PERGAMON

Computers & Graphics 26 (2002) 931–949

COMPUTERS
&GRAPHICS

www.elsevier.com/locate/cag

Technical section

Design and evaluation of a decision support system in a water balance application

Daniel C. Cliburn[a,*], Johannes J. Feddema[b], James R. Miller[c], Terry A. Slocum[b]

[a] *Department of Mathematics and Computer Science, Hanover College, PO Box 890, Hanover, IN 47243-0890, USA*
[b] *Department of Geography, The University of Kansas, Lawrence, KS 66045, USA*
[c] *Department of Electrical Engineering and Computer Science, The University of Kansas, Lawrence, KS 66045, USA*

Abstract

Visualization has become a vital tool for representing the results of scientific models in decision support applications. Both the raw data and the models from which these visualizations are derived usually have considerable uncertainty associated with them. Decision-makers are typically presented with results from these models with little or no insight as to the reliability of the information shown. For effective decisions to be made, a decision support system should allow collaborative participation from scientists and decision-makers, and it should display the locations, magnitudes, and sources of uncertainty in the results. This research work discusses a software application for visualizing the results of a water balance model and its associated uncertainty. The effectiveness of the application and its visual presentation methods were incrementally tested and improved through usability engineering principles.

Keywords: Visualization; Uncertainty; Usability engineering; Decision-making

1. Introduction

Visualization has been an active area of research ever since an issue of *Computer Graphics* was devoted to the topic in 1987 [1,2]. Visualization provides the means by which we can see a representation of data that otherwise would not have a visual form. For a human being, attempting to interpret raw data can be a very daunting task, especially if the data set is large and multidimensional. Since 50% of the brain's neurons are associated with vision [1], a visual representation of data allows us to utilize brainpower that we would not otherwise be accessing and allows us to see spatial relationships in data that may not otherwise be obvious.

An important application for visualization is in the area of decision-making. For example, as our understanding of natural phenomena has improved in recent times, scientists have developed computer simulations for representing these phenomena. Results from these simulations often form the basis for policy changes and new legislation. However, policy makers rarely have the time or background to fully interpret and understand the data produced from these simulations. An environment in which a scientist and a decision-maker can visualize the results of various model scenarios and discuss their significance can go a long way towards helping the policy maker minimize their time and maximize their understanding.

Unfortunately, even in the best environmental modeling and simulation systems there remain various types of uncertainties associated with data collection and manipulation, model accuracy, and people's interpretation of data and model results. A good example of such uncertainties is demonstrated in simulations of global climate change. Uncertainty in this data has had a profound impact on the interpretation of climate modeling results, ultimately resulting in stark international political conflicts as seen in the results of the Rio and Kyoto global climate change summit meetings [3].

---

*Corresponding author. Fax: +1-812-866-7229.
E-mail address: cliburn@hanover.edu (D.C. Cliburn).

PII: S0097-8493(02)00181-4

At the heart of the uncertainty problem is the fact that sophisticated computer graphics approaches allow us to produce attractive high-quality 3D color images that impart "truth" to simulations. We intuitively believe as humans that what we see is reality [4], however, this is often not the case in visualization as simulations are characterized by uncertainty both in the data and in the model used to create the simulations. When designing visualization tools to be used in a decision support context for policy related to natural phenomena, it is incumbent upon the system designer to display not only results of simulations, but also a reasonable estimate of the uncertainty associated with those results. That is, decision-makers must understand not only the results, but also the reliability of those results.

The goal of the system described in this paper is to allow decision-makers and their staffs to explore the results of a water balance model along with associated uncertainties to better understand the potential impacts of public policy decisions that might be under consideration. To effectively make decisions, decision-makers need to know about the uncertainty inherent in scientific models and, more importantly, to what that uncertainty can be attributed. The application developed as part of this research effort identifies sources and magnitudes of the uncertainties that can be quantified or approximated in the simulation. This is accomplished via a software tool that allows decision-makers to visualize both the model and its results (with uncertainty information).

Little has been published regarding the effectiveness of current uncertainty presentation techniques; formal evaluation of applications that attempt to quantify and present uncertainty has been largely neglected. In the work reported here, we include an extended discussion of a qualitative evaluation of our system through usability engineering practices [5–7]. Our goal was to ensure that we developed a product that was both useful and effective, and that we could quantify our level of success.

2. Water balance visualization system

We have created a prototype decision-making system that enables decision-makers to visualize results from a water balance model (a calculation of surplus or deficit in water supply) for terrestrial regions of the world. The water balance model [8,9] is representative of the many models that are considered when formulating policy. Results from this particular model are helpful when determining the environment's natural ability to support agriculture. The base water balance model can also be used to make predictions about future natural water supplies by using Global Circulation Models (GCMs). These GCMs provide predictions of how various climatic variables will change due to the effects of global warming [3].

2.1. Water balance model

A water balance model developed by Thornthwaite [8] and modified by Feddema [10,11] can be used to generate information about water surplus or deficit for regions of the world. Reflecting the nature of the water cycle, the water balance model has two major inputs: climatic data (climatology) and land parameter data.

The climatic data are gridded data sets consisting of precipitation and temperature information for particular locations around the world. There are several data sources freely available at the $0.5^\circ \times 0.5^\circ$ resolution employed in our application [12–16]. The land parameter information required by the model is the water holding capacity of the soil. In many cases, a constant value of the average soil holding capacity has been used, but a data set is also available with this information at a $0.5^\circ \times 0.5^\circ$ resolution [17].

The water balance model calculates potential evapotranspiration (E_p) based on temperature and latitude at a specific point location [8], using one of several methods [8,18]. Potential evapotranspiration is the amount of moisture that would be lost from the soil and land surface to evaporation and transpiration, if at least that much moisture was present. For example, if the amount of precipitation received in a region plus the amount of moisture available from the soil is equal to or exceeds E_p, then there is a surplus; otherwise there is a deficit. Actual evapotranspiration (E_a) is a measure of the actual amount of water lost to the environment. Moisture deficit and surplus can be calculated from the difference between E_p and E_a, and change in soil moisture storage can be determined from this surplus/deficit.

The model just described provides estimates of today's likely climate in selected regions. Considering future predictions about the water supply allows us to anticipate the environment's future ability to sustain plant and animal life. This can be done with the aid of the GCM data sets, which report delta values (amount of change) from the base conditions (today's climate) to the time period of interest. Adding these delta values to the input climatology data, and then running the model, allows us to create future water budget predictions. Unfortunately, the scale of the GCMs is larger (and varies depending on each GCM, see [3]) than that of the base input data. Therefore, we must scale the GCM data down to the finer $0.5^\circ \times 0.5^\circ$ resolution of the climatology data. Many interpolation methods exist to do this [19]; we currently support two such methods: a simple nearest neighbor approach and an inverse distance average of the nine closest neighbors.

2.1.1. Uncertainties in the model

We consider uncertainty to be the inverse of reliability for some point in a visualization and it is defined for all points in a visualization. Uncertainty can be due to many factors, but in this application, we quantify uncertainty as the difference in model results due to the choice of input data sets. As indicated in the preceding section, there are many data sources from which the temperature and precipitation information can be taken, and these sources do not always agree. It is frequently the case that different choices for input data sets in a given region can lead to markedly different model results. If we arrive at the same or similar values with each data set for a particular region, then we consider that region to have low uncertainty. However, it is important to note, that all results could be incorrect. Agreement among the model results (low uncertainty) does not necessarily indicate accuracy, as all results could be wrong.

There are several reasons that these data sets differ. Some use different station selection criteria, such as temporal constraints, resulting in different station networks. Additional uncertainty derives from different interpolation methods employed when generating the $0.5^\circ \times 0.5^\circ$ gridded data sets [20]. Furthermore, there are many other types of uncertainty associated with the raw data sets themselves—uncertainty in regard to data collection techniques and recording errors, to name two. It would not be feasible to quantify all of these sources of uncertainty, even if they were known. What we have found to be the best estimate for decision-making purposes is to look at the variation in model results when using different data sets.

Similar remarks apply to the analysis of uncertainty in climate change prediction. There are many GCM models, each providing different estimates of climate change. Moreover, the approach used to scale the GCMs down to the resolution of the base climatology data affects the resulting predictions. Our approach to quantifying the uncertainty in GCM predictions is therefore similar to that described for the water balance model itself. We record the results from several combinations of GCM predictions and scaling methods and use the variations among the results as a measure of uncertainty.

2.2. System architecture

The *Collaborative Visualization Room* is the visualization centerpiece of DesignLab, an interdisciplinary research laboratory established with an NSF Infrastructure grant. The room contains a 25 × 6 foot wall-mounted display driven by three SGI InfiniteReality2 graphics subsystems, projecting a total of 5760 × 1200 pixels (see Fig. 1). These graphics subsystems are in turn driven by a multiprocessor SGI Origin 2000 with 6

Fig. 1. Collaborative Visualization Room with an early version of the Water Balance Visualization System executing.

