VISUALIZATION OF IMAGE QUALITY IN DISTRIBUTED SPATIAL DATABASES

Isolde Schlaisich, Giorgos Mountrakis, Peggy Agouris

Department of Spatial Information Science and Engineering, University of Maine, Orono, ME 04469, USA
(isolde, giorgos, peggy)@spatial.maine.edu

KEY WORDS: Quality, Visualization, Imagery, Databases, Decision Support

ABSTRACT:

Users of GIS are offered an abundance of aerial and satellite imagery in distributed spatial databases. Available datasets can cover the same area, differing only in quality attributes, such as resolution and positional accuracy. The advantage of having a bigger pool of information at hand is counterbalanced by the fact that extracting data with the appropriate quality might be overwhelming. In addition, non-expert GIS users might not be aware of the problems that can arise while working with datasets of various qualities. Consequently, information about data quality should be provided to users along with the data. This paper presents a novel approach to communicating image data quality in a visual form, which allows users to grasp the quality of available datasets at a glance. The data attributes that are identified for holding relevant quality information are positional accuracy, scale, resolution, completeness, consistency, and currency. To communicate data quality, image quality information is visualized through 3D models and superimposed color. Delivering quality information alongside original data helps users choose the data best suited for their task.

KURZFASSUNG:

In räumlichen Datensammlungen stehen GIS Nutzern eine Fülle von Luft- und Satellitenbildern zur Verfügung. Für viele Regionen existieren mehrere Datensätze, die sich aber in Qualitätsmerkmalen wie Bildauflösung unterscheiden. Dem Vorteil, den GIS Benutzer durch diese grössere Auswahl an Daten haben, steht der Nachteil gegenüber, dass sie bei der Auswahl der für ihre Anwendung geeigneten Daten überfordert sein könnten. Vor allem Benutzer die keine GIS Experten sind, sind sich z.T. nicht der Probleme bewußt, die beim Bearbeiten von Datensätzen mit unterschiedlicher Qualität entstehen können. Daher sollten GIS Nutzern zusammen mit den Originaldaten auch Metadaten zur Verfügung stehen, die die Qualitätsbeurteilung erleichtern. Dieser Artikel stellt eine neue Methode zur Vermittlung der Qualität von Bilddaten in visueller Form vor. Die Visualisierung hat den Vorteil, dass der Benutzer die Qualität der Daten auf einen Blick erfassen kann. Die folgenden Datenattribute werden benutzt, um Qualitätsmaße zu vermitteln: Genauigkeit in der Position, Maßstab, Auflösung, Vollständigkeit, Konsistenz und Aktualität. Die Daten werden visualisiert mit Hilfe von 3D Modellen und dem Nutzen von Farben. Wenn Benutzer von räumlichen Bilddaten neben den Daten auch gleichzeitig Zugriff zu Qualitätsinformation haben, können sie leichter sinnvolle Daten für ihre Anwendung wählen.

1. INTRODUCTION

In this day and age users can choose appropriate image data for their GIS analysis from a substantial amount of aerial and satellite imagery. These image collections may be stored in distributed spatial databases. In these diverse collections many datasets cover the same area but have different quality attributes, offering users a large assortment of information to choose from, thus attracting a wide range of potential applications. As applicability of GIS increases, professionals who are not trained in dealing with geospatial image data, such as biologists and foresters, discover a useful tool for their tasks. To help users of varying degrees of expertise and heterogeneous backgrounds choose useful images from distributed geospatial libraries, information about data quality should be provided along with the data.

In this paper we present a novel approach to communicating image data quality in a visual form, which allows users to comprehend the quality of available datasets at a glance. As response to a database query users will get a side-by-side display of actual data and visualization of the corresponding data quality. Our environment generates visualizations dynamically, with parameters and methods changing according to the query. Visualization, as opposed to tables of quality specifications, is chosen for the following three reasons:

- The visual channel is the primary sensory input channel for most people.
- Images can convey a lot of information in condensed space (Beard and Buttenfield, 1999).
- People who work with GIS are skilled at absorbing information from images.

The value that data quality visualization adds to a query system of distributed spatial databases lies in the combination of spatial and attribute values and their displaying in map-space, which makes it easier for users to perceive the data's usefulness and applicability.

