

## Insights gained through visualization for large scale earthquake simulations

Chourasia. A. <sup>1</sup>, Cutchin. S. M. <sup>1</sup>, Olsen. K.B. <sup>2</sup>,  
Minster. B. <sup>3</sup>, Day. S. <sup>2</sup>, Cui. Y. <sup>1</sup>, Maechling. P. <sup>4</sup>,  
Moore. R. <sup>1</sup>, Jordan. T. <sup>4</sup>

<sup>1</sup>San Diego Supercomputer Center, UCSD

<sup>2</sup>Department of Geological Sciences, SDSU

<sup>3</sup>Scripps Institute of Oceanography, UCSD

<sup>4</sup>Southern California Earthquake Center, USC

### Abstract

Large weather and astrophysical simulations are performed on a regular basis. These simulations run on massively parallel machines and yield a wide variety of data products. The data products are then examined to gain insight and understanding of the simulated phenomena. Our study focuses on the visualization of a series of large earthquake simulations collectively called TeraShake. The simulation series aims to assess the impact of Southern San Andreas Fault earthquake scenarios on Southern California. The two simulation phases consist of seven production runs yielding 100 TB of data total. We discuss the role of visualization in gaining scientific insight and aiding unexpected discovery. We also cite instances of how visualization helped in finding simulation instabilities, and how these discoveries led to improvement in the simulation code and input parameters.

### 1.0 Introduction

Each year estimated losses from earthquakes are measured in billions of dollars and thousands of fatalities [2]. While earthquakes have been studied for a long period of time, recent advances in computation now enable large scale simulations of seismic wave propagation. The use of visualizations to interpret simulation output has been used extensively in the TeraShake project, which has completed one of the first large scale simulations done on the Southern San Andreas Fault.

### 1.1 Earthquake Threat

The San Andreas Fault has a history of producing large M8 earthquakes [6]. The most recent was the 1906 M7.9 event [1] that caused tremendous

damage to San Francisco. The 1857 (Fort Tejon) earthquake ruptured a 360km stretch from Parkfield to Wrightwood. However, the San Bernardino Mountains Segment and the Coachella Valley segment of the San Andreas Fault have not seen a major event since 1812 and 1690, respectively [8]. Average recurrence for large earthquakes with surface rupture on these segments is  $146 + 91 - 60$  yrs and  $220 \pm 13$  yrs, respectively [10]. A major component of the seismic hazard in southern California and northern Mexico comes from a large earthquake on this part of the San Andreas Fault [3]. Because no strike-slip earthquake of similar or larger magnitude has occurred since the first deployment of strong motion instruments in southern California, there is large uncertainty of the ground motions expected from such an event. Large population densities make this looming hazard a major threat in this region.

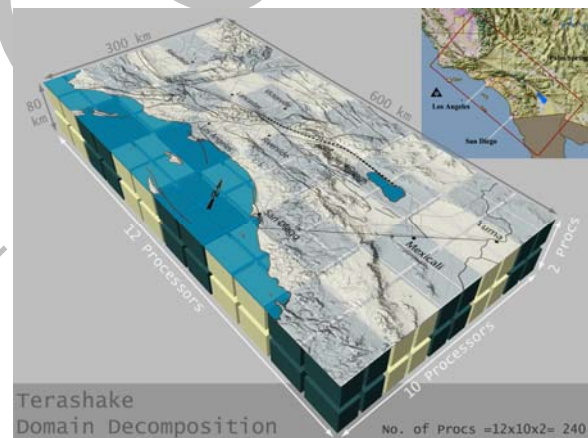


Figure 1: The top right inset shows the simulation region 600km long and 300 km wide, indicated by the red rectangle. In the center the topography, fault lines and city locations are visible. Domain decomposition of this region into 240 processors is also shown.

