Title
Visualization of the Quality of Spatial Information—NCGIA Research Initiative 7, Closing Report

Authors
Beard, M. Kate
Buttenfield, Barbara P.
Mackaness, William A.

Publication Date
1994-05-01

NCGIA

National Center for Geographic
Information and Analysis

Research Initiative 7

Visualization of the Quality of Spatial Information

Closing Report

By

M. Kate Beard
University of Maine - Orono

Barbara P. Buttenfield
State University of New York - Buffalo

William A. Mackaness
University of Maine - Orono

May 1994

Table Of Contents

CLOSING REPORT

VISUALIZATION OF THE QUALITY OF SPATIAL INFORMATION

NCGIA RESEARCH INITIATIVE 7

M. Kate Beard and Barbara P. Buttenfield, Initiative Leaders
William A. Mackaness, Research Assistant Professor

National Center for Geographic Information and Analysis

ABSTRACT

This report describes the results of NCGIA Initiative 7 on Visualization of the Quality of Spatial Information. It begins with a discussion of objectives and the process of developing a research agenda. The initiative research agenda identified eight topic areas which are discussed in detail. The initiative was active during the period 1990-1993. During that time, several research activities were completed. Major scientific outcomes of the initiative include linkage of error models and error propagation with visualization techniques, integration of communication theories into data quality acquisition and presentation, development of object-oriented approaches to data quality organization, and several novel graphical schemes for data quality display. Major deliverables include a refereed journal special issue, results of the data quality visualization challenge, and a bibliography.

OVERVIEW OF THE INITIATIVE

Scope of the Initiative

Technology currently allows us to process and display large volumes of information quickly. Large national spatial databases are becoming available, and regional and local databases continue to be developed. These databases are becoming accessible to large numbers of people and being called on to support a wide range of applications. In the course of these developments, users of spatial data have become removed from the details of data collection and processing which have historically been the basis for understanding data quality. As databases become distributed and shared by multiple users, the need to accumulate, store, and communicate data quality information to a large and growing pool of users becomes critical. In particular the growing interest in digital libraries and the National Spatial Data Infrastructure (NSDI) provides new impetus for both documenting spatial information quality and improving methods for its communication.

If we assume that data have been processed and checked sufficiently that gross errors have been removed, we still face the problem of presenting to GIS users the appropriateness of data with respect to their needs. The quality of spatial information and spatial information products is multidimensional and complex, being a function of both time and space. Communicating this potentially large and complex pool of information to users is a challenge. The primary medium of communication within GIS has been graphic displays, and therefore visualization seems an appropriate and efficient technique for making complex information more comprehensible. The focus of this initiative has been to investigate both impediments and solutions for making information on the quality of spatial information more robust, more pertinent, and more accessible to GIS users.

Objectives for the Initiative

A primary research objective of this initiative is to explore the tools of visualization for communicating the many dimensions of geographic information quality to users in meaningful ways. At present, access to data quality through visualization or other means is either unavailable in existing GIS packages or not well developed. The visualization of data quality implies that we can both reasonably identify and measure data quality. The NCGIA has made substantial progress in this area as an outcome of Initiative 1: Accuracy of Spatial Databases. This initiative extends the results of the first initiative by linking work on error models and error propagation with tools for making these more understandable, visible, and accessible to GIS users. As spatial data collection procedures, GIS operations, spatial analysis, and data displays themselves all impact the quality of GIS products, visualization in an exploratory or confirmatory context can help to inform users about quality throughout the production process or lifespan of the data.

A second objective of this initiative is to provide continuity with other NCGIA initiatives. As stated above the initiative builds most directly from the work of Initiative 1 on the Accuracy of Spatial Databases. In addition, the initiative has connections to Initiative 3: Multiple Representations, since the representation of data quality can be seen as another view of the data. Access to data quality information through visual tools influences the Use and Value of Spatial Information - Initiative 4. It supports work on Initiative 6: Spatial Decision Support Systems by considering the impact of data quality displays on spatial decision-making. It plays a role in Initiative 12: The Integration of GIS and Remote Sensing with respect to enhancing the visualization of data quality through images, and Initiative 13: User Interfaces for GIS, has direct implications for developing effective tools for user access to data quality information.

The lack of tools to visualize data quality in most GISs is partly due to the indifference of users towards quality issues. Therefore a third objective is to sensitize the GIS community to the need for collecting and maintaining data quality information, and to facilitate use and exchange of spatial data. The federal government through the Federal Geographic Data Committee (FGDC) has taken a lead role in promoting long term data quality management. Adoption of Spatial Data Transfer Standards (SDTS) as a Federal Information Processing Standard (FIPS-173) in 1992 set forth a mechanism for federal agencies to test various categories of spatial data quality. Current efforts by the Federal Geographic Data Committee (FGDC) to develop a Metadata Content Standard expands upon FIPS-173, establishing data definitions and formats for inclusion of data quality information with federal-produced data sets. Several state level agencies have shown corresponding initiative in adopting data quality documentation procedures and protocols. This NCGIA Initiative and its offshoot research are intended to serve as catalyst, to assist the GIS community in their research needs and to alert them to the long-term consequences of ignoring data quality representation and visualization.

Organization and Preparation of the Initiative

A research panel was organized at the Baltimore AUTO-CARTO 10 conference in March 1991. The panel was chaired by Barbara Buttenfield, and panelists included: Alan Saalfeld (US Census); Robin Fegeas (US Geological Survey); Ferko Csillag (Syracuse University); Alan MacEachren (Penn State University); and Nick Chrisman (University of Washington). Panelists presented position statements on themes including Components of Data Quality, Database and Data Modeling, Graphical Representation, and Assessment of Data Quality. The panel was well-attended, and audience participation was very good. Ideas stemming from the panel discussion were used to help focus and organize the subsequent specialist meeting discussions.

THE SPECIALIST MEETING

The specialist meeting was held from June 8-12, 1991 in Castine, Maine. The general format of the meeting was a series of group discussions, alternating between small and large groups and focusing on a single theme each time. Small groups were selected by the initiative leaders with the goal of mixing participants from different disciplines and affiliations. The meeting culminated with a discussion of specific research topics. Participants for the specialist meeting were drawn from academia, federal agencies, and industry and a diversity of disciplines including statistics, computer science, cartography and psychology.

The specialist meeting discussions were organized around four initial themes: data quality components, representational issues, data models and database issues, and evaluation of solutions. These themes served as the basis for directing discussion toward development of the initiative research agenda. Major points of the meeting discussions under each theme are summarized below.

Data Quality Components. Quality itself is application or context dependent. Its measures derive most commonly from statistical measures of error. A framework including positional accuracy, attribute accuracy, logical consistency, completeness and lineage was presented in the Proposed Standard for Digital Cartographic Data in 1988, and subsequently adopted in the data quality report section of the Spatial Data Transfer Standard (SDTS, FIPS 173).

Meeting participants discussed what quality components are important to users, how different quality components might be related, and methods for quantifying or qualifying quality components in ways that would make them amenable to visualization. The group concluded that at least two components, spatial and temporal resolution, were not well emphasized in the SDTS framework. The temporal component of data quality, what may be called currency, should be separated from the lineage category and rendered as a sixth component, due to the complexity of temporal data sources and database update operations.

Representational Issues. Discussions under this theme covered not only graphical representation but also cognitive and perceptual issues. Traditional cartographic techniques are now being supported and expanded by new visualization techniques, including hypermedia, animation, and virtual reality implementations. A good deal of attention was directed to real time user access to data quality displays during GIS operations. Significant improvements in display resolution, speed, parallel processing and support for color lead the list of technological improvements providing opportunities for other experimental representation techniques.

One of the significant points arising from specialist meeting discussions on visualization tools for representing data quality relates to the needs of users. Whereas many participants brought to the discussion the attitude that research should determine appropriate depiction techniques for specific data quality categories, it was realized that the more flexible solution is a toolbox approach, wherein users can select from a variety of display alternatives to suit their needs for data quality displays.

Data Models and Data Quality Management Issues. Management of data quality within a GIS database is required during manipulation and update, and the need to store quality information will likely impact upon the future architecture of such databases. Information about the information within a database, referred to as metadata, has recently become a research issue in its own right. Discussion at the specialist meeting revolved around several issues, including development of user interface designs facilitating data and metadata queries, the level of aggregation at which data and metadata should be linked to optimize database updates, and where in the data management process visualization can be used most effectively.

Evaluation Paradigms. The research questions considered important under this topic were not formulated in the stimulus-response paradigm of psychophysical research so common in the last quarter century of cartographic literature. Instead, participants suggested development of software tools to de-focus or fog imagery, with responses elicited by thinking aloud techniques such as protocol analysis to decipher user reactions to working with the tools. This approach lies more centrally in the domain of cognitive science than perceptual psychology, indicating a possible shift in the types of research questions to be reported on in future empirical research.

PROGRESS ON THE RESEARCH AGENDA

The specialist meeting generated a research agenda which was published in a center technical report (Beard et al. 1992) The research agenda covered eight major themes:

- *conceptual frameworks* -research on defining and matching quality components with visualization tools;
- *error modeling and monitoring*-research on error models and error propagation during GIS operations and their potential for subsequent visualization;
- *development of visualization techniques*-research on new or modified graphic displays for data quality components;
- *user needs and expectations*--research investigations on users requirements for data quality information;
- *role of data quality visualization in decision-making* -research on how availability and access to data quality information impacts decision making and dispute resolution;
- *educational tools and tutorials on spatial data quality*-research on and development of instructional aids for promoting understanding and use of data quality information
- *database support for quality information*-research on new or extended data models to support storage and access to data quality information
- *evaluation of visualization techniques*-user surveys of visualization techniques to assess their communication effectiveness.

Conceptual Frameworks

Visualization should serve the purpose of clarifying and enhancing understanding. To accomplish this objective, the phenomena to be visualized and the context of the visualization should be well understood. This research theme formed the foundation work for the initiative and concerned identification of data quality components, identification of visual expressions and tools, the logical mappings of quality components to visual expressions, a taxonomy for access to data quality, and a philosophical framework based on communication theory.

Under this topic, a framework for specifying logical connections between data quality components and visual variables for displaying those components was developed (Clapham and Beard 1991). This work was based on formal specifications of positional accuracy, attribute accuracy, and currency using a selected set of visual variables including color, shape, and size. Following Bertin (1983), scales of measurement were used as the link between the behavior of a quality component and behavior of a visual variable. Algebraic specifications were developed for each scale of measurement. These specifications become the modular building blocks for the quality and visual variable specifications. Sarah Clapham (1992a) completed her masters thesis on this topic and final results of her work were presented at GIS/LIS '92 in San Jose (Clapham 1992b).

Buttenfield and Beard (1993) in a paper published in Visualization in Geographic Information Systems (Hearnshaw and Unwin 1993), expanded on the five components of data quality outlined in the spatial data transfer standard FIPS 173. This paper also expanded on work presented at a special session on Visualization and GIS held at the Interface ’92: Graphics and Visualization conference, College Station, Texas.

Helen Couclelis developed a philosophical framework based on Searle's Speech Act Theory for understanding data quality needs and requirements for research. A key point in speech act theory is that conversation creates patterns of commitment between partners. The interaction involves giving partners information about the degree of validity of speech acts. In the analogy proposed by Couclelis, speech acts would have counterparts in graphical displays, icons, menus, messages and other interactive devices. The goal is a visual semiology analogous to the rich system of qualifiers of speech acts used in human conversation to indicate doubt, hesitation, uncertainty, or error. Couclelis suggests some of the following mappings of speech act qualifiers to graphic symbols. Adverbs, expressing precision or vagueness, would map to sharpness or fuzziness of symbols. Hedges, indicating force of belief, would map to color saturation. Proddings, to express disbelief, would map to special icons or prompts, and probes into hidden or misunderstood information would map to brushing techniques or possible occlusion of information. These concepts were published in an NCGIA technical report (Couclelis 1992) and in the International Journal of GIS (IJGIS) (Couclelis 1993).

