Recognition and Assessment of Error in Geographic Information Systems

Stephen J. Walsh
Department of Geography, University of North Carolina, Chapel Hill, NC 27514
Dale R. Lightfoot
Department of Geography, University of Colorado, Boulder, CO 80309
David R. Butler
Department of Geography, University of Georgia, Athens, GA 30602

ABSTRACT: Computer-generated maps, manipulated through geographic information systems (GIS), are powerful tools for analyzing complex spatial interactions. Depending upon the level of error inherent in the source data and the error operationally produced through data capture and manipulation, GIS products may possess significant amounts of error. Errors inherent in (1) land-cover maps derived through classification of Landsat digital data; (2) slope-angle and slope-aspect information derived from digital terrain tapes; and (3) soil-type data acquired from USDA Soil Conservation Service (SCS) soil survey reports were assessed. These data were compared to field information collected for 35 sample sites distributed throughout a four-square-mile test area. A 2.5-acre (100-sq.m) and 10.0-acre (200-sq.m) resolution cell were utilized to capture data using both the cell midpoint and cell-dominant methods of encoding. Inherent error ranged from 43 percent for the 2.5-acre (100-sq.m) land-cover map to 83 percent for the 10.0-acre (200-sq.m) slope-angle map. Operational error was also examined. When two or more data layers were combined, errors were found to range from 13 to 29 percent. Combined inherent and operational error ranged from 71 to 83 percent. Smaller cell sizes produced lower inherent and operational error levels, as did the use of the cell dominant area method of data capture. Error significantly increased when more than two data layers were analyzed in concert.

INTRODUCTION

ENVIRONMENTAL ANALYSIS is hampered by a host of informational problems. Lack of data, deficiencies in the quality of data, and incompatibility of data derived from different sources cause obvious difficulties in land management. Geographic information systems (GIS) technology can be successfully employed to lessen data integration problems and the time-consuming process of synthesizing tremendous amounts of information for problem analysis. In a GIS, one can convert analog information into digital data, and then edit, store, manipulate, and display the data as maps or color images for use in a variety of resource management applications (Walsh, 1985). Applications in modeling of regional evapotranspiration, hydrology, natural hazards, land cover, and wildlife habitat have been reported (Butterfield and Key, 1986; Cibula and Nyquist, 1987; More *et al.* 1984; Walsh and Gregory, 1985; Walsh and Stadler, 1983; Walsh *et al.*, 1984).

Two sources of error, inherent and operational, contribute to reduction in accuracy of the products that are generated by geographic information systems. Inherent error is the error present in source documents. Operational error is produced through the data capture and manipulation functions of a GIS.

This paper assesses errors inherent in (1) land-cover maps derived through the classification of Landsat digital data; (2) slope-angle and slope-aspect information derived from digital terrain tapes; and (3) soil-type data acquired from USDA Soil Conservation Service (SCS) county soil survey reports. A statistical analysis of map error is presented to show the error inherent in each of the three source data sets, and potential operational errors which may be produced through data manipulation within the GIS. The principal objective of this research is to document inherent sources of error in data types commonly utilized in resource management applications, and to trace the operational error produced through data manipulation scenarios involving data overlays, composite mapping, and model development.

To accommodate the accuracy assessment, field data were collected at sample sites randomly distributed throughout a four-square-mile test area. Both inherent error and operational error, introduced through the manipulation of data layers containing various amounts of inherent error, were compared to the field data in order to assess error.

BACKGROUND

Vitek *et al.* (1984) discussed the occurrence of inherent and operational errors in geographic information systems. They provided examples of inherent error and stressed the need for developing error statements for data contained within geographic information systems. They also provided examples of operational error present in simple products and posed the question "can we account for the error at various stages in the development of the final GIS product?" They concluded by stating that the integration of data from different sources and in different original formats (e.g., points, lines, and areas), at different original scales, and possessing inherent errors can yield a product of questionable accuracy. They suggested that every map will contain inherent error based upon the nature of the source map projection, map construction techniques, and symbolization of the data. They noted that operational error is introduced during data entry, data manipulation, data extraction, and data comparison within the GIS. Such error may result in error-filled maps which fail to impart the information intended or, even worse, mislead the user (Robinson and Jackson, 1985; Burrough, 1986). The results of empirical testing of the issues identified by Vitek *et al.* (1984), and discussed independently by Burrough (1986), are discussed in this paper.

Marble and Peuquet (1983) report that the accuracy of a GIS-derived product is dependent on characteristics inherent in the source products and on user requirements, such as scale of the desired output product and method and resolution of data encoding (vector or raster). They suggest that positional accuracies are of critical importance to many users (see also Stezhenskaya (1987)). Newcomer and Szajgin (1984) asserted that the highest accuracy of any GIS output product can only be as accurate as the least accurate data plane of information involved in the analysis. The final product, they note, will be less accurate than any of the individual layers utilized. Burrough (1986) suggests that results from some mathematical models in a GIS may have such error margins as to be useless for specific applications requiring stringent levels of accuracy. Manipulation of thematic

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 53, No. 10, October 1987, pp. 1423–1430.
0099-1112/87/5310–1423$02.25/0

overlays within the GIS to derive model variables are susceptible to inherent and operational errors.

Story and Congalton (1986) state that in order to aid the GIS user in assessing the inherent accuracy of source documents, accuracy statements or error evaluation matrices should be attached to the data. Other pertinent information should include the date of data compilation, method of data collection including sampling approach and frequency, format of the data, geographic coverage of the data, and name and address of the group responsible for data formulation. According to Mead (1982), quality of data within the GIS is affected by the age of data, areal coverage, source map scale, source map resolution, format, accessibility of the data, costs of data acquisition, degree of modification from the source data, and data accuracy.

