Data Visualization for Comprehensive Test Ban Treaty Monitoring

Randall W. Simons, Christopher J. Young, Tony L. Edwards
Sandia National Laboratories
work completed under DOE ST485D

ABSTRACT

The purpose of the Data Visualization Project at Sandia Labs is to prototype and evaluate new approaches to the presentation of data for CTBT monitoring applications. The great amount of data expected to be available, and the complex interrelationships in that data, make this a promising area for scientific data visualization techniques. We are developing a powerful and flexible prototyping environment with which to explore these possibilities. A user-friendly graphical user interface (GUI) should be an integral part of any data visualization tools developed. The GUI is necessary to select which data to visualize, and to modify and explore the displays that are the result of data visualization. Using our prototyping environment, we have produced data visualization displays of various kinds of data and have also experimented with different GUIs for controlling the visualization process. We present here an overview of that work, including promising results, lessons learned, and work in progress. To better understand what is needed, we have identified several data processing/analysis scenarios which we think will be important in CTBT monitoring. These scenarios help us identify what types of information we should display (together or in sequence), and help us focus on isolating the underlying goals. Each display we have produced is put in the context of one or more processing scenarios to help explain why and how it could be useful.

Keywords: data visualization, graphical user interface, GUI, Comprehensive Test Ban Treaty, CTBT, seismology

This work was supported by the United States Department of Energy under Contract DE-AC04-94AL85000.

1. Introduction

The purpose of the Data Visualization Project at Sandia Labs is to prototype and evaluate new approaches to the presentation of data for CTBT monitoring applications. We present here the status of that project, giving examples of the results achieved, including areas for further study and lessons learned. First, we briefly describe the hardware and software we are using to do this work, and some of what we have learned about the strengths and weaknesses of those tools. The next section details our results, including the various map projections used, displays based on data from a CSS database, and displays of other data. In the following section, we identify several data processing/analysis scenarios which we think will be important in CTBT monitoring. These scenarios help us identify what types of information we should display (together or in sequence), and help us focus on isolating the underlying goals. Each display we have produced is put in the context of one or more processing scenarios to help explain why and how it could be useful. Finally, we touch on expectations for future work.

2. Prototype Environment

We are developing prototypes and demonstrations, as opposed to production quality tools. This is a necessary first step, to explore a range of possibilities before recommending specific solutions. The display programs described here were developed on an SGI Onyx workstation and Sun Ultra 1 workstations with Creator3D frame buffers, using AVS Express software.

We have found some limitations with AVS/Express. Although it evolved from a mature product (AVS), Express is a major reorganization and reimplementation and it has not yet reached the stability of its predecessor. We think it works best as a prototyping tool, since it emphasizes flexibility over performance. By keeping functions modular, it simplifies prototyping with plug compatible modules, but combinations of modules cannot be optimized. For example, if two modules are invoked sequentially, each will have to loop through all the data separately, instead of merging the two loops as you would do if programming at a lower level. We continue to evaluate alternatives to cover the areas where Express is weak.

3. Visualization Results

We attempt to describe our work here, but the value of these programs can only be appreciated fully when used interactively on the workstation. Much is lost in the translation from the screen to black ink on white paper. First is the loss of color, which we use extensively to convey information. Second is the loss of dynamic behavior of the user interface, which allows the user to determine what he will look at when. Third is the loss of the ability to manipulate a 3D display to better understand the relative positioning of displayed objects in space.

3.1 Map Projections

Most of the data we are dealing with are inherently geographical, that is, they have locations on or in the earth. There are many ways of displaying maps of the earth. We have found that a few of these projections suffice for all of our display needs, but it is important to choose the right map for each purpose.

3.1.1 Cartesian Coordinates

This map simply uses longitude and latitude directly as (x, y) coordinates. The map shows the entire surface of the earth, but causes severe distortion near the poles. It also distorts shortest paths between points on the surface (great circles). On the plus side, it is easy to read longitude and latitude values off this type of map.

3.1.2 Spherical Earth

This is close to a true representation of the earth, which is nearly spherical. Thus it practically eliminates distortion and avoids "great circle" problems. It is only practical to view one side of the sphere at a time, however, thus requiring 3D rotation to view the back side.

