

VISUALIZING MULTIPLE UNCERTAINTY SOURCES

James R. Miller
Department of Electrical Engineering and Computer Science
University of Kansas
1520 W 15 Street
Lawrence, KS 66045-7621
USA
jrmiller@ku.edu

ABSTRACT

Scientists, engineers, and decision makers in a wide variety of fields are faced with the problem of visualizing the effect of multiple parameters on quantities of interest. Oftentimes a model of some physical system is involved in which the participants are interested not only in the model results, but also how the various driving parameters of the model have influenced that result. Moreover, the data sources and models have uncertainties associated with them. It is generally important to be able to visualize those uncertainties, and especially to be able to quickly focus on areas where there is considerable disagreement. In this paper, we review the use of a new visualization technique called Attribute Blocks and demonstrate how this can be used to visualize uncertainty in multiple driving model parameters simultaneously.

KEY WORDS

Multidimensional and multivariate visualization; Uncertainty

1. Introduction

Multidimensional multivariate visualization refers to the visualization of n attributes as a function of their position in d -dimensional space [1, 2, 3]. A variety of visualization and exploration tools have been investigated, both for navigating the d spatial dimensions as well as for presenting how the n attributes vary across the d -dimensional spatial domain.

Attribute Blocks form the basis of a visualization technique in which a regular array of cells are employed, one cell being used to visualize the variation of a single attribute across the small region covered by the cell [4]. Originally conceived as a tool for visualizing the continuous variation of multiple independent parameters across some region of interest, we noticed that visualizing the same parameter derived from different sources provided an effective mechanism for visualizing the

inherent uncertainty involved in using those data.

In this paper we briefly review the concept of uncertainty in scientific models, then we demonstrate the use of Attribute Block displays to visualize this uncertainty.

2. Previous Work

The backdrop for the work presented here is multidimensional multivariate visualization in scientific and engineering contexts. There is an extensive literature in this area which space does not permit us to review in depth here. Some starting points into the literature include the general surveys by Wong and Bergeron [1], de Oliveira and Levkowitz [2], and Love [3]. Slocum's text [5] provides a cartographic perspective as well as an extensive survey of a variety of methods.

The area of uncertainty modeling and visualization is also a well-studied problem. Early surveys by MacEachren [6], Pang [7], and Gershon [8] are good starting points. More recent methods are described by Johnson and Sanderson [9] and MacEachren [10].

One traditional method of displaying multiple attributes (or uncertainty in a single attribute coming from multiple independent sources) is to use side-by-side (or small multiple) displays. Each display covers the same region of space, but shows different attributes or data sources. The user forms a mental image of the combined displays in order to visualize the values of all the attributes at a given location [5].

When the entire visualization is present in a single display, we eliminate the need for the user to shift their gaze from display to display, but we pay for this by creating a generally more complex single display that presents n attributes across the entire spatial region of interest.

An icon-based representation of multiple attributes has been developed by Healey based on what he calls *perceptual texture elements* (pexels) [11]. A pexel is a polygon of fixed width whose height and color can be used to encode attribute values at a given location. Moreover, based on studies of how the human visual system perceives and analyzes texture and patterns,

Healey exploits the fact that density and regularity of pixels can also be used to encode additional attributes. Three levels of regularity are identified: regular, irregular, and random. Given sufficient numbers of pixels, it is possible to perceive changes in the density of pixel placement, independent of regularity [11].

We have developed an alternative approach based on *Attribute Block* displays, an interactive arrangement of all or a selected subset of the n attributes into a dynamically configurable $k_r \times k_c$ array [4]. Each element of the Attribute Block array encodes a single attribute value. All attributes are defined everywhere; the Attribute Block array simply acts like a “screen door” of lenses, each allowing a single attribute value to be seen at the specific location. The dimensions of each “lens” – denoted $b_r \times b_c$ – are also dynamically adjustable. The current $k_r \times k_c$ array of $b_r \times b_c$ lenses is then tiled across the entire region

3. Using Attribute Blocks for Uncertainty Visualization

Cliburn, et al. developed and described several techniques for visualizing uncertainty in computational climatology models, both with and without climate change predictions [12]. Here we demonstrate how Attribute Blocks can be used to rapidly identify areas of a display where significant differences in predicted behavior exist.

