Mapping “soft” data, data that are inaccurate, uncertain, or of dubious lineage, presents several challenges to cartography. A review of existing research shows that authors have focussed on issues of parallel vs. active display for uncertainty, which visual variable are appropriate for uncertainty data and the need to use additional display variables such as animation and focus to display uncertainty data on maps. Nevertheless, most research has shown that existing cartographic representational methods are adequate for the task. This study takes the case of point displacements between line maps, devises a six parameter affine fit between two possible maps, and then experiments with different static displays for the background display of uncertainty information using color and hillshading. Color and animation are thought to hold value for future active portrayal of uncertainty information, but this area of research needs a considerable amount of additional work.

INTRODUCTION

Cartographers have long been concerned with the accuracy and reliability of their source information for mapping. Some notorious historical examples of the consequences of map error and misinformation exist (Monmonier, 1995). Recently, attention has moved from the exclusive use of precise or “hard” data, with known accuracy, precision and lineage, to incorporate imprecise, uncertain or “soft” data. Soft data may contain positional inaccuracy, lack precision due to an inadequate source scale, has uncertainty associated with the attributes or features, or be of dubious lineage. Geographic Information Systems and cartographic display systems treat all data as hard, making no distinction either by attributes or over space. Automated systems, in fact, make it not only possible to mix data of varied lineage and accuracy, but desirable to do so. The power of the automated display, therefore, can be enhanced by applying the same cartographic tools now used for hard data to the cartographic portrayal of uncertainty.

A more realistic examination of the metrics and models used in building an understanding of error has shown that our assumptions of hardness are often misleading. Even a “hard” figure such as the root mean squared positional or elevation error (RMSE) leaves much to be desired, both as a metric and in application. Many studies, for example, assume that horizontal and vertical map error are independent, and can be measured as such using least squares methods. The recent rethinking of the National Map Accuracy Standard, and the pioneering studies in terrain analysis that have investigated the impact of random error are convincing demonstrations of the inadequacy of the single metric approach to hard map data (Fisher, 1993a).

Most of the existing work on error and uncertainty has understandably centered around both building models to describe the uncertainty and contriving metrics to make empirical measures of these values. This is an understandable order to the sequence of research, since without descriptive models and metrics, cartographers have little to map. Nevertheless, some work has focussed on which cartographic representational techniques would be suitable for use as cartographic means for portraying uncertainty. At the outset, this work poses a dichotomy. Many methods work in parallel to the map data, in that they use the method of multiple displays or “small multiples” to communicate the error information as metadata. This has a parallel in standard topographic mapping, such as the lineage source date maps on the collars of USGS quadrangles, and the air photo source coverage diagrams on the Canadian 1:50,000 series. Methods that work in parallel assume that the uncertainty portrayal is peripheral information that should be subjected to interpretation after the primary “hard” data have been read.

This approach does not match some of the more recent research findings. Automated mapping methods, and some of the more recent error metrics, make the level of interpretation of the data at the feature level. This means that for virtually every map object in the data base, we can compute and therefore display an error value. The symbolization decision then becomes one of deciding which level of uncertainty is acceptable as a broad generalization for
inclusion within the viewing geometry. This leads to the observation that a map is “a set of errors that have been agreed upon” (Clarke, 1995).

A superior approach would be one which seeks to recognize the degree of uncertainty associated with features and integrates them into the cartographic symbolization. We refer to such a map as one with “active” uncertainty. Should the user choose to ignore it, and there are many cases where this is essential, then traditional mapping methods can be used. Should the uncertainty be important for the interpretation task, then metric data should be viewable simultaneously with the hard data in the same map display geometry. The research challenge, then, is to do this in a way that is meaningful, does not create additional “use error” (Beard, 1987) and enhances the map’s utility for decision making. Deliberate or unintentional error in depiction of uncertainty information may be the ultimate way to “lie with maps.”