3.1.3 Azimuthal Equidistant Projection

This projection places one location at the center of the map and arranges everything else relative to it. To understand this projection, think of selecting a point on a very elastic globe. Now prick a hole in globe exactly opposite the point you picked and stretch the hole until you have pulled the globe out flat. This projection is good for seeing distance and direction from the center point, since these are not distorted (everything else is). When we place an event at the center of this projection, we call it an event-centered display.

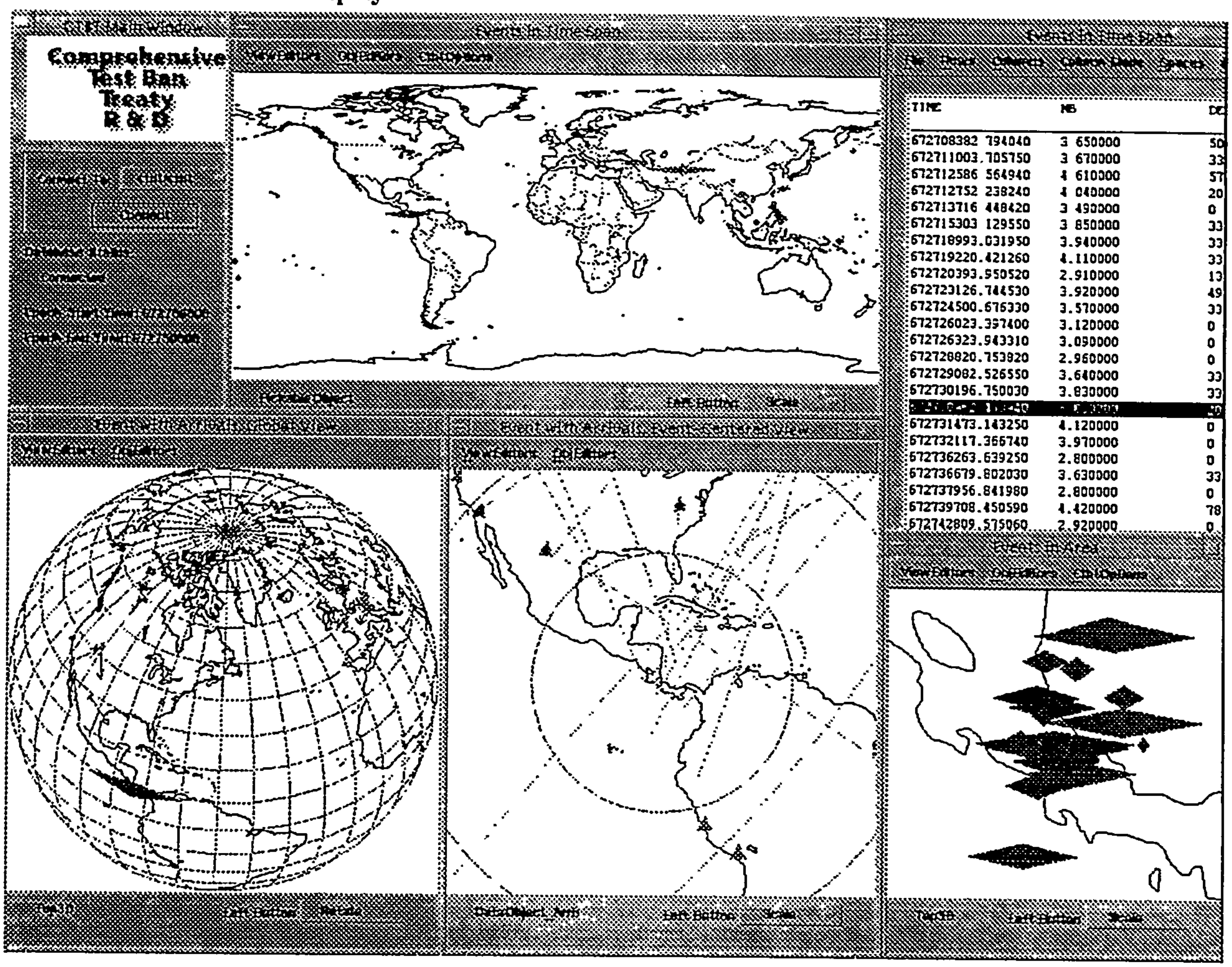
3.2 Visualizing Data from a CSS Database

Many of our visualization displays present data retrieved from a CSS database (Anderson, et. al. 1990). We have used GSETT II data (Group of Scientific Experts 1992), and some GSETT III reviewed event bulletins from the International Data Center.

