

Scalable Multi-variate Analytics of Seismic and Satellite-based Observational Data

Xiaoru Yuan, *Member, IEEE*, Xiao He, Hanqi Guo, Peihong Guo, Wesley Kendall, Jian Huang, *Member, IEEE*, and Yongxian Zhang



Fig. 1. Seismic and satellite-based observational data are visualized by our scalable multi-variate visualization system in a tiled display wall visualization environment with 32 LCD panels.

Abstract—

Over the past few years, large human populations around the world have been affected by an increase in significant seismic activities. Both for conducting basic scientific research and for setting critical government policies, it is crucial to be able to explore and understand seismic and geographical information obtained through all scientific instruments. In this work, we present a visual analytics system that enables explorative visualization of seismic data together with satellite-based observational data, and introduce a suite of visual analytical tools. Seismic and satellite data are integrated temporally and spatially. Users can select temporal and spatial ranges to zoom in on specific seismic events, as well as to inspect changes both during and after the events. Tools for designing high dimensional transfer functions have been developed to enable efficient and intuitive comprehension of the multi-modal data. Spreadsheet style comparisons are used for data drill-down as well as presentation. Comparisons between distinct seismic events are also provided for characterizing event-wise differences. Our system has been designed for scalability in terms of data size, complexity (i.e. number of modalities), and varying form factors of display environments.

Index Terms—Earth Science Visualization, Multivariate Visualization, Seismic Data, Scalable Visualization

1 INTRODUCTION

Earthquakes have tremendous impacts on the human society as one of the most devastating forces in nature. The energy released in a strong earthquake can wipe out entire cities and towns near its epicenter, causing permanent geological and geographical changes in the environment and the disastrous loss of human lives totalling in the hundreds of thousands. On May 12th, 2008, an earthquake of M_L 7.9 (Richter scale) hit the Sichuan province of China and resulted in at least 68,000 fatalities; On January 12th, 2010, an earthquake of M_L 7.0 struck Haiti with a death toll of over 230,000. As the earth en-

ters another era of increased seismic activity, there is an urgent global demand for a better understanding of earthquakes—a call to leave no stone unturned in exploring the novel sources of digital data from seismic events. This effort is critical not only for conducting scientific research on earthquake prediction and prevention, but also for evaluating the aftermath of earthquakes and preemptively designing better governmental policies in all areas.

Making earthquake predictions is currently very difficult, primarily due to highly sensitive, oftentimes unmeasurable, fine details deep underneath the ground with governing physical laws that are far from well-known. Fortunately, despite such great limits in human knowledge, occurrences of earthquakes are not entirely random.

Many instruments have been developed to monitor seismic events. An earthquake can be recorded by seismometers located at great distances from the quake's origin. Modern societies keep complete records of seismic events with information such as time of occurrence, location, magnitude and depth. It is fruitful to analyze these data spatially and temporally. For instance, it is generally known that seismic events show cluster patterns [28]. One sign that has been suggested to precede large mainshocks is a period of seismic quiescence [23, 33, 32], where seismicity rates appear reduced in and near the epicentral area but enhanced at farther distances. Other theories suggest patterns of seismic events like seismic gap and seismic belt.

More recently, scientists have begun to consider integrating obser-

- Xiaoru Yuan, He Xiao, Hanqi Guo and Peihong Guo are with Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University, E-mail: {xiaoru.yuan,xiaohex.pku,hanqi.guo,peihong.guo}@pku.edu.cn.
- Wesley Kendall and Jian Huang are with Department of Electrical Engineering & Computer Science, the University of Tennessee, E-mail: {kendall,huangj}@eecs.utk.edu.
- Yongxian Zhang is with China Earthquake Networks Center, E-mail: yxzhs@chinaearthquake.cn.

Manuscript received 31 March 2010; accepted 1 August 2010; posted online 24 October 2010; mailed on 16 October 2010.

For information on obtaining reprints of this article, please send email to: tcvg@computer.org.

vational satellite data with seismic event data to find out more reliable precursors to major earthquake strikes. Although not yet fully implemented, architectural studies about the Global Earthquake Satellite System (GESS [25]) have made great strides towards the ultimate goal of monitoring crustal deformation with high temporal and spatial resolutions. Similarly, an equatorial space mission (the ESPERIA project [29]) has just been proposed to observe near-Earth electromagnetic, plasma, and particle environments, as well as perturbations and instabilities in the ionosphere-magnetosphere transition region. It is also worth noting that, separate from seismic event prediction, population density and building codes also greatly affect an earthquake's outcome. For example, while the M_L 7.0 Haiti earthquake resulted in a death toll of 230,000, only four deaths were caused by a recent M_L 7.2 quake near Mexicali [7]. Understanding these differences requires a highly integrated visual analytics system that can handle data from heterogeneous sources.

In this work, we present a visual analytics system that enables explorative visualization (shown in Figure 1) of seismic data together with satellite-based observational data. Seismic and satellite data are integrated temporally and spatially. Users can select temporal and spatial ranges to zoom in on specific seismic events, as well as to inspect changes both during and after the events. Tools for designing high-dimensional transfer functions have been developed to enable efficient and intuitive comprehension of the multi-modal data. Spread-sheet style comparisons are used for data drill-down as well as presentation. Comparisons between distinct seismic events are also provided for characterizing event-wise differences.

Our system is scalable in terms of the number of dimensions in modality, the size of the data, and the display environment. We implement techniques for level-of-detail and out-of-core rendering to achieve interactivity with datasets over hundreds of gigabytes. We also implement a flexible data exploration tool for investigating multivariate as well as multi-modal data. Our system supports display environments ranging from a single monitor up to a large tiled display wall. In summary, the contribution of this work includes:

- Data Integration: we are able to integrate seismic catalog data with observational satellite data.
- Data Presentation: we are able to present multi-modal data through high-dimensional transfer functions.
- Scalable Visualization: we are able to handle complex large data in large-format display environments.

The remainder of this paper is organized as follows. We summarize related work in Section 2, overview our test datasets in Section 3 and overall system design in Section 4. We then describe functional components of our visualization system in Section 5. Sample usage of the system and the corresponding results are shown in Section 6. Finally, we go through important implementation details in Section 7, issues of discussion in Section 8, and conclusions in Section 9.

2 RELATED WORK

Many researchers have studied seismic visualization, covering both simulation and field-measured seismic data as well as earthquake catalog data.

For visualizing earthquake simulation data, 2D snapshot images [1, 22, 11] or videos [2, 5] have been applied to illustrate the spatio-temporal evolutionary behavior of the simulation results. The simulation results are often superimposed over a 2D map. More recent parallel methods implemented on supercomputers have tackled 3D earthquake simulation. In a related manner, parallel volume rendering methods have been developed to visualize 3D time-varying earthquake simulation data [1, 34, 6, 21]. In-situ visualization has also been applied to earthquake visualization, tightly coupling together simulation components and runtime visualization [30, 20]. In terms of innovative rendering systems for earthquake simulation data, Chopra et al. leveraged an immersive virtual environment to enhance collaboration between structural engineers, seismologists and computer scientists [4].