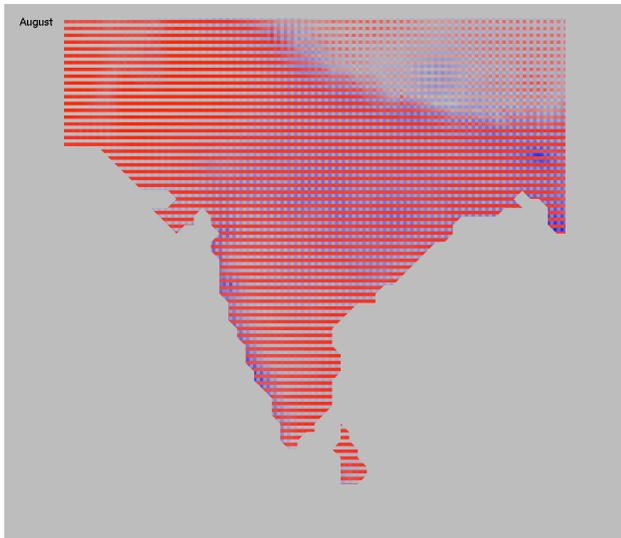


Figure 1: Visualizing uncertainty in temperature and precipitation inputs

Two critical driving parameters of the water balance model are temperature and precipitation. There are many different sources for this data, each typically starts from different raw data sources, and each typically uses different interpolation algorithms with various assumptions to obtain measures of these data at grid points throughout a region. To better understand how confident we can be with the predicted water balance

results in a region, we need to visualize the extent to which there is agreement or disagreement with the values for various critical input data. In figure 1, we see a display of two different temperature and two different precipitation data sources displayed across a region centered on India. The legend of figure 2 shows the template that is repeated across the display.

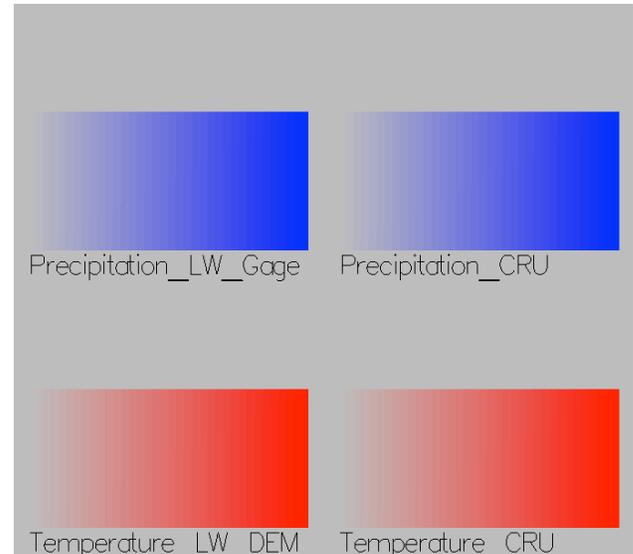


Figure 2: The legend showing how the two different temperature and precipitation data source displays are tiled across Figure 1.

From figure 2 we can see that alternating rows of the display will show temperature and precipitation inputs from the two different sources. Notice that we are using the same red color scale for the two temperature data sets. Similarly we are using the same blue color scale for the two precipitation data sets. Where they agree, we should therefore see a solid line segment (since the color scales are the same). Where they disagree, we will see alternating rectangles. By studying figure 1, we note that the two temperature predictions are in strong agreement across the bulk of India. In the mountainous region of the northeast, however, we see significant variations. This is not surprising since the two interpolation algorithms treat the altitude variations there very differently.