3.2.1 Multilevel Displays - Events, Stations, Arrivals

This is a continuation of work reported last year by Young, Pavlakos, and Edwards (1995). The key idea throughout all the many iterations of this prototype is that there are multiple displays simultaneously visible and logically linked.

FIGURE 1. Multilevel Display



The user begins by connecting to the database he wants and selecting the time span he wants to view events for. The top level display shows a symbol at the location of each event in the user-selected time span, with the color and size of each symbol representing user-selected values (such as magnitude and depth). In addition, there is a corresponding text list of information about the events displayed. Selecting an event brings up a spherical earth display, an event-centered display, and a display showing other close events in the database. It also highlights the corresponding entry in the text list of events. Selecting an arrival from the event-centered display brings up another display showing azimuth and emergence angle in three dimensions.

We have produced a variant of the multilevel display which could be used for the quick-look application (see next major section). It includes only the top level and event-centered displays, but adds text lists for arrivals seen on the event-centered display.

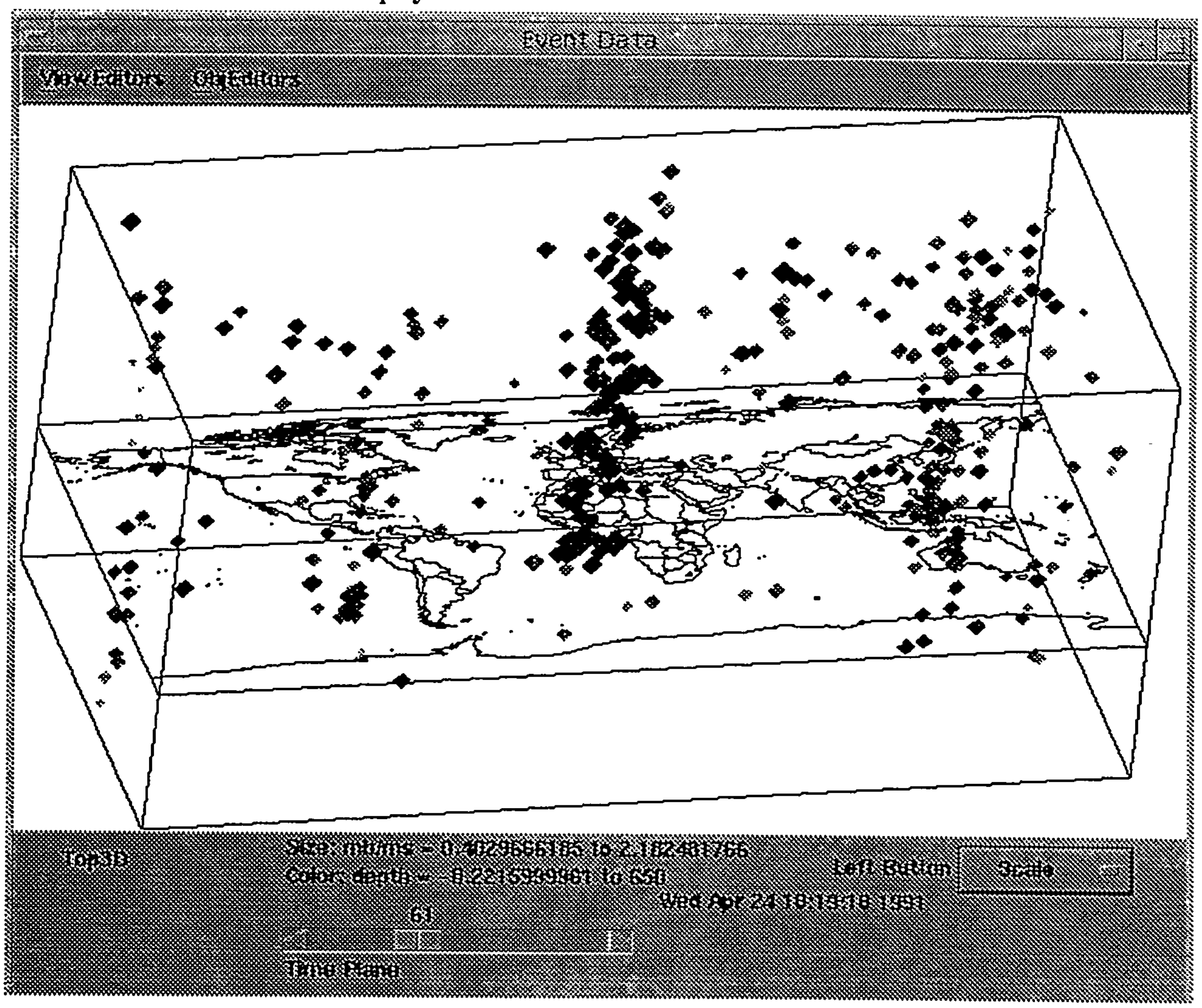
3.2.2 Animation of Seismic Waves

We wanted to display the propagation of seismic waves as they spread out from an event location. This turns out to be trivial when using an event-centered display. You simply draw a circle with its center at the event location and increment the scale of the circle in a loop. We have not yet identified an application for this utility.

3.2.3 Location/Time Display for Historical Trends

We noticed that when displaying events on a 2D map, it was hard to see them if there were many in the same area (e.g., a series of aftershocks). To make it easier to see all the events we added time as the third dimension. When you first see the display, you are looking straight down along the time axis, so it looks like the simple 2D version. If you rotate the display, however, you see that events are arranged at distances above the surface proportional to the time they occurred. This makes it possible to view events from longer time spans at once and still see each event distinctly. This turns out to be a good way to look for trends over time, such as sequences of events.

