value of epsilon has been replaced by the mean horizontal error.

This error band technique is attributed to Imhof [126], who has shown that an increase in terrain slope often results in overlapping of adjacent error bands. Hence the positions of contours on steep slopes may be anywhere within a zone of uncertainty whose width is greater than the horizontal distance between contours. While the error bands tend to widen as terrain slope decreases, the tendency for the bands to overlap is much diminished.

A variety of other techniques have also been proposed for assessing horizontal and vertical accuracy for topographic maps. Unfortunately, it is impossible to provide a detailed review of these methods as many are described only in general terms.

Rosenfield [189] presents a method for evaluating horizontal map accuracy for topographic maps. First, the latitude-longitude graticule is transformed to any desired plane-rectangular coordinate system. An affine transformation is then applied to fit the mapped graticule to the same coordinate system. The discrepancies between the locations of the actual and mapped graticule points on the coordinate system define the horizontal error in the graticule. The coefficients from the affine transformation are then applied to fit a sample of mapped test points to the coordinate system. The discrepancies between the assumed positions of these points and their positions on the coordinate system define the horizontal error in the points. The total horizontal error in the map is calculated as the square root of the sum of the squared horizontal errors in the graticule and the points.

Cook [57] presents two general methods for evaluating horizontal map accuracy for topographic maps. The first is based on an explicit mathematical expression of the errors introduced at each phase of map compilation and the propagation of these errors through to the finished map product. The second is based on controlled sampling in which a random control net containing points of known value and accuracy is carried through each phase of map compilation. Deformations of the net are computed after each phase is complete and a statistical appraisal of accuracy is performed to evaluate the accuracy of the finished map product.

Lee [142] presents two methods for assessing the horizontal accuracy of contour lines on topographic maps. Both are based on the superimposition of the mapped contour with a control contour of higher accuracy. The first method is based on the calculation of the area of the polygon or polygons encompassed by the two contours. Horizontal error is defined as the area of the polygon divided by the length of the control contour. The second method is based on the assumption that the end points of the two contours coincide. The deviations are then measured between points generated along the lengths of the contours. This method is analogous to that employed by Honeycutt [118] to evaluate digitizing error (see § 4.1.1).

Robinson [184] argues that error analysis for topographic maps may be more easily carried out in the frequency domain and therefore advocates the application of Fourier analysis to map profiles. Errors arise from imprecision in measured elevation values on the map and the measurement of these values at discrete sample locations. The first source of error is assumed to be distributed normally such that it is possible to compute the amplitude spectrum for any sample by taking the transform of a series of random normal numbers whose standard deviation is equal to the presumed standard error of the sample.

Measurement error therefore contributes a statistically constant value to the absolute amplitude spectrum over the entire profile. Reliance on discrete sample point locations, the second source of error, results in aliasing error in the absolute amplitude spectrum. That is, frequencies higher than that detectable by the sample may be aliased as additions to the amplitudes of lower frequencies. This error can only be reduced by additional sampling. Aliasing error is shown to be additive such that the absolute amplitude is increased as the distance between sample points rises, particularly at low frequencies. The sum of the constant measurement error and aliasing error defines the total error contained in the profile.

2.4. Planimetric Accuracy

This section reviews standards and evaluation methods for planimetric accuracy, or positional accuracy for planimetric and cadastral maps. Planimetric accuracy is closely allied to accuracy in surveying, which is not addressed in great depth in this report due to the volume of literature on this issue. To some extent, planimetric accuracy also parallels the issue of horizontal accuracy for topographic maps, as discussed in § 2.3. A number of mapping agencies have developed topographic map accuracy standards that define acceptable limits of horizontal error. Various standards are discussed by Merchant [160, 161], the American Society of Photogrammetry [4] and The American Society of Civil Engineers [3]. These standards include the National Map Accuracy Standard, the Engineering Map Accuracy Standard and the American Society of Photogrammetry and Remote Sensing spatial accuracy specification. Many of the issues addressed in § 2.5, including lineage and logical consistency, are also pertinent in evaluating planimetric accuracy.

In the literature on planimetric accuracy it is generally recognized that, if land information are to be utilized effectively, they must comply with a rigorous set of accuracy standards. In the case of maps and other graphical products, these standards should encompass not only positional accuracy, but map scale, format, data type and symbolization (Bennett [25] and Loelkes [146]). Much of the literature on cadastral systems focuses on the trade-off between the desirability of rigorous standards and the cost of acquiring data that are compatible with these standards. Burtch & Thapa [36], for example, argue that accuracy standards for multipurpose cadastral systems must be designed to meet the needs of the most demanding normal application of the system, while accounting for the fiscal capabilities of the agency responsible for system development. Applications such as planning, routing and facility mapping have much less stringent accuracy requirements than engineering applications and property boundary delineation. Due to the high cost of acquiring accurate data, the accuracy requirement of all potential applications of a multipurpose cadastral system must be realistically evaluated prior to the development of the system.

Donahue [73] maintains that the trade-off between accuracy and cost should be phrased in terms of the short- and long-term compatibility of the data with a set of accuracy standards. Long-term compatibility requires accurate geodetic control, which carries a large initial cost burden in system development. Short-term compatibility can be achieved without geodetic control by upgrading compatibility on a continual basis, but may eventually prove to be more costly.

The varying accuracy requirements of different applications of cadastral systems can be modeled in the framework of a hierarchical tessellation. Locational precision is an implicit attribute of any level of the hierarchy, such that features can be encoded and retrieved at any level of required precision. This framework might improve the efficiency of error tracking methods and formalize the occurrence of error in accordance with liability concerns (Smyth [198]). Dutton [77] explores the properties of a global hierarchical tessellation scheme based on recursive subdivision of spherical triangles.

Dahlberg [67] examines accuracy requirements in the context of cadastral systems designed to support legal property rights. The author asserts that the high positional accuracies demanded by such systems cannot realistically be achieved using data derived from existing maps and legal documents. This assertion has been verified empirically by Vonderohe [219], who observed high levels of positional error in existing cadastral maps and legal descriptions of property boundaries.

Dahlberg [67, 68] also examines the merging of parcel data with natural resource and socio-economic data to facilitate spatial analysis. Merging is problematic because of the discrepancy in the level of taxonomic and spatial resolution between these two types of data. Merging might be facilitated by aggregating parcel data, but a more promising solution lies in the acquisition of natural resource and socio-economic data that comply with a general set of accuracy standards. The application of accuracy standards to natural resource data is examined by Hsu [123]. The author argues that natural resource maps derived from remotely sensed satellite data can easily achieve the horizontal accuracy requirements established in the National Map Accuracy Standard (see § 2.3). Digital and photographic techniques may be applied to minimize positional error in polygon boundaries associated with geometric distortions and imprecision resulting from pixel size effects. Socio-economic data are often collected for a set of enumeration units whose boundaries may exhibit some degree of positional error (e.g., LaMacchia [138]). Guptill [107] suggests that indices of positional accuracy may be assigned to these boundaries or any other spatial object in a feature-based data model.

The boundaries between polygons on natural resource maps are often not defined independently of the thematic attribute itself. Rather, the boundaries are derived implicitly by assigning each point on the map to a class or category. On such "categorical coverages" (see § 2.2), errors in boundary location are attributable in part to classification error. Hord & Brooner [120] describe a method for assessing error that accounts for classification error, boundary line error and control point location error. Classification error is assessed with the normal approximation to the binomial distribution, as described in § 2.1.1. Boundary line error, or positional error in boundaries between polygons, is assessed by superimposing two classified images and calculating an index of agreement between the polygon boundary representations. (A polygon is defined as a set of adjacent pixels of the same class.) Control point location error refers to the geometric discrepancies between the map and a universal frame of reference, such as the latitude-longitude graticule. The test for control point location error is based on the normal approximation to the binomial distribution, by measuring control point error as a dichotomous variable (i.e., correct vs. incorrect). The authors maintain that map accuracy may be summarized by computing the mean of the lower 95 percent confidence limit of the three error measures. This approach is also reviewed by Dozier & Strahler [75].

Planimetric accuracy is also closely associated with distortions introduced by map projections. As described in many standard cartography texts, map projections introduce varying degrees of distortion in angles, areas and other characteristics. "Conformality" and "equivalence" refer respectively to the retention of correct angular and areal relationships by the map projection. In general a projection cannot simultaneously be both conformal and equivalent. Angular and areal distortions may give rise to planimetric errors that are particularly problematic when maps based on different projections are merged by some spatial operation. For example, in map overlay (see § 5.3), the use of different projections is frequently cited as a cause of positional discrepancies between features on different data layers. The degree of angular and areal distortion associated with a given map projection can be evaluated using Tissot's indicatrix. The indicatrix defines each point on the globe as a circle of unit radius. Angular distortion causes these circles to become ellipses on the map, while areal distortion results in changes in circle area. Angular and areal distortion can theoretically be measured at any point on a map using this approach. Laskowski [139, 140] presents an efficient method of computing Tissot's indicatrix based on the singular value decomposition of the appropriately scaled Jacobian matrix of the transformation equations. This approach requires linearization of the projection equations, but provides a simple one-step method of computing the distortion parameters. Various coordinate transformation methods may also be applied to reduce projection-related planimetric discrepancies between maps. Simple affine transformations or more complex "rubber-sheeting" functions are

frequently applied in remote sensing to facilitate the merging of remotely sensed imagery with ancillary spatial data (Pearson [177]).

2.5. Geographic Base File Standards

This section discusses some general issues of importance in developing standards and evaluating accuracy for multi-purpose, digitally-encoded geographic base files (GBFs). The discussion derives mainly from the data quality components identified by Chrisman [46] and in the Draft Proposed Standard for Digital Cartographic Data [72]. † This standard is designed to facilitate the documentation of data quality by the GBF producer and the transference of this documentation to the user along with the GBF itself. The standards are couched in terms of “truth in labeling” rather than simple expressions of accuracy or tests of compliance with accuracy standards (see § 2.3). As data quality requirements are application-specific (see § 2.4), it is the responsibility of the producer to document data quality but the responsibility of the data user to interpret this documentation and evaluate the fitness of the data for a particular application. Thus uncertainty is absorbed by the user rather than the producer of the data (see Bedard [23], § 4.6).

According to the proposed standard, documentation of data quality is achieved in the form of a quality report. If quality is known to vary over the area encompassed by the GBF, the quality report may take the form of a quality overlay, which is registered to the GBF (see Chrisman [46]). The quality report is composed of sections documenting five components of quality — data lineage, positional accuracy, attribute accuracy, logical consistency and completeness.

Data lineage refers to the data sources, methods of deriving and encoding the data and all transformations applied to the data. It includes the dates of all source and ancillary data, geodetic control information and the specific characteristics of the transformations applied at each step of database development. Documentation of transformations includes a listing of control point locations and a description of the transformation algorithm, including the nature of the computational steps and the numerical values of coefficients for a sample of computations.

Positional accuracy refers to the accuracy of feature locations after transformations have been applied. Possible tests of positional accuracy include deductive estimates, internal evidence checks, comparison to the source document and reference to an independent source of higher accuracy. Deductive estimates are based on knowledge and assumptions about the propagation of errors introduced at each step in database development. Internal evidence checks are based on standard methods of evaluating surveying accuracy, including repeated measurements and the computation of residuals from adjustments applied to the coordinate system. Comparison to the source document constitutes a visual inspection of the discrepancies between the encoded database and the document from which it was derived. Reference to an independent source of higher accuracy is the preferred method of testing positional accuracy and is based on the standards proposed by the American Society of Photogrammetry and Remote Sensing (Merchant [161]). Numerical test results are presented in ground units and the locations of test points are provided in the data quality documentation.

Attribute accuracy refers to the accuracy of the thematic attribute portrayed in the database. Methods of testing attribute accuracy include deductive estimates, tests based on independent samples of points and tests based on polygon overlay. Deductive estimates are qualitative evaluations based on prior experience and assumptions about the propagation of error through the various steps in database development. Tests based on independent samples rely on the construction of a classification error matrix (see § 2.1) to assess classification error in a sample of points relative to a source of higher

† This proposed standard was developed by a joint committee of the National Committee for Digital Cartographic Data Standards, the Federal Interagency Coordinating Committee on Digital Cartography and the Digital Cartographic Data Standards Task Force (Moellering [162]).

accuracy. The location of sample points is given in the data quality documentation. Tests based on polygon or map overlay (see § 5.3) use a classification error matrix showing the areas of agreement and disagreement in classification between the database and a source of higher accuracy.

Logical consistency refers to the fidelity of the relationships encoded in the data structure. Consistency checks include tests of the validity of assigned values to identify gross errors and tests of topological consistency (see Chrisman [47], § 4.1.3).

Completeness refers to the selection criteria, definitions and other relevant rules used in developing the database. Completeness describes the relationship between objects in the database and the abstract universe of all objects. Tests of completeness include checks of how consistently features have been assigned to classes and how exhaustive the classes are in an actual context.

Grady [104] provides a detailed description of a method of tracing data lineage based on techniques applied in auditing transactions in a management information system. In a management information system, documentation of lineage is achieved through an “audit trail,” whereby transactions that modify a database are recorded in a “transaction history file.” This file describes the data processing steps applied to the database, such that the effects of a given transaction can be traced through to the final product. In order to maintain the referential integrity of the data, a “data dictionary” is used that specifies the database contents and constraints that apply to each feature in the database. The data dictionary applies rules to transactions before new data are encoded or existing data are transformed, thus preventing conflicts between the original and transformed databases. This approach is thorough but may significantly increase the time and cost requirements for database development. Alternatively, “deferred checking” may be performed on the transformed database to check for compliance with rules defined by the data dictionary. Deferred checking resolves inconsistencies in the transformed database and therefore does not impede database development.

Grady also argues that the mandate of the data producer is an important component of data quality and should be incorporated in the lineage documentation. The mandate defines the reason for collecting data and therefore influences the nature of the data. Following Chrisman [46], the author also argues that as many apparent errors in a database are attributable to changes occurring over time, documentation of data quality must account for temporal changes in objects encoded in the database. This might be achieved by assigning temporal codes to all objects as the basis for performing tests of historical validity. The authors also discuss the importance of geodetic control in data quality. Geodetic control is the foundation for achieving positional accuracy, but the data structure requires the flexibility to accept changes in the coordinate values assigned to points resulting from temporal changes or improvements in the accuracy of geodetic control. In accordance with the suggestion made by White [227] (see § 6.3), this flexibility might be achieved by separating the locations of points from their code or identifier. Hudson [124] maintains that the management of lineage documentation might prove to be cumbersome and costly, and suggests instead that the computer system summarize data quality at each processing step and propagate these summary measures through subsequent steps.

2.6. Digital Elevation Models

Digital elevation models, or DEMs, are digital representations of topographic elevations for a part of the earth’s surface. The most common form of DEM produced by the US Geological Survey (USGS), the main producer of DEMs in the US, consists of a gridded array of elevation values conforming to the latitude-longitude graticule or the Universal Transverse Mercator (UTM) grid system. The former type of DEM is referred to as the “arc-second” format and is produced by the USGS at 3-arc-second and 1-arc-second grids. The 3-arc-second grid conforms to the USGS 1:250 000-scale topographic map series, while the 1-arc-second grid conforms to the standard 7 1/2-minute quadrangles used in the USGS 1:24 000 topographic map series. The arc-second format DEMs are non-rectangular due to the convergence of meridians in a poleward direction. In contrast, “planar” format DEMs conforming to the

UTM grid system consist of a rectangular array of elevation values.

Caruso [42] discusses the standards employed by the USGS to evaluate the accuracy of 7 1/2-minute arc-second DEMs. The USGS classifies errors in DEMs into three categories. "Blunders" are gross errors that are easily detected and are therefore usually edited from the DEM prior to general release. "Systematic errors" are non-random errors associated with specific procedures that introduce biases and artifacts into the DEM. An example is "striping," an artificially high level of spatial autocorrelation in elevation values along one axis of the DEM. Striping is associated with the practice of resampling elevation values along one axis at a higher spatial resolution (i.e., a smaller sampling interval) than that used to originally acquire the elevation data. Although such systematic errors are often easily detected, they are not always correctable. "Random errors," the third error category, result from measurement error and, unlike systematic errors, reduce precision but do not introduce bias.

Accuracy testing of DEMs by the USGS consists of comparing the known elevations of at least twenty control points to the elevations of these points as interpolated from the DEM. The RMSE is then calculated in the z dimension (see equation (2.52)). Based in part on the RMSE calculation, the USGS identifies three levels of DEM quality. Level I DEMs contain no points whose elevations are in error by more than 50 m. The maximum RMSE permitted is 15 m. DEMs acquired by manual profiling or the Gestalt Photo Mapper II (GPM2) typically fall within this level. Level II DEMs have a maximum RMSE of 7 m and contain no points whose elevations are in error by more than twice the contour interval of the source map. DEMs acquired by contour digitizing typically fall within this level. Level III DEMs have a maximum RMSE of 7 m and contain no points whose elevations are in error by more than the contour interval. Digital Line Graph (DLG) DEMs, which incorporate hypsographic and hydrographic data, fall within this level.

Carter [41] presents an alternate taxonomy of error for DEMs that distinguishes between "relative" and "global" errors. Relative error refers to a situation in which a number of elevation values are obviously inconsistent relative to neighboring elevation values which, as a group, give an adequate representation of the surface. Global error, in contrast, refers to a situation in which the DEM as a whole gives an adequate representation of the surface, but the total model departs significantly from the source document or the actual surface. Global errors are particularly problematic when concatenating adjacent DEMs, since overlaps can occur and the differences in elevation values across the "seam" may be unrealistically large. Global errors can often be corrected by applying a translation, scaling or non-linear transformation. The author provides numerous examples of relative errors in DEMs and demonstrates how these are often difficult to detect and correct.

Theobald [205] views DEM accuracy as a function of two distinct issues. First, accuracy is limited by the quality of the source document and the methods used to acquire elevation values from them. For example, DEMs acquired with the GPM2 often contain errors due to the inability to discern the actual ground surface, particularly in areas of dense tree cover. Semi-automatic profiling methods, in contrast, may produce striping on DEMs when the sampling interval is different in the x and y dimensions. The second issue affecting DEM accuracy pertains to the size of the sampling interval and its interaction with terrain surface variability. Sampling implies a certain level of generalization of the source document in order to create a DEM of manageable size while preserving a reasonable surface representation. The appropriateness of a given sampling interval depends on the variability of the terrain surface, but it is also dictated by the requirements of different applications of DEMs. Other important issues of DEM accuracy include the spatial distribution of systematic error and the propagation of error into products derived from DEMs, including measures of slope, aspect and curvature.

Bethel & Mikhail [26] present a method for detecting gross errors, or blunders, in DEMs based on mathematical modeling of the terrain surface by tensor-product B-splines. The authors view this procedure as the first stage of an on-line quality assessment system for DEMs. The procedure is based on fitting the tensor-product of two one-dimensional B-splines locally over the DEM. Residuals are then computed and a statistical test is performed to yield an overall assessment of the presence of outliers

(i.e., gross errors) in the DEM. This constitutes the detection phase of quality assessment. Specific outliers are then identified in a candidate subgroup of the residuals and flagged as gross errors. Tests performed on a set of synthetic and actual DEMs reveal that this approach is especially effective in the case of multiple blunders of relatively large magnitude.

Hannah [111] also presents a method for detecting and correcting gross errors in DEMs. The method is based on the calculation of slope and change-in-slope values for each grid point in the DEM. Error detection is performed in three phases. First, for each grid point, change-in-slope values across the point are computed in four directions corresponding to eight local neighbors and in eight directions corresponding to sixteen distant neighbors. Each of the resulting twelve change-in-slope values is thresholded by assigning it a value indicating whether or not it exceeds the maximum allowable change-in-slope. The mean of these threshold values is then calculated for each point as an index of change-in-slope reliability. The change-in-slope reliability indices for each point are further transformed via a "voting" procedure that assumes that a given point can be no more reliable than the least reliable of its local and distant neighbors. The voting procedure is applied iteratively until reliability indices converge.

The second phase of error detection involves calculation of the slope between each point and its eight local neighbors. Each of these slope values is thresholded by assigning it a value indicating whether or not it exceeds the maximum allowable slope. The mean of these thresholded values is then calculated for each point as an index of slope reliability. In the final phase of error detection, a composite reliability index is computed for each point as the minimum of its slope and change-in-slope reliability indices.

Error correction is based on change-in-slope analysis. A point is considered to be compatible with its neighbors when its elevation causes minimal changes in the slope surrounding and across the point. The surface is smoothed by assigning new elevation values that minimize changes in slope, using a weighting procedure in which the weights reflect the composite reliability index of each point. Experimental tests show that this method succeeds in identifying and correcting gross errors. However, problems may be encountered along steep ridge lines and an extreme level of surface smoothing sometimes occurs.

3. ACCURACY OF CARTOMETRIC ESTIMATES

This section focuses on the accuracy of cartometric estimates, or estimates of lengths and areas derived from maps. The effects of cartographic generalization on cartometric line length estimates are examined in § 3.1. In § 3.2, the accuracy of cartometric area estimates are discussed in the context of classification error and the method employed in area estimation.

3.1. Length Estimation

Estimates of the length of linear map features, including roads, political boundaries and coastlines, are inherently dependent on the level of cartographic generalization present in the map. As the level of generalization rises, lines become increasingly smooth and minor deviations from the general trend of the line are eliminated. Hence estimates of line length tend to decrease as the level of generalization rises. As cartographic generalization is a function of the limits on cartographic fidelity associated with map scale, the level of generalization is inversely related to scale. Estimates of line length therefore tend to decrease as map scale declines. Analogously, length estimates are dependent on the spatial resolution of the instrument used for estimating line length. For example, if length is estimated by stepping a pair of dividers along the line, then the resolution is defined by the divider "step-size," or the distance between the two divider points. For a given cartographic line, the estimate of line length decreases as step-size increases, since the line is in effect being defined at an increasing level of generalization. The discussion that follows reviews the empirical evidence supporting these assertions and then elucidates the relevance of fractal theory in exploring the relationship between length estimation and cartographic generalization.

3.1.1. Sources of Error

Baugh & Boreham [17] examined the effects of three sources of error on estimates of coastline length. The first source of error is the definition of the coastline itself, and in particular how far up-river the coast is assumed to extend. The second source arises from the digital representation of coastlines as sets of digitized points joined by straight line segments. This error source includes aspects of digitizing and generalization error (see § 4.1 and § 4.2). The third source of error is associated with map scale and the resulting level of cartographic generalization. The effects of these error sources were examined by comparing estimates of the length of different segments of the Scottish coastline. Lengths were estimated from maps of different scales using an opisometer (i.e., a calibrated wheel) and by summing the straight line distances between successive digitized points. Length estimates were observed to increase with map scale for both opisometer and digitizer methods, due to a decrease in the level of cartographic generalization. For maps of the same scale, the digitizer method yielded larger estimates, due mainly to instrumental error in opisometer estimates at coastline segment termini. The digitizer method was assumed to be the more accurate of the two methods. Comparison of different coastline segments revealed that discrepancies in length estimates associated with both map scale and estimation method were greatest for highly indented coastlines.

With reference to estimates of the length of the Australian coastline, Galloway & Bahr [88] examined the effects of map scale, method of estimating length, coastline smoothness and definition of the coastline. Coastline definitions differed in terms of the inclusion or exclusion of small inlets and islands. Coastline smoothness was examined by comparing the length estimates for coastline segments that exhibited varying degrees of indentation. Coastline lengths were estimated with divider, opisometer and flexible wire methods. Divider step-sizes varied between 0.1 and 1000 km ground distance. The effects of map scale were explored by estimating coastline lengths from maps with scales between 1:250 000 and 1:15 000 000.

The exclusion of small inlets and islands in the coastline definition was observed to reduce the estimate of coastline length. For the 1:250 000-scale map, this reduction was on the order of 12 percent. Coastline length estimates were also observed to decline as divider step-size increased and as map scale decreased, due to an effective rise in the level of cartographic generalization. Both the divider and flexible wire methods yielded accurate and consistent length estimates at a given scale, while the opisometer method showed considerable variability. The effects of step-size and map scale were observed to be more significant for highly indented coastlines. The authors calculated the relationship between coastline length estimates and divider step-size for the 1:250 000 scale map for step-sizes ranging from 10 to 1500 km as

$$L = 25\,000 s^{-0.12} \quad (3.1)$$

where: L = the coastline length estimate (in km); and

s = the divider step-size (in km).

Maling [154] provides a review of previous research on the effects of map scale and divider step-size on line length estimates and presents some equations that can be used to correct for these effects. The correction for step-size is based on the theoretical line length estimate when step-size is equal to 0. This theoretical line length should equal the “true” line length, since the divider points are infinitely close. Given L_1 and L_2 , two estimates of line length derived with step-sizes of s_1 and s_2 , respectively, then the true line length, L^* , can be estimated as

$$L^* = L_1 + k(L_1 - L_2) \quad (3.2)$$

An estimate of parameter k is given by

$$k = \frac{s_1^b}{s_2^b - s_1^b} \quad (3.3)$$

Parameter b is a measure of line sinuosity. It takes a value of 2 for a smooth line and a value of 0.5 for a sinuous line. Experiments show that this correction works well, although it may be difficult to empirically establish a reasonable value for parameter b .

The correction factor given in equation (3.2) is inconsistent the Steinhaus paradox, which states that as a line is measured with increasing accuracy (i.e., as step-size decreases to 0), the line length estimate increases to infinity. Maling’s argument, however, is that the only solution to this paradox is to define some order of measurement that represents the lower limit of the size of meaningful geographical features. This lower limit may be defined in terms of the epsilon band concept (see § 4.1.1). The epsilon band for a given cartographic line may be delineated by joining the center points of a circle of radius ϵ as it is “rolled” along both sides of the line. The value of ϵ determines the minimum size of cartographic features that may be resolved. Line length may be estimated by dividing the area of the band by its width (i.e., 2ϵ) and correcting for the area associated with the rounded ends of the band extending beyond the line segment termini. Experiments show, however, that this method tends to underestimate line length by about 2 percent.

The author also presents an equation that may be used to correct for variations in the level of cartographic generalization associated with map scale. Given a line length estimate of L derived from a map of scale 1: R , the corrected line length estimate, L^* , is given by

$$L^* = L + \alpha R^{1/2} \quad (3.4)$$

The parameter α depends on line sinuosity. Maling suggests that a mean value of α can be estimated empirically using a collection of maps of different scales.

Beckett [19] provides an alternate form of equation (3.5),

$$L / L^* = \beta R^\gamma \quad (3.5)$$

where: L^* = the actual length of the line;

L = the length of the line as estimated from a map; and

R = the scale (i.e., representative fraction denominator) of the map.

For a perfectly straight line, coefficient β should be equal to 1 and coefficient γ should be equal to 0. In this case the estimated length is identical to the actual length. If two length estimates are derived from maps of different scales then

$$\frac{L_1}{L_2} = \frac{\beta_1}{\beta_2} \left(\frac{R_1}{R_2} \right)^\gamma \quad (3.6)$$

where: L_i = the length of the line as estimated from map i ; and

R_i = the scale (i.e., representative fraction denominator) of map i .

Estimates of β_1/β_2 and γ may be derived from the regression of $\ln(L_1/L_2)$ on $\ln(R_1/R_2)$. The author performed regressions on five sets of data, where each set was composed of opisometer length estimates for several roads derived from British road maps of different scales. Map scales ranged from 1:2 500 to 1:2 000 000, although the range of scales within any one set of data was considerably smaller. For all five sets of data the ratio β_1/β_2 was approximately equal to 1, indicating that it is independent of map scale. The value of γ varied between 0.031 and -0.019 and the author suggests a mean value of -0.017 .

Regressions were also performed on two sets of coastline length estimates derived from maps with scales of 1:25 000 to 1:6 000 000 using dividers with step-sizes ranging from 2 to 15 mm. The ratio β_1/β_2 was again observed to be approximately equal to 1 and invariant over different step-sizes. Coefficient γ showed no consistent trend over different step-sizes, although it was considerably larger (in absolute value) than for the road map data. The value of γ varied from -0.051 to -0.031 and the author suggests a mean value of -0.04 .

Hakanson [110] examined the effects of map scale and line sinuosity on estimates of shoreline length for a set of Swedish lakes. Lakes varied in size from approximately 1 to 6000 km². Shoreline lengths were estimated from maps with scales ranging from 1:10 000 to 1:1 000 000, using an estimation method based on counting the number of intersections of the shoreline with the axes of a regular grid. A measure of "shoreline development," F , is defined as

$$F = \frac{L}{2 \pi^{1/2} A^{1/2}} \quad (3.7)$$

where: L = shoreline length; and

A = lake area.

The equation yields an index of lake shape representing the ratio of the length of the shoreline of a lake of area A to the circumference of a circle of the same area. Lake area is assumed to be independent of map scale since positive and negative errors in the cartographic representation of the shoreline will be cancelled out. Empirical curve-fitting showed that shoreline development at a scale of 1:1 (i.e., "normalized" shoreline development) was related to shoreline development at a given map scale of interest by the equation

$$F^* = \frac{F(K_2 - K_1)}{K_2 - \log(s + a)} \quad (3.8)$$

where: F^* = normalized shoreline development;

F = shoreline development at a given map scale, s ;

$K_2 = \log(s+a)$ for $s = 6\,000\,000$ (where $6\,000\,000$ is the “scale constant”);

$K_1 = \log(s+a)$ for $s = 1$ (where 1 is the “reference scale”);

$a = 10^5 \log(A)$ (where 10^5 is the “area constant”); and

A = lake area in km^2 .

Alternatively, for shoreline length,

$$L^* = \frac{L(K_2 - K_1)}{K_2 - \log(s + a)} \quad (3.9)$$

where: L^* = normalized shoreline length; and

L = shoreline length at a given map scale, s .

These equations were found to yield accurate results when A was held constant as F^* varied and when F^* was held constant as A varied.

3.1.2. Fractal Theory

The relationship between length estimation and cartographic generalization may be examined in light of fractal theory (Mandelbrot [155]). Consider a straight line segment of length L , for which the dimensionality, D , is equal to 1. For every positive integer n , the segment can be exactly decomposed into n nonoverlapping segments whose end points will be $(x-1)L/n$ and xL/n , where x ranges from 1 to n . Thus the length of each segment, L' , is equal to L/n . Each of the n segments can further be decomposed into n nonoverlapping segments of length L'' , where $L'' = L'/n$. This process may be repeated and at each step the length of the n new segments is equal to $1/n$ of the length of the next-largest segment. At each iteration, the sum of the lengths of all segments will equal the length of the original, undecomposed line.