EXISTING RESEARCH AND METHODS

Beard, Buttenfield and Clapham (1991) provided the first framework for the consideration of the cartographic portrayal of uncertainty in the published outcomes of NCGIA initiative 7. In their report, cartographic representational methods for uncertainty were mapped into a matrix of data types, one axis being the data quality categories of the Spatial Data Transfer Standard (positional accuracy, attribute accuracy, logical consistency, completeness and lineage) and the other axis being the feature type (discrete, categorical, or continuous). For point and line data, the Bertin variables of size, shape, and color (value and saturation) were suggested. For categorical data, texture, size, color value and mixing were outlined. For continuous data, color value and saturation, and possible size and shape were suggested. Lineage data seemed the hardest to represent, and no new methods were suggested other than showing source map bounding rectangles. The authors concluded that hypermedia, animation and virtual reality were worthy of further research for depicting uncertainty, the uncertainty data should be displayed in an interactive and proactive manner, and that user training may be necessary to gain acceptance of cartographic portrayal of uncertainty information.

The use of the Bertin visual variables was furthered in research by Alan MacEachren (MacEachren, 1992; 1994). His work concluded that of the visual variables outlined by Bertin, size and value were most suitable for numerical data, and that hue (color), shape, and orientation could be used for nominal information. While favoring color saturation for uncertainty portrayal (not a Bertin variable) MacEachren added a new variable of focus. Focus is changed for a feature by manipulating edge crispness, fill clarity, by adding fog, or by changing resolution. Adding image blur and obscuring detail may seem anathema to a trained cartographer, and further testing of this concept is clearly required to assess its utility. MacEachren tested three methods, map pairs, map sequences and bivariate maps for uncertainty presentation. McGranaghan (1993) instead favored the approach taken here, that data quality can be explicitly encoded and attached to point, line and area entities. McGranaghan favored creating visual and cognitive ambiguity in a display to depict uncertainty, and did so using even more Bertin variables, focus, realism, time and interaction. The latter two imply animation and interaction, and the work suggested that existing cartographic methods offered promise for the display of cartographic uncertainty. In further work, David Unwin and colleagues (Unwin et al., 1994) added to the Bertin variable projection, animation, time and sound for uncertainty work.

Peter Fisher, using soil maps as a case study, further refined the concept of the use of animation (Fisher, 1993b). His idea of animating the set of possible maps created by multiple simultaneous equiprobable outcome maps or realisations has a direct parallel in Journel’s concept of stochastic imaging (Journel, 1996). His methods combine qualitative and quantitative approaches to point, line and area data, and seem independent of attribute. Fisher used soil maps, imagery, dot maps and digital elevation models (DEMs) as test cases (Fisher, 1994;1996). The same approach was taken by Ehlschlaeger, Goodchild and Shortridge in three papers (Ehlschlaeger et al, 1994; 1996; 1997). Single frame sequential animation of equiprobable stochastic images were used for line (shortest paths) and DEM data. Davis and Keller considered animation as a separate suite of methods for uncertainty, arguing that animation pace or duration represented yet another visual variable for calibration against uncertainty (Davis and Keller, 1997).

In parallel work in computer science and graphics, several researchers at the Santa Cruz Laboratory for Visualization and Graphics have worked on the use of three dimensional display of uncertainty using vectors, and advocate the use of glyphs, multivariate point symbols that appear over objects in a 3D viewing geometry and change shape, size, color and texture according to the levels of uncertainty (Wittenbrink et al, 1995; 1996). While the goal of the group’s work is to integrate active uncertainty into their visualizations. With the exception of their glyphs, in many cases, uncertainty views are multiple displays separate from the hard data.
Finally, in the area of map projection distortion display, several recent papers have suggested the use of checkerboard grids (Steinwand et al., 1995), color variation (Clarke and Mulcahy, 1995), and other methods (Mulcahy and Clarke, 1995) on maps to show the geometric pattern of distortion. Unlike the classic Tissot method, few of these are suitable for direct display behind a hard map. The color method (Clarke and Mulcahy, 1995) shows most promise for this purpose, and was the basis of the further experiments in this work.

Our review allowed a diagramatic listing of the major Bertin variables and their utility for mapping uncertainty (Figure 1). Several themes are recurrent in this figure, and will be pursued in further research. Clearly animation is a very powerful suite of methods for uncertainty visualization. Of the static cartographic methods, the use of color seems to be equally versatile for uncertainty display. Color was chosen in the current work for an experiment in uncertainty visualization.