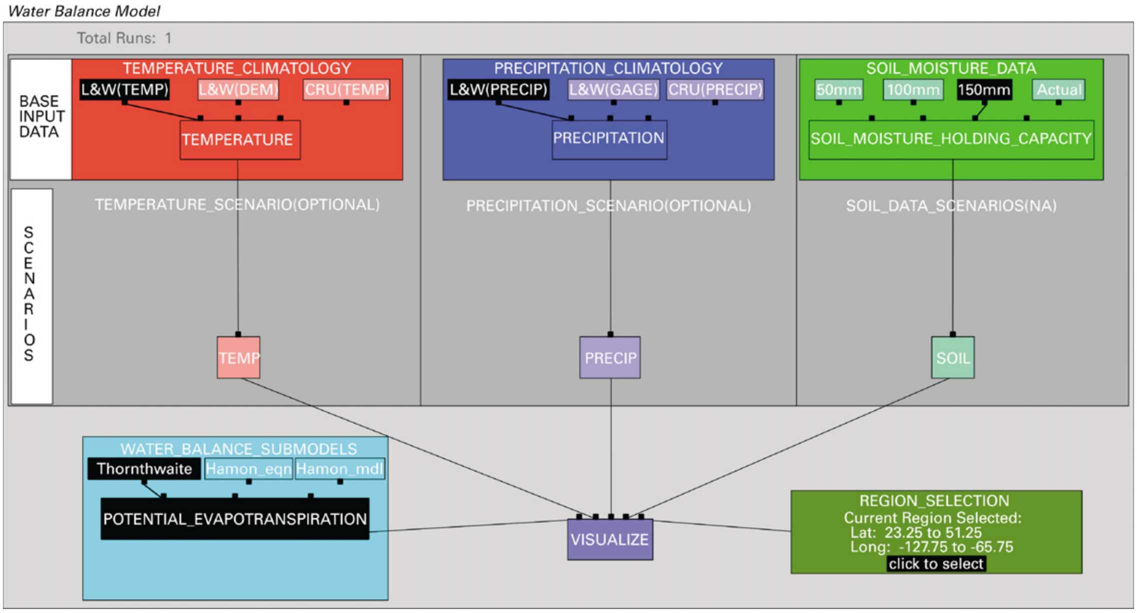
250 MHz R10000 MIPSpro processors and 1.024 GB RAM. The room comfortably accommodates ten or more people and could facilitate collaboration amongst groups of decision-makers. We plan future investigation into the potential advantages of running an application of this nature in a collaborative environment versus on a standard desktop PC.

The application was developed with C++ using OpenGL [21] for graphics rendering and the OpenGL Multipipe SDK [22] for event handling and window management. The OpenGL Multipipe SDK allows us to take advantage of the Collaborative Visualization Room's immense screen space by providing the needed functionality for simultaneous rendering on multiple graphics pipes. We dedicated the left one-third of the screen space for the model definition interface (see Section 2.3) and used the remainder for surface visualization (Section 2.4). An advantage with this architecture is that it allows users to see model definitions and their corresponding visualizations simultaneously. A desktop PC implementation of the current software would contain overlapping windows, requiring users to toggle back and force to see the information contained within.

2.3. Model definition

Decision-makers must have a good understanding of how scientific models work in general as well as an understanding that uncertainty in the data and models is an unavoidable fact of life. We believe that an important first step for learning about models and dealing with their uncertainty is to have an understanding of the model that produces the uncertain results. It is particularly crucial to understand which model parameters have an impact on uncertainty and to what extent. To this end, we have implemented a visual programming interface that provides a "road-map" of the model showing connections between model inputs and components (shown in Fig. 2). Users are given the ability to interactively manipulate model inputs and change model parameters in the visual interface. They get a feel for uncertainty in the results as they see how these results vary based on the interactive changes. This interface is presented in a separate window so the user can see the how the model was constructed while viewing the model results.

The model definition interface (Fig. 2) is divided into three rows. The first row deals with the basic model inputs of temperature, precipitation, and soil moisture holding capacity; the second row is used to define future water balance model scenarios using GCMs; the last row of the interface is for defining water balance submodels (potential evapotranspiration, for example) and region selection. Users can select any region of the world for which they would like to run the model by dragging a rectangle over the area in the region selection window (Fig. 3). When the user clicks on VISUALIZE, the water balance is computed and displayed for the designated area to the right of the visual programming window, as shown in Fig. 1.



Current Region Selected:
Lat: 23.25 to 51.25
Long: -127.75 to -65.75

Fig. 2. Model definition window depicting the water balance model.

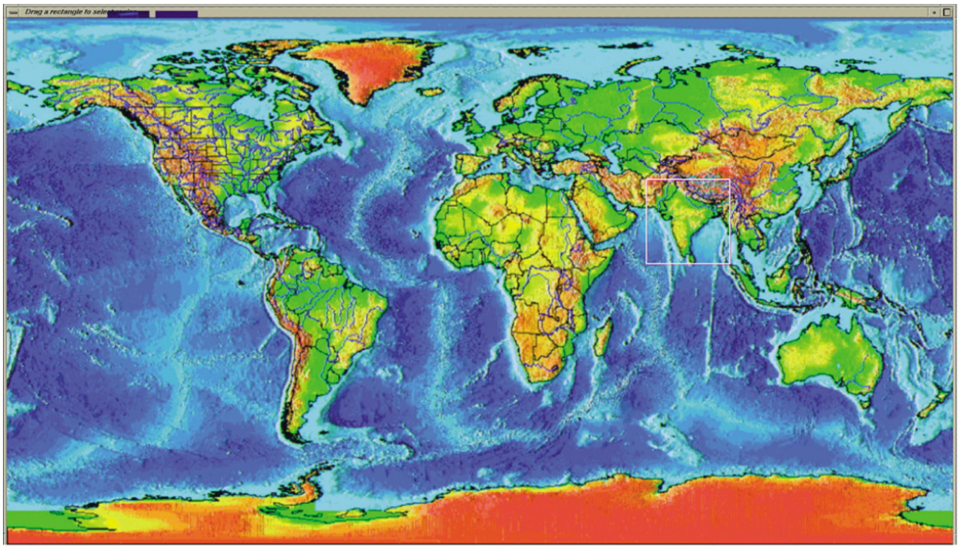


Fig. 3. Region selection window. Area to be visualized is outlined in white.

2.4. Surface visualization

Basic water balance results are displayed with a redundant coding scheme based on both color and height (see Fig. 4). High elevation and dark blue coloring represent surpluses, while low elevation and dark red coloring represent deficits. Each visualization includes a legend indicating the applicable coloring scheme. The legend is interactive allowing the user to click on the display to determine the actual value computed for that location.

Two shaded surface display modes are supported for these basic surface results. In the smooth surface shading mode illustrated in Fig. 4A, the surface is rendered by drawing triangle strips whose vertices are grid cell centers. The surplus/deficit results are linearly interpolated across the interior of the triangle. While visually appealing, this type of surface display is misleading in that model results are really not continuous as suggested by the display. As we explained above, the model results are actually cell-based, where each value represents the average value expected in the $0.5^\circ \times 0.5^\circ$ region covered by that cell. Our other shaded surface display mode displays directly this cell-based nature of the simulation results (see Fig. 4B). A prism is drawn showing the actual surplus/deficit as a constant height value at the location of each grid cell.

Yet another surface display mode we have used is the wire mesh surface (Fig. 4C). This surface is similar to "smooth" except that only lines are drawn between grid cell centers. This type of surface display is useful when viewing uncertainty information (discussed in the next section), as the amount of information can easily become visually overloading.

Predicted change in water availability due to climate change is visualized using an orange–purple diverging color scheme. Dark orange and dark purple indicate less water and more water, respectively, from the current climate value. We felt it was important to use colors not already associated with other types of information in the application, and this color pair is known to be an effective diverging scheme [23].

Fig. 5A shows a predicted change surface calculated using the nearest neighbor interpolator. The coarse resolution of the GCM grid cells is clearly visible in the coloring scheme on the surface. Fig. 5B depicts the same conditions, but calculated using an inverse-distance interpolator. This scheme produces much smoother results, but does not represent the GCM as accurately since water volume is not preserved in each grid cell.

2.5. Visualization of uncertainty

The visualization of uncertainty in data is not new [24,25]. However, previous systems have left much unaddressed in depicting uncertainty, particularly in representing it to decision-makers in a useful way. In a decision-making context, it is not only important to highlight where uncertainty exists, but we also wish to determine and report the source of and reasons for this uncertainty. Specifically, we describe how our system can be used to:

1. Allow users to explore uncertainty and access uncertainty in visualization applications by defining their own choices for model parameters.
2. Depict multiple sources of uncertainties simultaneously in a way that is manageable for users.

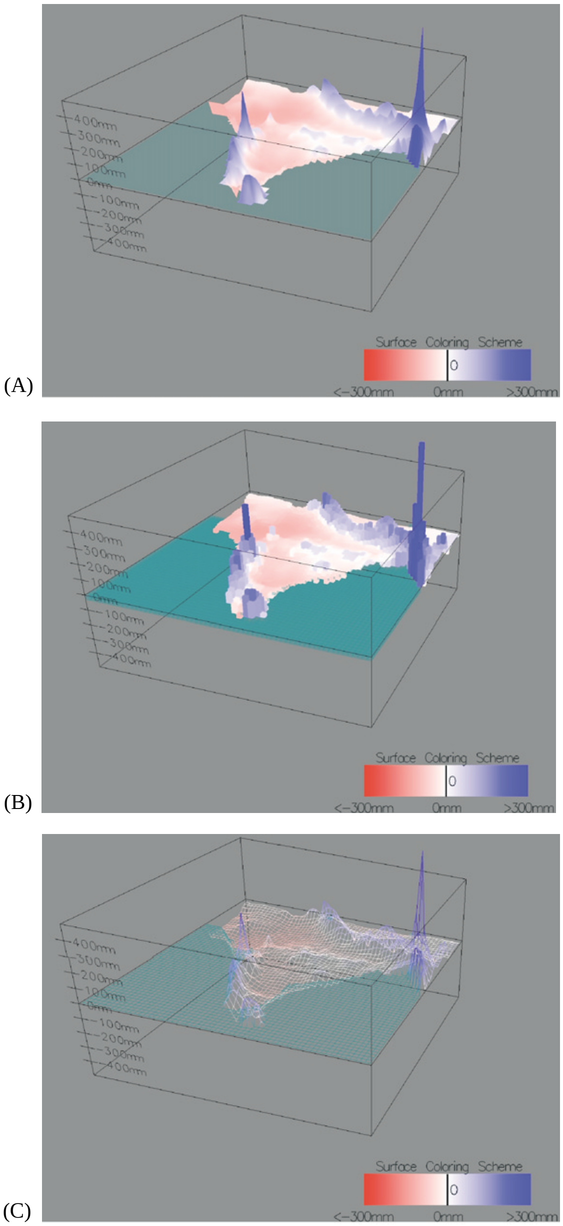


Fig. 4. Model results visualized using a redundant coding scheme (height and color): (A) smooth surface; (B) cell-based surface; and (C) mesh surface.