Analogously, consider a D -dimensional geometric figure ($1 < D < 2$) of length L consisting of n straight line segments of length s , where $s = L/n$. Each of these segments may be replaced by a geometric figure identical to the original and consisting of n straight line segments whose length, s' , is equal to $1/n^{1/D}$ of s , or $s/n^{1/D}$. The geometric figure is said to be “self-similar” since each segment is a reduced-scale image of the whole. The total length of each of the n new figures, L' , is equal to ns' and their combined length is equal to $n(ns')$, or n^2s' . In contrast to the one-dimensional case discussed above, note that $L' > L$. That is, at each iteration, the sum of the lengths of each segment is greater than the sum of the lengths of each segment at the previous iteration and, in the limit, the length is infinite.

Since the length of a segment at a given iteration is equal to $1/n^{1/D}$ of the length at the previous iteration, the fractional dimensionality, D , is defined by

$$D = \log(n) / \log(1/s) \quad (3.10)$$

Self-similarity may not hold for cartographic lines, but a statistical form is often encountered. That is, each portion of a cartographic line can be considered to approximate a reduced-scale image of the whole line. For a given cartographic line for which the length is to be estimated, variable s represents the step-size, or spatial resolution, of the length estimation method employed, while n represents the number of segments or “steps” that will be counted. A decrease in the value of s represents a decrease in the step-size and hence an increase in the amount of geographical detail that can be resolved. For lines for which $D = 1$, a given decrease in s will result in a proportional increase in n , such that the line length estimate, L , will remain the same (since $L = sn$). For cartographic lines, in contrast, a given decrease in s will result in a proportionally larger increase in n , such that D will tend to some limit as s approaches 0.

Goodchild [96] argues that for geographical lines, the limit to D is 2, which represents a line that completely fills two-dimensional space. If the length of a given line is measured using two different step-sizes, s_1 and s_2 , and the line is found to contain n_1 and n_2 such segments, respectively, then the fractional dimensionality of the line, D , can be estimated as

$$D = \log(n_2 / n_1) / \log(s_1 / s_2) \quad (3.11)$$

This is known as the “Hausdorff-Besicovitch dimension.” The relationship between step-size, s , and the length estimate, L (where $L = sn$), is given by

$$\log(L) = a + (1 - D) \log(s) \quad (3.12)$$

Coefficients a and $1 - D$ may be interpreted as the intercept and slope of the double-log regression of L on s , respectively. When $D > 1$, the slope will be negative. If the line under consideration has a constant value of D , indicating self-similarity, the slope of the regression line will be constant over the range of s . The greater the value of D , the greater the absolute magnitude of the slope, such that the reduction in the line length estimate with increasing step-size is greatest for lines with relatively high fractional dimensionality. Examination of coastline length estimates, however, indicates that the slope of the regression line tends to increase in absolute magnitude as s rises. This observation suggests that coastlines (and presumably other cartographic lines) are not strictly self-similar, since the fractional dimensionality changes over the range of s .

3.2. Area Estimation

Cartometric area estimation refers to the estimation of the area encompassed by a closed geometric figure, such as a polygon. Unlike errors in line length estimates, which are seen to arise primarily from the effects of cartographic generalization, errors in area estimates have typically been examined in the context of dot planimetry. This technique, common in forestry and photogrammetry, seeks to estimate polygon area by overlaying a transparent grid of dots on a map or aerial photograph and counting the number of dots falling within the polygon. Polygon area can then be estimated as na , where a is the ground area represented by a single dot and n is the number of dots counted.

There are strong parallels between this technique and the GIS operation of vector to raster conversion (see § 5.2). In this operation, a set of polygons possessing values for some thematic attribute are converted into a set of grid cells possessing the thematic attribute values (or some transformation thereof) of the polygon in which the cell is located. The area of a given polygon on a raster map can again be estimated as na , where a is the ground area of a single cell and n is the number of cells that have been assigned to the polygon. This area estimate is likely to be inaccurate, however, since cells located near the edges of the polygon are often bisected by the polygon boundary. When such cells are assigned to the polygon, they contribute to overestimation of area, and when they are not assigned, they contribute to underestimation. Research by Crapper [60–63], Crapper *et al* [64], Frolov & Maling [87] and Lloyd [144] shows that, when considering a single polygon, the relative standard error of the polygon area estimate derived from the rasterization process is proportional to the $3/2$ power of the linear dimension of the cells and the $-3/4$ power of polygon area. Other authors, including Goodchild [96, 97], Muller [168] and Switzer [203], have shown how the standard error may be computed when considering groups of polygons.

The discussion in this section reviews two main research issues on the accuracy of area estimates. The first is related to the effect of variations in the density of dots on the transparent overlay used in dot planimetry. This is the analog of the effect of variations in step-size on line length estimates, as discussed in § 3.1. Much of this research is reported in the forestry literature, since dot planimetry is often used to estimate the area of different forest types.

The second research issue discussed in this section pertains to the problem of classification error in a remote sensing context. In remote sensing, classification involves the assignment of a class (e.g., a land cover type) to each pixel in an image, based on the spectral response of the pixel in one or more spectral bands. The area encompassed by a given class can be estimated as the number of pixels in the image that have been assigned to that class. However, the presence of incorrectly classified pixels means that the area estimate for a given class may be biased. Much of the research on this issue has focused on methods of correcting area estimates for classification error, based on manipulation of the classification error matrix (see § 2.1). This issue is also related to other accuracy issues pertaining to “categorical coverages” (Chrisman [49]), a class of geographical data in which the values of the thematic attribute determine the structure of the map (see § 2.2).

3.2.1. Effects of Dot Grid Density

Tryon *et al* [211] examined how polygon area estimates derived by dot planimetry are affected by variations in the density of dots on the transparent dot overlay. A set of five aerial photographs at a scale of 1:31 680 were partitioned into polygons representing forest and non-forest land cover classes. The area of each polygon was estimated using transparent overlays with dot densities of 4, 8, 16 and 25 dots per square inch (dpi). Area estimates were summed over all polygons of each cover class to yield an estimate of the total area of each class. These estimates were observed to be fairly consistent over different dot densities for classes with relatively large areas, but dropped off sharply for small areas at densities of 4 or 8 dpi. Large areas exhibited high absolute errors, while small areas exhibited high relative errors.

Barrett & Philbrook [13] examined the relationship between dot density and the standard error of area estimates using a method of repeated trials. Each trial involved estimation of polygon area based on a random placement of the dot grid over the polygon. The authors performed 40 trials on two polygons of different sizes for a 16 dpi grid, 30 trials for a 25 and a 36 dpi grid and 20 trials for a 64 dpi grid. The mean area estimate of each polygon for a given dot density was computed and the standard error of the area estimate was then calculated as

$$s_i = \left[\frac{1}{k-1} \sum_{j=1}^k (a_{ij} - \bar{a}_i)^2 \right]^{1/2} \quad (3.13)$$

where: s_i = standard error of the area estimate for polygon i ;

a_{ij} = area estimate for polygon i in trial j ;

\bar{a}_i = mean area estimate for polygon i ; and

k = number of trials.

Both the standard error and the relative standard error (i.e., the standard error as a percentage of the mean area) declined as dot density increased. The authors also calculated the standard error according to the more conventional binomial method,

$$s_i = A (p_i q_i / n)^{1/2} \quad (3.14)$$

where: A = total area of the photograph on which the polygon was delineated;

p_i = proportion of A accounted for by the area of polygon i (i.e., the estimated area of polygon i divided by the total area of the photograph);

$q_i = 1 - p_i$; and

n = number of dots over the entire photograph.

In contrast to the repeated trial method, equation (3.14) is based on a single estimate of polygon area. The standard errors computed with equation (3.14) were significantly larger than those computed with equation (3.13). According to the authors, the repeated trial method yields more precise estimates of polygon area than the binomial method.

Gering *et al* [91] examined the degree to which area estimates are biased by the density of the dot grid. A set of aerial photographs at a scale of 1:12 000 were partitioned into polygons representing forest and non-forest land cover classes. The minimum polygon size was approximately 10 acres. The area of each polygon was estimated using three dot grids at 4, 16 and 49 dpi. Polygon area estimates were summed to yield an estimate of the total area of each cover class. In all but one case, the differences in cover class area estimates attributable to dot density variations were not statistically significant, indicating that a decline in dot density does not introduce bias into area estimates.

In a further test, a sample of polygons was selected and their areas repeatedly estimated using each of the three dot grids. In all but one case, differences in mean polygon areas estimate attributable to variations in dot density were not statistically significant. However, as density declined, the standard error of the area estimates was observed to rise. By extrapolating this trend and considering the trade-off between decreasing standard error and increasing cost associated with denser grids, the authors concluded that a density of 27 dpi is optimal. At densities of 27 dpi or greater, the binomial method (equation (3.14)) may appropriately be applied.

Frazier & Shovic [86] examined the accuracy of dot planimetry for situations in which area estimates are based on a sample of grid cells within which dots are randomly located. This procedure is a two-stage sampling design in which the cells represent the primary sample and the dots within each cell represent the secondary sample. Polygons corresponding to seven land cover classes were delineated on aerial photographs at two different dates. A systematic grid of 130 cells with 20 randomly located dots in each cell was used to estimate the area of each cover class and served as an accurate standard against which various sampling strategies were compared. In the authors' experimental design, cells were selected by random, systematic and stratified random sampling. Results indicated that 2.6 km² cells with 8 dots per km² (ground area) or 1.26 km² cells with 15 dots per km² provided relatively accurate area estimates with acceptable standard errors. No significant differences were observed in standard errors for random and stratified random sampling, but systematic sampling showed considerable potential for imprecision due to periodicity in the spatial distribution.

Hixson *et al* [117] investigated the effect of different sampling schemes on the bias and precision of crop area estimates derived from counts of Landsat pixels classified into crop types. This method of area estimation is analogous to dot planimetry except that the dots have been replaced by a grid of rectangular cells (i.e., Landsat pixels). Four sampling schemes were employed in which the number of pixels sampled remained constant but the size of the sampling frame (i.e., a set of one or more contiguous pixels) and the number of frames selected were permitted to vary. For experimental purposes, the area under wheat in the study area was estimated four times for each of the four sampling schemes. Hence the mean and standard error of the area estimates could be calculated for each sampling scheme. In terms of precision, results showed an increase in the standard error and the relative standard error as sampling unit size increased. Systematic bias was not observed in the mean area estimate for the different sampling schemes.

MacKay [152] argues that, whatever the method of area estimation employed, the accuracy of these estimates may be ascertained with reference to a "pattern map." The pattern map is composed of polygons of regular shapes whose area can therefore be determined geometrically with a high degree of accuracy. Dot planimetry estimates of individual polygons, for example, could be compared against their actual area to calculate bias and precision.

3.2.2. Correcting for Classification Error

In remote sensing, area estimates are often based on counts of the number of pixels assigned to different classes. As explained above, the presence of incorrectly classified pixels means that the area estimate for a given category may be biased. Hence, in estimating area by this method it is appropriate to attempt to correct for classification error.

Hay [114] describes a general method for correcting for the effects of classification error on area estimates. Consider an image for which the pixels have been assigned to k classes. The area encompassed by a given class is then proportional to the number of pixels assigned to that class. Let \mathbf{R} be a $k \times 1$ matrix where r_i is the number of pixels assigned to class i on the image. In order to assess classification accuracy, a sample of cells is randomly selected and their classes checked against more accurate information derived from ground survey. Let \mathbf{C} be the $k \times k$ classification error matrix for this sample (see § 2.1), such that c_{ij} is the number of pixels assigned to class i but according to the ground survey, actually belong in class j . Let \mathbf{E} be a $k \times k$ "normalized" classification error matrix, where each element of \mathbf{E} is the corresponding element of \mathbf{C} divided by its column total. That is,

$$e_{ij} = c_{ij} / \sum_{i=1}^k c_{ij} \quad (3.15)$$

Finally, let \mathbf{T} be a diagonal $k \times k$ matrix whose diagonal elements are the column totals of \mathbf{C} . That is,

$$t_{jj} = \sum_{i=1}^k c_{ij} \quad (3.16)$$

Since, by definition, $\mathbf{ET} = \mathbf{C}$, then \mathbf{E} may be calculated as $\mathbf{E} = \mathbf{CT}^{-1}$. The correction for classification error is performed as

$$\mathbf{A} = \mathbf{E}^{-1} \mathbf{R} \quad (3.17)$$

Matrix \mathbf{A} is a $k \times 1$ matrix where a_i is an estimate of the number of pixels in class i on the image after correcting for classification error. The elements of \mathbf{A} may be multiplied by pixel area to yield estimates of the area encompassed by each class. The technique ensures that the sum of the elements of \mathbf{A} will equal the sum of the elements of \mathbf{R} . The sum in both cases will equal the number of pixels in the image. In contrast, the more conventional "ratio" method of correcting for classification error does not preserve this equality, since it treats each row of \mathbf{C} independently. The ratio method simply scales each element of \mathbf{R} by a ratio of the corresponding row and column totals of \mathbf{C} . That is, the corrected estimate of the number of pixels in class i on the image, a_i , is given by

$$a_i = r_i \frac{\sum_{j=1}^k c_{ji}}{\sum_{j=1}^k c_{ij}} \quad (3.18)$$

The author presents a hypothetical example demonstrating that equations (3.17) and (3.18) do not necessarily yield the same result.

Maxim *et al* [158] present a variant of equation (3.17) that incorporates "detection" error, or error in identifying objects of interest. Let \mathbf{D} be a diagonal $k \times k$ matrix of detection errors. Specifically, $d_{ii} = 1/q_i$, where q_i is the probability of detection for class i . Area estimates may be corrected as

$$\mathbf{A} = \mathbf{DE}^{-1} \mathbf{R} \quad (3.19)$$

This generalizes to equation (3.17) when detection probabilities are equal to 1 for all classes. Other methods are also described that may be applied when knowledge of matrices \mathbf{D} or \mathbf{E} is incomplete or partial information about matrix \mathbf{A} is available.

Bauer *et al* [16] describe a variant of equation (3.17) in which the elements of \mathbf{A} and \mathbf{R} represent the proportions of pixels in each class rather than the total number of pixels. Hixson [116] maintains that the success of this method depends on obtaining unbiased estimates of the elements of \mathbf{C} . For example, stratified random sampling may be used to ensure that all classes in the image are adequately sampled. The approach also appears to be more successful when classification accuracy is relatively high for all classes.

Chrisman [44] and Prisley & Smith [180] present another variant of equation (3.17),

$$\mathbf{P} = \mathbf{R}^t \mathbf{F} \quad (3.20)$$

\mathbf{P} is a $k \times 1$ matrix where p_i is an estimate of the proportion of pixels on the image in class i after correcting for classification error, as in the approach described by Bauer *et al* [16]. As described by Hay [114], \mathbf{R} is a $k \times 1$ matrix where r_i is the number of pixels assigned to class i on the image. \mathbf{F} is similar to matrix \mathbf{E} described by Hay [114] except that normalization is performed over the columns of \mathbf{C} . Each element of \mathbf{F} is the corresponding element of \mathbf{C} divided by its row total. That is,

$$f_{ij} = c_{ij} / \sum_{j=1}^k c_{ij} \quad (3.21)$$

The advantage of this method is that matrix inversion is not required. Inversion may be infeasible due to singularity when sample sizes for specific categories are small (Hay [114]). Note, however, that equations (3.17) and (3.20) do not necessarily yield the same result.

4. ERRORS INTRODUCED DURING DATA COMPILATION

This section is concerned with errors introduced into spatial databases by data compilation methods. Models of digitizing error, or inaccuracies in the encoded positions of digitized points, are reviewed in § 4.1. The discussion in § 4.2 is devoted to generalization error, or error arising from the representation of cartographic lines as sets of digitized points joined by straight line segments. Errors associated with various mapping techniques are examined in § 4.3, § 4.4 and § 4.5. Choroplethic mapping error (see § 4.3) incorporates aspects of data symbolization and the interaction between attribute and positional error. In isometric mapping (see § 4.4) and isoplethic mapping (see § 4.5), errors are introduced by characteristics of the data that affect the accuracy of interpolation. Conceptual and perceptual error in mapping are discussed in § 4.6 in the context of the theory of cartographic communication.

4.1. Digitizing Error

Digitizing error refers to the positional error in a digitized point arising from inaccurate placement of the digitizer cursor during the data encoding process. Most models of digitizing error consider positional error in the context of lines or polygon boundaries defined by sets of points connected by straight line segments. Perhaps the most well-known of these error models is the epsilon band concept, in which positional error is represented by a "boxcar" distribution of probable "true" line locations around the digitized line. Variants of the epsilon band concept are based on alternate definitions of this distribution. Other authors have proposed that in certain cases, digitizing error may be modeled as a serially dependent process in which the error at a given point along a line is dependent on the error at previously digitized points. Other models are more qualitative and account for the physiological and psychological factors that affect the ability of the human operator to perform the digitizing task.

Digitizing error is closely related to generalization error and the two types of error are often considered together, since both result in positional error in cartographic lines. Following Amrhein & Griffith [5], generalization error may be defined as the error arising from the representation of a cartographic line as a set of points joined by straight line segments. In general, the fewer the number of points, the greater the degree of generalization error. The degree of generalization error will generally increase as line complexity rises (Burrough [34]). Generalization error is also associated with the post-digitizing process of line simplification, which is often applied to maintain cartographic fidelity as map scale is reduced. Generalization error is discussed in § 4.2 of this report, but many of the models discussed below are also pertinent to generalization error. Digitizing error also plays a role in error modeling for specific GIS operations. In map overlay, for example, digitizing error results in the generation of spurious polygons (see § 5.3.2). Digitizing error is likewise of some importance in assessing planimetric accuracy (see § 2.4) and in establishing map accuracy standards (see § 2.3 and § 2.5).

4.1.1. The Epsilon Band Concept

The epsilon band concept provides a means of modeling the degree of digitizing error associated with the points defining cartographic lines. This concept is described by Blakemore [27], Chrisman [45], Honeycutt [118] and others. It is based on the notion that a buffer zone having a width twice epsilon (ϵ) can be constructed around any cartographic line. The delineation of this zone may be visualized as the process in which a circle of radius ϵ is "rolled" along both sides of the line (Chrisman [45]). In modeling digitizing error, the value of ϵ (and hence the width of the buffer zone) reflects the degree of error in digitized point locations. The boxcar distribution of width 2ϵ around the digitized line is assumed to encompass the true line location and therefore represents a band of uncertainty associated with the position of the digitized line. As noted in § 4.2, this model may also be applied to account for generalization error, which is also associated with uncertainty in line location.

A number of authors have applied the epsilon band concept in an experimental context to determine the degree of uncertainty introduced by digitizing error and other error sources. Chrisman [43], for example, computed the epsilon band for a digital land use/land cover map in terms of the combined effect of line width drafting error, roundoff error and digitizing error. Line width drafting error was defined as the inaccuracy associated with the representation of a line as a feature of non-zero width. This may represent a relatively large zone of uncertainty in line location, particularly on small-scale maps. Roundoff error was defined in terms of the numerical precision of the hardware used for digitizing. Roundoff error may be propagated through subsequent calculations (see Burrough [34]). Digitizing error was assumed to hold the greatest potential for introducing errors during digital map encoding.

Each of these sources of error was treated independently, since each was assumed to be introduced at a distinct phase in the encoding process. The value of ϵ was therefore estimated as the sum of the errors associated with each source. For the digital land use/land cover map, ϵ was estimated as 15.2 m and increased to 20 m to account for interpretation and registration error. Interpretation error was defined as the uncertainty in the location of boundaries between land use/land cover categories associated with indeterminacy in taxonomic definitions. Imperfect alignment of data sources due to the dimensional instability of paper maps was assumed to be the primary cause of registration error. The epsilon band was defined for the polygon boundaries of each land use/land cover category on the map. The area within the band was observed to represent some 7 percent of the total map area. For individual land use/land cover categories, this zone of uncertainty was as large as 50 percent of the mapped area of the category. Since the value of ϵ was constant over the entire map, this percentage was inversely related to the mapped area of the polygon.

In a similar study, Blakemore [27] defined the epsilon band for a set of 115 polygons (employment office areas) as a function of digitizing error, roundoff error and generalization error. Generalization error was defined as the inaccuracy resulting from the representation of cartographic lines as a set of digitized points joined by straight line segments (see § 4.2). The value of ϵ was estimated as 0.7 km and a zone of this width was defined along the inner and outer edges of each polygon boundary. A set of 780 points (industrial establishments) was then assigned to each polygon based on a point-in-polygon algorithm. Each point was defined as being “definitely in” the polygon (i.e., inside the mapped polygon but outside of the epsilon band), “possibly out” of the polygon (i.e., inside both the mapped polygon and the inner epsilon band), “possibly in” the polygon (i.e., outside of the mapped polygon but inside the outer epsilon band) and “ambiguously defined” (i.e., on a digitized polygon boundary). Only half of the points were found to be uniquely assignable (i.e., “definitely in”). An increase in this percentage was observed for larger values of ϵ . Small and elongated polygons were found to be most affected. The attribute values associated with the points (employment in the industrial establishment) were summed for each polygon for all points in the “definitely in” category. Results showed considerable variation relative to the sums for all points within the mapped polygons (i.e., the “definitely in” and “possibly out” categories). These variations were not highly correlated with polygon size. Even small numbers of points in the “possibly in” category significantly affected the sum when the attribute values for these points were relatively large.

Honeycutt [118] applied a variant of the epsilon band concept in the context of positional error in digitized lines. Positional error was assumed to arise from digitizing error, interpretation error (i.e., uncertainty in line location due to indeterminacy in taxonomic definitions), sampling error (i.e., inaccuracy associated with locating ground control points) and compilation error (i.e., errors associated with line drafting, the dimensional instability of paper maps and map revisions). The degree of error was measured in terms of the discrepancies between the locations of a line digitized from maps of different scales. The map of the largest scale was assumed to be an accurate standard in the absence of information about “true” line location. Thus the approach may also be seen as a model of generalization error (see § 4.2).

In calibrating the model, eight stream segments were digitized from four topographic maps with scales of 1:24 000, 1:62 500, 1:100 000 and 1:250 000. The digitized versions from the three smallest-scale maps (the test or “compare” lines) were superimposed over the 1:24 000-scale version (the control or “base” line). The positional discrepancies were then calculated from the chain of polygons formed by joining the end points of each test and control line. The discrepancies were measured from the horizontal deviations between the test and control lines sampled at 10 m intervals. The mean and standard deviation were calculated for both positive and negative deviations of each test line from its corresponding control line. Results suggest that the probability distribution of positional error associated with each test line is bimodal in form. This result may be attributed to the tendency of human operators to undercut and overshoot during the digitizing process. The bimodal error distribution may be seen as a variant of the epsilon band in which probable line position is defined in terms of the union of the unimodal error distributions on either side of the digitized line. This concept may also be extended to map overlay (see § 5.3.2).

A similar model is described by Maffini [153]. Locational error in cartographic lines was ascribed to digitizing error, generalization error and interpretation error. Trials involving repeated digitizing of geometric figures revealed variations in digitizing error as a function of map scale and the speed of digitizing. The presence of a probability distribution of error around each digitized line suggests that lines may be mapped as a probability distribution rather than a simple linear feature. This model might also be applied in map overlay in terms of the intersection between overlapping probability distributions (see § 5.3.2).

Another variant of the epsilon band concept is described by Peucker [178]. Any cartographic line can be described by the general direction of the line and a corresponding band of a given width and length. The band is a bounding rectangle, the sides and ends of which are parallel and perpendicular respectively to the general direction of the line. The width of the band is defined by the maximum extent of the line perpendicular to its general direction. The band can be constructed at any level of abstraction, such that a line composed of n points can be represented by between 1 and $n-1$ subsets. In the latter case each band encloses a straight line segment between two adjacent points on the line and thus the width of the band is zero. In terms of digitizing error, this model suggests that any line may be represented by its corresponding band at a level of abstraction concomitant with its assumed accuracy. The implications of this model for map overlay is discussed in § 5.3.2.

A final variant of the epsilon band concept is presented by MacDougall [147]. As detailed in § 5.3.1, the total horizontal or positional error for a map is given by the product of the index of average horizontal error (e.g., the standard error) and the length of all lines on the map. This gives an estimate of the total area of uncertainty on the map and is analogous to the area encompassed by the set of epsilon bands for all lines. A more detailed description of MacDougall’s model and its application to map overlay is given in § 5.3.1.

4.1.2. Digitizing Error as a Serially Dependent Process

The notion that digitizing error results from the tendency of human operators to undercut and overshoot during the digitizing process (see Honeycutt [118], above) has led other researchers to argue that in certain cases digitizing error can be modeled as a serially dependent process. Keefer *et al* [132] show that the occurrence of errors in stream mode digitizing can be modeled as an autoregressive moving average (ARMA) process, in which the magnitude and direction of error at a digitized point are influenced by the magnitude and direction of error at previously-digitized points. The analogy to time series analysis exists because in stream mode digitizing, points are recorded at fixed time or distance intervals as the operator follows the mapped line with the cursor. The ARMA process is a combination of a pure autoregressive (AR) process, in which the error at a point is expressed as a weighted sum of p previous errors plus a random component, and a pure moving average (MA) process, in which the error at a point is expressed as an average of q previous random components. The ARMA(p,q) model assumes that the error series is stationary, such that the parameter d in the more

general autoregressive integrated moving average model, ARIMA(p,d,q), is equal to 0. A simpler case is an ARIMA(1,0) model, which accounts only for the influence of error in the preceding point and thus reduces to a simple AR(1) model.

In an empirical test, a set of sample lines were digitized under operational conditions and compared to more accurate representations of these lines digitized under carefully controlled conditions. Digitizing error was computed as the positive or negative discrepancy between each digitized point on the sample line and the point at which a perpendicular intersected the accurate reference line. Correlation coefficients for the AR(1) model ranged from 0.2 to 0.8, with the distribution centered on 0.55. The autoregressive model may also be used to simulate digitizing error. Simulation permits various map accuracy standards to be compared and facilitates evaluation of the effects of line complexity and the time or distance interval used in stream mode digitizing.

The more general ARIMA model can also be applied to model digitizing error for stream mode digitizing (Amrhein & Griffith [5]). In this model a unilateral dependence is assumed to exist in the error structure, such that the error at a given digitized point is dependent on the error at the previously digitized point, but not *vice versa*. Error in digitally encoded lines results from both digitizing and generalization error. Generalization error arises from the representation of cartographic lines as sets of digitized points joined by straight line segments. If the true line is defined by some mathematical function, then generalization error may be measured as the error variance of the digitization-error-free points relative to the functional form. Generalization error tends to zero as the number of digitized points increases to infinity, when the piece-wise approximation matches the functional form. In contrast, digitizing error rises as the number of points increases, and tends towards some upper limit corresponding to the error variance in point location. Hence the "optimal" number of points to digitize can be identified from the point at which digitizing and generalization error intersect, such that the total error is at a minimum for a relatively small number of points.

4.1.3. The Role of Human Factors

Other researchers have examined digitizing error in the context of human physiological and psychological variables that affect the ability of the human operator to perform the digitizing task. The results of such studies often suggest methods for reducing digitizing error, either during the digitizing process itself or at a subsequent editing phase.

While it is generally accepted that human factors contribute to digitizing error, the role of specific physiological and psychological variables has apparently not been studied extensively. Ottawa [176] maintains that personal habits and attitudes are often overlooked during the digitizing phase of database development. In an experiment conducted by the author, fourteen subjects were requested to digitize a set of twenty-four polygons derived from a 1:20 000-scale soils map. The area of each digitized polygon was then computed and compared across subjects. The standard deviation in area was generally within 7 percent of the mean polygon area, although standard deviations as large as 50 percent of the mean were observed for small polygons. More importantly, significant differences in the area of individual polygons were observed across subjects, indicating the importance of human factors in assessing digitizing error.

Jenks [130] argues that digitizing error is a function of physiological and psychological limitations of the human operator. Error in digitally encoded lines also arises from post-digitizing processes such as line simplification, a form of generalization error discussed in § 4.2. Different modes of digitizing give rise to different error structures. Stream mode digitizing, for example, is dominated by "latitudinal" errors, or positional discrepancies perpendicular to the trend of the line. These errors arise from several sources, including psychomotor errors in line-following, physiological errors producing "twitches and jerks" in the encoded line and the psychological tendency to undercut and overshoot once it is realized that an error has been made. In contrast, point mode digitizing, in which the

operator digitizes critical points along the line, is dominated by “longitudinal” errors, or errors parallel to the trend of the line. These errors are associated primarily with logical errors in the identification of critical points.

Digitizing error may be reduced during the digitizing process by enhanced training of human operators, or during a subsequent phase of data editing. Experimental results suggest that digitizing error is correlated with the direction of cursor movement and that operators tend to have a relatively consistent “error signature.” Error can be reduced by providing operators with feedback about the types of errors they have made, or by developing software that automatically corrects for the operator’s error signature. Error reduction during the editing phase can be carried out visually, by comparing the source document to the digitally encoded version, or automatically, by utilizing software that removes topological inconsistencies such as loops and unclosed polygons.

Chrisman [47] details a consistent set of digitizing procedures that may be used to establish quality control during the digitizing process. Digitizing error depends on the nature of the source document, the characteristics of the hardware used in digitizing, the digitizing mode adopted and the level of operator training. The source document should be plotted on a mylar base to prevent errors associated with the dimensional instability of paper maps. Hardware should be able to represent digitized coordinates at the required level of precision. Stream mode digitizing based on a fixed distance interval delegates the burden of selecting points to the computer. In contrast, point mode digitizing is based on subjective criteria for identifying critical points, while stream mode digitizing based on a fixed time interval is affected by the speed of digitizing. Experienced operators are essential to ensure high quality results and operators must understand the logic of the maps they are digitizing in order to ensure consistency. Requiring operators to edit their own work is a direct incentive to careful digitizing.

Error detection may be performed either visually or automatically. Visual error detection is based on a comparison of the source document to the digitally-encoded version. Automated error detection is performed by checking for topological consistency. This may be achieved with algorithms that define a tolerance, or epsilon band, for each digitized line. (Such an algorithm is described by Chrisman [45], Dougenik [74] and White [226], § 5.3.2.) Slivers between alternate representations of the same line can thus be eliminated. However, while an increase in the allowable tolerance reduces such errors, it simultaneously degrades positional accuracy and precision.