Users should be further cognizant of the spatial resolution involved in data encoding within the GIS, particularly the level of generalization in data capture and manipulation within a grid-cell data structure. The imposition of an arbitrary rectangular coordinate grid overlaid on the source document adds a second level of generalization to the already generalized original map source. In addition, the common practice of interpolating, from irregularly distributed sample points to nodes of a regular matrix for contouring in cartesian x and y coordinates, can introduce considerable error in data representation (Willmott *et al.*, 1985). Enumeration unit size, enumeration unit compactness, and variability of the distributions mapped are significant factors in determining choropleth map accuracy for data representation assessment (Maceachren, 1985).

Landsat digital data have been utilized for land-cover classification with mixed degrees of success (Walsh, 1985; Karaska *et al.*, 1986). The accuracy of Landsat digital classification can be assessed through use of error matrices (Rosenfield, 1986; Congalton *et al.*, 1983). Geometric registration of Landsat data to map projections has become a more routine procedure in digital classification. A geometric error of generally less than one-half of a pixel is deemed acceptable (Ford and Zanelli, 1985). As few as four ground control points may be sufficient to fit full- and quadrant-sized Landsat Thematic Mapper data scenes to the Universal Transverse Mercator map coordinate system with subpixel accuracies (Welch *et al.*, 1985). Surface complexity, level of detail sought, resolution of the source data, seasonality of the data, processing sophistication, and amount and quality of the ground control all contribute to the success of land cover classification.

Digital terrain and soils data are also known to possess inherent errors. The accuracy of digital terrain information is a function of the size of the sampling interval in relation to the variability of the surface (Burrough, 1986). Nichols (1975) compared computer-generated soils maps represented by the cell-dominant area method to detailed soils maps. He reported that, as cartographic detail on the original soil map increased, less accuracy existed between the original map and the computerized version. Cell size had an inverse relationship with data accuracy in characterizing actual soil conditions. Cell-dominant area is the process of assigning a landscape attribute to a cell based upon the largest area within that cell characterized by a specific landscape type or condition.

METHODS

Depending upon the level of error inherent in the source maps used to build a GIS and those errors produced operationally through digitizing, manipulation, and human errors, some GIS products may possess very high levels of error. Any decisions based on such products would thus be flawed. In order to quantify errors inherent in computer generated land cover, slope angle and aspect, and soil-type information, digital data sets of each of the three variables were produced and aligned as thematic overlays. To accomplish this task, map values of each of the three overlays were compared to field observations made within a four-square-mile area located in north-central Oklahoma (Figure 1). The basic objectives of the field data collection phase of this research were to identify control points within the study area; determine their position and landscape attributes; process disparate data to derive thematic overlays of land cover, soils, and terrain; and compare spatial and thematic characteristics of the control points to the processed data sets.

The study area exhibits sufficient physical diversity to adequately evaluate the spatial variations of the selected feature attributes. Characteristics of the study site include an undulating topography, clay to sandy-clay soils, and a blend of Post-Oak/Blackjack-Oak forest, short-grass prairie, and cultivated land. Thirty-five field sites were positioned to gather ground information (Figure 1). Points were randomly located by using a random numbers table to generate 35 pairs of x and y coordinates. These points were plotted onto a mylar grid divided into a 50 by 50 matrix. The boundaries of the matrix corresponded to the study area delineated on a 1:24,000-scale topographic map of the vicinity. The 35 study sites were located in the field by first locating the points on a 1:3,600-scale aerial photograph of the study area. A zoom transferscope and radial-line triangulation were used to position the points on the August, 1982 aerial photography. Historical aerial photography showed that no significant changes in land cover have occurred within the study area since 1982. The 35 points were located in the field through use of a transit. Numbered flags were used to indicate the position of the sample sites for subsequent data collection. The transit was used for measuring azimuths and a 100-foot tape was used to determine distances to each control point. Initial reference points, used to locate all 35 control points, were determined from the large scale aerial photography of the study area. At least two back-azimuth measures were made to known permanent surface features from each control point.

Land cover, terrain orientation (slope angle and aspect), and

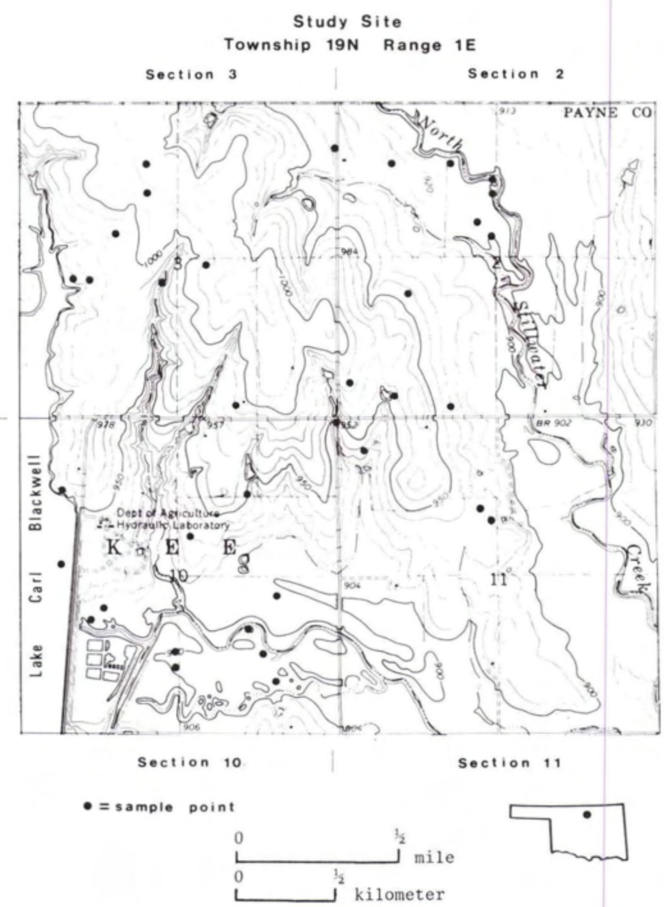

FIG. 1. Study area location, central Oklahoma.

soil-type were recorded at each of the 35 sample points. A Brunton compass was used to measure slope angle and slope aspect within a 3.0-metre radius around the sample site. Land-cover conditions were recorded within the same measurement radius. Soil-type information was collected at each site through soil coring and classification of the soil-type by the USDA Soil Conservation Service. The soil type found at each point was recorded regardless of whether it was a major mapping unit or only an inclusion soil within a mapping unit. Although soil inclusions are acknowledged and described within the SCS soil survey reports, their exact boundaries are not shown on SCS soil survey maps.