<table>
<thead>
<tr>
<th>DATA TYPE</th>
<th>POINT</th>
<th>LINE</th>
<th>AREA</th>
<th>VOLUME</th>
<th>TEMPORAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISCRETE</td>
<td>COLOR</td>
<td>ANIMATION</td>
<td>SOUND</td>
<td>VIRTUAL</td>
<td>POSITION</td>
</tr>
<tr>
<td>SIZE</td>
<td>SHAPE</td>
<td>TOUCH</td>
<td>GLYPHS</td>
<td>TRANSPARENCY</td>
<td></td>
</tr>
<tr>
<td>CATEGORICAL</td>
<td>COLOR</td>
<td>ANIMATION</td>
<td>SOUND</td>
<td>GLYPHS</td>
<td>SHAPE</td>
</tr>
<tr>
<td>SIZE</td>
<td>TOUCH</td>
<td>GLYPHS</td>
<td>TRANSPARENCY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONTINUOUS</td>
<td>COLOR</td>
<td>ANIMATION</td>
<td>SOUND</td>
<td>SIZE</td>
<td>TRANSPARENCY</td>
</tr>
</tbody>
</table>

Figure 1: Cartographic visual variables (Bertin variables) suitable for use as the basis of new representational techniques for the visualization of error and uncertainty on maps.

PURPOSE

The purpose of the current work was as follows. A suitable Bertin variable (color) was isolated for the purpose of exploring methods of static uncertainty display behind a hard image. We limited the consideration to positional error only, and made the assumption that point, line and area positional error can always be evaluated with a floating point metric at a set of points dispersed in space. Primarily, any cartographic representational method of uncertainty mapping should be suitable for printing or display “behind” the hard data without distracting the map reader in the primary data interpretation tasks. Secondly, any cartographic depiction of uncertainty should make use of existing rather than totally new methods, so that less training is involved and to minimize map interpretation and use error. Without this goal, it was felt that map users would simply ignore the uncertainty information because it was too difficult to interpret. Thirdly, cartographically it is desirable to depict not only the magnitude of expected error overall, as in the RMSE case, but the spatial distribution of both the magnitude and the direction of positional error. There are parallels in the decomposition of topographic information into a slope magnitude (slope) and direction of slope (aspect) for analysis. Finally, the use of color should be as close as possible both numerically sound and perceptually correct. Cartographers have long understood the relationship between the actual and perceived impact of colors. The Munsell color scheme takes the perceptual impact of colors into consideration. RGB and HSI color schemes can be mapped
ASSUMPTIONS

The following assumptions were made about the nature of the uncertainty and error involved. These assumptions are highly dependent on the type and nature of the error measurement in use. No claim is made that attribute error is covered by these assumptions, so the method is limited to positional accuracy in three dimensions. First, it is assumed that point, line and area measurements and therefore accuracies are spatially autocorrelated. This means that point error is impacted by adjacent points, parts of a line that are adjacent show similar errors, and that on a surface, errors are related to zones of similarity. This is most likely true for “random” spatial error. If the data were aspatial, it would be largely moot what the spatial error was because there would be no predictability in space that depiction could reveal and the most suitable cartographic method would be to estimate a suitable error level, select all data over this level as hard, and to report an aggregate error measure such as RMSE. Systematic error, cartographically the result of mismeasurement, blunder, incorrect rubber sheeting, incorrect edge matching or the like would be fairly spatially uniform throughout the area impacted. Other errors could be due to tiling, scan lines errors etc. In the majority of these cases, there is indeed significant spatial autocorrelation of the error over the map space. Autocorrelated error is most suitable for visualization.

Secondly, the assumption is made that whether the feature data are point, line or area, it is possible to interpolate the local error measurements over continuous space. Problems arise at the edges of the map extents, where clearly there are discontinuities. This assumption makes the parallel assumption of spatial continuity. That is, no set of errors can “tear” the space by say switching relative but not absolute locations.