4.2. Generalization Error

Generalization error arises from the representation of cartographic lines as sets of digitized points joined by straight line segments. As noted in § 4.1, if the true line is defined by some mathematical function, then generalization error may be measured as the error variance of the digitization-error-free points relative to the functional form. In general, generalization error tends to zero as the number of digitized points increases to infinity, when the piece-wise approximation matches the functional form (Amrhein & Griffith [5]). The degree of generalization error will generally increase as line complexity rises (Burrough [34]).

As the implications of generalization error tend to be application-specific, most models of generalization error are discussed in other sections of this report. Models of generalization error distinguish between inadvertent and intentional generalization. Intentional generalization results from the process of simplifying lines to maintain cartographic fidelity as map scale is reduced. Inadvertent generalization is associated with the errors in cartographic line location that inevitably occur when these lines are defined by a finite number of digitized points.

Inadvertent generalization can be modeled with the epsilon band concept, which is based on the notion that a buffer zone can be constructed around any cartographic line to represent the degree of uncertainty in line location. Uncertainty is a function of digitizing and generalization error, both of which result in discrepancies between the locations of the digitized and “true” lines. Blakemore [27] and Chrisman [43, 45] discuss the epsilon band concept in the context of digitizing and generalization error (see § 4.1.1). Implementation of the epsilon band concept for error detection and correction is described by Chrisman [45, 47], Dougenik [74] and White [226] (see § 4.1.3). Honeycutt [118] and Maffini [153] discuss variants of the epsilon band concept based on an observed distribution of error around the digitized line (see § 4.1.1). Further variants are detailed by Peucker [178] and MacDougall [147] (see § 4.1.1). The former variant is based on the notion that any cartographic line may be represented by an encompassing rectangle whose width reflects the degree of uncertainty in line location. The latter variant defines the total area of uncertainty on a map in terms of the product of an index of average positional error and the total length of all cartographic lines on the map. The extension of these models to map overlay is discussed in § 5.3.2.

The effect of intentional generalization on the accuracy of cartographic lines is discussed below. Its implications for cartometric line length estimation are detailed in § 3.1. Various researchers (Baugh & Boreham [17], Beckett [19], Galloway & Bahr [88], Hakanson [110] and Maling [154]) have demonstrated that length estimates tend to decline with increasing levels of generalization. This effect is tied to map scale, such that it is necessary to account for scale effects in the line length estimation procedure.

The remainder of this section is devoted to studies of the effects of line simplification. Jenks [130] argues that the effects of line simplification are poorly understood due to a paucity of studies on errors in the simplification process. Two types of line simplification methods are identified by the author. Point elimination methods are based on the exclusion of unwanted points to create a more generalized rendition of the line (e.g., Peucker [178]). Smoothing methods are based on the application of a filtering operator along the length of the line. Optimally, both methods should yield a line that is perceptually distinct from the original, but of a form recognizable as a generalization of the original line. Excessive levels of simplification produce lines that are perceptually unrecognizable as a generalization of the original line, while insufficient levels of simplification produce lines that are perceptually indistinguishable from the original.

Little [143] adopts an empirical approach to examine the effects of different line simplification algorithms on line accuracy. Accuracy was defined in terms of the National Map Accuracy Standard (NMAS) and the American Society of Photogrammetry and Remote Sensing (ASPRS) spatial accuracy specification (see § 2.3). NMAS defines an accurate map as one for which a maximum of 10 percent of a sample of points exhibit a positional error of more than a specified amount (defined as a function of map scale). The ASPRS standard is more rigorous and defines accuracy in terms of the root mean squared error in sample point locations. Accuracy was evaluated for a set of lines of varying complexity derived from topographic maps and simplified with five different point elimination algorithms. The algorithms differed in terms of the number of points examined in determining which points to eliminate. They were classified as independent point, local, extended local and global algorithms. One-way analysis of variance was used to test hypotheses related to the performance of different algorithms, the effects of line complexity and the ability to attain each of the two accuracy standards. A significant difference was observed in algorithm performance, with more global algorithms yielding more accurate results. Overall, the effect of line complexity on algorithm performance was observed to be insignificant, although the accuracy for independent point and local algorithms was often adversely affected by complex lines. The more rigorous ASPRS standards were observed to be more difficult to attain than NMAS. These results suggest that global algorithms should be employed in line simplification if an accurate generalization is required.

4.3. Choroplethic Mapping

The objective of choroplethic mapping is to “symbolize” the thematic attribute values associated with a set of areal units or polygons (e.g., census tracts, states, etc.). The thematic attribute may be categorical or numerical. In the latter case, symbolization is performed by defining a set of class intervals and assigning each polygon to the appropriate class interval based on its thematic attribute value. Often the polygons on choropleth maps are delineated solely for enumerative purposes and thus the positions of features on the map (i.e., the polygon boundaries) are defined independently of the thematic attribute. In this case, choroplethic mapping error is primarily a function of attribute error. In other cases, polygon boundaries are defined by the values of the thematic attribute themselves. For example, in remote sensing, maps of land cover classes are often derived by assigning a class to each pixel on the image based on its spectral response in one or more spectral bands. Polygons are defined by groups of contiguous pixels of the same class, and therefore map error is a function of both positional and attribute error. This type of choropleth map exemplifies a class of spatial data known as a “categorical coverage” (Chrisman [49]), which is discussed in greater detail in § 2.2.

The discussion of choropleth map accuracy in this section of the report reflects the relative importance of positional and attribute error. Research focusing primarily on attribute error is concerned with “data symbolization,” or the effect of class interval selection on the accuracy with which the choropleth map portrays the “true” spatial distribution of the thematic attribute. Other research focuses on the interactions between attribute and positional error in the context of categorical coverages.

4.3.1. Data Symbolization

Jenks [127] introduces the “data model” concept as a means of assessing the adequacy of the selected set of class intervals for portraying the spatial distribution of the thematic attribute on the choropleth map. Consider a choropleth map composed of n polygons, where each polygon i possesses a data value, z_i (i.e., a numerical thematic attribute value). The set of n data values represents the data model. In choroplethic mapping, the data values are symbolized by defining a set of k class intervals, or ranges of data values, and assigning each polygon to the appropriate interval based on its data value. The purpose of the symbolization process is to generalize the n data values into k classes, where $k < n$. Hence the choropleth map may be referred to as a “generalized model.”

The error introduced by symbolization is defined in terms of the differences between the data model and the generalized model, and is referred to as the “blanket of error.” The thickness of this blanket and its distribution over the choropleth map depends on the particular set of class intervals selected, since the class intervals define the generalized model. Class intervals may be selected to enhance certain properties of the blanket of error. For example, it might be appropriate to distribute the blanket of error as uniformly as possible over the choropleth map, or to vary the thickness as a function of the data values. To achieve a uniform blanket of error, each polygon is assigned to a class interval and the mean data value for class interval j , m_j , is calculated as

$$m_j = \frac{\sum_{i \in j} z_i}{\sum_{i \in j} 1} \quad (4.1)$$

Next the mean absolute deviation between the polygon data values and the mean data value for class interval j , s_j , is calculated as

$$s_j = \frac{\sum_{i \in j} |z_i - m_j|}{\sum_{i \in j} 1} \quad (4.2)$$

A polygon is shifted from the class interval with the highest mean absolute deviation to one with a lower mean absolute deviation. Equations (4.1) and (4.2) are reapplied and the process continues until the mean absolute deviations for all class intervals are approximately equal. To achieve a blanket of error that varies in thickness as a function of data value, the same process is applied except that the

mean absolute deviation for each class interval (i.e., s_j) is first divided by the mean data value for the class interval (i.e., m_j) to yield the relative mean absolute deviation.

Jenks & Caspall [131] further elaborate upon the data model concept. Three types of error in choropleth maps are identified. "Tabular error," e_t , is defined as the sum of the absolute deviations between the data values of the polygons and the mean data values of their corresponding class intervals. That is,

$$e_t = \sum_{j=1}^k s_j \quad (4.3)$$

where: s_j = the mean absolute deviation between the polygon data values and the mean data value for class interval j (see equation (4.2)); and

k = the number of class intervals.

The mean tabular error is computed as e_t divided by n , the number of polygons. The minimum tabular error will occur when there are as many class intervals as polygons (i.e., $k = n$), while the maximum tabular error will occur when there is only one class interval. The tabular accuracy index, TAI, is computed by dividing e_t by the maximum tabular error and subtracting the resulting quotient from 1.

"Overview error," e_o , is defined as the sum of the tabular error of each polygon weighted by polygon area. That is,

$$e_o = \sum_{j=1}^k \left(\sum_{i \in j} a_i |z_i - m_j| / \sum_{i \in j} 1 \right) \quad (4.4)$$

where: a_i = the area of polygon i .

The overview accuracy index, OAI, is computed by dividing the e_o by the maximum overview error and subtracting the resulting quotient from 1. As in the case of tabular error, the maximum overview error will occur when there is only one class interval.

"Boundary error," e_b , refers to the degree to which the choropleth map accurately portrays sharp boundaries, or "cliffs," between the data values of neighboring polygons. The boundary accuracy index, BAI, is estimated by summing the absolute differences in data values for neighboring polygons on the choropleth map and dividing by the sum in an "ideal" case. The ideal case is one in which the highest cliffs (i.e., the greatest differences in the data values of neighboring polygons) in the original data model are preserved, subject to a constraint on the maximum number of cliffs that may be portrayed on the choropleth map.

Based on the computed values of TAI, OAI and BAI for a given choropleth map, the composite accuracy index, CAI, is calculated as

$$CAI = (OAI^2 + TAI^2 + BAI^2)^{1/2} \quad (4.5)$$

Since $CAI \leq 1.732$ (i.e., $3^{1/2}$), CAI may be divided by 1.732 to yield the map accuracy index, MAI, which ranges from 0 to 1. The authors describe a procedure to obtain near-optimal values of TAI, OAI and BAI. This procedure is an iterative one in which polygons are shifted among different class intervals to obtain successively higher accuracy index values. As the number of class intervals increases, MAI increases at a decreasing rate for choropleth maps produced with the optimization procedure. However, a threshold is reached after which the addition of more class intervals causes a decline in map readability due to an excessive level of map detail.

Monmonier [164] argues that while the optimization of class intervals increases choropleth map accuracy, the resulting class breaks (i.e., the end points of the class intervals) may contain more significant digits than the user would prefer or the accuracy of the data warrant. A solution to this problem is to add constraints to the optimization procedure that ensure that class breaks are round numbers and permit inherently meaningful values (e.g., the overall mean of the polygon data values) to be defined as class breaks. By the principle of "flat laxity," the resulting sub-optimal, constrained solutions will generally increase choropleth map error only marginally over the optimal set of class breaks.

Implementation of the round number constraint requires that polygon data values be converted to their round number representations. Data values are arranged in rank order and one class break is selected per iteration. Potential class breaks are defined in terms of the rounding base used to round off the polygon data values. All potential breaks at any iteration are examined and the break yielding the greatest increase in accuracy (as defined by an index such as TAI) is provisionally selected. Improved solutions are attempted by repeatedly shifting the provisional class break as far as possible without passing a break defined at an earlier iteration. Iteration continues until the desired number of class breaks is obtained. Empirical evidence presented by the author shows a slight decrease in map accuracy relative to unconstrained solutions. However, this decrease is offset by an increase in map readability.

MacEachren [148, 149] examined the effects on choropleth map accuracy of polygon geometry and the complexity of the underlying statistical surface being mapped. Experiments were performed in which 54 polygons of varying size and compactness were overlaid randomly on four different statistical surfaces of varying complexity to simulate choroplethic mapping. Each surface was composed of a regular grid of 112×75 z-coordinate values. Surface complexity was defined by the autocorrelation of the z-values at a lag equal to the mean length of the long axis of the set of 54 polygons. The mean and standard deviation of the z-values falling within each overlaid polygon was computed and the coefficient of variation, or the ratio of the standard deviation to the mean, was used as an index of error. Multiple regression analysis was performed with the coefficient of variation as the dependent variable and polygon size, polygon compactness and surface complexity as the explanatory variables. Bivariate regressions revealed that surface complexity was the most critical variable, followed by polygon size and polygon compactness. An increase in accuracy was found to correspond to a decrease in surface complexity, a decrease in polygon size and an increase in polygon compactness.

A second set of experiments (MacEachren [150]) were performed to simulate a situation in which the underlying statistical surface is unknown. A set of 36 contiguous polygons was selected and the size and compactness of each polygon was measured. The coefficient of variation associated with each polygon was computed for the four statistical surfaces used in the first set of experiments and served as an index of observed error. In order to calculate predicted error, multiple regression analysis was performed with the coefficient of variation as the dependent variable and polygon size and compactness as the explanatory variables, using the data for the 54 polygons employed in the first set of experiments. The resulting regression equation was used to predict the coefficient of variation (i.e., error) for each of the 36 polygons employed in the second set of experiments. The differences between observed and predicted error were found to be statistically significant for only one surface. Correlations between observed and predicted error ranged from 0.64 to 0.88. It may therefore be possible to predict choropleth map error when the underlying statistical surface is unknown.

4.3.2. Interactions between Attribute and Positional Error

For the class of choropleth maps known as categorical coverages, polygon boundaries on the map are defined in terms of the values of the thematic attribute themselves and thus attribute and positional errors are not independent. Goodchild & Dubuc [100] present a model that may be used to simulate error in categorical coverages (which the authors refer to as "natural resource data"). The model is based on a "phase-space" concept. If m continuous variables, z_1 through z_m , are distributed in two-dimensional space, then the m -dimensional space defined by these variables is denoted as the phase-space. To construct a choropleth map, the phase-space is divided into domains, each associated with

one of a set of k classes. When $m = 1$, this process is equivalent to the choropleth map symbolization procedure, as discussed above. In the more general case when $m > 1$, the domains in phase-space define a transformation of the set of m variables to a set of k classes. Since the variables are continuous, if two classes share a boundary on the map, they must also share a boundary in phase-space. The model may be applied to simulate the effects of various sources of error on choropleth map accuracy. Examples include the effects of cell size, generalization and digitizing error, measurement error in variables and classification error associated with the delineation of domains in phase-space.

Goodchild & Wang [102] discuss “source” errors in categorical coverages, or discrepancies between the mapped data values and data acquired by ground survey. In terms of remotely sensed data, source errors result from pixel misclassifications associated with within-pixel variations in class membership, class heterogeneity and uncertainty in class definitions. Information about these errors is lost, however, when remotely sensed data input to GIS consist of a single class membership for each pixel. The authors describe a model that is based on a set of k data layers, where k is the number of classes, and the value for pixel i on any data layer denotes the probability that the pixel is a member of that class. That is, each pixel is associated with a class membership probability vector, $[p_{i1}, p_{i2}, \dots, p_{ik}]$, where p_{ij} is the probability that pixel i is a member of class j . The probability vector provides a method for estimating the uncertainty inherent in responses to queries and is analogous to the notion of membership functions in fuzzy set theory (see § 6.1).

Given the k data layers for a given scene, the authors define a “realization” as the process whereby pixel probabilities are converted into specific classes. That is, the probability that a pixel i will be assigned to a class j is the probability p_{ij} . For example, the realization implied by the maximum likelihood classifier is one in which each pixel is assigned to the class with the highest probability. The authors describe a simulation method in which pixels may be assigned to one of two classes (i.e., $k = 2$). Realizations were acquired by independent trials for each pixel in a regular, rectangular matrix in which the probability of membership in class 1 decreased monotonically over the columns of the matrix. The realizations were smoothed with a spatial filter to introduce an appropriate degree of spatial autocorrelation. The resulting smoothed realizations were assumed to portray different representations of the boundary between the two classes based on the same set of pixel membership probabilities. This simulation method can be used to study cartographic generalization error. For example, a measure similar to the epsilon band (see § 4.1.1) can be computed from the mean displacement of the expected and observed boundaries for each realization. (The expected boundary is the column of the matrix for which pixel probabilities are equal for the two classes.) The simulation method might also be used to examine the spurious polygon problem (see § 5.3.2) by overlaying different realizations and computing the areas of agreement and disagreement.

Goodchild & Wang [103] and Goodchild [99] present two additional methods for generating realizations of pixel membership probabilities. The first swaps pixel class memberships between realizations until a target level of spatial autocorrelation is reached. The second posits a spatially autoregressive process that generates a realization based on a presumed level of spatial autocorrelation.

4.4. Isometric Mapping

Isarithmic maps depict a set of non-intersecting lines connecting points of equal value for some continuous variable distributed over space. If the attribute can exist at discrete points in space, these lines are referred to as “isometric” lines. “Contours” are isometric lines joining points of equal topographic elevation. In contrast to isometric lines, “isopleths” are isarithmic lines for attributes that cannot exist at discrete points (e.g., population density). Isoplethic mapping accuracy is discussed in § 4.5.

There are two distinct processes involved in isometric mapping that have some impact on the accuracy of the derived isometric map. "Interpolation" refers to the process in which the spatial distribution of the attribute is estimated from the observed values of the attribute at a set of sample points. The estimated spatial distribution, or "interpolated surface," commonly consists of a regular grid of points for which the attribute values have been interpolated from a set of randomly-located sample points. "Contouring," the second process in isometric mapping, involves the derivation of a set of isometric lines (or contours) from the interpolated surface. It is important to maintain the distinction between these two processes, since each is associated with different sources and types of error.

The discussion in this section of the report focuses primarily on errors associated with interpolation, due to an emphasis on this procedure in research articles concerned with isometric map accuracy. Most of the articles reviewed here are empirical studies demonstrating that isometric map accuracy is a function of the method of deriving the interpolated surface from a sample of points, the nature and complexity of the surface, the number and spatial distribution of sample points. Other studies focus on the interactions between interpolation error and the degree of measurement error in the observed values associated with these points. The problems associated with contouring (e.g., the classic saddle-point problem and the visual appearance of contours derived with different contouring methods) are described in many standard cartography texts.

Other sections of this report also complement the discussion of isometric map accuracy. For example, § 2.6 highlights some aspects of sampling and spatial resolution for digital elevation models that have direct parallels in isometric mapping. Methods of measuring the accuracy of contour lines as a function of measurement error in observed elevation values is discussed in § 2.3. Isometric mapping also has parallels to isoplethic mapping (in which isarithmic lines are derived from the values associated with a set of polygons), vector to raster conversion (in which the values associated with a set of polygons are interpolated to a regular grid of cells) and areal interpolation (in which the values associated with a set of polygons are interpolated to another set of polygons). These issues are discussed in § 4.5, § 5.2 and § 5.1.

4.4.1. Factors Affecting Interpolation Accuracy

Many of the earliest studies on interpolation accuracy focus on the effects of the number and spatial distribution of sample points. According to Morrison [166], the ability of a given set of sample points to capture the variation present in the surface is inversely related to the degree to which the points are clustered in space. Hence the relative accuracy of interpolation can be assessed with reference to the nearest neighbor statistic, or the ratio of the mean distance between each point and its nearest neighbor to the expected distance for a random distribution of points. The nearest neighbor statistic ranges from 0 (i.e., all points coincide spatially) to 2.1491 (i.e., all points are at their maximal dispersion). The author examined variations in the nearest neighbor statistic for six sampling methods — unaligned random (UR), aligned random (AR), unaligned stratified random (USR), aligned stratified random (ASR), unaligned systematic (US) and aligned systematic (AS). Empirical results indicate that for samples derived with any one of these methods, the nearest neighbor statistic tends to cluster around a unique modal value. The relative differences in the modal value for each method exhibited the following relationship:

$$AR < UR < ASR < USR < US < AS \quad (4.6)$$

According to the author, systematic sampling is preferable to stratified random sampling, and stratified random sampling is preferable to random sampling. Moreover, given a sample of points derived by any one of these methods, it is possible to infer the relative accuracy of an interpolated surface. These conclusions are based on the notion that a dispersed sample is more capable of capturing surface variation.

Rhind [182] argues that these conclusions are unwarranted, since there is no guarantee that a more dispersed sample will yield a more accurate interpolated surface. Surface-specific sampling, in which sample points are selected if they define critical points in the surface, may exhibit a high degree of spatial clustering but produce more accurate results than systematic sampling. Accuracy may also be affected in systematic sampling if periodicity is evident in the surface. Thus the appropriateness of a given sampling method depends in part on the nature of the surface.

The effects of the nature of the surface on interpolation accuracy have been explored by Morrison [167]. In an empirical test, four different synthetic surfaces were constructed, each continuous and each completely described by a finite set of mathematical terms. Each of the four surfaces was sampled using the six sampling methods described in Morrison [166]. For each surface/sampling method combination, four samples of different sizes were selected, yielding a total of 96 sets of sample points. For each of these sets, interpolation was performed with ten different interpolation methods. These methods consisted of a fifth-order polynomial fitted to the entire surface, a trigonometric or Fourier series and a set of first- and second-order polynomials fitted to a local neighborhood of points (in which the neighborhood was variously defined as containing between three and ten points).

Interpolation accuracy for each of the resulting 960 interpolated surfaces was assessed by comparing observed and interpolated values for a set of 100 grid points. Two indices of accuracy were defined — the correlation between the observed and interpolated values and the standard deviation of the residual values. The latter index, s , is defined as

$$s = \left[\frac{1}{n} \sum_{i=1}^n (r_i - \bar{r})^2 \right]^{1/2} \quad (4.7)$$

where: r_i = the residual value at point i , or the difference between the observed value, z_i , and the interpolated value, \hat{z}_i ;

\bar{r} = the mean residual value; and

n = the sample size.

Note that this index is similar to the RMSE (equation (2.52)), but will yield the same result only when the mean residual value is equal to 0.

For each of the two accuracy indices, three-way analysis of variance was employed to test the effects of interpolation method, sample size and sampling method on interpolation accuracy. For the second index of accuracy (i.e., equation (4.7)), the sampling and interpolation methods were observed to have the greatest effect on accuracy. Sample size and the first-order interactions between the three factors were observed to be only marginally significant. Similar results were obtained for the alternate index of accuracy (i.e., the correlation coefficient), except that sample size was observed to have a significant effect.

Comparison of interpolation accuracy for different sampling methods revealed that higher levels of accuracy were associated with unaligned methods (i.e., UR, USR and US). Considering unaligned and aligned methods separately, systematic methods produced more accurate results than stratified random methods, which in turn produced more accurate results than random methods. These findings are in accordance with Morrison [166], although they are partly attributable to the use of synthetic surfaces exhibiting an unrealistically high degree of smoothness. Variations in accuracy associated with sampling method were also observed to have a less significant impact as surface complexity increased, due to an overall decline in accuracy for more complex surfaces. An increase in accuracy was also observed as sample size increased, although this effect again was found to be less significant for complex surfaces. Interpolation methods based on fitting polynomials to a neighborhood of points yielded relatively low interpolation accuracies, although improvements were noted as the number of points in the neighborhood increased. The polynomial and Fourier series computed for the entire surface yielded more accurate results, although this finding may again be attributed to a high degree of

smoothness in the synthetic surfaces.

The factors considered by Morrison in this study have since been examined in greater detail by other authors. Shepard [196], for example, examined the effects of variations in sample size on interpolation accuracy. The author interpolated the grid point values for a 60×64 grid using a sample of between 4 and 272 points randomly selected from the grid. The relative RMSE was computed for all 3840 grid points on each interpolated surface. The relative RMSE is defined as the ratio of the RMSE (see equation (2.52)) to the standard deviation of the observed grid point values. Regression results indicated a close fit ($r^2 = 0.997$) between the relative RMSE and sample size according to a relationship of the form

$$\text{RMSE}_r = 2.83 n^{-0.56} \quad (4.8)$$

where: RMSE_r = the relative RMSE; and

n = the sample size.

Hence as n increases, RMSE_r decreases at a declining rate.

In a similar study, MacEachren & Davidson [151] examined the effects of sample size on interpolation accuracy for surfaces of varying complexity. Six surfaces were defined, each of which portrayed topographic elevation values for a 103×103 grid. Eight samples of points were obtained for each surface using unaligned stratified random sampling, in which sample size varied between 100 and 2025 points. Accuracy was calculated as the mean absolute deviation between actual and interpolated values for all 10 609 grid points on the surface. The mean absolute deviation, d , is defined as

$$d = \frac{1}{n} \sum_{i=1}^n |z_i - \hat{z}_i| \quad (4.9)$$

Regression results revealed a close fit between d and n according to an equation of the form

$$d = a n^{-b} \quad (4.10)$$

Coefficient b was observed to be relatively constant over all surfaces, with a mean value of approximately -0.3 . Coefficient a was found to be directly related to the range of elevation values on the surface, such that the value of a was lower for less complex surfaces. The authors examined the spatial distribution of error (i.e., the absolute deviation between observed and interpolated values at the grid points) and observed that errors tended to be more clustered in space for smaller sample sizes.

The authors also computed the appropriate contour interval based on the absolute deviations between observed and interpolated values. For each surface/sample size combination, the minimum absolute deviation was determined such that no more than 10 percent of the interpolated values deviated by more than this amount from their corresponding observed values. This minimum deviation corresponds to one-half the contour interval according to the National Map Accuracy Standard (see § 2.3). Considerable variation in the minimum deviation was observed as a function of interpolation accuracy, due to the effects of sample size and the nature of the surface. This approach facilitates contouring of interpolated surfaces using a contour interval that reflects interpolation accuracy.

The effects of interpolation method have been explored by other authors as well. Battle [15], for example, evaluated the effects of varying the weighting function in interpolation methods based on distance-weighted averaging. According to this method, the interpolated value at grid point j , \hat{z}_j , is given by

$$\hat{z}_j = \frac{\sum_{i \in N} z_i}{\sum_{i \in N} d_{ij}^\gamma} \quad (4.11)$$

where: z_i = the observed value at sample point i ;
 d_{ij} = the distance between sample point i and grid point j ;
 γ = a weight defining the influence of sample point i on grid point j as a function of d_{ij} ; and
 N = the set of points defining the neighbors of grid point j .

Sample points for the evaluation were selected if they defined critical points in the surface (i.e., surface-specific sampling). This method of sampling was also compared to aligned systematic sampling, in which sample points corresponded to points in a coarse grid. Two aligned systematic samples were obtained to assess the effects of grid orientation. In the evaluation of sampling effects, the value of γ in equation (4.11) was fixed at 1.5. Accuracy was assessed for each interpolated surface as the mean absolute error (see equation (4.9)).

Variations in γ for surface-specific sampling were observed to produce small but systematic changes in the accuracy of the interpolated surface. A value of 1.5 yielded the smallest mean absolute error and produced a relatively high frequency of small errors at individual points. Thus the optimal value of γ depends on the variability of the surface. Depending on grid orientation, the aligned systematic sample of points yielded a mean absolute error less than or approximately equal to the surface-specific sample. The frequency distribution of absolute errors for the surface-specific sample showed a higher concentration of both small and large errors than the aligned systematic samples. That is, the errors associated with aligned systematic sampling were more consistent and of an intermediate range. According to the author, the optimal sampling method is a regular grid where the spacing of grid points is a function of surface variability.

Braile [30] examined the effects of four different interpolation methods on interpolation accuracy. Two of the methods were based on distance-weighted averaging of neighboring sample points (see equation (4.11)), but differed in terms of the definition of the neighborhood, N . For the first method N was defined as the set of three points closest to the grid point that form a triangle encompassing the grid point. For the second method N was defined as the set of three to nine points closest to the grid point. For the third interpolation method an n^{th} -order polynomial was fitted to segments of the surface. The values at grid points were then derived from the fitted equations. For the final method sample points were arranged into irregular profiles and a cubic polynomial was fitted along lines perpendicular to the profiles and passing through each grid point.

These interpolation methods were applied to a sample of 200 random points derived from a digitized aeromagnetic map. The RMSE (see equation (2.52)) was computed for each interpolated surface as an index of accuracy. The RMSE was observed to vary significantly for the four interpolated surfaces, with the n^{th} -order polynomial interpolation method yielding the lowest error and the second distance-weighted averaging method yielding the highest error. Areas of local minima and maxima, as well as areas along map borders, were interpolated inaccurately by both distance-weighted averaging methods. None of the methods was able to reproduce short wavelength variations in the surface.

When interpolating over large portions of the earth's surface, the interpolation method must also account for the non-Cartesian distance and directional relationships between points (Willmott *et al* [228]). Interpolation of climatological data typically begins with the collection of data associated with a set of irregularly-distributed weather stations. These are then transformed into two-dimensional space by assuming an equivalence between the longitude and latitude of each station and the x - and y -coordinate values in two-dimensional space. Similarly, the latitude-longitude graticule is transformed into two-dimensional space as a regular grid. Interpolation and contouring are then performed by assuming that Cartesian distance and directional relationships hold between all stations and all grid points. To avoid the distortion in isometric lines resulting from this approach, interpolation and contouring should be performed on the sphere and only then transformed into two-dimensional space.

In an empirical demonstration, the authors employed two different methods to interpolate and contour temperature data for 100 randomly-selected weather stations located between 2 and 90° N and 50 and 170° W. In the first method, the locations of the stations and the grid points in a 4° latitude by 5° longitude lattice were transformed into two-dimensional space using two different cylindrical projections. Temperature values were then interpolated from the stations to the grid points using an interpolation method based on distance-weighted averaging (see equation (4.11)). (The interpolation method employed also corrects for “directional isolation” of points, such that points with a small angular separation from a grid point contribute less to the interpolated value than points with a large angular separation, under the assumption of positive autocorrelation.) Cartesian distance and directional relationships between points were assumed to hold. Isometric lines were then derived from the interpolated grid. The second method employed by the authors involved interpolation and contouring on the sphere, followed by the transformation of isometric lines onto the two cylindrical projections used in the first method. Great-circle distances between points were used in the interpolation, and the directional isolation of points was calculated as a function of the latitude and longitude of the points. A comparison of the results of the two interpolation methods revealed significant errors in contour lines derived with the first method. Errors included such anomalies as multiple contours passing through the pole (represented as a line on the cylindrical projections). Errors in isometric line locations were observed to be largest where data points were most sparse or projection-related distortions were greatest.

Peucker [179] and Rhind [182, 183] maintain that the method of interpolation is perhaps the most critical factor affecting interpolation accuracy. The appropriateness of a given interpolation method depends on the nature of the actual surface, the number and spatial distribution of sample points, the degree of measurement error in the values associated with these points and the desired level of accuracy (which is application-specific). In addition to quantitative accuracy, the interpolation method should ensure surface “representativeness” or truthful depiction of the overall form of the surface.