By varying the selection of input data sets (temperature, precipitation, and soil moisture holding capacity) and choice of potential evapotranspiration method, we can produce several scenarios along with the water surpluses and deficits they predict. It is the variation between these predictions that we use as a measure of "uncertainty". If we make multiple model runs, modifying the same input parameter each time, we can attribute variation in results to that parameter. If we do this for all parameters, we can approximate the uncertainties associated with each parameter choice. These uncertainties can then be visualized.

Gershon [4] states that information pertaining to objects in a visual scene (in this case, uncertainty information) can be presented with techniques falling into two general categories: intrinsic and extrinsic. Intrinsic techniques vary an object's appearance to show uncertainty associated with it, while extrinsic techniques rely on additional geometry in the scene to highlight areas and levels of uncertainty. We have developed both intrinsic and extrinsic uncertainty presentation schemes for showing the uncertainty in base and future predictions.

2.5.1. *RGB Scheme*

The RGB Scheme is an intrinsic tri-variate uncertainty visualization scheme. The RGB approach has been utilized before [26], and other coloring schemes have been applied with tri-variate data [27,28], but not in the context of uncertainty. The idea is to assign a color to each location on a surface indicating how much uncertainty is located there, with respect to three different variables. The color shown is based on the three primary colors of light: red, green, and blue, with each input variable assigned one of these colors. The intensity of each color is based on the magnitude of uncertainty for its corresponding variable.

We applied this scheme to the water balance model by mapping red, green, and blue to temperature, soil, and precipitation, respectively (staying consistent with the coloring scheme employed in the model definition window of Fig. 2). Fig. 6A illustrates the RGB Scheme, with the surface height indicating the base run moisture conditions for July over India.

As can be seen in Fig. 6A, black areas on the surface indicate regions with little or no uncertainty with respect to any of the variables, since the intensity of each color component is very low. Bright green regions indicate areas where the choice of a value for soil moisture holding capacity makes a big difference in model outcomes. Blue regions indicate areas where precipitation has the biggest impact on uncertainty. Likewise, red areas have the greatest amount of temperature uncertainty. Of course, some areas of the surface have colors that are combinations of red, green, and blue, indicating significant contributions to uncertainty from two or more of the variables. Users can click on the surface and the legend will indicate the amount of variation between model runs with white lines in the horizontal bars corresponding to each variable (shown in Fig. 6B).

2.5.2. *Line Glyph scheme*

Another scheme for visualizing model uncertainty is an extrinsic presentation method, adding geometry to a

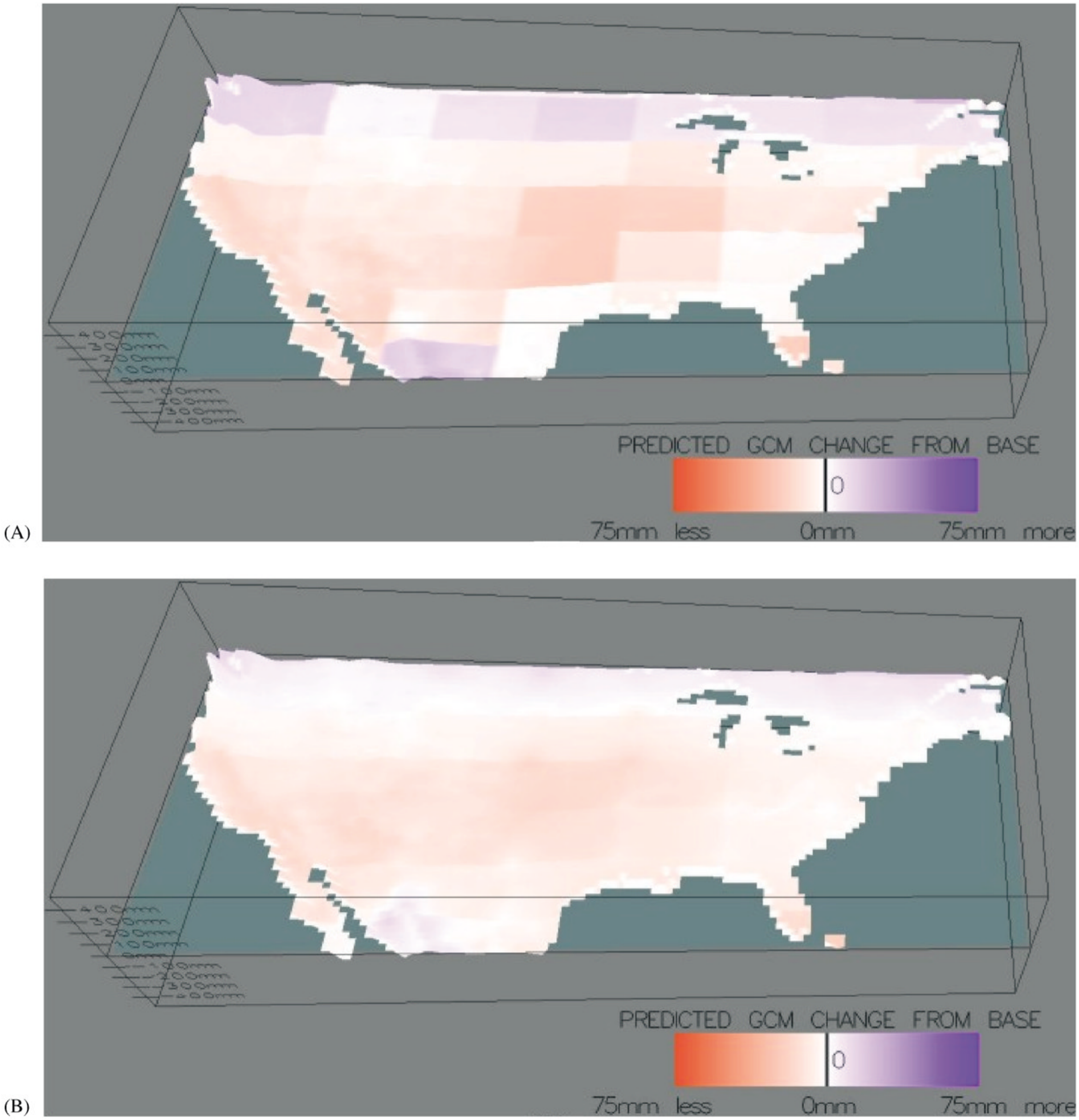


Fig. 5. Base surface colored with an orange–purple coloring scheme to show less water or more water, respectively: (A) using a nearest interpolator and (B) using an inverse-distance interpolator.

scene to indicate areas of uncertainty. In this method, vertical bars are placed at each grid cell location for each variable. The height of a bar indicates the magnitude of the corresponding variable's uncertainty. Each bar must touch the base surface, but the base could be at the bottom, top, or pass through the middle of the bar. Thus, the bars show the range of variation around the base, for each variable. Fig. 7A shows this scheme for the state of Kansas in the Midwestern United States.

Clearly, this type of display can overload the user with information. Clicking on the surface turns all glyphs off except those indicating uncertainty for the selected location (see Fig. 7B) allowing the user to focus the display on a region of interest.

