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

# 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, Zachary Ross, Egill Hauksson and Robert Clayton

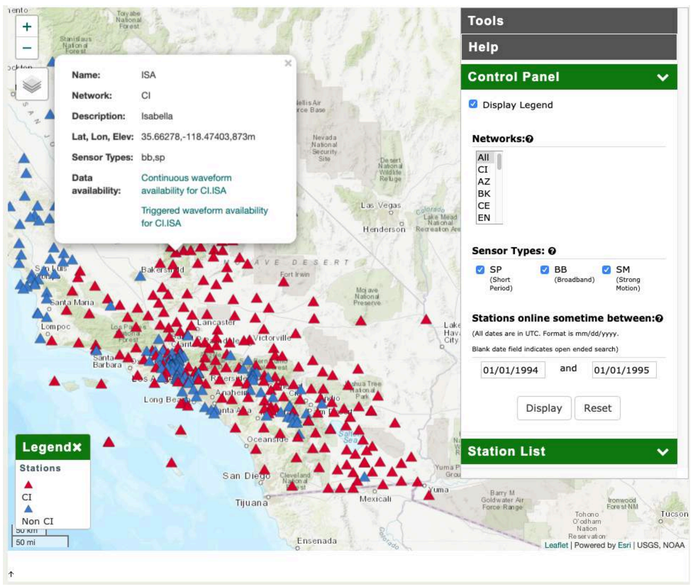
SCEC 2019 Annual Meeting; poster 301

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.

## *New Tools and Services*

### Find what data is available at the SCEDC using web services or map

*http://service.scedc.caltech.edu/SCSNStationMap/station.html*




### Data Availability Web Service

*https://service.scedc.caltech.edu/scedcws/availability/1/*

Shows what time spans of channels are available in the archive for both continuous and triggered data.
*/extent* method: earliest and latest entries
*/query* method: continuous time spans

## *New Researcher Provided Datasets*

### Quake Template Matching (QTM) Catalog

*https://scedc.caltech.edu/research-tools/QTMcatalog.html*
Catalog spans 2008-2017 (Ross et. al., Science 2019)

### Double Difference and Focal Mechanism Catalogs

*https://scedc.caltech.edu/research-tools/altcatalogs.html*
Catalog spans 1981-2018 (Hauksson et.al 2011; Yang et. al. 2011)

### Training Sets for Deep Learning

*https://scedc.caltech.edu/research-tools/deeplearning.html*
P Wave Picking and First Motion Polarity (Ross et.al JGR 2018)
Generalized Seismic Phase Detection (Ross et al, BSSA 2018)
Seismic Signal/Noise Discrimination (Meier et al, JGR 2018)

### What's there?

- ~20 years of continuous waveform data. (Broadband 1,20,40 sps); 10 years 100 sps
- Day long files of single seismic channels
- miniSEED format

### I'm New to AWS. How Do I Get Started?

Tutorials, including simple demos, Docker and Amazon Machine Images.
*https://scedc.caltech.edu/cloud/*

### Want to see a real life example?

See: Clements and Denolle "Cactus to Clouds: Processing the SCEDC Open Data Set on AWS" Poster #302

# SCEDC WAVEFORM ARCHIVE NOW IN AWS CLOUD

# s3://scedc-pds us-west-2

### What is the cost?

The archive is part of the AWS Public Dataset Program. This means there is no cost to download data. You will need an AWS account which is also free.
Whether you will be charged will depend what resources you use, but the AWS Free Tier that comes with each AWS account provides much opportunity for experimentation with little or no cost.

**Example: Decimate a month in 11 minutes for $0.68**
**Task:** Decimate Jan 2016 40 sps to 10 sps. (122GB) and write results into S3. (27GB)
**How:** Deployed and ran AWS Lambda functions in parallel containing ObsPy decimate function.

**Total time and cost: 11 minutes and $0.68**
Cost Breakdown:

| | |
|---|---|
| I/O: | $0.11 |
| Storage: | $0.57 (includes 5GB-mo free tier) |
| CPU: | Free Tier |

### Why Cloud Computing?

- ***Increased performance and scalability***: Amazon's infrastructure provides higher download rates.
- ***Spend more time analyzing data, not retrieving it***: Previously, users needed to download a copy of the data before beginning analysis. Now you can analyze in the cloud, and remove that time consuming, download step.
- ***Computer resources customized to research, not the other way around:*** Use exactly what resources are required, for as long as required. Stop being a slave to the hardware life cycle!
- ***Lower the barriers for investigation***: Ever thought of of a research question, but didn't bother because it was too onerous finding a server or storage space? Cloud computing lowers or removes that hurdle.

## 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 and s3://scedc-pds

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)