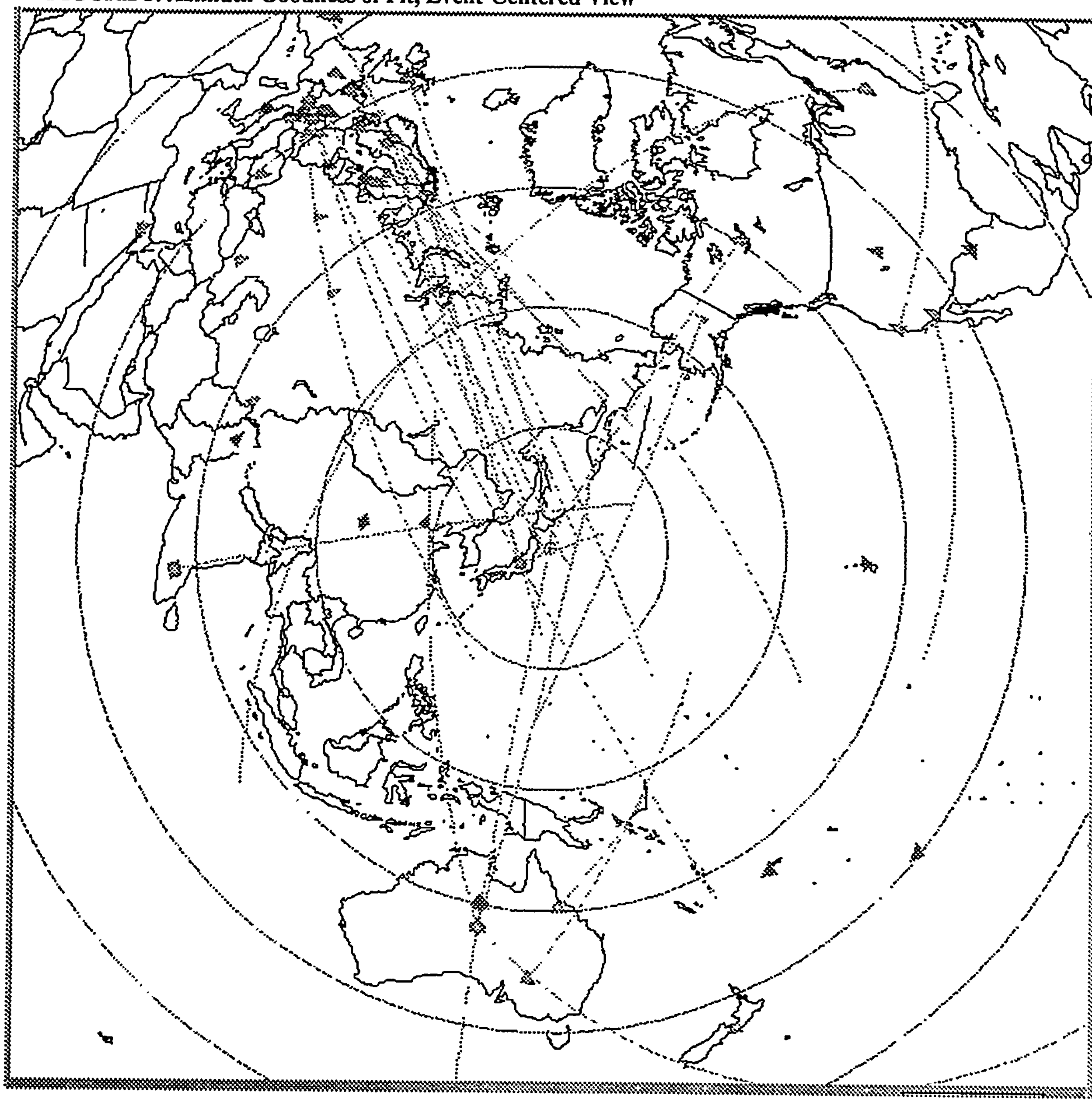
FIGURE 2. Location/Time Display



Viewing three dimensions on a two dimensional screen causes some difficulty in seeing how objects are arranged relative to each other. Using a perspective transformation does not seem helpful, perhaps because we are viewing disconnected simple objects instead of complex objects with edges and right angles. We use depth cueing, which dims more distant objects, to help the user's eye sort out what it sees. It also helps to put a wire frame bounding box around the data points, as this seems to give the eye another reference to relate to. Another technique is to drop a line from each object to the base plane. If the base plane contains the map, this helps in locating the object. However, with many objects, the lines can obscure more than they clarify. Instead, we have implemented a moving base plane, which snaps to the origin time of the selected event, so you can easily see exactly where and when that event occurred. 3D glasses might help significantly, but AVS Express does not support generation of stereo views. We are looking into other software to try this.

3.2.4 Goodness of Fit Measures

FIGURE 3. Azimuth Goodness of Fit, Event-Centered View



Anderson and Anderson (1995) proposed several goodness of fit measures for evaluating the strength of event association. We developed a program which computes these measures from data in a CSS database and displays the results. In addition to Anderson and Anderson's proposed maximum likelihood magnitude and azimuth goodness of fit measures, we also provide slowness and simplified magnitude goodness of fit measures. Since these measures are one per event, we use a Cartesian map projection with time added as the third dimension. We tried two variations, one which shows each measure in its own separate display, and one which combines selected measures, showing one as size and one as color. We also provide an event-centered display to show details of an event selected from the top level display. This is useful for understanding all the individual station contributions that are combined in a single goodness of fit measure. For example, one can easily see which stations are reporting azimuths with large residuals, and thus cause the azimuth goodness of fit to be large.

All these goodness of fit measures depend on standard deviations determined for the data. Although GSETT II contains such values, it appears they were somewhat arbitrarily assigned, and may be significantly off in many cases. This means goodness of fit measures derived using these erroneous standard deviations must also be considered unreliable.

3.2.5 Generalized Event and Arrival Data Display

The generalized event display allows the user to set the size and color of all event glyphs to any two database values or certain computed values (e.g., magnitude goodness of fit) for all events in a time span. The user can also select an event to show an event-centered display. The generalized arrival display is similar, but displays database values relating to arrivals, and positions the glyphs over the station where the arrival was recorded.

3.2.6 Event and Station Correlation Display

This display shows which stations saw a selected event, or which events were seen by a selected station. Trends visible in this data may help identify good and bad stations, or areas that are more or less visible from given stations.

3.3 Visualizing Other Data

In addition to CSS data, we have created visualization displays of other types of data. Some of these results support other CTBT research projects.

