A Diagnostic Test for Error in Categorical Maps

Nicholas Chrisman and Marcus Lester
CHRISMAN@max.u.washington.edu & MARCUSL@max.u.washington.edu
Department of Geography DP 10, University of Washington
Seattle, Washington 98195 USA

ABSTRACT
A test based on exhaustive overlay of two categorical maps provides a description of error distinguished into the likely sources of that error (a diagnosis of the error). The results of the overlay are characterized by geometric, topological and attribute criteria to separate the most likely positional errors from the attribute errors. This paper applies the proposed test to a simple land cover map, which was replicated by a second interpreter. Results diagnose the positional inaccuracy and misidentifications common in such a GIS layer. Adopting this test will target the efforts of a producer's quality control functions, and it will also clarify fitness for the particular uses contemplated by others.

Preamble

A great quantity of geographic information, including maps of land use, soils, geology, property ownership and other phenomena, are represented in the form of categorical maps. A test for categorical maps is required to understand their fitness for use. Beyond a simple accuracy figure, a test should provide an indication (a diagnosis) of which component of the map might need correction or quality control attention.

For many years, cartographers, remote sensing experts and others have made do without tests or with tests that provide much less diagnostic information than a comprehensive test. The test developed here uses polygon overlay, not point sampling. Such a test is specifically mentioned in the US Proposed Standard for Digital Cartographic Data [Part III, 4.3.3].

> "4.3 Attribute Accuracy
> ... Accuracy tests for categorical attributes can be performed by one of the following methods. All methods shall make reference to map scale in interpreting classifications.
> ...
> 4.3.3 Tests based on Polygon Overlay
> A misclassification matrix must be reported as areas. The relationship between the two maps must be explained; as far as possible, the two sources should be independent and one should have higher accuracy." (Morrison, 1988, p. 133)

Despite this explicit reference in the standard, there is no complete specification for such a test. Furthermore, the standard does not discuss the diagnostic results which are possible.

This paper presents a new test for the accuracy of categorical maps. Rather than examining previously studied alternatives, this paper presents the case for the new test, using a worked example. In the conclusions, the paper will generalize beyond the specific case.

The Example

The example for this paper derives from the efforts of the Dane County, Land Records Project to generate a Soil Erosion Control Plan (Ventura, 1988). A more complete description of the project and its products has been presented in a number of publications (Chrisman and others, 1984; Niemann and others, 1987). In producing this plan, the project could rely on many layers of existing mapping, but land cover was not readily available. For the purposes of the plan, land cover requirements were relatively simple, leading to five categories:

Row Crop	Row planted crops, particularly corn and soybean
Meadow	Pasture and crops such as alfalfa and hay
Coop	Cooperator fields (in row crops or meadow)
Woods	Forested areas in rural use (not including housing)
Other	all non-rural uses plus wetlands, water, etc.

The plan needed to separate its realm of interest, rural agricultural land use, from the non-agricultural (suburban and urban). Thus, the general purpose cover category of Other included wetlands, roads, subdivisions, golf courses, industrial and commercial uses. The Woods category applies to area out of crop use – whole woodlots, not single trees. The other three categories deal with the active agricultural uses. Cooperator fields cover those areas with existing soil conservation agreements between the farmer and the conservation agencies. In any particular year, a cooperator field would be in either a row crop or meadow.

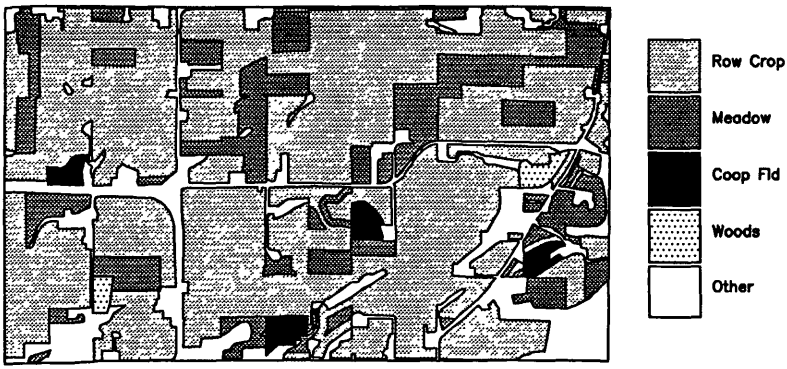
Once the conservation staff developed the categories for mapping, they had to acquire photography. The US Agricultural Stabilization and Conservation Service (ASCS) takes a color 35mm slide of each section (square mile) in Wisconsin (and many other agricultural states) each crop year to verify compliance with various federal programs. These photos are not strictly controlled photogrammetric products, but they offered color, more timely coverage and greater detail than the higher altitude products available from other sources. The project decided to use the 1982 ASCS slides to identify land cover categories and to map the results on the photographic base produced for the Dane County Soil Survey.

Map 1: Land Cover by Interpreter 1

The County staff made the maps (in pencil on prints of the photobase) and digitized them. Map 1 shows the map product for one soil sheet (the unit of original compilation). The University of Wisconsin–Madison team assisted in verifying the topological consistency and related operations. By summer 1983, one township (out of 35) was mapped. This pilot stage was used to demonstrate the capabilities required to complete the county plan. Much of the investigation dealt with the economics of data preparation and digitizing (Chrisman and others, 1984, p. 33-37).

So far, this process is uneventful. A local government group was making do, producing a map product to fill a project need without a large appropriation. This stage should also have included a test of accuracy in order to determine that the product was fit for the intended use. Most applications teams, being sure of their own work, forge ahead without such testing. It is the purpose of this paper to describe how a testing process could assist in the operational decisions of the GIS user. Developing such a test is, of course, a matter for theory and research. Much of the paper concentrates on the development of important ramifications of the test.