We can also visualize uncertainty due to varying precipitation inputs from the same display. The ability to see both in a single display is one of the strengths of this approach. Specifically, we see very little disagreement in the northwest – both sources predict very little precipitation there – however we see rather a lot of disagreement throughout much of the rest of the country. This is most noticeable in the center of the country and the leading edge of the mountains in the northeast.

One problem with the simple 2x2 arrangement used in figure 1 is that it is hard to detect which prediction is higher than the other. By altering the structure of the

attribute block display, we can make the shapes of the cells sufficiently distinct so that it is easy to tell at a glance which is which.

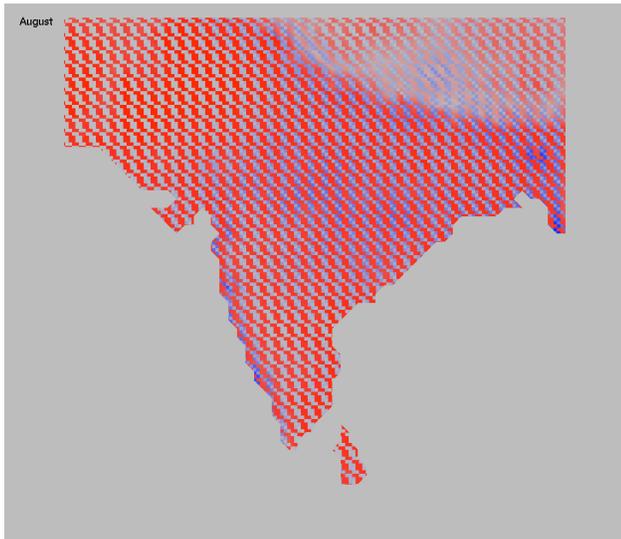


Figure 3: Using a 4x4 cell layout to make individual data sources more visually obvious.

Consider the example of figure 3 in which corner sections, diagonals, and broken vertical bars are used to make the two pairs of data sources more visually obvious.

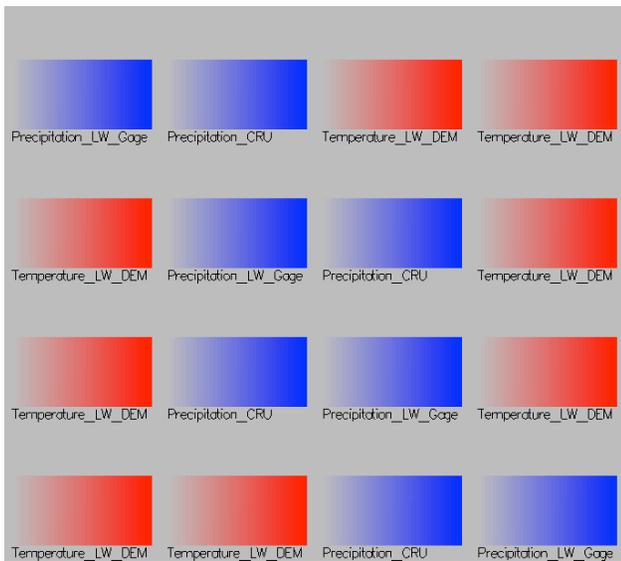


Figure 4: The assignment of data sources to cells used in figure 3.

The legend in figure 4 shows the actual assigned layout. The two temperature sources are displayed in two opposing corner layout orientations. One of the precipitation sources is laid out along a diagonal; the other appears as a vertical bar which is “broken”. That is, it shifts alternately to the right and left as it progresses from top to bottom.

By comparing the legend with the main display, it is immediately obvious that in areas of precipitation uncertainty, the “LW_Gage” data is uniformly drier than the “CRU” data. The diagonal pattern stands out very well.