3.3.1 Time/Distance/Depth of Seismic Phases

This display plots arrivals of different phases in time, distance, and depth, color coded by phase. This is simply an extension of the traditional time-distance plot of seismic phases, using results from real events. The addition of depth as the third dimension, and the ability to rotate the 3D scene can lead to better understanding of the data.

3.3.2 Travel Time Surfaces and Cones

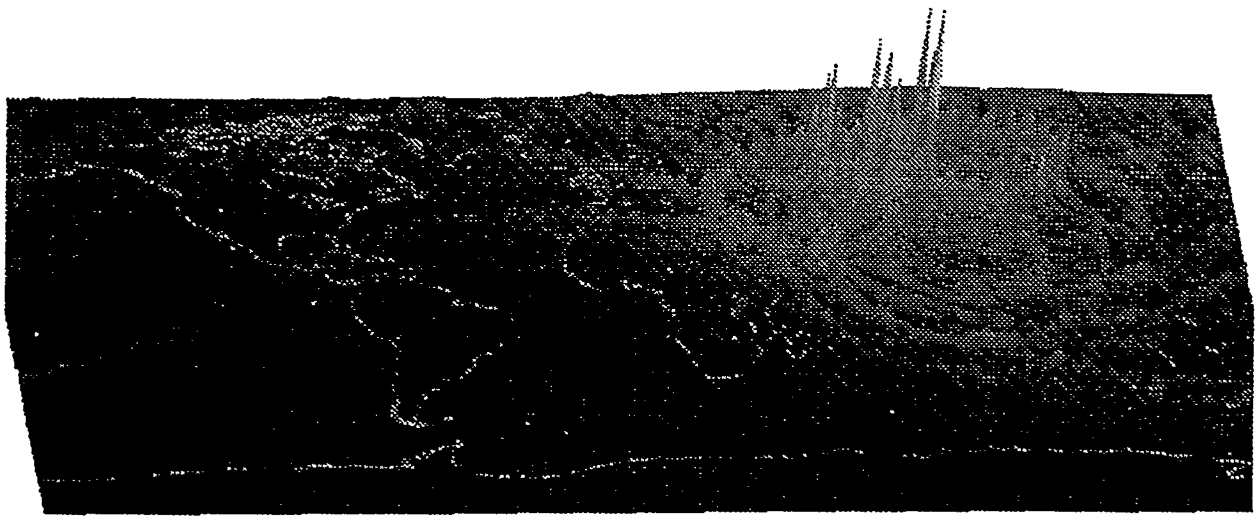
This display can plot travel time data (e.g., IASP91, Kennett and Engdahl 1991) for a single seismic phase as a surface in time, distance and depth. Color shading based on the slope of the curve (slowness) at each point makes it easy to see subtle distortions in the surface and better understand how travel times vary with distance and depth.

A single travel time curve for a single seismic phase at one depth can be rotated about the time axis to display as a cone in time and distance. When superimposed on an event-centered display, the surface is the locus of points in space and time where arrivals for that phase may be seen.

3.3.3 WCEDS Output

The Waveform Correlation Event Detection System (WCEDS) (Young, et. al. 1996) is a CTBT research project which attempts to associate and locate events using a novel method for combining full waveform seismic data. Data visualization can help us understand whether and how this new algorithm is working. We first tried plotting a colored glyph at each mesh computation point, and this gave reasonably good results. We then found that converting the data points to a 3D surface with sufficiently high resolution (200x200) produced a striking image that showed subtle details that were not visible in the colored glyphs, because it is easier to see small differences in a third spatial dimension than small differences in color.

FIGURE 4. WCEDS Output



3.3.4 Knowledge Base

The knowledge base is another CTBT research project (Keyser, et. al. 1995) that can benefit from data visualization. Much of the data in the knowledge base is tied to points on the earth's surface. A tessellation of these points creates a mesh which allows searching and interpolating the data (e.g., travel time corrections). Placing and tessellating these control points is an important part of the research, so it is useful to see a clear picture of their locations and the mesh connecting them. A flat map proved unsatisfactory, since mesh lines that crossed 180 degrees West longitude became long lines running the length of the map. A sphere works better, since it avoids introducing this artificial boundary. AVS/Express is especially well suited to this problem, since this type of mesh data is a good match to its internal data types.