Map 2: Land Cover by Interpreter 2



In summer 1984, another person reproduced the interpretation, according to the same rules and using the same materials (see Map 2). The second interpreter was a graduate student with some years experience as a conservationist in a nearby county. Such a test would be most authoritative based on an independent source of higher accuracy, but that would require simultaneous acquisition of another source of photography, imagery or field reports which cannot be mobilized in retrospect.

Thus, this test began with two maps of the same scale. As a test, it provides a measure of the deviation between two trials. If both interpreters agree, it shows that the classification can be reproduced reliably. When they differ, it may be due to error on either part, but the first interpreter had somewhat more field experience with this specific area.

Polygon Overlay

Map 3 shows the result of overlaying Map 2 onto Map 1. Areas where the two interpreters agreed are not shaded, while the disagreement is dark. The individual polygon boundaries have been suppressed to permit small and thin features to show in the printed format.

Map 3: Disagreement between Interpreters

The areas of the polygons created by the overlay can be crosstabulated by the categories from the two source maps in the form of a matrix called a misclassification matrix by the Proposed Standard for Digital Cartographic Data Quality (Morrison, 1988, p.133), as in Table 1.

Table 1: Misclassification Matrix (hectares)

Interp. 2:	Row Crops	Meadow	Coop Field	Woods	Other
Interp. 1:					
Row Crops	1110.9	82.5	.9	.8	57.8
Meadow	17.5	212.3	.04	.2	35.8
Coop Fld	4.0	3.6	32.6	.03	2.0
Woods	.9	.3		10.2	2.1
Other	32.7	11.6	.1	7.5	212.3

From Map 3 and Table 1, it is clear that the two sources are in rough agreement. 85.8% of the area covered by the two maps falls into the diagonal of the matrix, meaning that the two interpreters agreed. For the purposes of the USGS program of land use and land cover mapping, 85% correct was set as a standard (Fitzpatrick-Lins, 1978). Thus, this example demonstrates a result at the low end of acceptability (though many remote sensing products fall much below this threshold). These overlay results were reported earlier (Chrisman, 1987) with a verbal interpretation of the reasons for the particular errors. This paper reports on an analytical procedure to decompose the error detected by overlay. The basis for such a test has been described in earlier publications (Chrisman, 1989a; 1989b); this paper reports actual results and some extensions which developed from this trial.

Outline of the Test Procedure

The basic theory behind the test distinguishes positional error from attribute (classification) error. The first arises from uncertainty in the location of a boundary, while the second arises from lack of agreement in the categories mapped. Some researchers deny the utility of this distinction. To them, a categorical map is too much of a fiction to merit the attention otherwise attached to map error for continuous surfaces. While some categorical maps may contain dubious elements, the political and administrative requirements for GIS continue to specify sharp distinctions in a fuzzy world. A test for categorical maps, such as the land cover example presented above, is sorely needed.

This section describes the procedure applied to separate the positional errors from the attribute errors. The following sections explain the rationale for these decisions, using the example as illustration.

Figure 1: Flow of test applied to each overlay polygon

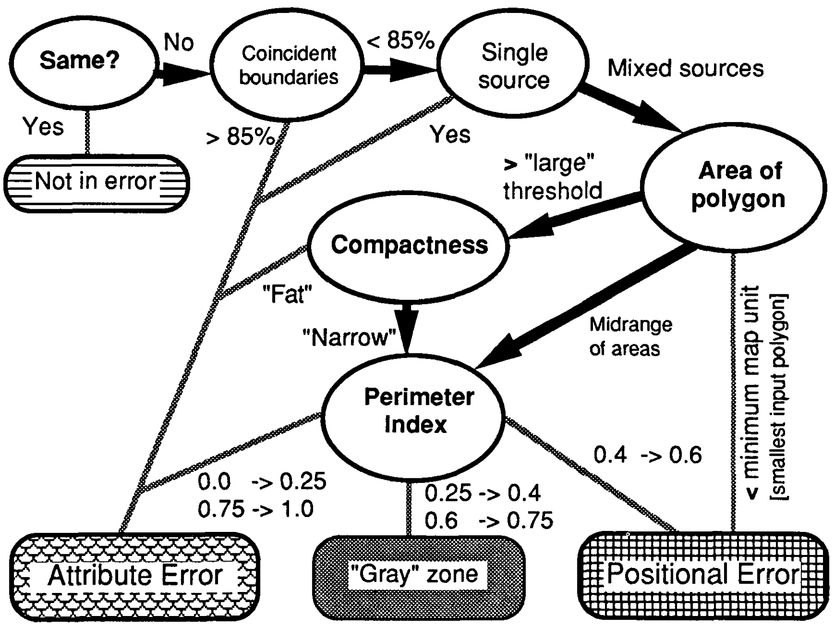


Figure 1 shows the steps involved in this test. Sequentially, each polygon is examined and fit into one of four resultant categories. There are a number of numerical parameters involved; each one may be adjusted, but the diagram shows the particular values used in this application. The first step simply decides if it represents an error. For a simple test, the attributes must be identical, in more complex cases this decision may require more information. Second, if both sources coincide to form the bulk of the polygon's boundary, then the error must be in attributes (there is not enough linework in disagreement). The parameter used here was more than 85% of the perimeter. Third, if all the non-coincident lines come from one source, it is attribute error since it must be a whole "island" of different thematic attribute, and not an area where polygons partially overlap. Fourth, polygons with area less than a "minimum map unit" are judged to be positional in nature, since they are too small to have been identified on the input layers. This threshold may be the smallest area mapped on the input layers (the procedure adopted here) or some other minimum area parameter which may be appropriate for a given test.