In another framework, published in Cartographica, Beard and Mackaness (1993) define components of spatial data quality and identify three levels of access to quality (notification, qualification, and quantification) to these components. Notification offers the simplest, least intrusive access by just notifying users that there may be quality concerns. Qualification expands this by indicting the nature of the problem. Quantification provides quantitative assessment of the magnitude of the problem. Each access level implies different visual expressions. This framework also distinguished three contexts for data quality assessment: a production context, a decision making context, and an exploratory context. Each of these contexts also warrants different levels of visual interaction with data quality, ranging from simple, routine techniques for the data producer to intensive levels of interaction for the data explorer.

Wu (1993), in his PhD thesis, addressed the domain-dependent differences in requirements for data quality information in digital soils, digital terrain, digital satellite, and digital vector products. His research designed a query structure for data quality that incorporated knowledge about the types of objects stored within each data product, and explored the database storage requirements needed to respond to each type of query. He focused on lineage information, extending Lanter's (1990) research on an object oriented framework.

Error Modeling and Monitoring

Work under this theme concerned development of error models whose outcomes would be conducive to visual expression and effective communication to GIS users. Much of the work under this theme extended work from Initiative 1. Work by Goodchild and graduate students at Santa Barbara considered three possible options for alternative visualizations of error models. These were: 1) displays which omit any reference to quality, for example by displaying maximum likelihood classes; 2) displays which include descriptions of data quality in the form of error model parameters; or 3) displays which present a sample of realizations of the error model. The last option was determined to be the more dramatic, and in the spatial context more intelligible to non-statistical users. Techniques were developed for displaying realizations under the error model developed by Goodchild and others (1992) for Initiative 1. Preliminary results were published in (Hearnshaw and Unwin 1993).

Yee Leung (a visiting researcher from the Chinese University of Hong Kong) and Goodchild studied the display and analysis of fuzzy-classified scenes. Although the advantages of fuzzy classification are generally accepted, no consensus appeared in the literature on how to display these effectively, or analyze them within a GIS context. Following from this work, Chi-Chang Lin built a working system for the visualization of fuzzy classified images. A multinomial probability field (a vector of class membership probabilities) was computed for each image. The system supports the ability to display several individual realizations of the error model with user determined levels of spatial dependence. This display method focuses attention on variation between realizations rather than on the average or expected case. A display of four to six realizations in different windows on the screen allows graphic illustration of the implications of uncertainty in spatial data and draws attention to its influence on analysis, modeling, and decision making. This has recently been packaged as a module within GRASS and will be available over ftp.

This work is being continued by graduate student Chuck Ehlschlager, who is developing a series of GRASS modules for error modeling, using various random field models. The work will be extended to the distortion of vector objects, and measurement of vector object positional uncertainty, in Ehlschlager's dissertation research. Goodchild gave presentations on this research at the University of Colorado, at the Second International Conference/Workshop on Integrating GIS and Environmental Modeling in Breckenridge, and at the Institute of Navigation Conference in Salt Lake City in 1992.

David Lanter also studied error in spatial inferencing software, induced by choice of data structure, and the effects on resulting spatial analysis. Results indicate that data structure induced error can be made explicit, measured, visualized, and reduced in grid cell-based viewshed algorithms. This provides a basis for determining and representing lateral dimensions of error in the viewshed. Combining this with a statistical vertical error index opens up the possibility of 3-dimensional error visualization and analysis. Algorithms for this study were developed with Paul Sorensen in LISP on IBM RS/6000s and reported in Sorensen and Lanter (1993).

At the University of Zurich, Rob Weibel is leading a multi-year research project funded by the Swiss National Science Foundation, dealing with the use of environmental models where information must be integrated from (often incompatible) multiple data sources. Data quality forms a large component of the project, which in the first years of the project emphasizes climate change and its impact on ecosystems in alpine regions. Barbara Buttenfield spent a week in Zurich in May 1993, funded by the Swiss to advise on the data quality aspects of the project, and to inform the Swiss research team on the NCGIA I-7 research agenda. Project MINT (Multiple dataset INTegration) seeks to implement techniques to resolve incompatibilities arising from differences in georeferencing, the temporal framework, and in differing attribute categories or category definitions. The first paper reporting this work to an American conference appeared in the AUTO-CARTO 11 proceedings (Stephan et al. 1993). While this work was not funded by NCGIA, it has funded NCGIA I-7 work through travel, and its focus emphasizes topics of relevance to Data Quality representation and display. For these reasons, and to inform the GIS community, it is cited in this report.

Development of Visualization Techniques

The initiative undertook development of specific techniques for visualizing data quality. Following from the earlier conceptual framework research, visualization techniques were designed to capture the behavior of both existing error measures and new error models developed under this initiative. Bouchra Miri (1993), a graduate student at the University of Maine, developed a visualization technique to display homogeneity in soil maps. The visualization technique focused on attribute error and particularly considered the constraint of scale on the graphic display. Miri defined a range of scales and corresponding visual presentations. Under these schemes, soil inhomogeneity at small scales was represented by variation in gray scale value or color saturation as a function of soil map unit impurity. At medium scales, the visual display changed to point symbols indicating the rough location of soil inclusions within map units. At the largest scale, areal delineations of soil inclusions were displayed and shaded to indicate inclusion type. Figures 1 through 3 show this progression.Thus with increasing scale more detailed quality information was made available to users. It should be noted that detail on soil inclusions is not reported beyond the percentage for map units or point symbols, so the addtional quality information displayed under this scheme was simulated. This work, implemented as a macro program within Arc/Info, was the winning student entry in the Visualization Challenge, held at GIS/LIS'93.

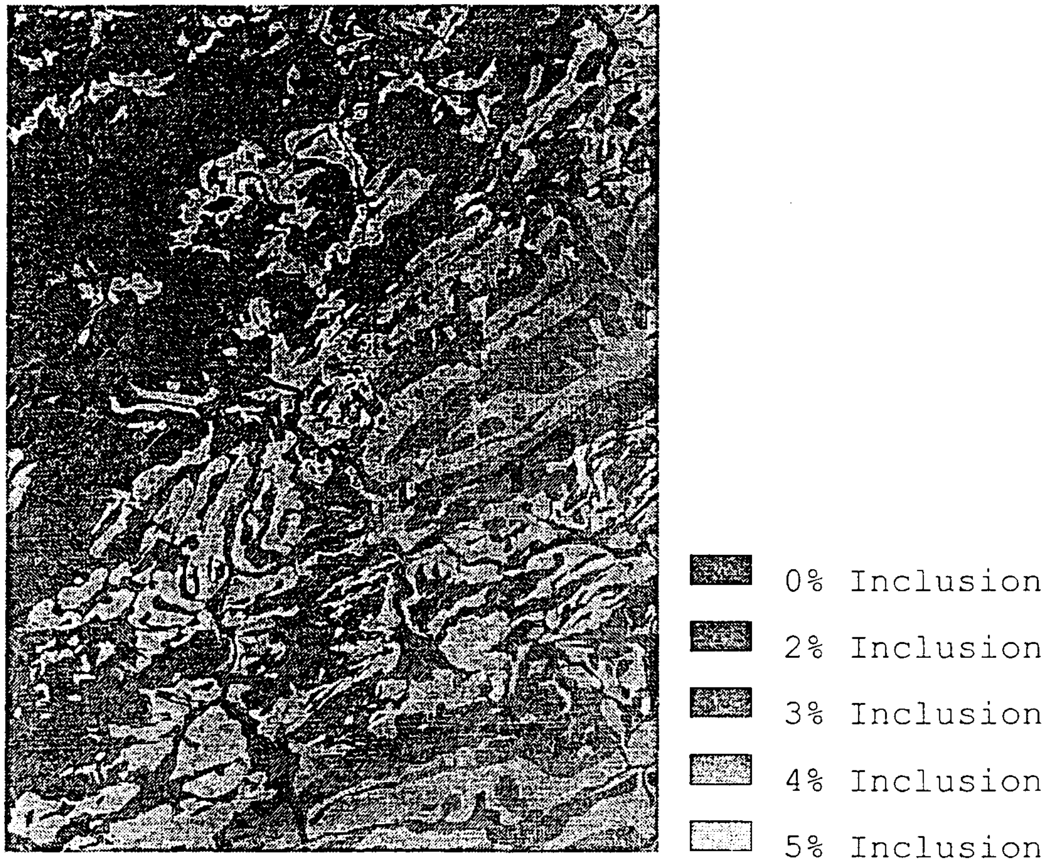


Figure 1. Small scale display: Soil map inclusions displayed as a percentage by shading applied to the map units.

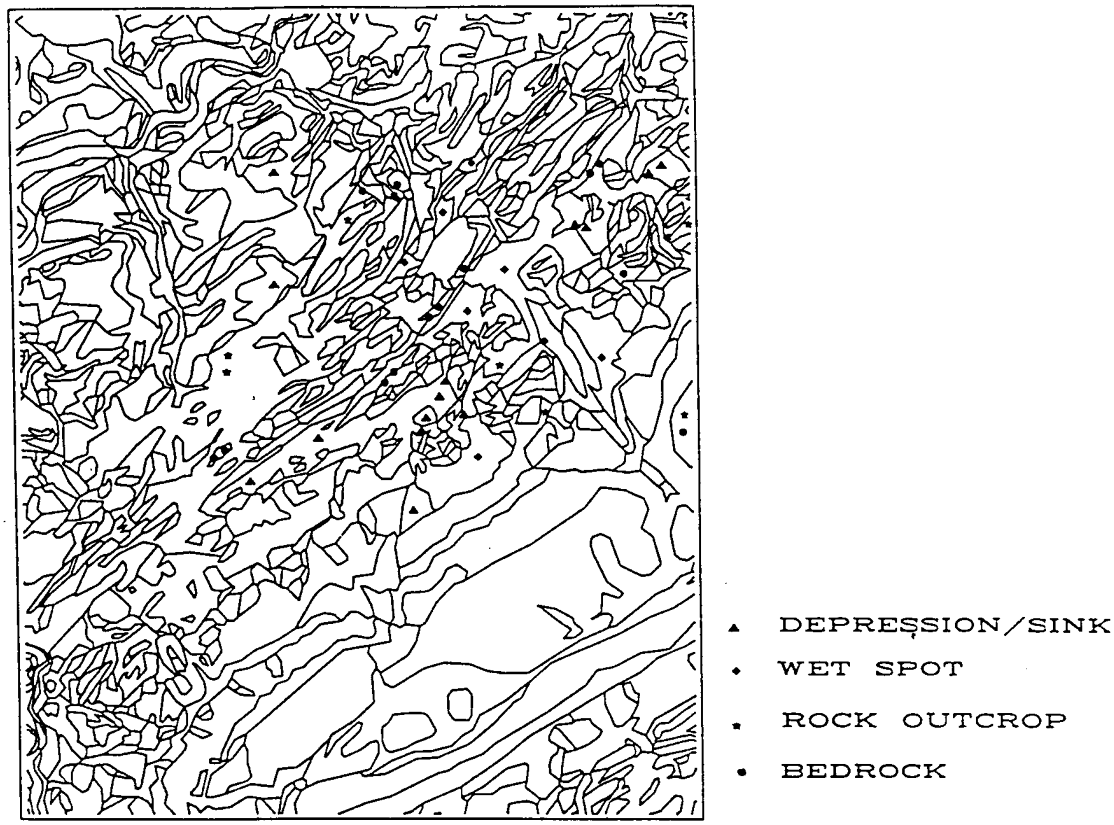


Figure 2. Medium scale display: Soil map inclusions displayed as point symbols indicating location and type of inclusion.