Field-measured seismic data have been another focus for visualization research. Wolfe et al. visualized the seismic volume data generated from ultrasound reflections [31]. Patel et al. developed the Seismic Analyzer for interpreting and illustrating 2D seismic data [26].

In their toolbox design, illustrative visualization, scale invariant visualization, and multi-attribute visualization worked together to provide both top-down and bottom-up methods for exploring 2D slices of seismic volumetric reflection data. Hsieh et al. visualized the field-measured seismic data of time-varying multidimensional earthquake accelerograph readings of the M_L 7.6 Chi-Chi earthquake in Taiwan [13, 12]. The temporal information of the wave-field volume was illustrated through volume rendering. Researchers, such as Komatitsch et al., have also compared simulation vs. field-collected seismic data to superimpose simulated seismic wave propagation data over field-measured waveform data [19].

Besides simulation and field-collected seismic data, the earthquake catalog data provides a third source of information. The catalogs contain detailed information about occurrences of earthquakes. Naka et al. visualized seismic-center distribution by showing the distributed points on a stereo display and applying a 3-D correlation graph for the time sequence analysis [24]. Dzwiniel et al. applied cluster analysis on the observed and synthetic seismic catalogs to discover the multi-resolution structure of earthquake patterns, with clustering results illustrated by 3D visualization [9]. Yuen et al. [36] used a web client-server service to cluster and visualize seismic catalog data in a grid environment. Very recently, scientists have begun to study the relationship between seismic events and other observational data. For example, earth observation satellite images can help rescue efforts by providing updated views of the landscape and infrastructure changes after a powerful earthquake [8]. However, we are not aware of any previous work with the aim of discovering intrinsic patterns in observational data related to the occurrences of major earthquakes.

3 DATA DESCRIPTION

Ideally, in this work we should study earthquake event datasets together with satellite data dedicated for studying seismic events like data from NASA's planned GESS system [25] or the newly proposed ESPERIA project [29]. Unfortunately neither project has been implemented yet. To proactively start our research, herein we chose to use NASA's MODIS (Moderate Resolution Imaging Spectroradiometer) satellite dataset as the source of the observational data. From applied system development perspective, MODIS and GESS datasets are of similar sizes, complexity and resolutions.

3.1 MODIS Data

NASA MODIS data comes from sensors on NASA's Earth Observing System (EOS) Aqua and Terra satellites. MODIS is able to view the entire Earth in two days and continuously collects data at up to 250-meter samplings. MODIS data are stored in 36 wavelength bands, and many variables (i.e. MODIS products) can be computed from combinations of wavelength bands. Among the many to choose from, we used 6 computed variables that capture properties of ground vegetation and soil, which are widely studied and most likely related to earthquake effects.

These computed variables are: Leaf Area Index (LAI), a measure of the extent of leaf canopies (the lowest value would result from barren land, while the highest value would result from a dense forest); Land Surface Temperature (LST), the temperature of land; Normalized Difference Vegetation Index (NDVI), a measure of the vegetation content of the area; fraction of absorbed photosynthetically active radiation (FPAR), a measure of the water content of the soil; Enhanced Vegetation Index (EVI), similar to NDVI but improves on quality by correcting for some distortions in the reflected light caused by the particles in the air as well as the ground cover below the vegetation; and Gross Primary Productivity (GPP), a measure of the compounds created by carbon dioxide. Our dataset is a 1-kilometer sampling spanning 222 timesteps in 16 day intervals from February 2000 to September 2009. It is approximately 60 GB.

3.2 Earthquake Catalog Data

Our earthquake catalog dataset is provided by China Earthquake Networks Center. Events are recorded on Richter magnitude scale (also known as the local magnitude (M_L) scale). Our dataset includes all

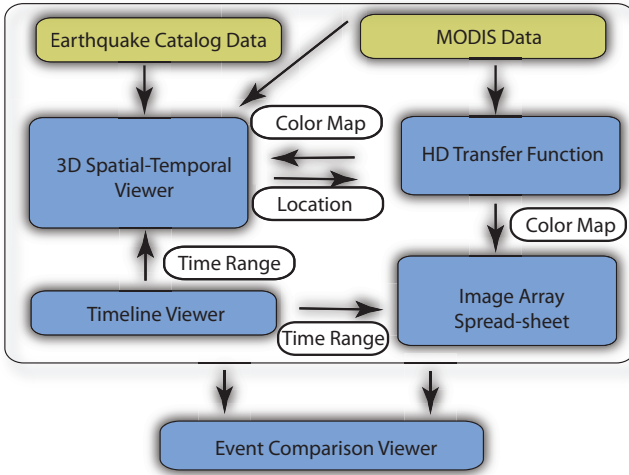


Fig. 2. Our visual analytics system with five major viewable components: 3D Spatial-temporal Viewer, Time-line Viewer, High Dimensional Transfer Function, Image Array Spread-sheet, and Event Comparison Viewer.

earthquakes in China and the bordering regions from January 1, 1970 to January 31, 2010. The data for all regions are complete with earthquakes over M_L 3.0, except in the area of Tibet where instrument placement is sparse and the corresponding catalog is only complete with earthquakes over M_L 4.0. In addition, for many major cities the dataset is M_L 2.0 complete. Altogether, 494,956 seismic events have been recorded.

In our system, the seismic catalog data can be spatially registered with the MODIS data in a common latitude and longitude coordinate system. MODIS is available in a much shorter time period and more sparsely allocated in the time axis when compared to the catalog data. Considering the time range, nearest neighboring sampling is applied to MODIS data.

4 OVERVIEW

The components of our system are illustrated in Figure 2 in the context of how visualizations of seismic catalog data and satellite data are integrated. The five major components are described as follows.

- **3D Spatial-temporal Viewer:** In a 3D space, satellite images are plotted on the ground. The catalog event data are rendered as 3D points, where the z-axis is time and animated. Location selection is enabled and linked with selections in other displays.
- **Time-line Viewer:** From the total 30-year time span, a focus+context-enabled time-line viewer is used to select a short period of time to explore.
- **High Dimensional Transfer Function:** Color maps for multiple MODIS variables are controlled through parallel coordinates and multidimensional scaling (MDS) methods.
- **Image Array Spreadsheet:** MODIS images of multiple dimensions or dimension combinations that are selected in time and space can be displayed as an image array.
- **Event Comparison Viewer:** Multiple seismic events with different space and time coordinates can be visually compared.

The five different views are linked dynamically. For example, users can specify a time range in the time-line viewer and the corresponding changes are displayed in the 3D viewer window where the earthquake catalog data and the satellite images are also updated. When the user applies a new transfer function to the satellite image(s), the color-map changes are simultaneously applied to the image array spreadsheet and the 3D viewer windows.