Thirdly, to compare two maps even on the same coordinate system, one has to be assumed the control and one the measurement. In the example given here, both maps were made from digital coordinates on the same coordinate system, at the same scale and with the same extent. Neither can then be regarded as an “independent source of higher authority.” Two approaches can therefore be used. First, one geometry can be assumed correct and the other in error. In this case, the errors can be rather simply measured and depicted, for example as vectors. A second case evolves when neither map is assumed as control. In this case, both can be seen as a sample of a set of all possible maps, the case of stochastic imaging (Journel, 1996).
Many methods for the cartographic portrayal of error and uncertainty are possible, and some from prior studies meet the criteria set in the section on assumptions. An extensive review of the methods conducted as part of this research resulted in the summary table shown as figure 1. Of these, the Bertin variable of color seemed to be the most uniformly applicable to static maps. It was therefore selected for further work. A test data set was selected from the VITAL case study data for Goleta, California streets. The two maps were street centerline data from two commercial vendors of digital map data for vehicle guidance systems. If the data are simply overdrawn, then the resultant map appears as in figure 2 If significant street intersections are taken as point data, then a set of comparison points can be
Two problems with the vector line depiction in figure 3 are (1) that it assumes one correct and one incorrect map in the direction of the vector (although the same line fits both cases) and (2) systematic, uncorrelated error is overrepresented because the error variance has not been normalized. To circumvent the latter problem, a least squares six parameter affine fit between the two datasets was conducted. Results are given in the following section. The data were entered into a spreadsheet, with map A set as the source (u,v) and map B (x,y) as the erroneous data. A computer program for the six parameter affine was used as given in Clarke (1995). A total of 304 points in the UTM projection were used. Each point’s coordinates were entered to two decimal places, that is to the nearest centimeter. This level was used to facilitate centimeter level GPS measurement for accuracy testing at a later time.

RESULTS

Of the displacement vectors for the two data sets, the magnitude of error varies over a range from 0 to 366.75 meters, with a mean of 41.56 meters and a high standard deviation of 55.33 meters. The distribution of the errors as shown in figure 4 is clearly not random but shows a tendency for systematic error. This is echoed by the computed values for the six affine constants. Units were UTM zone 11 meters. Computed values for the coefficients were:

![Figure 3. Section of the Goleta map databases with highlighted corresponding nodes shown in both their locations, connected by a vector. Map courtesy of Kevin Curtin, NCGIA.](image3)

![Figure 4: Distribution of distance errors for the 304 point matches in figure 3. Figure courtesy of Jeannette Candau.](image4)
\[
\begin{align*}
  u &= 1.000000 \times + 0.000000 \times + 0.000000 \\
  v &= 0.004842 \times + 0.986290 \times + 51123.742188
\end{align*}
\]

This indicates a perfect on average match between the maps in the \( x \) direction, but a significant mismatch in \( y \), involving scaling, rotation, and translation. This could be due to the use of a different datum, projection mismatch, a digitizing error, and any one of many other causes. For every point, then, a residual aggregate error vector was computed. This error vector represents the distance and direction each node would be moved during an affine fit between the two data sets.

An experiment was then conducted. Using the computer program referred to above, at each point in the map the distance vector was calculated from the inverse of the six parameter affine transformation. These values were interpolated into a grid, and decomposed into (1) east-west distance (2) a north-south distance and (3) a direction as the aspect of the vector, from zero to 360 degrees. The distributions are shown in figure 5.

The radically different random error in \( x \) and systematic error in \( y \) are clearly shown. Obviously, the “\( y \)” match has one very incorrect point in the upper left of the display area. The aspect image shows the impact of outliers and the map edges on the interpolation, but generally shows the patchiness of the distortion over the map space. The first three of these images are hill shaded in figure 6.

Clearly, the most information-rich image backdrop for the data would combine the two different (\( x \) and \( y \)) magnitude components of the error and the information about vector direction. To accomplish this, the east-west component was scaled and mapped onto the red image component, the north-south onto the green, and the direction onto blue, in the same manner as in the map projection distortion method of Clarke and Mulcahy (1995) discussed above (Figure 7). Note that the scaling is not by HSI, as is required in the purpose statement. Conversion to HSI and absolute color scaling would be necessary to balance the colors perceptually. For example, a zero value on all should result in a neutral grey rather than a black, as it does in the current scaling. Note that the extremes are identifiable, and are color sepa-
CONCLUSIONS

Clearly much work on the cartography of uncertainty remains to be done. This work has considered the range of existing research on the cartographic portrayal of uncertainty. The modified Bertin variables of color and animation (time) seem to hold the greatest potential for the cartography of uncertainty display. A single example was pursued, using line data abstracted as points and interpolated assuming autocorrelated error. As such, an error “surface” can be built that yields to some forms of standard cartographic display such as hill shading. A color combination method, devised as an experiment misrepresents the variable of hue by failing to use the Munsell color scheme. Nevertheless, the methods used as experiments all reveal the segregated random and systematic directional nature of the error in the case study. Further work will continue this static color method, and will further examine the variables of animation and depth of field as ways of communicating map and attribute uncertainty together as active map layers.