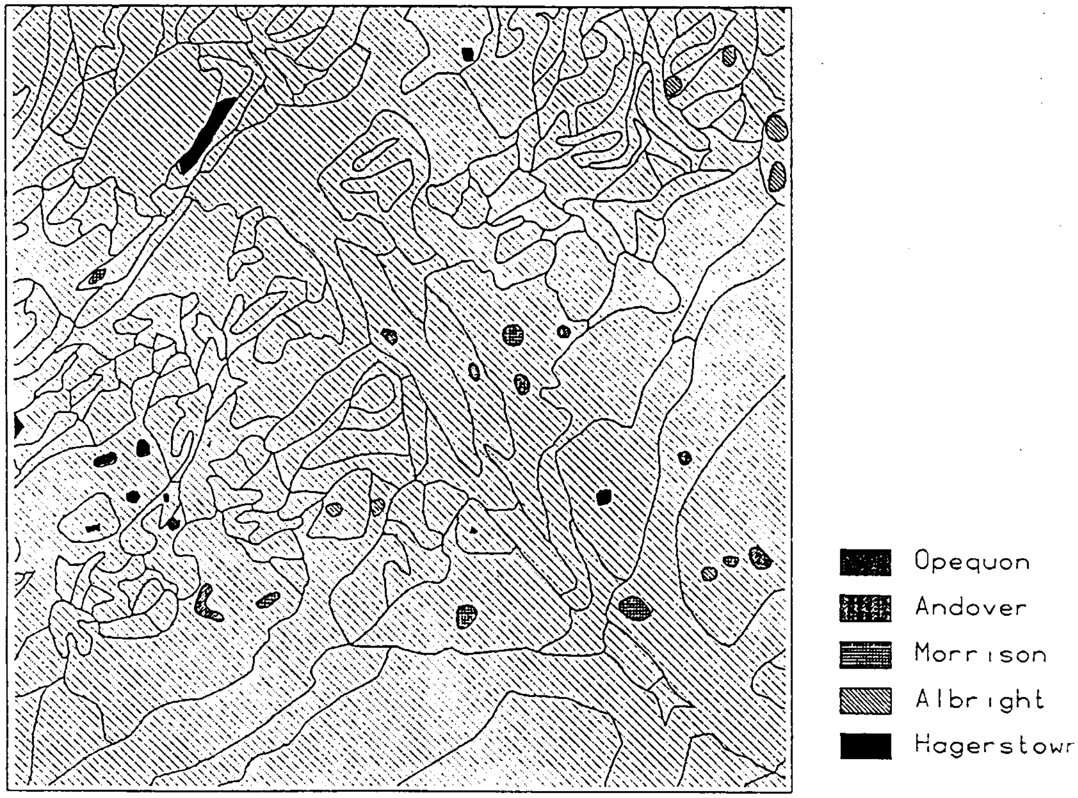
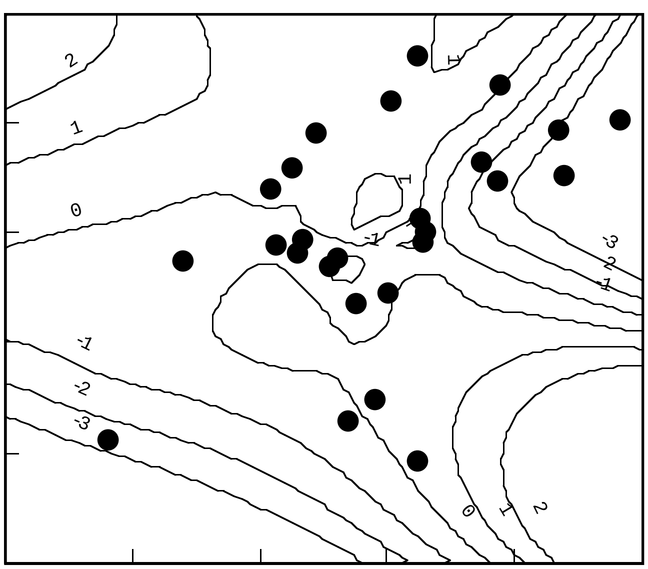


Figure 3. Large scale display: Soil map inclusions displayed as polygons indicating location, areal extent, and type of inclusion.

Lori Dolby (1993) incorporated several techniques within her master's thesis work to display uncertainty in seabottom sediments. This work was implemented as an Arc/Info macro. Bottom sediment facie (bedrock, glacio-marine and Holocene) were interpolated from seismic profile data. These facie surfaces can be displayed and queried with respect to depth, and sediment type. Two overlay techniques were added to the display to help inform users of uncertainty in the interpolated surfaces. In one case the uncertainty (defined here as estimated kriging variance) was displayed as color shades with data values as contours. The opposite display is also available to users in which data values are displayed as color shades and uncertainty is displayed as a contour overlay.

William Mackaness and Kate Beard expanded on this work and looked at visualization techniques to display interpolation uncertainty more generally. The primary objective was to capture behavior of the interpolation technique in the visual display. As interpolation uncertainty is a function of sample density, sample point distribution, interpolation method, as well as the distribution of the phenomena being observed, the contribution of each of these factors to uncertainty, excluding the spatial distribution of the phenomena, should be displayable. A paper presented at Auto Carto 11 (Mackaness and Beard 1993) introduced these concepts. Assuming the interpolated data are visually displayed as contours, the contributions of the uncertainty factors are included by 1) displaying the sample point locations as an overlay, 2) displaying the sample points and contours by values. By this approach large deviations between observed and predicted values become apparent, 3) displaying a gray scale uncertainty surface computed using the same distance function as the interpolation distance function 4) by mapping uncertainty intervals to contours through contour symbol changes. These techniques are illustrated in Figures 4 through 6.



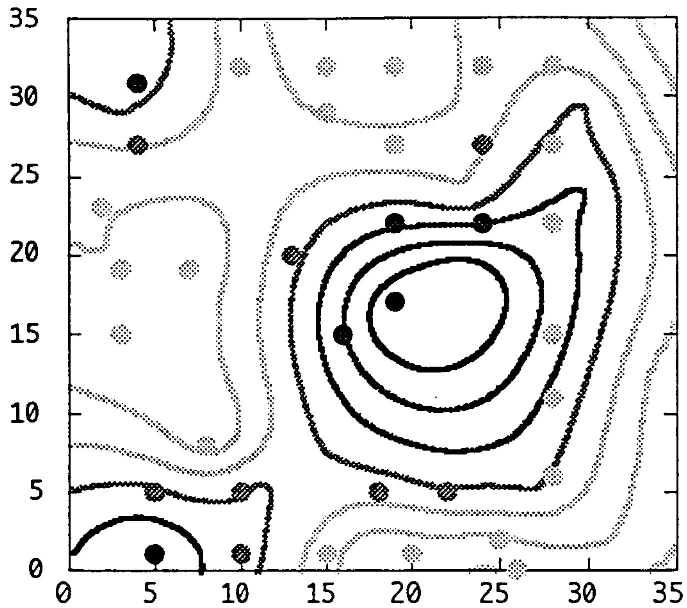


Figure 5. Display of sample points and contours using corresponding symbol shades.

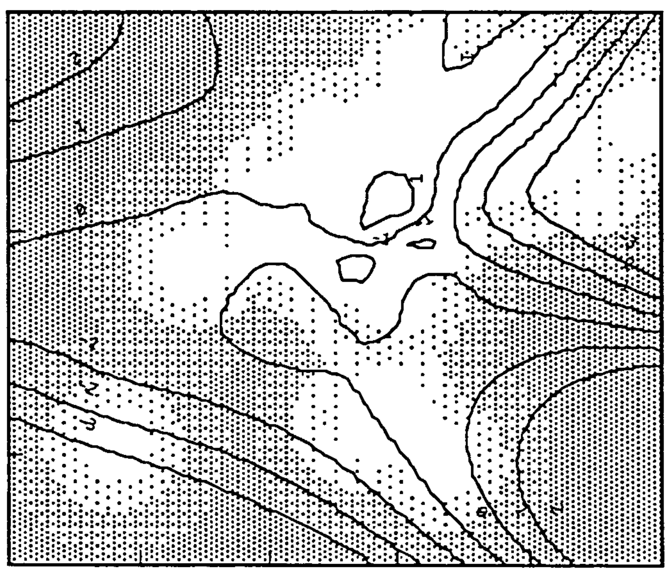
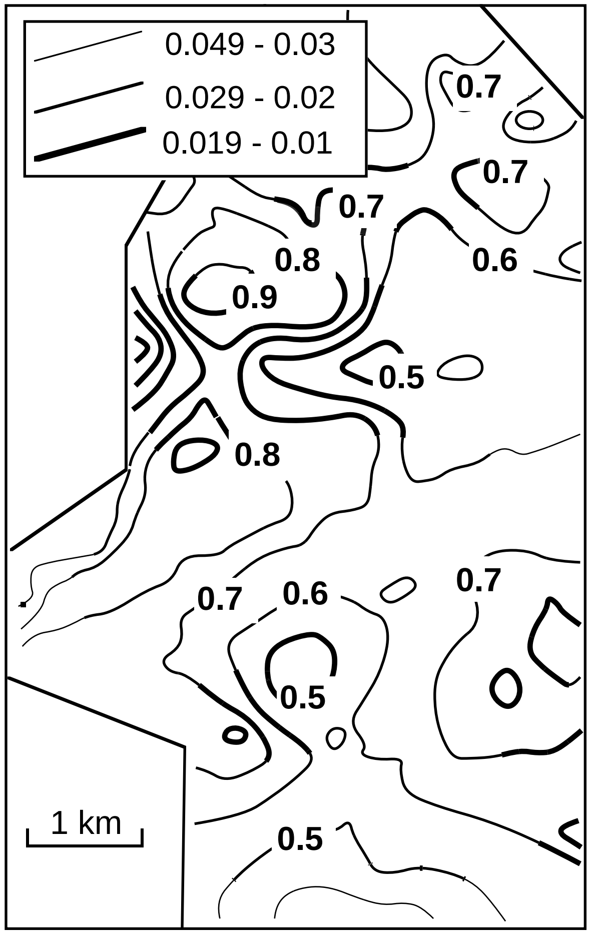


Figure 6. Display of a gray shade uncertainty surface which captures the behavior of the distance function employed by the interpolation method.



Combination of kriging error estimates and data variable. Data variable displayed by contours. Uncertainty displayed by line width.

Gary Hunter and Mike Goodchild proposed visualization techniques for displaying uncertainty in 7.5 minute DEM data. Using the published 7m RMSE for the DEM data, probabilities were computed for each cell in the raster surface exceeding or being exceeded by a given threshold. Various displays were then created using the resulting probability values. For example a color ramp from blue to brown was generated to illustrate the probability of cells being inundated (blue) and those likely to remain above water level (brown). This work has been published in the journal Photogrammetric Engineering and Remote Sensing (PE&RS).