Fifth, for those polygons whose area is greater than a different "large" threshold (the square of a given "minimum discrimination distance" parameter), a minimum compactness value is calculated, and the compactness of the polygon is compared to this calculated minimum. This compactness value is a measure of polygon shape based upon Unwin's S_2 (1981). It is reformulated to allow a single calculation from a polygon's area and perimeter:

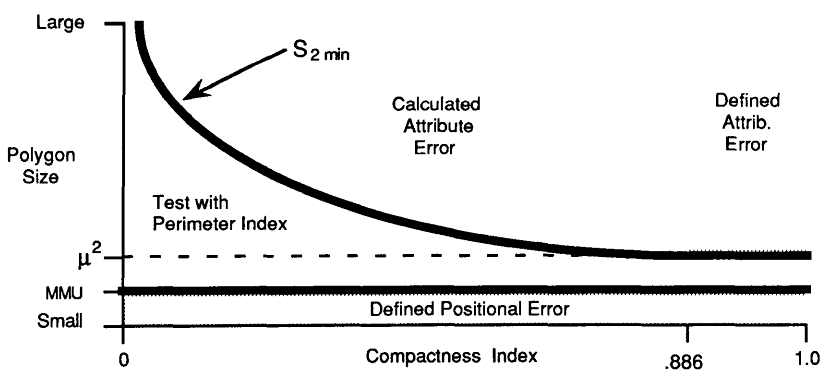
$$S_2 = 2(\pi a/p^2)^{0.5} \quad (1$$

The minimum compactness is that of a rectangle with area equal to the polygon's area and one side equal to the minimum discrimination distance:

$$S_{2\ min} = (\pi a)^{0.5} / (\mu + (a/\mu)) \quad (2$$

where μ = minimum discrimination distance (in this case μ is 1/8 inch on the original maps). Any polygon that is both larger than μ^2 and more compact than $S_{2\ min}$ is at once large enough and wide enough to have been identifiable on the input layers, so it is judged to be an attribute error. Essentially, this creates a sliding scale. Relatively less compact polygons can fall into the attribute error category if they are relatively large.

Figure 2: Classification by size and compactness.



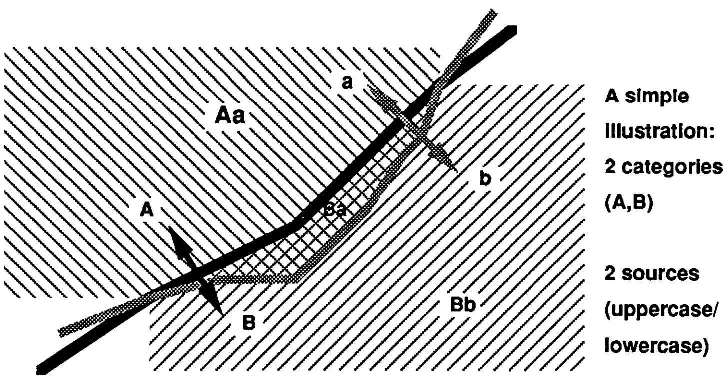
For the sixth and final stage of the test, a perimeter index is calculated for all polygons not previously classified. The perimeter index, discussed more fully below, is a ratio of the perimeter from one source to the total of the non-coincident perimeter from both sources. With this index, each remaining polygon falls into one of three categories: attribute, ambiguous (gray zone) or positional error.

Explanation of the test procedures

In applications of polygon overlay, it has long been known that "slivers" can fill up computer storage and clutter the analytical procedures (see for example, Goodchild, 1978; Cook, 1983). In this case, however, the slivers provide a clue to the origin of errors. Slivers have been identified in the past by their size and by their shape, being generally small and narrow. Narrowness is usually interpreted by human visual pattern recognition, which is difficult to quantify for complicated map features. One analytical approximation of narrowness is compactness. Compactness indices, typically ratios of perimeter to area, (see Unwin, 1981) are unreliable measures of very narrow shapes for this purpose. Because perimeter increases dramatically with line sinuosity and with inner rings of polygons, a large polygon may have a compactness index similar to a sliver.

The purpose of the test is not solely to isolate those polygons commonly called slivers. The purpose is to test the accuracy of the categorical map. Slivers are simply one form of commonly recognized error which serve as indicators of positional differences. Each of the components mentioned above; size, compactness and narrowness play a role in the proposed test, alongside a measure of "perimeter contribution". The measure of shape described in Formula 1 and 2 above serves well in distinguishing relatively compact polygons, but it is not reliable for other circumstances.

Figure 3: A typical positional error



In Figure 3, the classical sliver has some clear distinguishing characteristics. It is "small"; it is narrow; it is not compact, but these criteria are either scale-specific or could falsely identify complex features as described above. Other distinguishing characteristics might be proposed. For example, one sliver tends to engender another, in a sequence along the "true" line (Goodchild, 1978). Thus, slivers would have two possible identifying characteristics. Slivers should be topologically adjacent and the nodes at either end should be four valent (be formed by the geometric intersection of two straight lines). Both of these criteria are relatively difficult to implement for a number of reasons. First, slivers will occur at different positions, sometimes near true nodes (usually three valent for non-parcel data). It may be difficult to separate these true nodes from the sliver nodes. Second, polygon overlay is a messy business, involving the vagaries of floating point hardware (see Douglas, 1974; Dougenik, 1980). While a pure sliver might have four valent nodes at either end, the calculations of the intersection might discover a coincident section, creating two three valent nodes. Additionally, instead of random fluctuations, a long narrow sliver may be produced by a uniform (or one-sided) misinterpretation of a thematic boundary. Such a sliver may have no adjacent slivers. These same difficulties make the topological criterion difficult to manage, as well.

Another measure of slivers is required. Observing Figure 3, the linework from the two sources is nearly equal in length. In general, a sliver is an area bounded by one line from one source and a second line from the other source that are both intended to represent the same feature. In addition, other forms of positional error which do not exhibit other special characteristics of slivers also show the same balance between sources. Working backwards from the results of the test, those overlay polygons whose boundaries come from the two sources in approximately the same amounts are more likely to be positional errors. For sources of equivalent scale, the perimeter from the two sources will be very close to equality. Of course, when one source records much more detail, the perimeter may be much longer to enclose the same area. This form of cartographic texture has been related to fractal measures. In those cases, there should be more perimeter from the detailed source.