ACKNOWLEDGEMENTS

This work was funded by the National Imagery and Mapping Agency under the NURI University Research Initiative. Support is also acknowledged from the National Science Foundation as part of the “Advancing Geographic Information Science” award to NCGIA.

REFERENCES


Mapping “soft” data, data that are inaccurate, uncertain, or of dubious lineage, presents several challenges to cartography. A review of existing research shows that authors have focused on issues of parallel vs. active display for uncertainty, which visual variable are appropriate for uncertainty data and the need to use additional display variables such as animation and focus to display uncertainty data on maps. Nevertheless, most research has shown that existing cartographic representational methods are adequate for the task. This study takes the case of point displacements between line maps, devises a six parameter affine fit between two possible maps, and then experiments with different static displays for the background display of uncertainty information using color and hillshading. Color and animation are thought to hold value for future active portrayal of uncertainty information, but this area of research needs a considerable amount of additional work.

INTRODUCTION

Cartographers have long been concerned with the accuracy and reliability of their source information for mapping. Some notorious historical examples of the consequences of map error and misinformation exist (Monmonier, 1995). Recently, attention has moved from the exclusive use of precise or “hard” data, with known accuracy, precision and lineage, to incorporate imprecise, uncertain or “soft” data. Soft data may contain positional inaccuracy, lack precision due to an inadequate source scale, has uncertainty associated with the attributes or features, or be of dubious lineage. Geographic Information Systems and cartographic display systems treat all data as hard, making no distinction either by attributes or over space. Automated systems, in fact, make it not only possible to mix data of varied lineage and accuracy, but desirable to do so. The power of the automated display, therefore, can be enhanced by applying the same cartographic tools now used for hard data to the cartographic portrayal of uncertainty.

A more realistic examination of the metrics and models used in building an understanding of error has shown that our assumptions of hardness are often misleading. Even a “hard” figure such as the root mean squared positional or elevation error (RMSE) leaves much to be desired, both as a metric and in application. Many studies, for example, assume that horizontal and vertical map error are independent, and can be measured as such using least squares methods. The recent rethinking of the National Map Accuracy Standard, and the pioneering studies in terrain analysis that have investigated the impact of random error are convincing demonstrations of the inadequacy of the single metric approach to hard map data (Fisher, 1993a).

Most of the existing work on error and uncertainty has understandably centered around both building models to describe the uncertainty and contriving metrics to make empirical measures of these values. This is an understandable order to the sequence of research, since without descriptive models and metrics, cartographers have little to map. Nevertheless, some work has focussed on which cartographic representational techniques would be suitable for use as cartographic means for portaying uncertainty. At the outset, this work poses a dichotomy. Many methods work in parallel to the map data, in that they use the method of multiple displays or “small multiples” to communicate the error information as metadata. This has a parallel in standard topographic mapping, such as the lineage source date maps on the collars of USGS quadrangles, and the air photo source coverage diagrams on the Canadian 1:50,000 series. Methods that work in parallel assume that the uncertainty portrayal is peripheral information that should be subjected to interpretation after the primary “hard” data have been read.

This approach does not match some of the more recent research findings. Automated mapping methods, and some of the more recent error metrics, make the level of interpretation of the data at the feature level. This means that for virtually every map object in the data base, we can compute and therefore display an error value. The symbolization decision then becomes one of deciding which level of uncertainty is acceptable as a broad generalization for
inclusion within the viewing geometry. This leads to the observation that a map is “a set of errors that have been agreed upon” (Clarke, 1995).

A superior approach would be one which seeks to recognize the degree of uncertainty associated with features and integrates them into the cartographic symbolization. We refer to such a map as one with “active” uncertainty. Should the user choose to ignore it, and there are many cases where this is essential, then traditional mapping methods can be used. Should the uncertainty be important for the interpretation task, then metric data should be viewable simultaneously with the hard data in the same map display geometry. The research challenge, then, is to do this in a way that is meaningful, does not create additional “use error” (Beard, 1987) and enhances the map’s utility for decision making. Deliberate or unintentional error in depiction of uncertainty information may be the ultimate way to “lie with maps.”