Barrett & Rhind [12] describe a technique that may be used to enhance the representativeness of the interpolated surface based on filtering of spatial frequencies. Many interpolation methods are unable to account for local minima and maxima (i.e., boundary constraints) in the interpolated values. Arbitrary digitization of sample points at which such minima and maxima occur offers a partial solution, but has the disadvantage that these points may condition the overall appearance or representativeness of the map. The technique presented by the authors is based on the construction of the spectral field, F , from a regular grid of sample point values, Z , using the Fourier transform. In this case, Z contains sample points defining local minima. A second spectral field, F' , is then constructed in the same way, but using a matrix of values Z' for which all minima values have been replaced with a higher value (which is application-specific). The differences between F and F' indicate those elements of the spectral field which are contributed to by the minima values, and are deleted from F to produce a third spectral field F'' . The inverse Fourier transform is then applied to F'' to yield Z'' , a filtered matrix of sample point values. Empirical tests reveal enhanced trend representation in the isometric map constructed from Z'' .

4.4.2. Interactions between Interpolation and Measurement Error

The preceding studies have considered interpolation accuracy to be a function of sampling, surface and interpolation method effects. A fourth effect relates to the degree of measurement error in the values associated with sample points and is closely related to topographic map accuracy (see § 2.3). A number of authors have proposed techniques for assessing the interactions between interpolation and measurement error. These techniques often suggest ways in which interpolation methods can be devised to minimize the effects of such interactions.

Stearns [200] presents a method for estimating interpolation error for isometric maps constructed from observations that vary in space and time. This method accounts for measurement error (including observational, time and positional error), interpolation error and “synopticity error” (i.e., the discrepancy between the recorded and reference time of an observation). All types of error are converted to map error (i.e., error in interpolated values at points) by means of conversion factors. Three types of map error are defined — error associated with point observations, primary interpolation error (i.e., error associated with points interpolated along lines between observations, or the primary interpolation axes) and secondary interpolation error (i.e., error associated with points interpolated between the primary interpolation axes). Error associated with point observations, ϵ_1 , is defined as

$$\epsilon_1 = e_o + e_t h_t + e_p g_p \cos(\gamma_p) + e_a h_a \quad (4.12)$$

where: e_o = observational error (i.e., the deviation between the actual and observed value at a point);
 e_t = time error (i.e., the difference between the recorded time of an observation and the actual time);
 h_t = the rate of change over time in the mapped variable in the vicinity of the point;
 e_p = positional error (i.e., the distance between the actual and observed position of a point);
 g_p = the local ascendant (i.e., the positive rate of change in space of the mapped variable in the vicinity of the point);
 γ_p = the angle defined by e_p and g_p ;
 e_a = synopticity error (i.e., the difference between the recorded time of the observation and the time selected as the reference time of the map); and
 h_a = the rate of change in the mapped variable in the interval between the recorded and reference times.

The first three terms of equation (4.12) define measurement error and thus ϵ_1 represents the sum of measurement and synopticity errors. Parameters h_t , g_p and h_a are conversion factors to convert time, positional and synopticity errors to map error.

The second type of map error, primary interpolation error, ϵ_2 , is defined as

$$\epsilon_2 = e_l g_l \cos(\gamma_l) + \epsilon_1 \quad (4.13)$$

where: e_l = truncation error (i.e., the distance between the true position of a given value and its interpolated position);
 g_l = the local ascendant; and
 γ_l = the angle defined by the primary interpolation axis and g_l .

ϵ_2 represents the sum of the primary truncation errors and propagated measurement and synopticity errors.

The third type of map error, secondary interpolation error, ϵ_3 , is defined as

$$\epsilon_3 = e_l' g_l' \cos(\gamma_l') + \epsilon_2 \quad (4.14)$$

where parameters are defined as in equation (4.13) but for points interpolated between the primary interpolation axes.

Overall map reliability is expressed in terms of the mean and variance of ϵ_3 , under the assumption that all parameters are independent. The choice of different patterns of interpolation points (e.g., transects, triangular grids and regular and irregular grids) influences overall map reliability as a function of the angle between the local ascendant and the interpolation axes. Several of the parameters employed in the technique (e.g., e_o , e_t and e_p) are context-dependent and must be estimated uniquely for different surfaces. Other parameters (e.g., h_t , h_a , g_p and g_l) can be estimated from average rates over the entire map.

Rattray [181] also details a method of assessing accuracy as a function of interpolation error and measurement error in observed values at sample points. The method described is applicable to interpolation in one dimension (i.e., along horizontal or vertical profiles). The interpolated value at a point is expressed as a weighted mean of two Lagrangian interpolation polynomials. Consider four points along a profile, x_1 through x_4 , and their corresponding values, z_1 through z_4 . If $x_1 < x_2 < x_0 < x_3 < x_4$, then the interpolated value at x_0 , \hat{z}_0 , is given by

$$\hat{z}_0 = p_{ijk}(x_0) + R_{ijk}(x_0) \quad (4.15)$$

where $p_{ijk}(x_0)$ is given by the Lagrangian interpolation formula

$$p_{ijk}(x_0) = \frac{(x_0 - x_j)(x_0 - x_k)}{(x_i - x_j)(x_i - x_k)} z_i + \frac{(x_0 - x_i)(x_0 - x_k)}{(x_j - x_i)(x_j - x_k)} z_j + \frac{(x_0 - x_i)(x_0 - x_j)}{(x_k - x_i)(x_k - x_j)} z_k \quad (4.16)$$

and the error, $R_{ijk}(x_0)$, is given by

$$R_{ijk}(x_0) = \frac{(x_0 - x_i)(x_0 - x_j)(x_0 - x_k)}{3!} z'''(\xi_j) \quad (4.17)$$

where: $z'''(\xi_j)$ = the third derivative evaluated at ξ_j , where $x_i \leq \xi_j \leq x_k$.

Two interpolation estimates may be obtained with equation (4.16), one with $ijk = 123$ (i.e., x_1, x_2 and x_3) and one with $ijk = 234$ (i.e., x_2, x_3 and x_4). The interpolated value at point x_0 may then be expressed as a weighted mean of these two estimates,

$$\hat{z}_0 = [r p_{123}(x_0) + (1 - r) p_{234}(x_0)] + [r R_{123}(x_0) + (1 - r) R_{234}(x_0)] \quad (4.18)$$

where: r = a weighting factor, usually equal to 0.5.

The second term in this equation is referred to as $R(x_0)$. Rearranging the above equations yields

$$R(x_0) = [p_{123}(x_0) - p_{234}(x_0)] [r r_0(x_0)] \quad (4.19)$$

where $r_0(x_0)$ is defined as

$$r_0(x_0) = \left[1 + \frac{(x_0 - x_1) z'''(\xi_2)}{(x_4 - x_0) z'''(\xi_3)} \right]^{-1} \quad (4.20)$$

Hence the interpolation error (equation (4.19)) is defined as the difference in the two Lagrangian interpolation estimates times a factor depending on the third derivative of the function z in the vicinity of the interpolated point.

The author also presents a means of separating interpolation errors from measurement errors in the observed values at sample points. The ratio of error in interpolated values associated with these measurement errors to the total measurement error in sample points is shown to be wholly determined by the sample and interpolation point locations and the value of r . (The reader is referred to the article for a complete proof.) Hence it is possible to devise sampling methods that will yield relatively low measurement-related errors in interpolated values. The effects of interpolation error, however, can only be assessed after interpolation has been performed. These techniques are illustrated by the author with reference to ocean temperature and salinity data along depth profiles.

Switzer [204] analogously argues that, in two dimensions, covariance modeling may be used to determine the optimal locations of sample points to reduce error in interpolated values. Accuracy depends only on sampling design, non-stationarity in the surface, temporal effects and the sensitivity of interpolated values to changes in sample point location and value. Numerous techniques exist for assessing interpolation accuracy, including cross-validation and kriging.

Newton [173] describes an interpolation method based on the premise that the observed value at each sample point is not unique, but rather has been drawn from a probability distribution around the point. The author posits a simple radial dependence probability distribution for each sample point i , P_i , such that

$$\begin{aligned} P_i &= \exp \{ -A_i [(x_i - x_p)^2 + (y_i - y_p)^2]^{1/2} \} \\ &= \exp(-A_i r_i) \end{aligned} \quad (4.21)$$

where: A_i = a weight associated with the influence of sample point i ;

x_i = the x-coordinate of sample point i ;

y_i = the y-coordinate of sample point i ;

x_p = the x-coordinate of an interpolated point;

y_p = the y-coordinate of an interpolated point; and

r_i = the radius, or distance, from sample point i to an interpolated point.

Replacing x_p and y_p with the actual location x_j and y_j of an interpolated point j , the predicted value at point j , \hat{z}_j , is given by

$$\hat{z}_j = \frac{\sum_{i=1}^n z_i P_i}{\sum_{i=1}^n P_i} \quad (4.22)$$

where: z_i = the observed value at sample point i ; and

n = the number of sample points.

The value of A_i represents the influence of sample point i , since the larger the value of A_i , the less the influence of point i on a given interpolated point. The value of A_i may be established for each sample point such that it reflects a measure of confidence in the value associated with the sample point.

4.5. Isolethic Mapping

Isoleth maps depict a set of non-intersecting lines called isopleths that join points of equal value on a statistical surface. Isolethic and isometric maps are both classes of isarithmic maps. The distinction between the two lies in the nature of the statistical surface. In isometric mapping, it is assumed that the z -values defining the surface can exist at discrete points in space (e.g., topographic elevation). In isoplethic mapping, the z -value at any point is assumed to represent a quantity distributed over space (e.g., population density). Isoleth maps are constructed by interpolating between z -values associated with a set of areal units, or polygons. In practice, the distinction between isometric and isopleth maps is often blurred, since it is always possible to aggregate or average the z -values at discrete points over a set of areal units.

As in the case of isometric maps, the accuracy of isopleth maps is affected by the nature of the underlying statistical surface, the interpolation technique and the sampling method. For isopleth maps, the sampling method refers to the number and shape of areal units to whose centroids z -values have been assigned. The studies reviewed in this section, although few in number, focus exclusively on sampling methods and their interactions with surface complexity.

Hsu & Robinson [122] examined the effects of areal unit size and shape on isopleth map accuracy. Accuracy was assessed for a set of isarithmic maps constructed from different statistical surfaces and different sets of areal units. Four isarithmic maps of varying complexity were defined, one of which was synthetic and the other generalized topographic surfaces. Four sets of areal units were also defined in which unit size was held relatively constant but unit shape was permitted to vary. For each isarithmic map/areal unit combination, the z -value associated with each areal unit, \bar{z} , was calculated as

$$\bar{z} = \frac{\sum_{i=1}^{n+1} a_i m_i}{\sum_{i=1}^{n+1} a_i} \quad (4.23)$$

where: a_i = the area within the areal unit between isarithm i and isarithm $i-1$;
 m_i = the midpoint of the z -values of isarithms i and $i-1$; and
 n = the number of isarithms in the areal unit.

To examine the effects of areal unit size, each of the isarithmic map/areal unit combinations was transformed to each of five sets of hexagons, where the average size of the hexagons was different for each set. The centroid of each hexagon was assigned a z -value based on the proportion of its area overlapping each areal unit weighted by the z -value calculated for that areal unit. In contrast to the centroids of irregular areal units, hexagon centroids describe a regular triangular pattern. The authors argue that this pattern is more conducive to interpolation, and thus isopleth maps were constructed by interpolating between hexagon centroids for each of the eighty isarithmic map/areal unit/hexagon combinations.

Accuracy was evaluated by obtaining a stratified random sample of thirty points for each combination. The discrepancy between each original isarithmic map and each of the twenty corresponding isopleth maps (i.e., four sets of areal units by five sets of hexagons) was then computed for each sample point. Four-way analysis of variance was employed to test the significance of areal unit size and shape, hexagon size and sample point location. All main effects and all two-way interactions were observed to be significant, but sample point location was found to be a more critical factor for more complex isarithmic maps. The overall discrepancy between the isarithmic and isoplethic z -values, d , was computed for each of the eighty isopleth maps (see equation (4.9)). An increase in the value of d was associated with an increase in isarithmic map complexity, hexagon size and the irregularity of areal unit shape. An identical pattern was observed for the standard deviation of d , a measure of precision.

Hsu [121] examined isopleth map accuracy in the context of a spatial filtering paradigm. Aggregation was performed by filtering the original continuous surface, Z , to produce a stepped statical surface, Z' , containing a z -value at each areal unit centroid. Interpolation was then performed by filtering the stepped statistical surface, Z' , to produce the isopleth map, Z'' , containing a set of isopleths joining points of equal value.

The author examined the effects of areal unit size on the derivation of z -values for unit centroids in the aggregation process. Two synthetic surfaces were constructed for which z -values were defined for an array of 115×115 grid cells. Aggregation was performed for each surface over a set of rectangular areal units. Each areal unit was composed of an array of $(2m+1) \times (2n+1)$ grid cells, where the values of m and n depend on areal unit size and shape. The aggregation process was performed with the two-dimensional filter

$$\bar{z}_{ij} = \frac{1}{N} \left(\sum_{k=-m}^m \sum_{l=-n}^n z_{i+k,j+l} \right) \quad (4.24)$$

where: \bar{z}_{ij} = the filtered z -value for cell ij corresponding to the centroid of a particular areal unit;
 $z_{i+k,j+l}$ = the z -value for cell $i+k,j+l$ on the original surface; and
 $N = (2m+1) \times (2n+1)$.

This filter averages all z -values within a given areal unit and assigns the result to the unit centroid.

The effects of aggregation were assessed by computing an index of relative error, e_{ij} , for each unit centroid as

$$e_{ij} = r_{ij} / z_{ij} \quad (4.25)$$

where:

$$r_{ij} = -\frac{1}{N} \left[\left(\sum_{k=-m}^m \sum_{l=-n}^n z_{i+k,j+l} \right) - z_{ij} \right] + \frac{N-1}{N} z_{ij} \quad (4.26)$$

phenomenon to the map reader, should optimally strike a balance between map complexity and readability. On choropleth maps, for example, map accuracy rises as the number of class intervals is increased, but a threshold is reached at which additional class intervals cause a decline in map readability due to an excessive level of map detail (see Jenks & Caspall [131], § 4.3.1).

Jenks [128] argues that the most important factors determining the effectiveness of cartographic communication are conceptual and perceptual error. Conceptual error results from the cognitive transformation of real-world observations by the cartographer and the translation of these concepts into graphic form on the map. Perceptual error, in contrast, refers to the degree to which the map reader fails to duplicate the mapped concepts in the construction of the reader's cognitive model. Jenks maintains that these types of error are poorly understood but often more important than positional error in cartographic features and measurement error in data values.

Jenks [129] argues further that the effectiveness of cartographic communication depends on the cartographer's clarity as to the intended purpose of the map, the symbology employed in map compilation, the fidelity of the mapped representation of the phenomenon and the distortions associated with the reader's perception of the map. Confusion over map purpose may occur because the map is an inherently inaccurate source of information, since it is intended only to depict the general form of some phenomenon. The map is a graphic generalization of a set of data values and cartographers must recognize and work within this framework rather than overloading a map with excessive detail. The conception of a map as a graphic generalization requires that symbols be used to depict the phenomenon on the map. These symbols must be selected in accordance with the nature of the phenomenon. Similarly, map fidelity is largely a function of decisions made during map compilation. In choropleth mapping, for example, fidelity is enhanced by the selection of class intervals that minimize within-class variance and maximize between-class variance (see Jenks & Caspall [131], § 4.3.1). The distortions associated with map perception interact with each of these factors to determine the ability of a given map to communicate its message clearly.

Herzog [115] similarly argues that in the context of computer mapping, "visualization" or map perception is an important component of map accuracy since the map represents the human interface with the computer. Visualization incorporates numerous aspects of graphic design and the technical methods employed in map construction. Map generalization is a particularly important variable since it may have a direct impact on the success of cartographic communication. The advent of computer technology means that traditional cartographic techniques must be applied in new ways. However, many commercial computer mapping packages are not adaptable to such techniques since they present too few graphic design options.

Other authors maintain that the standard representation of cartographic communication given above must be extended to incorporate the notion of liability. That is, cartographers must be aware of the ways in which a given map may be misinterpreted and must design maps that minimize the chance of such misinterpretations. Gersmehl [93] illustrates how misinterpretation may occur with reference to a series of small-scale maps of the nine major soil orders in the US. Despite the relatively low spatial and taxonomic resolution of these maps, they were eventually applied in a planning context in which a much higher degree of resolution was assumed to exist. Gersmehl [92] maintains that such problems occur because the positional and attribute features of maps are often represented at higher levels of implied precision than the actual precision of the data would warrant. Cartographers should therefore make use of uncertainty bands around data values, fuzzy class intervals with overlapping class boundaries, or additional information in the map legend, in order to portray non-homogeneity in areal units mapped as homogeneous. Alternatively, locationally ambiguous symbols can be employed such that the general form of the phenomenon is preserved on the map, but the reader is prevented from making inappropriate inferences about the characteristics of individual sites.

Beard [18] also examines the question of liability associated with the misapplication of maps. The author refers to such misapplications as “use errors,” and suggests that these errors are commonly associated with the attempt to extract detailed information from a map with low spatial or taxonomic resolution, a high level of cartographic generalization, or a graphic language that is unfamiliar to the map reader. Use errors often arise when time and budget constraints do not permit more detailed data to be collected, particularly when data collection is costly or time-consuming. For example, use errors may occur when small-scale maps are used for detailed planning, maps are out-of-date, or quantitative analysis is performed without considering scale and generalization effects (e.g., the variations in line length estimates as a function of map scale, as discussed in § 3.1). The author argues that computer technology affords an opportunity to reduce the incidence of use error. Computers allow data to be stored in a detailed, disaggregated manner. Updating is much less expensive and time-consuming than in the era of paper maps. Documentation of the quality of each map component is also possible. Moreover, data may be structured to prevent inappropriate operations, such as the arithmetic manipulation of ordinal data.

Bedard [22, 24] examines the issue of liability for land information system (LIS) databases in the context of cartographic communication. The author views this process as the construction of a sequence of cognitive and physical models, as described above. Uncertainty is introduced whenever a model is constructed. Uncertainty results both from the limitations of human cognitive processes and the inherent distortions associated with modeling processes. These distortions include the loss of detail and context-dependent distortions associated with the equivocacy of model-building rules. Models are approximations of reality in which measurements are imprecise and indeterminacy exists in the identification and labeling of feature positions and attributes. LIS databases may therefore be said to contain at least four types of uncertainty. “Conceptual uncertainty” refers to indeterminacy in the identification of features. “Descriptive uncertainty” and “locational uncertainty” refer respectively to imprecision in feature attributes and positions. “Meta-uncertainty” refers to uncertainty in the preceding types of uncertainty.

Liability concerns demand that uncertainty be either reduced or absorbed. Uncertainty reduction is facilitated by establishing modeling and communication rules and standards to reduce indeterminacy in identification and to enhance precision in feature attributes and positions. Uncertainty absorption occurs when the developer of the database guarantees it and agrees to compensate the user for damages associated with the presence of error. In this case, the database becomes “official truth” to which the data developer subscribes. Alternatively, uncertainty may be absorbed by the user who employs an unguaranteed database and receives no compensation for damage associated with the presence of error. In either case, the degree of uncertainty absorption is defined in terms of the level of risk in providing or using the database. An illustration of the notion of uncertainty absorption is provided by Bedard [23], who examined the degree of uncertainty absorption in different land registration systems in North America and Europe. The author evaluated the degree to which each system guarantees an individual’s rights to land and the spatial location of these rights. Each system was ranked on a “user uncertainty absorption spectrum,” which ranged from complete absorption by the user to complete absorption by the system. Results indicate that different systems vary in terms of their willingness to accept responsibility for errors. Hence liability is a function of institutional arrangements and the degree of coincidence between the views of the user and the system as to the intended function of the database.

5. PROPAGATION OF ERROR THROUGH GIS OPERATIONS

This section reviews models of error for operations that might be applied in a GIS environment. These models account for error propagation, or the ways in which errors present in input data are transformed by the operation. Areal interpolation (see § 5.1) is the operation in which the values for some variable of interest are estimated for a set of target areal units based on the corresponding values for a different set of source areal units. Vector to raster conversion (see § 5.2) is an analogous operation in which the target zones are regular grid cells. Map overlay (see § 5.3) involves the superimposition of two or more data layers to produce a composite map depicting combinations of the attributes from each data layer.

5.1. Areal Interpolation

Areal interpolation or cross-area estimation refers to the operation in which the values for some variable of interest are estimated for a set of m "target" zones based on the corresponding values for a set of n "source" zones. The target and source zones represent two different partitionings of the same geographical area. The region bounded by the set of target zones is not necessarily identical to that bounded by the set of source zones. Areal interpolation may be necessary when zonal boundaries have been redefined or one wishes to perform quantitative or statistical analysis on variables that have been collected for different sets of areal units.

In general, $m \neq n$ and the boundaries of individual target and source zones do not coincide. Aggregation may be viewed as a special case of the more general areal interpolation model. In aggregation, $m < n$ and each source zone is wholly encompassed within a single target zone. Areal interpolation is a more complicated operation than aggregation and is affected by the myriad of ways in which the value for a given source zone might be apportioned between the target zones that overlap it. Accuracy in areal interpolation often requires that the underlying spatial distribution of the variable of interest be taken into account by the areal interpolation method, yet information about this distribution is often unknown for data that have been tabulated by areal units.

Markoff & Shapiro [156] describe a number of techniques for performing areal interpolation. These techniques are differentiated according to the class of the variable of interest. Classes include 1) absolute figures (or "spatially extensive" data) characterizing some aspect of area (e.g., area of land under cultivation), 2) proportions (or "spatially intensive" data) characterizing some aspect of area (e.g., percentage of land under cultivation), 3) absolute figures characterizing some aspect of population (e.g., total population) and 4) proportions characterizing some aspect of population (e.g., population density).

For the first two classes, "common area" techniques are used. For case 1, values for target zones may be estimated from source zone values weighted by the overlapping area of the source and target zones.

$$v_i = \sum_{j=1}^n \frac{a_{ij}}{a_j} u_j \quad (5.1)$$

where: v_i = estimate of the variable of interest for target zone i ;

a_{ij} = area of overlap between target zone i and source zone j ;

a_j = area of source zone j ; and

u_j = value of the variable of interest for source zone j .

For case 2, a weighted mean is calculated for each target zone, where the weights are the proportions of the target zone that overlap the source zone.

$$v_i = \sum_{j=1}^n \frac{a_{ij}}{a_i} u_j \quad (5.2)$$

where: a_i = area of target zone i .

These techniques assume that the variable of interest is distributed uniformly over each source and target zone, an assumption that is clearly inconsistent with the distribution with data types 3 and 4. A more appropriate areal interpolation technique for such data is based on shared population rather than shared area. This is referred to as the "common population" technique. For case 3, the technique involves estimation of the urban population of an area of overlap using known locations and populations of cities, and then estimation of the rural population of the area of overlap based on equation (5.1). That is,

$$v_i = \sum_{j=1}^n \left(\frac{p'_{ij}}{p_j} + \frac{p_j - p'_j}{p_j} \frac{a_{ij}}{a_j} \right) u_j \quad (5.3)$$

where: p'_{ij} = urban population of the overlap between target zone i and source zone j ;

p_j = total population of source zone j ;

p'_j = urban population of source zone j ; and

$p_j - p'_j$ = rural population of source zone j .

For case 4, the appropriate equation is

$$v_i = \sum_{j=1}^n \left(\frac{p'_{ij}}{p_i} + \frac{p_i - p'_i}{p_i} \frac{a_{ij}}{a_i} \right) u_j \quad (5.4)$$

where: p_i = total population of target zone i ;

p'_i = urban population of target zone i ; and

$p_i - p'_i$ = rural population of target zone i .

The authors present empirical results for population data showing that the common population technique is more accurate than the common area technique and that both techniques are more accurate when estimating "upwards" from small source zones to large ones (i.e., $m < n$) than "downwards" from large source zones to small target zones (i.e., $m > n$). However, the improvement in accuracy associated with the common population technique is less dramatic when $m < n$, since the total population of each source zone, both urban and rural, tends to be allocated to a single target zone.

Crackel [59] argues that these techniques assume that the area of the target zone is equal to the sum of the areas of the source zones that overlap it. That is, for a given target zone i ,

$$a_i = \sum_{j=1}^n a_{ij} \quad (5.5)$$

This assumption will not hold if the region bounded by the set of target zones is not identical to that bounded by the set of source zones. The value for target zone i is likely to be underestimated when

$$a_i > \sum_{j=1}^n a_{ij} \quad (5.6)$$

and overestimated when

$$a_i < \sum_{j=1}^n a_{ij} \quad (5.7)$$

To account for this problem, it is necessary to weight the numerical result of equation (5.1) or (5.3) by $a_i / \sum_{j=1}^n a_{ij}$ and to replace the a_i denominator in equations (5.2) and (5.4) by $\sum_{j=1}^n a_{ij}$.

Goodchild and Lam [101] focus on the common area technique of areal interpolation, which the authors refer to as the “direct overlay method.” In matrix notation, the direct overlay method may be expressed as

$$\mathbf{V} = \mathbf{WU} \quad (5.8)$$

where: \mathbf{V} = an $m \times 1$ matrix where v_i is the estimate of the variable of interest for target zone i ;

\mathbf{W} = an $m \times n$ matrix where w_{ij} is the standardized area of overlap between target zone i and source zone j ; and

\mathbf{U} = an $n \times 1$ matrix where u_j is the value of the variable of interest for source zone j .

For spatially extensive data, or absolute figures,

$$w_{ij} = a_{ij} / \sum_{i=1}^m a_{ij} \quad (5.9)$$

For spatially intensive data, or proportions,

$$w_{ij} = a_{ij} / \sum_{j=1}^n a_{ij} \quad (5.10)$$

The denominators in equations (5.9) and (5.10) are equivalent to a_j and a_i respectively when the area encompassed by the set of target zones is identical to that encompassed by the set of source zones.

According to the authors, the accuracy of the target zone estimates is reflected in the structure of \mathbf{W} . This matrix tends to be sparse and its elements are always non-negative. Small elements are often spurious and may arise from digitizing errors and discrepancies between the cartographic representations of zonal boundaries on the source and target maps. The amount of error in the target map may therefore be deduced from the number and magnitude of elements of \mathbf{W} that are neither 1 nor 0. An error-free target map can be assured only when each column of \mathbf{W} contains only one non-zero element (i.e., the aggregation operation, in which each source zone is wholly encompassed within a single target zone).

A measure of the accuracy of the target map could presumably be obtained by reversing the interpolation. Taking equation (5.8), the reverse interpolation is given by

$$\mathbf{U}^* = \mathbf{BV} \quad (5.11)$$

where: \mathbf{U}^* = an $n \times 1$ matrix where u_j^* is an estimate of the variable of interest for source zone j ;

\mathbf{B} = an $n \times m$ matrix where b_{ji} is the standardized area of overlap between source zone j and target zone i ; and

\mathbf{V} = an $m \times 1$ matrix where v_i is the estimate of the variable of interest for target zone i .

In the case of spatially extensive data,

$$b_{ji} = a_{ij} / \sum_{j=1}^n a_{ij} \quad (5.12)$$

Since the summation is performed over source zones rather than target zones, as in equation (5.9), in general $\mathbf{B} \neq \mathbf{W}^t$.

In an empirical test of the direct overlay method, the authors estimated the population for a set of target zones for which the actual populations were known in advance. This made it possible to directly compute the accuracy of target zone estimates. The accuracy of the direct overlay method was compared to that of pycnophylactic interpolation (based on Tobler, as cited by the authors) and interpolation based on distance-weighted averaging. In general the direct overlay method was found to yield the most accurate results. Comparison of the \mathbf{U} and \mathbf{U}^* matrices revealed that by reversing the interpolation, target zones with inaccurate estimates can readily be identified. In general, the largest errors

were found to occur where the population was least homogeneously distributed within source zones.

Two methods for approximating the W matrix that are computationally less demanding were also examined. The first consists of assigning each source zone value to the target zone with which it has the greatest overlap (in the case of spatially extensive data). The second consists of apportioning the source zone value equally among all overlapping target zones. Both methods yielded accuracies higher than pycnophylactic interpolation but lower than the direct overlay method. The authors also experimented with the aggregation of source zones into new zones with maximum internal homogeneity with respect to the variable of interest. Target zone estimates based on these new source zones were found to be relatively poor, although they were substantially improved by the addition of a compactness or a size constraint.

Lam [136, 137] argues that for the direct overlay method the accuracy of the target map is primarily a function of the presence of "split" source zones, or source zones that overlap more than one target zone. The underlying spatial distribution of the variable of interest determines the degree to which split source zones will contribute to error, since the direct overlay method assumes that this distribution is homogeneous. The size and shape of the source and target zones contribute to the degree of variation discernible on the source and target maps, but their effects on accuracy are less clear. The author posits an error model in which the error in a given target zone estimate is a function of the number and area of split source zones overlapping the target zone and the mean absolute difference between the value of each split source zone and its neighbors.

This model was tested for four fractal surfaces for which D ranged from 2.1 to 2.9 (see § 3.1.2). Each surface was partitioned into sets of rectangles of varying sizes representing the source and target zones. Error was measured in terms of the deviations between the fractal surface and the target zone estimates. Correlation coefficients were computed between the error associated with each target zone estimate and the three factors postulated in the model to account for error. Correlations of between 0.6 and 0.99 were observed, with the highest correlations occurring for surfaces of low dimensionality and a large number of source zones relative to target zones (i.e., aggregation). The model was also tested for pycnophylactic interpolation but was found to be a relatively poor predictor for this method.

Flowerdew & Openshaw [85] discuss the data problems associated with areal interpolation and present a typology of potential problems based on the nature of the areal unit, the scale of measurement and the relationship between the areal unit and the variable of interest. Areal units may be either "natural" (defined by the phenomenon under consideration), "imposed" (defined for statistical or enumerative purposes), or "arbitrary" (bearing no relationship to real-world phenomena). Measurement scales include "categorical" (both dichotomous and polychotomous), "count," "continuous" and "rankings." A distinction is also drawn between absolute figures (or spatially extensive data), proportions (or spatially intensive data) and measurements which do not correspond to either of these two categories. The relationship between the areal unit and the variable of interest may be either "summary" (e.g., the population of the areal unit, since the variable summarizes some aspect of the areal unit) or "functional" (e.g., political representation of an electoral district, since the variable is related to a function of the areal unit).