Khaled Hassen, a PhD student at Maine, has been developing a visualization technique, referred to as reference grid, to display positional change generated by GIS processes. The reference grid is initially a uniform grid which can be displayed with a data layer. As GIS processes which alter geometry are applied to the data, the grid is distorted to visually capture the nature (location and magnitude) of this distortion. Reference grid mappings have been developed for edge-matching, line generalization (Douglas-Peucker) and fuzzy polygon overlay operations. In each case the positional change generated by the GIS

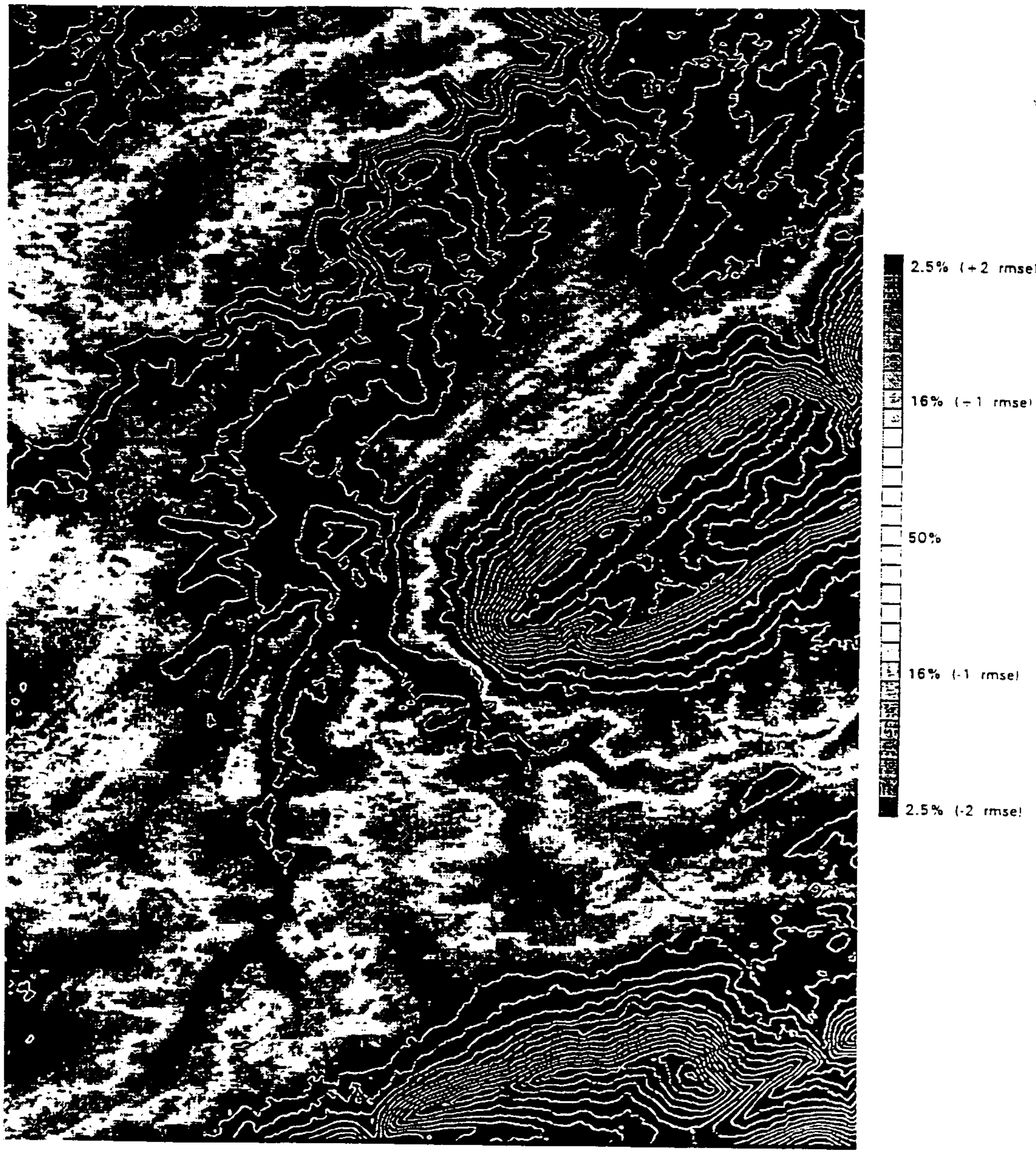


Figure 8. Uncertainty of 350m contour based on published 7m RMSE. Other contours shown normally for reference.

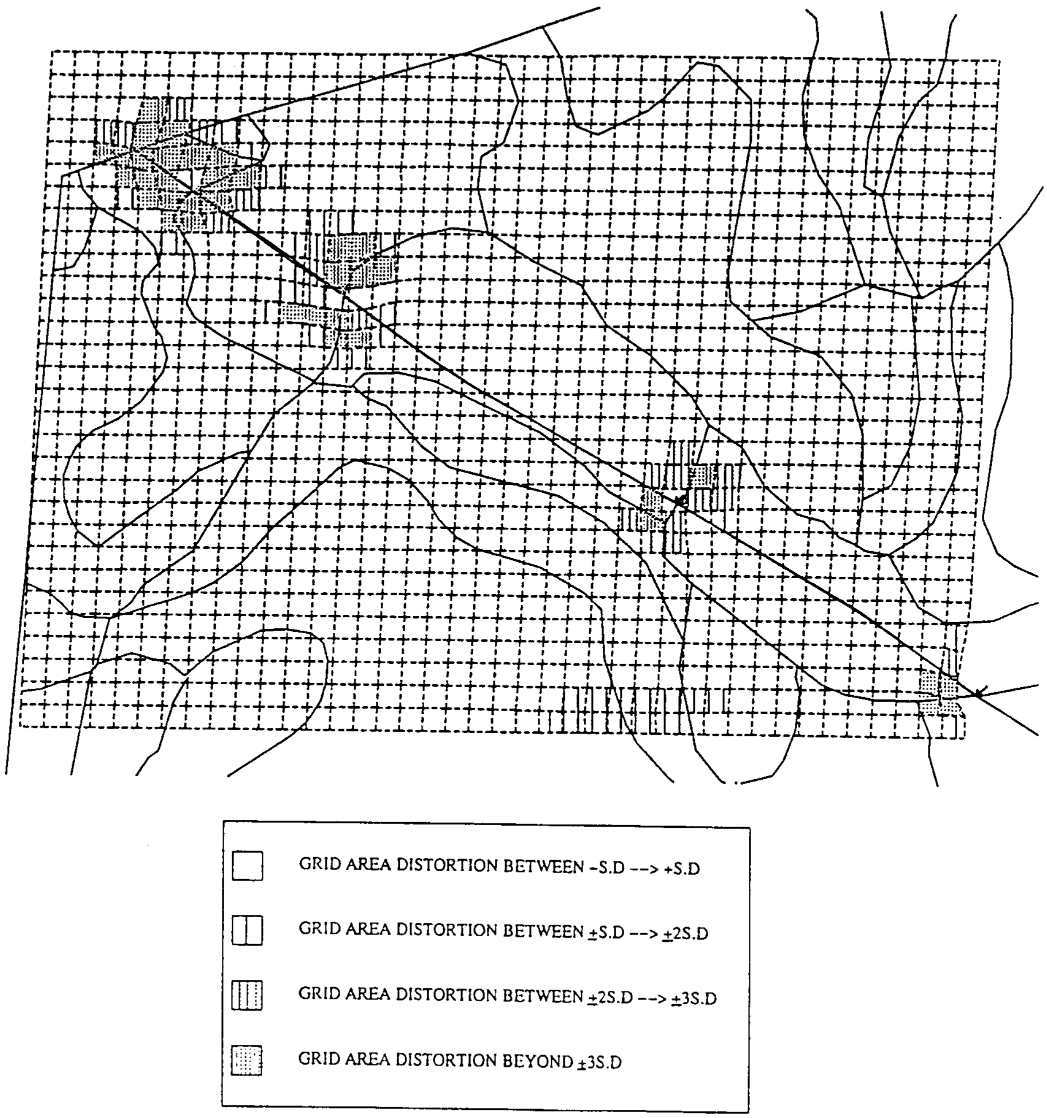


Figure 9. Shading of a reference grid to highlight the areal distortion in the grid from the initial uniform cell area.

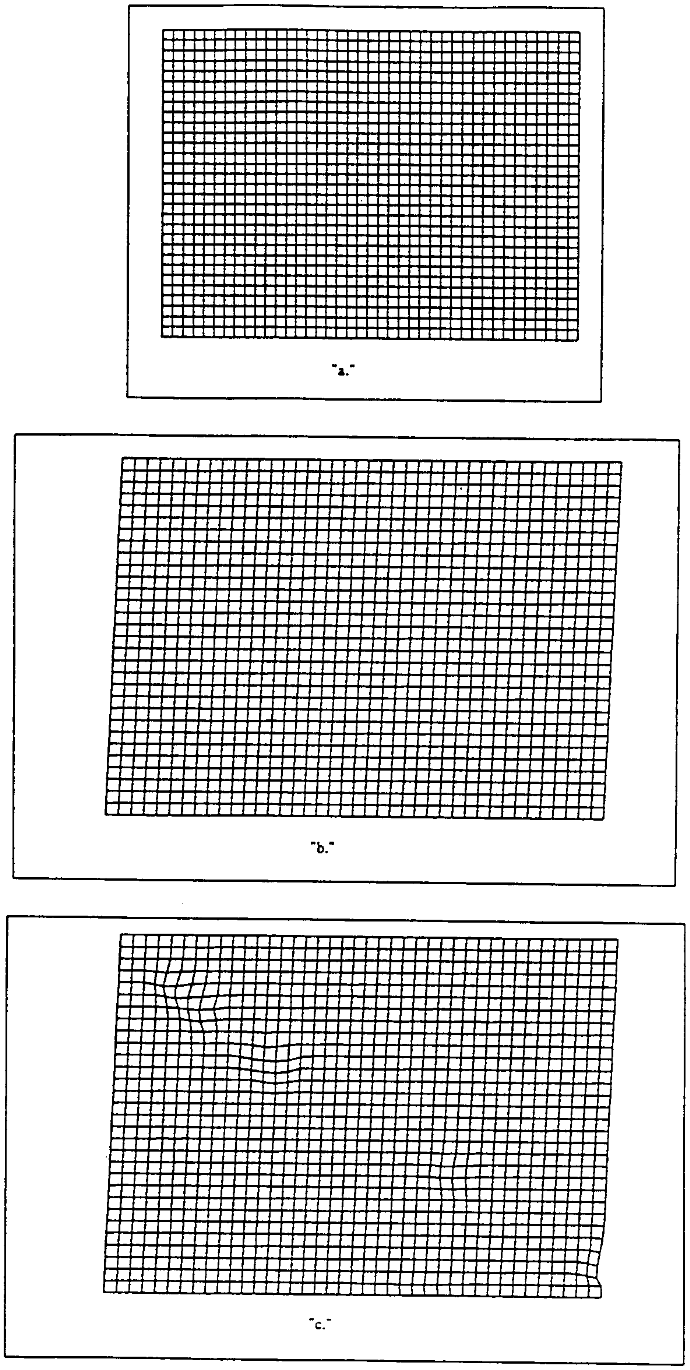


Figure 10. A reference grid summary sequence showing a) initial grid, b) grid after linear coordinate transformation, c) grid after rubber sheet edgematching.

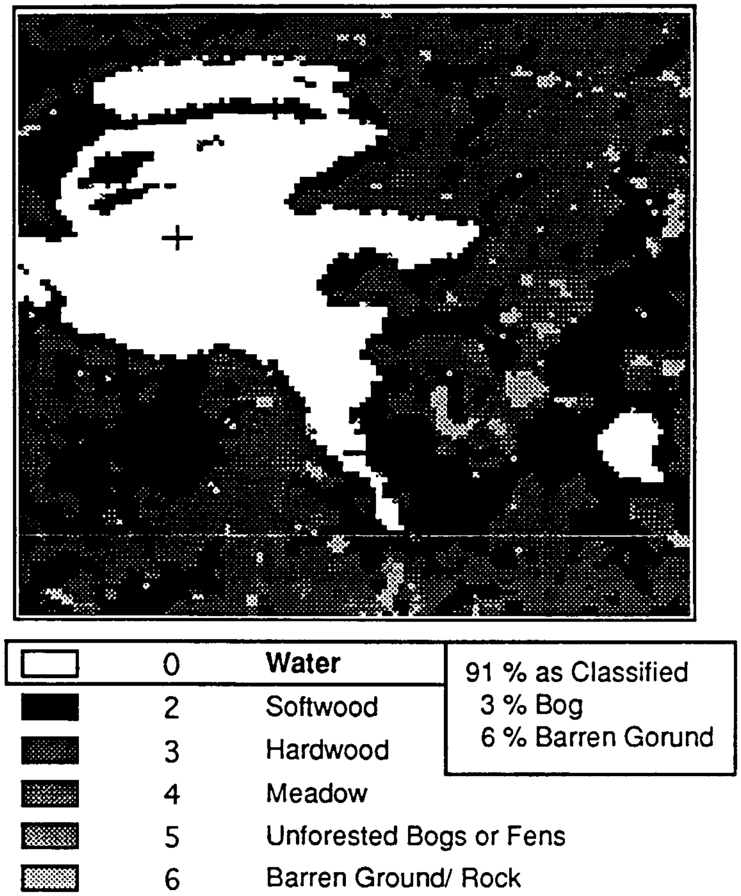


Figure 11. Selection of a location within a class highlights all instances of the class (water in this example) to indicate the uncertainty measure applies at the class level.