### 1.2 TeraShake Simulation

TeraShake is a large-scale finite-difference (fourth-order) simulation of an earthquake event based on Olsen's Anelastic Wave Propagation Model code, conducted in the context of the Southern California Earthquake Center Community Modeling Environment. The 600km x 300km x 80km simulation domain [Fig 1] extends from the Ventura Basin and Tehachapi region to the north and to

Mexicali and Tijuana to the south. It includes all major population centers in southern California, and is modeled at 200m resolution using a rectangular, 1.8 giganode, 3000 x 1500 x 400 mesh. The simulated duration is 250 seconds, with a temporal resolution of 0.011 seconds, maximum frequency of 0.5Hz, for a total of 22,727 time steps. The study was conducted in two phases. TeraShake-1 was modeled using a kinematic source description [Fig 2b] for the San Andreas Fault rupture. Four production runs with different parameter and input settings were made for phase-one. TeraShake-2 added a physics-based dynamic rupture component to the simulation, which was run at a very high 100 meter resolution, to create the earthquake source description of the fault. This is physically more realistic than the kinematic source description [Fig 2b] used in TeraShake-1. Three productions runs were done for TeraShake-2 with various parameter settings.

## 2.0 Data Products and Challenges

To date, all of the TeraShake simulations have consumed 800,000 CPU hours on the 15.6 teraflop Datastar supercomputer at San Diego Supercomputer Center (SDSC) and have generated 100 TB of output. The TeraShake application runs on 240-1600 processors and generates up to about 100,000 files per simulation, for more than one million files in total. The output data consist of a range of data products including surface velocity components (2D in space, 1D in time) and volumetric velocity components (3D in space, 1D in time). The surface data are recorded for every timestep while the volumetric data are recorded for every 10<sup>th</sup> or 100<sup>th</sup> timestep of the simulation. Other products including velocity magnitude, peak cumulative velocity and displacement are derived from the output data.

This simulation data poses significant challenges for analysis. The foremost need is for computational verification of simulation progress at runtime and thereafter seismological assessment of data computed. On completion of the simulation run, the multivariate data need to be intuitively and quickly accessible to scientists at different geographical locations. Features of interest or key regions need to be identified for further analysis. This requires that the geographical context be integrated with data when presented.

## 3.0 Visualization Techniques

We utilized existing visualization techniques and combined them with off-the-shelf software to create meaningful imagery from the dataset. We classify our visualization process into four categories: Surface, Topography, Volumetric visualizations and Maps.

### 3.1 Surface Visualization

In this technique the 2D surface data are processed via direct 24 bit color map and overlaid with contextual geographic information. Annotations and captions provide additional explanation. The temporal sequence of these images is encoded into an animation for general dissemination. We utilized Adobe's After Effects for compositing and encoding the image sequences. Layering the geographic information with high resolution simulation results provides precise, insightful, intuitive and rapid access to complex information. This is required for seismologists to clearly identify ground motion wave-field patterns and the regions most likely to be affected in San Andreas Fault earthquakes. Surface visualizations were created for all 2D data products using this method.

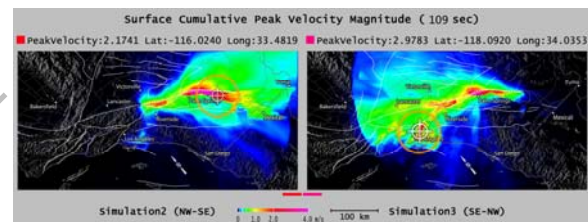


Figure 2a: Image comparing the effect of ground shaking produced by two different earthquake scenarios on the San Andreas Fault. The left image shows the areas affected by a rupture traveling north to south, the right image shows the corresponding results for a rupture moving south to north. Both images have geographical and contextual information (fault lines, freeways) overlain. Seismic information of time, peak velocity and instantaneous location are shown visually via a graphical cursor and in text. The goal is to gain an understanding of the areas most heavily impacted by such an earthquake and the degree of damage. The side-by-side character of this visualization also provides the scientists an important visual intuition regarding similarities and differences between fault rupture scenarios.

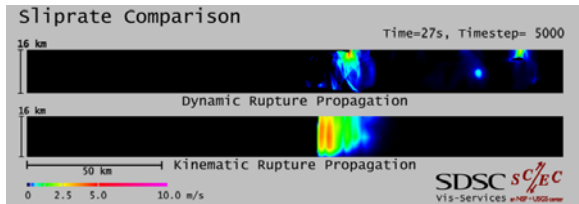


Figure 2b: Comparative snapshot of source description for TeraShake 1 and 2 simulations. With use of animations like this scientists were able in some instances to find instabilities in the dynamic rupture model, and in other instances to note specific physical behaviors, such as that the rupture velocity was exceeding the S wave speed at some locations.

