1. Introduction to the TeraShake 2.1 Simulation

TeraShake 2.1 represents a physics-based simulation of a magnitude 7.7 earthquake on the Southern San Andreas Fault. The TeraShake 2.1 simulation was performed by scientists at the Southern California Earthquake Center (SCEC) and researchers at San Diego Supercomputer Center (SDSC).

In the TeraShake 2.1 simulation, an Mw 7.7 earthquake begins south of Palm Spring and propagates north-west on the San Andreas Fault. The rupture begins at time 0 in the simulation and continues for approximately 60 seconds. At time 60 seconds, the fault rupture stops but the earthquake waves continue to propagate throughout the region for several minutes. The simulation stops at time 250 seconds after the earthquake started by which time most strong ground motions have subsided.

TeraShake 2.1 was performed by propagating earthquake waves through an anelastic volume using finite difference codes on San Diego Supercomputer Center (SDSC) DataStar computer. The material characteristics of the volume are specified prior to running the simulation using a 3D velocity model of the southern California region, SCEC Community Velocity Model 3.0, a realistic 3D velocity structure for the region that was developed by SCEC. The edges and the bottom of the simulation volume are designed to act as absorbing boundaries to reduce or eliminate reflections of waves as they reach the edges of the simulation volume. The surface of the ground is treated as a free surface, not an absorbing boundary. Ground motions at the surface are of particular interest because that is where people live.

This particular earthquake, an Mw 7.7 southern San Andreas earthquake, was simulated because such earthquakes are known to have occurred in the past and they are reasonably likely to occur again. The simulation results are of significant interest to scientists and the public in the southern California region. The simulation produces high resolution ground motion records for the entire simulation area including the many large cities in the region. The distribution of strong ground motions and the levels of ground motion in areas of significant populations have important societal implications.

This type of simulation has been validated and shown to produce accurate results for earthquake waves at 0.5Hz and longer periods. Validation has been done by running similar simulations for historic earthquakes and comparing the resulting synthetic seismograms produced by the simulation to the seismograms there were recorded by instruments in the region during the actual earthquake.

Figure 1: Map showing area in southern California that was used in the TeraShake 2.1 simulation. The red box indicates the extent of the region used in the simulation.

2. Region:

TeraShake 2.1 simulates earthquake propagating through a volume that represents a region in southern California. The region used in the TeraShake simulation is a 600x300x80 km3 area of Southern California. The surface area of this region is shown in Figure 1. The geographically origin of this region is Latitude: 34.5, Longitude: -121.0 at depth of 80km. The four coordinate coordinates of the box are (121W, 34.5N), (118.9511292W, 36.621696N), (116.032285W, 31.082920N), 113.943965W, 33.122341N).

The result is a rectangular region 600 km by 300 km in extent, with long axis oriented N50W and short axis N40E. We extend the domain to 80 km depth, well into the upper mantle. The X coordinates increase moving to the south-east. The Y coordinates increase moving to northeast. The Z coordinates increase moving up towards the surface. This is a right-handed coordinate system.

In this simulation, the region is treated as a flat Cartesian box. The curvature of the earth and the topography are ignored.

This volume is divided into regular squares with a 200 meter spacing in all directions creating a 3000x1500x400 point regular mesh. This results in a mesh consisting of 1.8 billion regular cubes. Material properties are specified for each cube at the start of the simulation. The output data is saved as ground velocity for each cube.

The region is specified using geographical coordinates to begin. The geographical coordinates are projected onto a flat surface and then the geographically-based material properties are determined.

Given the geo-referenced material properties, the simulations are then conducted in the internal mesh coordinate system with point 0,0,0 as the origin at the south, west, corner at the bottom of the volume.

3. Simulation Time:

The simulation calculates the state of the entire volume every 0.011seconds.

This is calculated for a total of 250.03 seconds which requires the calculation of 22730 time steps.

4. Output Data

TeraShake 2.1 output data is time-varying velocity vector data. This data can be configured, or formatted, in different ways. There are three main types of TeraShake 2.1 output data.