As a first step in our approach we had to select which attributes are most useful for conveying image data quality. It is rather important to establish balance between showing attributes that help the user choose the appropriate data, and at the same time not displaying too many attributes that would increase the cognitive workload of the user and even cause confusion. The quality attributes that we chose to convey the overall image data quality are: positional accuracy, scale, resolution, completeness, consistency, and currency. The paper presents arguments for selecting these specific quality attributes to describe image data quality. To convey the selected quality attributes we developed visualizations using representations in 3D and color. 3D representations appear like a DEM, but instead of terrain height the corresponding data quality attribute is displayed.

The remainder of this paper is structured as follows: section two explores previous work on spatial data quality, the theory behind visualization methods and their combination, the use of visualization to convey data quality, and it describes existing projects on data quality visualizations. Section three discusses the selection of the quality attributes we chose in our approach. In section four we discuss effective visualization methods and introduce the visualizations that we have developed. Section six provides conclusions and future work.

2. RELATED WORK

In the last two decades data quality has become an important research topic. Scientists argued that users of spatial data should have access to data quality information (McGranaghan, 1993; Buttenfield and Beard, 1994; Beard, 1997). Soon it became obvious that the nature of spatial data lends itself perfectly to the communication of quality parameters by visualization in the form of images and graphics. As a result, the call for visualization of data quality surfaced (Beard and Mackaness, 1993; van der Wel et al., 1994).

Since the early nineties researchers took formal approaches to the visualization of spatial data quality (Clapham, 1992). The National Center for Geographic Information and Analysis devoted a lot of energy in exploring this area and spearheaded a research initiative on "Visualization of the Quality of Spatial Information" (Beard et al., 1991). Results from this initiative are introduced in (Buttenfield and Beard, 1991).

In the remaining part of our literature review we present various terms used to describe data quality aspects, and we discuss related visualization approaches and past project efforts.

2.1 Discussion of Terminology

In the literature a substantial number of expressions are used to describe data quality, namely quality, error, reliability, uncertainty, validity, accuracy, vagueness precision and fitness for use.

The term *quality* is used as an umbrella-term that covers all aspects of the issue. It is used by practically everybody in the field (Beard, 1997; Veregin, 1999). The use of the term *error* is also widely used, and there is broad consent on what the word describes when used for image data, namely the difference between true value and the value stored in the database (Hunter and Goodchild, 1995; Buttenfield, 1993). *Reliability* can be defined as the level of confidence a data provider has that the data are correct (Evans, 1997).

The term *uncertainty* is used in various ways, one being that the resolution of the data does not allow a user to make an assured decision about the content of the data. For example, pixels in remotely sensed images might contain uncertain information because of sub-pixel mixing or sensor sampling bias (Bastin et al., 2002). Worboys and Duckham (2004) use the term uncertainty to describe the doubt that users have about the right use of data. In this sense it is a measure that describes the state of mind of the user.

Other terms that are used to describe different outlooks on data quality are *validity* (Goodchild et al., 1994), and *accuracy* (Veregin, 1999). *Vagueness* describes the impossibility to determine the exact location or boundary of an object in space (Duckham et al., 2001). For example 'the East of Maine' is a vague area in that its boundaries are not exactly determinable. *Precision* denotes the exactness with which the measurement is made that led to the entry in the database (MacEachren, 1992). An overall phrase that is used frequently is *fitness for use*. It indicates whether the data has the specifications that the users need to solve their task (Paradis and Beard, 1994).

2.2 Visualizations

Beard and Buttenfield (1999) listed the following challenges in the visualization of data: graphic design, metadata, error analysis, and user satisfaction. In this research we concentrate on the graphic design issues. For the combined display of data and data quality three possible forms are mentioned in the literature (MacEachren, 1994; Beard and Buttenfield, 1999). First, there are side-by-side images, where one picture shows the data and the other one the quality of the data. The second approach is to generate composite images that display data quality superimposed on the visualization of the data. Thirdly, sequenced images of data and data quality can be presented, either affording the user to toggle between the displays or providing an animation (Evans, 1997).

The following two visualization approaches have also been discussed: variation in color hue and saturation to convey the quality of data (Schweizer and Goodchild, 1992; Howard and MacEachren, 1996), and, showing quality attributes as the z-axis in a 3D elevation model, which was mentioned as a worthwhile endeavour by van der Wel et al. (1994) without any follow-up projects implementing the idea. We take up the concept of the latter approach and incorporate it in our 3D visualizations.