### 3.2 Topography Visualization

This process utilizes the dual encoding of the surface velocity data as both color mapping and as displacement mapping. The surface velocity data are used to create a color mapped image and the displacement magnitude calculated from the surface velocity data are used to generate a grey scale image. The grey scale image is used as a displacement map to create 3D terrain deformation from the data along the vertical axis [Fig 3a] in Autodesk's Maya. Maya offers state of the art rendering, lighting and animation capabilities. We have created a custom pipeline to integrate our visualization tools with Maya. Currently we are working to develop a method to display true 3D deformation [Fig 3b] based on three component data to provide more realistic insight.

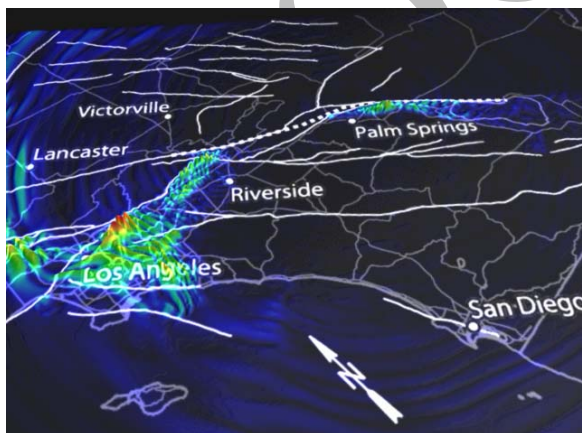


Figure 3a: Deformation along the vertical axis of the terrain using displacement mapping to show velocity magnitudes along with color.

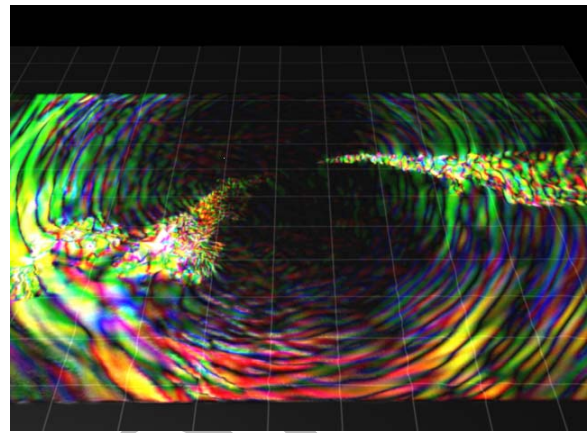


Figure 3b: 3D deformation of the terrain showing x, y and z components of velocity as red, green and blue respectively.

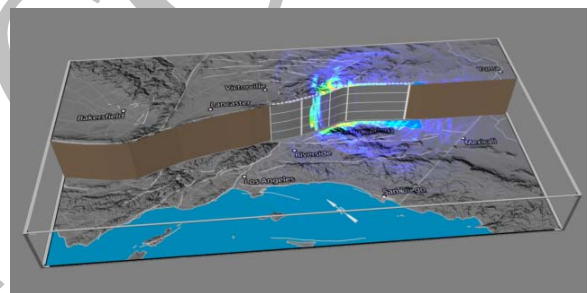


Figure 3c: A cross sectional view showing slip-rate on the fault plane and velocity magnitude on the surface topography.

The use of multiple data sets with different visual representations [Fig 3a-c, 4, 5a] helps the seismologists to understand key earthquake concepts like seismic wave propagation, rupture directivity, peak ground motions, and the duration of shaking. The strong visual impact leads the viewer from the global context of earthquake hazards to the hazards in a specific region and then into the details about a specific earthquake simulation. Viewing the earthquake simulation evolve over time leads viewers to gain insights into both wave propagation and fault rupture processes, and illustrates the earthquake phenomena in an effective way to non-scientists.

### 3.3 Volumetric Visualization

The bulk of the data yielded by simulation is volumetric. This is by far the most significant for

analysis as it holds the most information content. We perform direct volume rendering [4] of the volumetric dataset and composite it with contextual information to provide a holistic view of the earthquake rupture and radiates waves to the scientists. Our initial work has helped the seismologists to see the features in general but more work needs to be done to represent multivariate data in a unified visual format.

