



Seismological Laboratory,
California Institute of Technology
Pasadena, CA 91125

<https://scedc.caltech.edu> doi: 10.7909/C3WD3xH1
<http://scsn.org> doi: 10.7914/SN/CI



southern california

earthquake data center



Cloud Computing and Big Data – Using the Southern California Earthquake Data Center (SCEDC) and Southern California Seismic Network (SCSN) for Earthquake Research

Ellen Yu, Prabha Acharya, Aparna Bhaskaran, Shang-Lin Chen, Jennifer Andrews, Valerie Thomas, Egill Hauksson and Robert Clayton

The Southern California Seismic Network is **one of the largest regional seismic networks in the United States**. The **Southern California Earthquake Data Center is the SCSN data archive**. SCSN/SCEDC operations generate products vital to emergency response and earthquake research such as event catalogs, focal mechanisms, moment tensors, ShakeMaps and Recent Earthquake Maps.

Cloud Computing & Your Research

The SCEDC is developing a pilot project to host its 100 TB repository of continuous and triggered waveforms in a publicly available cloud archive using Amazon S3 storage.



Why? Spend less time preparing the data so you have more time to analyze it.

Get data faster

- Higher download speeds. Users can spin up EC2 instances configured to have up to 25 Gbps of bandwidth.

Process data faster

- Process data in parallel.
- Make the data available to GPUs or to other high performance, large scale resources.

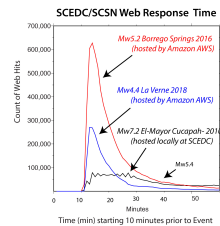


Figure 1. Plot of SCEDC web response for the 2010 Mw 7.2 El Mayor Cucapah event (black), the 2018 Mw 5.4 Barro Colorado event (red), the 2018 Mw 4.4 La Verne event (blue). Hosting in AWS S3 has proven effective in distributing information to users efficiently.

Customize your computational needs for what your research needs now. Don't be limited by old hardware. Don't stress out trying to predict what you might need 2 years from now when purchasing hardware. With cloud computing users only pay for resources used.

We Need Your Input

Share with us common use cases for how you download waveforms and process them.

Do you do any convert to other formats? (ex – SAC, ASCII)
Do you do any filtering? Windowing?
Do you have previous experience with AWS or other cloud providers?

Send us feedback at scedc@gps.caltech.edu

The SCEDC staff will take this feedback to develop documentation and tutorials, software libraries, and tools such as docker images, and Amazon Machine Images (AMIs) that will allow the users to start analyzing data with minimal training or software development.

New Datasets

Training Sets for Deep Learning

<http://scedc.caltech.edu/research-tools/deeplearning.html>

The SCEDC is hosting supplemental material that was used in the research publications below. Questions about these files should be directed to the first author.

P Wave Picking and First Motion Polarity

hdf5 files of the training and validation data sets

Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018).

P-wave arrival picking and first-motion polarity determination with deep learning. J. Geophys. Res. Solid Earth, 123, 6, 5120-5129, <https://doi.org/10.1029/2017JB015251>.

Generalized Seismic Phase Detection

The training dataset contains 4.5 million seismograms evenly split between p-waves, s-waves, and pre-event noise classes. The authors encourage the use of this dataset for training deep learning models and use as a benchmark.

Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). *P-wave arrival picking and first-motion polarity determination with deep learning*. J. Geophys. Res. Solid Earth, 123, 6, 5120-5129, <https://doi.org/10.1029/2017JB015251>

Alternate Catalogs

The SCEDC also distributes Southern California earthquake catalogs where location, magnitude, and mechanisms are derived from different methods than routine SCSN processing. The standard SCSN catalog does not account for local variations in the seismic velocity structure, which limits the accuracy of its event locations. In addition to the STP client, SCEDC alternate location catalogs are also available via FDSN web services at <https://service.scedc.caltech.edu/fdsnws/event/1/>

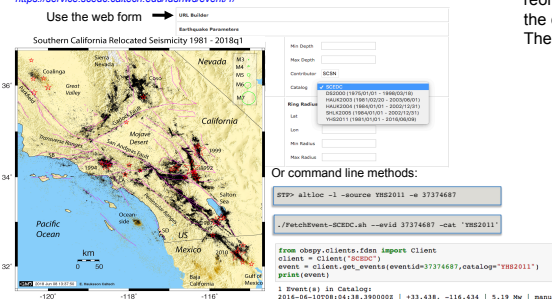


Figure 2. Waveform relocated southern California seismicity from 1981 to April 2018 (black color) and 3D model only relocated events (tan color). Red stars are events of M>5.5. Data recorded by SCSN and distributed by SCEDC.

New Access Tools

New Web Services: Data Availability

<https://service.scedc.caltech.edu/scedcws/availability/1/>

The SCEDC is developing web services that serve out information on what time spans of channels are available in the archive for both continuous and triggered data. This web service is compatible to the data availability service at IRIS.

In Beta Testing:

<https://service.scedc.caltech.edu/scedcws/availability/1/extent>

Returns lists of available time extents (earliest to latest) for selected channels and time ranges.

Future development:

We are currently developing a /query endpoint as well which would produce lists of contiguous time spans for selected channels. The web service will also be integrated with the SCEDC clickable map. This will allow users to see availability geographically.

The Future of SEED?

The SCEDC is considering discontinuing distributing full SEED. Web services now exist that can produce components of this format. **Do you use full SEED? We want to talk to you!** scedc@gps.caltech.edu

Improving Download Speeds

The SCEDC staff has been analyzing ways to improve download speeds from its FDSN web services. A major finding is that by reorganizing data in order of the seismic channel (how users retrieve the data) improves download speeds by a factor of 4. The SCEDC is in the process of implementing this reorganization.

About the SCEDC Archive

Currently the SCEDC archives continuous and triggered seismic waveforms from 528 SCSN recorded stations. The SCEDC catalog of earthquakes detected by SCSN spans 1932-present. On average SCSN processes on average 16,000 events a year. The archive provides a valuable resource for research topics benefiting from large datasets such as template matching, noise correlation, and earthquake detection algorithms to name a few. The SCEDC releases updates to metadata for stations operated by SCSN within 24 hours of the field operation in FDSN StationXML and dataless SEED formats.

To access SCEDC web services by ObsPy:

```
from obspy.clients.fdsn import Client
Client = Client("SCEDC")
```

Acknowledgements

SCEDC is supported by the Southern California Earthquake Center (SCEC) and the USGS/ANSS.

SCSN is supported by the USGS/ANSS and California Office of Emergency Services (Cal OES)

Citing the Data

Please cite SCEDC doi:10.7909/C3WD3xH1 and SCSN doi:10.7914/SN/CI when using SCSN data downloaded from the SCEDC