THEMATIC MAP OVERLAYS

An April, 1981 Landsat MSS digital data set was acquired for land-cover classification of the study area. An unsupervised classification technique was used to produce a map which portrayed water, cropland, exposed soil, grassland, sparse forest (less than 50 percent crown closure), and dense forest (greater than 50 percent crown closure). The digital data were geographically referenced to the Universal Transverse Mercator coordinate system prior to classification. The land-cover data were aggregated to a 2.5-acre (100 sq.m) cell and 10.0-acre (200 sq.m) cell to correspond to the two spatial resolutions under investigation. Data aggregation levels were achieved through resampling of the data matrix derived from the initial Landsat data set (Figure 2). Output from the classification was printed, on an electrostatic printer/plotter, at the same scale (1:24,000) as

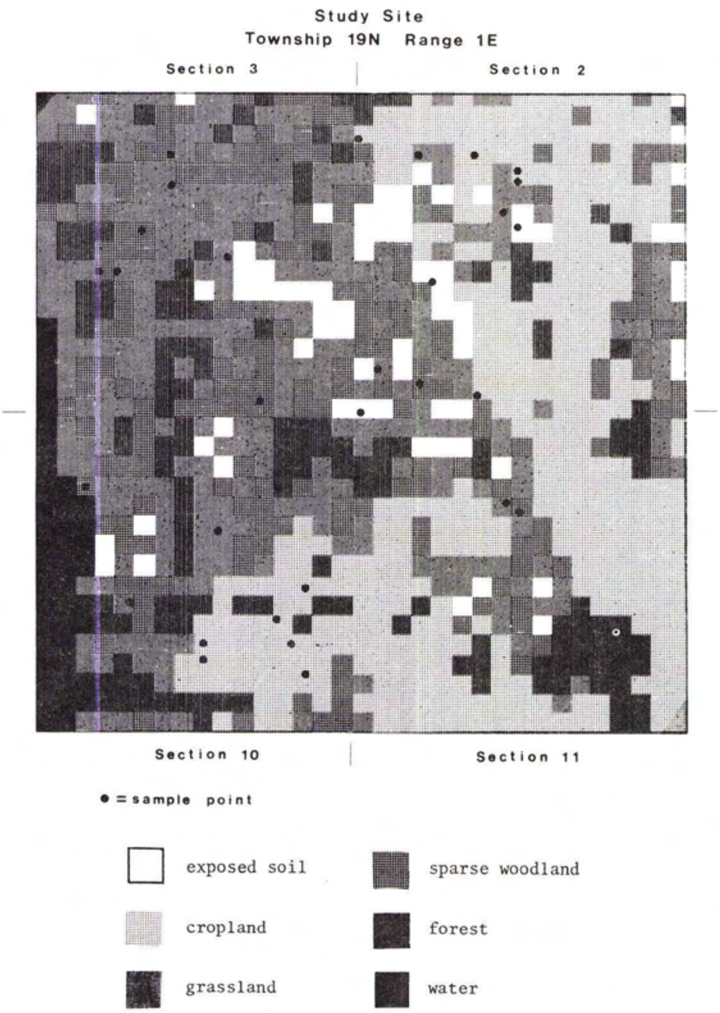

FIG. 2. Landsat derived land-cover map, 2.5-acre (100-sq.m) cell resolution.

the U.S. Geological Survey (USGS) topographic base map. Land-cover information was determined from the plots for the 35 study sites by superimposing the land-cover plots onto the 1:24,000-scale topographic map where the sample points had previously been delineated.

Digital terrain data from the USGS (1:250,000-scale source document) was utilized to derive slope-angle and slope-aspect values for the 35 field sites. Digital elevation models derived from 1:24,000-scale source documents were not available for the study area, nor are they available for significant portions of the United States. Given, therefore, that most land planners would, at best, have 1:250,000-scale source data available, use of this scale was believed likely to mirror "real-world" uses of digital terrain data.

The 1:250,000-scale source data are produced by interpolating elevations at intervals of 3 arc-seconds from contour lines from USGS 1:250,000-scale topographic maps. Three seconds of arc represents approximately 90 metres in the north-south direction and a variable amount in the east-west direction (approximately 45 metres at this study site in north-central Oklahoma). The accuracy of the digital elevation models is partially dependent upon the scale of the map utilized for digitizing the elevation data (Elassal and Caruso, 1983).

The digital terrain data were used to produce separate maps for slope angle and slope aspect. Maps were produced at grid-cell resolutions of 2.5 acres (100 sq.m) and 10.0 acres (200 sq.m) through matrix resampling of the source data. Slope increments of 1.0 percent and eight primary compass directions were used to characterize the topography of the study area. Slope angles ranged from 0 to 6 percent.

The terrain information was output to the printer/plotter at a scale of 1:24,000 for overlay onto the topographic base map. Slope angle and aspect information was extracted from the plots for the 35 sites. Data were aggregated to 100-sq.m and 200-sq.m cells, respectively (Figures 3 and 4). Note that the steeper slopes were eliminated through data aggregation (Figure 4). A loss of sensitivity in characterizing slope aspect when data were aggregated from the 100-sq.m to the 200-sq.m cell resolution was also noted.