5 SYSTEM DETAILS

The interface for our scalable multi-variate analytics system is shown in Figure 3. The right part shows the 3D spatial-temporal viewer of seismic events, along with the associated observational data and the

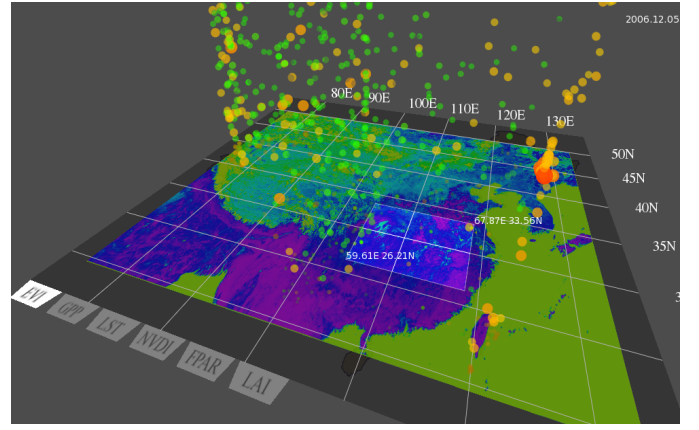


Fig. 4. 3D Spatial-temporal Viewer. Seismic catalog dataset is displayed as 3D points, together with satellite images in a common temporal-geographical coordinates system.

time-line control. The top left of the figure shows the image panel array, and bottom left is the interface for the multi-dimensional transfer function. The top center depicts the event comparison viewer for two selected seismic events.

5.1 3D Spatial-temporal Viewer

In this viewer, seismic catalog data are displayed together with satellite images in a common temporal-geographical coordinate system as shown in Figure 4. The seismic catalog data are displayed as points in 3D space above a ground plane, and each point represents a seismic event recorded in the catalog. The 3D location of the point is determined by the latitude and longitude coordinates on the ground plane (i.e. the X and Y dimensions) and time (i.e. the Z dimension).

We implemented animation by shifting the z-axis such that seismic events appear as though they are “falling from the sky” as time progresses. This option effectively facilitates data overview by users. In the animation, the colors of the points of seismic events are composited onto the ground plane after falling from space. The magnitude of the earthquake is used to dual-code point size and color. Satellite images are plotted and overlaid on the ground plane. Since the MODIS data contains multiple variables, the system allows the selection of the variable modality by clicking the tag along the border of the map as shown in Figure 4. Location selection is enabled in this window and the information of the selection made is also used to update other display windows, which are discussed in the following subsections.

5.2 Time Line Viewer

Supported by modern technology, researchers have been able to collect comprehensive records of seismic event around the world over long periods of time. In this work, we dealt with the over 30 years of seismic catalog data of China. Users typically want to inspect earthquakes in a particular region in a limited time span. To facilitate such studies, a time line viewer window, employing context+focus techniques, has been designed to set the observation time window.

As shown in Figure 5, the bottom part of the window is the full time range covering all of the seismic data available. A histogram of the earthquake occurrence is plotted over time. The small rectangular window indicates a sub-range of time for detailed view, which is shown in the top portion of the window. Each individual earthquake is plotted along the timeline. Events with large magnitudes are plotted above those with smaller ones. The vertical lines indicate which satellite images are available since the MODIS data are sampled on a 16-day interval. In this part, another rectangle widget can be tuned by the user, which indicates the time range of the window upon further examination. Three time lines are specified: beginning time, end time, and current time. For example, in the image arrange window, all satellite images included in the specified range will be displayed and compared.

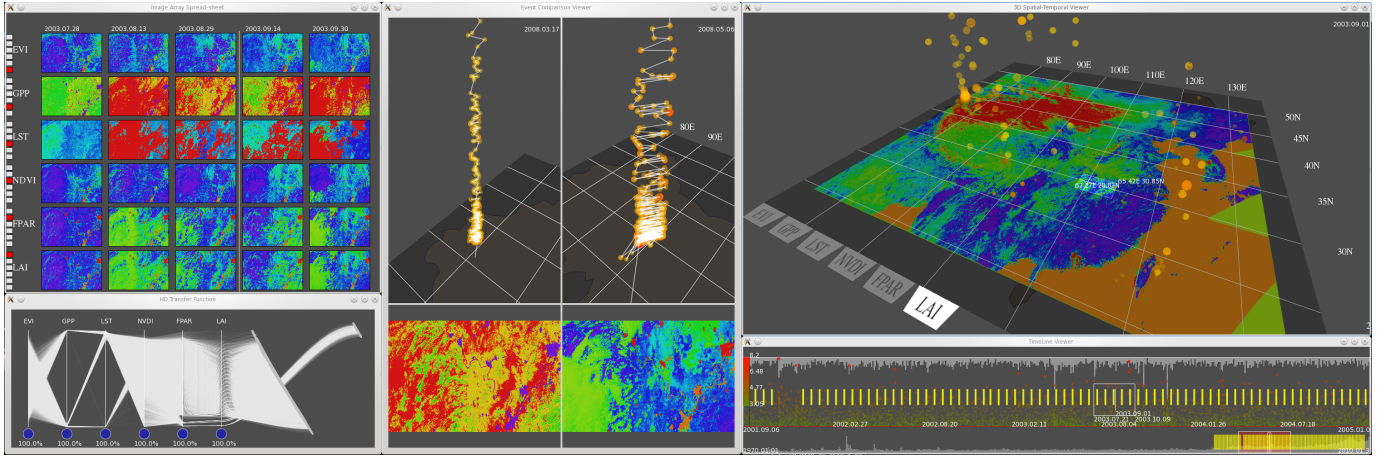


Fig. 3. Interface components of our system. Right-top, 3D view of seismic events and associated observational data; right-bottom, time-line control; left-top, image panel array; left-bottom, interface for multi-dimensional transfer function; top-center, comparison of two seismic events.

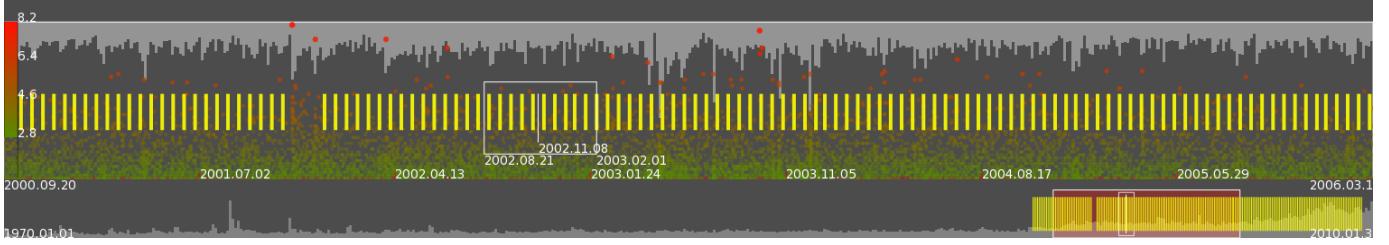


Fig. 5. Time-line control for selecting region of interest. Histogram and scatterplot of seismic events are shown.