First, the 4D volumetric velocity data represents the vector velocity state at each point in the volume throughout the simulation. This type of data provides insights into the subsurface paths through which the earthquake waves propagate. Because of the large storage requirements, this data is decimated in both space and time. The 4D volumetric data is decimate by 100 in time (1 volume every 1.1 seconds) and is decimated by a factor of 4 in each spatial dimension. The resulting volume meshes are 750x375x100 for a total of 28,125,000 cubes per time step and a total of 227 time steps.

Second, the surface data represents the vector velocity state at each point at the surface throughout the simulation. This is redundant. The surface data is also included in the volume data, but since the surface ground motions are of particular interest, this data format is often provided. The surface data has been saved at a higher time sample rate. Surface data was decimated by a factor of 10, producing a total of 2270 surface files.

Third, the surface data has been transposed to produce surface seismogram files. These files represent the seismograms at each point at the surface. Seismograms are of interest to seismologist because it is a very familiar data representation for them.

Further processing of data into displacement, or calculation of data types such as gradient, divergence, and curl are possibly of interest. The frequency contents of the original and decimated data should be considered if such processing is considered.

5. File Formats:

All TeraShake 2.1 simulation data is big-endian, 4 byte, IEEE floating point binary format data. The data is saved in binary files with no header. In order to read and interpret the data, you must know the x,y,z dimensions of the mesh, the geo-referenced coordinates for the mesh, and data types that the floating points represent, the units of the data, and the order in which the indices change.

The TeraShake 2.1 data values are velocity in cm/sec.

The velocities created by the simulation are broken into 3 components (x,y,z) and are divided into files by component and time step. Thus, there are 3 files (x,y,z) for each time step.

6. Volume Data:

The volume data was sampled at every 100th time step so there are 227 files for each component and three components for a total of 681 files. Each file represents a volume of earth at a particular time in the simulation.

The files are reduced twice in each spatial direction producing files that are 1/32nd the size of the original (750x375x100). These reduced volume files are named TS21z_*Component*_R2_*Timestep*.

The data in these are 4 byte floats in the format fast x, then y, then z. Following the format fast x, then y, slow z, the floats in the file will be ordered like this:

X Mesh Coordinate, Y Mesh Coordinate, Z Mesh Coordinate

0,0,0 <--- 1st 4 bytes of data
1,0,0 <--- 2nd 4 bytes of data
2,0,0
...
1499,0,0
0,1,0
1,1,0
2,1,0
...
1499,749,0
0,0,1
1,0,1
2,0,1
...
1499,749,199

Important info:
250.00 TMAX = propagation time (.011 x 22730)
200.00 DH = spatial step for x, y, z (meters)
0.011 DT = time step (seconds)
3000 NX = x model dimension in nodes
1500 NY = y model dimension in nodes
400 NZ = z model dimension in nodes

7. TeraShake Science Questions:

(a) Are earthquake waves guided from the San Andreas into the Los Angeles Basin?

(b) Do strongly shaken basins act as wave sources?

(c) Do anelastic models produce long-period strains that are larger than what the earth is believed capable of supporting?

(e) At which locations do the largest conversion between wave types occur?

(f) Do the waves in the Whittier-Narrows area from the SE-NW ruptures follow a pronounced sediment channel defined in the crustal structure?

(g)Is there a depth-dependent asymmetry of the waves across the fault due to velocity contrast across the fault?

(i) Which regions produce wave reflections?

(j) Do waves systematically focus toward the centers of basins, thus providing a physical explanation of the correlation of amplification factors with basin depth (on average).

(k) Do wave amplitudes increase or decrease, on average, at the far sides of the basins where the depths are decreasing (an important question with respect to using basin depth as a proxy for amplification).

(l) Are ground motions intrinsically more variable near basin edges? In other words, should anomalous ground motion like seen in Santa Monica during Northridge, when averaged over events in different locations, be treated as a systematic effect (amplification factor) or as more intrinsic variability (increase in sigma).

8. Info about histograms

Each histogram files contains the Min and Max found in that data

The min/max range is divided into 1536 bins of equal size. Each row with 5 columns is a bin.

Histogram file has five columns: Index Perc Offset Min Max
Perc is percentage of data found in a particular bin
Offset is cumulative total of the percentage up till that bin
Min is the minimum value found in the bin
Max is the maximum value found in the bin (when max=0.0 then min=max)
For the bins that were empty (no data found in that bin, i.e. perc=0.0) only "offset" column will have non zero entry.