SCEC Simulation Data Access

16 February 2018

Philip Maechling (maechlin@usc.edu)
Fabio Silva, Scott Callaghan, Christine Goulet, Silvia Mazzoni,
John Vidale, et al.

SCEC
an NSF+USGS center
SCEC Open Data Policy:
SCEC has an open data policy for research projects and results. Consistent with this, all workshop datasets are available to workshop participants on request, as distributed to workshop researchers.

Simulation Data Access Planning:
Observational ground motion data centers provide data access tools (CESMD, PEER).

Ground motion simulation data centers are beginning to develop specialized simulation data access tools.

In Collaboration with Natural Hazards Engineering Research Infrastructure (NHERI) DesignSafe project, SCEC is developing simulation data access tools based on observational data access tools.
PEER Ground Motion Database
(https://peer.berkeley.edu/products/strong_ground_motion_db.html)

Center for Engineering Strong Motion Data (CESMD)
(https://www.strongmotioncenter.org/)
**SCEC Simulation Data Sets**

**Broadband Platform Part A and Part B from the latest BBP release (v17.3):**
- Data Types: seismograms for 6 methods with site response, and without site response
- Data Set Size: 7TB
- Data Access Methods: Web-based tar files

**Broadband Platform Part A Northridge:**
- Data Types: simulated seismograms for 50 variations, 5 methods, 38 sites, in trimmed format
- Data Set Size: 500GB
- Data Access Methods: Web-based tar files

**Broadband Platform GMSV Study 17.3:**
- Data Types: ruptures, and simulated seismograms for scenario events: Northern San Andreas (M8.0), Hayward (M7.0), Elysian Park (M6.6), Southern San Andreas (M7.9), and San Jacinto (M7.8) at sites San Francisco Downtown, Los Angeles Downtown, San Bernadino
- Data Set Size: 87 GB
- Data Access Method: DesignSafe Data Access Tool
- Data Access: Planned 2018

**CyberShake 15.4:**
- Data Types: maps, ruptures, amps, and simulated seismograms for 336 sites to 1Hz
- Data Set Size: ~15TB
- Data Access Methods: SQL Queries, Python/C Code, USC Acct Required

**CyberShake 15.12:**
- Data Types/Formats: seismograms for 5 sites (LADT, STNI, WNGC, PAS, SBSM) to 10Hz
- Data Set Size: 500GB
- Data Access Methods: SQL Queries, Python/C Code, USC Acct Required
Observational Seismic Data Request Parameters

Ground motion simulation data sets can be significantly different that observational data sets.

Observational ground motion datasets are often organized around an specific earthquake (e.g. Network.Event ID), with several data types in the event data collection.

Simulation ground motion datasets are often a collection of data. Simulation data sets are less frequently organized around a single earthquake. SCEC describes simulations as Studies, so simulation results are identified by Platform Name (e.g. Broadband Platform, CyberShake) and Study ID (which is year and month the study began (in format 17.3)).
- “Broadband Platform” (Focus of current workshop)
- “CyberShake” (used by UGMS Committee)

<table>
<thead>
<tr>
<th>Current Differences</th>
<th>CyberShake</th>
<th>Broadband Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>PSHA</td>
<td>Scenarios</td>
</tr>
<tr>
<td>Methods</td>
<td>Graves &amp; Pitarka</td>
<td>Several (7)</td>
</tr>
<tr>
<td>Basin effects</td>
<td>3-dimensional</td>
<td>1-dimensional</td>
</tr>
<tr>
<td>Frequency band</td>
<td>&lt; ~1 Hz</td>
<td>0-100 Hz</td>
</tr>
<tr>
<td>Computer needed</td>
<td>Supercomputer</td>
<td>Personal computer</td>
</tr>
<tr>
<td>Validations</td>
<td>Relatively limited</td>
<td>Relatively extensive</td>
</tr>
</tbody>
</table>
SCEC Broadband Platform (BBP)

- Simulates Broadband ground motions from 0.1-100Hz (deterministic up to ~ 1Hz)
- 6 alternative simulation codes
- Uses simple 1D seismic velocity structure for source and path
- Methods and software fully validated for spectral response
  - Multiple rounds of validation/improvements
  - Independent review panel
- Used for large ground motion characterization projects
  - Southwestern U.S. utilities (NPPs)
  - PEER NGA-East for Central and Eastern North America