Simple arrangement of the cells does not always work that well, however. For example, it is still a little difficult to determine which of the two temperature predictions is uniformly higher than the other. This is because the two opposing corner notches are blended together when the pattern is tiled across the display. We can overcome this problem by forcing a column and/or row divider line to appear in the background color. In the case of Figure 3, it suffices to introduce a single pixel wide column of background color between each column of Attribute Blocks. The image of figure 5 shows the result.

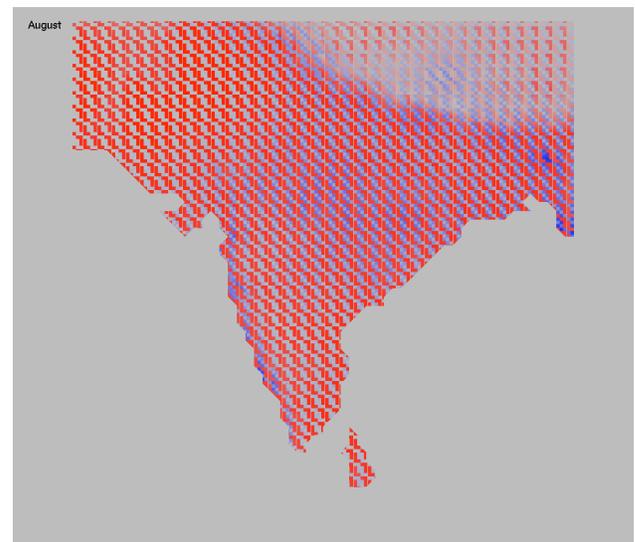


Figure 5: Introducing a single pixel wide column between Attribute Block cells to make opposing notches clear.

Notice that it is now immediately obvious that it is the “CRU” temperature data prediction that is uniformly lower than that of the “LW_DEM” data in areas such as the northeast where they differ.

We can study more than two types of data at once and/or more than two data sources for each type. In the system described by Cliburn, et al. [12], for example, we actually alternated between using three different temperature and three different precipitation data sets. They each differed in terms of how they selected weather stations from which to obtain raw data, whether they applied corrections for known errors in the data, the interpolation algorithms used, and whether the interpolation took into account geographical differences between the weather stations such as the interposition of mountains.

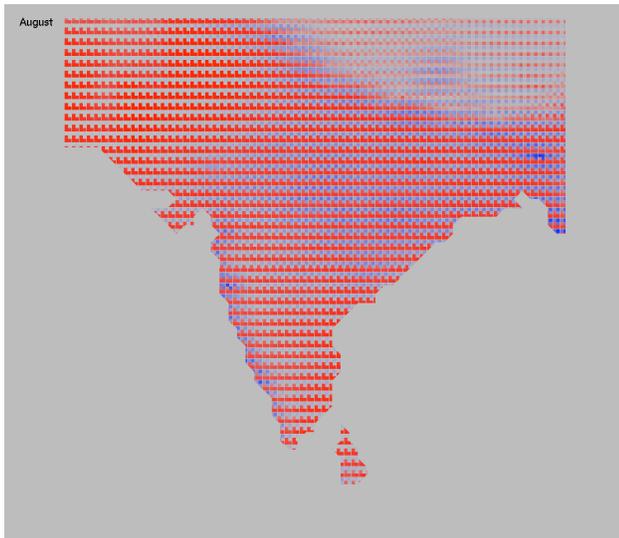


Figure 6: Using a 3x2 pattern of attribute block cells to show three temperature inputs and three precipitation inputs to the water balance model.

The legend in figure 7 shows the assignment of data sets to cell locations in the Attribute Block display.