EXISTING RESEARCH AND METHODS

Beard, Buttenfield and Clapham (1991) provided the first framework for the consideration of the cartographic portrayal of uncertainty in the published outcomes of NCGIA initiative 7. In their report, cartographic representational methods for uncertainty were mapped into a matrix of data types, one axis being the data quality categories of the Spatial Data Transfer Standard (positional accuracy, attribute accuracy, logical consistency, completeness and lineage) and the other axis being the feature type (discrete, categorical, or continuous). For point and line data, the Bertin variables of size, shape, and color (value and saturation) were suggested. For categorical data, texture, size, color value and mixing were outlined. For continuous data, color value and saturation, and possible size and shape were suggested. Lineage data seemed the hardest to represent, and no new methods were suggested other than showing source map bounding rectangles. The authors concluded that hypermedia, animation and virtual reality were worthy of further research for depicting uncertainty, the uncertainty data should be displayed in an interactive and proactive manner, and that user training may be necessary to gain acceptance of cartographic portrayal of uncertainty information.

The use of the Bertin visual variables was furthered in research by Alan MacEachren (MacEachren, 1992; 1994). His work concluded that of the visual variables outlined by Bertin, size and value were most suitable for numerical data, and that hue (color), shape, and orientation could be used for nominal information. While favoring color saturation for uncertainty portrayal (not a Bertin variable) MacEachren added a new variable of focus. Focus is changed for a feature by manipulating edge crispness, fill clarity, by adding fog, or by changing resolution. Adding image blur and obscuring detail may seem anathema to a trained cartographer, and further testing of this concept is clearly required to assess its utility. MacEachren tested three methods, map pairs, map sequences and bivariate maps for uncertainty presentation. McGranaghan (1993) instead favored the approach taken here, that data quality can be explicitly encoded and attached to point, line and area entities. McGranaghan favored creating visual and cognitive ambiguity in a display to depict uncertainty, and did so using even more Bertin variables, focus, realism, time and interaction. The latter two imply animation and interaction, and the work suggested that existing cartographic methods offered promise for the display of cartographic uncertainty. In further work, David Unwin and colleagues (Unwin et al., 1994) added to the Bertin variable projection, animation, time and sound for uncertainty work.

Peter Fisher, using soil maps as a case study, further refined the concept of the use of animation (Fisher, 1993b). His idea of animating the set of possible maps created by multiple simultaneous equiprobable outcome maps or realisations has a direct parallel in Journel’s concept of stochastic imaging (Journel, 1996). His methods combine qualitative and quantitative approaches to point, line and area data, and seem independent of attribute. Fisher used soil maps, imagery, dot maps and digital elevation models (DEMs) as test cases (Fisher, 1994;1996). The same approach was taken by Ehlschlaeger, Goodchild and Shortridge in three papers (Ehlschlaeger et al, 1994; 1996; 1997). Single frame sequential animation of equiprobable stochastic images were used for line (shortest paths) and DEM data. Davis and Keller considered animation as a separate suite of methods for uncertainty, arguing that animation pace or duration represented yet another visual variable for calibration against uncertainty (Davis and Keller, 1997).

In parallel work in computer science and graphics, several researchers at the Santa Cruz Laboratory for Visualization and Graphics have worked on the use of three dimensional display of uncertainty using vectors, and advocate the use of glyphs, multivariate point symbols that appear over objects in a 3D viewing geometry and change shape, size, color and texture according to the levels of uncertainty (Wittenbrink et al, 1995; 1996). While the goal of the group’s work is to integrate active uncertainty into their visualizations. With the exception of their glyphs, in many cases, uncertainty views are multiple displays separate from the hard data.
Finally, in the area of map projection distortion display, several recent papers have suggested the use of checkerboard grids (Steinwand et al., 1995), color variation (Clarke and Mulcahy, 1995), and other methods (Mulcahy and Clarke, 1995) on maps to show the geometric pattern of distortion. Unlike the classic Tissot method, few of these are suitable for direct display behind a hard map. The color method (Clarke and Mulcahy, 1995) shows most promise for this purpose, and was the basis of the further experiments in this work.