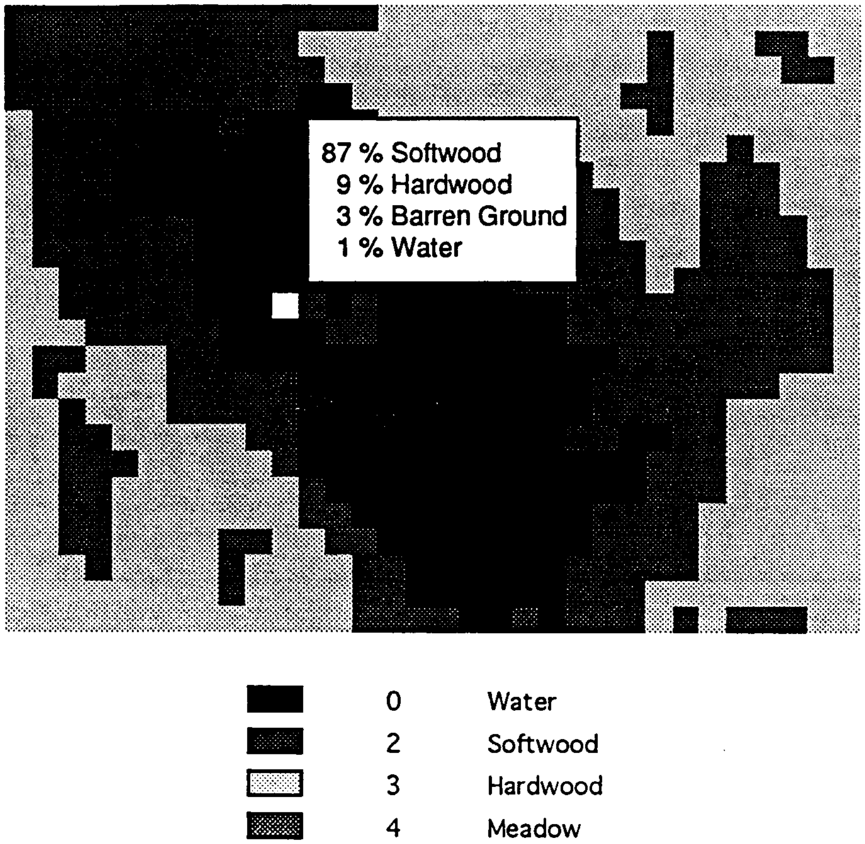


Figure 12. Using a more refined categorical error model uncertainty can be queried at the pixel level as indicated by a single highlighted pixel.

operation is simulated in the grid by a piecewise linear rubber sheet transformation. The role of the grid is to indicate transformations of the data which may affect quality but are typically not made explicit. A journal article reporting this work is currently under review.

Responses to queries within a GIS should include uncertainty estimates, but currently few if any do. Beard (1994) has proposed uncertainty responses for a basic GIS query ' What is the value of a variable at location X?' In the near term, the uncertainty response can exploit existing error measures (e.g. RMSE, PCC, or variance estiamtes from kriging) as these are more likely to be immediately available. As more spatially explicit measures become routinely available, these can be substituted. The concept works as an interim solution since the query responses are supported by visualization displays which illustrate the behavior (limitations) of particular error measures. For example if PCC, which only reports uncertainty at the class level, is the available measure, user probes of a location will highlight all instances of the probed class and report the class uncertainty. If the Goodchild et al. (1992) model is available, a probe to a location highlights a single pixel and reports the vector of probaibilites at that pixel.

Several other visualization techniques for data quality were generated as a result of the NCGIA Visualization challenge. Seven entries were submitted and results were published in the Proceedings of GIS/LIS'93. These are discussed in more detail under the section on the Visualization Challenge.

User Needs and Expectations

Victor Wu and Barbara Buttenfield studied the impacts of various GIS overlay operations upon specific types of data quality. The domain of interest centered on natural resources data, particularly soils maps. An interview with staff in the New York State SCS offices in Syracuse uncovered a need for more detailed access to data quality information than is commonly available in layer-based GIS data models, and Wu's doctoral dissertation presented basic requirements for an object-based system supporting analysis of data quality information. A journal article is in preparation on this topic.

Role of Data Quality Visualization in Decision-Making

Quality of spatial information is a function of the specific purpose of each application so absolute thresholds should not be imposed in a system. In a decision making context, users should be able to specify the minimum requirements for quality that satisfy their application needs. Part of the research under this theme investigates how visualization techniques can be made flexible for users to explore and evaluate data sets with respect to their requirements or conversely how visualizations may help educate users on data limitations. A second topic of concern under this theme is to better understand the impact of availability of quality information on decision--

making. The question being, Does access to this information, in fact, make a difference?

Barbara Buttenfield and Victor Wu considered the role of metadata acquisition, following from Couclelis's discussion of Searle's Speech Act theory at the Castine meeting. Extending this into the realm of decision-making during database query, it becomes evident that several types of decision-making can occur, depending upon the risk involved in the decision and the amount of available information. Interpretation of positional accuracy for example depends upon deterministic tests with fixed tolerance thresholds, and users can rely upon established conventions of inferential statistics to interpret fitness-for-use. Lineage evaluation however may be associated with a large amount of information that is nonetheless descriptive and may require semi-structured interpretations, thus attaching a higher risk to deciding on fitness-for-use. They presented commonly encountered human-computer dialog examples to demonstrate that more flexible dialog schemes can improve the possibilities for decisions based upon fitness for use. An article on this research is in press for the journal Computers, Environment and Urban Systems.

In 1993, Gary Hunter (University of Melbourne) spent six months at Santa Barbara working with other I7 researchers. Several papers were written on the management of uncertainty in spatial data and GIS, and on its impact on decision-making when uncertainty is visualized. A series of case studies were made of the impact of uncertainty on specific GIS modeling and decision making projects. The first of these papers by Hunter and Goodchild appeared in the URISA Proceedings for 1993; it won the Horwood Critique Prize, and was subsequently published in the URISA journal. This work is being continued by Chuck Ehlschlager, who has been experimenting with animations of error model realizations for DEMs. Much of Hunter's work also appears in his PhD thesis, recently accepted by the University of Melbourne.

Jeff Paradis, a graduate student at the University of Maine, is working on the concept of a Data Quality Filter for his masters thesis. The data quality filter is designed as an aid to decision-making. As data quality is context dependent, the filter allows a decision maker to specify a data quality component, measure, and threshold (e.g. positional accuracy, RMSE, .5) and to test a data set against this specified criteria. Only data which pass the filter are displayed, so when user quality criteria exceed data quality (a set of measures) the result is a blank screen. This objective is largely educational as the user is enlightened as to the quality of the data with respect to their expectations or requirements. If a data set does not meet their requirements they can test an alternate data set or modify their criteria. This work has been accepted for publication in the URISA journal (Paradis and Beard 1994).

Educational Tools and Tutorials on Spatial Data Quality

Kate Beard, Sarah Clapham, and David Steiner developed The Data Quality NoteBook (a Hypercard stack which provides a visual and animated tutorial on spatial data quality) as part of the initiative. The Data Quality Notebook uses the metaphor of a small spiral notebook in which the users flip through the pages to find the information they need. It provides definitions of quality, spatial data quality, and spatial data quality components. The main purpose of the Notebook is to track specific data collection methods and illustrate how collection and processing activities affect spatial data quality. Two case studies in the notebook (remote sensing and soil mapping) document the steps of data collection and compilation into products which could be incorporated in a GIS. A matrix is used to show how each compilation step affects positional accuracy, attribute accuracy, completeness, and logical consistency.

Andre Skupin and Barbara Buttenfield designed a prototype SDTS Browser containing the text of the SDTS FIPS-173 standard, embedding hypertext links to provide definitional cross-references and indexing. Andre began the project in a graduate seminar on knowledge representation. Barbara Buttenfield completed the prototype and demonstrated it to the staff of the Standards and Technology Branch at USGS National Mapping Division. They are currently implementing a version on MOSAIC to be accessible via internet. Skupin is currently studying the geometry of the Browser's hypertext linkage map to uncover logical flaws and redundancies in the Browser for future versions. A journal paper is in preparation co-authored with Buttenfield.

Data Structures and Models for Quality Information

There are two aspects to this research theme both focused on more explicit capture of data quality information. Current GIS maintain a minimal level of quality information usually consisting of limited documentation on source data characteristics such as scale, producer, sensor, datum, etc. Some information can be captured on processing steps (Lanter 1991) but rarely are the effects of these processes on quality documented. As GIS databases mature through update cycles, the quality will only become more heterogeneous and complex to represent. New database structures are needed to capture quality information more explicitly and efficiently so neither producers nor users are over burdened by this task.

At Buffalo, Victor Wu's dissertation considered the role of metadata acquisition and management in environmental databases. He presented metadata acquisition in static and operational states. Static state acquisition refers to accuracy information, missing values determination and similar measures which often can be archived quantitatively and which rely on established conventions for testing. Operational state acquisition refers to some measures of logical consistency and to lineage which form a richer source of data quality information but require a different, more flexible database architecture to incorporate qualitative and sometimes open-ended descriptive or chronological information.

Geoff Dutton and Barbara Buttenfield analyzed coordinate conversion between world cartographic feature databases commonly used for mapping and GIS applications. Working with two databases, the World Data Bank II and the database supplied with Atlas Pro (a commonly used microcomputer mapping package) they established ranges of variation in positional accuracy that may accrue when converting latitude-longitude pairs into and out of QTM space at differing levels of resolution. Results of this research were published in the ICA-Cologne Proceedings as guidelines for protection of positional accuracy (Buttenfield and Dutton 1993).

A second aspect of this research theme relates to the uncertainty and thus quality implications introduced by the data modeling process itself. A GIS depends on conversion of geographic reality to finite digital representations. Each implementation of a finite representation theoretically imposes a specific model of uncertainty. Because many GIS databases are still based on a map model, the uncertainty information is ill defined because of subjective components in the map making process. To support visualization of model uncertainty, new models are required which explicitly capture the associated uncertainty.

Related to this area, Bud Bruegger (1994), in his PhD dissertation, developed a new spatial theory based on a resolution-limited space. As part of this theory, he defined a set of associated representations, metadata, and transformations which capture the uncertainty introduced by limited resolution sensors. In the thesis, he demonstrated how these limited resolution representations and their corresponding uncertainty can be displayed. The limited resolution model relies on disks and sets of disks rather than Euclidean point sets. The attributes of disks are attribute mixtures in the certain case and mixture balls (intervals of mixtures) in the uncertain case. Mixtures and mixture balls are partitioned to create certain and uncertain feature partitions respectively. In both cases, features of a partition are separated by areal transition zones rather than the traditional sharp boundaries. The complement of mixtures associated with a feature is known with certainty and all uncertainty is absorbed in the transition zone. As uncertainty imposed by the resolution limitation is explicit within the structure (representations and metadata), it is more directly available for visualization.

Evaluation of Visualization Techniques

Diane Schweizer (1992) investigated the use of color spaces to display uncertainty in the values assigned to areas in a choropleth map. An empirical research design tested groups of students in attempting to determine whether progressions of color saturation or color value (darkness) are associated with variations in certainty about perceived magnitude on choropleth maps. The results indicate that viewers do not associate color saturation with data magnitude, nor do they associate color value with certainty. When a map legend is added to the display, viewers made correct choices more often when saturation is used to represent certainty. Diane published this work as an MA thesis (Schweizer 1992) and conference paper (Schweizer and Goodchild 1992).