Detailed soils maps were manually encoded by visually selecting the dominant soil based upon areal extent within a cell and the soil type occurring at the midpoint within each cell, and then entering the information into a computer file using a graphic digitizer (Nichols, 1975). This process often introduces error by reducing the spatial detail and number of soil mapping units and changing the configuration of soil unit boundaries from the original map.

A 1:20,000-scale county soil survey sheet of the study area was photographically reduced to a scale of 1:24,000 (the scale of the topographic base map). Soils information, recorded for the 35 field points, included soil-type at each sample point for the 2.5-acre (100 sq.m) and the 10.0-acre (200 sq.m) area plotted by both the cell midpoint and cell dominant area methods. These were plotted as individual thematic overlays. Data aggregation was accomplished through a data resampling scheme in which a cell size of user specification is passed through the original data, yielding a derived data set of the desired spatial resolution.

INHERENT ERROR ASSESSMENT

In order to assess the errors inherent in each of the thematic data bases, a series of 11 accuracy evaluation matrices was produced by comparing field data to the digital data sets of land cover, terrain, and soils, respectively. Table 1 presents an accuracy matrix for land cover as derived from Landsat digital data. Overall accuracy was calculated by dividing the sum of the entries that form the major diagonal (number of correct responses) by the number of samples taken (Table 2).

Accuracy decreased as spatial resolution became more coarse. There was also a drop in accuracy when encoding map data by grid cell midpoint. Out of a combined 175 sample points tested

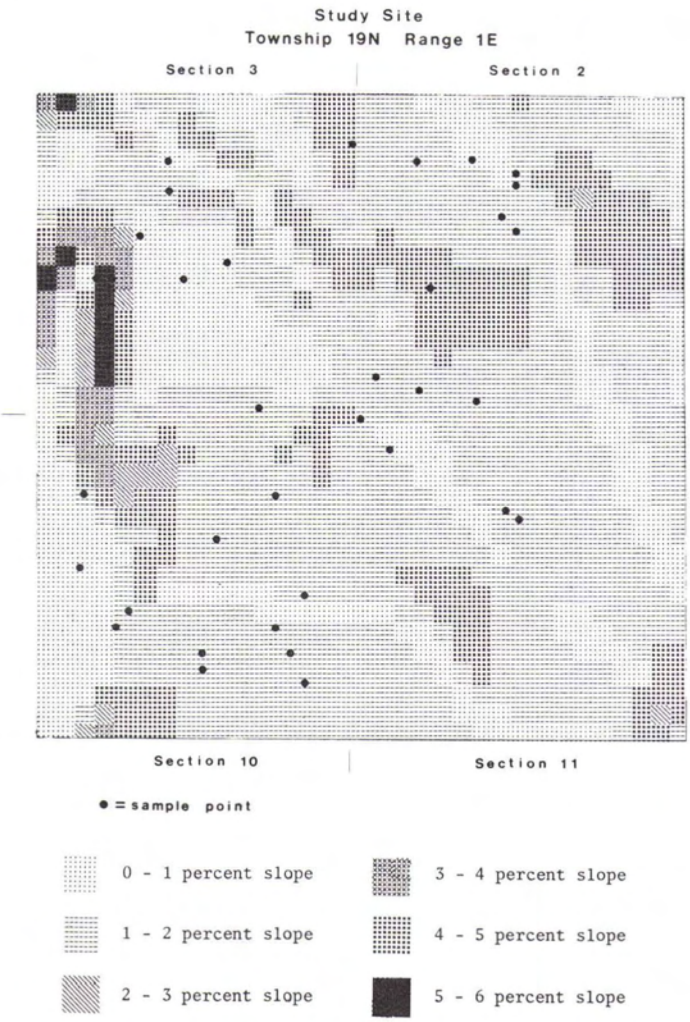

FIG. 3. Digital terrain derived slope-angle map, 2.5-acre (100-sq.m) cell resolution.

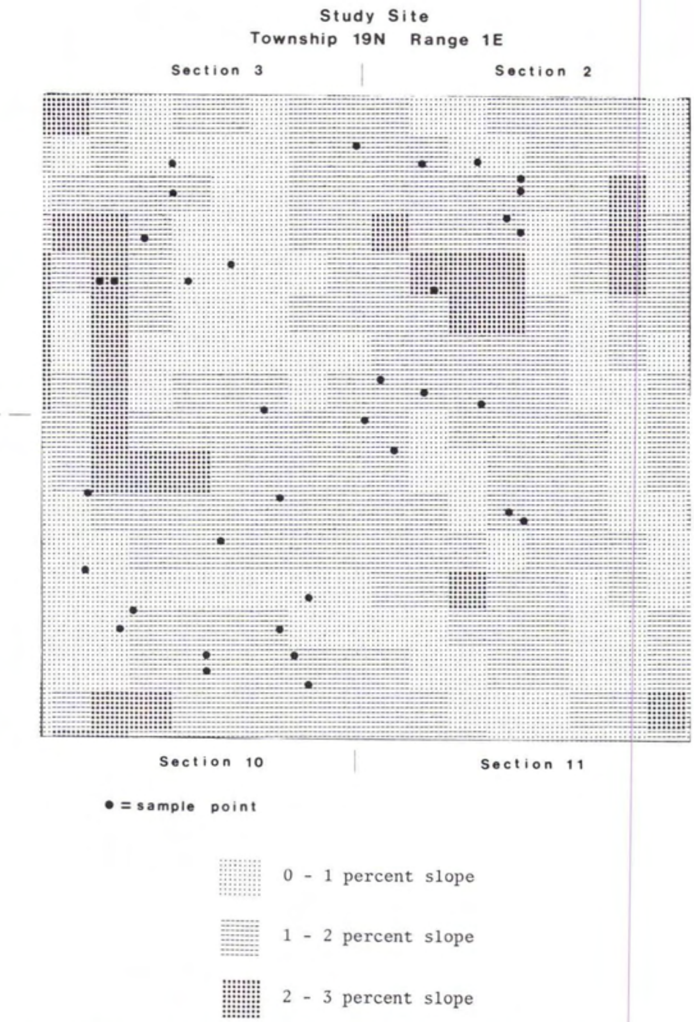

FIG. 4. Digital terrain derived slope-aspect map, 10.0-acre (200-sq.m) cell resolution.