2.3 Previous Projects

The following works concentrate on the communication of quality of geospatial data. The R-VIS project introduces a model which shows the reliability of water quality data (Howard and MacEachren, 1996). A visualization of uncertainty in meteorological forecast models was also developed showing the discrepancy of multiple weather forecasts over time (Fauerbach et al., 1996). Various graphs, bivariate images and animations are used in the FLIERS project to visualize uncertainty in multi-spectral remotely sensed imagery (Bastin et al., 2002). Davis and Keller (1997) offer quality information for risk management decisions. Spatial data uncertainty was also communicated using animation (Ehlschlaeger et al., 1997).

3. DATA QUALITY ATTRIBUTES

Metadata contain a wealth of information about the data at hand. From the attributes that are typically described by metadata information we selected the ones that convey data quality, and more specifically, those which pertain to geospatial image quality. Our goal has been to display the optimum number of essential data attributes, avoiding redundancies which could confuse the user. We based our selection on the US Spatial Data Transfer Standard's (SDTS) section on data quality (NIST, 1992), which is quite consistent with the data quality parameters that are used by the GIS community (Veregin, 1999).

Accordingly, to visualize the quality information of geospatial image data we selected the following data attributes: positional accuracy, logical consistency, completeness, scale, resolution, and currency (also called up-to-dateness).

One attribute of the US STDS that we are not using in our approach, and that is also mentioned as an important quality attribute for spatial data in other projects, is lineage. We agree that it is a significant attribute for describing spatial data in a GIS, but image data are typically not modified after their collection as other spatial data usually are. Apart from that it is also a non-quantifiable measure and in our approach we are focusing on quantifiable attributes.

Positional Accuracy: In the generally accepted definition, accuracy of spatial data describes how much the data deviates from a certain model or from reality. A point in an image does not necessarily lie exactly in the place where it is shown, but depending on its positional accuracy it can lie anywhere within an area around its displayed position. The most common way to describe positional accuracy is to define an error ellipse within which the point lies with a certain probability (Figure 1).

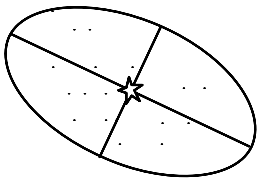

Figure 1: Positional Accuracy: Point might lie anywhere within the ellipse.

Resolution: For geospatial image data resolution is an important quality measure. Digital images are stored in the form of raster data of a certain resolution, e.g. a single pixel can cover a ground area of 1x1km, 30x30m, or 0.6x0.6m. The attribute of resolution is hardware-dependent. It relates to sensor specifications in the case of digital images, and the granularity and settings of the scanner in the case of digitized photographs. The difference in information detail and quality that images of various resolutions offer is clear. Different resolutions are useful for different tasks, but users have to be aware of the impact that resolution has on the usability of the data. As a simple example, figure 4 shows two satellite images of different resolutions covering a contiguous area.

Scale: The scale of an image describes the correspondence between image space and the captured ground area. This geometric relationship is dependent on the selected flight mission parameters such as flying height and camera focal length. Therefore, it is a measure that is affected by image formation geometry, as opposed to the hardware-dependent resolution. Image scale is one aspect expressing the image's fitness of use for a desired application.

Completeness: Completeness communicates whether data is present. Data might be missing completely for some areas, for example in cases where there have been no flight missions to produce aerial photographs. Also, parts of a picture can be missing due to cloud cover.

Currency: In the NCGIA specialist meeting on the visualization of spatial data quality, the currency of data has been defined as the time between the production of images and the date of a query (Beard et al., 1991). One could also call it image timestamp. Information on currency can be a decision factor for users. Several applications ask for the most recent available data, while other users may be looking for older data or specific time series for change detection. Two images of different currency can be seen in figure 2. A comparison of the images illustrates change that has taken place in the University of Maine between 1939 and 1972.

Figure 2: Part of the University of Maine in 1939 and 1972.

Logical Consistency: Logical consistency checks if data are contradicting each other. The visualization of the results of these checks can be a powerful warning sign for users. For example, a satellite image with 5m resolution and a currency of 1972 should raise an inconsistency flag.