In the 3D viewer window, the MODIS data of the current time is displayed on the ground plane; catalog data within the beginning time and the end time are displayed. The points between the beginning time and the current time are composited on the ground plane; the rest are above the ground plane. Animation can be achieved by dynamically changing the current time. After the selection of the time window, our system enables the user to compare the satellite sampled images in other views.

5.3 High Dimensional Transfer Function

Since the satellites can take the ground plane images with many different channels, it is necessary to develop a strategy to enable the user to view the multi-modality images conveniently. In our system, a high dimensional transfer function has been designed to color map multiple images into an integrated view, taking advantage of high dimensional data visualization techniques on parallel coordinates and dimension reduction [35].

Parallel coordinates, introduced by Inselberg and Dimsdale [15, 16], represents an N-dimensional data tuple as one polyline crossing parallel axes. For a large multidimensional dataset, parallel coordinates can turn the tuples into a compact two-dimensional visual representation. In Figure 6, the left side of the the transfer function design interface is a parallel coordinate plot. Pixel values of each modality are mapped to a corresponding axis. Such techniques of using parallel coordinates to generate color maps for volume rendering have been exploited [27].

Although parallel coordinates can show the data distribution of all dimensions clearly, it is not very efficient in showing the correlation or clustering over more than 3 dimensions simultaneously. In this work, we augment the high dimensional transfer function design with one additional plot of MDS. Data are also projected to the 2D plane with multi-dimensional scaling. The difference between each data item resembles their distance in the original high dimensional space. As show in Figure 6(a), the MODIS images of the 6 variables at a specified time can be projected and clustered into different groups. It is very intuitive for the scientists to select a particular group of pixels and apply desirable colors. After the selection in MDS region, further modification can also be applied in the parallel coordinate plot. In Figure 6, the right side of the transfer function interface shows clusters selected and

refined via the MDS interface.

More specifically, the process of color mapping multivariate images via the MDS interface has three major steps:

1. Pick up sample points from input images, usually user selected ROI (region of interest);
2. Project the sample points from multi-dimensional feature space to a 2D plane;
3. Estimate the parameters of a Gaussian transfer function with the brushed sample points as input.

The sample points are picked up randomly from the input images with a heuristic sampling rate of 5%, which is sufficient to preserve important data features based on our experiments. These sampled points are then projected to a 2D plane with pivot MDS [3]. We adopted pivot MDS for the projection mainly because it produces very nice projection results with relatively low computation complexity. Besides, pivot MDS is also well suited for data sets with large amount of homogeneous items, which is exactly the case in our problem.

The pivot MDS algorithm can be summarized as follows:

- Randomly pick up k pivot item from input data set;
- Construct the double-centered dissimilarity matrix $C(c_{ij})$ between pivot items and all input items, whose elements are defined as

$$c_{ij} = -0.5(\delta_{ij}^2 - \frac{1}{n} \sum_{r=1}^n \delta_{rj}^2 - \frac{1}{k} \sum_{s=1}^k \delta_{is}^2 + \frac{1}{nk} \sum_{r=1}^n \sum_{s=1}^k \delta_{rs}^2) \quad (1)$$

where k is the number of pivot points, and n is the total number of data points. δ_{ij} stands for the dissimilarity of the i th item of input data with the j th pivot item, which is the Euclidian distance between the two items by default [3];

- Calculate the eigenvalues and eigenvectors of the matrix $C^T C$;
- Pick up the largest d eigenvectors $\{v_i\}$, the low dimension embedding is achieved by

$$x_i = C v_i, i \in \{0, 1, 2, \dots, d\} \quad (2)$$

Users can perform free selection on projected sample points with a lasso tool. The selected sample points are considered as interesting features. A multivariate Gaussian transfer function is then constructed for each interested feature: for a set of selected sample points $P\{p_i\}$, $i = 1, 2, \dots, n$, assume their distribution in the feature space

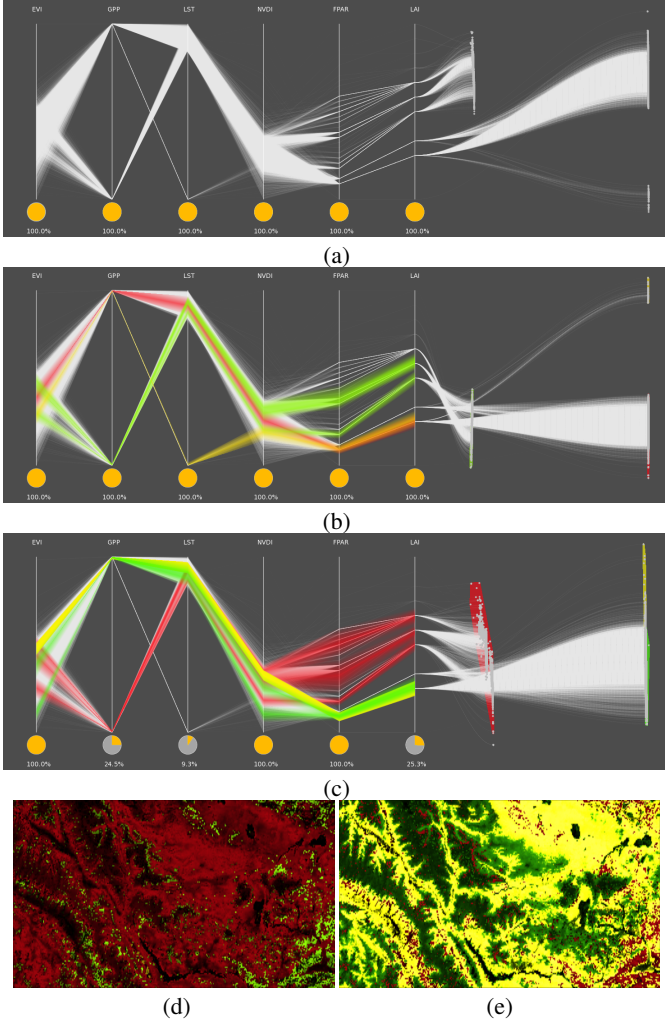


Fig. 6. High dimensional transfer function. (a) The transfer function design interface integrating parallel coordinates and MDS plots, 6 variables are considered; (b) One color mapping by the user; (c) Weighting of selected variables have been changed. (d) and (e) the color mapped multivariate satellite image using transfer function specified in (b) and (c) respectively.

satisfies the Gaussian Mixture Model (GMM), i.e., the probability of appearance of the points is governed by the following function:

$$P(\mathbf{p}) = \sum_{i=0}^n \omega_i \exp(((\mathbf{p} - \mu_i)^T \sigma_i^{-1} (\mathbf{p} - \mu_i))) \quad (3)$$

The Gaussian transfer function is then constructed as follows: the chromatic components of the transfer function is specified by user, and the opacity component of the multivariate Gaussian transfer function is written as [18]:

$$GTF(\mathbf{v}) = \sum_{k=0}^{n'} \alpha_k P_k(\mathbf{v}) \quad (4)$$

where \mathbf{v} is the input data vector, α_k is the maximum opacity value for scaling each Gaussian bolb. The final color and the opacity values are:

$$\alpha = \sum \alpha_i, C = \frac{\sum \alpha_i C_i}{\sum \alpha_i}, \quad (5)$$

where C_i is the color of each Gaussian transfer function. Gaussian transfer functions are applied to the images using GPU acceleration technology. The images of different dimensions and the parameters of the Gaussian transfer functions are stored as textures in the video memory, and then the per-pixel fragment shader generates the result images in real time for the image array.