By combining these three dimensions, the authors were able to produce a typology of areal interpolation problems. Different areal interpolation methods are shown to be suitable for each component of the typology. For example, the direct overlay method is appropriate for natural areal units, since these units are defined by the spatial distribution of the variable of interest and hence each areal unit should be relatively homogeneous. For summary-count data for imposed areal units, one might apportion the count for the source zone into counts for each of the target zones overlapping the source zone. This might be achieved by weighting the count by the area of each overlapping target zone portion (i.e., case 1 of Markoff & Shapiro [156]). Information on other, related variables might usefully be employed to improve the accuracy of this procedure (i.e., case 3 of Markoff & Shapiro [156]). For functionally related data, a different type of problem exists because the variable of interest, while

relevant to the source zones, may not apply directly to the target zones. The authors argue that for certain components of the typology, areal interpolation is not logically justifiable.

The geographical aspects of areal interpolation derive from the configuration and size of the source and target zones. Areal interpolation is especially problematic when the source zones are large relative to the target zones (i.e., disaggregation). The size of the source and target zones also complicates the homogeneity assumption of the direct overlay method. An alternate approach might rely on a dasymetric method in which the distribution of some limiting variable might be used to model spatial heterogeneity in the underlying distribution and thus facilitate accurate estimation of target zone values. For example, the percentage of cropland in a target zone might be more accurately estimated given knowledge of the spatial distribution of altitude, slope, rainfall, soil type, or some other variable limiting the location of cropland.

Application of the dasymetric method as a means of improving the accuracy of target zone estimates has been further explored by Flowerdew [84]. This approach uses a binary limiting variable whose spatial distribution is known for each target zone. This variable may be entered into a Poisson regression model of the form

$$\lambda = \lambda_1 A_1 + \lambda_2 A_2 \quad (5.13)$$

Variables A_1 and A_2 represent the area of each source zone covered by categories 1 and 2 of the limiting variable. The dependent variable in this model is the variable of interest for each source zone.

In an illustrative example presented by the author, the limiting variable was land use (grassland vs. woodland) and the variable of interest was population. Thus population estimates were desired for target zones for which land use was known. The regression model was calibrated using area and population data to yield estimates of parameters λ_1 and λ_2 . These estimates represent the expected population per unit area of grassland and woodland respectively. Since population density tends to be higher for grassland than woodland, $\lambda_1 > \lambda_2$. The estimated population for the portion of each source zone covered by each land use category was calculated by multiplying the parameter estimates for each category (i.e., λ_1 and λ_2) by the area covered by that category within each source zone. To preserve the pycnophylactic constraint (i.e., the sum of the estimated populations for the portions of each source zone covered by each land use category must equal the actual population of the target zone), the estimated populations were scaled by the ratio of the actual populations to the fitted populations derived from the regression model. The resulting values were summed so as to yield population estimates for each target zone.

This approach was also applied to a set of real data in an effort to estimate the population for a set of parliamentary constituencies (target zones) using population data derived from administrative districts (source zones). The limiting variable, political affiliation (Labor vs. Conservative), was known for each target zone. Despite the poor fit of the regression model, the approach yielded substantial improvements in the accuracy of target zone population estimates compared to the direct overlay method.

5.2. Vector to Raster conversion †

The vector and raster models represent two ways in which spatial data may be encoded in the construction of a spatial database. The vector model represents spatial data as a set of objects. These objects, whether points, lines or areas, are encoded by recording their positions in space. Thus points may be defined as a single x,y-coordinate pair. Lines may be encoded as strings of x,y-coordinate pairs joined by straight line segments. Polygons are composed of strings of x,y-coordinate pairs which, when joined

† The discussion in this section derives largely from Veregin [216].

by straight line segments, define a closed geometric figure.

In the raster model spatial data are represented by the values assigned to a matrix of grid cells that cover the region of interest. Typically these cells are regular, non-overlapping and spatially exhaustive, but they need not always be square. Spatial objects may be encoded as the set of cells that demarcate the locations of the objects in space. Points may be encoded as single cells, while lines and polygon boundaries may be defined as the set of adjacent cells that most closely approximate these linear features. Clearly the degree of approximation is a function of the linear dimension of the cell, or its spatial resolution. In this case cells may take one of two possible values, indicating whether or not the cell is located on a point, line or polygon boundary. If the raster model is extended to allow cells to take more than two values, then the spatial distribution of some variable of interest may be encoded as a set of values associated with the cells. These values may be categorical, as in the case of land use, or numerical, as in the case of remotely sensed spectral data.

Vector to raster conversion refers to the operation whereby spatial data encoded under the vector model are transformed into a grid of cells. Scanning digitizers, for example, assign a binary value to each cell that indicates whether or not the cell falls on some object of interest on the source map. Vector to raster conversion may also be carried out to produce a grid in which the cells take more than two values. By accessing the attribute values corresponding to a set of polygons, a raster data set may be produced in which each cell is assigned the value of the polygon in which it is located. This would be appropriate for categorical data, such as land use, and for proportions, such as population density. For other types of data, such as population counts, allowance must be made for the different areas of the polygons and cells, and the count must be appropriately apportioned to the cells. Allowance must also be made for cells that fall on a polygon boundary and thus might be reasonably assigned to either polygon.

Vector to raster conversion may be performed for a variety of purposes. The speed of scanning digitizers over non-mechanical digitizing methods explains the rationale for performing vector to raster conversion in this case. Raster data are also often easier to display and analyze, and they may facilitate data compression and computational speed. When vector and raster databases must be combined to facilitate data analysis, vector to raster conversion may be applied to yield a common base for the analysis. The opposite operation, raster to vector conversion, would seem to be less commonly applied, due in part to the difficulties of deriving meaningful x,y-coordinate values from the step-like raster representation of spatial objects. In effect, vector to raster conversion results in a loss of precision because the spatial resolution of raster data is typically coarser than that of vector data. In many applications, however, this loss of precision may be less deleterious than the creation of spurious representations of spatial objects.

In discussing the errors associated with vector to raster conversion, it is important to keep in mind that the conversion may be performed to rasterize cartographic detail (e.g., digitizing source maps by scanner) or to produce a raster representation of some thematic attribute distributed over space (e.g., assigning to each cell in the grid the value of the attribute for the polygon in which the cell is located). This distinction facilitates a broad classification of error models for vector to raster conversion as models of positional error and models of attribute error. This classification also reflects the fact that in many cases cartographic errors may be treated independently of thematic errors. Independence exists when the thematic attribute has been measured for areal units that have been delineated solely for enumerative purposes (e.g., census data). However, for other types of data known as "categorical coverages" (Chrisman [49]), independence cannot be assumed because the locations of polygon boundaries are determined by the values of the thematic attribute themselves (see § 2.2).

Models of positional error tend to focus on inaccuracies in polygon area arising from vector to raster conversion. These inaccuracies result because vector to raster conversion tends to cause shifts in the location of polygon boundaries. Models of attribute error, in contrast, focus on the accuracy of the raster representation of the spatial distribution of the thematic attribute. Accuracy is affected by the

often arbitrary nature of polygon boundaries which may mask the underlying spatial distribution of the attribute. Models of error for vector to raster conversion are reviewed by Burrough [34], Goodchild [97], Muller [168] and Veregin [213, 216].

5.2.1. Models of Positional Error

Frolov & Maling [69] examined changes in polygon area associated with vector to raster conversion. Their analysis derives from a desire to model the error associated with cartometric estimates of polygon area (see § 3.2.1). The model developed by Frolov & Maling is based on the identification of “boundary cells,” or cells that are bisected by a polygon boundary. The authors assume that the linear dimension of the cells is sufficiently large to allow the boundary segment passing through each cell to be represented as a straight line segment without any appreciable loss of precision. Thus each boundary cell is bisected by a straight line segment that subdivides the cell into two, usually unequal, parts. The authors contend that the set of all possible locations of this line segment can be determined from the set of all possible values of two variables, x and a , which represent the point and angle at which the line segment intersect one edge of the cell. Given a cell of unit dimension, x may vary from 0 to 1 and a may vary from 0 to π . By integrating over these two ranges, the authors computed \bar{w} , the mean area of the smaller subdivision of the cell. The standard error, s , of the estimate of polygon area is then given as

$$s = \bar{w} n^{1/2} \quad (5.14)$$

An estimate for n , the number of boundary cells for the polygon, is given by

$$n = k_2 B \quad (5.15)$$

An estimate for B , the estimated boundary length, is given by

$$B = \sum_{i=1}^n b_i \quad (5.16)$$

where: b_i = the length of the straight line segment in boundary cell i .

Equations (5.15) and (5.16) may be combined to yield

$$k_2 = 1 / \bar{b} \quad (5.17)$$

where: \bar{b} = the mean length of the straight line segments in the boundary cells (computed in a manner analogous to \bar{w}).

The value of B in equation (5.15) may be computed as

$$B = k_3 A^{1/2} \quad (5.18)$$

where: A = polygon area; and

k_3 = a measure of polygon shape.

Combining the above expressions, the standard error of the estimate of polygon area is

$$s = k_1 k_2^{1/2} k_3^{1/2} A^{1/4} \quad (5.19)$$

(For consistency in notation, k_1 has replaced \bar{w} in this equation.) The relative standard error, or the standard error as a percentage of polygon area, is obtained by dividing through by A to yield

$$s = k_1 k_2^{1/2} k_3^{1/2} A^{-3/4} \quad (5.20)$$

These results indicate that the relative standard error is a function of polygon size. Relative error declines as polygon size increases. Empirical tests conducted by the authors show close agreement between actual and theoretical levels of error.

Lloyd [144] describes an alternate interpretation of the set of all possible locations of line segments passing through boundary cells. Rather than a value of x varying from 0 to 1 and a value of y varying from 0 to π , Lloyd considered the rotation through angle π of a set of parallel, infinitely close lines. This approach yields values of 0.8621 and 0.2469 for \bar{b} and \bar{w} , respectively. (Frolov & Maling obtained values of 0.7979 and 0.2127.) Lloyd's results change slightly the values of the parameters in equations (5.19) and (5.20) but do not alter the overall meaning of these equations or the significance of the results.

Crapper [63] and Crapper *et al* [64] focus on the estimation of n , the number of boundary cells. The number of boundary cells depends on polygon shape and the degree of boundary contortion. The polygon shape factor, k_3 , is defined as

$$k_3 = \frac{P}{2 \pi^{1/2} A^{1/2}} \quad (5.21)$$

where: P = the actual perimeter of the polygon; and
 A = the actual area of the polygon.

The mean within-cell contortion parameter, k_4 , is given by

$$k_4 = \frac{\frac{1}{n} \sum_{i=1}^n p_i}{\bar{b}} \quad (5.22)$$

where: p_i = the actual length of the boundary segment in boundary cell i ; and
 \bar{b} = the mean length of the straight line segment in boundary cells (as defined above).

Examination of polygons with homogeneous biophysical characteristics by Crapper [61] shows that, on average, $k_3 = 1.82$. The value of k_4 may often be assumed to be 1.0, which is equivalent to the assumption of Frolov & Maling that the boundary segment passing through any boundary cell can be approximated by a straight line segment. Given these values, the number of boundary cells is given as

$$n = \frac{2 k_3 \pi^{1/2} A^{1/2}}{\bar{b}} \quad (5.23)$$

Crappier [60, 62] extends this model to the case of rectangular cells, which are commonly encountered in remote sensing. The standard error of the estimated area of a polygon in this case is given by

$$s = \left(\frac{2 k_1^2 k_3 \pi^{1/2} A^{1/2}}{k_4 \bar{b}} \right)^{1/2} \quad (5.24)$$

In this equation, $k_1 = \bar{w}$ as defined by Frolov & Maling, and all other parameters are defined above. For Landsat pixels, the author gives values of 1.82 for k_3 and 1.0 for k_4 . Values for k_1 and \bar{b} were computed in the manner proposed by Frolov & Maling, but refer to a rectangular cell with dimensions equal to a Landsat pixel. The resulting equation for the relative standard error is

$$s_r = 0.39 A^{-3/4} \quad (5.25)$$

The preceding models show that the relative standard error is proportional to the $-3/4$ power of polygon area, A . When the linear dimension of the cell, c , is not equal to 1, s_r is also proportional to the $3/2$ power of c (Goodchild [97]). This relationship holds when considering individual polygons, but not ensembles of polygons or entire maps. In the latter case, the relative standard error is linearly related to c and proportional to the $-1/2$ power of A (Muller [168]).

Goodchild [96] argues that the model developed by Frolov & Maling assumes serial independence in the errors associated with each boundary cell. If serial correlation is present, however, its effect will be to increase the relative standard error. Serial correlation will tend to be highest when the boundary segments within boundary cells are closely approximated by straight lines (i.e., the boundary is relatively smooth). Goodchild defines a measure of polygon boundary smoothness, β , as

$$\beta = \log(n + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n r_{ij}) / \log(n) \quad (5.26)$$

where: r_{ij} = the correlation between the errors associated with boundary cells i and j ; and
 n = the number of boundary cells.

Serial independence (i.e., the Frolov & Maling model) corresponds to a case in which $\beta = 1$ and perfect correlation to a case in which $\beta = 2$. The resulting general form of the equation for relative standard error is

$$s_r = k_1 k_2^{\beta/2} k_3^{\beta/2} A^{(\beta/4)-1} c^{2-(\beta/2)} \quad (5.27)$$

where: A = the actual area of the polygon; and
 c = the linear dimension of the cell.

(Parameters k_1 , k_2 and k_3 are defined above. See equations (5.14), (5.17) and (5.18).)

5.2.2. Models of Attribute Error

In contrast to error models for vector to raster conversion that focus on positional error, those models that focus on attribute error are concerned with inaccuracies in the spatial representation of the thematic attribute on the raster map. Such inaccuracies may arise from the often arbitrary nature of polygon boundaries, which mask the underlying spatial distribution of the thematic attribute. The manner in which values of the attribute are assigned to cells therefore becomes an important determinant of the accuracy of vector to raster conversion.

Many of the models of relevance in this discussion are likewise important for other GIS operations. For example, the models developed for areal interpolation (see § 5.1) may also be applied to vector to raster conversion if the target zones are grid cells. The model described by Goodchild & Lam [101] and Lam [136, 137] shows that errors may result if cells are assigned the value of the polygon in which they are encompassed, since this procedure makes no allowance for heterogeneity in the attribute value within any polygon.

Several researchers have developed error models that focus explicitly on vector to raster conversion. Switzer [203] presents a model of error in polygon area for entire maps based on errors of omission and commission in the assignment of cells to polygons. It is assumed that a cell is assigned to a polygon based on the location of the center point of the cell. The model is based on map complexity defined in terms of the differences in "subdivision membership" (i.e., attribute values) for cells separated by a specified distance. In general, larger errors are expected as map complexity rises. According to Switzer, the area A_{st}^* that belongs to subdivision s on the true map, but has been assigned to subdivision t , is given by the intersection of subdivision s on the true map and the collection of cells that have been assigned to subdivision t . Thus the total misclassified area, A^* , is given as

$$A^* = \sum_{s=1}^S \sum_{t=1}^T A_{st}^* \quad (5.28)$$

where: S = the number of subdivisions on the true map; and

T = the number of subdivisions on the raster map.

The operation $P_{st}(d)$ is then defined as the probability that a random point is in subdivision s on the true map and a given cell center point is in subdivision t on the true map, when the two points are separated by distance d . Switzer proposed that $P_{st}(d)$ be approximated as a Taylor series expansion of its derivatives with respect to d (Goodchild [97]). The first two terms may be estimated from the frequency with which cells separated by distances of d_1 and d_2 on the raster map fall into subdivisions s and t . For square cells, d_1 and d_2 are equal to $N^{-1/2}$ and $2N^{-1/2}$ respectively, where N is the number of cells in the entire map.

$P_{st}(d)$ is estimated as

$$P_{st}(d_k) = \frac{\sum_{i=1}^N \sum_{j=1}^N f_s(c_i) f_t(c_j) g(c_i, c_j, d_k)}{\sum_{i=1}^N \sum_{j=1}^N g(c_i, c_j, d_k)} \quad (5.29)$$

where: $f_s(c_i)$ = an indicator function equal to 1 when cell c_i is in subdivision s and 0 otherwise;
 $f_t(c_j)$ = an indicator function equal to 1 when cell c_j is in subdivision t and 0 otherwise; and
 $g(c_i, c_j, d_k)$ = a proximity function equal to 1 when the distance between cells c_i and c_j is equal to d_k and 0 otherwise.

This equation is evaluated twice, once for $k=1$ ($d_k = d_1 = N^{-1/2}$) and once for $k=2$ ($d_k = d_2 = 2N^{-1/2}$). The denominator of the equation is the number of cell pairs on the map that are separated by distance d_k . The numerator is the number of cells pairs separated by distance d_k for which the subdivisions are s and t .

The misclassified area is then given as

$$A_{st}^* = a_1 P_{st}(d_1) - a_2 P_{st}(d_2) \quad (5.30)$$

where $a_1 = 0.76$ and $a_2 = 0.19$. (Goodchild [97] gives values of 0.6 and 0.11.) The values obtained from equation (5.29) for a given pair of subdivisions s and t are inserted into equation (5.30) and the values of A_{st}^* are then summed for all s and t , as in equation (5.28), to yield the total misclassified area on the map. Empirical work by Muller [168] shows that this model accurately predicts the observed area of misclassification.

Wehde [224, 225] presents an error model for vector to raster conversion in which error is defined as the incorrect assignment of thematic attribute values at sampled points on the map. The model is based on the notion that the accuracy of the raster map is dependent on the frequency distribution of distances between points on the boundaries of the polygons on the source map. The "interboundary distance" frequency distribution is defined as the distribution of distances between boundaries for a set of randomly-placed lines. For raster maps, the positions of these lines are controlled by the grid matrix and the distribution is referred to as the "span" distribution. In this case interboundary distances are multiples of the linear cell dimension. As cell size increases, the frequency distribution becomes more skewed, with small interboundary distances (defined in terms of the number of cells) predominating. In this form the model describes aggregation, in which cells on the source map are combined to produce cells on the target map. However, it is applicable to vector to raster conversion by replacing the span distribution with the interboundary distance distribution.

The error model is expressed as

$$E = FG \quad (5.31)$$

where: E = a $1 \times t$ matrix where e_j is the predicted mapping error as a function of target cell size j ;
 F = a $1 \times s$ matrix where f_i is the relative frequency of span size i on the map being tested; and
 G = a $s \times t$ matrix where g_{ij} is the estimate of the mapping error for target cell size j and span size i .

The value of t is the number of different target cell sizes being tested and the value of s is the number of possible span sizes for the map. This model can be used to predict the mapping error that would result in a particular target map given a specified target cell size and the particular span distribution on the source map.

In order to estimate G the author enumerated the possible positions of a target cell over different span sizes to derive the probability of incorrectly assigning the target cell. When target cell size is a multiple of source cell size, there is a finite number of such positions. The author enumerated possible positions in only one dimension by placing a line with a length equal to the linear dimension of a given target cell over sets of source cells with different span sizes. In an empirical test, the model predicted the correct trend in incorrect assignments, but overestimated the actual error by a factor of two.

Clarke [51] employed an experimental approach in which a set of five different spatial distributions of a thematic attribute, each mapped onto six different sets of polygons which varied in size, shape and degree of convolution, were rasterized according to four different vector to raster conversion methods. These methods included inverse distance weighting of the five nearest polygon centroids to the cell (using weights of d^{-2} and d^{-4} , where d is the distance), trend-fitting in which a polynomial surface was fitted to the spatial distribution and inverse distance weighting then applied to the estimated surface, pycnophylactic interpolation (based on Tobler, as cited in Goodchild & Lam [101]) and spatial smoothing (based on Hsu [121]).

The author constructed thirty polygon maps, each of which was then rasterized using each of the four methods. Since the polygon maps were based on a known spatial distribution of the thematic attribute, it was possible to measure the degree of error associated with each raster representation. The author posited an error model based on the size and shape of the polygons, the complexity of the spatial distribution and the vector to raster conversion method employed. Fourteen variables related to the first three factors were employed in multiple regression analysis in which the dependent variable was the absolute error between the actual spatial distribution and the raster representation of this distribution. The regression models, stratified by vector to raster conversion method, exhibited high explanatory power (R^2 values ranged from 0.78 to 0.98). When variables related to the complexity of the spatial distribution were excluded, explanatory power dropped significantly (R^2 values ranged from 0.16 to 0.43).

5.3. Map Overlay

In the map overlay operation, two or more data layers are superimposed to produce a composite map depicting combinations of the thematic attributes on each data layer. For numerical attributes, attribute values can be combined by arithmetic operators. Boolean operators are employed for categorical attributes. If the data are in vector format the mapping units are polygons, and if the data are in raster format they are single cells. Raster data may also be used to delineate polygonal mapping units as sets of contiguous cells whose attribute values are the same. Only in the raster case are the boundaries of the mapping units on different data layers likely to coincide perfectly. Moreover, this will occur only when the data layers are registered to the same grid. In the vector case, the mapping units or polygons on different data layers may share common boundary segments, but perfect boundary coincidence is unlikely to occur.

The technical details of the map overlay operation are therefore dependent on the nature of the input data. A straightforward application of map overlay in remote sensing is the calculation of band ratios for a set of pixels. Similarly it is a relatively simple task to apply Boolean operators once remotely sensed spectral data have been classified into cover types. For vector data, in contrast, map overlay involves the identification of all points at which the boundaries of the polygons on each data layer intersect. Polygons on the composite map are then delineated as sets of boundary segments which, when chained together, create closed geometric figures.

Error models for the map overlay operation may be differentiated according to their emphasis on positional and attribute error. As noted previously in § 5.2, these two error sources may often be assumed to be independent when polygon boundaries have been imposed *a priori* for enumerative purposes. If attribute and positional error may be treated independently, then both error components contribute to total composite map error. That is, composite map error is a function of error in the thematic attribute values attached to the polygons on each data layer, as well as error in the positions of polygon boundaries. The latter source of error has received considerably more attention in the literature. Errors in polygon boundaries may be attributed largely to digitizing and generalization error (see § 4.1 and § 4.2). Since map overlay involves the superimposition of two or more data layers, positional error is also associated with differences in the scale, age or map projection of the data layers. In the context of map overlay, these sources of error result in discrepancies in the cartographic representations of polygon boundaries such that these boundaries may not coincide on different data layers even if they happen to coincide in reality. The resulting “gaps” or “slivers” that appear on the composite map are often referred to as “spurious polygons” since they do not correspond to any observable or readily agreed-upon feature of the real world.

Thematic and cartographic error cannot be so easily separated when the positions of features on a given data layer are defined by the values of the thematic attribute. For such categorical coverages (see § 2.2), positional error, to the extent that it may be said to exist, is not separable from attribute error. Models for attribute error therefore tend to emphasize the manner in which error in thematic attribute values on individual data layers are propagated through the map overlay operation.

The following discussion first focuses on factors affecting the level of error associated with map overlay. Models of positional error are then described with reference to the spurious polygon problem. Finally, models of attribute error are discussed that focus on the propagation of errors in thematic attribute values through the map overlay operation. Reviews of many of these models are provided in Burrough [34] and Veregin [213, 214, 215].

5.3.1. Factors Affecting Accuracy

MacDougall [147] argues that composite map accuracy is a function of three parameters — “horizontal” or positional error in polygon boundaries, polygon “purity” or attribute error, and error introduced by the operation itself (error propagation). Horizontal error is seen to arise, not only from imprecisions associated with drafting and digitizing, but the prevalence of indeterminate polygon boundaries for such attributes as soil and vegetation types. Polygon purity refers to the internal homogeneity of a polygon and is calculated as the proportion of the total area of the polygon that is correctly classified. Errors associated with map overlay itself are seen to interact with these two sources of error. Note that since the author is concerned primarily with thematic attributes such as soils and vegetation, his assumption that horizontal error is distinguishable from polygon purity is difficult to justify, since the polygon boundaries are defined by the values of the thematic attribute (Chrisman [48]).

The total horizontal error for a given data layer i , H_i , can be estimated by

$$H_i = h_i t_i / a_i \quad (5.32)$$

where: h_i = a measure of horizontal error for data layer i ;
 t_i = the total length of all polygon boundaries on data layer i ; and
 a_i = the total area of data layer i .

This equation yields an estimate of the proportion of the total area of data layer i that is uncertain or unreliable. If the polygon boundaries are known to have different levels of horizontal error, then equation (5.32) may be replaced by

$$H_i = \frac{1}{a_i} \sum_{j=1}^{m_i} h_{ij} t_{ij} \quad (5.33)$$

where: h_{ij} = a measure of horizontal error for polygon boundary j on data layer i ;
 t_{ij} = the total length of polygon boundary j on data layer i ; and
 m_i = the number of polygon boundaries on data layer i .

According to MacDougall, the lower limit of composite map accuracy, A_{\min} , is given by

$$A_{\min} = f\left(\sum_{i=1}^n H_i, \prod_{i=1}^n P_i, \epsilon\right) \quad (5.34)$$

where: P_i = the purity of data layer i ;
 n = the number of data layers; and
 ϵ = the error introduced by the map overlay operation.

This equation represents a case in which the errors on each data layer are independent. MacDougall argues that when errors are correlated, the horizontal error of the composite map will approach the mean horizontal error of data layers 1 through n , \bar{H} , while the purity of the composite map will be approximated by the purity of the least pure data layer, P_{\min} . Thus the upper limit to composite map accuracy, A_{\max} , is given as

$$A_{\max} = f(\bar{H}, P_{\min}, \epsilon) \quad (5.35)$$

Equation (5.35) corresponds to a case in which the areas of horizontal error and impurity tend to coincide spatially. Hence it would be inappropriate to sum horizontal errors, or multiply purities (since the area of impurity on each data layer is contained within the area of impurity on the least pure data layer). Sources of error arising from the map overlay operation itself (i.e., ϵ in equations (5.34) and (5.35)) are not dealt with extensively by MacDougall. These errors are seen to arise primarily from scale and map projection differences between data layers, as well as the generation of spurious polygons due to discrepancies in the cartographic representation of the same polygon boundary segment on different data layers.

Chrisman [48] argues that MacDougall's analysis of error is pessimistic, because errors associated with map overlay can be substantially lower than MacDougall's model would suggest. Chrisman re-examines MacDougall's analysis in light of advances in map overlay techniques and recent empirical findings. Empirical tests, for example, demonstrate that positional error in polygon boundaries may be much lower than expected, since data often exceed the minimum accuracy standard reported in map legends. Testing can also be performed to establish site-specific estimates of positional error for indeterminate boundaries. These boundaries are often artifacts of imprecise attributes, rather than imprecise boundary positions as MacDougall suggests. Fuzzy boundaries also acquire a sharper character if they obtain legal status (i.e., planning or zoning). In certain applications, this may be equivalent to an actual increase in precision. MacDougall's method of estimating horizontal error (i.e., equations (5.32) and (5.33)) also ignores the effects of line curvature. Without such an adjustment, lines will seem to increase in length with more detailed measurement, resulting in an increase in the estimate of error (see § 3.1).

As in the case of positional accuracy, polygon purity may also be much higher than the minimum accuracy standard reported in the map legend. MacDougall's analysis ignores the distinction between "identification" errors (i.e., errors in assigning the correct attribute value) and "discrimination" errors (i.e., errors in separating adjacent values). The latter type of error is essentially synonymous with the indeterminate boundary problem and, despite MacDougall's separation of positional and attribute error, is impossible to distinguish from positional error in polygon boundaries. In defining polygon purity as the proportion of the polygon that is correctly classified, MacDougall also ignores the fact that certain misclassifications are less serious than others. Chrisman argues that in the context of a classification error matrix approach (see § 2.1), the entire matrix, and not just the main diagonal, should be considered. Moreover, since maps are often constructed using rules of minimum polygon size, impurity is often intentional and designed to ensure effective cartographic communication (see § 4.6). These impurities are often conscious choices tied to the concepts of scale and generalization.

Errors associated with the map overlay operation itself are also capable of being managed with modern computational approaches to map overlay. Examples include the merging of data from diverse sources to select the most accurate cartographic detail, the identification of "integrated terrain units" (ITUs) as basic mapping units for map overlay and the use of an error tolerance in map overlay algorithms to reduce the incidence of spurious polygons. Chrisman also argues that MacDougall's error model is unrealistic in its treatment of horizontal error and polygon purity. In particular, the model allows the total horizontal error to exceed the total area of the composite map. The estimate of composite map purity ignores correlations between data layers as it omits a covariance term. Finally, MacDougall does not incorporate the most common cause of composite map error — the lack of source maps of sufficient accuracy.

Chrisman [49, 50] describes an empirical approach to error modeling for categorical coverages. For categorical coverages, issues of positional and attribute error tend to interact, but established methods of accuracy assessment tend to treat these two error components as separable. Assessment of positional accuracy, for example, focuses on sets of well-defined points, thus ignoring the impact of attribute misclassification. Conversely, methods of assessing attribute accuracy, such as the classification error matrix approach adopted in remote sensing, fail to distinguish between positional and attribute components of error.

Chrisman argues that for such accuracy assessments to be more comprehensive, they must compare complete maps rather than just sets of sampled points. One approach might be to overlay two categorical coverages purporting to show the same phenomenon. This approach could be used to detect spurious polygons and artifacts of error which may seem perfectly distinct, but which in reality are often difficult to disentangle. Moreover, spurious polygons and classification error may be seen as extreme cases between which exist transitional forms of error that are neither purely positional nor purely attribute. These issues cannot be divorced from the effects of map scale, which impact both positional and attribute accuracy due to the distortions inherent in cartographic generalization. Total error may therefore be viewed as a composite of a set of stochastic processes operating simultaneously within the boundaries set by scale effects.

5.3.2. The Spurious Polygon Problem

Some of the issues raised by Chrisman [48] pertaining to positional error have been examined in greater detail by other authors. Roller [188] suggests that ancillary data (in this case provided by remote sensing) can be used to improve the accuracy of polygon boundary positions. The methodology is based on the identification of credible polygons within the polygons delineated in the original survey, and is conceptually equivalent to Chrisman's suggestion that data sources may be merged to acquire more accurate cartographic detail. This is referred to as "templating" by Dangermond [69]. Dangermond also suggests that ITUs might be identified as basic mapping units for map overlay. ITUs possess multiple attributes derived from several data layers and as such they avoid the problem of alternate representations of the same boundary line on different data layers. Chrisman [45] and Dougenik [74] describe the

implementation of a map overlay algorithm that incorporates an error tolerance band in defining boundary intersection points. This tolerance defines the maximum distance that any point on a polygon boundary is permitted to move. By allowing the algorithm to move points, spurious polygons can be removed if they are narrower than the defined tolerance. The algorithm is described in detail by White [226].