4. VISUALIZATION METHODS

Of the display methods for the combination of data and data quality that were introduced in section 2.2, in this research we elected to use the side-by-side display. Overlaying data and quality information in one composite image would put too much information into one image and overload the user's sensory input. General visualization methodologies that interchange between the data themselves and their quality information in the form of blinking or user toggling do not seem optimal for our purpose, as they may be distracting for the user. Therefore we opted for the side-by-side displays. Users navigating through a collection of geospatial images will get the actual data and at the same time a graph showing the corresponding quality information using our visualization methods.

Due to space limitations, in this paper we show only the quality visualizations and not the data beside them. The techniques that we use are 3D visualizations and color properties such as hue and saturation.

4.1 Single Quality Attribute

In the following section we present two visualization techniques that communicate available quality information on a single attribute.

4.1.1 Single Attribute Using 3D Surface: Figure 3 depicts a visualization of a single quality attribute represented on the z-axis of the 3D image. The x- and y-axes show the spatial extent of the area of interest, and the z axis corresponds to the resolution (in meters) of the images in the database collection. One can see that available images vary in resolution from 2m to a resolution of 30m. In this example, if the data collection contains multiple images covering the same area, the one with the highest resolution is chosen for visualization in the 3D model. A method to convey multiple values in resolution from overlapping images is shown in section 4.1.2.

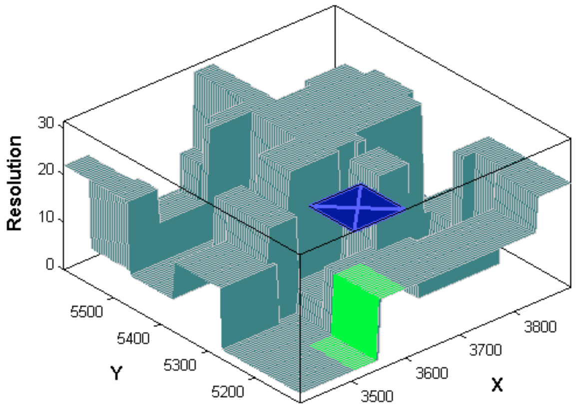

Figure 3: 3D representation of image resolution.

As an applicability example of the above visualization, imagine a forester conducting a tree vitality evaluation. Our visualization in figure 3 gives her an at-a-glance overview of available resolutions within the spatial extent of her interest.

For better navigation in the 3D model, the user can zoom into an area of interest to further explore the available resolution. This is done by a sliding window, which is marked by an X in figure 3. The zoom-in version overlapping the selected window area is shown in greater detail in figure 4. This figure shows two images, with the left image having a resolution of 2m and the right one a resolution of 20m. To enhance the selection process the original images can be superimposed on the 3D illustration. The user can also rotate the 3D surface to investigate hidden areas.

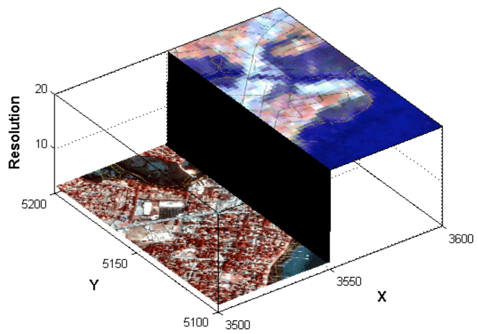

Figure 4: Zoom-in of selected area from figure 3 with superimposed images.

4.1.2 Multiple Instances of a Single Attribute: An extensive data collection is likely to contain multiple images of varying quality covering the same area. In the following visualization (figure 5) the data with the highest quality data value is chosen for the z-axis. Color is used to convey whether there are additional datasets of lower resolution available for the same area. The lighter (yellow) color shows that there are no additional datasets, while the darker (red) color shows that there are additional data (i.e. high multiplicity).