The perimeter index is a ratio which compares the perimeter from one source to the sum of the perimeters from each source (Equation 3).

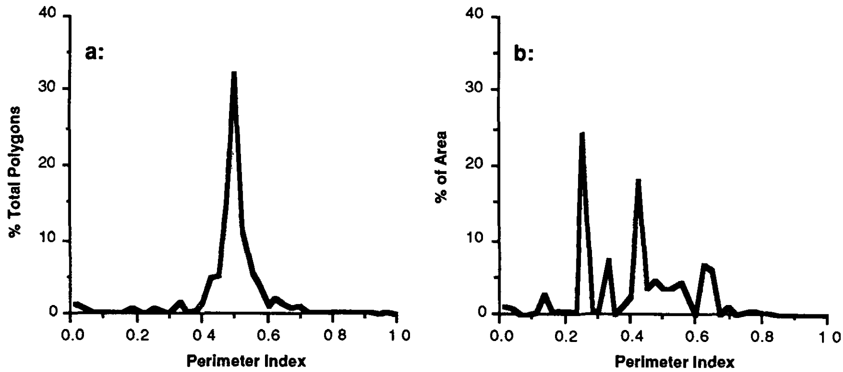
$$\text{Perimeter index} = a \,/\, (a+b) \qquad (3$$

where: a = length of chains from source A only;
b = length of chains from source B only.

In this formula, perimeter does not include those sections of a polygon's border which come from both sources (those lines which the overlay process finds to be coincident).

This index falls into the range from zero to one, with 0.5 as the result for the pure sliver (subject to the concern about scale discussed above). The index can be calculated with either map as Source A (the numerator of the ratio), which will yield values reflected around 0.5. The two versions are equivalent as long as the interpretation of the index is symmetric around 0.5. Figure 4 shows the observed distribution of the perimeter index for the land cover test presented earlier. In Figure 4 and all subsequent diagrams, the vertical axis represents percentage of the relevant total, in order to standardize the presentations.

Figure 4: Distribution of perimeter index (all polygons)
a: by percentage of count of polygons; b: by area of polygons



The pattern in Figure 4 (particularly part b) is somewhat clouded because it includes all polygons. While there is a central tendency between .4 and .6 in the number of polygons, much of the total map area comes from a few large polygons. The overlay generates 681 polygons, 106 of which are not errors. As shown in Map 3, these 106 polygons have 85.8% of the area, hence they dominate Figure 4b.

Figure 5: Distribution of index (tested polygons only)
a: by percentage of tested polygons; b: by percentage of area tested

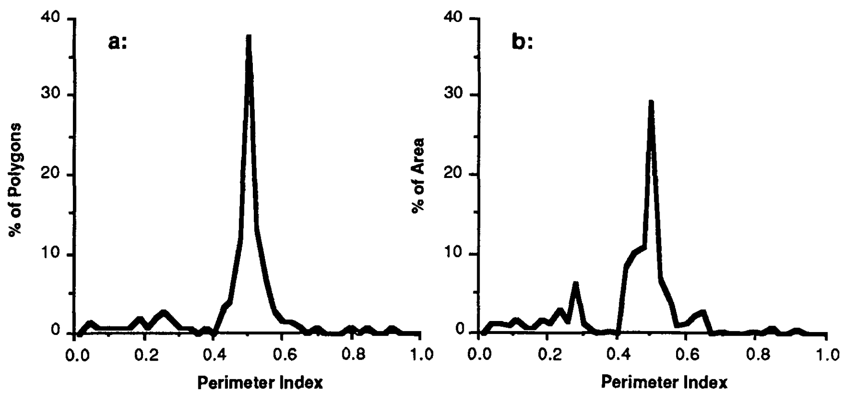
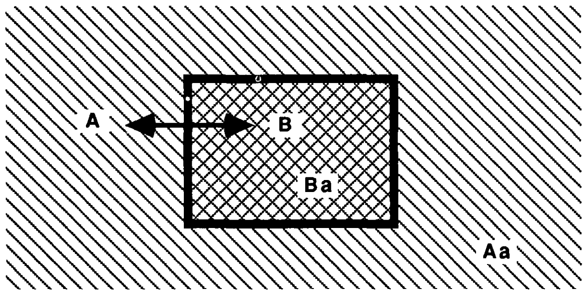


Figure 5 only tabulates the 153 polygons actually subjected to the perimeter index test, as described above. These are the polygons which are not classified by the earlier parts of the test. The largest number of overlay polygons (Figure 5a) fall near the center of the index, around 0.5. In both diagrams, the distribution is similar, and the values of .4 and .6 fall near obvious breakpoints in the distribution.

Figure 6 is a diagram of an area which is misclassified by one interpreter; the linework comes from just one source. In the diagram, the uppercase letters show a distinction on one source, but the lowercase letters show that the whole region is classified "a" in the other source. The line comes completely from one source. These errors are classified as attribute errors by the single source test (Figure 1).