<table>
<thead>
<tr>
<th>Rupture Generator</th>
<th>0 to 1Hz Deterministic Ground Motions</th>
<th>1 to 10Hz Stochastic Ground Motions</th>
<th>Common Ground-Motion Post-processing</th>
<th>Common Goodness of Fit Post-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP</td>
<td>Gen_Slip</td>
<td>JB_Sim</td>
<td>HF_Sim</td>
<td>RotD50</td>
</tr>
<tr>
<td>SDSU</td>
<td>Gen_Slip</td>
<td>JB_Sim</td>
<td>BB_Toolbox</td>
<td>Bias Plot</td>
</tr>
<tr>
<td>UCSB</td>
<td>UCRMG</td>
<td>Syn1D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXSIM</td>
<td>EXSIM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSM</td>
<td>Simula</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Released by SCEC for download Github: https://github.com/SCECcode/BBP
Validation Data Products
Seismogram Comparison

LOMAP vs. run 10000021, station 8001-CLS
Comparison Data Products

*Goodness-of-fit*

*Spectral response comparison*

*Seismogram comparison*
Validation Data Products: Map GOF

GOF Comparison between LOMAP and simulation 10000021
R < 85 km

RotD50

PSa North 5%

PSa East 5%

GOF Comparison between LOMAP and simulation 10000021
# Broadband Part A

<table>
<thead>
<tr>
<th>Event</th>
<th>Mw</th>
<th>Variations</th>
<th>Site Response</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alum Rock</td>
<td>5.45</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>Chino Hills</td>
<td>5.39</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>Landers</td>
<td>7.22</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>Loma Prieta</td>
<td>6.94</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>Mineral</td>
<td>5.68</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, UCSB</td>
</tr>
<tr>
<td>Niigata</td>
<td>6.65</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>North Palm Springs</td>
<td>6.12</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>Northridge</td>
<td>6.73</td>
<td>50</td>
<td>No / Yes for GP*</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>Riviere-du-Loup</td>
<td>4.6</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>Saguenay</td>
<td>5.81</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>Tottori</td>
<td>6.59</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>Whittier Narrows</td>
<td>5.89</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
</tbody>
</table>

Note: Northridge was calculated both with and without site response (only for the GP method), using the GP 2014 version of the site response module and removing the correction coefficients applied to the observation data files.
### Broadband Part B GMPEs

Part-B rupture variations may differ by hypocenter location and by slip distribution.

<table>
<thead>
<tr>
<th>Event</th>
<th>Region</th>
<th>Distance (km)</th>
<th>Variations</th>
<th>Site Response</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5 Reverse</td>
<td>LABasin, Northern California</td>
<td>20, 50</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>6.2 Strike-Slip</td>
<td>LABasin, Northern California</td>
<td>20, 50</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>6.6 Reverse</td>
<td>LABasin, Northern California</td>
<td>20, 50</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
<tr>
<td>6.6 Strike-Slip</td>
<td>LABasin, Northern California</td>
<td>20, 50</td>
<td>50</td>
<td>No</td>
<td>ExSim, GP, SDSU, Song, UCSB</td>
</tr>
</tbody>
</table>
Data Collection includes Ground Motion at Selected Sites for these Ruptures including:

1. San Francisco Downtown (8029-RIN, 37.786 N, 122.391 W)
2. Los Angeles Downtown (LADT, 34.052 N, 118.257 W)
3. San Bernadino (S688, 34.104 N, 117.288 W)
DesignSafe is the web-based research platform of the NHERI Network that provides the computational tools needed to manage, analyze, and understand critical data for natural hazards research.

https://www.designsafe-ci.org/
What is CyberShake?

- SCEC’s 3D physics-based PSHA platform calculates long-term site-oriented seismic hazard estimates
- 7,000 ERF ruptures -> 500,000 events
  - Slips and hypocenter locations varied
- Simulates seismograms from UCERF earthquake rupture forecast (<200 km)
- Intensity measures extracted from seismograms
- Hazard curves created for individual locations in region of interest, interpolated for map
- Engineers using CyberShake results to inform ground motion predictions
- UGMS Committee: “Use of 3-D Physics-Based Numerical Simulations in the Development of Long Period Ground-Motion Maps for Los Angeles”, Thursday at 4:15
CyberShake Multi-layer Simulation Data

Structure of the CyberShake Hazard Model

1. Hazard map
2. Hazard curves
3. Hazard disaggregation
4. Rupture model
5. Seismograms
CyberShake Study 15.4 sites (336)
Study 15.4 Data Products

• CVM-S4.26 Los Angeles-area hazard maps
  • RotD100 2, 3, 4, 5, 7.5, 10 sec
  • RotD50 2, 3, 4, 5, 7.5, 10 sec
  • Geometric mean 2, 3, 5, 10 sec
• Hazard curves for 336 sites, at 2s, 3s, 5s, 10s
• 336 sets of 2-component SGTs
• Seismograms for all ruptures (~160M)
• Peak amplitudes in DB for 2s, 3s, 5s, 10s
  • RotD100, RotD50 and geometric mean SA
CyberShake Seismic Hazard Models