on all four thematic map overlays (land cover, soil-type, slope aspect, and slope angle), 68 points (39 percent) were correctly classified on the 2.5-acre (100-sq.m) resolution maps (cell midpoint and cell dominant area). By contrast, only 56 of 175 points (32 percent) were correctly classified on the 10.0-acre (200-sq.m) resolution maps (cell midpoint and cell dominant area). This drop in accuracy (7 percent) was attributed to a decrease in spatial resolution of the encoded data. Comparison of cell dominant area and cell midpoint accuracies was performed only with the soils data aggregated at both the 2.5-acre (100-sq.m) and the 10.0-acre (200-sq.m) spatial resolutions. Only 27 of 70 sample points (39 percent), representing the cell dominant area method of data encoding, were correctly assigned. Only 22 of 70 points (29 percent) of the sample points were correctly assigned through the cell midpoint method of data encoding. There was an average drop in accuracy of 10 percent when using the cell midpoint method of data capture.

Landsat land-cover maps proved to have less inherent error than any of the other overlays, partly due to the digital nature of the source document and the relative lack of landscape complexity within the study area. Slope-angle maps had the poorest accuracies of any of the four thematic map overlays because of the coarseness of the original source product. In all but one case, smaller grid cell size (i.e., 2.5 acre) resulted in higher accuracies of encoding when compared to larger grid cell sizes. Accuracy increased an average of 7 percent when smaller grid cells were used.

During field work, data were recorded at specific points. One of the major problems with misclassification resulted from data aggregation in which, for example, a point recorded in the field as grassland, located adjacent to a large concentration of forest, was categorized as forest because of the relatively coarse spatial resolution of the cells. Grid cells had to be included or excluded in their entirety, because no spatial information existed at the sub-grid cell level (Crapper, 1984). Small irregular-shaped features can be more effectively portrayed through the vector encoding approach, given certain resolution stipulations.

TABLE 1. LAND-COVER ACCURACY MATRIX; 2.5 ACRE ($100m^2$) SPATIAL RESOLUTION. [FIELDCHECKED LAND COVER]

	Exposed Soil	Cropland	Range	Sparse Woodland	Forest	Water
Exposed Soil	1	2				
Cropland		5		2	3	
Range		3	5	1		
Sparse Woodland			4	4		
Forest					4	
Water						1

20 of 35 sample points correctly assigned. Overall Accuracy: 57.1%

The low levels of accuracy found to be inherent in products created from the digital terrain data likely stem from generali-

TABLE 2. SUMMARY OF THEMATIC OVERLAY ACCURACIES.

Thematic Overlay	Number of Correct Responses (%)
Land Cover (2.5 ac)	20 of 35 (57%)
Land Cover (10.0 ac)	15 of 35 (43%)
Slope angle (2.5 ac)	8 of 35 (23%)
Slope angle (10.0 ac)	6 of 35 (17%)
Slope aspect (2.5 ac)	12 of 35 (35%)
Slope aspect (10.0 ac)	14 of 35 (41%)
Soil-type (2.5 ac, cm)	13 of 35 (37%)
Soil-type (10.0 ac, cm)	9 of 35 (26%)
Soil-type (2.5 ac, cd)	15 of 35 (43%)
Soil-type (10.0 ac, cd)	12 of 35 (35%)

Note: cd = cell dominant area; cm = cell midpoint area

zation of original surface data that results from the creation of these data. Relatively small scale (1:250,000) topographic maps, already generalized and therefore possessing considerable inherent error, are further generalized by sampling and digitizing points at 60-metre intervals. Additional inaccuracies result from the aggregation of the terrain information into grid cells during resampling of the data to fit user-specified spatial resolutions.

The SCS soil overlay had an accuracy of 60 percent as computed from the field sites and the encoded data from the soil surveys. If soil inclusions are not treated as distinct soil mapping units, the accuracy increases to 94 percent. Although inclusions are acknowledged within the soil surveys, they are not mapped because of their small areal extent relative to the scale of the soil maps. The soil maps exclude soil inclusions as large as four acres, constituting up to 20 percent of the total area of any soil mapping unit. Sample points "placed" within these non-mapped inclusions were, therefore, fairly common and contributed greatly to the amount of error detected within these soil overlay products.

OPERATIONAL ERROR ASSESSMENT

Operational errors may be categorized as positional errors and identification errors (Newcomer and Szajgin, 1984). Positional errors stem from inaccuracies in the horizontal placement of boundaries. Identification errors occur when there is mislabeling of areas on thematic maps. Additional possible sources of operational errors include human error in digitizing, categorizing (classification) and delineating data (boundaries), GIS algorithm inaccuracies, and human bias.

Newcomer and Szajgin (1984) found that, as the number of layers in an analysis increases, the number of possible opportunities for error increases. In order for a correct assignment to result in any particular grid cell on a GIS product, every vertically aligned cell in each data layer used in the GIS must also be correctly assigned (Figure 5). Any other combination of assignments in the data stack of thematic overlays will result in an incorrect assignment in that cell (Figure 6). Therefore, the highest accuracy possible in any GIS product can never be better than the accuracy of the least accurate individual map layer. The worst case could occur when mislabeled cells are found at different locations throughout all data layers (Figure 6) (Newcomer and Szajgin, 1984).