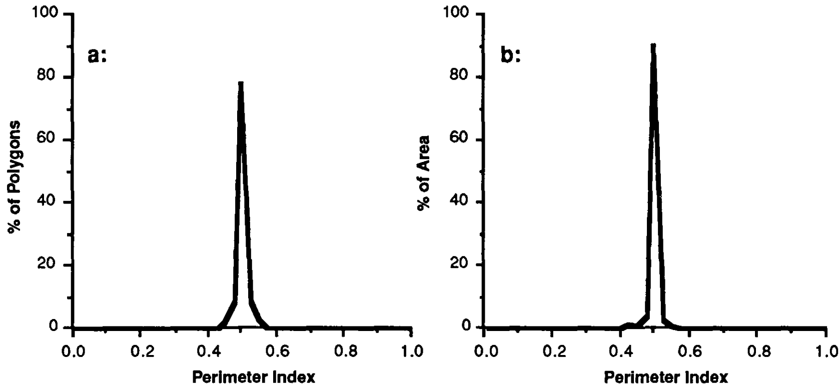
Figure 6: Attribute error (all lines from one source)



Even without a single source test, a perimeter index near zero or near one indicates an attribute-like error. However, while it is a more reliable classifier of error than the compactness index, larger polygons will occasionally, by chance, have a nearly equal proportion of perimeter from each source. Knowing where to draw the line between the ends and the positional error in the middle is not immediately obvious. One way to proceed is to introduce a known type of error, then see how the perimeter index varies.

To test the reliability of the index for classifying identification and discrimination error, positional error was introduced by translating (shifting) a map relative to itself. The original was used as the source of higher accuracy. Figure 7 shows the distribution of the perimeter index (for polygons of disagreement) after translating Map 1 south by a distance corresponding to .8 meters on the ground. This distance is quite small relative to the line width of the map, and an entirely possible registration error. These maps were digitized on a tablet with a least count resolution of about .4 meters on the ground at map scale. Variations in registration will arise from the hardware as well as the visual placement of the cursor over the registration marks.

Figure 7: Perimeter Index distribution from shifting Map 1 (distribution for all polygons (a) and area (b) in disagreement)

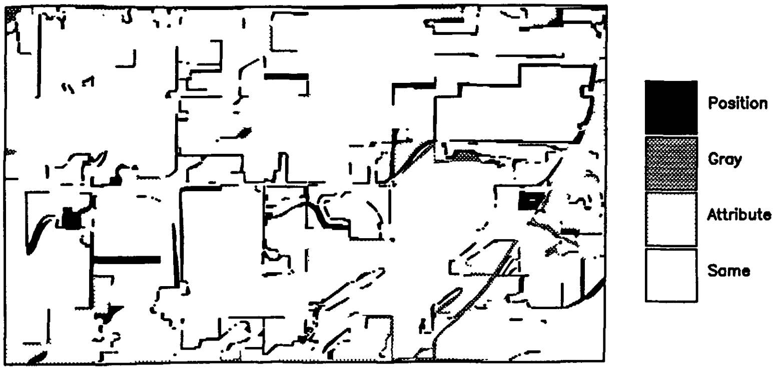


As Figure 7 shows, this purely positional error is closely packed around the theoretical value of 0.5. This indicates that the index correctly interprets the translation as a positional error.

Translation error, which would be due to misregistration or similar causes, is only one kind of positional error, but other forms of positional error also generate similar overall distributions. The result of the test of Map 1 on Map 2, shows a peak in Figure 4a much like the peak in Figure 7a, but the distribution by area in 4b is significantly different from 5b. When the restrictions are applied prior to applying the perimeter index, Figure 5a still shares the dominant central spike of 7a. Artificially pure positional error produces an unmistakable signature in the distribution of the perimeter index.

The distribution of the test results (Map 3, Table 1 and Figure 4) is the combination of all the error processes that distinguish the two sources. To decompose the polygons into two broad categories, position-like error and attribute-like error, they were analyzed for area, coincident boundary segments, compactness index and perimeter index. The compactness index is used to classify the more compact polygons, while the less compact polygons are classified with the perimeter index. While as yet there is no theory to guide the selection of a threshold between the medial and extreme values of the perimeter index, the range of values in Figure 7 and the breakpoints in Figures 4 and 5 indicate that from 0.4 to 0.6 is very likely to be positional error. Thus, for this application, any error with an index between 0.4 and 0.6 is termed positional, while those from 0 to 0.25 and from 0.75 to 1.0 are termed attribute-like. The remaining "gray" zone between 0.25 and 0.4 and between 0.6 and 0.75 (which contained 14 polygons and about 1% of the total area of the map) is ambiguous. In Map 4 it appears in a separate gray category. A symmetric set of thresholds is easily justified for a case, such a this one, of maps at the same approximate scale and level of detail. If one source has much more detailed lines, the midpoint might be biased due to the well-known effects of resolution on perimeter (Perkal, 1966).

Map 4: Positional and Attribute Errors



Map 4 shades the same polygons as Map 3, but classifies them into the position-like and the attribute-like categories. Again, the polygon borders are suppressed to show the small polygons. In general, the distinction of the two types seems reasonable, although there are some problems. A few large areas which were not due to uncertain boundaries were classified as position-like because they are classified as "narrow" by the compactness index and the perimeter contribution happened to be relatively balanced. Further refinement of the model may separate these from the more clearly position caused errors. A topological study of the chains may serve this purpose.

Table 2 shows the matrix crosstabulating the position-like error.

Table 2: Position-like Error (hectares)

Interp. 2:	Row Crops	Meadow	Coop Field	Woods	Other
Interp. 1:					
Row Crops	—	22.8	.9	.8	34.0
Meadow	11.9	—	.04	.2	10.2
Coop Fld	.5	1.2	—	.03	1.3
Woods	.9	.3		—	1.4
Other	14.5	4.1	.1	.09	—

Table 3 shows the areas of polygons classified as attribute errors. In this situation, there are many fewer attribute errors, in terms of polygons, but the total area is greater. The nature of the attribute error should be interpreted in terms of the five categories. The largest attribute cell is 59.8 hectares which interpreter #1 classed as Row Crop and #2 classed as Meadow. This error will have little impact on the soil erosion plan. The errors involving the Other category will have greater impact, but are relatively small.