This error tolerance model derives largely from the epsilon band concept described by Blakemore [27], Chrisman [45], Honeycutt [118] and others (see § 4.1.1). The epsilon band concept suggests that for any cartographic line a buffer zone having a width of twice epsilon (ϵ) can be constructed around the line. This zone may be delineated by joining the center points of a circle of radius ϵ as it is "rolled" along both sides of the line (or in the case of polygons, along the inner and outer edges of the polygon boundary). The primary role of the epsilon band concept in error modeling for map overlay is to describe a zone of uncertainty in the position of a given polygon boundary. For example, if polygon boundaries have been digitized according to a specified level of accuracy, such that the digitized boundary cannot deviate more than x mm from the boundary on the source map, then $\epsilon = x$ and the boundary on the source map will fall somewhere within the zone of width 2ϵ surrounding the digitized boundary.

The epsilon band concept and its variants may also be applied in map overlay to identify spurious polygons (Honeycutt [118]). If the maximum width of a given composite map polygon is less than 2ϵ , then the polygon is likely to be spurious since it is wholly within the zone of uncertainty in polygon boundary location. The epsilon band is a boxcar distribution in which the probability of error is equal at all distances up to ϵ on either side of the boundary and drops to zero at distances greater than ϵ . If the distribution of error around boundaries is known to follow some other probability distribution, then this distribution can be used in place of the epsilon band model (Hudson [124] and Maffini [153]). An example is the bimodal distribution of error around digitized lines observed by Honeycutt [118] (see § 4.1). The distributions associated with all lines are superimposed and the combined probability that a given point is actually contained within a mapped polygon is computed as a function of the individual probability distributions. The combined probability may be computed for a sample of points within a given polygon and if this probability is consistently below some threshold, then the polygon may be flagged as spurious. Honeycutt suggests a number of methods for combining probabilities, including the mean probability and the probability associated with the line closest to the point.

Peucker [178] describes another variant of the epsilon band concept that may be applied to map overlay. This variant is based on the notion that any cartographic line has a corresponding band defined by the maximum perpendicular deviation of the line from its general trend (see § 4.1). In map overlay, this model may be applied to reduce the computations required to identify points at which polygon boundaries intersect, since the intersection point for any two lines must be located within the parallelogram defined by the intersection of the two corresponding bands. Since Peucker's model also has implications for cartographic generalization, it might also be applied in map overlay to reduce the incidence of spurious polygons. Any line may be generalized by partitioning it into subsets until each subset has a band with a width less than or equal to a specified threshold value. At each step, partitioning is performed by selecting, as the subset end points, those points that define the maximum extent of the line perpendicular to its general direction. The original line is then replaced by the lines defining the general direction of each subset. In map overlay, this procedure might be applied to the cartographic lines defining the boundaries of each composite map polygon, with the aim of eliminating spurious polygons caused by slight discrepancies in the representations of the same feature on different data layers.

The spurious polygon problem has also been approached from the viewpoint that these polygons tend to have relatively small areas. Thus composite map polygons might be flagged as spurious if their area is below a specified threshold value. McAlpine & Cook (cited in Burrough [34]) found that the number of composite map polygons, m_c , could be reasonably well predicted by the equation

$$m_c = \left(\sum_{i=1}^n m_i^{1/2} \right)^2 \quad (5.36)$$

where: m_i = the number of polygons on data layer i ; and
 n = the number of data layers.

(Note that the exponent 1/2 is missing in Burrough [34].) Thus the number of composite map polygons tends to rise exponentially as the number of data layers increases. Empirical tests indicate that a large percentage of these composite map polygons tend to be small and, by implication, spurious as well.

Cook [56] further explored the implications of composite map polygon size for the identification of spurious polygons. According to the author, the reliability of a given polygon on any data layer will in general decline as the area of the polygon decreases. The magnitude of this decline, however, is a function of the thematic attribute portrayed on the data layer. This gives rise to the concept of a "size-probability function" (SPF), which describes the relationship between polygon reliability and polygon area. The SPF may be estimated for each of the data layers involved in map overlay, and the joint SPF for the composite map may be computed as the product of the SPFs of all individual data layers (assuming independence between data layers). Given an acceptable minimum probability of reliability, it is then possible to flag those composite map polygons whose area, according to the joint SPF, corresponds to a probability below this acceptable minimum. Once flagged, these spurious polygons presumably could be deleted.

An alternate approach to the identification of spurious polygons is given by Goodchild [95, 97]. The author maintains that the number of spurious polygons on the composite map is a function of the number of polygon vertices on each data layer rather than the number of polygons. Goodchild's approach is based on the identification of points of intersection between the "true" cartographic line and the digitized representations of this line on different data layers. Each point of intersection along the true line is assigned a value of either 1 or 2, corresponding to the data layer containing the representation of the line at that point. Spurious polygons are then identified by certain sequences of 1s and 2s.

Based on runs of binary symbols, the number of spurious polygons that can occur on the composite map ranges from a minimum of 0 to a maximum of $2 \min(v_1, v_2) - 4$, where v_i is the number of vertices on data layer i . In the random case, the expected number of spurious polygons, $E[s]$, is given by

$$E[s] = \frac{2 v_1 v_2}{v_1 + v_2} - 3 \quad (5.37)$$

As this equation indicates, the number of spurious polygons will tend to be higher for data layers with a high degree of cartographic detail. Simulations performed by Goodchild show that, on average, equation (5.37) overestimates the observed number of spurious polygons by approximately 20 percent. The number of spurious polygons tend to be only one-half of the theoretical maximum.

5.3.3. Models of Attribute Error

Newcomer & Szajgin [172] present a model for categorical raster data based on the attribute accuracy of each data layer. The accuracy of data layer i , $P[\bar{E}_i]$, is defined as the proportion of cells on data layer i that are correctly classified. If two data layers are involved, the accuracy of the composite map, $P[\bar{E}_c]$, is given by

$$\begin{aligned} P[\bar{E}_c] &= P[\bar{E}_1 \cap \bar{E}_2] \\ &= P[\bar{E}_1] P[\bar{E}_2 | \bar{E}_1] \end{aligned} \quad (5.38)$$

This is the intersection of the correctly classified cells on the two data layers. That is, a given cell may be misclassified on more than one data layer, but it need be misclassified on only *one* of the data layers to be misclassified on the composite map. Equation (5.38) bears some similarity to equation (5.34) but incorporates a covariance term in the form of the conditional probability, $P[\bar{E}_2 | \bar{E}_1]$. This term represents the proportion correctly classified cells on data layer 1 that are also correctly classified on data layer 2. However, as noted by Chrisman [48] and Veregin [215], this term cannot be computed from the classification error matrix, the standard method of accuracy assessment in remote sensing and other fields where raster data are employed.

Veregin [215] gives the general form of equation (5.38) when there are n data layers as

$$\begin{aligned} P[\bar{E}_c] &= P[\bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_n] \\ &= P[\bar{E}_1] P[\bar{E}_2 | \bar{E}_1] \prod_{i=1}^n P[\bar{E}_i | \Theta(\bar{E}_i)] \end{aligned} \quad (5.39)$$

where:

$$\Theta(\bar{E}_i) = \bar{E}_1 \cap \bar{E}_2 \cap \dots \cap \bar{E}_{i-1} \quad (5.40)$$

Minimum composite map accuracy, $P[\bar{E}_c^{\min}]$, is given by

$$P[\bar{E}_c^{\min}] = \max \left\{ 0, \left(1 - \sum_{i=1}^n P[E_i] \right) \right\} \quad (5.41)$$

where: $P[E_i]$ = the proportion of cells on data layer i that are misclassified.

This will occur when the misclassified cells on each data layer do not coincide spatially. Maximum composite map accuracy, $P[\bar{E}_c^{\max}]$, is given by

$$P[\bar{E}_c^{\max}] = \min \left\{ P[\bar{E}_i] \right\} \quad i=1,2,\dots,n \quad (5.42)$$

This will occur when the misclassified cells on all data layers coincide spatially with the misclassified cells on the least accurate data layer.

Equations (5.39) through (5.42) indicate that composite map accuracy will generally be lower than the accuracy of the individual data layers. As the number of data layers increases, composite map accuracy will initially decline sharply and then tend to level off. Moreover, composite map accuracy will tend to be higher if the misclassified cells on each data layer coincide spatially, in which case the conditional probabilities will tend towards 1.

Walsh *et al* [221] applied the Newcomer & Szajgin model to a set of land cover, slope-angle, slope-aspect and soil type data layers. "Inherent" error was measured for each data layer by field-checking a sample of cells and computing the proportion of cells that were correctly classified. In order to calculate the conditional probabilities for the model, the same cells were sampled on each data layer. Inherent error ranged from 43 to 83 percent. Map overlay was performed on various combinations of two or three data layers based on cell sizes of both 2.5 and 10.0 acres. Due to high levels of inherent error, minimum composite map accuracy (i.e., equation (5.41)) was often equal to 0, while maximum composite map accuracy (i.e., equation (5.42)) ranged from a low of 32 percent to a high of only 41 percent. Observed composite map accuracy (i.e., equation (5.39)) ranged from a low of 6 percent to a high of only 29 percent. "Operational" error was defined as the difference between the theoretical upper limit to composite map accuracy (i.e., equation (5.42)) and observed composite map accuracy (i.e., equation (5.39)). Operational error thus measures the degree to which misclassified cells on different data layers do not coincide spatially. Levels of operational error were observed to range from 12 to 27 percent.

Veregin [215] argues that the Newcomer & Szajgin model is applicable only to the Boolean AND operator. For this operator, a cell must be accurate on all data layers in order to be accurate on the composite map. For example, if two data layers depicting land cover at different dates are superimposed with the aim of producing a composite map showing land cover change over time, then the land cover class for a given cell must be accurate on both data layers. In contrast, if map overlay is applied to identify cells with a particular cover class on either of the two dates, then the requirement is only that the land cover class for a given cell be accurate on *one* of the data layers.

The latter example is consistent with the application of the Boolean OR operator, in which case composite map accuracy for n data layers is given by

$$\begin{aligned} P[\bar{E}_c] &= 1 - P[E_1 \cap E_2 \cap \dots \cap E_n] \\ &= 1 - P[E_1] P[E_2 | E_1] \prod_{i=1}^n P[E_i | \Theta(E_i)] \end{aligned} \quad (5.43)$$

where:

$$\Theta(E_i) = E_1 \cap E_2 \cap \dots \cap E_{i-1} \quad (5.44)$$

Thus composite map accuracy for the OR operator is defined in terms of the intersection of the misclassified cells on each data layer. Minimum composite map accuracy, $P[\bar{E}_c^{\min}]$, is given by

$$P[\bar{E}_c^{\min}] = \max \left\{ P[\bar{E}_i] \right\} \quad i=1,2,\dots,n \quad (5.45)$$

This will occur when the misclassified cells on the most accurate data layer are also misclassified on every other data layer. Maximum composite map accuracy, $P[\bar{E}_c^{\max}]$, is given by

$$P[\bar{E}_c^{\max}] = \min \left\{ 1, \left(\sum_{i=1}^n P[\bar{E}_i] \right) \right\} \quad (5.46)$$

This will occur when the misclassified cells on each data layer do not coincide spatially. Note that the situation that produces the lowest composite map accuracy for the AND operator produces the highest composite map accuracy for the OR operator.

Equations (5.43) through (5.46) indicate that for the OR operator, the accuracy of the composite map can never fall below that of the most accurate data layer. Moreover, as the number of data layers increases, composite map accuracy will tend to increase, rather than decline as for the AND operator. This may result in a composite map that is more accurate than any of the data layers from which it was constructed.

The impact of different Boolean operators on composite map accuracy is also illustrated by Robinson & Strahler [187] in the context of map overlay for “fuzzy” data (see § 6.1). Fuzziness or indeterminacy in attribute values may be represented by assigning a membership value between 0 and 1 to each cell on each data layer, reflecting the presumed validity of the attribute value for that cell. If the AND operator is applied, then the membership value for a cell on the composite map is defined as the minimum membership value of the corresponding cells on all data layers (see equation (6.14)). Equation (6.14) directly parallels equation (5.39). It represents a more general form of the Newcomer & Szajgin model in which cell accuracy is not confined to 0 (misclassified) or 1 (correctly classified), but may range anywhere between these two values. For the OR operator, the membership value for a cell on the composite map is defined as the maximum membership value of the corresponding cells on all data layers (see equation (6.15)). Equation (6.15) directly parallels equation (5.43).

In the Newcomer & Szajgin model, accuracy is defined as the proportion of cells correctly classified, and thus the model is appropriate only for categorical data. An analogous model may also be developed for numerical data, in which error is defined in terms of the deviations between actual and estimated cell values (Veregin [215]). This model, like that of Newcomer & Szajgin, is based on the accuracy of each data layer as well as a measure of covariance. The error covariance for any two data layers i and j , s_{ij} , is defined as

$$s_{ij} = \frac{1}{M} \sum_{m=1}^M (z_{mi} - \hat{z}_{mi})(z_{mj} - \hat{z}_{mj}) \quad (5.47)$$

where: z_{mi} = the actual value of cell m on data layer i ;

\hat{z}_{mi} = the estimated value of cell m on data layer i ;

z_{mj} = the actual value of cell m on data layer j ;

\hat{z}_{mj} = the estimated value of cell m on data layer j ; and

M = the number of cells on each data layer.

When $i = j$ the equation defines the error variance of a data layer. Based on this equation the error variance of the composite map can be computed as a function of the arithmetic operator applied in map overlay. For example, when n data layers are added, the error variance of the composite map, s_c^2 , is given by

$$s_c^2 = \sum_{i=1}^n \sum_{j=1}^n s_{ij} \quad (5.48)$$

Equations can also be derived for other arithmetic operators or sets of operators.

While most models of thematic error for map overlay focus on the propagation of error in thematic attribute values through the map overlay operation, it is also important to consider the error introduced by the operation itself. Even if the input data layers are of relatively high quality, the misapplication of map overlay may introduce error into the composite map. Some examples of the misapplication of map overlay have already been discussed, including the use of data layers of different dates, scales or projections. In the case of attribute error the misapplication of map overlay is primarily the result of applying arithmetic or Boolean operators to data that do not support the assumptions behind these operations.

Hopkins [119] compares different methods of computing the attribute values of composite map polygons and identifies the major difficulties and errors associated with each method. The author's analysis focuses on suitability mapping, in which data layers representing various dimensions of the landscape are superimposed to create a composite map in which each polygon is assigned a rating reflecting its suitability for some proposed activity. Error may be introduced in the construction of suitability maps when the data do not support the assumptions of the arithmetic operator applied to them.

In the "ordinal combination" method, polygons on data layers depicting different factors are rated according to their suitability for the proposed activity. The data layers are then superimposed to yield a composite map in which the rating for a given polygon is equal to the sum of the ratings for the polygons on the individual data layers whose intersection is the composite map polygon. Hopkins argues that this is an invalid arithmetic operation since suitability ratings are measured on an ordinal, not an interval, scale. Moreover the method implies that each data layer is independent, whereas suitability may be a nonlinear or multiplicative function of the factors portrayed on the data layers. The "linear combination" method is identical to the ordinal combination method except that, before they are summed, the suitability ratings on each data layer are multiplied by a weight reflecting the importance of the factor portrayed on the data layer. The resulting suitability ratings on the composite map are normalized by dividing by the sum of the weights. The effect of weighting is to change the unit of measurement for each factor such that all factors are on the same interval scale. According to the author, this method is valid arithmetically but still suffers from the assumption of independence.

Other methods of suitability mapping do not assume independence between data layers. These include the “nonlinear combination method” (in which data layers are combined according to known mathematical functions), the “factor combination method” (in which data layers depicting categorical factors are combined with Boolean operators and each unique combination of categorical values on the composite map is assigned a suitability rating), and cluster analysis (in which sites are grouped according to their similarity across factors). Alternatively, suitability mapping may be approached in terms of logical rules of combination. For example, “critical” data layers might be identified whose suitability rankings override those of other data layers in certain situations.

6. GENERAL ISSUES OF SPATIAL DATABASE ACCURACY

This section discusses a set of issues with general implications for assessing accuracy in spatial databases. The discussion in § 6.1 focuses on models of uncertainty, or methods of incorporating information about inexactness into operations applied to spatial data. Accuracy issues of importance to specific geographical models are briefly reviewed in § 6.2. General data quality issues pertaining to spatial databases are discussed in § 6.3.

6.1. Modeling Uncertainty

Many operations are applied to spatial data under the assumption that features, attributes and their relationships have been specified *a priori* in a precise and exact manner. However, as revealed by the studies reviewed in this report, this assumption is generally not justifiable, since inexactness is almost invariably present in spatial data. Inexactness exists in the positions of features and the assignment of attribute values (see § 2) and may be introduced at various stages of data compilation and database development (see § 4). Moreover, inexactness may be propagated through GIS operations to appear in modified form on tabular and graphic output products (see § 3 and § 5). Inexactness is often inadvertent, as in the case of measurement error or imprecision in taxonomic definitions, but may also be intentional since generalization methods are frequently applied to enhance cartographic fidelity (see § 4.2).

The discussion that follows focuses on methods of modeling uncertainty, or incorporating information about inexactness into operations applied to spatial data. Research has tended to focus on certain classes of operations, including map overlay, querying and classification. Models of uncertainty incorporate ideas from natural language processing, the value of information concept, non-monotonic logic and fuzzy set, evidential and probability theory. In contrast to many of the models discussed elsewhere in this report, models of uncertainty seek to redefine spatial operations to account for inherent inexactness in spatial data.

Stoms [201] reviews four models of uncertainty based on probability theory, Shafer's theory of evidence, fuzzy set theory and non-monotonic logic. The author describes how each model is appropriate for a different type of inexactness in spatial data. Inexactness is seen to arise primarily from three sources. "Randomness" may occur when an observation can assume a range of values. "Vagueness" may result from imprecision in taxonomic definitions. "Incompleteness of evidence" may occur when sampling has been applied, there are missing values, or surrogate variables have been employed.

The first model of uncertainty is based on probability theory and models uncertainty as the conditional probability that a hypothesis is true given some observation. That is,

$$P[H_i | X] = P[X | H_i] \frac{P[H_i]}{P[X]} \quad (6.1)$$

where: $P[H_i | X]$ = the conditional probability that hypothesis H_i is true given observation X ;
 $P[X | H_i]$ = the conditional probability of observing X given that hypothesis H_i is true;
 $P[H_i]$ = the probability that hypothesis H_i is true; and
 $P[X]$ = the probability of observing X .

As an example, in classifying remotely sensed data, H_i might signify the hypothesis that a given pixel belongs to class i , where X is the vector of responses for the pixel in different spectral bands. Inexactness is therefore reflected in the conditional probability $P[H_i | X]$, which indicates the degree to which one might assume the class to be correct given the vector of responses. As the author notes, this model assumes that probabilities can be assigned correctly based on prior knowledge, which is often untrue in an operational context. Unfortunately, the model provides no mechanism for weighting the assigned probabilities as a function of their reliability. This probabilistic model is appropriate primarily for

inexactness associated with randomness, since it focuses on inexactness in factual information. Maxim & Harrington [157] employ an analogous model in the analysis of classification error in remote sensing (see § 2.1).

The second model of uncertainty discussed by Stoms derives from Shafer's theory of evidence. In this model, the probability measure that hypothesis H_i is true is replaced by a measure of the probability that the available evidence supports the truth of the hypothesis. Evidential theory is based on the formulation of an "evidential interval" between a "belief function," which measures the degree to which the evidence supports the hypothesis, and a "plausibility function," which measures the degree to which the evidence fails to refute the hypothesis. The belief and plausibility functions represent, respectively, the lowest and highest degree of evidential support in favor of the hypothesis. Assume that set A represents the set of possible hypotheses and the associated probability that each hypothesis is true.

$$A = \{ P[H_1], P[H_2], P[H_3], \dots, P[H_h], \Theta \} \quad (6.2)$$

where: $P[H_i]$ = the probability that hypothesis H_i is true;

h = the number of hypotheses; and

Θ = the uncommitted or distributed support.

By definition,

$$\Theta + \sum_{i=1}^h P[H_i] = 1 \quad (6.3)$$

The belief function, $\text{Bel}[H_i]$, is simply the probability $P[H_i]$. The plausibility function, $\text{Pl}[H_i]$, is given by

$$\begin{aligned} \text{Pl}[H_i] &= 1 - \sum_{j=1}^h P[H_j] \quad j \neq i \\ &= \text{Bel}[H_i] + \Theta \end{aligned} \quad (6.4)$$

Note that when $\Theta = 0$, $\text{Pl}[H_i] = \text{Bel}[H_i] = P[H_i]$, which shows that evidential theory is a generalization of probability theory. The inferential interval, $[\text{Bel}[H_i], \text{Pl}[H_i]]$, represents the incompleteness of evidence for hypothesis H_i due to uncommitted support (i.e., Θ). Evidential theory is most appropriate for inexactness associated with incompleteness of evidence. Despite its pragmatic value, however, the model suffers from an inability to account for conflicting evidence and to choose between different hypotheses with similar levels of support.

The third model discussed by Stoms is fuzzy set theory. Fuzzy logic is based on the assignment of a "membership function," $\mu_A(X)$, which indicates the degree to which observation X belongs to set A . The membership function can be used to model the vagueness inherent in geographical concepts (e.g., "near"), such that a response to a query will be defined in terms of the degree to which the concept is satisfied for the specified locations. According to the author, the primary strength of fuzzy logic lies in its ability to handle inexactness associated with vagueness.

The final model of uncertainty, non-monotonic logic, takes assertions about the validity of a hypothesis to be true until evidence is found to prove it false. The logic is based on a list of the evidential factors that would prove the hypothesis to be true and those that would prove it to be false. If additional evidence shows a previous hypothesis to be false, it is possible to backtrack to the point at which the false inference was made and establish an alternate hypothesis that is consistent with the new evidence. This model is appropriate for inexactness associated with incomplete evidence, in which case default or expected values can be used until additional evidence suggests some alternate hypotheses.

Lee *et al* [141] further explore the roles of probability and evidential theory in the context of classification of remotely sensed data. The authors focus specifically on the merging of data from different sources to produce a final classified image. It is assumed that there are n data sources, each providing a measurement x_s for each pixel, where $s = 1, 2, \dots, n$. In the case of multiple spectral bands, x_s will be a vector for each pixel. From these measurements, a set of m_s data classes are derived by classifying each source s . The i^{th} data class for source s is denoted as d_{si} , where $i = 1, 2, \dots, m_s$. Finally, a set of M information classes are derived by merging the data classes from each source. Information class j is denoted as w_j , where $j = 1, 2, \dots, M$.

A global membership function is then defined for each information class w_j as

$$F_j = P[w_j]^{1-n} \prod_{s=1}^n P[w_j | x_s]^{\alpha_s} \quad (6.5)$$

where: $P[w_j]$ = the probability of observing information class w_j ;

$P[w_j | x_s]$ = the conditional probability of observing information class w_j given measurement x_s ;
and

α_s = an index of reliability related to the inverse of the uncertainty associated with source s .

The conditional probabilities are estimated as

$$P[w_j | x_s] = \frac{\sum_{i=1}^{m_s} P[x_s | d_{si}, w_j] P[d_{si}, w_j]}{\sum_{j=1}^M \sum_{i=1}^{m_s} P[x_s | d_{si}, w_j] P[d_{si}, w_j]} \quad (6.6)$$

where: $P[d_{si}, w_j]$ = the joint probability of information class j and data class i from source s .

(The reader is referred to the article for the derivation of these equations.) The joint probabilities may be obtained from prior knowledge or may be estimated by tabulating the joint occurrence of the information classes against each data class in a training area for which ground truth data (i.e., the information classes) are available. To simplify the calculations of the conditional probabilities $P[x_s | d_{si}, w_j]$, the authors argue that they may be estimated independently of the information classes as $P[x_s | d_{si}]$.

The membership function in equation (6.5) is used as a discriminant function to classify pixels. This is achieved by assigning each pixel to the information class with the maximum membership function value. The reliability index α_s for source s determines the influence of the source on the discriminant function, since an increase in α_s , which implies a decrease in uncertainty, results in an increase in the influence of the source. The authors suggest that in practice, α_r should be set to 1 for the most reliable source, r . The α_s values for each of the remaining sources are then computed as

$$\alpha_s = u_r / u_s \quad s=1, 2, \dots, n \quad (6.7)$$

where: u_r = the uncertainty associated with source r ; and

u_s = the uncertainty associated with source s .

An empirical test shows that varying the value of u_s for different sources can result in enhanced classification accuracy (expressed as the proportion of pixels correctly classified).

The evidential model presented by the authors focuses on uncertainty in the assignment of classes to pixels and permits different data sources to be merged in order to reduce this uncertainty. Consider set S_1 , which is associated with the first data source and gives the set of probabilities that a pixel belongs to classes A through D.

$$S_1 = \{ P_1[A], P_1[B], P_1[C], P_1[D], \Theta_1 \} \quad (6.8)$$

Θ_1 is the uncommitted or distributed support, as defined in equation (6.2). $P_1[D]$ represents a mixture of classes A and B, which are not individually resolvable. The belief functions for classes A through C (i.e., $Bel_1[A]$, $Bel_1[B]$ and $Bel_1[C]$) can be calculated simply as their associated probabilities from S_1 (i.e., $P_1[A]$, $P_1[B]$ and $P_1[C]$). (See the discussion following equation (6.2).) However, in calculating the belief function for class D, $Bel_1[D]$, it is necessary to account for the fact that classes A and B are subsets of class D. Hence $Bel_1[D] = P_1[A] + P_1[B] + P_1[D]$. More generally,

$$Bel_i[U] = \sum_{V \cap U = V} P_i[V] \quad (6.9)$$

Analogously, the plausibility function is defined as

$$Pl_i[U] = \Theta_i + \sum_{V \cap U \neq \emptyset} P_i[V] \quad (6.10)$$

As a consequence of the inability to resolve classes A and B, the evidential interval for class D may overlap those of the other classes, making it difficult to confidently establish the correct class assignment for the pixel. Assume, however, that a second data source exists for which all classes are individually resolvable. Then

$$S_2 = \{ P_2[A], P_2[B], P_2[C], \Theta_2 \} \quad (6.11)$$

Dempster's orthogonal sum rule may then be applied to combine S_1 and S_2 . Let $P_c[Z]$ be the combined probability for class Z. Then the rule states that

$$P_c[Z] = \frac{1}{k} \sum_{X \cap Y = Z} P_1[X] P_2[Y] \quad (6.12)$$

where:

$$k = 1 - \sum_{X \cap Y = \emptyset} P_1[X] P_2[Y] \quad (6.13)$$

Using the values generated by equation (6.12) to yield a set of combined probabilities, S_c , the belief and plausibility functions may be computed by equations (6.9) and (6.10).

The authors argue that the probabilities in sets S_1 and S_2 may be computed from average classification errors incurred for each source. Several different rules may then be used in the classification process. The support-based rule is based on the assignment of a pixel to the class with the highest belief function value (or the highest plausibility function value, since the authors maintain that the belief and plausibility functions will have the same rank ordering). Alternatively, the absolute rule may be employed, in which a pixel is assigned to the class whose belief function value exceeds the plausibility function values of all other classes (although this can lead to situations where decisions are not possible). Empirical results show that the evidential method is able to incorporate information from different data sources, since classification based on the combined set of probabilities, S_c , was observed to yield a higher classification accuracy than classification based on either data source alone.

Robinson & Strahler [187] focus on models of uncertainty based on fuzzy set theory. As noted above, the basis of fuzzy set theory lies in the assignment of a membership function, $\mu_A(X)$, which indicates the degree to which observation X belongs to set A. Alternatively, the membership function may be viewed as a measure of belief that X is an element of A, or an index of the relative accuracy associated with assigning observation X to class A. The authors describe three approaches for incorporating the membership function within a non-fuzzy schema, such as a GIS in which the semantics of the data model are expressed as precise logical constraints.

The first approach is based on a fuzzy relational database management system, where each observation in the database has an attached membership function value. In this approach, the membership function might represent the degree to which the observation belongs to the class to which it has been assigned. The authors demonstrate how these membership function values may be combined when performing map overlay. (As described in § 5.3, map overlay involves the superimposition of two or more data layers to produce a composite map depicting the joint distribution of attributes from each data layer.) Consider a case in which n data layers are superimposed and each data layer is composed of a set of pixels or cells. Let μ_{mi} be the membership function for cell m on data layer i . If the AND operator is applied (i.e., the intersection of the attributes from each data layer), the membership function for cell m on the composite map, μ_{mc} , is given by

$$\mu_{mc} = \min(\mu_{mi}) \quad i=1,2,\dots,n \quad (6.14)$$

Hence the composite map cell is only as accurate as the least accurate of the corresponding cells on all data layers. If the OR operator is applied (i.e., the union of the attributes from each data layer), the membership function is given by

$$\mu_{mc} = \max(\mu_{mi}) \quad i=1,2,\dots,n \quad (6.15)$$

Thus the composite map cell assumes the accuracy of the most accurate of the corresponding cells on all data layers. As noted in § 5.3.2, these equations represent a generalized form of an error model for map overlay based on the classification error matrix (see § 2.1), where cell accuracies must be either 0 (misclassified) or 1 (correctly classified).

The second approach discussed by the authors is based on the assignment of a “similarity relation” over the elements of each domain set. For example, given a set of slope classes, one can assign a value between 0 and 1 to each pair of classes representing the degree of similarity in the class values. A query that requests the locations in the data base of a certain class can then be answered in terms of the similarity relation between this class and the class associated with each location. Similarity relations can also be assigned over multiple domains and the results combined for queries involving subsets of these domains. When the query incorporates the AND operator (e.g., all locations of class A in domain set 1 AND class B in domain set 2), the combined similarity relation for a given location in the data base is defined as the minimum of its similarity relations for each domain set. When the query incorporates the OR operator, the combined similarity relation is defined as the maximum of the individual similarity relations.