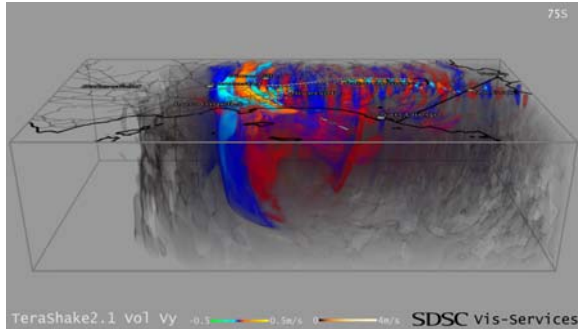


Figure 4: Snapshot showing volume rendered velocity in y (shorter horizontal) direction.

### 3.4 Maps

Some data products like spectral acceleration (see Fig. 5a), peak ground velocities and peak ground displacements are non-temporal and require visual representation for better understanding. Spectral acceleration maps depict the vibration characteristic of ground at different frequencies.

#### 3.4.1 Self Contoured Maps

We developed a technique to highlight features in 2D by using bump mapping. The detailed treatment of this approach is described in detail by Wijk [8]. Encoding the spectral acceleration levels using both color maps and bump maps reveals subtle transitions between ground motion levels within localized regions with similar spectral acceleration properties.

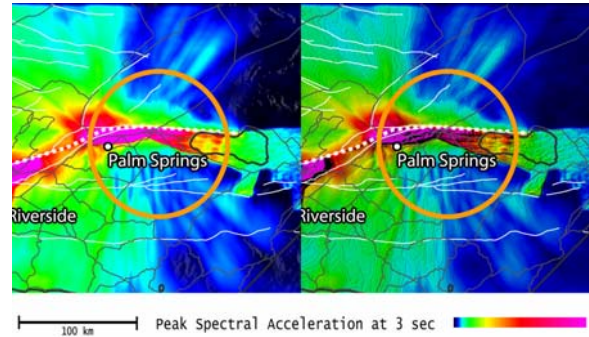


Figure 5a: Map showing the advantage of the self contouring technique in contrast to simple color mapping. The color and bump encoding technique brought out variations in the data that were not previously visible. In the region highlighted by the orange circle, scientists identified a star burst pattern, indicating an unusual radiation of energy worthy of further investigation, which went unnoticed with simple color mapping.

#### 3.4.2 Web-Portal for surface data visualization

Scientists want to conduct hands on analysis in an attempt to gain a better understanding of output data. The size of TeraShake data poses a significant problem for accessibility and analysis. We developed a web front end where scientists can download the data and are able to create custom visualizations over the web directly from surface data. The portal uses LAMP (Linux, Apache, MySQL, PHP) and Java technology for web middle-ware and on the back-end compute side relies on specialized programs to fetch data from the archive, visualize, composite, annotate and make it available to client browser.

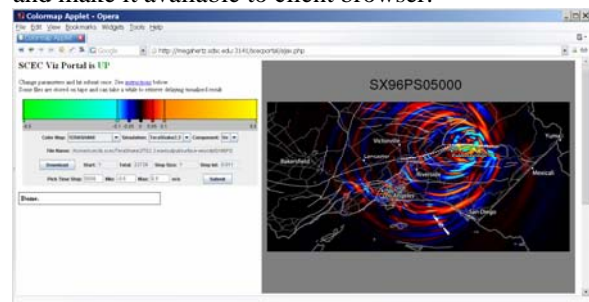


Figure 5b: Screenshot of the visualization portal.

### 4.0 Visualization Tools and Results

SDSC's volume rendering tool Vista, based on the Scalable Visualization Toolkit (SVT), was

used for visualization rendering. Vista employs ray casting [4] with early ray termination for performing volumetric renderings. Surface and volume data have been visualized with different variables (velocities and displacements) and data ranges in multiple modes. The resulting animations have proven valuable not only to domain scientists but also to a broader audience by providing an intuitive way to understand the results. Visualization required significant computational resources. So far visualizations alone have consumed more than 10,000 CPU hours and over 30,000 CPU hours on SDSC's Datastar and TeraGrid IA-64 respectively. More than 100 visualization runs were performed, with each run utilizing 8 to 256 processors in a distributed manner. The results have produced over 130,000 images and more than 60 unique animations [7].