Table 3: Attribute Error (hectares)

Interp. 1: \ Interp. 2:	Row Crops	Meadow	Coop Field	Woods	Other
Row Crops	—	59.8			15.2
Meadow	5.6	—			22.2
Coop Fld	3.5	2.4	—		.3
Woods				—	
Other	13.2	7.4		7.4	—

Attribute error can be decomposed directly into the cells of the table. An error between Meadow and Row Crop does not depend on the error between Woods and Other. Thus, any analysis of the attribute error relates to the particular classification and the ability to identify it on the source material.

In the spirit of a diagnostic test, there is a need to decompose more completely the error in Table 2, above. The positional error can have two systematic components which can be studied separately: bias in position – caused by misregistration, and filtering (or generalization). These two cases will be considered separately.

Translation

The experiment of translating a map against itself was described above in order to show a pure case of positional error. Such an experiment can be used as a diagnostic tool as well. By translating the map of higher accuracy, one may obtain a simulation of the amount of error which would arise from a given error. This distribution can be compared to the actual error discovered. Some goodness-of-fit procedure could be applied to pick out the best fitting translation. Error matrices are not random samples to which the usual tools of linear regression apply, but estimation tools from the toolkit of robust statistics would be the most applicable. The Least Median Square is one such method (Shyue, 1989).

Table 4: Misclassification Matrix
resulting from a .8 m South translation of Map 1 (figures in hectares)

Interp. 1: \ Interp. 2:	Row Crops	Meadow	Coop Field	Woods	Other
Row Crops	1234.2	5.9	.4	.5	12.6
Meadow	7.2	256.9	.2	.2	2.1
Coop Field	.9	.2	40.9		.2
Woods	.7	.1	.03	12.6	.2
Other	11.3	3.1	.7	.3	250.4

For this test, no such fit was examined. The error polygons showed a set of narrow polygons elongated east-west (Map 3). These could have been generated by a north-south translation error. A small translation was performed, which produced the matrix shown in Table 4. The error of the translation was completely classified as position-like (see Figure 7). The distance of .8 meters was chosen as a reasonable (and small) number which produced an error matrix whose values nearly fit the actual errors discovered (see Table 2).

The translation error can be subtracted from the total positional error to produce a matrix of those errors which could not be ascribed to a simple registration problem (Table 5). A subtraction of the attribute error is not produced because translation produces only position-like error. In this case, the ambiguous error was aggregated with the position-like error, to judge the overall effect of translation. The subtraction does not remove the large instances of error, but it removes much of the small quantities from the matrix. In some cases, the translation produces a fraction of a hectare more than observed. Such small negative numbers are, in aggregate magnitude, less than the areas which they supplant, and should not detract from the use of translation in a larger model of error. It is not proven that any particular translation occurred, but such a test could assist in discovering the amount of a misregistration in a very specific manner. The residual error in Table 5 is not strictly proportional to the error reported in Table 2. For instance the errors involving Row Crop are all reduced, due to long boundaries, but Meadow/Other is much less affected.

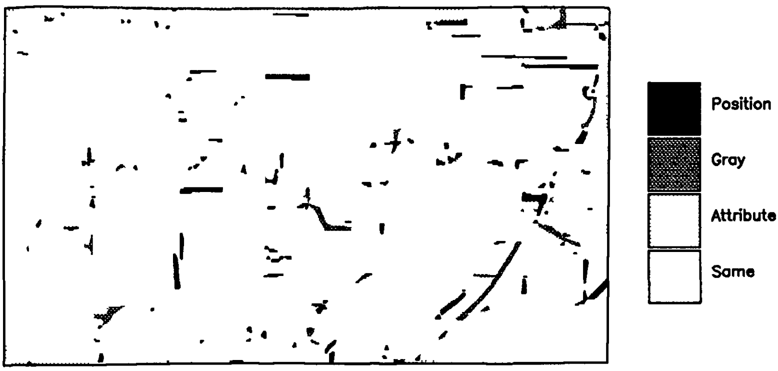
Table 5: Residual Positional Error [Table 2 minus Table 4] (hectares)

Interp. 2:	Row Crops	Meadow	Coop Field	Woods	Other
Interp. 1:					
Row Crops	–	16.9	.5	.3	30.0
Meadow	4.7	–	-.2		11.5
Coop Field	-.4	1.0	–	.03	1.8
Woods	.2	.2	-.03	–	1.6
Other	8.2	1.0	-.6	-.2	–

Filtering

An alternative view of positional error does not seek a uniform translation, but it recognizes that maps are often digitized in greater resolution than is warranted by their accuracy. The overlay test in Map 3 and Table 2 was carried out at a very exacting tolerance (.4 m). This is essentially an exact overlay. Of course, it is critical to apply an exact test to determine the total error between two digital maps. But in addition, it is useful to study how much of the error observed comes from the classical slivers which are entirely unintentional (Goodchild, 1978). Cook (1983) included a distribution showing the areas of the objects, as shown indirectly in the graphs of Figures 2 &3. But the important characteristic of a sliver is narrowness, not area. A distance filter is much more appropriate. The epsilon filter (Dougenik, 1980; Chrisman, 1983; Beard, 1987) was applied during another overlay run. A series of tolerances were tried from 1 meter to 20 meters. The results of the test are reported from the most drastic filter, 20 meters. Map 5 presents the test results. This figure is a better estimate of the accuracy required for a land cover map for a soil erosion plan in this landscape. Any point on either map found within 20 meters of another point causes the cluster analysis of the WHIRLPOOL algorithm to ensure that only one will survive. This process does not average coordinates, it selects points.

Map 5: Test results after a 20 m filter



The first observation about Map 5 and Table 6, the misclassification produced by a 20 meter filtering overlay, is that the area in the diagonal increased from 85.8% to 88.4%. Thus, some area which was found to be in error with an exact overlay, was placed into the correct classifications if the positional tolerance was broadened to 20 meters. Some small error interactions discovered by the exact test disappear. Many categories are reduced substantially, while others are unaffected. This difference in behavior begins to discover the structure of error. The error below the 20 meter threshold may have essentially random distributions, but what survives may point at specific problems to correct.