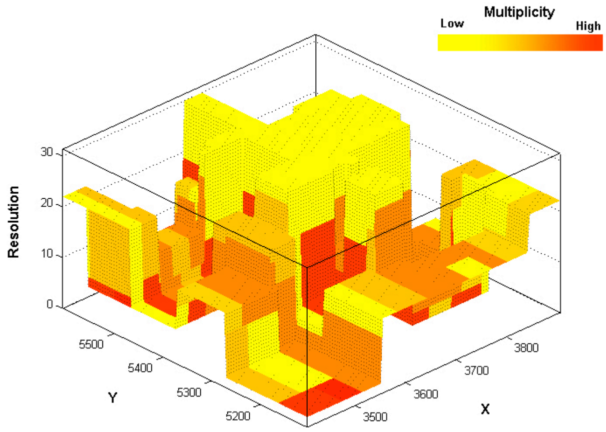

Figure 5: Representation of additional available imagery in lower resolutions (multiplicity).

Applying the above representation to our previous scenario gives our forester access to additional information that she would not be able to get solely from the 3D model. The user would like to know what other datasets exist to facilitate more detailed processes. This would be especially useful when multiple images of different resolutions covering the same area would be required. For example she might be interested in a low resolution satellite image to get an overview of forest density areas, and based on that, subsequently use a high resolution aerial photograph to extract the vitality of single trees.

4.2 Combination of Attributes

The next step is to combine information on multiple quality attributes in one image. For this we use 3D representations and superimposed color. Unlike the previous example that used color and 3D to communicate different aspects of the same attribute, below we discuss how to combine different attributes in a single visualization.

4.2.1 Combination of Two Attributes: In order to effectively combine the visualization of two attributes, the first attribute is depicted along the z-axis of the 3D image. The second attribute is conveyed by the color that is overlaid. Figure 6 shows the combination of the attributes of resolution and currency. The images of the highest available resolution are displayed on the z-axis (as before), while the currency is communicated with the help of variations in color hue. Lighter shades of green represent older data, while the more the color changes towards blue (darker) the more recent the data is. When multiple images of the same resolution but taken at different times are present, we communicate the most recent available image.

This visualization supports users that are interested in the currency of the data in addition to resolution. A city planner, for example, whose task is to identify a site for a new housing development wants to use geospatial images for a first overview of the area. One of the important parameters he considers in his evaluation is road access to the site. In order to effectively do so, he can use the above visualization to extract the most current data with sufficient enough resolution to identify roads.

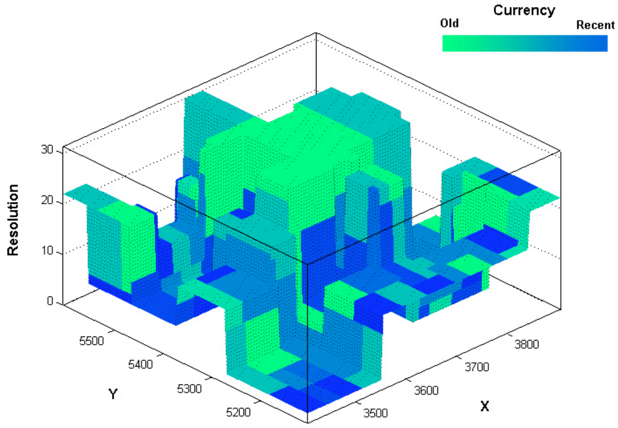

Figure 6: Combination of quality attributes resolution (3D) and currency (color).

4.2.2 Combination of Three Attributes:

Color saturation can be used as a third graphical component to display another quality attribute of the data. The more saturated the color is, the better the data quality that it represents is.

Figure 7 shows an example of how saturation can be added to express an additional attribute. Building on the example of figure 6 we can create the multi-variable color legend of figure 7. Variations from blue to green show timestamp information (x-axis) and changes in saturation depict available positional accuracies (y-axis).

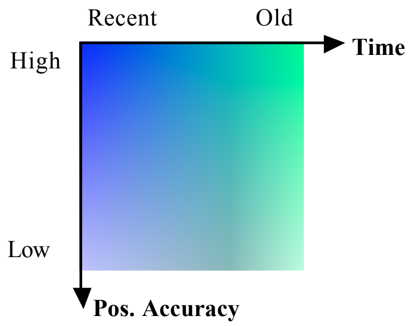

Figure 7: Interaction of color hue and saturation.

Figure 8 provides an example of using the legend of figure 7 to add positional accuracy information to the graph of figure 6. Note how the appearance of the graph changed, compared to figure 6. Figure 8 uses the same dataset for resolution and currency as figure 6, but also takes the positional accuracy of the data into account.