The third approach described by the authors is based on the notion that the imprecision intrinsic to natural language is possibilistic in nature. A possibilistic relational uniform fuzzy (PRUF) model then provides a means of facilitating approximate machine inference. In the PRUF model, queries and propositions are processed by identifying constraints induced by the query or proposition, performing tests on each constraint and then aggregating the individual test results to yield an overall test score. Consider a proposition stating that a specified location is on gentle slopes and is near a certain city. The constraints induced by the proposition, “gentle” and “near,” are tested using a possibility distribution yielding test results indicating the degree to which the specified location satisfies each constraint. The two test results are then aggregated to produce an overall test score indicating the degree to which the proposition is satisfied.

Tests of queries and propositions based on natural language processing have been explored in greater detail by Robinson [186]. The author focuses on the development of a linguistic approximation to the geographical concept of “near” that accounts for the vagueness inherent in this concept. Most previous models of natural language processing do not adequately account for this vagueness, assuming instead that the user has the ability to translate inexact natural language concepts into exact concepts for the computer. The author proposes a question-answer procedure to facilitate machine acquisition of a fuzzy representation of the concept of near and permit automated inference.

Denote as C the concept to be acquired by the computer. Let $F_{k-1}(X)$ be the acquired concept, or fuzzy set, at the $k-1^{\text{th}}$ step in the question-answer procedure. At the k^{th} step, the computer asks whether a given location in the database (e.g., a city), x_i , belongs to set C (i.e., is near to another city, z). If the user responds affirmatively, the computer constructs $F_k(X)$ as a function of the distance between x_i and z , and a parameter α that determines the spread of the fuzzy set. Parameter α is adaptive and is modified in the concept acquisition procedure. The concept C at step k , C_k , is then defined as

$$C_k = C_{k-1} \cup F_k(X) \quad (6.16)$$

If the user responds negatively, the computer constructs $F'_k(X)$ as $1-F_k(X)$ and computes C_k as

$$C_k = C_{k-1} \cap F'_k(X) \quad (6.17)$$

The author describes a method for automating the question-answer procedure such that fuzziness in the concept can be reduced in an efficient manner. A set of locations is selected for testing against location z based on an index of fuzziness. This index also provides the basis for a stopping rule. Empirical results show that concept acquisition is dependent on the stopping rule and the initial value of α .

Smith & Honeycutt [197] describe the value of information concept as a means of estimating the value of reducing uncertainties in data. The concept uses a decision tree to characterize a decision problem and the ways in which the problem might be modified with the incorporation of additional information. The decision tree is composed of nodes and arcs. Nodes represent branching points in the decision process. At "decision nodes," the decision maker must choose between two or more alternatives. Decision nodes are followed by additional decision nodes or by "chance nodes." Chance nodes are points at which uncertainty is resolved. Chance nodes are followed by outcomes of the decision process which may represent the final outcome or additional decision nodes. Arcs represent either alternatives (i.e., the arc joins two decision nodes or a decision node to a chance node) or outcomes (i.e., the arc joins a chance node to a decision node or a chance node to a final outcome). Arcs that represent outcomes are referred to as "chance arcs," each of which possesses an associated probability. The probabilities for the set of chance arcs emanating from a given chance node must sum to 1.

A number is also attached to each final outcome as a function of the desirability of that outcome. The preferred decision path through the tree is then determined by "rolling back" the tree. This is achieved by computing the value at each node at successively higher levels of the tree. In the case of a decision node, the value of the node, v_i , is computed as

$$v_i = \max(u_j) \quad j=1,2,\dots,n \quad (6.18)$$

where: u_j = the value at which arc j terminates; and

n = the number of arcs emanating from node i .

In the case of a chance node, v_i is computed as

$$v_i = \sum_{j=1}^n u_j p_j \quad (6.19)$$

where: p_j = the probability associated with arc j .

This procedure is repeated successively for each level in the tree by substituting the computed v_i values for each u_j . The procedure continues until the highest level of the tree is reached, which corresponds to the root node. The arc emanating from the root node with the highest value of v_i indicates the preferred decision and v_i itself represents the value of this decision.

The expected value of perfect information (EVPI) is defined as the increase in the value of a decision node associated with completely resolving the uncertainty about a chance node. To compute the EVPI, the chance node is moved one level above the decision node and the decision node is replicated for each of the arcs emanating from the chance node. The value of the decision tree is then computed as described above. The EVPI is calculated as the difference between the values of the modified and original trees. It is also possible to compute the expected value of imperfect information using an analogous approach. The authors provide several simple examples in which these procedures are applied. They suggest that the value of information concept can be used in developing spatial databases to answer questions about the preferred scales, sources and sampling methods for collecting and encoding data.

6.2. Geographical Modeling

This section briefly reviews studies that consider the implications for geographical modeling of errors in spatial databases. Some of these studies refer to specific models and are not discussed in great detail below. For example, Davis [70] discusses the reliability of predictive land classification models as a function of cartographic and ecological factors. Cartographic factors include characteristics such as spatial resolution, precision and bias, while ecological factors include the strength, consistency and scale-dependent nature of relationships between ecological and terrain variables. Griffith [106] examines errors in estimates of distance or separation in network line lengths or between areal unit centroids. The expected values and variances of these estimates can be established in the context of both additive and proportional error structures for different probability distributions of error. Kennedy [135] discusses the small number problem, which occurs when small fluctuations in a measured variable cause large changes in a calculated percentage, ratio or rate. This may occur for small areal units, for which the value of the measured variable is small, or for large areal units that are sparsely populated. Vitek & Richards [217] examine the implications of error for flood-hazard analysis. Vertical and horizontal error in topographic maps associated with coordinate transformations and data compilation practices can seriously reduce the reliability of the line defining the flood limit.

In a more general context, errors associated with geographical modeling may be classified as measurement and specification errors (Anselin [6]). Measurement error refers in part to inexactness in the coordinate values of digitized points. Other forms of measurement error relate to the aggregation of attribute values over a set of areal units. The interdependence between location and value in spatial databases gives rise to spatial dependence and heterogeneity. Specification error, in contrast, is particular to a given model and includes the use of an incorrect model, functional form or set of variables. In the context of geographical modeling, specification error results in spatial patterns of error. The presence of errors with a distinctive pattern affects the validity of estimation and prediction with standard linear models and the assessment of model validity. Summary measures of accuracy are likely to imperfectly reflect the partitive accuracy associated with individual observations. A meaningful loss or risk function is needed that accounts for the relative significance of error at different locations (see Aronoff [9, 10], § 2.1.2).

In a non-spatial context it is a relatively straightforward task to evaluate the propagation of measurement error in input variables (Alonso [2] and Burrough [34]). Consider a set of m input variables, x_1 through x_m , which are combined by some arithmetic operator (e.g., addition, multiplication, etc.) to produce an output variable, z . The error propagated to z , ϵ_z , is defined in terms of the measurement error associated with each input variable, ϵ_x , as

$$\epsilon_z^2 = \sum_{i=1}^m \sum_{j=1}^m \frac{\partial z}{\partial x_i} \frac{\partial z}{\partial x_j} \epsilon_{x_i} \epsilon_{x_j} r_{ij} \quad (6.20)$$

where: r_{ij} = the correlation between x_i and x_j .

This equation indicates that propagated error tends to be higher for correlated variables. Moreover, propagated error will be higher for certain arithmetic operators (e.g., multiplication) than others (e.g., addition) as a function of the partial derivatives. Increasing the number of variables will tend to reduce error associated with model misspecification, but will simultaneously increase the amount of propagated measurement error. Therefore model complexity should reflect the reliability of the data. The optimal model is one for which the sum of propagated and specification error is at a minimum. Variants of equation (6.20) can be applied to assess error propagation in estimates of elevation, area and distance (Neumyvakin & Panfilovich [171]).

As noted above, specification error includes the use of an incorrect set of variables. This often occurs when surrogate variables are employed because the real variable of interest cannot be measured. Empirical evidence suggests that reliability of surrogate variables may often be distributed non-uniformly over space (Napton & Luther [170]). Moreover, surrogate variables are frequently collected at scales that are incompatible with the specified statistical model. Hence the reliability of surrogate variables is highly dependent on the context in which they were originally collected.

Several authors have presented general frameworks for assessing the implications of error on spatial transformations or operations. For example, the "spatial data configuration" framework defines the exact relationship between pre- and post-transformed parameters (e.g., means, variances, correlations, etc.) by specifying the manner in which these parameters are modified by the transformation (Arbia [7]). This framework may be applied to transformations such as the delineation of areal units, the aggregation of existing areal units and areal interpolation (see § 5.1). Alternatively, transformations can often be modeled as a spatial filtering operation with an estimable response function (Tobler [210]). To the degree that transformations affect the results of geographical modeling, it might be appropriate to develop modeling techniques that are "frame independent," or unaffected by the spatial coordinate system.

Other authors have proposed that the effects of transformations be evaluated in the context of sensitivity analysis (Openshaw [175]). Formal models of error propagation are not always necessary, since for many purposes precise quantitative estimates of error are not necessary. Sensitivity analysis based on stochastic simulation rests on few assumptions and is computationally tractable. It facilitates calculation of error given a plausible set of assumptions about errors in the untransformed data. In suitability mapping (see Hopkins [119], § 5.3.3), sensitivity analysis may be used to evaluate the effects of different weights or the exclusion of different data layers (Lodwick [145]). These effects may be expressed in the form of confidence limits on the composite map.

6.3. Data Quality Issues

This section briefly reviews some issues of general importance in developing quality standards and methods of error assessment for spatial data. These issues are pertinent in a variety of contexts and are not specific to any single data type, data compilation method or GIS operation.

Errors in spatial databases include inherent and operational error (Vitek *et al* [218] and Walsh [220]). Inherent error arises from errors in source documents as a function of the methods applied in data collection, compilation and representation. Operational error refers to error arising from data processing operations. It is the responsibility of the data producer to document the levels of inherent error and describe the probable nature of operational error and its interaction with inherent error. Data utility, or the level of accuracy required, is application-specific (Morehouse [165]). Hence data users are responsible for evaluating the appropriateness of the data for a specific application.

The critical accuracy issues encountered in the development of databases for specific applications may not conform to theoretical notions about quality and accuracy. The nature of the operational environment must be considered in the development of quality standards and methods of accuracy assessment for spatial data. In developing and implementing large databases for market and social research, many operational concessions must be made that have an impact on data quality (Saalfeld [195]). To the degree that these databases rely on existing data, quality concerns include the integration of data from different sources and methods of aggregation and disaggregation (Brusegard & Menger [31]).

Although the nature of the operational environment has a direct effect on the relevance of different types of error, practitioners must not ignore basic quality standards solely for reasons of cost and expedience. For example, Blakemore [28] argues that enthusiasm for microcomputer technology has led to an erosion of concern for fundamental aspects of data quality. Despite the publication of quality standards by all major mapping agencies (see § 2.3 and § 2.5), concern for data quality has been eroded by the low resolution display devices common to microcomputers. Technological innovation has not responded to quality concerns. Rather, technological limitations have shaped notions of data quality as practitioners have been forced to cope with the crude nature of display devices. The limited storage capabilities of microcomputers may exacerbate the apparent unwillingness of practitioners to incorporate quality information within digital databases.

The concept of error in cartography has evolved over time as a function of the development of new types of maps and new technologies for encoding and displaying spatial data (Muller [169]). The oldest tradition is the concept of horizontal and vertical precision in topographic maps and geodetic surveys (see § 2.3 and § 2.4). Methods of accuracy assessment are relatively well-established and account for errors associated with locating ground position, interpreting data from surveys and air photos, generalizing to small-scale maps and employing different map projections and drafting methods. For thematic or "communication" maps, precisely locating points is relatively unimportant. Error concepts pertinent to such maps include conceptual and perceptual error (see § 4.6) and method-produced error, or error attributable to data compilation (see § 4). These maps often contain "controlled" error, or error intentionally introduced to facilitate cartographic communication. Modern spatial data handling maps are produced for a variety of purposes and are often merged and combined in an automated environment to produce new maps. For these maps, error is defined relative to map use and objectives. Hence the importance of horizontal and vertical precision, conceptual and perceptual error, and method-produced error depend on the application of the map.

Dutton [76] argues that the development of data quality standards must consider the ontology of space, since the character of spatial data is tied to notions of what constitutes reality. Ontology is closely allied with the purpose or mandate whereby data have been collected. Discrepancies in the representations of the same feature in different databases often result because the feature has been encoded to fulfill a different purpose. Topological data structures are required that allow for multiple versions of the same feature. Data structures must also be capable of handling features whose locations are indeterminate or defined solely by cultural and institutional consensus. Accurate ground control is important in developing accurate databases, but current technological developments encourage *ad hoc* measurements unrelated to existing survey networks. Such measurements will have no general utility without agreed-upon cataloging procedures. Thus evaluation of data quality does not amount simply to estimating global and local error parameters and documenting their incidence in a quality report. Rather than relying on external documentation of quality, data structures must intrinsically qualify their information content. Important qualifiers include location, time, scale and purpose, the specification of which must be seen as an implicit goal of the development of data quality standards.

White [227] similarly argues that many errors in spatial databases are fundamentally attributable to ontological issues. Correcting such errors amounts to identifying the elementary geometrical features that are assumed to exist. For example, a boundary between two polygons is a shared attribute of these polygons. The boundary should therefore be represented by encoding the shared points along this boundary rather than tracing the boundary of each polygon independently. This approach eliminates

gaps and slivers that inevitably occur between independently traced polygon boundaries. The generation of spurious polygons in map overlay (see § 5.3.2) results because the automated process does not adequately mimic human geometrical intuition. (A similar point is made by Goodchild [98].) Again the solution to this problem is not to maintain the separate data layers but to define the elementary features assumed to exist on the composite map. At a more basic level, successful development and implementation of spatial databases is often thwarted by the failure to separate points from their coordinate values. In consequence, precise coordinate values must be obtained before the database can be implemented and future revisions of coordinate values cannot easily be achieved. Points should be consistently and absolutely distinguishable but coordinate values should be unconstrained variables. Successful automation of spatial data requires understanding of the nature of space afforded by topological, graph and model theory.

7. SUMMARY

The issue of spatial database accuracy encompasses a variety of concepts, methods and models. The significance of different dimensions of accuracy is a function of data type, application and the sources of error deemed to be important in a particular context. This report reviews a diverse range of research and provides a logical structure for addressing different dimensions of spatial database accuracy. On the broadest level, the error taxonomy developed in this report identifies five main dimensions of spatial database accuracy — measurement of error in spatial databases (§ 2), accuracy of cartometric estimates (§ 3), errors introduced during data compilation (§ 4), propagation of error through GIS operations (§ 5) and general issues of spatial database accuracy (§ 6). The taxonomy organizes different dimensions of accuracy into a practical, user-oriented structure that reflects the interests of those conducting both pure and applied research.

Methods for detecting and measuring errors in spatial databases are presented in § 2. Various aspects of classification accuracy are examined in § 2.1. In a remote sensing context, classification accuracy is often evaluated with a classification error matrix, or a cross-tabulation of the assigned and actual classes for a sample of points or pixels. The PCC, or the percentage of pixels correctly classified, is frequently employed as an index of classification accuracy. In § 2.1.1, methods are described for constructing confidence limits around a sample PCC, estimating the minimum sample size required to meet a specified level of confidence and performing hypothesis tests on a sample PCC based on the binomial distribution or the normal approximation to this distribution. Alternatives to the PCC as an index of classification accuracy are introduced in § 2.1.2. These alternatives are designed to circumvent some of the limitations of the PCC, including its sensitivity to row and column totals, its inability to distinguish between errors of omission and commission and its inability to account for correct classifications occurring by chance alone. Alternatives include producer's and user's accuracy, minimum class accuracy, the κ statistic and its variants, transformations of the classification error matrix and a variety of statistical models. Some alternatives to the classification error matrix are described in § 2.1.3. These alternatives are based on the agreement between the mapped and actual areas of different classes. Other issues pertaining to classification accuracy are addressed in § 2.1.4, including the implications of spatial autocorrelation and errors in ground survey data for accuracy assessment, sources of error in classification methods and the development of standards for accuracy assessment procedures.

In § 2.2, the nature of error in soil maps is addressed in the context of spatial variations in soil properties. Soil properties exhibit relatively high-frequency spatial variations that result in heterogeneity within mapping units and indeterminacy in mapping unit boundaries. Empirical studies of the nature of soil property variations are reviewed in § 2.2.1. Mathematical models of these variations are presented in § 2.2.2. Some of these models account for soil property variations as a set of superimposed processes operating at different spatial scales that give rise to definitive soil patterns. Other models focus on the changes in soil properties at mapping unit boundaries and the use of ancillary data to improve the accuracy with which mapping unit boundaries are delineated.

Data quality standards and methods for evaluating horizontal and vertical error for topographic maps are described in § 2.3. Standards based on compliance testing, including the National Map Accuracy Standard, are compared with standards based on statistical expressions of accuracy, including the Engineering Map Accuracy Standard and the American Society of Photogrammetry and Remote Sensing spatial accuracy specification. Methods that have been proposed as the basis for alternate standards are also described in this section. An example is Koppe's formula, which accounts for the effects of terrain slope on vertical error, permits horizontal and vertical accuracy standards to be combined into a single expression and facilitates computation of bands of horizontal error around contour lines.

The issue of horizontal accuracy for large-scale planimetric and cadastral maps is addressed in § 2.4. This section of the report examines the trade-off between the desirability of rigorous standards and the cost of acquiring data that conform to these standards. This section also addresses the problems associated with merging data with different levels of spatial and taxonomic resolution and the application of hierarchical tessellations that define horizontal precision as an implicit attribute of the level in the hierarchy. Planimetric accuracy is also closely associated with distortions introduced by map projections. These distortions can be evaluated with techniques such as Tissot's indicatrix and reduced with various linear and non-linear coordinate transformations.

General data quality standards for multi-purpose digital geographic base files are discussed in § 2.5. These standards are based on the notion of "truth in labelling" rather than compliance testing or statistical expressions of accuracy. Concepts such as data lineage, positional accuracy, attribute accuracy, logical consistency and completeness are discussed in this section.

Methods for classifying and correcting vertical errors in digital elevation models are presented in § 2.6. This section describes error classifications, accuracy testing procedures and aspects of spatial resolution that affect accuracy as a function of terrain variability. Several techniques are also presented for detecting and correcting gross errors based on surface modeling.

The accuracy of cartometric estimates, or estimates of lengths and areas derived from maps, is examined in § 3. The effects of cartographic generalization on line length estimation are discussed in § 3.1. In § 3.1.1, empirical studies are reviewed that focus on the interaction between map scale and the method employed to estimate line length. This section also details some techniques that may be used to correct for these interactions. The application of fractal theory in this context is addressed in § 3.1.2.

The accuracy of cartometric estimates of area is discussed in § 3.2 in the context of classification error and the method used to derive area estimates. Empirical studies of the effects of grid density in dot planimetry are reviewed in § 3.2.1. In § 3.2.2, methods are presented that may be used to correct for the effects of classification error on area estimates. These methods are based on transformations of the classification error matrix.

Errors introduced into spatial databases by data compilation methods are discussed in § 4. Models of digitizing error, or inaccuracies in the encoded positions of digitized points, are presented in § 4.1. These models typically focus on positional error in the context of cartographic lines or polygon boundaries defined by sets of digitized points joined by straight line segments. The epsilon band concept and its variants define error in terms of a distribution of probable true line locations around the digitized line (§ 4.1.1). Digitizing error may also be modeled as a serially dependent process, such that the error at a given point along the line is dependent on the error at previously digitized points (§ 4.1.2). Various physiological and psychological factors also affect the ability of the human operator to perform the digitizing task (§ 4.1.3).

The issue of generalization error, or error arising from the representation of cartographic lines as sets of digitized points joined by straight line segments, is discussed in § 4.2. This section focuses on the epsilon band concept and its variants, the implications of cartographic generalization for line length estimation and the effects of line simplification on the positional accuracy of cartographic lines.

Errors associated with choroplethic mapping are discussed in § 4.3. The focus of § 4.3.1 is data symbolization, or the transformation of the numerical attribute values associated with a set of areal units to a set of classes. Map accuracy depends on the selection of class intervals that reliably depict the numerical values of the areal units. Methods are described that facilitate near-optimal class interval selection. These methods are dependent on areal unit geometry and the nature of the underlying statistical surface. Interactions between attribute and positional errors are examined in § 4.3.2.

Errors in isometric maps, the focus of § 4.4, are introduced primarily by factors affecting the accuracy of interpolation. As described in § 4.4.1, these factors include the nature of the underlying surface, the interpolation algorithm and the sampling method (including the number and spatial distribution of sample points). The interaction between interpolation and measurement error is discussed in § 4.4.2. Various methods are described whereby the effects of measurement error on interpolation accuracy can be evaluated and reduced.

For isopleth maps (§ 4.5), accuracy is also largely a function of the nature of the underlying surface, the interpolation algorithm and the sampling method (including the number and geometry of the areal units to which attribute values have been assigned). This section of the report reviews a number of empirical studies that focus on these factors.

Conceptual and perceptual errors in mapping are addressed in § 4.6 in the context of cartographic communication. The cartographic communication model posits that cognitive and physical filters introduce distortions at each phase of the communication process. Effective communication requires that these distortions be controlled to enhance, rather than detract from, the communicative potential of the map. Important considerations include overall map design, symbology, map projection and class interval selection. This section also addresses the question of liability, or responsibility for conceptual and perceptual errors. The notion of uncertainty absorption is introduced, whereby the map producer guarantees the map as official truth or the user accepts responsibility for all misinterpretations.

Models of error propagation for spatial operations applied in a GIS environment are detailed in § 5. The focus of § 5.1 is areal interpolation, or cross-area estimation. In this operation, the values of some variable of interest are estimated for a set of target areal units based on the corresponding values for a set of source units. Accuracy is primarily a function of areal unit geometry, the nature of the underlying statistical surface and the degree of coincidence in target and source areal unit boundaries. Various techniques are discussed that can be used to model the effects of these factors and improve the accuracy of areal interpolation using ancillary data.

Vector to raster conversion, discussed in § 5.2, is analogous to areal interpolation when the target areal units are regular grid cells. Models of positional error for vector to raster conversion (§ 5.2.1) focus on inaccuracies in estimates of polygon area based on counts of cells. These errors arise from the presence of boundary cells, or cells intersected by a polygon boundary. A variety of error models are presented that account for the effects of boundary cells as a function of cell size, polygon size and boundary smoothness. Models of attribute error (§ 5.2.2) focus on errors in the raster representation of the thematic attribute. These errors arise from the often arbitrary nature of polygon boundaries, which masks the nature of the underlying statistical surface. Models of attribute error focus on errors of omission and commission in assigning cells to polygons.

Map overlay (§ 5.3) involves the superimposition of two or more data layers to produce a composite map showing combinations of the thematic attributes on each data layer. The accuracy of this operation is a function of positional error in polygon boundaries, attribute error and error introduced by the map overlay operation itself (§ 5.3.1). The spurious polygon problem is discussed in § 5.3.2. Spurious polygons arise from positional errors in polygon boundaries that give rise to discrepancies in the cartographic representations of the same feature on different data layers. Error models for spurious polygons are based on the epsilon band concept and its variants, the geometrical properties of polygon boundaries and the notion that spurious polygons may be identified in terms of their size. Models of attribute error for map overlay are described in § 5.3.3. These models use indices of classification error for individual data layers to estimate the level of classification error on the composite map. Analogous models exist for numerical data. This section of the report also addresses the impact of different arithmetic operators on the reliability of composite maps used in suitability mapping.

The issues addressed in § 6 are more general in nature and have broad implications for assessing accuracy in spatial databases. Methods of incorporating information about inexactness into spatial operations are described in § 6.1. These methods derive from natural language processing, the value of information concept, non-monotonic logic and fuzzy set, evidential and probability theory.

Some of the implications of errors in spatial databases for geographical modeling are addressed in § 6.2. This section focuses on the trade-off between measurement and specification error, error propagation through simple mathematical models and the role of sensitivity analysis in evaluating the effects of different types of error.

In § 6.3, accuracy issues of general significance are discussed in the context of the data quality standards and methods of accuracy assessment for spatial data. These issues include inherent and operational error, the nature of the user environment, the evolution of the concept of error in cartography and ontological dimensions of accuracy.

8. REFERENCES

1. Adejuwon, O. (1975). A note on the comparison of chorochromatic surfaces. *Geographical Analysis*, 7(4), 435-440.
2. Alonso, W. (1968). Predicting best with imperfect data. *American Institute of Planners Journal*, 34(4), 248-255.
3. American Society of Civil Engineers (Committee on Cartographic Surveying, Surveying and Mapping Division). (1983). *Map uses, scales and accuracies for engineering and associated purposes*. New York, NY: American Society of Civil Engineers.
4. American Society of Photogrammetry (Committee for Specifications and Standards, Professional Practice Division). (1985). Accuracy specification for large-scale line maps. *Photogrammetric Engineering and Remote Sensing*, 51(2), 195-199.
5. Ainrbein, C. & Griffith, D. (1988). GISs, spatial statistics, and statistical quality control. Unpublished manuscript.
6. Anselin, L. (1988). What is special about spatial errors? Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
7. Arbia, G. (1988). Statistical effects of spatial data transformations. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
8. Aronoff, S. (1982). Classification accuracy: A user approach. *Photogrammetric Engineering and Remote Sensing*, 48(8), 1299-1307.
9. Aronoff, S. (1982). The map accuracy report: A user's view. *Photogrammetric Engineering and Remote Sensing*, 48(8), 1309-1312.
10. Aronoff, S. (1984). An approach to optimized labeling of image classes. *Photogrammetric Engineering and Remote Sensing*, 50(6), 719-727.
11. Aronoff, S. (1985). The minimum accuracy value as an index of classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 51 (1), 99-111.
12. Barrett, A. N. & Rhind, D. W. (1974). Retrospective application of boundary constraints in automated contouring by digital spatial filtering. *Mathematical Geology*, 6(1), 83-89.
13. Barrett, J. P. & Philbrook, J. S. (1970). Dot grid area estimates: Precision by repeated trials. *Journal of Forestry*, 68(3), 149-151.
14. Bascomb, C. L. & Jarvis, M. G. (1976). Variability in three areas of the Denchworth soil map unit: L Purity of the map unit and property variability within it. *Journal of Soil Science*, 27(3), 420-437.
15. Battle, J. E. (1977). Coding and interpolation of topographic surfaces. *Proceedings, American Congress on Surveying and Mapping*, 288-300.
16. Bauer, M. E., Hixson, M. M., Davis, B. J. & Etheridge, J. B. (1978). Area estimation of crops by digital analysis of Landsat data. *Photogrammetric Engineering and Remote Sensing*, 44(8), 1033-43.
17. Baugh, I. D. H. & Boreham, J. R. (1976). Measuring the coastline from maps: A study of the Scottish mainland. *The Cartographic Journal*, 13(2), 167-171.
18. Beard, K. (1989). Use error: The neglected error component. *Proceedings, Auto Carto 9*, 808-817.

19. Beckett, P. (1977). Cartographic generalization. *The Cartographic Journal*, 14(1), 49-50.
20. Beckett, P. H. T. & Burrough, P. A. (1971). The relation between cost and utility in soil survey: IV. Comparison of the utilities of soil maps produced by different survey procedures, and to different scales. *Journal of Soil Science*, 22(4), 466-480.
21. Beckett, P. H. T. & Burrough, P. A. (1971). The relation between cost and utility in soil survey: V. The cost-effectiveness of different soil survey procedures. *Journal of Soil Science*, 22(4), 481-489.
22. Bedard, Y. (1986). A study of the nature of data using a communication-based conceptual framework of land information systems. *The Canadian Surveyor*, 40(4), 449-460.
23. Bedard, Y. (1986). Comparing registration systems using a communication-based criterion: Uncertainty absorption. XVIII Congress, International Federation of Surveyors, Commission 7, 565-579.
24. Bedard, Y. (1987). Uncertainties in land information systems databases. *Proceedings, Auto Carto 8*, 175-184.
25. Bennett, H. C. (1977). The geographic data base: Reliability or chaos? *Proceedings of the American Congress on Surveying and Mapping 37th Annual Meeting*, 675-680.
26. Bethel, J. S. & Mikhail, E. M. (1983). On-line quality assessment in DTM. *Technical Papers, American Congress on Surveying and Mapping-American Society of Photogrammetry Fall Convention*, 576-584.
27. Blakemore, M. (1983). Generalisation and error in spatial data bases. *Cartographica*, 21(2/3), 131-139. (See also: Blakemore, M. (1983). Generalisation and error in spatial data bases. *Proceedings, Auto Cart 6*, 313-322.)
28. Blakemore, M. (1985). High or low resolution? Conflicts of accuracy, cost, quality and application in computer mapping. *Computers and Geosciences*, 11 (3), 345-48.
29. Blakney, W. G. G. (1968). Accuracy standards for topographic mapping. *Photogrammetric Engineering*, 34(10), 1040-1042.
30. Braile, L. W. (1978). Comparison of four random to grid methods. *Computers and Geosciences*, 4(4), 341-349.
31. Brusegard, D. & Menger, G. (1988). Real data with real problems: Dealing with large spatial databases. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
32. Burrough, P. A. (1983). Multiple sources of spatial variation in soil: I. The application of fractal concepts to nested levels of soil variation. *Journal of Soil Science*, 34(3), 577-597.
33. Burrough, P. A. (1983). Multiple sources of spatial variation in soil: H. A non-Brownian fractal model and its application in soil survey. *Journal of Soil Science*, 34(3), 599-620.
34. Burrough, P. A. (1986). *Principles of geographical information systems for land resources assessment*. Oxford: Clarendon.
35. Burrough, P. A. (1987). Multiple sources of spatial variation and how to deal with them. *Proceedings, Auto Carlo 8*, 145-154.
36. Burtch, R. & Thapa, K. (1989). Developing accuracy standards for multipurpose cadastral systems. *Technical Papers, American Society of Photogrammetry and Remote Sensing/American Congress on Surveying and Mapping Annual Convention*, 5, 67-75.
37. Campbell, J. B. (1977). Variation of selected properties across a soil boundary. *Soil Science Society of America Journal*, 41(3), 578-82.
38. Campbell, J. B. & Edmonds, W. J. (1984). The missing geographic dimension to soil taxonomy. *Annals of the Association of American Geographers*, 74(1), 83-97.
39. Card, D. H. (1982). Using known map category marginal frequencies to improve estimates of thematic map accuracy.