Figure 6(b) shows a coloring scheme for the interface. Note our system also allows a user specified weight for each dimension, by allowing the user to click and change the weight wheel below each axis. A clustering with different weights for several dimensions is shown in Figure 6(c).

By user specifying different weighting on different dimensions, we could achieve user configurable clustering through our transfer function interface. In Figure 6(d), the yellow color indicate the region with low LST, low LAI and low FPAR values. While after re-weighting, we are able to select a clustering with high values of LST with low LAI and low FPAR values, resulting in different color mapping highlighting different features in Figure 6(e). We are currently collaborating with domain scientists to obtain their feedback.

5.4 Image Array

In our visualization system, one window is created and dedicated to satellite image visualization and comparison. In the image array window shown in Figure 7, color mapped observational satellite data are aligned for easy comparison. Each row in the image array corresponds to one observational dimension or a combination of dimensions. The color map function is supplied from the transfer function design interface. In each row, images are placed according to their time stamp, while the time range range is specified in the time-line interface as discussed previously. Each individual image can be set as the focus and enlarged for detailed inspection. Easy side by side comparison can be achieved by enlarging selected images.

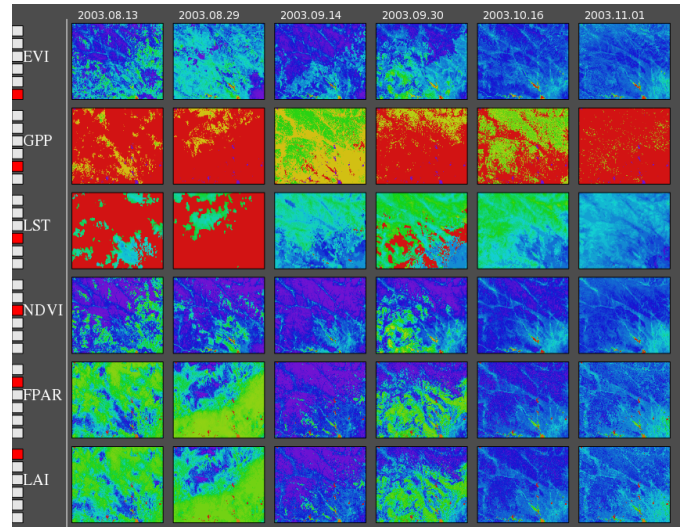


Fig. 7. Image Array Window. Color mapped satellite observational data are aligned for easy comparison. Images in the same row belong to the same observational dimension or the same dimensional combination and aligned according to the observation time from the left to the right.

5.5 Comparison

In the event comparison window, the user can specify two events with different spatial and temporal ranges. The parameters are defined in the time-line window for the time range and in the 3D spatial-temporal viewer window for the spatial range. The user can first navigate the earthquake catalog histogram, overview in the time line window, and identify a specific seismic event to be visualized. Then, the user brushes on the ground plane in the 3D spatial-temporal viewer window to circle a geographic area. Only earthquakes within the area will be displayed. The user can take the same approach to select the second event. The selected events are then shown side by side in the comparison window as illustrated in Figure 11. Earthquake shocks are linked with line segments according to the time occurrence. The earthquake plot is aligned in in two images so that their time axis is in parallel. The user can adjust an optimal viewing setting for better understanding and comparing the events.

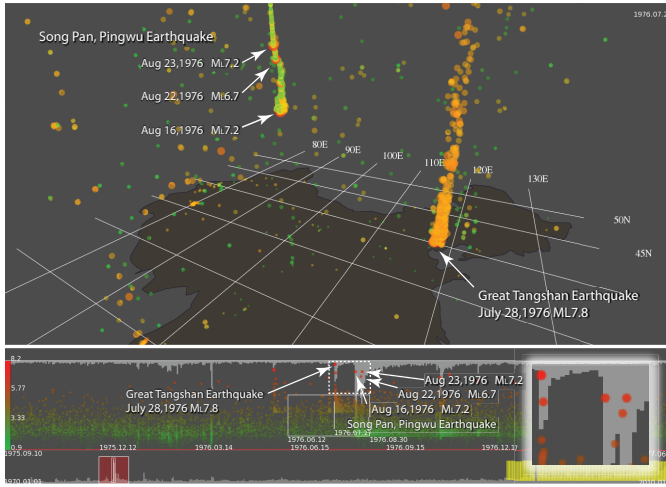


Fig. 8. Spatial-temporal exploration of Great TangShan Earthquake and SongPan, Pingwu Earthquake, two major seismic event series occurred in China in 1976.

6 SAMPLE USE

In this section, we discuss the results of visualizing the seismic catalog data in China from 1970 until Jan. 2010, together with MODIS satellite data spanning from 2000 to 2009 with our system.

6.1 Spatial-Temporal Exploration on Seismic Catalog Data

The first seismic event we explore with our tools is the Great TangShan Earthquake which is believed to be the largest earthquake in recent a few centuries by death toll. The outbreak of Tangshan Earthquake can be easily identified as an outstanding peak in the overview histogram bar of the time-line view in our visualization system as shown in Figure 8. After placing the time focus near the August, 1976, a few peaks appear in the histogram in the zoom-out view. The 3D view, as shown as the top plot of Figure 8 reveals there were two major sequences. The first that occurred is the TangShan Earthquake, which consists many large seismic strikes. The second sequence, located in SiChuan province of China, can be identified as the Song Pan, Pingwu Earthquake. Although it has large peaks (higher number of earthquake shocks) in the histogram, other than the three major strikes, the majority of the earthquake shocks have much lower magnitudes and appear in green in the display as indicated by the magnitude color map. From the enlarged part of the time line shown, the major hits and the different characteristics of these two events are clearly illustrated: In TangShan, the second largest hit, believed to contribute greatly to the increase of the death toll, closely following the first hits; while in the later earthquake, three major hits can be identified and other aftershocks are relatively small in magnitudes.

From the overview histogram bar, a trend of increasing over the seismic events recorded can be observed. This is due to the advancement in detection technologies and an increase in the number of record stations, both of which contribute to better detection of small magnitude events. As illustrated in the middle snapshot of Figure 9, it is difficult to identify the large Sichun Earthquake that occurred on May 12, 2008; by increasing the magnitude threshold, the event is clearly discernable in the bottom snapshot of Figure 9.