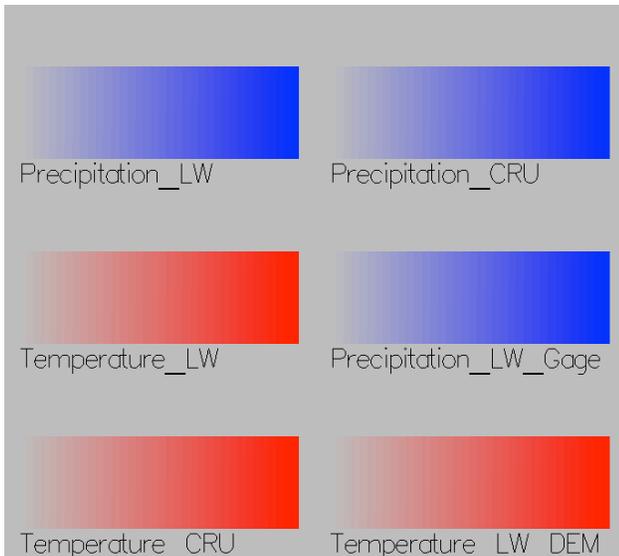


Figure 7: The assignment of data sets to attribute block cell locations used in figure 6.

Notice that the color scale creation strategy used throughout the paper is based on mapping minimal attribute values to background color and maximum values to the full red or blue. This has been found to be helpful in the context of our climatology visualizations because the greater the impact of a given parameter on the result, the more visually apparent it is in the display.

In the context of the uncertainty presentation being described here, however, that makes it slightly more difficult in some areas of the display to see the corner notch structures. Specifically, it is sometimes harder to

clearly identify which of the three temperature or three precipitation data values is the outlier. In the display of figure 6, we have used a one-pixel wide row and column separator to overcome this problem. Now it is apparent that, with the exception of a few areas (for example, in the central northern border region of the data set), the “CRU” predictions for precipitation are uniformly wetter than those for the other two.

Similarly, with a few exceptions, it is clear that the “LW_DEM” data is generally warmer than the other data sets in areas where they disagree.

This tool is able to point out systematic differences such as these. Notice that there is no claim that one data set or the other is “more correct” in areas of difference. This tool can simply be used to highlight high areas of uncertainty, thereby allowing the scientist to quickly be able to determine areas that perhaps require additional study.

4. Implementation

This system is designed to be used as a highly interactive exploratory tool. One needs to be able to rapidly alter the configuration of cells (i.e., the number of rows and columns in the template) as well as the specific assignment of attributes to cells. Moreover, we need to be able to dynamically alter the display size of each cell. Most of the images in this paper were generated using attribute block cells of size 6x6 pixels.

Our approach is based on using the programmable shader facility of the OpenGL graphics API [13]. We define OpenGL/GLSL attributes for each type of data in a given data set. We then define their values at vertices of polygons sent to the display. All attributes are therefore defined everywhere throughout the display. In our fragment shader, we calculate where we are in the generic Attribute Block template based on the current pixel coordinates, the pixel size of the attribute block cells, and the current numbers of rows and columns in the Attribute Block template.

The ability to relegate all these operations to the vertex and especially the fragment shader enabled the decoupling of Attribute Block cells from the geometry definition and was critical to the successful implementation of these ideas. Moreover, the entirety of the C++ program itself is completely independent of the actual data. Our data file begins with a short header that describes the data. The vertex and fragment shaders are then in separate files and employ the variable names identified in the data file. To use Attribute Blocks in a different context, one needs only design the data file header and modify the vertex and fragment shader files accordingly.

Even modifying the vertex and fragment shaders is straightforward. Much of the code for operations like mapping a pixel location to an attribute index is itself independent of the data. For those portions that are data-dependent (e.g., anything that involves the actual name of an attribute) we have a tool that automatically generates the code from the information in the data file header.

The current Attribute Block template is communicated to the fragment shader via uniform variables passed from the OpenGL program to the GLSL shader. On the OpenGL side, each entry of the Attribute Block template has an associated popout menu that allows the user to dynamically alter the attribute assigned to that location. This triggers a redisplay, and the image is then immediately redrawn with the new attribute assignment.