Various combinations of two and three layers were used to determine the amount of identification-type operational error which might be present in output products. Both 100-sq.m and 200-sq.m resolution data for the land-cover, terrain, and soils layer were analyzed. Output product accuracies were determined by counting sample points which were correctly labeled throughout each of the data layers analyzed. Theoretical upper and lower accuracy limits were computed for each GIS product according to the technique proposed by Newcomer and Szajgin (1984). The upper and lower accuracy limits of a composite map indicate the range of accuracy probabilities when analyzing two

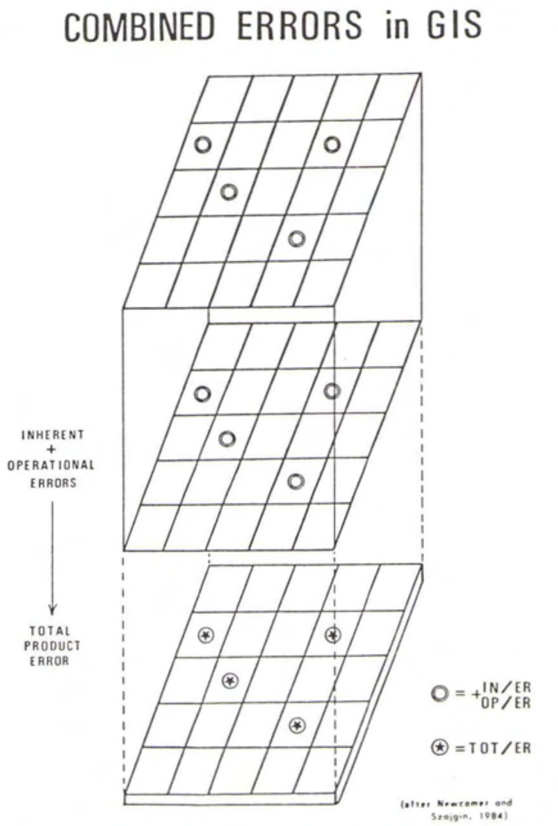

FIG. 5. GIS thematic overlays: best case combined error: Inherent error (IN/ER), operational error (OP/ER), and total error (TOT/ER).

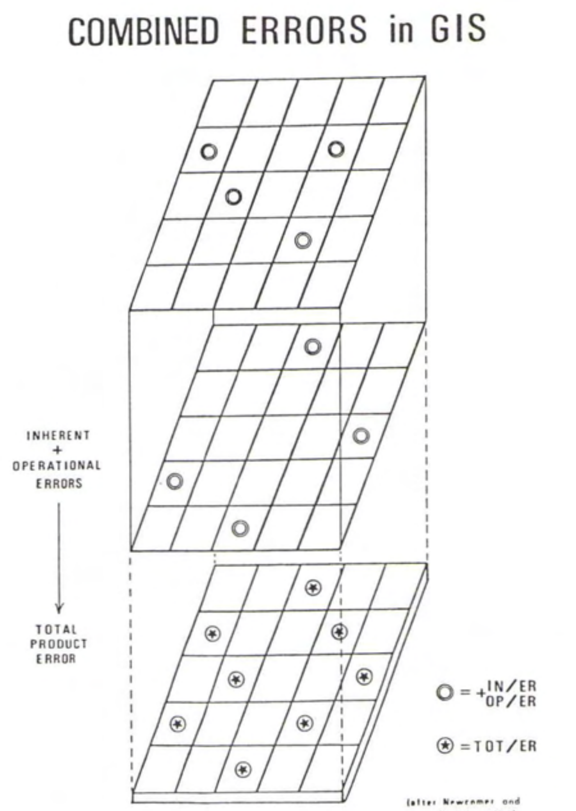

FIG. 6. GIS thematic overlays: worst case combined error: Inherent error (IN/ER), operational error (OP/ER), and total error (TOT/ER).

or more thematic overlays, given the accuracies of the source data utilized for data encoding. The lower limit is calculated as

$$1 - \sum_{i-1}^{n} \Pr(Ei), \tag{1}$$

where n = the number of layers used in the GIS; i = data layers 1, 2, 3,. . .n; and Pr (E) = the probability of error in a data layer. The upper limit is calculated as the maximum Pr (Ei), or the percent accuracy of the least accurate input or output map used in the analysis.

The most accurate output product possible, using the thematic data bases prepared for this study, was a two-layer map created by compositing the 2.5-acre (100-sq.m) spatial resolution land-cover map and the 2.5-acre (100-sq.m) cell dominate soil-type map. Only 10 of the 35 sample points on the 2.5-acre (100-sq.m) resolution output maps had both land cover and soils correctly labeled (combined accuracy of 29 percent). Upper and lower limits to accuracy were calculated as 41 percent and 3 percent, respectively (Table 3). The least accurate output map was a three-layer product derived from a combination of the 2.5-acre (100-sq.m) resolution map of land cover, slope aspect, and soil type (cell dominant method). In this instance, only 4 of 35 sample points (11 percent) were correctly classified through all three map overlays (Table 4). A 35 percent upper limit and 0 percent lower limit were calculated. Only one combination of a four-layer scheme resulted in a product having an accuracy above 0 percent.

Operational error reduces the accuracy of a GIS output from its theoretical best. A two-layer composite of 2.5-acre (100-sq.m) resolution land-cover map and 2.5-acre (100-sq.m) cell dominant soil type yielded a 13 percent increase in error as compared to the least accurate overlay in this analysis (soil overlay). An operational error of 13 percent degraded the map composite to the level of total error indicted. A three-layer data set comprised

TABLE 3. COMBINED ERROR ANALYSIS: TWO THEMATIC MAP LAYERS.