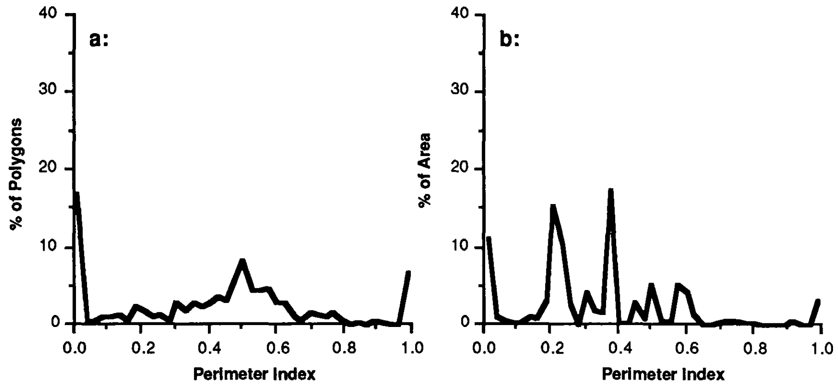
Table 6: Misclassification Matrix after filtering [20 m] (hectares)

Interp. 1: / Interp. 2:	Row Crops	Meadow	Coop Field	Woods	Other
Row Crops	1136.6	76.5		.2	41.0
Meadow	12.1	221.6			32.2
Coop Fld	3.5	2.9	34.1		1.1
Woods	.04	.1		11.0	1.9
Other	24.1	8.9	.1	7.6	222.6

It is interesting to compare the effects of the 20 meter filtering and the .8 meter translation. The two figures seem to be radically different, but the error matrices are surpisingly similar. The difference is that the .8 meter translation occurs everywhere. All lines (at least those going east-west) generate slivers in proportion to their length. The 20 meter filtering actually moves things less. Compared to the exact overlay conducted in Map 3 which used 6048 points, the 20 meter filtering produces a representation in Map 5 with only 3241 points. It is a radical filtering, yet the basic message about the error between Map 1 and Map 2 is still there.

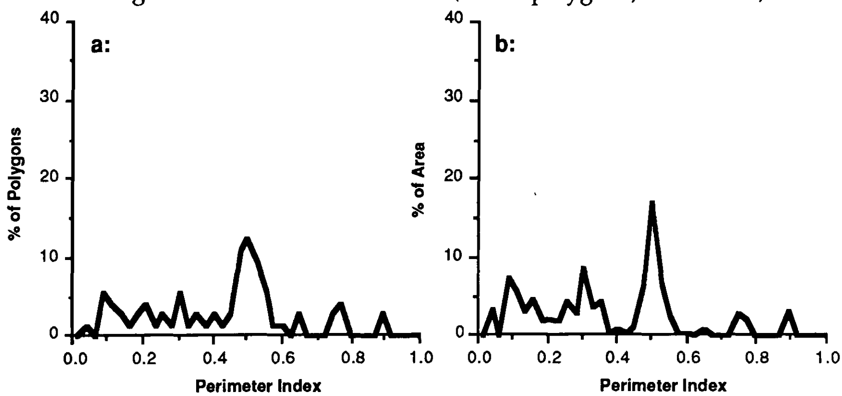
The filtering has an impact on the distribution of the perimeter index. Figure 8 shows the distribution for all polygons, as Figure 4 did for the exact case.

Figure 8: Distribution of perimeter index (all polygons 20 m filter)
a: by percentage of number of polygons; b: by percentage of area



There are many fewer polygons, but the distribution in 8a is well centered around 0.5. The filter removed many small slivers, reducing the numbers, and increased the discovery of coincident lines, making the distribution of indeces more balanced. Figure 9 shows the distribution of the tested polygons only, just as Figure 5 did above. Figure 9 shows a dramatic reduction in the positional (central) spike, though it still remains the mode of the distribution. Figure 9 involves many fewer polygons and less area than Figure 5.

Figure 9: Distribution of index (tested polygons, 20 m filter)



The filter has reduced the positional error from 6% to near 1% of the total area. Thus, we can infer that the positional accuracy of the features on these maps match to within 20 meters, except for 1% of more gross blunders. Such a statement could be refined by an iterative use of the filter.

Table 7 tabulates the errors reported in the positional category by the test at 20 meter tolerance. This is the positional error residual after the filtering. It seems to discover some of the same residual effects found by the translation. A few figures in this table (such as Row Crop/Meadow and Row Crop/Other) are roughly symmetrical around the diagonal, meaning that the one error was about as likely as the other. If the user is interested in overall figures for area, a finding of balanced error is similar to a finding of an unbiased estimator in statistics. However, if the user needs site-specific figures, the errors still are errors.

Table 7: Filtered Positional Error [20 m] (hectares)

Interp. 2:	Row Crops	Meadow	Coop Field	Woods	Other
Interp. 1:					
Row Crops	—	6.3		.2	8.9
Meadow	4.4	—			2.2
Coop Fld	.004		—		.4
Woods	.04	.1		—	.4
Other	5.3	1.0	.1	.04	—

Some of the entries in Table 7 are not symmetric. These point out specific discrimination biases between the interpreters. This information, if produced as a test during a normal GIS production sequence would provide information beyond the typical accuracy assessment that would diagnose the specific pair of categories. Such information should enhance quality control.

The filtering procedure has less effect on attribute-like errors. The matrix of attribute error is presented in Table 8. The large propensity for interpreter 2 to see Meadow when #1 sees Row Crops carries over from the misclassification matrix. The size has risen from 59.8 ha. in Table 3 to 68.1. Most of the figures in this table have increased from Table 3. This matrix is notably less symmetrical compared to Table 7.