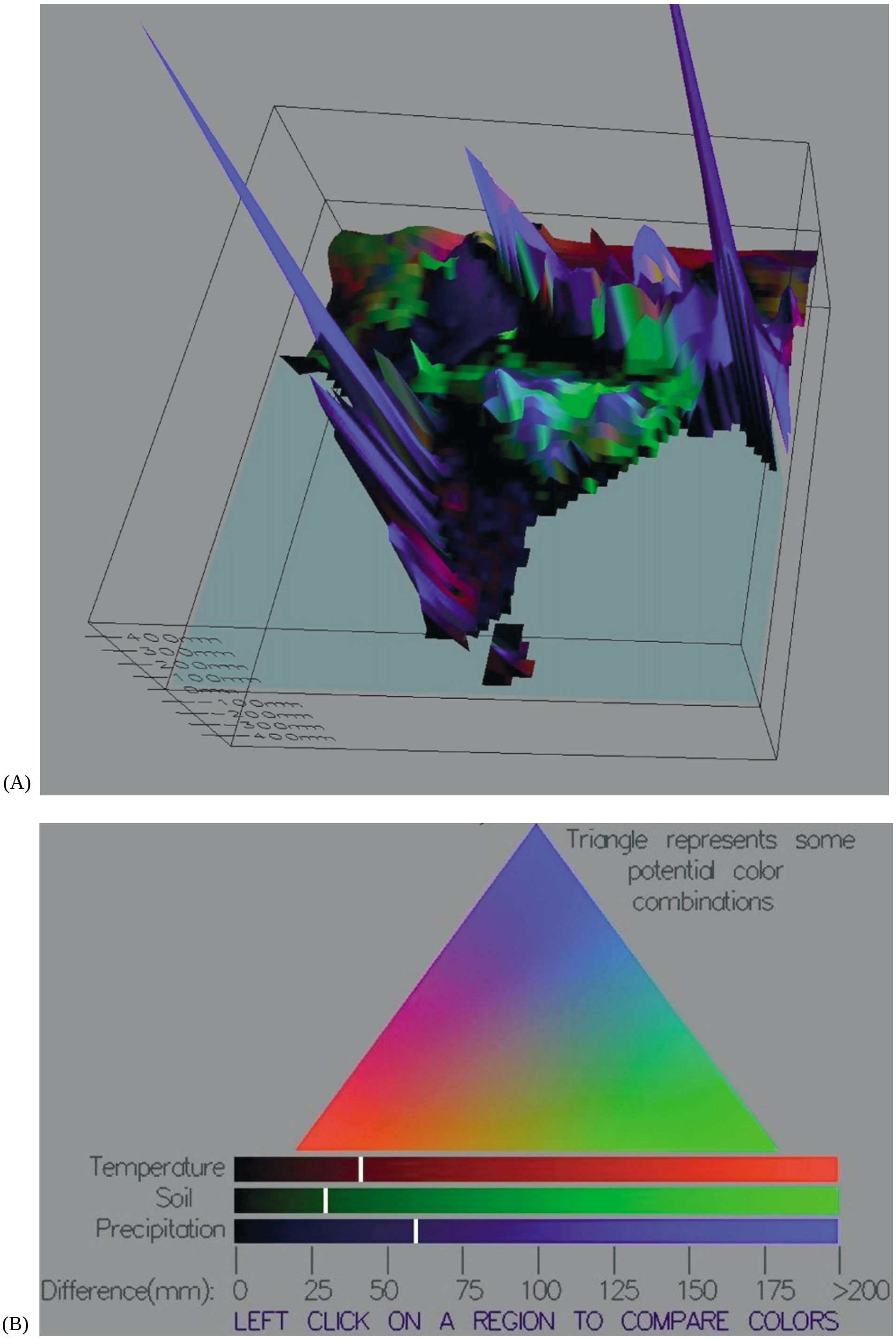


Fig. 6. (A) Surface colored with the RGB Scheme and (B) RGB Scheme legend.

2.5.3. Visibility scheme

This intrinsic scheme for depicting uncertainty in future climate prediction uses transparency to make uncertain data hard to see. Transparency can be applied to a surface, making some points visible, and others invisible. This technique can actually be applied in two ways. One is to make the surface invisible in areas of high certainty. The rationale here is that we want to

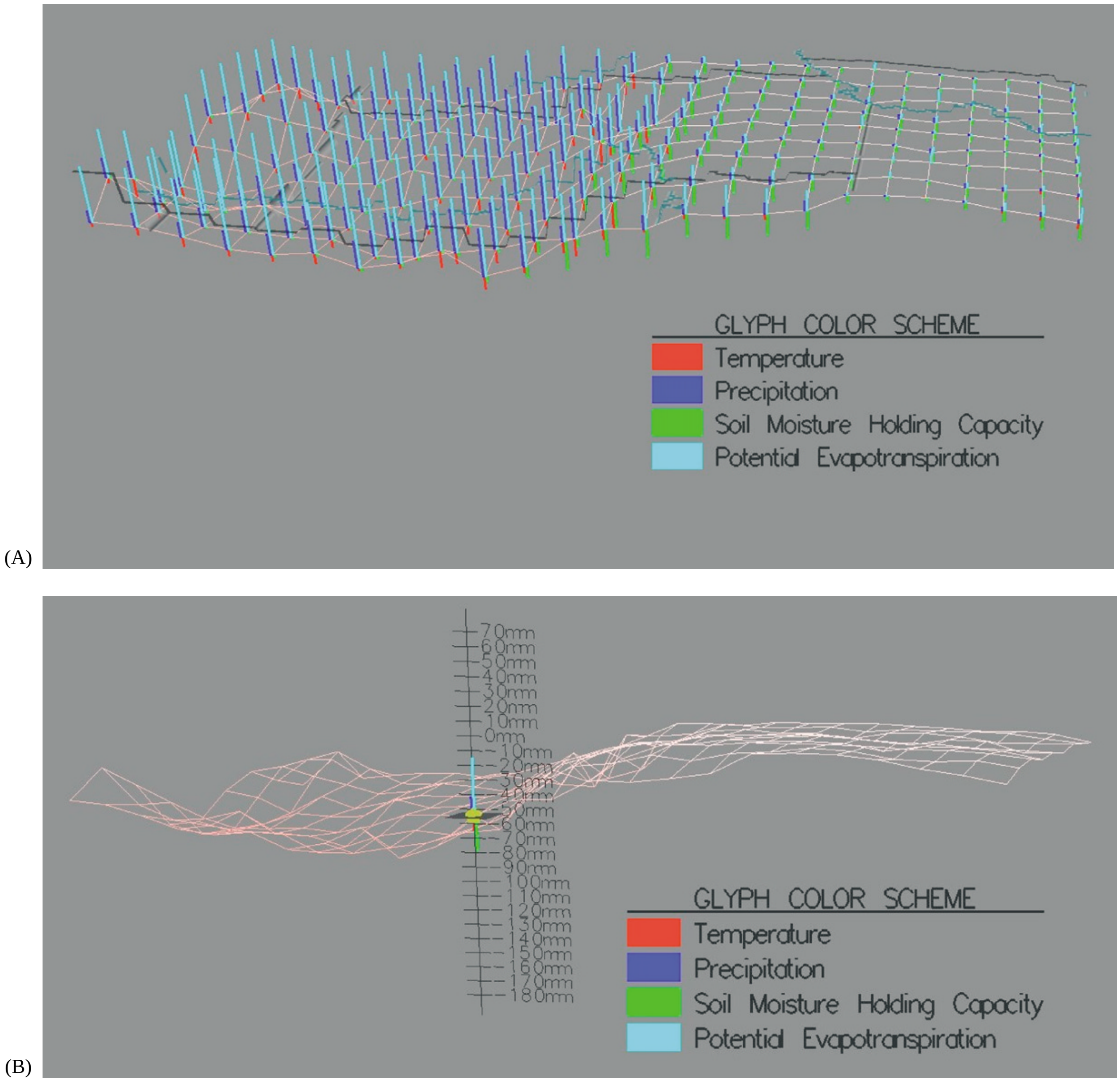


Fig. 7. Line Glyph Scheme: (A) separate glyphs are displayed for all input variables and (B) glyphs are only displayed for the currently selected location.

make uncertain points stand out. The second strategy is to make uncertain points invisible. It is this second idea that we implemented with the rationale that we intend this as an application to aid in decision-making—we choose not to present information that is considered unreliable for making decisions. This technique (Fig. 8b) was applied to the visualizations depicting the average predicted change, calculated from the deviation of all selected GCM models, with the orange/purple coloring scheme (shown in Fig. 8A).

The advantage of the Visibility Scheme is that it can easily allow users to see the uncertainty in large regions (as is the case in Fig. 8B with India). The disadvantage is it can be hard for users to get a grasp of the relative amount of uncertainty corresponding to various levels of visibility. While the legend is interactive and users can find exact values at every location, it can require significant time and interaction to determine exact uncertainty amounts for large regions.

2.5.4. *GCM Glyph Scheme*

The GCM Glyph Scheme is an extrinsic uncertainty presentation method for displaying uncertainty associated with future water budget predictions. This