As illustrated in above two cases, our visualization tool, with only the time line view and 3D viewer, can provide a very convenient way of spatial-temporal exploration of seismic events and enable close inspection of their spatial-temporal characteristics. The spatial-temporal distribution is an important tool for understanding the overall story of an earthquake. As illustrated the earthquake hits with magnitudes within a certain range are linked. The distribution along time is clearly displayed. It is not only useful for researchers working on earthquake understanding and prediction, but also an easy to use tool for presentation and educational purpose.

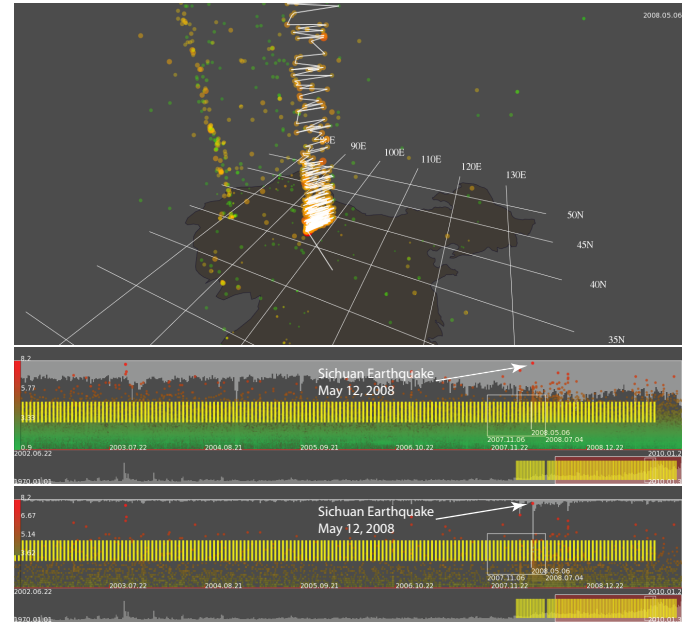


Fig. 9. Spatial-temporal exploration of the Sichuan Earthquake that occurred in China in May 2008.

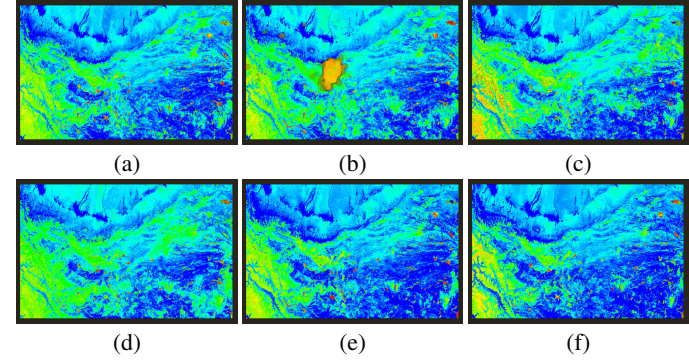


Fig. 10. Regional MODIS data variation before and after a M_L 7.1 earthquake in Hetian, China in 2008. In the first row, images are satellite data obtained on (a) March 5th, (b) March 21st, (c) April 6th, 2008, respectively. In (b), earthquake event dot are overlaid; (d)-(f) are Satellite data obtained on exactly one year before the earthquake as (a)-(c).

6.2 Correlation Between Seismic Event and Satellite Observational Data

The MODIS data has been studied near a region where an M_L 7.1 earthquake in Hetian, Xinjiang Uygur Autonomous Region of China. The NDVI data before and after the earthquake shows evidence of traceable changes around the earthquake region as shown in Figure 10. As indicated in Figure 10(c), after the earthquake there are feature variations west of the seismic center as illustrated in Figure 10(b). The data obtained one year prior is shown in the second row for comparison. Although more specific reasons are under further investigation, one hypotheses is that the earthquake changed the regional vegetation coverage.

In our visualization, the user can specific a time range to explore in the time line viewer, corresponding images of each dimension are colored and shown in the image array viewer. The system also provides functionality that allows the user to specify a range and observe the average value over the time range.

6.3 Seismic Event Pattern Comparison

Our visualization enables quick comparison and evaluation of seismic events. The example shown in Figure 11 demonstrate the different foreshock activity behaviors between Haichen earthquake, which occurred in 1975 and was able to be successfully predicted, and Tang-

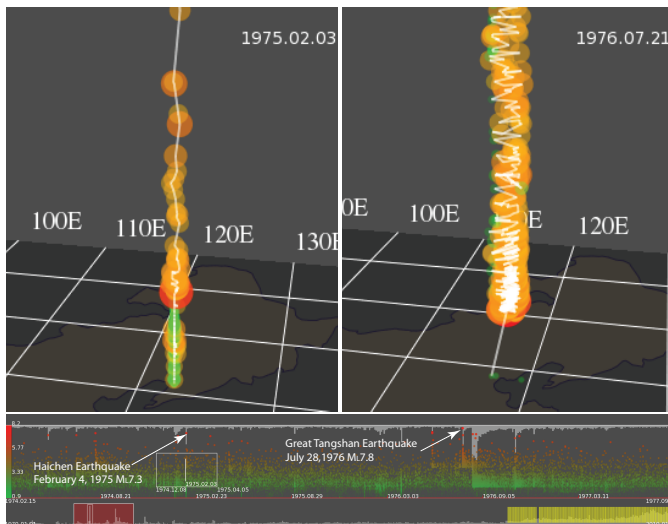


Fig. 11. Seismic event pattern comparison between Tangshan earthquake (Right) and Haichen earthquake (Left). It is obvious from the visualization that two earthquakes have very different patterns in foreshock activities.

shan earthquake, which occurred in 1976 and no prediction was made. For the Haichen earthquake, there are a series of foreshocks as illustrated in green dots in the left image. Very differently, there are no foreshocks at all in the Tanshan event. The side by side comparison of these two earthquakes can clearly demonstrate such differences. With our visualization tool, it is very intuitive to explore the difference or similarity between earthquakes.

7 IMPLEMENTATION DETAILS

Our PC is a Dell T3400 workstation having a 2.66GHz CPU, 4GB Memory, and Nvidia Geforce GTX 275 graphics card with 896MB Memory. Our display wall consists of 32 tiled 32-inch LCD panels.

The multi-dimensional transfer function is computed on GPU using Cg fragment shaders. The input of the shader program include the satellite images treated as texture images, and the parametric Gaussian transfer functions. Since the space for uniform variables of the shader is limited, we composite the parameters to textures as the input instead of uniform variables, in order to ensure the scalability of the program. The colors, centers and covariance matrices of the Gaussian components are stored as 1D, 2D and 3D textures. The output image of the shader is bound to a framebuffer object, and the system can directly obtain the output textures for next operations and user interaction. The GPU-accelerated image transfer is about 10 times faster than CPU works.