David Lanter studied the use of alternative contour symbologies in bivariate mapping. Bivariate maps provide a useful display of relationships between two statistical variables generated in spatial decision support systems. Lanter's research focused on two alternatives for using contour lines to focus attention on statistical interrelationships. Experiments tested the effectiveness of these two techniques and compared them to standard bivariate mapping techniques.

Barbara Buttenfield and Tom Kress developed an electronic tutorial simulating a slide show on major categories of spatial data quality. Each category (positional accuracy, attribute accuracy, completeness, etc.) is represented by a 'hot' piece of type. Selecting the hot word brings up displays illustrating a variety of methods by which each type of data quality might be displayed. Tom Kress (1991) completed this project as an MA thesis.

At Syracuse University, Mark Monmonier has supervised student research on design of multivariate map symbols that incorporated data and data quality measures. Results of this work were presented in John Hancock's MA thesis (Hancock 1994).

RESEARCH PRODUCTS AND DELIVERABLES

Specific initiative deliverables which were identified by the specialist meeting included: an initiative newsletter, journal special issue, conference panels, an initiative bibliography, and a competition to stimulate creation of new visualization techniques.

Newsletter

A newsletter was organized to provide notice of announcements, publications, and events related to the research topics. The newsletter was intended to provide a forum for updates on recent research activities. Two issues per year (May and November) were scheduled to be available through electronic mail and/or page layout form with graphics. Geoff Dutton agreed to act as editor. The first issue came out on schedule and on budget. Although he solicited feedback and input for subsequent issues, no input was received, and subsequent issues were not generated.

Special Journal Issues

Barbara Buttenfield guest-edited a special issue of the journal **Cartographica** on Mapping Data Quality; the issue is Summer 1993 but did not appear in print until February 1994. Articles in that issue provide a representative sampling of research ideas on representational issues and development of visual tools for data quality display. Authors include Buttenfield, McGranaghan, Fisher, Monmonier, and Beard and Mackaness. Their work is summarized above in the section on Progress on the Research Agenda.

A second journal special issue of the journal **Cartography and GIS** was guest-edited by Kathryn Wortman, USGS NAtional Mapping Division, and Barbara P. Buttenfield, NCGIA-Buffalo. The issue is titled "Current Developments and Use of the Spatial Data Transfer Standard". The focus of this issue is on acceptance of SDTS in various GIS communities (state, national and international) on implementation of the standard in specific software contexts (GRASS), on profile development for vector and raster data, and on evaluation and conformance testing procedures. Papers were solicited from public, private and academic sectors. The issue was funded by USGS National Mapping Division. Paper solicitation, peer review and editorial revision were completed during Barbara Buttenfield's sabbatical leave at USGS in Reston. Except for the paper by Kottman and Buttenfield, none of the papers cite the initiative explicitly, and thus they are not included in the bibliography at the end of this report. Papers include:

Wortman, K.C., USGS
Developing Standards for a National Spatial Data Infrastructure

Tom, H., National Institute for Standards and Technology (NIST),
The Spatial Data Transfer Standard (SDTS): A Standard for Change

Kottman, C.A., Intergraph Corporation and Buttenfield, B.P., NCGIA-Buffalo,
Standards for Spatial Data Use: Similes Improve Our Understanding.

Hogan, R.L. and Wortman, K.C., USGS, Roswell, C. , DMA, and Winn, W.A., NOAA, Joint Leadership for Spatial Data Standards

Szemraj, J.A., Fegeas, R.G., and Tolar, B.R., USGS
Profile Development for the Spatial Data Transfer Standard

Cascio, J, USGS
A Spatial Features Register: Toward Standardization of Spatial Features

Lazar, R. A., USGS
Conformance Testing for the Spatial Data Transfer Standard

Stigberg, D. Construction Engineering Research Laboratories (CERL)
AN SDTS Implementation for GRASS

Musselman, C. Prince William County Office of Information and Mapping Resources
Implementation of SDTS in the Commonwealth of Virginia

Miller, D. and Hume, R, Australasian Spatial Data Exchange Centre (AUSDEC), Internationalizing SDTS: An Australasian Experience.

Moellering, H., Ohio State University
Continuing Research Needs Resulting from the SDTS Development Effort

Bibliography

An annotated bibliography has been compiled jointly at Maine and Buffalo and is published as an NCGIA Technical report (Mackaness et al 1994). The list explores literature across disciplines considering notions of data quality and visualization concepts and techniques from statistics, scientific computing, medical imaging, automated manufacturing, dynamic process analysis, and graphic design.

Conferences

Many presentations were made on I7 research during the lifetime of the initiative. A conference on Spatial Issues in Statistics was hosted by Statistics Canada in Ottawa, November 12-14, 1991. The keynote presentation was made by Michael Goodchild, and Barbara Buttenfield discussed the I7 research agenda in a paper session organized by Joel Yan of Statistics Canada.

Several presentations were made on I7 research topics at the 1992 Interface Conference in College Station Texas. Interface meetings are devoted to presentations at the boundary between statistics and computer science. In a session organized by Barbara Buttenfield and chaired by David Mark the following papers were presented: Spatial, Statistical and Graphical Dimensions of Data Quality (Beard and Buttenfield 1992) Visualizing the Uncertainty in Multinomial Fields (Goodchild 1992), Integrating Maps and Statistical Graphics in Graphic Scripts (Monmonier 1992). These papers were subsequently published in the conference proceedings.

At the Fifth International Symposium on Spatial Data Handling in Charleston, SC, the following papers related to I-7 were presented: Goodchild, M. F. , Chihchang, L. and Leung, Y. "Visualization of Fuzzy Scenes and Probability Fields", Fisher, P. F. "Real-Time Randomization for the Visualization of Uncertain Spatial Information", Dutton, G. "Handling Positional Uncertainty in Spatial Databases".

Barbara Buttenfield made an invited presentation on visualization at the Second International Conference /Workshop on Integrating GIS and Environmental Modeling in Breckenridge, CO in September 1993 and Michael Goodchild's presentation at the same conference also addressed I7 issues. Buttenfield gave a keynote presentation on Visualization at the first Education and GIS Conference in Washington, DC in January 1994. Goodchild made presentations on visualizing uncertainty in workshops at the ESRI Users conference in 1991, at the Ontario ministry of Natural Resources in 1993 and in a keynote presentation on spatial database accuracy at a symposium organized by the interagency Committee on Geomatics in Ottawa in 1993. He also gave invited lectures in May 1993 at the University of California, Riverside, in April 1993 at Iowa State University, and at the Joint Statistics Conference in Boston in August 1992.

Two special sessions were organized for the April 1992 meeting of the American Association of Geographers. Barbara Buttenfield organized a Cartography Specialty Group special paper session on Cartographic Issues. Participants included Mark Monmonier (Syracuse), Alan MacEachren (Penn State), and Matt McGranaghan (Hawaii).

A special session was also organized by Dan Griffith (Syracuse) on Statistical Issues and visualization as a GIS Specialty Group paper session. Papers in this session included “Heterogeneity of sampling error in spatial data sets” (Griffith), “Controlling error propagation in spatial databases” (Giuseppe Arbia, U. la Sapienza, Rome, Italy), “Bootstrap prediction intervals for kriging” (Subhash Lele, Johns Hopkins U., and Noel Cressie, Iowa State U.) and “Decomposition of information content as a quality descriptor” (Ferenc Csillag, Syracuse University).

Visualization Challenge

Kate Beard organized the I-7 Challenge on the Visualization of Spatial Data Quality. The objective of the Challenge was to generate external interest in the problem of visualizing spatial data quality. The challenge was co-sponsored by US Environmental Protection Agency and USDA Soil Conservation Service (who each supplied data sets) and by the American Statistical Association. The Visualization Challenge was concluded at GIS/LIS'93 in Minneapolis. Eight final papers were submitted to the challenge. The winning entry was by Alan MacEachern, David Howard, Tim Taormino, Martin Von Wyss and David Askov of the Department of Geography, Penn State University. The winning student entry was from Bouchra Miri a masters student at the University of Maine. Winners were selected by a panel of six judges; Richard Becker from AT&T, Dennis Lytle from SCS, Charles Spooner from the Chesapeake Bay Program, Dr. Phillip Ross, Director of the Center for Environmental Statistics at EPA, Waldo Tobler, from the University of California, Santa Barbara, and Leland Wilkinson from Systat.

Participants in the challenge were required to use one of two data sets. The two data sets were provided by the US EPA Chesapeake Bay Program Office. and the USDA Soil Conservation Service. The US EPA Chesapeake Bay Program data set included the concentrations of dissolved inorganic nitrogen (DIN) measured at 49 stations in the main stem of the Chesapeake Bay. The sample collection period covered October, 1985 through September, 1991. Data were collected 20 times each year, biweekly in March through October, and monthly thereafter. Two or four water samples were collected at fixed depths at each station for DIN analysis. The complete data record included the latitude and longitude of the field station, date, depth of sampling, and the concentration of dissolved inorganic nitrogen. Ancillary data included digital boundary files of the Chesapeake Bay, descriptions of the data formats, and associated metadata.

The winning entry by MacEachren et al. developed a prototype exploratory analysis environment to analyze the dissolved inorganic nitrogen in Chesapeake Bay. The project was implemented in IMSL/IDL to address two questions related to uncertainty: 1) which graphic variables are appropriate for showing different kinds of uncertainty and 2) what kind of user interface is most effective.

The other challenge entries included a submission by Helena Mitasova, William Brown, David P. Gerdes, Irina Kosinovsky and Tom Baker from the Spatial Analysis and Systems Team of U.S. Army Construction Engineering Research Laboratories entitled Interactive Dynamic Visualization And Analysis Of Spatial Distribution Of Nitrogen In Chesapeake Bay. They developed a prototype interactive visualization technique to allow scientists to communicate the reliability of the interpolated spatial distribution of DIN and analyze its dependence on the spatial and temporal accuracy of the data. They developed tools to allow simultaneous visualization of the data, GIS product (interpolated volume and its change in time) and its reliability.

Howard Veregin, Paul Krause, Robin Pandya and Rebecca Roethlisberger from the Department of Geography at Kent State University developed a project entitled Design And Development Of An Interactive "Geiger Counter" For Exploratory Analysis Of Spatial Data Quality. Their system implemented in C, provided tools for manipulating spatial data displayed in different windows and allowed for interactive querying of the data quality component of a database. This project used the SCS soils database.

Dr. Matthew Ward and Jun Zheng from the Computer Science Department at Worcester Polytechnic Institute designed a prototype visualization system to display data quality in the US EPA Chesapeake Bay data. They framed the problem in terms of five dimensions (x, y, z, time, data quality), where each dimension could be specified as a point, a set of points (sampled), or a continuous range of values.

Brand L. Niemann from the Center for Environmental Statistics at U.S. EPA used MapInfo, Voyager, and Movie software to display the EPA Chesapeake Bay data set. Combining this software allowed users to start, for example, with a map of monitoring locations, select a parameter and a site, click on that site to display the time series of that parameter in another window, and then click on a point in that time series to display the textual metadata about that data in a third window.

Three student projects developed for the challenge included an interactive visualization technique that allowed users to examine uncertainty in soil map unit classes in more detail as scale increased (Miri 1993), an interactive technique to visualize soil boundary uncertainty using a raster model (Hassen 1993) and a grid technique to visualize uncertainty for a linear interpolation of dissolved inorganic nitrogen for Chesapeake Bay (Frank 1993).