2-Layer Products	Theoretical Lower Limit	Actual Composite Map Accuracy	Theoretical Upper Limit
1. 2.5-acre (100-m^2) Land Cover and Cell Dominant Soils	2.9% accuracy (1/35)	28.6% accuracy (10/35)	41.2% accuracy (14/34)
2. 10.0-acre (200-m^2) Land Cover and Cell Dominant Soils	0.0% (0/35)	17.1% (6/35)	32.4% (11/34)
3. 2.5-acre (100-m^2) Cell Dominant Soils and Slope Aspect	0.0% (0/35)	17.1% (6/35)	35.3% (12/34)
4. 10.0-acre (200-m^2) Cell Dominant Soils and Slope Aspect	0.0% (0/35)	11.4% (4/35)	32.4% (11/34)
5. 2.5-acre (100-m^2) Land Cover and Slope Aspect	0.0% (0/35)	22.9% (8/35)	35.3% (12/34)

TABLE 4. COMBINED ERROR ANALYSIS: THREE THEMATIC MAP LAYERS.

3-Layer Products	Theoretical Lower Limit	Actual Composite Map Accuracy	Theoretical Upper Limit
1. 2.5 acre ($100\ m^2$) Land Cover, Slope Aspect, and Cell Dominant Soils	0.0% accuracy (0/35)	11.4% accuracy (4/35)	35.3% accuracy (12/34)
2. 10.0 acre ($200\ m^2$) Land Cover, Slope Aspect, and Cell Dominant Soils	0.0% (0/35)	5.7% (2/35)	32.4% (11/34)

TABLE 5. ANALYSIS OF OPERATIONAL ERROR: TWO AND THREE THEMATIC MAP LAYERS. OPERATIONAL ERROR = THEORETICAL UPPER LIMIT ACCURACY - ACTUAL COMPOSITE ACCURACY.

2-Layer Products	Operational Error
1. 2.5-acre (100-m^2) Land Cover and Cell Dominant Soils	12.6%
2. 10.0-acre (200-m^2) Land Cover and Cell Dominant Soils	15.3%
3. 2.5-acre (100-m^2) Cell Dominant Soils and Slope Aspect	18.2%
4. 10.0-acre (200-m^2) Cell Dominant Soils and Slope Aspect	21.0%
5. 2.5-acre (100-m^2) Land Cover and Slope Aspect	12.4%
3-Layer Products	**Operational Error**
1. 2.5-acre (100-m^2) Land Cover, Slope Aspect, and Cell Dominant Soils	23.9%
2. 10.0-acre (200-m^2) Land Cover, Slope Aspect, and Cell Dominant Soils	26.7%

of the 2.5-acre (100-sq.m) resolution land cover, slope aspect, and cell dominant soil-type information resulted in an operational error of 24 percent (Table 5). As Newcomer and Szajgin (1984) suggest, error increases rapidly as the number of layers used to produce the composite map increases.

CONCLUSIONS

By recognizing inherent errors within the input (source) products and the operational error created by combinations of input data, total error may be minimized. GIS output products will contain errors. Such products are properly used to make general assumptions about spatial patterns and trends rather than for site-specific applications. This will contribute to fewer user errors and, therefore, better decisions. Burrough (1986, p. 132) states that "Current practices of thematic mapping tend to hide natural variation behind a smokescreen of smoothly drawn lines and homogeneously presented mapping units." To lessen the frequency and reduce the impact of such practices, land managers and computer specialists must become more critical of the data bases assembled for inclusion into a GIS. Questions need to be asked regarding the methods of data encoding, validity of the source products, accuracy standards employed during the generation of the source document, and types and magnitudes of operationally-induced error.

GIS users also need to apply the assembled data to appropriate applications. For some uses, additional data sets may have to be generated and encoded into the GIS, while other applications may require different data structures, changes in the spatial and temporal resolutions of the encoded data, and more sophisticated analytical procedures. Through project planning and an awareness of possible problems that can be encountered during the GIS process, the land manager can obtain the quality products capable within the GIS approach.

Reliability diagrams can be employed in map legends to provide guidance regarding accuracy of analog representations of spatial data. Chrisman (1984) has suggested that the accuracy of digital data might be indexed by developing a reliability overlay that is registered to each thematic data base and integrated into the data structure, thereby indicating locations on the overlay where the user can apply the data with the greatest confidence. Also, one layer of attribute data might be used to verify another thematic overlay through logical collaboration of landscape characteristics. For example, lake surfaces should not occur in areas showing irregular topography within a digital terrain file. Control points can be employed to establish a probability of error that may occur within the thematic overlay. Figure 7 shows a hypothetical surface representing a thematic overlay. Six control points serve as the basis for assessing the accuracy of spatial location and attribute characteristics of the thematic overlay. Isolines summarize the probability of error in a linear mode outward from the control points. As distance from the control points increases, the level of uncertainty as to data quality also increases.

Additional research in inherent and operational error assessment is warranted. Research in methods for cartographic display of errors, and the setting of statistical confidence limits of data bases, are particularly required. In addition, research should be continued on resampling algorithms that permit the most effective translation of data through aggregation levels. Traditional transformation algorithms (vector to raster, and raster to vector) need to be evaluated regarding their speed and accuracy, and improvements need to be made where appropriate. Finally, universities, professional organizations, and private entities must be involved in the basic education and continued training of GIS users and computer specialists. Information regarding basic cartographic principles and spatial concepts needs to be presented as part of GIS training in order to facilitate the appropriate use of geographic information systems and the quality of data incorporated in a GIS.

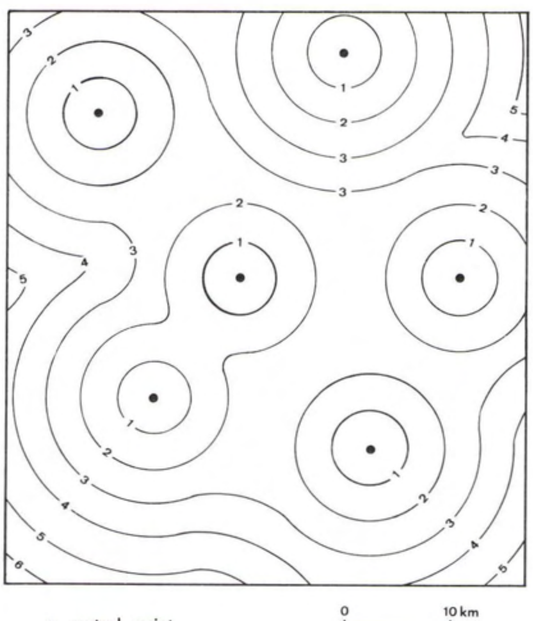

FIG. 7. Error probability isolines.