FIGURE 5. Tessellation of Knowledge Base Data Points

4. Scenarios/Possible Applications

We have identified several data processing/analysis scenarios within CTBT monitoring where we feel advanced visualization techniques have potential for significant impacts. Existing seismic data processing systems, such as the International Data Center (IDC), National Data Center (NDC), and the AFTAC Distributed Subsurface Network (ADSN), already provide tools that meet some of these needs, but we believe that well-designed data visualization tools can significantly improve efficiency. For each scenario we give a description of the requirements, discuss some possible approaches to meeting these requirements, and then explain how our data visualization results fit with these approaches.

4.1 *Monitoring Data Flow, State of Health, and Alerts*

4.1.1 Requirements

Within a seismic data processing system, data flow must be monitored from acquisition in the field through automatic processing to review by analysts. This task is made difficult both by the large number of programs which operate on the data during this flow and by the (potentially) large number of stations involved for a global monitoring network. It may not be possible to show information about all stations at once. There is also a need to acquire contextual information about the data presented. Increasingly it looks like questions about processing will be directed towards specific events so a flow monitoring tool should be able to show how the status affects processing of specific events.

Another important task is to monitor the state of health (SOH) of the data acquisition (both in the field and the transfer to local machines). SOH monitoring has some of the same requirements as data flow monitoring. It, too, needs to handle a large number of stations and be able to display contextual information.

Finally, there is a need to quickly identify events that meet immediate reporting requirements as soon as the data becomes available. Much of the information that the operator needs to see regarding calling alerts is in a database, so well-designed database access is a must. A map-type display is a natural because the problem is so fundamentally geographical in nature.

The requirements to monitor data flow, state of health, and alerts are considered together because it has been suggested they could all be done by one person.

4.1.2 Approaches

Using our map displays, we could show symbols for origins and allow the operator to access processing information by interacting with the map(s). At some level more detailed status displays (such as automatic processing status) might be accessed, but perhaps at a lower level than the map. Selecting an origin could trigger an event-level display which would show the level of processing for both the network and the stations. All of the information needed should be available in the database. We could also include some higher-level processing information about the event such as depth and other discriminants which the operator might use to decide how an event should be dealt with (call an alert, send to an analyst, initiate special processing, etc.).

The different versions of the multilevel displays we have experimented with all have this two-level organization with the top level giving information about multiple events, and the next level giving details about a single event.

4.2 *Preliminary Event Assessment*

4.2.1 Requirements

The preliminary event assessment or "quick look" application requires that for every event in a prioritized list a quick summary of useful information will be presented to an analyst who will then decide how the event should be handled. It must be a quick display because the analyst may have to look at a display for every event (or at least all the important ones), and it must present critical information properly because it is essential that the analyst spend significant time only on the events of interest and deal with the others as quickly as possible.

The type of information presented could be goodness of fit (error ellipse, various residuals) which would allow the analyst to assess the quality of the event association, and discrimination information which would allow the user to assess the potential importance of the event for the mission. Both types of information are needed for the analyst to make the proper decision. Much of the information presented lends itself to graphical displays (especially maps) which could be assessed more quickly by an analyst than simple text displays. Some textual information is certainly useful, but it would be better to have it summoned via a pop-up from a selectable objects on a map display.

4.2.2 Approaches

The multilevel displays start at a global map of events and allow the user to access information about an event at an increasingly fine level of detail. For the quick look application, we could provide a mechanism to sequentially count through the events (e.g., in order of time of occurrence, but other sortings, such as by magnitude, are possible), changing the subdisplays as the selected event is displayed. This would allow the user not only to see how well an event has been put together but also how it fits with other events occurring before and after.

The goodness of fit displays have options for displaying several measures for multiple events.

4.3 *Knowledge Base Review/Editing*

4.3.1 *Requirements*

The knowledge base will be the repository of most if not all of the information needed by automatic data processing software and by analysts for CTBT monitoring. Because most of the data to be stored in the knowledge base is tied to a geographic location, it makes sense to use a tool which is well-suited to displaying geographic information to aid in the display, organization, and editing of the information.