An application of this visual combination method is the selection of appropriate information for cadastral updates. For this task, a town official needs the most recent information with high resolution and positional accuracy. All this information can be communicated using the single graph of figure 8.

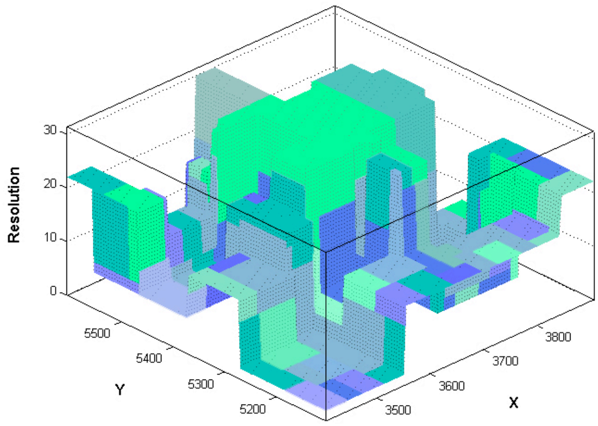

Figure 8: Combination of 3D, color, and saturation.

4.2.3 Addition of Other Attributes:

The previously discussed images offer an illustration of our visualization approach. In the included examples we convey resolution, currency, and positional accuracy. The remaining data quality attributes, namely completeness, scale and logical consistency, are conveyed using the following methods:

- Incompleteness of the data is shown by holes in the 3D illustrations, as can be seen in figure 9. To avoid the blocking of clear views of holes in the 3D surface, figure 9 shows a top view.

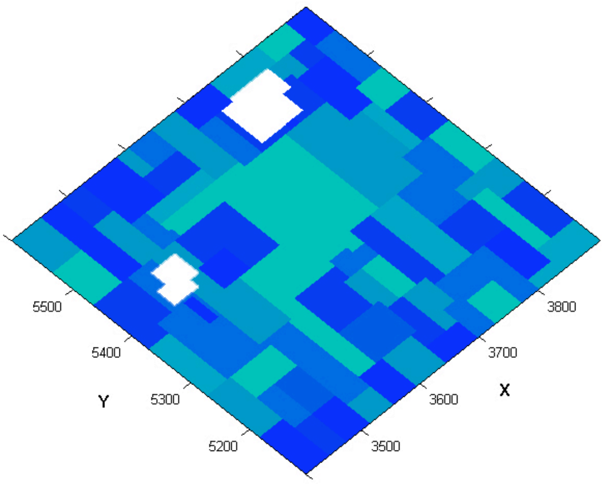

Figure 9: Top view of the visualization in figure 6 – white areas show incomplete data.

- Scale can be depicted on the z-axis in lieu of resolution.
- For communicating logical inconsistency in the data simple animations such as blinking can draw the user's attention to the inconsistent area.

5. CONCLUSIONS AND FUTURE WORK

This paper presented a visualization environment for spatial image data quality, that provides an additional tool for users to browse and choose the right data for their tasks and improve their work efficiency.

The data attributes positional accuracy, scale, resolution, completeness, consistency, and currency are visualized using a combination of 3D and color representations. We demonstrated how multiple attributes can be conveyed in a single visualization. Use of 3D and hue and saturation of color proved to be efficient tools to convey data quality of available datasets to users of various expertise.

Part of our future work is to further the development of advanced combined visualizations coupled with an intuitive user interface for the communication of geospatial image data quality. In addition to the quality attributes that were discussed in this paper, the benefit and visualization feasibility of additional data attributes will also be explored.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation through grant ITR-0121269 and by the National Geospatial-Intelligence Agency (NGA) through grants NMA 401-02-1-2008 and NMA501-03-BAA-0002. Their support is highly appreciated.

REFERENCES

Bastin, L., P. Fisher, and J. Wood, 2002. Visualizing Uncertainty in Multi-spectral Remotely Sensed Imagery. *Computers & Geosciences,* 28(3), pp. 337-350.

Beard, M. K., 1997. Representations of Data Quality. In: M. Craglia and H. Couclelis (Eds.) *Geographic Information Research - Bridging the Atlantic*. Taylor & Francis, London, pp. 280-294.