From the perspective of the fragment shader, Attribute Block cell sizes are always specified in pixel units. In some applications, the user wants to fix these cell sizes in model space. When this is desired, the application computes at the beginning of each display callback what the pixel dimensions are based on the current field of view and viewport dimensions. In either event, the fragment shader simply gets (via “uniform variables” [13]) the current b_x b_y dimensions in pixel units.

5. Further Applications

In the original water balance application from which the examples of this paper were derived, the primary parameters were temperature, precipitation, soil holding capacity, and potential evapotranspiration. These were all known (or computed) throughout a region at a certain time resolution (monthly averages, in our case). These quantities fed our computational water balance model, allowing us to approximate the water balance in the region at the same time and spatial resolution.

The original visualization of this model was developed as the key portion of a collaborative decision support system [12]. That application focused on visualizing the computed water balance as well as our level of confidence in that final measure. But to adequately address the underlying goal of the effort – collaborative policy and decision making – it became clear we needed to understand not only *where* there was a deficit or surplus of water, but also get an idea as to *why* the surplus or deficit was there. From a policy making perspective, this understanding could provide insight into appropriate remediation strategies, if any. For example, we might be able to better understand whether irrigation might be appropriate to address a periodic chronic deficit in a region, or whether instead alternative land usage decisions should be made. Coupled with the ability to incorporate climate change predictions along with their uncertainties (which the system also modeled [12]), the tool could be especially helpful for policy makers.

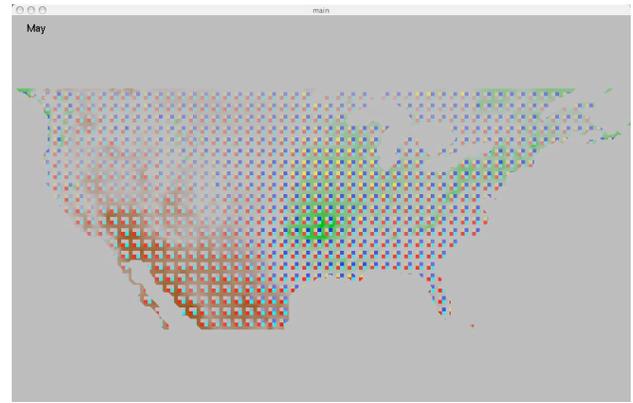


Figure 8: Visualizing the Water Balance Model Prediction Along With its Primary Driving Parameters

The example of figure 8 uses a 3x3 attribute block display in which the leftmost column and bottommost row display the water balance using a diverging brown to background to green color scale. (See figure 9.) Brown indicates a water deficit; green indicates a water surplus; background colors indicate a balance between need and supply. The interior 2x2 portion of the 3x3 attribute block display shows the four primary driving parameters of the model: temperature (red), precipitation (blue), soil holding capacity (yellow), and potential evapotranspiration (cyan).

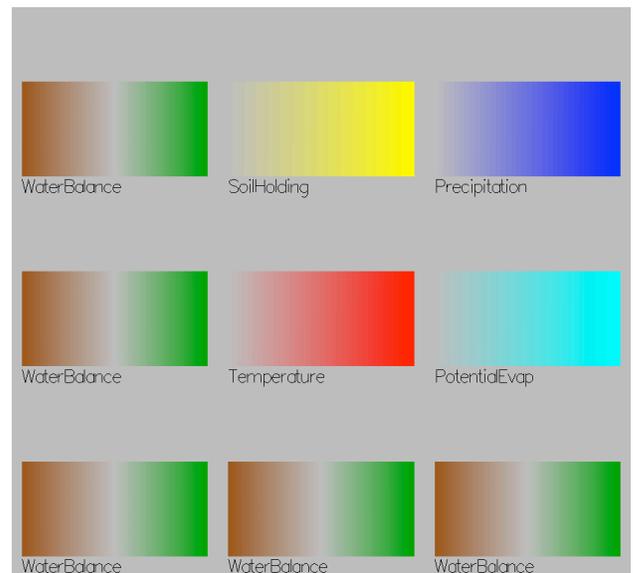


Figure 9: The Attribute Block Cell Assignments for the display of Figure 8.