CyberShake Results are used by OpenSHA to calculate Maximum Considered Earthquake Response (MCER)
DesignSafe is the web-based research platform of the NHERI Network that provides the computational tools needed to manage, analyze, and understand critical data for natural hazards research.

https://www.designsafe-ci.org/
Ground-Motion Portal

Description:
- Provide user interface to search and retrieve SCEC BBP record files

Scope
- Develop a web-API-based search tool with access limited to basic input-output parameters and existing data for NEHRI application.
- The input will be limited to the following basic parameters:
  - Magnitude, Distance, Vs30, IM, IM level tolerance
  - Simulation platform
  - Study (set of simulation data)
- The output will be limited to the BBP results, using the SCEC flatfile as the prototype.
- The search tool will be open-source, written in Python. It will be able to run within the NEHRI CI system.
- The basic framework of the tool will be a web-services system where the simulation data will be accessed by the tool via a web-application API.
**SCEC BBP Ground-Motion Manager for Simulation Ground Motions**

**Introduction**

The SCEC Broadband Platform is a software system that can generate 0-100 Hz seismograms for historical and scenario earthquakes in California, Eastern North America, and Japan using several alternative computational methods.

The goal of the SCEC Broadband Platform is to generate broadband (0-100 Hz) ground motions for earthquakes. The SCEC Broadband Platform is a collaborative software development project involving SCEC researchers, research engineers, graduate students, and the SCEC/Qualis software development group. SCEC scientific groups have contributed modules to the Broadband Platform including rupture generation, low- and high-frequency seismogram synthesis, non-linear site effects, and visualization. These complex scientific codes have been integrated into a system that supports easy on-demand computation of broadband seismograms. The SCEC Broadband Platform is designed to be used by both scientific and engineering researchers with some experience interpreting ground motion simulations. More can be found at [www.scec.org](http://www.scec.org).

**Instructions**

**Input Constraints**

- **Record Information**
  - Input: Minimum and Maximum values. The data limits are given in (s)
  - Earthquake Moment Magnitude
  - Source-Station Joyner-Boore Distance (Rj)
  - Station Vs30

- **Intensity Measure, Spectral Ordinance**
  - Select the Spectral Ordinance from the option menu
  - Select the Period for the Spectral Ordinance (use 0.01 for PGA)
  - Select Minimum and Maximum values.

**Perform Search**

- Press the "Search" button to perform search using user input constraints.
- Processing time is proportional to the number of records meeting the search criteria.

**Output**

The output consists of the following:

- Number of Records meeting search Criteria.
- Reply of Search Criteria.
- Response-Spectra Plots (Rs2D00, Rs2D100, H1, H2)
- Record Intensities Table
- Response Spectra Tables (Rs2D00, Rs2D100, H1, H2)
- Record Manifolds Table (these data are also found in the metadata table)

---

**Export Search Results**

- The search results consist of the metadata tables, not the time-series themselves.
- Input a Search Label to identify this individual search.
- This data file will be placed in a folder with the SearchLabel as its name, in the SearchResults Directory.
- Press the "Export Search Results" button to save the search results in a set of files.
- All Search Result files are compressed into a tar.gz file for easy download.
- The program overwrites existing files that have the same Search Label.
- After each export _"_" is appended to the Search Label to maintain file uniqueness.

**SEARCH CRITERIA**

**Earthquake Magnitude (Range 6.73-8.73):**
- Min: 4
- Max: 8

**Rj, Distance, km (Range 0.0-105.33):**
- Min: 10
- Max: 20

**V_530, m/s (Range: 688-888):**
- Min: 100
- Max: 1700

**Intensity Measure:**
- Include Yes
- Period: 0.01
- I_min: 0.1
- I_max: 0.2

**Label (for csv export):**
- MiB: "mysearch"
User Interface

SEARCH CRITERIA

Earthquake Magnitude (Range: 6.73-6.73):

- Mmin: 4
- Mmax: 8

Rjb Distance, km (Range: 0.0-105.33):

- Rjimin: 10
- Rjimax: 20

V_S30, m/s (Range: -888--888):

- VS30min: -969
- VS30max: 1760

Intensity Measure:

- Include: Yes
- IM: PSA_RotD50_D0pt05
- IMmin: 0.1
- IMmax: 0.2

Period: 0.010

Label (for csv export):

- Label: mySearch

Search
Clear Output
Export Search Results

Plot/Refresh Trace
Press after performing Search
Search results
New trace
Thank you

Questions?
SCEC Community Velocity Model – Harvard (CVM-H)

Grid-based southern California velocity model with high resolution area in Los Angeles region constructed as part of a Unified Structural Representation (URS) for southern California that aims for consistency between CVM and CFM.