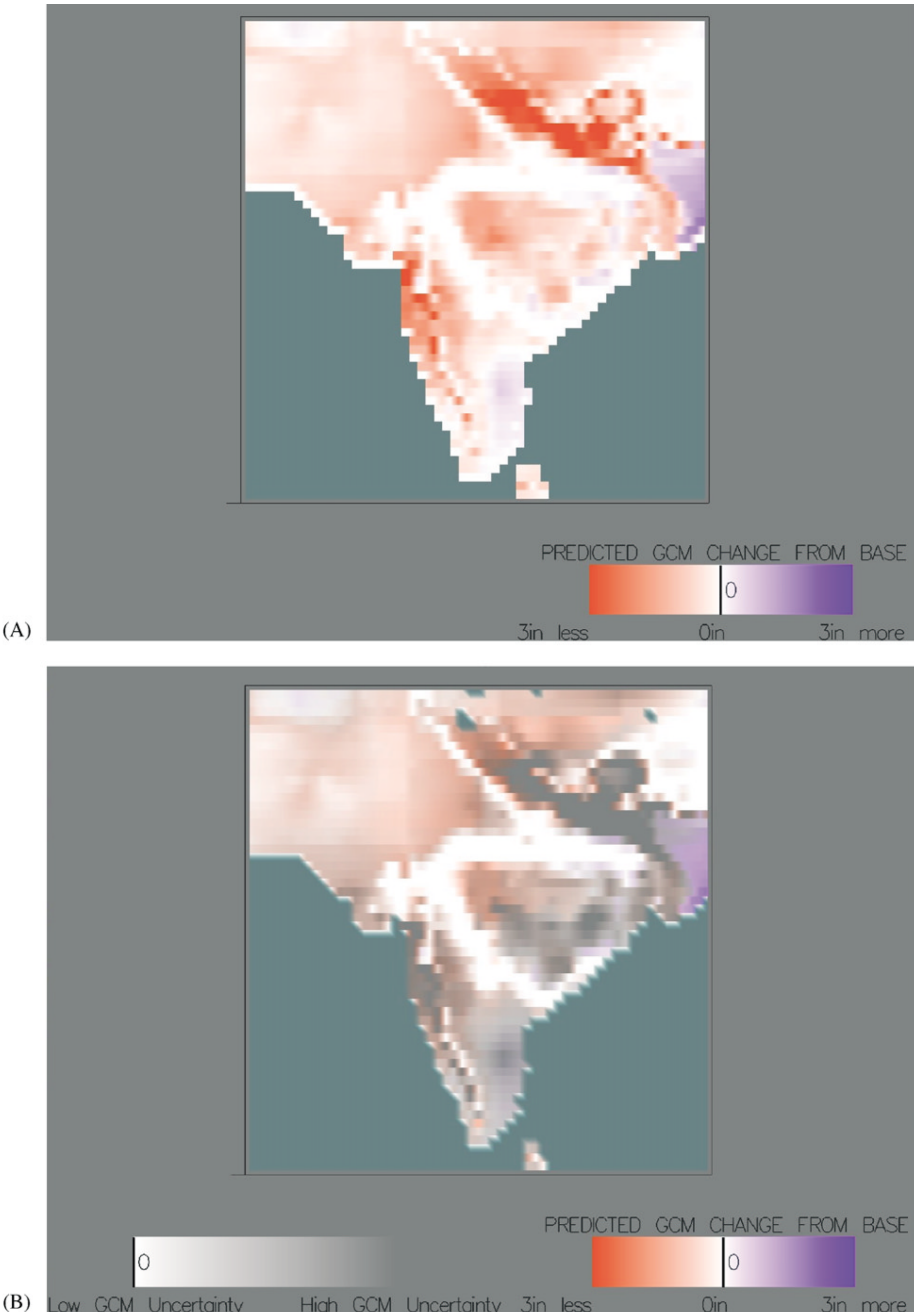


Fig. 8. (A) GCM predicted change shown through an orange/purple coloring scheme. (B) Uncertainty associated with making future water budget predictions shown through level of visibility by coloring the base condition surface to show GCM predicted change and then making it invisible to highlight areas of uncertainty.

method is similar to the Line Glyph method discussed in Section 2.5.2 as it adds vertical bars to a visualization to communicate uncertainty information. These glyphs, however, show both future water availability change and uncertainty information.

The technique places vertical bars at a surface grid cell center indicating whether the GCMs predict more or less water for each location and the amount of that change. To help highlight whether a change is positive or negative, the bars are colored orange and purple to

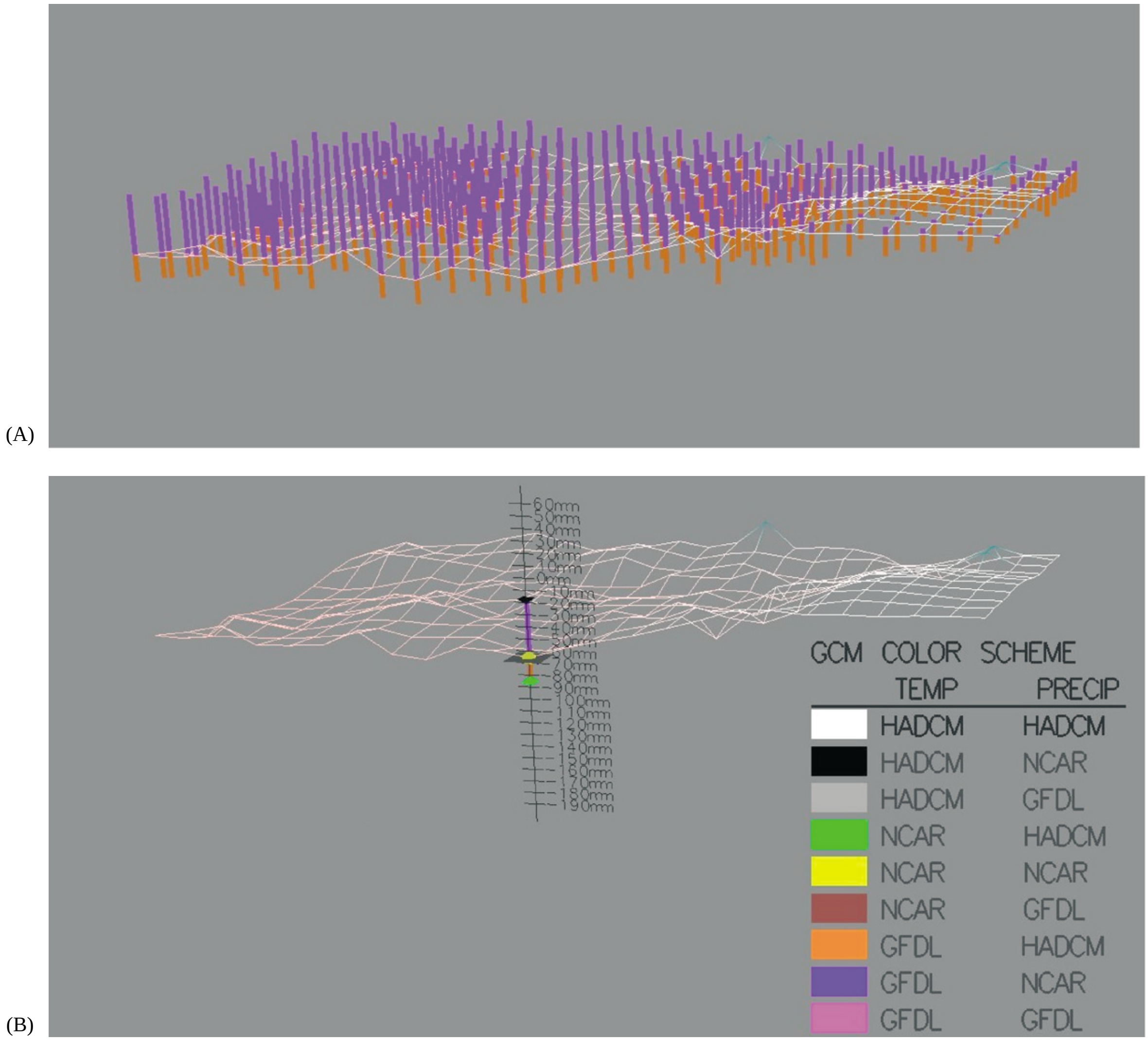


Fig. 9. GCM Glyphs depicting future water balance change predictions: (A) all glyphs are shown as orange/purple bars; (B) a single glyph shown with "pyramids" representing the min and max predictions with associated GCM indicated by the GCM Color Scheme legend.

indicate less and more water, respectively. Bars are drawn from the surface to the maximum and minimum heights predicted by the GCMs for each location. Orange bars are drawn below the surface and purple bars above.

Several GCMs can be selected simultaneously and we define uncertainty with respect to future water balance conditions as the difference between the predictions from each selected GCM. Therefore, there are multiple heights for which we could draw each bar (heights are determined from GCM runs). Because of this, it is possible to have a bar that is both purple and orange if a location is predicted to get more water in the future by one GCM, and less by another. Fig. 9A shows a surface with GCM Glyphs drawn to indicate the range of predicted change.

Users can click on the surface to select individual glyphs. This hides all other glyphs and draws a scale indicating the amount of predicted change. Fig. 9B shows a single glyph display. Also drawn are "pyramids" at the glyph's ends indicating the GCMs having the min and max predictions. The GCM Color Scheme legend in the bottom right corner of Fig. 9B indicates the GCM corresponding to each possible pyramid color.

This method contains the most information about GCM predictions and their uncertainty. Its primary drawback is the same as for the Line Glyph Scheme, the visual complexity of the scene can make it overwhelming for the user.

2.6. System capabilities

The system supports many interactive capabilities allowing for the exploration of visualizations. Using the mouse, visualizations can be rotated around the horizontal and vertical axis. As can be seen on the left of Fig. 10, a button toolbar contained within the surface visualization window supports options such as: zooming in to and out from a surface and orthogonal verses perspective views. The menus at the top of the window (seen in Fig. 10) allow for selection of the various visual presentation schemes discussed in the previous sections.

The button toolbar was added to the software at the request of several test subjects, who wanted a more traditional GUI feel for the application. We found that several users seemed hindered by not having buttons to push when they wanted some action performed or feature toggled. Other features controlled through the toolbar and suggested by users include: quick view buttons for common orientations (North, South, East, West, Top, Bottom), different units of measure (millimeters and inches), base map information and latitude longitude lines.

3. Evaluation

The evaluation of a decision support system and the methods for presenting uncertainty can be a difficult endeavor as they can be evaluated in both a quantitative and qualitative manner. Much of uncertainty presentation evaluation to date has been of the quantitative variety—testing users to determine how accurately particular methods convey uncertainty information [29,30]. Given our research objectives—the development of visualization techniques that allow users to explore and access uncertainty, and to depict multiple uncertainties—coupled with our goal to create uncertainty visualization tools useful for decision-makers attempting to understand model data and its reliability, we chose to evaluate this project qualitatively. Primarily, we were interested in determining the effectiveness of uncertainty presentation methods in the context of decision-making and if these methods could improve the decision-making process.