### **5.0 Insights Gained Through Visualization**

Since the very beginning of the project the simulation results have been a bit of mystery to the seismologists. It was unknown what the simulation would predict. Scientists were able to gain the following insights from the visualizations:

- In the north-westward propagating rupture scenario the wave propagation is strongly guided toward the Los Angeles basin after leaving the San Andreas Fault (unexpected)
- The sediment filled basin acts as an amplification source for trapped waves. A Strong amplification is observed in the LA basin long after the initial rupture (unexpected)
- Contiguous basins act as energy channels, enhancing ground motion in parts of the San Gabriel and Los Angeles basins.
- Identification of regions of particularly strong shaking
- Validation of input rupture model and instability identification
- Observation of star burst patterns in the SA maps (unexpected)

### **5.1 Discussion**

The role of contextual information has been pivotal; the surface visualizations have proven to be

very helpful to the scientists. Encoding of the rendered image sequence into animations has been a bottleneck since it is serial, time consuming and lossy compression process. Further it requires careful selection of codecs for broader accessibility by scientists on different platforms. Large temporal datasets pose significant challenges for analysis. Automation of visualization techniques and methods applied in a carefully planned way is desirable. Interactive visualization of large temporal datasets seems useful but is non-trivial and often impractical. Domain specific feature capturing algorithms coupled with visualization can play an important role for analysis. Animations though non-interactive can often serve the purpose for gaining insight when created in an thoughtful manner. Off-the-shelf software's such as Maya can augment scientific visualization tools. Multiple representations of the same data products with different techniques can be valuable assets. Use of HDR imagery has seemed promising but lack of HDR display hardware and plethora of tone mapping methods make this task difficult.

### **6.0 Future Direction**

Planned future work includes online analysis of surface data by remote web clients plotting synthetic seismograms. Data mining operations, spectral analysis and data sub setting are planned as future work. The TeraShake simulation project has provided some insights on the IT infrastructure needed to advance computational geosciences, which we will examine further. We would like to use 16bit imagery instead of current 8bit imagery to increase visualization fidelity. In addition integration of GIS databases to overlay a variety of contextual information would aid further analysis. Integration of surface imagery with virtual globes like Google Earth, NASA's World Wind, etc is being considered.

### **7.0 Acknowledgements**

We would like to thank Marcio Faerman, Yaunfang Hu, Jing Zhu and Patrick Yau for their contributions. We would also like to express our thanks to scientists and engineers from SDSC and SCEC who made this work possible. The work was funded by SDSC and SCEC.

### **References**

1. Bolt, B. A. (1968), The focus of the 1906 California earthquake, *Bull. Seismol. Soc. Am.*, 68, 457– 471.
2. Federal Emergency Management Agency, *HAZUS@99 Estimated Annualized Earthquake Losses for the United States*, FEMA Report 366, Washington, D.C., 33 pp., February 2001.
3. Frankel, A. D., et al. (2002), Documentation for the 2002 update of the National Seismic Hazard maps, U.S. Geol. Surv. Open File Rep. 02-420.
4. Levoy, M., *Display of Surfaces from Volume Data*, *IEEE Computer Graphics & Applications* 8, 5 (May 1988).
5. Olsen, K. B., Day, S. M., Minster J. B., Cui, Y., Chourasia, A., Faerman, M., Moore, R., Maechling, P., Jordan, T.(2006), Strong shaking in Los Angeles expected from southern San Andreas earthquake, *Geophys. Res.*, 33, L07305, doi:10.1029/2005GL025472. 2006
6. Sieh, K. (1978), Prehistoric large earthquakes produced by slip on the San Andreas fault at Pallett Creek, California, *J. Geophys. Res.*, 83, 3907–3939.
7. TeraShake Visualization webpage, <http://visservices.sdsc.edu/projects/scec/TeraShake/>
8. Weldon, R., K. Scharer, T. Fumal, and G. Biasi (2004), Wrightwood and the earthquake cycle: What a long recurrence record tells us about how faults work, *Geol. Seismol. Am. Today*, 14, 4 – 10.
9. Wijk, J.J. van, A.C. Telea. Enridged Contour Maps. Proceedings of the conference on Visualization 2001, IEEE Computer Society, p. 69-74.
10. Working Group on California Earthquake Probabilities (1995), Seismic hazards in southern California: Probable earthquakes, 1994 to 2024, *Bull. Seismol. Soc. Am.*, 85, 379– 439.