Visiting Scholars to NCGIA-Maine

Carl Amrhein, University of Toronto, spent three months at Maine in 1993. He and Kate Beard examined the aggregation or modifiable areal unit problem using radon data for the State of Maine. Radon which has been a suspected contributor to lung cancer has shown no positive correlations with lung cancer rates when aggregated to demographic units for this comparison. Extensive radon measurements within schools allow for testing aggregation effects within schools, within zip codes, town and counties and for comparison of these aggregations against natural aggregations based on soil and geologic units. Beard and Amrhein along with Dan Griffith and others submitted a proposal to Centers for Disease Control to fund this research ($274,630). The proposal was ranked meritorious but not funded. Amrhein will return to Maine for the one month in 1994 to continue this work. Howard Veregin, Kent State University, visited Maine in June 1994 to pursue I7 research on modeling soil uncertainty and visualization techniques for displaying this uncertainty.

Visiting Scholars to NCGIA-Buffalo

Geoff Dutton spent two periods in residence at Buffalo during the early months of 1993, working on research refining the QTM Hierarchical Tesselation system he has developed. QTM, or Quaternary Triangular Mesh, subdivides the globe recursively into triangles that locate position with accuracies on the order of centimeters or better resolution. Geoff's previous work has argued that QTM can be substituted for vector coordinates and that some GIS operations become computationally simpler in QTM space. His specific project at Buffalo was to determine variations in positional accuracy that may accrue when converting latitude-longitude pairs into and out of QTM space.

Visiting Scholars to NCGIA-Santa Barbara

Yee Leung of the Chinese University of Hong Kong spent two months at Santa Barbara in 1992 working on I7 research. In 1993 Gary Hunter of the University of Melbourne spent six months at Santa Barbara working with Michael Goodchild and others. In July, 1992 a small workshop was organized at Santa Barbara with Sucharita Gopal (Boston University) and Ferko Csillag (Syracuse University) to pursue I7 topics.

External Funding

During the active period of I7, several NCGIA researchers have been involved in applied projects, using results of I7 and prior initiatives to address practical problems concerned with uncertainty in spatial databases. At Santa Barbara, the California Department of Forestry has funded an Accuracy Assessment Task Force to design improved methods of accuracy assessment for forest land mapping in the state. Members of the committee included Nick Chrisman(University of Washington), Russ Conglaton, (University of New Hampshire), and Greg Biging (University of California-Berkeley). Goodchild was a member of an accuracy assessment workshop funded under the NOAA C-CAP program chaired by Siamak Khorram of North Carolina State University.

ASSESSMENT

This section presents an assessment of the initiative according to the five criteria established by the NCGIA Board of Directors.

How Has the Research Agenda Been Affected?

Prior to the Specialist Meeting, four themes were identified to focus discussion:

- Data Quality Components (error modeling and indices of data quality),
- Data Models and Database Issues (management of data quality within databases),
- Representational Issues (visual tools), and
- Evaluation of User Needs (assessment of the tools and of user needs).

At the meeting, these were revised and the research agenda was modified to include:

- conceptual frameworks for matching quality components with visualization tools;
- modeling and monitoring error propagation during GIS operations;
- investigations of user needs and expectations about data quality;

- the role of data quality visualization in decision-making and dispute resolution;
- development of educational tools and tutorials on spatial data quality;
- database support for quality information; and
- evaluation of visualization techniques.

The initiative made progress in each of these areas, with substantially more progress in some areas than others. It is interesting to note that there was essentially no research agenda on visualizing data quality prior to the initiative. Since publication of the NCGIA research agenda in 1989 and more recently with the active phase of this initiative, substantial research in this area has appeared both nationally and internationally as evidenced by journal articles, papers in conference proceedings, and graduate student thesis. The most significant effect on our internal agenda, has been a circular one. The in-depth research in visualizing spatial information quality has served to illustrate many of the inadequacies in our present error models, error measures and data quality management efforts. The data quality information to support visualization is simply not there in most of our current spatial data products. Quality measures which do exist, take the form of mapping standards or summary measures (means, variance, RMSE, PCC, see Veregin 1989) which do not yield substantial information for useful visualization of data quality within a GIS. The initiative has clearly demonstrated that we can visualize these measures and at least provide GIS users better access to these measures of data quality. The major limitation is that disaggregate or localized behavior of error or uncertainty is not captured by these models or measures. The error model developed under NCGIA Initiative 1 for categorical data is one of the few models which provides sufficient data to visualize the spatial structure of uncertainty at a refined (sub-polygon) level. Thus one of the outcomes of this initiatives in terms of influencing the GIS research agenda is to reinforce the need for more research on data models and error models that can provide more spatially refined information as input for visualization techniques.

What Do We Know Now That is New?

Research in every topic has been pursued and published, and much of the work remains in progress as the initiative final report is drafted. The advances in the state of knowledge fall into four categories, including extension of theory, information management, and development of new graphical tools.

We have learned much from incorporation of current communication theory (Searle), behavioral science theory (Coseir and Dalton), and principles for human-computer interaction (Winograd and Flores) into data quality representation. Presentation of data quality information in a GIS environment requires a dialog that may differ from other database queries in structure and in content; the work of Wu and Buttenfield addresses the first aspect and that of Couclelis addresses the second. Access to data quality also requires flexible as opposed to rigid structures to serve a broad range of user needs and experience levels. Work by Paradis on the Data Quality Filter addresses this area.

We have learned that spatial theories which explicitly capture and model uncertainty lend themselves more effectively to visualization of data quality. Bruegger's PhD dissertation work clearly illustrates this. Current models (many still map based models) have severe limitations for visualization.

We know more about techniques for communicating an understanding of error models to users of spatial data. We know that maps of mean estimates and estimation variance, common in kriging, create a flawed understanding and in general that presentation of such error descriptors is inappropriate since maps of error descriptions are not possible realizations of the error model. Instead presentation of a sample of realizations, by animation or simultaneous display may create the only sound understanding of uncertainty and its implications.

We have learned about the advantages of object-oriented and feature-oriented approaches to data quality representation, but need to learn more about the computational and data volume overhead that their inclusion in a database may generate. Layer-based GIS functions must continue to proliferate in part because error modeling remains based upon layer data models, and in part because the layer model is most efficient for raster data, such as digital terrain and remotely sensed imagery. However, it is clear that data quality information is rarely homogeneous across a single map layer, and further that variations in quality that accrue with specific GIS operations (buffering, overlay, coordinate conversion, etc.) may be unique depending on the data theme, resolution, and time frame.

The work on visualization has brought to the fore the need to know the behavior of error generating processes more explicitly. For example to capture the behavior of a GIS operation visually and communicate it unambiguously to a user, its behavior must be thoroughly known. Work by Hassen on the reference grids as a process for documenting postional change generated by GIS operations has made progress in this area.

We have learned that the existing graphical framework (Bertin's visual variables) may not be sufficient to manipulate aspects of data quality information. Several new visual variables have been explored, including defocusing of features and symbols ("fog" is the name given this by Alan MacEachren), development of multivariate symbols (John Hancock and Mark Monmonier), and existing visual variables have also been tested empirically, including various dimensions of color (hue, saturation and value).

Additionally, we have learned that application of emerging technologies for visualization (animation, multimedia, and hypermedia) enrich the fabric of visual display to allow data quality information to be embedded directly in a GIS data display. Animation has received the most attention so far, and this may be happening because animation technology is farther advanced than either multimedia or hypermedia at present. These novel forms of visualization can also improve user involvement in learning about data quality. Online tutorials developed at Maine and at Buffalo have demonstrated this is the case.

A knowledge which is not new but sometimes forgotten particularly in the thrill and potential of new technologies is that simplest is best. We now know that simple visualizations are best.

What recommendations would the NCGIA make in this area?

Several recommendations could be put forward as a result of work on this initiative. The initiative clearly points to inadequacies in current data sets for support of quality information. Despite the on-going standards efforts little information is still available in current data sets to thoroughly document quality for a wide range of users needs. Until such quality information is available, vendors will be reluctant to invest in even some of the simpler prototype visualization techniques developed under this initiative. Thus substantial support and encouragement is recommended to institutionalize collection and maintenance of data quality information.

Until better error measures and models are more generally available, vendors should make an effort to support some level of access to quality information using existing measures or models. Some level of awareness of quality is better than none. Although we suggest that some current error measures may be misleading such as the error variances generated by kriging, even displays of these measures can warn users of some level of uncertainty. If displays using current measures are properly constructed to communicate their limitations as expressions of uncertainty (as demonstrated by work under this initiative) they improve on no communication of uncertainty.

The current proposed metadata standard calls for more extensive quality information than SDTS. Such requirements stand to pose a substantial burden on data producers and higher data costs for users. To counteract this side effect and negative repercussions we recommend research to support for more automated capture of quality information.This implies new data collection methods in which information which was previously only available to field scientists is captured and maintained for GIS users. Electronic filed data collectors now make this effort more realistic. This recommendation also implies new data model or database support for storing and managing data quality information. Measurement based systems in which original measurements are stored or an entirely new spatial theory such as that proposed by Bruegger offer solutions for making quality information more directly available.

Maps will not disappear in the near future as a source of information for GIS so the necessity to encode map information accurately and automatically will still exist. Quality of a database derived from maps is largely a measure of accurate replication and internal consistency. Methods to verify the metric and topological information stored on maps are well developed. Methods to encode and verify attribute data are less automated and less robust. Research in this area is recommended to provide more complete quality coverage for map-based databases.

How has the Education of GIS Scientists Been Enhanced?
We have learned that adoption of practices for data quality management and presentation have been slow in part because many GIS users do not understand concepts or components of data quality. It is common to hear GIS experts speak of data quality as being composed only of positional accuracy, without regard to the multifaceted characteristics of attribute definition, logical consistency, or temporal currency. Research teams at Maine and at Buffalo have built and disseminated online tutorials defining data quality components and exemplifying their display.

Several prototype visualization techniques have been developed that simplify access to data quality information. These techniques are designed to make quality issues more explicit and thus less easy to ignore, and hopefully more easy to understand by lay users. Many of these prototypes will hopefully serve as inspiration to vendors such that they are readily implemented and made available to GIS users.

Federal adoption has moved ahead more quickly perhaps due to the establishment of mandated standards for data use and exchange. Continuing efforts to convince the nonfederal community must include identification of cost savings, user benefits (such as demonstration of more efficient database queries), and identification of incentives for adoption. One incentive may be found now in the FGDC Spatial Data Clearinghouse, part of the National Spatial Data Infrastructure (NSDI). Clearinghouse data will incorporate data quality information as a matter of course, and as GIS users discover this can facilitate data use and exchange, they may come to rely upon its inclusion in the data they generate.

What has NCGIA Learned about the Initiative Process?
At a general level, this initiative has made it abundantly clear that an initiative agenda does not advance in isolation. Work on this initiative is particularly indebted to work from Initiative 1 and 13 and in turn it reinforces need for new work in both areas.

We confirmed that small group / full group interaction on specific topics provides a manageable if exhausting venue for specialist meetings. We confirmed that breaking the Initiative Meeting with a half- to full-day of unstructured discussion (in the case of Castine, this took the form of a sailing excursion and picnic on an island) offers time for meeting participants to collect their thoughts, discuss the research agenda, and plan for possible collaboration on the research agenda. We also learned that it is nearly impossible for a single individual (in this case, Barbara Buttenfield) to lead or co-lead two initiatives simultaneously, even if the motivation and good intentions are in place at the outset. The administrative overhead and the dual scientific focus tend to distract attention from one initiative or the other. Scheduling of future initiatives should take this into account.

SUMMARY

The quality of spatial data and databases is a major concern for developers and users of Geographic Information Systems (GIS). All information which is subject to display in GIS can be characterized in terms of its quality, which may vary with location, theme, and time. The quality of spatial information products is multidimensional, and relates to accuracy, error, consistency and reliability. Implications for spatial analysis and for spatial decision-making can be identified in theoretical work (for example in spatial statistics) as well as in GIS applications (for example in resource management). Our goal in NCGIA Research Initiative 7 has been to focus on effective means of acquiring, managing and visually communicating components of data quality to researchers, decision-makers and users of spatial information, particularly in the context of GIS.