CVM-H Available From SCEC: https://scec.usc.edu/scecpedia/CVM-H
SCEC software developers have worked closely with USGS scientists and software developers to produce the OpenSHA seismic hazard platform since 2001.

OpenSHA is a Java-based computational platform that implements the Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3).

OpenSHA is used as a data source, and a data processing system for SCEC CyberShake platform.

OpenSHA software is being extended to include an implementation of an ETAS short-term earthquake forecast method.

OpenSHA Available at: https://www.opensha.org/
### BBP Regression Test Suites – Part A and Part B

#### Estimates for Part A and Part B Simulations

<table>
<thead>
<tr>
<th>Method</th>
<th>CPU Hours</th>
<th>Storage (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP</td>
<td>150</td>
<td>620</td>
</tr>
<tr>
<td>SDSU</td>
<td>150</td>
<td>620</td>
</tr>
<tr>
<td>EXSIM</td>
<td>36</td>
<td>500</td>
</tr>
<tr>
<td>CSM</td>
<td>71</td>
<td>390</td>
</tr>
<tr>
<td>Irikura</td>
<td>52</td>
<td>472</td>
</tr>
<tr>
<td>UCSB</td>
<td>150</td>
<td>620</td>
</tr>
</tbody>
</table>

| Total   | 609       | 3222         |
|         | 19488     |

**Estimate per Part A and Part B runs**

- **20k SU**
- **3.2 TB**
CyberShake Data Flow

1. Uniform California Earthquake Rupture Forecast

2. UCVM (poster 302)
   - Mesh generation
   - 1 job per site
   - MPI, 1500-4000 cores

3. AWP-ODC (poster 279)
   - SGT computation
   - 2 jobs per site
   - MPI, 200-800 GPUs

4. Graves-Pitarka kinematic rupture generator

5. Seismogram Synthesis
   - Post-processing
   - 500,000 jobs per site
   - MPI master/worker, 3712 cores

6. Data Product Generation
   - Populate DB, construct images
   - 6 jobs per site

7. Community Velocity Model

8. 500,000 Seismograms
   - 75M intensity measures

9. Data Product
   - 12 TB data transfer

10. CyberShake Hazard Map

Most recent CyberShake study took 4 weeks of real time and used 21 million core-hours
Southern California Earthquake Center

Seismological Data Centers

IRIS (https://ds.iris.edu/ds/nodes/dmc/data/)

USGS (https://earthquake.usgs.gov/data/)
Seismological Data Centers

Download Data

These pages describe what software and data sets are used to access data from the SCEDC archives.

- Waveform Data - Search, download or display waveforms from the SCEDC archive.
- Web Services - Search, download data from the SCEDC archive using web services.
- Special Data Sets - Download data sets produced by SCEC scientists
- Downloads Page - Table of all downloadable information from SCEDC archive

SCEDC
(http://scedc.caltech.edu/research-tools/index.html)
SCEC Simulation Data Sets

Broadband Platform Part A and Part B from the latest BBP release (v17.3):
- Data Types: seismograms for 6 methods with site response, and without site response
- Data Set Size: 7TB
- Data Access Methods: Web-based tar files

Broadband Platform Part A Northridge:
- Data Types: simulated seismograms for 50 variations, 5 methods, 38 sites, in trimmed format
- Data Set Size: 500GB
- Data Access Methods: Web-based tar files

Broadband Platform GMSV Study 17.3:
- Data Types: ruptures, and simulated seismograms for scenario events: Northern San Andreas (M8.0), Hayward (M7.0), Elysian Park (M6.6), Southern San Andreas (M7.9), and San Jacinto (M7.8) at sites San Francisco Downtown, Los Angeles Downtown, San Bernadino
- Data Set Size: 87 GB
- Data Access Method: DesignSafe Data Access Tool
- Data Access: Planned 2018

CyberShake 15.4:
- Data Types: maps, ruptures, amps, and simulated seismograms for 336 sites to 1Hz
- Data Set Size: ~15TB
- Data Access Methods: SQL Queries, Python/C Code, USC Acct Required

CyberShake 15.12:
- Data Types/Formats: seismograms for 5 sites (LADT, STNI, WNGC, PAS, SBSM) to 10Hz
- Data Set Size: 500GB
- Data Access Methods: SQL Queries, Python/C Code, USC Acct Required