A current trend in the field of software engineering is *usability engineering*, which provides structured methods for achieving usability in user interface design during product development [5–7]. Usability engineering principles involve potential users early in the design process in an attempt to create an interface and features that are "useful" rather than a cryptic system understood by designers but not end users. We decided that the

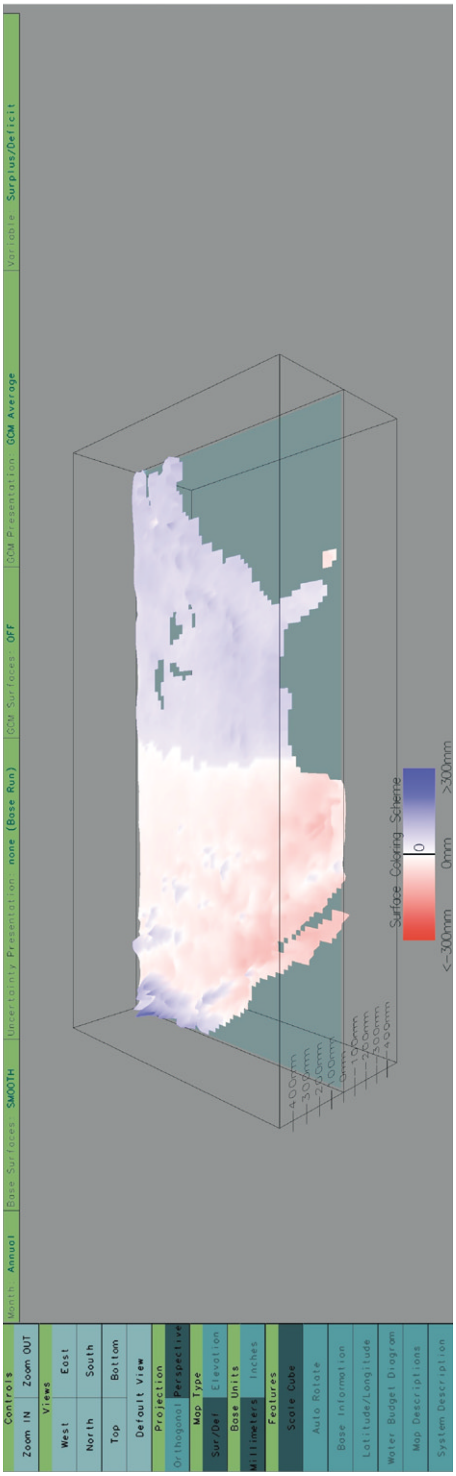


Fig. 10. Surface Visualization Window with Button toolbar on the left side and menus at the top.

iterative process of usability engineering—several cycles of user evaluation and appropriate software modification—was ideal for our project, as we did not know a priori what the important information would be nor how to utilize this information once it was determined.

In an effort to create useful and verifiable software, researchers have begun applying usability engineering principles to Virtual Environment (VE) applications [31]. These researchers are developing a structured, iterative methodology for user-centered design and evaluation of user interaction based on Nielsen's work [5]. Researchers at Penn State University have adapted this work for use in development and testing of geographic visualization applications for the purposes of determining their usefulness towards facilitating collaborative decision-making [32,33]. The environment used in the latter group's testing sessions allowed multiple users to both view and manipulate multivariate climatic data simultaneously, and to share knowledge and insight they gained from the visualization. The four stages employed by these researchers to design and evaluate software are: User Task Analysis, Expert Guidelines-Based Evaluation, Formative User-Centered Evaluation, and Summative Comparative Evaluation.

Our testing methodology was based on these works and involved testing sessions with three separate subject groups: Domain Experts, Usability Experts and Decision-Makers. The testing results for each group will be discussed in turn.

3.1. Domain experts

The purpose of the *domain expert* group was to fulfill the *User Task Analysis* stage. This first step of usability engineering is often very difficult, particularly when scientists are developing an application they think will be useful, but which involves new and untested ideas. If decision-makers (our end user group) had come to us with a description of what they needed in order to make better decisions, user task analysis would have been much easier. Interviews could have been performed to find out what the common tasks were in this type of decision-making situation. However, we were not approached with a specific need for better uncertainty visualization and exploration tools, so we had to conjecture as to what would be important. Given that our intended audience of potential users was likely to have little or no knowledge of the water budget model and its associated uncertainties, it would have been extremely difficult for them to specify what common or useful tasks might be when exploring this type of data. Furthermore, they probably would not have had the technical expertise to know what would have been useful and needed. We decided that the group most likely to know what information was pertinent in water balance models would be water balance experts (from here on we will call this group the domain experts).

Interviews were arranged with six local domain experts we hoped could offer suggestions for improvement to our system. This group included two civil engineers, two atmospheric scientists, and two geographers with significant experience in water balance models.

We began testing the system before development was completed. We assumed that our domain experts would not have enough background in uncertainty presentation methods to give good suggestions on how to present reliability information, so we created versions of the methods described in Section 2.5 before they arrived. We hoped that after studying our visualizations of uncertainty, they would be able to offer comments for improvement or offer opinions on alternative techniques.

In our domain expert testing sessions, we stepped the subjects through the main features of the system and asked them to comment on effectiveness and clarity issues. The sessions were interview style and informal. Because of the system's complexity, it was not possible to test users on every feature or every possible use.

We asked the domain experts to evaluate the software system as a whole and to compare and contrast uncertainty presentation methods for both base climate predictions and future climate predictions. When asked to compare the RGB Scheme to the Line Glyphs scheme for representing uncertainty in base climate prediction, half of the test subjects had no uniform preference. For the most part, this group found the RGB Scheme most useful when looking at a region as a whole, and the Line Glyph Scheme best for determining exact amounts of uncertainty and in greater detail, as we hypothesized. One user, who preferred the RGB Scheme to the Line Glyphs, thought the Glyphs were good for more detail as well, but that they were extremely hard to interpret. The user who preferred the Line Glyph Scheme simply preferred it for its greater amount of information. One subject thought both methods were complicated and unusable since they showed too much information. In hindsight, we showed the domain experts a very large region of the world (the whole country of India) in a very uncertain month (July) making the scenes, for the glyphs in particular, quite visually complex. Domain experts who independently chose to explore smaller regions later on in the testing session seemed to find the Line Glyphs much more manageable. This observation suggests that more work is needed to develop new uncertainty visualization schemes that would be better able to cope with large areas.

The domain experts also compared the Visibility Scheme to the GCM Glyph Scheme for visualizing uncertainty in future water budget predictions. At the time of the testing sessions, we had only devised the Visibility Scheme to work in the context of a base surface and a separate transparent future water budget prediction surface.

Two of the experts did not like either scheme for visualizing uncertainty in future water balances, stating that the methods were too confusing and depicted too much information. One subject said that both methods were equally good, and really served separate purposes: the GCM Glyphs give more specific information and the Visibility Scheme was good for getting an overall picture of the information in a region. The remaining experts found the GCM Glyphs to be most effective, citing that they give more information. Several in this group never seemed to fully understand the Visibility Scheme. Part of the problem may have been that we showed surfaces without much uncertainty making discrepancies in levels of visibility hard to see.

Many of the suggestions made by the domain experts for improvement to the system turned out to be usability issues. We had the domain experts "pilot" the application themselves, which in hindsight was not a good idea. Some of the experts who were less computer literate did not want to interact with the visualizations. This was problematic as most of the methods work best when users explore the data through the application's interactive capabilities. Those in the group who did not choose to interact with the visualizations seemed to develop an immediate bias to the application, as they were obviously uncomfortable using the mouse.

Several features were added to the system as a result of the domain expert testing sessions. These include:

- Orthogonal views of surfaces (producing flat map like displays);
- Options to add country/state boundaries and rivers to the surfaces;
- Options to add latitude and longitude grids over the surfaces;
- Interactive legends on every display;
- Reorganizing the interface to put the Region Selection Module at the bottom (this made the interface seem more intuitive to users);
- Banners describing the visualizations;
- Different Colored Meshes for Min/Max Surfaces;
- Name of the Transparency Scheme change to Visibility Scheme (some test subjects were confused by the terminology);
- Visibility Scheme applied the Orange/Purple coloring scheme for representing future change.

3.2. Usability experts

The second testing phase employed *usability experts*, those who were experts in software usability issues, but who had little or no knowledge about water balance issues. This testing group was in fulfillment of the Expert Guidelines-Based Evaluation stage [31]. The purpose of this group was to detect and offer suggestions for improvement of major usability violations found in the software's features.

The usability experts ran through a tutorial of system features while providing feedback as to their effectiveness. The number of features this group explored, however, was smaller than that of the domain experts (we wanted these subjects to have time for additional stages in the testing process that the domain experts did not do). In this vein, the tutorial was designed to give the experts general familiarity with the system and its features.

The usability experts were asked to comment on which method they preferred (between the RGB Scheme and Line Glyph method) for visualizing uncertainty associated with base climate prediction. This group was split down the middle in terms of method preference, with two choosing the RGB Scheme and two choosing Line Glyphs. Actually, most in the group seemed to have a hard time choosing between the two methods. The general consensus was that Line Glyphs provided more information but were harder to learn and interpret. One expert, however, commented that Line Glyphs would be better for colorblind people as they could be drawn in shades of gray instead of different colors. No modification of the RGB Scheme for colorblind users seems immediately apparent.

When asked to choose between the Visibility Scheme and GCM Glyphs for visualizing uncertainty in future water budget prediction, two experts preferred the Visibility Scheme, one liked both methods equally, and one did not seem to like either method. It should be noted that the Visibility Scheme tested by the usability experts was the Visibility Scheme applied to the surface employing the Orange/Purple coloring scheme showing predicted water budget change (shown in Fig. 8) and not what the domain experts viewed.