One particular problem which we have identified is the review and editing of the knowledge base grid. This grid must be checked for adequate coverage, for proper interpolation, and referenced to contextual information such as geologic and political boundaries. This implies a fairly complex display which can allow the user to turn on and off several different display options and view in different ways for clarification.

4.3.2 *Approaches*

We currently have several map displays which could be used to show a grid, including various options to change color and viewing parameters. These displays do not allow editing of the objects displayed however, so modifications would have to be made if the tool was to be used to edit the knowledge base grid.

The WCEDS displays show a grid of output from the WCEDS program, and could be used to display other grids as well. The travel time surface and phase time/distance/depth displays show the power of visualization to help understand large data sets. The knowledge base control point tessellation display is a first attempt to display proposed knowledge base grids overlaying political boundaries.

4.4 *Event Reporting*

4.4.1 *Requirements*

In the CTBT monitoring scenario, a decision must be made about the course of action for events of interest. To make this decision, qualified personnel will want access to data from a variety of sources (including seismic) and to a great wealth of contextual information which will allow them to assess the probable significance of the data. Contextual information would include data on seismicity, geologic structure, mines, industrial complexes, population centers, etc. The requirements would be for detailed, high resolution information good down to as little as a few kilometers. The system must be able to fuse many different types of data together and should do this in a flexible manner because so little is known about how mixed data should be presented. An ability to relatively easily handle a new data source is a big plus.

4.4.2 *Approaches*

Some or all of the AVS/Express map display screens which we have developed could prove useful for event reporting, though their ability to access GIS data sets is not clear (i.e. they may not be able to access the full wealth of knowledge available to a GIS). Express software is flexible, allowing us to relatively easily test whatever options we deem worthwhile. Once concepts have been prototyped using Express, we may want to utilize some of the capabilities provided by a GIS, particularly the availability of lots of contextual information. It may be challenging to deal with the relative inflexibility of the GIS interface. Customization can be difficult and it is not unusual for a user to want a modification which simply cannot be made due to software constraints.

5. Future Work

There are other types of data we expect to be used in monitoring a CTBT, but lack of a sample database populated with complete examples has prevented us from extending to these possibilities. The knowledge base work being done by SNL will give us other types of data to visualize, so this is an area for future expansion. A better understanding of the event reporting scenario requirements will probably lead to new ideas for visualizing the results of seismic data processing and other outputs.

6. Acknowledgments

We would like to thank Kevin Anderson and Dale Anderson of PNNL, and Henry Swanger of SAIC, who visited us to see our displays on the screen, and gave us much insightful feedback. We also thank Ralph Keyser, SNL's CTBT Project Leader, for his ideas and support.

7. References

Anderson, J., W. E. Farrell, K. Garcia, J. Given, and H. Swanger (1990), CSS Version 3 Database: Schema Reference Manual, Science Applications International Corporation, Arlington, VA.

Anderson, K. K., and D. N. Anderson (1995), Statistical Issues in CTBT Verification: Seismic Monitoring Performance Measures, Technical Report PNL-10805, Pacific Northwest Laboratory, Richland, Washington.

Group of Scientific Experts (1992), Second Technical Test, Seismic Parameter and Waveform Data, Washington Experimental International Data Center, Arlington, VA.

Kennett, B. L. N. and E. R. Engdahl (1991), Traveltimes for Global Earthquake Location and Phase Identification, *Geophys. J. Int.* **105**, 429-465.

Keyser, R. G. and H. M. Armstrong (1995), Proposed Conceptual Requirements for the CTBT Knowledge Base, *Proceedings of the 17th Seismic Research Symposium on Monitoring a Comprehensive Test Ban Treaty*, Scottsdale, AZ.

Young, C. J., M. L. Harris, J. I. Beiriger, S. G. Moore, J. R. Trujillo, M. M. Withers, R. C. Aster (1996), The Waveform Correlation Event Detection System Project, Phase I: Issues in Prototyping Development and Testing, SAND Report in press.

Young, C. J., C. Pavlakos, and T. L. Edwards, "A Top-down Hierarchical Approach to the Display and Analysis of Seismic Data", *Proceedings of the 17th Seismic Research Symposium on Monitoring a Comprehensive Test Ban Treaty*, Scottsdale, AZ.