The parallel version of our system is based on CentOS Linux 5.4 64-bit and NVidia X servers. Xdmx and Xinerama combine the individual X servers on each node into a uniform large one display environment. Chromium [14] streams the OpenGL command from the control node to the rendering nodes. Since Chromium only supports OpenGL 1.5, and it has not been updated for a long time, many modern OpenGL extensions cannot be used when migrate the system to the tiled display wall. For example, `GL_EXT_framebuffer_object` extension is not supported in Chromium system, but we can avoid using framebuffer objects by rendering the contents to the buffers and then reading pixels from them, we can then add application-only extensions to Chromium system, which can be only used on the application node. Our shading program of multi-dimensional image transfer used 'for' statement, which requires a relatively modern shader profile, does not work on the tiled display. We run the shading program on the control node instead. In the future we plan to improve the capability of the parallel visualization environment and make it compatible to more advanced graphics features.

We implemented level-of-detail (LoD) and out-of-core rendering to achieve interactivity with datasets on 100GB level. The system can

handle multi-variate data as well as multi-modal data with scalability, and can also support display environments as small as a single monitor and as large as a 32-panel tiled display wall.

MODIS satellite images are preprocessed into a quad-tree data structure to facilitate quick display in the 3D view. The displaying LoD is determined by the zooming factor of earth surface. Images of the specified regions are extracted from the tree structure to allow further processing; by doing so we carefully restrict runtime processing and computation on the satellite images to only the small areas selected by the user. As the images could be larger than the memory available, we also implemented out-of-core rendering. In addition, as done in many applied visualization system, we implement caching and prefetching. To ensure smooth interaction, while images of the current time step is always loaded in-core, a few neighboring time steps are also prefetched from hard drive.

8 DISCUSSION

Our visualization system can adapt to various display environments. The users can explore the data with a workstation or just a laptop. Several users can also work in a collaborative visualization environment with large display setup. As shown in Figure 1, our system has been implemented on a tiled display wall system with over 130 million pixels. Multiple seismic scientists can work side by side to explore the data. Large display enables simultaneous viewing of the different aspects of the data. Our visualization consists of many individual components, which can be customized into different combinations on the fly, according to special needs by domain users. For example, the time line viewer alone can be a very effective tool for exploring earthquake catalog data. It can also work with the 3D spatial-temporal viewer to provide geographical information of seismic events with user interaction enabled.

The image array component, together with high dimensional transfer function, also provides scalability in dimensionality. At the current stage, our test dataset only includes earthquake catalog data and satellite data that approximately covers China and surrounding areas. Our seismic scientists actually strong urged that we add the capability to handle datasets that cover the entire globe. Especially more recently, seismologists are very interested in exploring the relationship between certain types of low frequency electromagnetic activity (e.g. the ionospheric electron density) and the seismically most active zones on the Earth. We will closely work with seismologists and include more data sources in our visualization system to fully explore the potentials of the tool we have developed. The key limiting factor in our prototype system is unfortunately with I/O and other system bandwidth related bottlenecks. We plan to leverage cutting edge parallel visualization techniques [17] that are closely integrated with parallel I/O.

Currently, our prototype system is deployed in a academic laboratory environment and have been evaluated by the a few scientists working in the field of seismology, including a coauthor of this paper. The feedback is in general very positive and encouraging. At current stage, although visualization techniques have been widely applied in many geologically relevant domains, and used for understanding earthquake simulation data, there are few existing tools for exploring abnormal seismic events and their correlation with other observational earth surface data. Due to the extremely complex nature of the seismic activities, it is also premature to assess our system's potential of helping to make correct seismic event predictions. However, our system has revealed spatial-temporal correlations that are unique and interesting. These findings provide a stepping stone for seismic scientists to further hypothesize and explore for potential relationships between foreshock patterns and mainshock (Figures 9, 11), as well as between earthquake events and the aftermath impacts (Figure 10).

9 CONCLUSION

This paper proposes a visual analytics framework to explore seismic event catalog data together with satellite imagery data. Our prototype system allows users to visually zoom in on interested time periods and spatial regions around seismic events, while revealing temporal changes in satellite imagery data in nearby areas. We have designed

special transfer functions to handle multi-modal observational data, which can be further analyzed through a spread-sheet style interface. Our system enables data drill-down. Users can simultaneously explore and understand seismic and geographical information.

In current usage, the users can explore correlations between the heterogeneous MODIS and seismic catalog data sets by specifying time and spatial ranges. It works for the users who already have knowledge of the possible auspicious locations. A thorough blind search is not possible in currently working mode. We plan to introduce more powerful analysis tools, for example regular expression based query driven visualization [10], in this system to help the seismic scientists.

In the future, we also plan to integrate more data modalities into the system. For example, to add an ability to visualize changes in terrain surface due to a seismic event. Population density and building codes can be integrated for policy evaluations based on studies about correlation between earthquakes and fatality consequences. We will also test our system using larger computing systems for the purpose of handling datasets with even more details. We also plan to perform user studies with seismic scientists to quantitatively evaluate the effectiveness of our system.

ACKNOWLEDGMENTS

We wish to thank anonymous reviewers for their suggestive comments. Authors Yuan, He, Guo and Guo were supported in part by National Natural Science Foundation of China Project No. 60903062, Beijing Natural Science Foundation Project No. 4092021, 973 Program Project No. 2009CB320903, Chinese Ministry of Education Key Project No. 109001. The project was also supported by the startup funding awarded to Dr. Xiaoru Yuan by the National 985 Project Phase II at Peking University. Work by authors Kendall and Huang was supported in part by USA National Science Foundation grants CNS-0437508 and OCI-0906324. Zhang was supported in part by Key Projects by China MOST grant No. 2008BAC35B05 and No. 2010DFB20190.