Photogrammetric Engineering and Remote Sensing, 48(3), 431-439.

40. Carstensen, L. W., Jr. (1987). A measure of similarity for cellular maps. *The American Cartographer*, 14(4), 345-358.
41. Carter, J. R. (1989). Relative errors identified in USGS gridded DEMs. *Proceedings, Auto Carlo* 9,255-265.
42. Caruso, V. M. (1987). Standards for digital elevation models. *Technical Papers, American Society of Photogrammetry and Remote Sensing-American Congress on Surveying and Mapping Annual Convention*, 4, 159-166.
43. Chrisman, N. R. (1982). A theory of cartographic error and its measurement in digital data bases. *Proceedings, Auto Carlo* 5, 159-168.
44. Chrisman, N. R. (1982). Beyond accuracy assessment: Correction of a misclassification. *Proceedings, International Society for Photogrammetry and Remote Sensing Commission IV*, 123-132.
45. Chrisman, N. R. (1983). Epsilon filtering: A technique for automated scale changing. *Technical Papers of the 43rd Annual Meeting of the American Congress on Surveying and Mapping*, 322-331.
46. Chrisman, N. R. (1983). The role of quality information in the long-term functioning of a geographic information system. *Cartographica*, 21(2/3), 79-87. (See also: Chrisman, N. R. (1983). The role of quality information in the long-term functioning of a geographic information system. *Proceedings, Auto Carto* 6, 303-312.)
47. Chrisman, N. R. (1987). Efficient digitizing through the combination of appropriate hardware and software for error detection and editing. *International Journal of Geographic Information Systems*, 1(3), 265-277.
48. Chrisman, N. R. (1987). The accuracy of map overlays: A reassessment. *Landscape and Urban Planning*, 14(5), 427-439.
49. Chrisman, N. R. (1988). Modeling error in overlaid categorical maps. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
50. Chrisman, N. R. (1989). Error in categorical maps: Testing versus simulation. *Proceedings, Auto Carto* 9, 521-529.
51. Clarke, K. C. (1985). A comparative analysis of polygon to raster interpolation methods. *Photogrammetric Engineering and Remote Sensing*, 51(5), 575-582.
52. Congalton, R. G. (1988). Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 54(5), 587-592.
53. Congalton, R. G. (1988). A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 54(5), 593-6W.
54. Congalton, R. G. & Mead, R. A. (1983). A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*, 49(1), 69-74. (See also: Congalton, R. G. & Mead, R. A. (1981). A quantitative method to test for similarity between photo interpreters. *Technical Papers, American Society of Photogrammetry 47th Annual Meeting*, 263-266.)
55. Congalton, R. G., Oderwald, R. G. & Mead, R. A. (1983). Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering and Remote Sensing*, 49(12), 1671-1678.
56. Cook, B. G. (1983). Geographic overlay and data reliability. *Proceedings, United States/Australia Workshop on Design and Implementation of Computer-Based Geographic Information Systems*, 64-69.
57. Cook, H. R. (1976). Accuracy/quality control of digital data in the digital topographic information bank. *Proceedings of the American Congress on Surveying and Mapping Annual Meeting*, 483-488.
58. Court, A. (1970). Map comparisons. *Economic Geography*, 46(2) (supplement), 435-438.
59. Crackel, T. J. (1975). The linkage of data describing overlapping geographical units: A second iteration. *Historical Methods Newsletter*, 8(3), 146-150.

60. Crapper, P. F. (1980). Errors incurred in estimating an area of uniform land cover using Landsat. *Photogrammetric Engineering and Remote Sensing*, 46(10), 1295-1301.
61. Crapper, P. F. (1981). Geometric properties of regions with homogeneous biophysical characteristics. *Australian Geographical Studies*, 19(1), 117-124.
62. Crapper, P. F. (1981). The relationship between region shape and error variance. *Proceedings, Second Australian Remote Sensing Conference*, 6.3.1-6.3.5.
63. Crapper, P. F. (1984). An estimate of the number of boundary cells in a mapped landscape coded to grid cells. *Photogrammetric Engineering and Remote Sensing*, 50(10), 1497-1503.
64. Crapper, P. F., Walker, P. A. & Manninga, P. M. (1986). Theoretical prediction of the effect of aggregation on grid cell data sets. *Geo-Processing*, 3(2), 155-166.
65. Crosson, L. S. & Protz, R. (1974). Quantitative comparison of two closely related soil mapping units. *Canadian Journal of Soil Science*, 54(1), 7-14.
66. Curran, P. J. & Williamson, H. D. (1985). The accuracy of ground data used in remote-sensing investigations. *International Journal of Remote Sensing*, 6(10), 1637-1651.
67. Dahlberg, R. E. (1984). The public land survey system: The American rural cadastre. *Computers, Environment and Urban Systems*, 9(213), 145-153.
68. Dahlberg, R. E. (1986). Combining data from different sources. *Surveying and Mapping*, 46(2), 141-149.
69. Dangermond, J. (1988). A review of digital data commonly available and some of the practical problems of entering them into a GIS. *Technical Papers, American Congress on Surveying and Mapping-American Society of Photogrammetry and Remote Sensing Annual Convention*, 5, 1-10.
70. Davis, F. (1988). Impacts of map errors on predictive land classification. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
71. Dicks, S. E. & Lo, T. H. C. (1989). Evaluation of thematic map accuracy in a land use and cover mapping program. *Technical Papers, American Society of Photogrammetry and Remote Sensing-American Congress on Surveying and Mapping Annual Convention*, 4, 7-20.
72. Digital Cartographic Data Standards Task Force. (1988). Draft proposed standard for digital cartographic data. *The American Cartographer*, 15(1).
73. Donahue, J. G. (1988). Land base accuracy: Is it worth the cost? *American Congress on Surveying and Mapping Bulletin*, 117, 25-27.
74. Dougenik, J. (1980). Whirlpool: A geometric processor for polygon coverage data. *Proceedings, Auto Carto 4*, 2, 304-311.
75. Dozier, J. & Strahler, A. H. (1983). Ground investigations in support of remote sensing. In R. N. Colwell (Ed.), *Manual of Remote Sensing* (2nd ed.) (pp. 959-986). Falls Church, VA: American Society of Photogrammetry.
76. Dutton, G. (1984). Truth and its consequences in digital cartography. *Technical Papers, FortyFourth Annual Meeting, American Congress on Surveying and Mapping*, 273-283.
77. Dutton, G. (1988). Modeling locational uncertainty in hierarchical tessellations. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
78. Estes, J. (1988). GIS product accuracy. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.

79. Fairchild, D. (1981). The effect of enumeration unit shape on isopleth map accuracy. *Technical Papers of the American Congress on Surveying and Mapping*, 423-432.
80. Fisher, P. F. (1988). Knowledge-based approaches to determining and correcting areas of unreliability in geographic databases. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
81. Fitzpatrick-Lins, K. (1978). Accuracy and consistency comparisons of land use and land cover maps made from high-altitude photographs and Landsat multispectral imagery. *Journal of Research, US Geological Survey*, 6(1), 23-40.
82. Fitzpatrick-Lins, K. (1978). Accuracy of selected land use and land cover maps in the greater Atlanta region, Georgia. *Journal of Research, US Geological Survey*, 6(2), 169-173.
83. Fitzpatrick-Lins, K. (1981). Comparison of sampling procedures and data analysis for a land-use and land-cover map. *Photogrammetric Engineering and Remote Sensing*, 47(3), 343-351.
84. Flowerdew, R. (1988). Statistical methods for inference between incompatible zonal systems. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California. (See also: Flowerdew, R. (1988). Statistical methods for areal interpolation: Predicting count data from a binary variable (Research Report No. 16). Northern Regional Research Laboratory.)
85. Flowerdew, R. & Openshaw, S. (1987). A review of the problems of transferring data from one set of areal units to another incompatible set (Research Report No. 4). Northern Regional Research Laboratory.
86. Frazier, B. E. & Shovic, H. F. (1980). Statistical methods for determining land-use change with aerial photographs. *Photogrammetric Engineering and Remote Sensing*, 46(8), 1067-1077.
87. Frolov, Y. S. & Maling, D. H. (1969). The accuracy of area measurement by point counting techniques. *The Cartographic Journal*, 6(1), 21-35.
88. Galloway, R. W. & Bahr, M. E. (1979). What is the length of the Australian coast? *Australian Geographer*, 14(4), 244-247.
89. van Genderen, J. L. & Lock, B. F. (1977). Testing land-use map accuracy. *Photogrammetric Engineering and Remote Sensing*, 43(9), 1135-1137.
90. van Genderen, J. L., Lock, B. F. & Vass, P. A. (1978). Remote sensing: Statistical testing of thematic map accuracy. *Remote Sensing of Environment*, 7(1), 3-14.
91. Gering, L. R., Bailey, R. L. & Shain, W. A. (1985). Estimation of area using optimum dot-grid density with aerial photographs. *Technical Papers, American Congress on Surveying and Mapping-American Society of Photogrammetry Fall Convention*, 811-820.
92. Gersmehl, P. J. (1981). Maps in landscape interpretation. *Cartographica*, 18(2), 79-115.
93. Gersmehl, P. J. (1985). The data, the reader, and the innocent bystander: A parable for map users. *The Professional Geographer*, 37(3), 329-334.
94. Ginevan, M. E. (1979). Testing land-use map accuracy: Another look. *Photogrammetric Engineering and Remote Sensing*, 45(10), 1371-1377.
95. Goodchild, M. F. (1978). Statistical aspects of the polygon overlay problem. In G. Dutton (Ed.), *Harvard Papers on Geographic Information Systems*, 6. Cambridge, MA: Laboratory for Computer Graphics and Spatial Analysis, Harvard University.
96. Goodchild, M. F. (1980). Fractals and the accuracy of geographical measures. *Mathematical Geology*, 12(2), 85-98.
97. Goodchild, M. F. (1980). The effects of generalization in geographical data encoding. In H. Freeman & G. G. Pieroni (Eds.), *Map data processing* (pp. 191-205). New York, NY: Academic Press.
98. Goodchild, M. F. (1982). Accuracy and spatial resolution: Critical dimensions for geoprocessing. In D. H. Douglas & A. R.

- Boyle (Eds.), Computer assisted cartography and geographic information processing: Hope and Realism (pp. 87-90). Ottawa, Ontario: Canadian Cartographic Association.
99. Goodchild, M. F. (1988). Towards models of error for categorical maps. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
 100. Goodchild, M. F. & Dubuc, O. (1987). A model of error for choropleth maps, with applications to geographic information systems. *Proceedings, Auto Carto 8*, 165-174.
 101. Goodchild, M. F. & Lam, N. S.-N. (1980). Areal interpolation: A variant of the traditional spatial problem. *Geo-Processing*, 1(3), 297-312.
 102. Goodchild, M. F. & Wang, M.-H. (1988). Modeling error in raster-based spatial data. *Proceedings, Third International Symposium on Spatial Data Handling*.
 103. Goodchild, M. F. & Wang, M.-H. (1989). Modeling errors for remotely sensed data input to GIS. *Proceedings, Auto Carto 9*, 530-537.
 104. Grady, R. K. (1988). Data lineage in land and geographic information systems (LIS/GIS). *Proceedings, GISILIS '88*, 722-730.
 105. Greenland, A., Socher, R. M. & Thompson, M. R. (1985). Statistical evaluation of accuracy for digital cartographic data bases. *Proceedings, Auto Carto 7*, 212-221.
 106. Griffith, D. (1988). Distance calculations and errors in geographic data bases. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
 107. Gupta, S. C. (1988). Inclusion of accuracy data in a feature based, object oriented data model. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
 108. Gustafson, G. C. (1977). Koppe errors in automatic contouring. *Proceedings of the American Congress on Surveying and Mapping 37th Annual Meeting*, 305-317.
 109. Gustafson, G. C. & Loon, J. C. (1981). Updating the national map accuracy standards. *Technical Papers of the American Congress on Surveying and Mapping*, 466-482.
 110. Hakanson, L. (1978). The length of closed geomorphic lines. *Mathematical Geology*, 10(2), 141-167.
 111. Hannah, M. J. (1979). Error detection and rectification in digital terrain models. *Joint Proceedings of the American Society of Photogrammetry and American Congress on Surveying and Mapping Fall Meeting*, 152-164.
 112. Havens, K. A., Minter, T. C. & Thadani, S. G. (1977). Estimation of the probability of error without ground truth and known a priori probabilities. *IEEE Transactions on Geoscience Electronics*, GE-15(3), 147-152.
 113. Hay, A. M. (1979). Sampling designs to test land-use map accuracy. *Photogrammetric Engineering and Remote Sensing*, 45(4), 529-533.
 114. Hay, A. M. (1988). The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing*, 9(8), 1395-1398.
 115. Herzog, A. (1988). Modeling reliability on statistical surfaces by polygon filtering. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
 116. Hixson, M. M. (1981). Techniques for evaluation of area estimates. *1981 Machine Processing of Remotely Sensed Data Symposium*, 84-90.
 117. Hixson, M. M., Bauer, M. E. & Davis, B. J. (1979). Sampling for area estimation: A comparison of full-frame sampling with the sample segment approach. *1979 Machine Processing of Remotely Sensed Data Symposium*, 97-103.

118. Honeycutt, D. M. (1986). Epsilon, generalization and probability in spatial data bases. Unpublished manuscript.
119. Hopkins, L. D. (1977). Methods of generating land suitability maps: A comparative evaluation. *American Institute of Planners Journal*, 43(4), 386-400.
120. Hord, R. M. & Brooner, W. (1976). Land-use map accuracy criteria. *Photogrammetric Engineering and Remote Sensing*, 42(5), 671-677.
121. Hsu, M.-L. (1975). Filtering process in surface generalization and isopleth mapping. In J. C. Davis & M. J. McCullagh (Eds.), *Display and analysis of spatial data* (pp. 115-129). London: Wiley.
122. Hsu, M.-L. & Robinson, A. H. (1970). *The fidelity of isopleth maps: An experimental study*. Minneapolis, MN: University of Minnesota Press.
123. Hsu, S.-Y. (1978). The national map accuracy standards in the context of mapping with imagery data. *Proceedings of the American Congress on Surveying and Mapping Fall Convention*, 202-207.
124. Hudson, D. (1988). Some comments on data quality in a GIS. *Technical Papers, American Congress on Surveying and Mapping-American Society of Photogrammetry and Remote Sensing Annual Convention*, 2, 203-210.
125. Hudson, W. D. & Ramm, C. W. (1987). Correct formulation of the kappa coefficient of agreement. *Photogrammetric Engineering and Remote Sensing*, 53(4), 421-422.
126. Imhof, I. (1982). *Cartographic reliefpresentation*. New York, NY: de Gruyter.
127. Jenks, G. F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, 7, 182-190.
128. Jenks, G. F. (1970). Conceptual and perceptual error in thematic mapping. *30th Annual Meeting, American Congress on Surveying and Mapping*, 174-188.
129. Jenks, G. F. (1976). Contemporary statistical maps: Evidence of spatial and graphic ignorance. *The American Cartographer*, 3(1), 11-19.
130. Jenks, G. F. (1981). Lines, computers, and human frailties. *Annals of the Association of American Geographers*, 71(1), 1-10.
131. Jenks, G. F. & Caspell, F. C. (1971). Error in choroplethic maps: Definition, measurement, reduction. *Annals of the Association of American Geographers*, 61(2), 217-244.
132. Keefer, B. J., Smith, J. L. & Gregoire, T. G. (1988). Simulating manual digitizing error with statistical models. *Proceedings, GISILIS '88*, 475-483.
133. Kellie, A. C. & Bryan, D. G. (1981). A comparison of field methods for testing the vertical accuracy of topographic maps. *Technical Papers of the American Congress on Surveying and Mapping*, 275-284.
134. Kenk, E., Sondheim, M. & Yee, B. (1988). Methods for improving accuracy of thematic mapper ground cover classifications. *Canadian Journal of Remote Sensing*, 14(1), 17-31.
135. Kennedy, S. (1988). The small number problem and the accuracy of spatial databases. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
136. Lam, N. S.-N. (1982). Areal interpolation using map overlay. *Modeling and Simulation*, 13, 953-959.
137. Lam, N. S.-N. (1982). An evaluation of areal interpolation methods. *Proceedings, Auto Carto 5*, 471-479.
138. LaMacchia, R. A. (1982). Geographic areas and computer files from the 1980 decennial census. *Proceedings, Auto Carto 5*, 491-498.
139. Laskowski, P. H. (1987). Map distortions and singular value decomposition. *Technical Papers, American Society of Photogrammetry and Remote Sensing-American Congress on Surveying and Mapping Annual Convention*, 4, 42-51.

140. Laskowski, P. H. (1988). The traditional and modern look at Tissot's Indicatrix. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
141. Lee, T., Richards, J. A. & Swain, P. H. (1987). Probabilistic and evidential approaches for multisource data analysis. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(3), 283-293.
142. Lee, Y. C. (1985). Comparison of planimetric and height accuracy of digital maps. *Surveying and Mapping*, 45(4), 333-340.
143. Little, A. R. (1989). An evaluation of selected computer-assisted line simplification algorithms in the context of map accuracy standards. Technical Papers, American Society of Photogrammetry and Remote Sensing/American Congress on Surveying and Mapping Annual Convention, 5, 122-132.
144. Lloyd, P. R. (1976). Quantisation error in area measurement. *The Cartographic Journal*, 13(1), 22-25.
145. Lodwick, W. (1988). Confidence limits in geographic analysis-suitability analysis. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
146. Loelkes, G. L., Jr. (1977). Specifications for land use and land cover and associated maps (Open File No. 77-555). Reston, VA: US Geological Survey.
147. MacDougall, E. B. (1975). The accuracy of map overlays. *Landscape Planning*, 2(1), 23-30.
148. MacEachren, A. M. (1982). Choropleth map accuracy: Characteristics of the data. *Proceedings, Auto Carto 5*, 499-507.
149. MacEachren, A. M. (1982). Thematic map accuracy: The influence of enumeration unit size and compactness. Technical Papers, American Congress on Surveying and Mapping, 512-521.
150. MacEachren, A. M. (1985). Accuracy of thematic maps: Implications of choropleth symbolization. *Cartographica*, 22(1), 38-58.
151. MacEachren, A. M. & Davidson, J. V. (1987). Sampling and isometric mapping of continuous geographic surfaces. *The American Cartographer*, 14(4), 299-320.
152. McKay, C. J. (1967). The pattern map: A reliability standard for area sampling methods. *The Cartographic Journal*, 4(1), 114.
153. Maffini, G. (1988). Observations and comments on the generation and treatment of error in digital GIS data. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
154. Maling, D. H. (1968). How long is a piece of string? *The Cartographic Journal*, 5(2), 147-156.
155. Mandelbrot, B. (1967). How long is the coast of Britain? *Science*, 156, 636-638.
156. Markoff, J. & Shapiro, G. (1973). The linkage of data describing overlapping geographical units. *Historical Methods Newsletter*, 7(1), 34-46.
157. Maxim, L. D. & Harrington, L. (1983). The application of pseudo-Bayesian estimators to remote sensing data: Ideas and examples. *Photogrammetric Engineering and Remote Sensing*, 49(5), 649-658.
158. Maxim, L. D., Harrington, L. & Kennedy, M. (1981). Alternative "scale-up" estimators for areal surveys where both detection and classification errors exist. *Photogrammetric Engineering and Remote Sensing*, 47(8), 1227-1239.
159. Mead, R. A. & Szajgin, J. (1982). Landsat classification accuracy assessment procedures. *Photogrammetric Engineering and Remote Sensing*, 48(1), 139-141.
160. Merchant, D. C. (1982). Spatial accuracy standards for large scale line maps. Technical Papers of the American Congress on Surveying and Mapping, 222-231.
161. Merchant, D. C. (1987). Spatial accuracy specification for large scale topographic maps. *Photogrammetric Engineering and Remote Sensing*, 53(7), 958-61. (See also: Merchant, D. C. (1987). Spatial accuracy specification for large scale topographic

- maps. Technical Papers, American Society of Photogrammetry and Remote Sensing-American Congress on Surveying and Mapping Annual Convention, 2, 200-207.)
162. Moellering, H. (1982). The goals of the National Committee for Digital Cartographic Data Standards. *Proceedings, Auto Carto 5*, 547-554.
 163. Monmonier, M. S. (1975). Maps, distortion, and meaning (Resource Paper No. 75-4). Washington, DC: Association of American Geographers.
 164. Monmonier, M. S. (1982). Flat laxity, optimization, and rounding in the selection of class intervals. *Cartographica*, 19(1), 16-27.
 165. Morehouse, S. (1988). Experimental model of GIS. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
 166. Morrison, J. L. (1969). Control point spacing as an indicator of the accuracy of isarithmic maps. *Papers from the 29th Annual Meeting of the American Congress on Surveying and Mapping*, 187-194.
 167. Morrison, J. L. (1971). Method-produced error in isarithmic mapping (Technical Monograph No. CA-5). Washington, DC: American Congress on Surveying and Mapping.
 168. Muller, J.-C. (1977). Map gridding and cartographic errors: A recurrent argument. *Canadian Cartographer*, 14(2), 152-167.
 169. Muller, J.-C. (1987). The concept of error in cartography. *Cartographica*, 24(2), 1-15.
 170. Napton, D. E. & Luther, J. (1981). Transferring resource interpretations: Limitations and safeguards. *Proceedings, Pecora VII Symposium*, 175-186.
 171. Neumyvakin, Yu. K. & Panfilovich, A. I. (1982). Specific features of using large-scale mapping data in planning construction and land farming. *Proceedings, Auto Carto 5*, 733-738.
 172. Newcomer, J. A. & Szajgin, J. (1984). Accumulation of thematic map errors in digital overlay analysis. *The American Cartographer*, 11(1), 58-62.
 173. Newton, R. (1973). A statistical prediction technique for deriving contour maps from geophysical data. *Mathematical Geology*, 5(2), 179-189.
 174. Nichols, J. D. (1975). Characteristics of computerized soil maps. *Soil Science Society of America Proceedings*, 39(5), 927-932.
 175. Openshaw, S. (1988). Learning to live with errors in spatial databases. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
 176. Ottawa, T. (1987). Accuracy for digitizing: Overlooked factor in GIS operations. *GIS '87*, 295-299.
 177. Pearson, R. (1988). Approaches to detecting and correcting areas of unreliability in geographic databases. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
 178. Peucker, T. K. (1976). A theory of the cartographic line. *International Yearbook of Cartography*, 16, 134-143.
 179. Peucker, T. K. (1980). The impact of different mathematical approaches to contouring. *Cartographica*, 17(2), 73-95.
 180. Prisley, S. P. & Smith, J. L. (1987). Using classification error matrices to improve the accuracy of weighted land-cover models. *Photogrammetric Engineering and Remote Sensing*, 53(9), 1259-1263.
 181. Rattray, M., Jr. (1962). Interpolation errors and oceanographic sampling. *Deep-Sea Research*, 9, 25-37.
 182. Rhind, D. W. (1971). Automated contouring: An empirical evaluation of some differing techniques. *The Cartographic Journal*, 8(2), 145-158.
 183. Rhind, D. W. (1975). A skeletal overview of spatial interpolation techniques. *Computer Applications*, 2(3/4), 293-309.

184. Robinson, J. E. (1973). Frequency analysis, sampling, and errors in spatial data. In J. C. Davis & M. J. McCullagh (Eds.), *Display and Analysis of Spatial Data* (pp. 78-95). London: Wiley.
185. Robinson, J. W., Gunther, F. J. & Campbell, W. J. (1983). Ground truth sampling and Landsat accuracy assessment (Document N83-26161). National Technical Information Services.
186. Robinson, V. B. (1984). Modeling inexactness in spatial information systems. *Modeling and Simulation*, 15, 157-161.
187. Robinson, V. B. & Strabler, A. H. (1984). Issues in designing geographic information systems under conditions of inexactness. 1984 Machine Processing of Remotely Sensed Data Symposium, 198-204.
188. Roller, N. E. G. (1984). Effective integration of data sources for optimizing resource inventories. Eighteenth International Symposium on Remote Sensing of Environment, 1577-1603.
189. Rosenfield, G. H. (1971). On map accuracy specifications: Part II. Horizontal accuracy of topographic maps. *Surveying and Mapping*, 31(1), 60-64.
190. Rosenfield, G. H. (1981). Analysis of variance of thematic mapping experiment data. *Photogrammetric Engineering and Remote Sensing*, 47(12), 1685-1692.
191. Rosenfield, G. H. (1986). Analysis of thematic map classification error matrices. *Photogrammetric Engineering and Remote Sensing*, 52 (5), 681-686.
192. Rosenfield, G. H. & Fitzpatrick-Lins, K. (1986). A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 52(2), 223-227.
193. Rosenfield, G. H., Fitzpatrick-Lins, K. & Ling, H. S. (1982). Sampling for thematic map accuracy testing. *Photogrammetric Engineering and Remote Sensing*, 48(1), 131-137.
194. Rosenfield, G. H. & Melley, M. L. (1980). Applications of statistics to thematic mapping. *Photogrammetric Engineering and Remote Sensing*, 46(10), 1287-1294.
195. Saalfeld, A. (1988). Census Bureau research concerns for accuracy of spatial data. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
196. Shepard, D. S. (1984). Computer mapping: The SYMAP interpolation algorithm. In G. L. Gaile & C. J. Willmott (Eds.), *Spatial statistics and models* (pp. 133-145). Dordrecht: Reidel.
197. Smith, G. R. & Honeycutt, D. M. (1987). Geographic data uncertainty, decision making and the value of information. *GIS '87*, 300-312.
198. Smyth, S. (1988). Implicit positional accuracy in a large online spatial database: Data model translation project. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
199. Star, J. (1988). Sources of error in thematic classification of remotely sensed data. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
200. Stearns, F. (1968). A method for estimating the quantitative reliability of isoline maps. *Annals of the Association of American Geographers*, 58(3), 590-6w.
201. Stoms, D. (1987). Reasoning with uncertainty in intelligent geographic information systems. *GIS '87*, 693-700.
202. Story, M. & Congalton, R. G. (1986). Accuracy assessment: A user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52(3), 397-399.
203. Switzer, P. (1975). Estimation of the accuracy of qualitative maps. In J. C. Davis & M. J. McCullagh (Eds.), *Display and analysis of spatial data* (pp. 1-13). London: Wiley.

204. Switzer, P. (1988). Precision of interpolated maps. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
205. Tleobald, D. M. (1988). Accuracy and bias issues in surface representation. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
206. Thomas, I. L. & Allcock, G. McK. (1984). Determining the confidence level for a classification. *Photogrammetric Engineering and Remote Sensing*, 50(10), 1491-1496.
207. Tompson, M. M. (1973). Standards for modern maps. *Proceedings of the American Congress on Surveying and Mapping Fall Convention*, 140-147.
208. Thompson, M. M. (1971). On map accuracy specifications: Part I. Practical interpretation of map-accuracy specifications. *Surveying and Mapping*, 31(1), 57-60.
209. Tikunov, V. S. (1986). Some issues in the modeling approach to cartography. Unpublished manuscript.
210. Tobler, W. (1988). Frame independent spatial analysis. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
211. Tryon, T. C., Hale, G. A. & Young, H. E. (1955). Dot gridding air photos and maps. *Photogrammetric Engineering*, 21(5), 737.
212. Turk, G. (1979). The GT index. *Remote Sensing of Environment*, 8(1), 65-75.
213. Veregin, H. (1987). Error modeling in geographic information systems: A review. Unpublished manuscript.
214. Veregin, H. (1988). Error modeling in geographic information systems. Paper presented at the Canadian Association of Geographers Annual Meeting, Halifax, Nova Scotia.
215. Veregin, H. (1988). Error modeling for the map overlay operation. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
216. Veregin, H. (1989). A review of error models for vector to raster conversion. *The Operational Geographer*, 7(1), 11-15.
217. Vitek, J. D. & Richards, D. G. (1978). Incorporating inherent map error into flood-hazard analysis. *Professional Geographer*, 30(2), 168-173.
218. Vitek, J. D., Walsh, S. J. & Gregory, M. S. (1984). Accuracy in geographic information systems: An assessment of inherent and operational errors. *Proceedings, PECORA IX Symposium*, 296-302.
219. Vonderohe, A. (1988). An experimental field lab for testing accuracies of spatial data bases. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
220. Walsh, S. J. (1988). Inherent and operational error within a GIS: Evaluation from a user perspective. Paper presented at First Specialist Meeting on Accuracy of Spatial Data Bases, National Center for Geographic Information and Analysis, Montecito, California.
221. Walsh, S. J., Lightfoot, D. R. & Butler, D. R. (1987). Recognition and assessment of error in geographic information systems. *Photogrammetric Engineering and Remote Sensing*, 53(10), 1423-1430. (See also: Walsh, S. J., Lightfoot, D. R. & Butler, D. R. (1987). Assessment of inherent and operational errors in geographic information systems. *Technical Papers, American Society of Photogrammetry and Remote Sensing-American Congress on Surveying and Mapping Annual Convention*, 5, 24-35.)
222. Webster, R. (1968). Fundamental objections to the 7th approximation. *Journal of Soil Science*, 19(2), 354-366.
223. Webster, R. & Beckett, P. H. T. (1968). Quality and usefulness of soil maps. *Nature*, 219, 680-682.
224. Welde, M. (1979). Spatial quantification of maps or images: Cell size or pixel size implications. *Joint Proceedings of the American Society of Photogrammetry and American Congress on Surveying and Mapping Fall Meeting*, 45-51.

225. Welde, M. (1982). Grid cell size in relation to errors in maps and inventories produced by computerized map processing. *Photogrammetric Engineering and Remote Sensing*, 48(8), 1289-1298.
226. White, D. (1978). A design for polygon overlay. In G. Dutton (Ed.), *Harvard Papers on Geographic Information Systems*, 6. Cambridge, MA: Laboratory for Computer Graphics and Spatial Analysis, Harvard University.
227. White, M. (1983). Tribulations of automated cartography and how mathematics helps. *Cartographica*, 21(2/3), 148-159. (See also: White, M. (1983). Tribulations of automated cartography and how mathematics helps. *Proceedings, Auto Carto 6*, 408-418.)
228. Willmott, C. J., Rowe, C. M. & Philpot, W. D. (1985). Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *The American Cartographer*, 12(1), 5-16.
229. Yoeli, P. (1984). Error bands of topographical contours with computer and plotter (Program KOPPE). *Geo-Processing*, 2(3), 287-297.