Our review allowed a diagramatic listing of the major Bertin variables and their utility for mapping uncertainty (Figure 1). Several themes are recurrent in this figure, and will be pursued in further research. Clearly animation is a very powerful suite of methods for uncertainty visualization. Of the static cartographic methods, the use of color seems to be equally versatile for uncertainty display. Color was chosen in the current work for an experiment in uncertainty visualization.

![Diagram of Bertin variables suitable for use as the basis of new representational techniques for the visualization of error and uncertainty on maps.](image)

Figure 1: Cartographic visual variables (Bertin variables) suitable for use as the basis of new representational techniques for the visualization of error and uncertainty on maps.

**PURPOSE**

The purpose of the current work was as follows. A suitable Bertin variable (color) was isolated for the purpose of exploring methods of static uncertainty display behind a hard image. We limited the consideration to positional error only, and made the assumption that point, line and area positional error can always be evaluated with a floating point metric at a set of points dispersed in space. Primarily, any cartographic representational method of uncertainty mapping should be suitable for printing or display “behind” the hard data without distracting the map reader in the primary data interpretation tasks. Secondly, any cartographic depiction of uncertainty should make use of existing rather than totally new methods, so that less training is involved and to minimize map interpretation and use error. Without this goal, it was felt that map users would simply ignore the uncertainty information because it was too difficult to interpret. Thirdly, cartographically it is desirable to depict not only the magnitude of expected error overall, as in the RMSE case, but the spatial distribution of both the magnitude and the direction of positional error. There are parallels in the decomposition of topographic information into a slope magnitude (slope) and direction of slope (aspect) for analysis. Finally, the use of color should be as close as possible both numerically sound and perceptually correct. Cartographers have long understood the relationship between the actual and perceived impact of colors. The Munsell color scheme takes the perceptual impact of colors into consideration. RGB and HSI color schemes can be mapped...
ASSUMPTIONS

The following assumptions were made about the nature of the uncertainty and error involved. These assumptions are highly dependent on the type and nature of the error measurement in use. No claim is made that attribute error is covered by these assumptions, so the method is limited to positional accuracy in three dimensions. First, it is assumed that point, line and area measurements and therefore accuracies are spatially autocorrelated. This means that point error is impacted by adjacent points, parts of a line that are adjacent show similar errors, and that on a surface, errors are related to zones of similarity. This is most likely true for “random” spatial error. If the data were aspatial, it would be largely moot what the spatial error was because there would be no predictability in space that depiction could reveal and the most suitable cartographic method would be to estimate a suitable error level, select all data over this level as hard, and to report an aggregate error measure such as RMSE. Systematic error, cartographically the result of mismeasurement, blunder, incorrect rubber sheeting, incorrect edge matching or the like would be fairly spatially uniform throughout the area impacted. Other errors could be due to tiling, scan lines errors etc. In the majority of these cases, there is indeed significant spatial autocorrelation of the error over the map space. Autocorrelated error is most suitable for visualization.

Secondly, the assumption is made that whether the feature data are point, line or area, it is possible to interpolate the local error measurements over continuous space. Problems arise at the edges of the map extents, where clearly there are discontinuities. This assumption makes the parallel assumption of spatial continuity. That is, no set of errors can “tear” the space by say switching relative but not absolute locations.

Thirdly, to compare two maps even on the same coordinate system, one has to be assumed the control and one the measurement. In the example given here, both maps were made from digital coordinates on the same coordinate system, at the same scale and with the same extent. Neither can then be regarded as an “independent source of higher authority.” Two approaches can therefore be used. First, one geometry can be assumed correct and the other in error. In this case, the errors can be rather simply measured and depicted, for example as vectors. A second case evolves when neither map is assumed as control. In this case, both can be seen as a sample of a set of all possible maps, the case of stochastic imaging (Journel, 1996).
Many methods for the cartographic portrayal of error and uncertainty are possible, and some from prior studies meet the criteria set in the section on assumptions. An extensive review of the methods conducted as part of this research resulted in the summary table shown as figure 1. Of these, the Bertin variable of color seemed to be the most uniformly applicable to static maps. It was therefore selected for further work. A test data set was selected from the VITAL case study data for Goleta, California streets. The two maps were street centerline data from two commercial vendors of digital map data for vehicle guidance systems. If the data are simply overdrawn, then the resultant map appears as in figure 2. If significant street intersections are taken as point data, then a set of comparison points can be
determined, as shown in figure 3.