REFERENCES

Burrough, P. A., 1986. *Principles of Geographical Information Systems for Land Resources Assessment*. Clarendon Press, Oxford, 193 p.

Butterfield, B. R., and C. H. Key, 1986. Mapping Grizzly Bear Habitat in Glacier National Park Using a Stratified Landsat Classification: a Pilot Study. *Proceedings - Grizzly Bear Habitat Symposium*, USDA Forest Service General Technical Report INT-207, pp. 58–66.

Chrisman, N. R., 1984. The Role of Quality Information in the Longterm Functioning of a Geographic Information System. *Cartographica*, 21(2): 79–87.

Cibula, W. G., and M. O. Nyguist, 1987. Use of Topographic and Climatological Models in a Geographical Data Base to Improve Landsat MSS Classification for Olympic National Park. *Photogrammetric Engineering and Remote Sensing*, 53(1): 67–75.

Congalton, R. G., R. G. Oderwald, and R. A. Mead, 1983. Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques. *Photogrammetric Engineering and Remote Sensing*, 49(12): 1671–1678.

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(1): 1497–1503.

Elassal, A. A., and V. M. Caruso, 1983. *Digital Elevation Models*. Geological Survey Circular 895-B, USGS Digital Cartographic Data Standards, 3–6.

Ford, G. E., and C. I. Zanelli, 1985. Analysis and Quantification of Errors in the Geometric Correction of Satellite Images. *Photogrammetric Engineering and Remote Sensing*, 51(11): 1725–1734.

Karaska, M. A., S. J. Walsh, and D. R. Butler, 1986. Impact of Environmental Variables on Spectral Signatures Acquired by the Landsat Thematic Mapper. *International Journal of Remote Sensing*, 7(12): 1653–1667.

Maceachren, A. M., 1985. Accuracy of Thematic Maps: Implications of Choropleth Symbolization. *Cartographica*, 22(1): 38–58.

Marble, D. F., and D. J. Peuquet, 1983. Geographic Information Systems and Remote Sensing: *Manual of Remote Sensing*, 2nd edition, Falls Church, Virginia, American Society of Photogrammetry, pp. 923–958.

Mead, D. A., 1982. Assessing Data Quality in Geographic Information Systems. *Remote Sensing for Resource Management*, edited by C. J. Johannsen and J. L. Sanders, Soil Conservation Society of American, Ankeny, Iowa, Chapter 5, pp. 51–59.

More, G., O. Niemann, and G. Langford, 1984. Avalanche Information Systems in Kananaskis County, Alberta, Canada. *Proceedings, International Snow Science Workshop*, Aspen, Colorado, pp. 8–11.

Newcomer, J. A., and J. Szajgin, 1984. Accumulation of Thematic Map Error in Digital Overlay Analysis. *The American Cartographer*, 11(1): 58–62.

Nichols, J. D., 1975. Characteristics of Computerized Soil Maps. *Proceedings of the Soil Science Society of America*, 39: 927–932.

Robinson, G., and M. Jackson, 1985. Expert Systems in Map Design. *Proceedings of Auto-Carto VII*, pp. 430–439.

Rosenfield, C. H., 1986. Analysis of Thematic Map Classification Error Matrices. *Photogrammetric Engineering and Remote Sensing*, 52(5): 681–686.

Stezhenskaya, I. N., 1987. A Seminar on Geographic Information Systems. *Mapping Sciences and Remote Sensing*, 24(1): 59–62.

Story, M., and R. G. Congalton, 1986. Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing*, 53(3): 397–399.

Vitek, J. D., S. J. Walsh, and M. S. Gregory, 1984. Accuracy in Geographic Information Systems: An Assessment of Inherent and Operational Errors, *Proceedings, PECORA IX Symposium*, pp. 296–302.

Walsh, S. J., 1985. GIS for Natural Resource Management. *Journal of Soil and Water Conservation*, 40(2): 202–205.

Walsh, S. J., D. R. Lightfoot, and D. R. Butler, 1987. Assessment of Inherent and Operational Errors in Geographic Information Systems. *Proceedings, American Society for Photogrammetry and Remote Sensing - American Congress on Surveying and Mapping*, Baltimore, Maryland, 5: 24–35.

Walsh, S. J., and M. S. Gregory, 1985. Hydrologic Modeling Through Remote Sensing and Geographic Information Systems. *Proceedings, Applied Geography Conferences*. Denton, Texas, 5: Volume 109–116.

Walsh, S. J., and S. J. Stadler, 1983. Regional Evapotranspiration Modeling of Oklahoma's Little Washita Drainage Basin. *Proceedings, Ninth International Symposium, Machine Processing of Remotely Sensed Data*, West Lafayette, Indiana, pp. 389–397.

Walsh, S. J., S. J. Stadler, and M. S. Gregory, 1984. Modeling of Regional Evapotranspiration Through a Geographic Information System Approach. *Proceedings, International Symposium on Spatial Data Handling*, Zurich, Switzerland, pp. 252–266.

Welch, R., T. R. Jordan, and M. Ehlers, 1985. Comparative Evaluations of the Geodetic Accuracy and Cartographic Potential of Landsat-4 and Landsat-5 Thematic Mapper Image Data. *Photogrammetric Engineering and Remote Sensing*, 51(11): 1799–1812.

Willmott, C. J., C. M. Rowe, and W. D. Philpot, 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.