Beard, M. K. and B. P. Buttenfield, 1999. Detecting and Evaluating Errors by Graphical Methods. In: P. A. Longley et. al. (Eds.) *Geographical Information Systems - Priciples and Technical Issues*. John Wiley & Sons, New York, pp. 219-233.

Beard, M. K., B. P. Buttenfield, and Clapham, S. B., 1991. NCGIA Research Initiative 7 - Visualizing the Quality of Spatial Information. *NCGIA Technical Report 91-26*, Santa Barbara, California, pp. 170.

Beard, M. K. and W. Mackaness, 1993. Visual Access to Data Quality in Geographic Information Systems. *Cartographica* 30(2+3), pp. 37-45.

Buttenfield, B. and M. K. Beard, 1994. Graphical and Geographical Components of Data Quality. In: H. M. Hearnshaw and D. J. Unwin (Eds.*): Visualization in Geographical Information Systems*. Wiley & Sons, Chichester, England, pp. 150-157.

Buttenfield, B. P., 1993. Representing Data Quality. *Cartographica* 30(2-3), pp. 1-6.

Buttenfield, B. P. and M. K. Beard, 1991. Visualizing the Quality of Spatial Information. In: *Auto-Carto 10*, Baltimore, MD, pp. 423-427.

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

Davis, T. and C. Keller, 1997. Modelling and Visualizing Multiple Spatial Uncertainties. *Computer & Geosciences* 23(4), pp. 397-408.

Duckham, M., K. Mason, J. Stell and M. Worboys, 2001. "A Formal Approach to Imperfection in Geographic Information." *Computer, Environment and Urban Systems,* 25, pp. 89-103.

Ehlschlaeger, C., A. Shortridge, and M. Goodchild, 1997. Visualizing Spatial Data Uncertainty Using Animation. *Computers in GeoSciences***,** 23(4), pp. 387-395.

Evans, B., 1997. Dynamic Display of Spatial Data-reliability: Does it Benefit the Map User? *Computers & Geosciences*, 23(4), pp. 409-422.

Fauerbach, E., R. Edsall, D. Barnes, and A. MacEachren, 1996. Visualization of Uncertainty in Meteorological Forecast Models. *7th International Spatial Data Handling Conference*, Delft, The Netherlands, pp. 8A.17-28.

Goodchild, M., B. Buttenfield, and J. Wood, 1994. Introduction to Visualizing Data Validity. In: H. Hearnshaw and D. Unwin (Eds.) *Visualization in Geographical Information Systems*. Wiley & Sons, Chichester, England, pp. 141-149.

Howard, D. and A. MacEachren 1996. Interface Design for Geographic Visualization: Tools for Representing Reliability. *Cartography and Geographic Information Systems,* 23(2), pp. 59-77.

Hunter, G. and M. Goodchild, 1995. Dealing with Error in Spatial Databases: A Simple Case Study. *Photogrammetric Engineering & Remote Sensing,* 61(5), pp. 529-537.

MacEachren, A., 1992. Visualizing Uncertain Information. *Cartographic Perspectives* 13, pp. 10-19.

MacEachren, A., 1994. *Some Truth with Maps: A Primer on Symbolization and Design*. Association of American Geographers,Washington, DC, pp.129.

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

NIST, 1992. Federal Information Processing Standard. Washington, DC, US Department of Commerce, National Institute of Standards and Technology.

Paradis, J. and K. Beard, 1994. Visualization of Spatial Data Quality for the Decision Maker: A Data Quality Filter. *URISA Journal* 6(2), pp. 25-34.

Schweizer, D. and M. Goodchild, 1992. Data Quality and Choropleth Maps: An Experiment with the Use of Color. In: *GIS/LIS '92*, San Jose, CA, pp. 686-699.

van der Wel, F., R. Hootsmans, and F. Ormeling, 1994. Visualization of Data Quality. In: A. MacEachren and D. Taylor (Eds.) *Visualization in Modern Cartography*. Pergamon, pp. 313-331.

Veregin, H., 1999. Data Quality Parameters. In: P. A. Longley et.al. (Eds.): *Geographical Information Systems - Priciples and Technical Issues*, John Wiley & Sons, New York, pp. 177-189.

Worboys, M. and M. Duckham, 2004. *GIS: A Computing Perspective*, CRC Press.