The two experts who preferred the Visibility Scheme chose it mainly because the glyphs were visually overwhelming and displayed too much information. The expert who did not like either method said that the numerous details made it difficult to focus on which was best. The remaining expert who thought the methods were about the same commented that the Visibility Scheme was good for an initial pass to determine where uncertainty was high, while the Glyphs provide details as to why.

The subjects were next asked to use the system to answer a few water budget questions in a region of their choosing. In this phase of the testing sessions, the goal was to determine how well the usability experts had learned the system and if they could correctly interpret the uncertainty presentations. Only one expert was able to answer all the questions without some reminders on

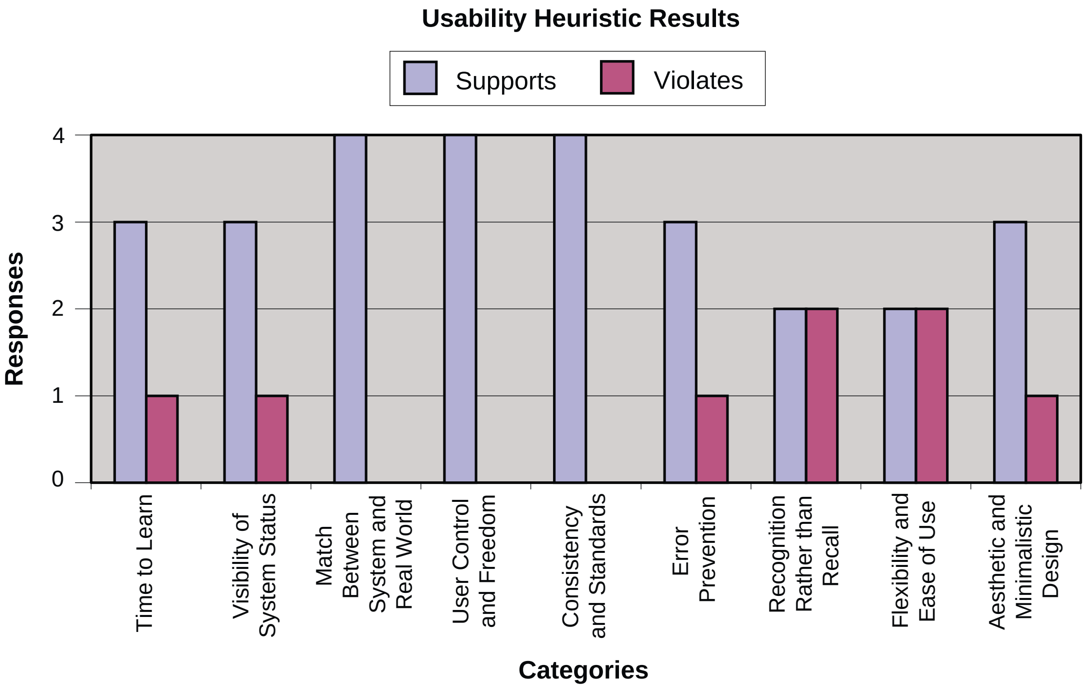


Fig. 11. Usability Heuristic results obtained from usability expert testing sessions.

how to use the system and its features. All of the subjects, however, correctly interpreted the displays once they managed to create appropriate ones. As was the case with the domain experts, this group seemed to like the glyphs much more when they looked at them in the smaller regions they had chosen during the task.

These test sessions concluded with the usability expert's comments on the heuristics described in [33] (the results are depicted in Fig. 11). The expert who said the application violated the *Time to Learn* principle commented that the software certainly was hard to learn, but users could manage. It seemed apparent that potential users would need more than the tutorial we offered the usability experts in order to become proficient with the software and uncertainty visualization techniques. The same expert said the application violated the Visibility of System Status heuristic citing that menus were confusing and the text needed to be easier to read. As was the case with the domain experts, many users had a hard time recalling what various visualizations were communicating and needed more reinforcement than what was provided through various map description banners we had implemented.

The expert who said the application violated the *Error Prevention* heuristic stated that we should provide a more extensive help system. According to two usability experts, the application violated the *Recognition Rather than Recall* heuristic. One expert wanted to see pop-up menus when users left clicked with the mouse (as in windows based operating systems). These menus could provide information such as uncertainty values and real world location without requiring the user to take their eyes off the display. This data is available from the system, but requires users to make other types of option selections.

The system also performed poorly with respect to *Flexibility and Ease of Use*, with two experts stating that it violated this principle. The consensus was that the program presented a lot of information that was hard to digest without considerable time spent. Two users also wanted to know if they could supply their own data sets, which currently is not supported by the system. The final category was Aesthetic and Minimalistic design, which all the experts thought the system supported except that one mentioned the display would become confusing if implemented on a smaller screen (i.e. a single workstation or PC).

Several features were added and modifications made to the system as a result of usability expert testing including:

- map legends moved closer to the surface;
- current selections in menus made bolder and more obvious;
- terminologies made more consistent in system messages;
- indicators added to surface displays highlighting the most recently selected area;

- map Description Banners made more clear;
- module colors in the Model Definition window were made less visually distracting.

3.3. Decision-makers

The final group of test subjects was decision-makers. This group included two employees of a state water management office, a state representative, a legislative aid to a United States representative, an employee of a state agricultural office and four scientists from an African university. When asked to explain their role in the decision-making process, one of the African scientists stated that they act as "consultants" to the policy makers on scientific issues. The remaining subjects all held positions with the government related in some way to water balance and/or environmental issues. Some of these test subjects were allowed to participate in pairs, with the rationale that many if not all decision-making problems typically involve multiple decision-makers and interest groups.

The purpose of this testing group was to provide formal evaluation of the application and uncertainty presentation methods in the context of decision-making. This took the place of the third stage of usability engineering guidelines, *Formative User-Centered Evaluation.* This step is actually an iterative process of evaluation and revision. We plan on implementing this iterative process further as we consider future directions for the software.

In contrast to the first two testing groups, the decision-makers were not taken through a tutorial of system features. Instead, we decided to "drive" the software and have the test subjects participate in a decision-making scenario with a water balance expert (the second author). We hypothesized that the system would be most effective when used as a tool for facilitating group decision-making as opposed to an application with a single decision-maker working alone and attempting to learn about a model and its uncertainty. These test sessions were very informal and the testing script was allowed to be "flexible" based on subject interest and response.

The tests began by asking subjects to describe the decision-making process they are involved in. For the most part, the members of this final testing group agreed on how decisions were made. Interestingly enough, the Africans described a very similar process for decision-making as the Americans. Typically, certain groups influence legislators or policy makers in an effort to get new legislation introduced. These groups can be other government agencies or lobbyist organizations. Often, it is scientists from these groups who explain information to policy makers and their aides from which the policy makers are then expected to form a decision. These types of information sharing sessions are often done collaboratively, with several scientists presenting to several policy makers. At the end of our testing sessions we asked subjects if the application they had seen could be useful in these collaborative settings between scientists and decision-makers. The application received very positive reviews with a unanimous vote that it could be helpful towards facilitating group decision-making.

During the testing sessions, decision-makers were shown visualizations of the current water supply in their home region of the world and then future climate predictions for this region. We developed a scenario were decision-makers considered the impacts of climate change on their region's ability to grow crops in the future. The economy of several regions was based on agriculture making this a particularly interesting scenario to some test subjects.

The uncertainty of these future conditions was then shown through the Visibility Scheme and GCM Glyph methods. The subjects were asked to choose which method they preferred for visualizing uncertainty in future climate prediction. Subjects with a strong scientific background tended to choose the Line Glyphs method, citing that it presented data in a clearer manner. Those test subjects with a less technical background preferred the Visibility Scheme because it was "simpler".

Some subjects were later shown how other uncertainties could be visualized in the model (uncertainty with respect to input data set choices) through the RGB Scheme and Line Glyphs methods. These subjects were asked to choose which method they preferred for visualizing uncertainty in base climate prediction. No subjects exclusively preferred the RGB Scheme, but half thought both methods were equally effective. One decision-maker commented that the advantage with glyphs is that you do not have to direct your eye towards a scale in order to understand the relative amounts of uncertainty as you do with RGB. Another subject, who actually preferred the glyphs, mentioned that they could get so complicated they would be hard to understand.

These sessions concluded by asking the decision-makers to comment on the application in terms of its usefulness towards decision-making and its presentation of uncertainty. The decision-makers all thought the application could be very helpful in the decision-making process, particularly in helping them to understand model data. One subject in particular really seemed to enjoy exploring possible scenarios in various places of the world in an effort to understand the global impacts of the problem presented. Another subject stated that decision-makers "want to see something quick that tells the whole picture" and that the application was able to do that.

The drawback with the application, in many subjects' minds however, was the presentation of uncertainty. Not that the presentations were unclear or ineffective, but that uncertainty information could be received negatively amongst decision-makers. One subject stated flatly that, "Politicians have trouble with uncertainty." Another test subject conveyed a story about how uncertainty in water quality data in their area was used as an argument for having fewer regulations and lowering standards. This person went on to say that the possible impacts of model predictions could be negated by the uncertainty. Decision-makers need to know if something is, or is not, a problem. They have little time for pondering confidence levels in data. A shortcoming of the way uncertainty is presented in our system is that it provides little advice on how to deal with it. This gives rise to a new area of research in uncertainty presentation, how to present uncertainty so that it is helpful to decision-makers.