The value of this display is that one can simultaneously visualize the calculated water balance along with getting insight into why there is a surplus or deficit in each area. In the Midwest, for example, the high soil moisture holding capacity has a clearly obvious effect on the high water surplus. In the southwest, we see how high temperatures and high water demand (potential

evapotranspiration) combine with relatively low precipitation to produce a significant water deficit.

6. Summary

We have described the Attribute Block visualization method, focusing on its applications in uncertainty visualization. As demonstrated in section 5, however, it is also generally useful for any sort of multivariate visualization application.

While it is admittedly difficult to obtain a feel for the precise value of any given attribute at a particular location in an Attribute Block display, the strength of this method seems to be the ability to rapidly locate areas where one attribute or another is most dominant. In an uncertainty visualization context, we are able to quickly see areas where various predictions differ considerably. This is certainly an important visualization task. Moreover, it is clearly straightforward to address the need for quantitative measures by supplementing this technique with a “click and probe” tool which would allow precise values to be displayed in any of a number of ways for a given attribute at a selected location.

References

- [1] P. C. Wong and R. D. Bergeron, 30 Years of Multidimensional Multivariate Visualization, in *Scientific Visualization: Overviews, Methodologies, & Techniques*, G. M. Nielson, H. Mueller, and H. Hagen (editors), (IEEE Computer Society Press, Los Alamitos, CA, 1997).
- [2] M. C. F. de Oliveira and H. Levkowitz, From Visual Data Exploration to Visual Data Mining: A Survey, *IEEE Transactions on Visualization and Computer Graphics*, 9(3), July-September 2003, 378-394.
- [3] A. L. Love, A. Pang, and D. L. Kao, Visualizing Spatial Multivalued Data, *IEEE Computer Graphics and Applications*, 25(3), May/June 2005, 69-79.
- [4] J. R. Miller, Attribute Blocks: A Tool for Visualizing Multiple Continuously-Defined Attributes, *IEEE Computer Graphics & Applications*, 27(3), May/June 2007, 57-69.
- [5] T. A. Slocum, R. B. McMaster, F. C. Kessler, and H. H. HOWARD, *Thematic Cartography and Geographic Visualization*, second edition, (Upper Saddle River, NJ: Pearson Prentice Hall, 2005).
- [6] A. M. MacEachren, Visualizing Uncertain Information, *Cartographic Perspective*, 13, 1992, 10-19.
- [7] A. Pang, C. Wittenbrink, and S. Lodha, Approaches to Uncertainty Visualization, *The Visual Computer*, 13(8), November 1997, 370-390.
- [8] N. Gershon, Visualization of an Imperfect World, *IEEE Computer Graphics and Applications*, 18(4), July/August 1998, 43-45.
- [9] C. R. Johnson and A. R. Sanderson, A Next Step: Visualizing Errors and Uncertainty, *IEEE Computer Graphics and Applications*, 23(5), September/October 2003, 6-10.
- [10] A. M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler, Visualizing Geospatial Uncertainty: What We Know and What We Need to Know, *Cartography and Geographic Information Science*, 32(3), July 2005, 139-160.
- [11] C. G. Healey and J. T. Enns, Large Datasets at a Glance: Combining Textures and Colors in Scientific Visualization, *IEEE Transactions on Visualization and Computer Graphics*, 5(2), April-June 1999, 145-167.
- [12] D. C. Cliburn, J. J. Feddema, J. R. Miller, and T. A. Slocum, Design and Evaluation of a Decision Support System in a Water Balance Application, *Computers & Graphics*, 26(6), December 2002, 931-949.
- [13] R. Rost, *OpenGL Shading Language*, second edition (Upper Saddle River, NJ: Addison-Wesley, 2006).