Two problems with the vector line depiction in figure 3 are (1) that it assumes one correct and one incorrect map in the direction of the vector (although the same line fits both cases) and (2) systematic, uncorrelated error is overrepresented because the error variance has not been normalized. To circumvent the latter problem, a least squares six parameter affine fit between the two datasets was conducted. Results are given in the following section. The data were entered into a spreadsheet, with map A set as the source \((u,v)\) and map B \((x,y)\) as the erroneous data. A computer program for the six parameter affine was used as given in Clarke (1995). A total of 304 points in the UTM projection were used. Each point’s coordinates were entered to two decimal places, that is to the nearest centimeter. This level was used to facilitate centimeter level GPS measurement for accuracy testing at a later time.

RESULTS

Of the displacement vectors for the two data sets, the magnitude of error varies over a range from 0 to 366.75 meters, with a mean of 41.56 meters and a high standard deviation of 55.33 meters. The distribution of the errors as shown in figure 4 is clearly not random but shows a tendency for systematic error. This is echoed by the computed values for the six affine constants. Units were UTM zone 11 meters. Computed values for the coefficients were:
\[ u = 1.000000 \times + 0.000000 \times + 0.000000 \]
\[ v = 0.004842 \times + 0.986290 \times + 51123.742188 \]

This indicates a perfect on average match between the maps in the \( x \) direction, but a significant mismatch in \( y \), involving scaling, rotation, and translation. This could be due to the use of a different datum, projection mismatch, a digitizing error, and any one of many other causes. For every point, then, a residual aggregate error vector was computed. This error vector represents the distance and direction each node would be moved during an affine fit between the two data sets.

An experiment was then conducted. Using the computer program referred to above, at each point in the map the distance vector was calculated from the inverse of the six parameter affine transformation. These values were interpolated into a grid, and decomposed into (1) and east-west distance (2) a north-south distance and (3) a direction as the aspect of the vector, from zero to 360 degrees. The distributions are shown in figure 5.

![Figure 5: The Goleta, CA test map transformed into “error” space using the six parameter affine. From the left: a. Raw distance of the rubber-sheeting transformation interpolated from 304 match points. b. Scaled north-south distance. c. Scaled east-west distance. d. Scaled aspect (direction) of the error vector.](image)

The radically different random error in \( x \) and systematic error in \( y \) are clearly shown. Obviously, the “\( y \)” match has one very incorrect point in the upper left of the display area. The aspect image shows the impact of outliers and the map edges on the interpolation, but generally shows the patchiness of the distortion over the map space. The first three of these images are hill shaded in figure 6.

![Figure 6: a. Magnitude and b. north-south and c. east-west components of the error, depicted as hill shading.](image)

Clearly, the most information-rich image backdrop for the data would combine the two different (\( x \) and \( y \)) magnitude components of the error and the information about vector direction. To accomplish this, the east-west component was scaled and mapped onto the red image component, the north-south onto the green, and the direction onto blue, in the same manner as in the map projection distortion method of Clarke and Mulcahy (1995) discussed above (Figure 7). Note that the scaling is not by HSI, as is required in the purpose statement. Conversion to HSI and absolute color scaling would be necessary to balance the colors perceptually. For example, a zero value on all should result in a neutral grey rather than a black, as it does in the current scaling. Note that the extremes are identifieable, and are color separ-
rated by error type.

CONCLUSIONS

Clearly much work on the cartography of uncertainty remains to be done. This work has considered the range of existing research on the cartographic portrayal of uncertainty. The modified Bertin variables of color and animation (time) seem to hold the greatest potential for the cartography of uncertainty display. A single example was pursued, using line data abstracted as points and interpolated assuming autocorrelated error. As such, an error “surface” can be built that yields to some forms of standard cartographic display such as hill shading. A color combination method, devised as an experiment misrepresents the variable of hue by failing to use the Munsell color scheme. Nevertheless, the methods used as experiments all reveal the segregated random and systematic directional nature of the error in the case study. Further work will continue this static color method, and will further examine the variables of animation and depth of field as ways of communicating map and attribute uncertainty together as active map layers.

ACKNOWLEDGEMENTS

This work was funded by the National Imagery and Mapping Agency under the NURI University Research Initiative. Support is also acknowledged from the National Science Foundation as part of the “Advancing Geographic Information Science” award to NCGIA.

REFERENCES