Table 8: Filtered Attribute Error [20 m] (hectares)

Interp. 2:	Row Crops	Meadow	Coop Field	Woods	Other
Interp. 1:					
Row Crops	—	68.1			23.9
Meadow	5.6	—			27.9
Coop Fld	3.5	2.9	—		.6
Woods				—	1.5
Other	18.0	7.9		7.6	—

Considering the size of the map sheet, most of these attribute errors may be tolerable. Quality control efforts might apply to correct the positional errors as a higher priority, but some attention might also be given to Interpreter 2's propensity to classify Row Crops as Meadow.

Limitations

The test procedure developed in this paper is provisional. It does not classify all the errors entirely correctly. For example, there are some large errors declared to be position-like because the compactness index is relatively small and the perimeter index is balanced between the two sources, but the nature of the error seems to be much more a disagreement over classification, not position. These cases occur when a relatively compact polygon has an attached "tail" or when a polygon is a rectangle with a relatively narrow width (see the extreme left of Map 4). In these cases, unlike more typical position-like errors, the model will require further refinement.

In a more general sense, this test simply reports on the results for the area studied. It has no mechanism to estimate what would happen in some other, even nearby region. It does not have any particular statistical distribution or measure of goodness-of-fit. However, separation of the distinct forms of error is a first step towards the construction of such models.

Conclusions

This paper has attempted to demonstrate that a polygon overlay test is indeed possible and useful. It is possible to conduct a test using a replication of a map product, not necessarily a source of known higher accuracy. Differences of minimum mapping units and classification schemes are not a hinderance, but they are the very goal of a test. This test offers a chance to diagnose specific forms of mapping error. With some development and fine tuning it may come to replace the more standard point sampling methods used in the remote sensing discipline.

Acknowledgements
US National Science Foundation Grant SES 87-22084 provided partial support for this paper. The data used for the tests was generated by the Dane County Land Records Project during 1984 by Marcel Thoma and Kevin Connors. The University of Washington Statistical Consulting Service has provided a sounding board for some of the concepts, though the authors are responsible for any errors.

References Cited

Chrisman, N.R. 1982: Methods of spatial analysis based on error in categorical maps. PhD thesis, U. of Bristol.

Chrisman, N.R. 1983: The role of quality information in the long-term functioning of a GIS. *Proceedings AUTO-CARTO 6, 2,* 303-321.

Chrisman, N.R. 1987: The accuracy of map overlays: a reassessment. *Landscape and Urban Planning, 14,* 427-439.

Chrisman, N.R. 1989a: Modeling error in overlaid categorical maps. In: Goodchild and Gopal (ed.) *Accuracy of Spatial Databases,* London, Taylor and Francis: 21-34.

Chrisman, N.R. 1989b: Error in categorical maps: testing versus simulation. *Proceedings AUTO-CARTO 9,* 521-529.

Chrisman, N., Mezera, D., Moyer, D., Niemann, B., Vonderohe, A. 1984: Modernization of routine land records in Dane County, Wisconsin: implications to rural landscape assessment and planning, *URISA Professional Paper 84-1.*

Cook, B.G. 1983: Geographic overlay and data reliability, Proceedings IGU US/Australia Workshop: 64-80.

Douglas, D. 1974: It makes me so CROSS, Harvard Laboratory for

Computer Graphics, reprinted p. 303-307 in Peuquet and Marble 1990: *Introductory Readings in GIS*, Taylor & Francis.
Dougenik, J.A. 1980: WHIRLPOOL: A geometric processor for polygon coverage data, *Proceedings AUTO–CARTO IV*, *2:* 304-311.
Fitzpatrick-Lins, K. 1978: Comparison of sampling procedures and data analysis for a land use and land cover map, *Photogrammetric Engineering and Remote Sensing*, *47*, 343-351.
Goodchild, M. 1978: Statistical aspects of the polygon overlay problem, in vol. 6 G. Dutton, ed., *Harvard Papers on Geographic Information Systems*, Addison Wesley.
Mark, D.M. and Csillag, F. 1989: The Nature of Boundaries in 'Area-Class Maps', *Cartographica*, *26*(1), 65-78.
Morrison, J. (ed.) 1988: The Proposed Standard for Digital Cartographic Data, *The American Cartographer*, **15**: 9-140.
Niemann, B. Sullivan, J. Ventura S., Chrisman, N. Vonderohe, A. Mezera, D. and Moyer, D. 1987: Results of the Dane County Land Records Project, *Photogrammetric Engineering and Remote Sensing*, *53*, 1371-1378.
Perkal, J. 1966: An attempt at objective generalization. *Discussion Paper 10*, Michigan InterUniversity Community of Mathematical Geographers.
Rosenfield, G. and Melley, M. 1980: Applications of statistics to thematic mapping. *Photogrammetric Engineering and Remote Sensing*, *46*, 1287–1294.
Shyue, S-W. 1989: High breakdown point robust estimation for outlier detection in photogrammetry, PhD dissertation, University of Washington.
Sinton, D. 1978: The inherent structure of information as a constraint to analysis: mapped thematic data as a case study , in vol. 7 G. Dutton, ed., *Harvard Papers on Geographic Information Systems*, Addison Wesley.
Unwin, D. 1981: *Introductory Spatial Analysis*, Methuen, London.
Ventura, S. 1988: *Dane County Soil Erosion Plan*, Land Conservation Committee, Madison WI.
Ventura, S. Sullivan, J.G. and Chrisman, N. 1986: Vectorization of Landsat TM land cover classification data. *Proceedings URISA*, *1*, 129-140.
Vonderohe, A.P. and Chrisman, N.R. 1985: Tests to establish the quality of digital cartographic data: some examples from the Dane County Land Records Project, *Proceedings AUTO–CARTO 7*, 552-559.