REFERENCES

- [1] V. Akcelik, J. Bielak, G. Biros, I. Epanomeritakis, A. Fernandez, O. Ghattas, E. J. Kim, J. Lopez, D. O'Hallaron, T. Tu, and J. Urbanic. High resolution forward and inverse earthquake modeling on terascale computers. In *SC'03: Proc. of ACM/IEEE Supercomputing*, page 52, 2003.
- [2] M. Antolik, C. Stidham, D. Dreger, S. Larsen, A. Lomax, and B. Romanowicz. 2-D and 3-D models of broadband wave propagation in the San Francisco Bay region and North Coast ranges. *Seismological Research Letters*, 68(328), 1997.
- [3] U. Brandes and C. Pich. Eigensolver methods for progressive multidimensional scaling of large data. In *Graph Drawing*, pages 42–53, 2006.
- [4] P. Chopra, J. Meyer, and A. Fernandez. Immersive volume visualization of seismic simulations: a case study of techniques invented and lessons learned. In *Proc. of IEEE Visualization*, pages 497–500, 2002.
- [5] A. Chourasia, S. Cutchin, and B. Aagaard. Visualizing the ground motions of the 1906 San Francisco Earthquake. *Computers and Geosciences*, 34(12):1798–1805, 2008.
- [6] A. Chourasia, S. Cutchin, Y. Cui, R. W. Moore, K. Olsen, S. M. Day, J. B. Minster, P. Maechling, and T. H. Jordan. Visual insights into high-resolution earthquake simulations. *IEEE Comput. Graph. Appl.*, 27(5):28–34, 2007.
- [7] CSMonitor. <http://www.csmonitor.com/World/Americas/2010/0405/Why-Mexicali-earthquake-damage-is-nothing-compared-to-Haiti>, 2010.
- [8] S. Daily. Science news, first satellite map of haiti earthquake. <http://www.sciencedaily.com/releases/2010/01/100114143323.htm>.
- [9] W. Dzwiniel, D. A. Yuen, K. Boryczko, Y. Ben-Zion, S. Yoshioka, and T. Ito. Nonlinear multidimensional scaling and visualization of earthquake clusters over space, time and feature space. *Nonlinear Processes in Geophysics*, 12:117–128, 2005.
- [10] M. Glatter, J. Huang, S. Ahern, J. Daniel, and A. Lu. Visualizing temporal patterns in large multivariate data using textual pattern matching. *IEEE Trans. on Visualization & Computer Graphics*, 14(6):1467–1474, 2008.
- [11] K. Hirahara, N. Kato, T. Miyatake, T. Hori, M. Hyodo, J. Inn, N. Mitsui, Y. Wada, T. Miyamura, Y. Nakama, T. Kanai, and M. Iizuka. Simulation of earthquake generation process in a complex system of faults. Technical report, Earth Simulator Center, 2004.
- [12] T.-J. Hsieh. Understanding earthquakes with advanced visualization. *ACM SIGGRAPH Comput. Graph.*, 44(1):1–13, 2010.
- [13] T.-J. Hsieh, C.-K. Chen, and K.-L. Ma. Visualizing field-measured seismic data. In *Proc. of IEEE Pacific Visualization Symp.*, pages 65–72, March 2010.
- [14] G. Humphreys, M. Houston, R. Ng, R. Frank, S. Ahern, P. D. Kirchner, and J. T. Klosowski. Chromium: a stream-processing framework for interactive rendering on clusters. *ACM Trans. Graph.*, 21(3):693–702, 2002.
- [15] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.
- [16] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proc. of IEEE Visualization*, pages 361–378, 1990.
- [17] W. Kendall, M. Glatter, J. Huang, T. Peterka, R. Latham, and R. Ross. Terascale data organization for discovering multivariate climatic trends. In *SC'09: Proc. of ACM/IEEE Supercomputing*, page 1, 2009.
- [18] J. Kniss, S. Premoze, M. Ikits, A. Lefohn, C. Hansen, and E. Praun. Gaussian transfer functions for multi-field volume visualization. In *Proceedings of IEEE Visualization 2003*, pages 65–73, 2003.
- [19] D. Komatitsch, S. Tsuboi, C. Ji, and J. Tromp. A 14.6 billion degrees of freedom, 5 teraflops, 2.5 terabyte earthquake simulation on the earth simulator. In *SC'03: Proc. of ACM/IEEE Supercomputing*, page 4, 2003.
- [20] K.-L. Ma. In situ visualization at extreme scale: challenges and opportunities. *IEEE Comput. Graph. Appl.*, 29(6):14–19, 2009.
- [21] K.-L. Ma, A. Stompel, J. Bielak, O. Ghattas, and E. J. Kim. Visualizing very large-scale earthquake simulations. In *SC'03: Proc. of ACM/IEEE Supercomputing*, page 48, 2003.
- [22] G. P. Mavroudis and A. S. Papageorgiou. Simulation of long-period near-field ground motion for the great 1906 San Francisco earthquake. *Seismological Research Letters*, 72(2):227, 2001.
- [23] K. Mogi. Some features of recent seismic activity in and near japan (2), activity before and after great earthquakes. *Bull. Earthq. Res. Inst.*, pages 395–417, 1969.
- [24] T. Naka, M. Yamada, M. Endo, S. Miyazaki, and J. Hasegawa. Visualization of seismic-center distribution data for earthquake prediction. In *Proc. of Nicograph Intl, Session V Simulation*, 2006.
- [25] NASA. Global Earthquake Satellite System (GESS). <http://solidearth.jpl.nasa.gov/gess.html>, 2010.
- [26] D. Patel, C. Giertsen, J. Thurmond, J. Gjølberg, and M. E. Gröller. The seismic analyzer - interpreting and illustrating 2D seismic data. *IEEE Trans. on Visual. & Comput. Graph.*, 14(6):1571–1578, Oct. 2008.
- [27] S. Potts, M. Tory, and T. Moller. A parallel coordinates interface for exploratory volume visualization. In *Proc. of IEEE Vis*, page 102, 2003.
- [28] J. Rundle, W. Klein, and S. Gross. Physical basis for statistical patterns in complex earthquake populations: Models, predictions and tests. *Pure and Appl. Geophys.*, 155(2-4):575–607, 1999.
- [29] V. Sgrignal, C. Information, R. Console, L. Conti, A. M. Galper, V. Malvezzi, M. Parrot, P. Picozza, R. Scrimaglio, P. Spillanti, and D. Zilpimiani. The esperia project: a mission to investigate the near-earth space. *Earth Observation with CHAMP*, pages 407–412, 2005.
- [30] T. Tu, H. Yu, L. Ramirez-Guzman, J. Bielak, O. Ghattas, K.-L. Ma, and D. R. O'Hallaron. From mesh generation to scientific visualization: an end-to-end approach to parallel supercomputing. In *SC'06: Proc. of ACM/IEEE Supercomputing*, page 91, 2006.
- [31] R. H. Wolfe, Jr. and C. N. Liu. Interactive visualization of 3D seismic data: a volumetric method. *IEEE Comput. Graph. Appl.*, 8(4):24–30, 1988.
- [32] M. Wyss, P. Bodin, and R. E. Habermann. Seismic quiescence at parkfield: an independent indication of an imminent earthquake. *Nature*, 345:426–428, 1990.
- [33] M. Wyss and R. E. Habermann. Precursory quiescence. *Pure and Appl. Geophys.*, 126:319–332, 1988.
- [34] H. Yu, K.-L. Ma, and J. Welling. A parallel visualization pipeline for terascale earthquake simulations. In *SC'04: Proc. of ACM/IEEE Supercomputing*, page 49, 2004.
- [35] X. Yuan, P. Guo, H. Xiao, H. Zhou, and H. Qu. Scattering points in parallel coordinates. *IEEE Trans. on Visual. & Comput. Graph.*, 15(6):1001–1008, 2009.
- [36] D. A. Yuen, B. J. Kadlec, E. F. Bollig, W. Dzwiniel, Z. A. Garbow, and C. R. S. da Silva. Clustering and visualization of earthquake data in a grid environment. *Visual Geosciences*, pages 1–12, 2005.