We identified four themes for the Specialist Meeting. These four were reorganized and refined during the meeting, to incorporate conceptual frameworks for matching quality components with visualization tools; modeling and monitoring error propagation during GIS operations; investigations of user needs and expectations about data quality; the role of data quality visualization in decision-making and dispute resolution; development of educational tools and tutorials on spatial data quality; database support for quality information; and evaluation of visualization techniques.

Literature Cited

Coseir, R.A. and Dalton, D.R. 1988. Presenting information under conditions of uncertainty and availability: Some recommendations. **Behavioural Science**, 33, 272 - 281.

Searle, J.R. 1969. **Speech Acts Cambridge**. UK: Cambridge University Press.

Searle, J.R. 1979. **Expression and Meaning**. Cambridge UK: Cambridge University Press.

Winograd, T. and Flores, F. 1987. **Understanding computer and cognition**. Reading MA: Addison- Wesley.

NCGIA PUBLICATIONS

Refereed Journal Articles

1994 Goodchild, M. F. Integrating GIS and Remote Sensing for Vegetation Analysis and Modeling: Methodologies Issues. **Journal of Vegetation Science** (in press).

1994 Hunter G. and Goodchild M. F. Dealing with Error in Spatial Databases: a Simple Case Study. **Photogrammetric Engineering and Remote Sensing** (in press).

1994 Paradis, J. and Beard, M. K. Visualization of Spatial Data Quality for the Decision Maker: A Data Quality Filter. **URISA Journal** (in press).

1994 Wu, C.V. and Buttenfield, B.P. Evaluation of Spatial Data Quality. **Computers, Environment and Urban Systems** (in press).

1994 Kottman, C.A., and Buttenfield, B.P. Standards for Spatial Data Use: Similes Improve Our Understanding. **Cartography and GIS** (in press).

1993 Beard, M.K. and Mackaness, W.A. Visual Access to Data Quality in Geographic Information Systems. **Cartographica** vol.30(2/3): 37-46.

1993 Buttenfield, B.P. Representing Data Quality. **Cartographica** vol.30(2/3): 1-7.

1993 Fisher, P.F. Visualizing Uncertainty in Soil Maps by Animation. **Cartographica.** vol.30(2/3): 20-27.

1993 Fisher, P.F., Dykes, J. and Wood, J. Map Design and Visualization. **The Cartographic Journal**, vol.30 (Dec 93): 136-142.

1993 Hunter, G and Goodchild, M. Managing Uncertainty in Spatial Databases: Putting Theory into Practice. **URISA Journal**. Vol. 5 (2): 55-63.

1993 Lanter, D.P. A Lineage Meta-Database Approach Towards Spatial Analytic Database Optimization. **Cartography and GIS**. vol. 20(2): 112-121.

1993 McGranaghan, M. A Cartographic View of Spatial Data Quality. **Cartographica.** vol.30(2/3): 8-19.

1993 Monmonier, M. Exploring the Quality of Enumeration-Area Data With Graphic Scripts. **Cartographica.** vol.30(2/3): 28-36.

1993 Sorensen, P. and D. Lanter. Two Algorithms for Determining Partial Visibility and Reducing Data Structure Induced Error in Viewshed Analysis. **Photogrammetric Engineering and Remote Sesnsing.** Vol 59. (7): 1149-1160.

1992 MacEachren, A.M. Visualizing Uncertain Information. **Cartographic Perspectives.** no. 13: 10-19.

1991 Lanter, D.P. Design of a Lineage Meta-Database for GIS. **Cartography and GIS**. vol. 18(4): 255-261.

Journal Issues

1994 Buttenfield, B.P. (ed.) “Mapping Data Quality”. Special Issue of **Cartographica** vol.30(2/3).

1994 Wortman, K.C. and Buttenfield, B.P. (eds.) "Current Developments and Use of the Spatial Data Transfer Standard". Special Issue of **Cartography and GIS** (in press).

Book Chapters

1993 Buttenfield, B.P. and Beard, M.K., Graphic and Geographic Components of Data Quality. In Hearnshaw, H. and Unwin, D. (Eds.) **Visualization in Geographic Information Systems**. London: Belhaven Press 150-157.

1993 Goodchild, M. F., L. Chih-Chang, an d YLeung. Visulizing Fzzy Maps. In Hearnshaw, H. and Unwin, D. (Eds.) **Visualization in Geographic Information Systems**. London: Belhaven Press. 158-167.

1993 Goodchild, M.F., Buttenfield, B.P., and Wood, J. Issues of Data Validity. In Hearnshaw, H. and Unwin, D. (Eds.) **Visualization in Geographic Information Systems**. London: Belhaven Press. 141-149.

1992 MacEachren, A. E., Buttenfield, B. P., Campbell, J. C., DiBiase, D. and Monmonier, M. S. Visualization. In: Abler, R. A., Olson, J. M. and Marcus, N. G. **Geography's Inner World**. New Brunswick, NJ: Rutgers University Press: 99-137.

1991 Buttenfield, B.P. and Mackaness, W.A. Visualisation. In **GIS: Principles and Applications**. (Eds. D. MacGuire, M.F. Goodchild, and D. Rhind) London: Longman Publishers Ltd., Volume 1, Chapter 28: 427-443.

also Lanter, D.P. Comparison of Spatial Analytic Applications of GIS

Monographs and Technical Reports

1994 MacEacheren, A. M. Some truth with Maps: a Primer on Symbolization and Design. **Resource Publications in Geography.** Washington DC: American Association of Geographers.

1992 Couclelis, H., M. K. Beard and W. A. Mackaness. Two perspectives on Data Quality. **NCGIA Technical Report 92-12.**

1991 Beard, M.K., Buttenfield, B.P. and Clapham, S.B. Scientific Report on the Specialist Meeting on Visualizing the Quality of Spatial Information. **NCGIA Technical Report 91-26.** 170 pp.

1990 Lanter, D.P. Lineage in GIS; the Problem and a Solution. **NCGIA Technical Report 90-6.**

Conference Proceedings

1994 Beard, M. K. Accommodating Uncertainty in Query Response. **Proceedings** 6th International Symposium on Spatial Data Handling.

1994 Hancock, J. Defining Multivariate Regions: A Flexible Approach Using Interactive Statistical Visualization, **Proceedings** AAG. San Fancisco.

1994 Hunter, G. and M. Goodchild,

1993 Beard, M. K. and Mackaness, W. Visualization of Interpolation Accuracy, **Proceedings** Auto Carto 11. 228-238.

1993 Dutton, G. and Buttenfield, B.P. Scale Change Via Hierarchical Coarsening: Cartographic Properties of Quaternary Triangular Meshes. **Proceedings**, 16th Conference of the International Cartographic Association, 3-8 May 1993, Cologne, Germany, vol. 2: 847-862.

1993 Frank, S. Using Visualization Techniques to Examine the Effects of Random Control Points on Grid Interpolation. **Proceedings** GIS/LIS'93. Minneapolis, MN. Vol. 1 226-232.

1993 Hassen, K. M. Visualization of Soil Boundaries in a Cell Based Model. **Proceedings** GIS/LIS'93. Minneapolis, MN. Vol. 1 292-301.

1993 Hunter, G. and Goodchild, M.F. Managing Uncertainty in Spatial Databases: Putting Theory into Practice. **Proceedings** URISA. Vol.1. 1-14.

1993 Miri, B. Visualization of Soil Map Inclusions with Change in Scale. **Proceedings** GIS/LIS'93. Minneapolis, MN. Vol. 2. 540-549.

1993 Stephan, E.M., Bucher, F. and Vckovski, A. Virtual Data Set -- An Approach for the Integration of Incompatible Data. **Proceedings**, AUTO-CARTO 11, Minneapolis, MN, October, 1993: 93-102.

1992 Beard, M.K. and Buttenfield B.P. Spatial, Statistical, and Graphical Dimensions of Data Quality. **Proceedings**, 24th Symposium on the Interface-Graphics and Visualization Conference, College Station Texas, March 19-21, 1992 408-415.

1992 Clapham, S. B. A Formal Approach to the Visualization of Spatial Data Quality. **Proceedings** GIS/LIS'92 San Jose CA. Vol.1. 138-149.

1992 Leung, Y. Goodchild, M.F. and Lin, C-C., Visualization of Fuzzy Scenes and Probability Fields. **Proceedings** of the 5th International Symposium on Spatial Data Handling.

1992 Schweizer, D.M. and Goodchild, M.F. "Data Quality and Choropleth Maps: An Experiment with the Use of Color. **Proceedings** GIS/LIS'92. Vol.2. 686-99.

1991 Buttenfield, B.P. and Beard, M.K .Visualizing the Quality of Spatial Information. **Proceedings**, AUTO-CARTO 10, Baltimore, Maryland, March, 1991 vol. 6: 423-427.

1991 Clapham. S. and Beard M. K. The Development of an Initial Framework for Visualization of Spatial Data Quality. **Proceedings** Annual Meeting American Congress on Surveying and Mapping, Baltimore, Maryland, March, 1991 vol.2: 73-82.

1990 Buttenfield, B.P. and Ganter, J.H. Visualization and GIS: What Should We See? What Might We Miss? **Proceedings**, 4th International

Symposium on Spatial Data Handling, Zurich Switzerland, July 1990, vol. 1: 307-316.

Other Publications

1994 Wortman, K.C. and Buttenfield, B.P. (Guest Editors). Preface to Special Issue on SDTS. **Cartgoraphy and GIS** (in press).

1994 Buttenfield, B. P. Mapping Data Quality. **Cartographica**, Preface to Special Issue on Mapping Data Quality (B.P. Buttenfield, Guest Editor), vol.30(2/3): vii-viii.

1992 Buttenfield, B. P. Quality and Availability of Spatial Information: Cartography and The Year in Review. **ACSM Bulletin**, No. 135 January 1992, p. 30-33.

1992 Clapham, S.B. A Formal Approach to the Visualization of Spatial Data Quality; Visual Variables and the Spatial Data Transfer Standard. IEEE Visualization 92. Position Paper. Boston MA.

1992 Griffith, D.A. and Csillag, F. EXPLORHO: Simulation Package for Exploration of Spatial Autocorrelation. Copyright 1992 by Griffith and Csillag. Department of Geography, Syracuse University.

Graduate Thesis Research

Bruegger, B. 1994. A Spatial Theory for the Integration of Resolution Limited Space. Unpublished PhD thesis, Department of Surveying Engineering, University of Maine.

Carey, D. 1992. Use of Federal Digital Data Products with ARC/INFO. Unpublished MA thesis, Department of Geography, SUNY-Buffalo. (May 1992)

Clapham, S. A. 1992b. A Formal Approach to the Visualization of Spatial Data Quality. Unpublished MA thesis, Department of Surveying Engineering, University of Maine.

Dolby, L. A. 1994. Mapping Subsurface Geology in Eastern Casco Bay Using Seismic Reflection Profiles and GIS. Unpublished MA thesis, Department of Surveying Engineering, University of Maine.

Fly, P. 1993. Hypermedia and Ethics in Cartography. Unpublished MA thesis, Department of Geography, SUNY-Buffalo.

Kick, J. (ongoing) Refinement of Soil Modeling Parameters Using GIS. Unpublished MA thesis, Department of Geography, SUNY-Buffalo.

Kress, T. 1991. Visualization of Spatial Data Quality. Unpublished MA thesis, Department of Geography, SUNY-Buffalo.

McKinney, J. 1992 Evaluation of Visual Tools for Location Modeling. Unpublished MA thesis, Department of Geography, SUNY-Buffalo.

Schweizer, D. 1992. Visualizing Data Quality for Choropleth Maps. Unpublished MA thesis, Department of Geography, University of California-Santa Barbara.

Wu, C. Victor 1993 Object-Based Queries on Spatial Data Quality. Unpublished Ph.D. thesis, Department of Geography, SUNY-Buffalo.