A great deal of insight was gained into the visualization of uncertainty in the context of decision-making. Many decision-makers often want the big picture first, i.e. they want to know what the general patterns are and what can be expected. The intrinsic methods for presenting uncertainty seem best suited for this (the RGB Scheme for current climate conditions and the Visibility Scheme for future predictions). The drawback with approaches of this type is that extracting exact values of uncertainty at particular locations is difficult with these schemes. We developed interactive legends allowing exact values to be ascertained, but decision-makers often have little time for exploring data. They need to know results and if those results are reliable. One decision-maker suggested they would not want to see the uncertainty, only our best guess. Decisions cannot be made from uncertain data; it only leads decision-makers to discount the results. Unfortunately, not considering uncertainty may lead to inappropriate decisions. A potential collaborator, who viewed the application in its later stages, suggested incorporating a reasoning network of potential actions to problems presented by the visualizations. This would be a good tool, we believe, to help decision-makers deal with uncertainty and would require a great deal more research in logical reasoning and decision-making processes.

4. Future work

The Collaborative Visualization Room, where the application currently executes, is excellent for facilitating group decision-making. Unfortunately, not many decision-makers have access to this type of environment. A standard PC application could be developed, but the interactivity of the visualizations would suffer due to less processing power and local storage capacity. Another possibility is to develop a web deliverable application. The interface of the application, with Model Definition Window and Surface Display Window, could execute locally as a Java Applet with the model calculations executing on a multiprocessor machine. This avenue would be most effective for delivering the application to many more users.

We evaluated the application qualitatively, determining which methods were preferred in various situations, but the uncertainty presentation methods could be evaluated quantitatively as well. Tests could be conducted to determine, for various tasks, which methods outperform others in terms of task completion time and/or accuracy. For example, users could be shown surfaces and then asked to determine what areas have the greatest magnitudes of uncertainty, using both the intrinsic and extrinsic methods.

Another potential avenue of testing could be to compare the utility of the collaborative visualization laboratory's environment to that of a single PC or workstation. We hypothesize that the large screen and collaborative laboratory improve collaboration, but as yet, this remains a hypothesis.

Perhaps the solution to the uncertainty problem, however, is not in new or better ways to present uncertainty, but in finding ways to help decision-makers better cope with the uncertainty. Many subjects we spoke with seemed to think that decision-makers would prefer to have the system suggest possible ways of minimizing or dealing with uncertainty. In this regard, it might be best to study how decision-makers will act when presented with uncertainty and attempt to give them options for finding solutions for the problem before them.

5. Conclusion

The goal of the system described in this paper was to allow decision-makers and their staffs to explore the results of a water balance model along with associated uncertainties to better understand the potential impacts of public policy decisions that might be under consideration. A prototype application was developed that visualizes water budgets for terrestrial regions of the world. Reactions were unanimous among decision-makers that the interactive capability provided by the interface of this application was very helpful for exploring a problem.

We developed many techniques to show uncertainty in the predictions created by the application. Two methods were created for visualizing uncertainty in base predictions: the RGB Scheme—an intrinsic, tri-variate uncertainty presentation method; and Line Glyphs—an extrinsic method for communicating uncertainty information for an indefinite number of variables through vertical bars. Test subjects with scientific backgrounds tended to prefer the Line Glyphs, citing that they provide more information. However, the utility of glyphs decreases as the number of glyphs in a scene increases. Subjects with less experience in the sciences and models seemed to prefer the RGB Scheme.

Two analogous methods were described to show uncertainty in future climate prediction: the Visibility Scheme, an intrinsic method showing uncertainty through level of visibility, and GCM Glyphs, an extrinsic method showing uncertainty through vertical bars and pyramids. Again, the scientists among our test subjects preferred the GCM Glyphs, but the decision-makers in large part preferred the Visibility Scheme stating that it shows the magnitude of uncertainty in a simpler manner.

Reactions were unanimous among the decision-makers tested that the application could help in the decision-making process. A primary critique of our system and its uncertainty presentation methods was that decision-makers do not like to see uncertainty. In the legislative process, decision-makers typically must choose a side of an issue, so they need to know what choice is best. The interactive nature of the application was very helpful for explaining the data and its potential impacts, but the decision-makers wanted more advice on what could be done to reduce the uncertainty or how to make a decision with it present.

A potential research direction could be the incorporation of a visual abstract reasoning network that helps decision-makers choose a course of action when presented with uncertainty. Visualization is a tremendous tool for helping us to understand complex and abstract concepts as most humans can process information in a picture much easier than through data printed on a page. However, to maximize the effectiveness of visualization, uncertainty must be represented *and ways to deal with it must be provided.*

Acknowledgements

The work described here was performed in DesignLab, a multidisciplinary research laboratory at the University of Kansas funded in part by NSF Grant CDA-94-01021. The University of Kansas Center for Research Development Fund, grant 2300502, provided additional support. Special thanks should also go to the many test subjects who traveled to our lab to participate in the evaluation of this software.

References

[1] McCormick B, DeFanti T, Brown M, Visualization in scientific computing. Computer Graphics, 1987;21(6):1–14.

[2] Slocum T. Thematic cartography and visualization. Upper Saddle River, NJ: Prentice-Hall, 1999.

[3] Intergovernmental Panel on Climate Change website, 2001. http://www.ipcc.ch

[4] Gershon N, Visualization of an Imperfect World. IEEE Computer Graphics and Applications, 1998;8(4):43–5.

[5] Nielsen J. Usability engineering. San Diego, CA: Morgan Kaufmann, 1993.

[6] Mayhew D. The usability engineering lifecycle. San Diego, CA: Morgan Kaufmann, 1999.

[7] Constantine L, Lockwood L. Software for use. New York: Addison-Wesley, 2000.

[8] Thornthwaite C. An approach to a rational classification of climate. Geography Revue 1948;38(1):55–94.

[9] Mather J. The climatic water budget in environmental analysis. Lexington, MA: Lexington Books, 1978.

[10] Feddema J. Estimated impacts of soil degradation on the African water balance and climate. Climate Research 1998;10:127–41.

[11] Feddema J. Future African Water resources: interactions between soil degradation and global warming. Climate Change 1999;42:561–9.

[12] Legates D, Willmott CJ. Mean seasonal and spatial variability global surface air temperature. Theoretical and Applied Climatology 1990;41:11–21.

[13] Willmott C, Matsuura K. Smart interpolation of annually averaged air temperature in the united states. Journal of Applied Meteorology 1995;34:2577–86.

[14] New M, Hulme M, Jones P. Representing twentieth-century space-time climate variability. Part I: development of a 1961-90 mean monthly terrestrial climatology. Journal of Climate 1999;12(3):829–56.

[15] Willmott C, Matsuura K, Legates D. Global air temperature and precipitation: regridded monthly and annual climatologies, http://climate.geog.udel.edu/~climate/html_pages/README.lw.html

[16] Legates D, Willmott CJ. Mean seasonal and spatial variability in gauge-corrected, global precipitation. International Journal of Climatology 1990;10:111–27.

[17] Dunne K, Willmott C, Global distribution of plant-extractable water capacity of soil. International Journal of Climatology, 1996;16:841–59.

[18] Hamon W. Estimating potential evapotranspiration. Journal of the Hydraulics Division: Proceedings of the American Society of Civil Engineers, 1961;87(HY 3): 107–20.

[19] Climate Change, 1995. Impacts, adaptations and mitigation of climate change: scientific-technical analysis. Cambridge: Intergovernmental Panel on Climate Change (IPCC), Cambridge University Press, 1996.

[20] Robeson SM. A spatial resampling perspective on the depiction of global air temperature anomalies. Bulletin of the American Meteorological Society, 1995;76: 1179–83.

[21] OpenGL Architectural Review Board. OpenGL Programming Guide. Reading, MA: Addison-Wesley, 1993.

[22] OpenGL Multipipe SDK, http://www.sgi.com/software/multipipe/sdk/index.html

[23] Brewer C. Guidelines for selecting colors for diverging color schemes on maps. Cartographic Journal 1996; 33(2):79–86.

[24] MacEachren A, Visualizing uncertain information. Cartographic Perspective 1992;(13):10–19.

[25] Pang A, Wittenbrink C, Lodha S, Approaches to uncertainty visualization. The Visual Computer, 1997:13(8):370–90.

[26] Byron J, Spectral encoding of soil texture: a new visualization method. GIS/LIS Proceedings, Phoenix, AZ, 1994. p. 125–32.

[27] Brewer C. Color use guidelines for mapping and visualization. In: MacEachren A, Taylor D, editors. Visualization in modern cartography. Oxford: Pergamon, 1994. p. 123–47.

[28] Dorling D. Map design for census mapping. Cartographic Journal 1993;30(2):167–83.

[29] Evans B. Dynamic display of spatial data-reliability: does it benefit the user. Computers & Geosciences, 1997;23(4):409–22.

[30] Leitner M, Buttenfield B. Guidelines for the display of attribute certainty. Cartography and Geographic Information Systems, 2000;27(1):3–14.

[31] Gabbard J, Hix D, Swan J, User-centered design and evaluation of virtual environments. IEEE Computer Graphics and Applications 1999;19(6):51–9.

[32] Brewer I, MacEachren A, Abdo H, Gundrum J, Otto G, Collaborative geographic visualization: enabling shared understanding of environmental processes. IEEE Information Visualization Symposium, Salt Lake City, UT, 2000. p. 137–41.

[33] Brewer I. The Design, Implementation of Temporal and Attribute Query Tools for Geovisualization. The Association of American Geographers, Annual Meeting, New York, NY, 2001.