12th National Conference on Earthquake Engineering
Salt Lake City, Utah
27 June - 1 July 2022
Hosted by the Earthquake Engineering Research Institute

Modernization of Data Processing and Review at the U.S. Geological Survey's National Strong Motion Project (NSMP)

J. Steidl[1], P. Hegarty[2], L. Schleicher[3], J. Brody[4] and L. Gee[5]

ABSTRACT

The USGS has a long history of providing processed strong motion (SM) data to the engineering community for use in designing earthquake-resistant structures to reduce property loss and casualties from future earthquakes. These recordings are also critical for improving our understanding of the earthquake source, the generation and propagation of ground motions, and the performance of the built environment when subject to these motions. In the four years since the 11NCEE, significant changes have taken place within the NSMP with respect to processing and disseminating strong motion data. Prior to the 11NCEE, the processing of most SM data from the NSMP was performed manually. Given the establishment of the Advanced National Seismic System (ANSS), followed by the Earthquake Early Warning system, the increase in the number of SM stations in the first two decades of the new millennia made manual processing of SM data untenable. This manuscript describes the modernization of the NSMP processing workflow to generate the Consortium of Organizations for Strong Motion Observation Systems (COSMOS) V0 through V3 data products and make them available to the engineering community at the Center for Engineering Strong Motion Data (CESMD).

Introduction

Prior to the 11NCEE the NSMP processing of strong motion data was performed manually, where depending on the magnitude and number of stations, could take days or weeks to complete. A new software package was developed at the USGS Earthquake Science Center (ESC) to facilitate automated data processing. The software [1] is called **P**rocessing and **R**eview **I**nterface for **S**trong **M**otion Data (PRISM). PRISM takes raw acceleration time histories in digital counts (COSMOS V0 format) and automatically produces acceleration time histories in physical units (COSMOS V1 format), processed acceleration, velocity, and displacement time histories (COSMOS V2 format), and acceleration, velocity, and displacement response spectra at standard damping values (COSMOS V3 format) [2]. The PRISM software was designed with both a GUI interface for the desktop user, and a command line processing engine interface for programmatic use when dealing with large amounts of data. This allows even large magnitude events with 100's of stations to be processed within minutes to hours. Once data has been processed by PRISM, it can be reviewed and archived at the CESMD [3].

[1] Earth Research Institute, University of California, Santa Barbara, CA (email: steidl@ucsb.edu)
[2] Earth Research Institute, University of California, Santa Barbara, CA (email: paul.hegarty@ucsb.edu)
[3] Earthquake Science Center, U.S. Geological Survey, Moffett Field, CA (email: lschleicher@usgs.gov)
[4] Earthquake Science Center, U.S. Geological Survey, Moffett Field, CA (email: jbrody@usgs.gov)
[5] Earthquake Science Center, U.S. Geological Survey, Moffett Field, CA (email: lgee@usgs.gov)

Steidl J, Hegarty P, Schleicher L, Brody J., and Gee L. Modernization of Data Processing and Review at the U.S. Geological Survey's National Strong Motion Project (NSMP). *Proceedings of the 12th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Salt Lake City, UT. 2022.

NSMP/AQMS Automated System for Data Processing and Review

In 2018 a workflow was designed to integrate the PRISM processing engine within the ANSS Quake Monitoring System (AQMS). As event notifications are distributed by the Product Distribution Layer (PDL) of the AQMS system, if the events are large enough and close enough to strong motion stations of the ANSS, the data (and metadata) from these stations are automatically acquired from waveservers (and metadata information systems) and converted to COSMOS V0 format. These waveservers are located locally within the NSMP, at Regional Seismic Network (RSN), at the Incorporated Research Institutions for Seismology (IRIS) Data Management Center (DMC), or could be at any location that is known and network accessible globally. Typically, this data is in raw miniseed format, which contains very little metadata, so it is combined with metadata from station information systems and converted into COSMOS V0 raw data. At this point, the V0 data is handed off to the PRISM processing engine, which then automatically generates the COSMOS V1, V2, and V3 data products and can be automatically uploaded to the CESMD (strongmotioncenter.org). The mechanism and workflow for this automatic upload of data is built, but currently an additional human review step occurs prior to upload and dissemination. With the PRISM software the NSMP is now processing SM data from many ANSS partner networks as well as some international partners, so a web-based tool for rapid review of these products was developed to ensure that ground motion values made sense when comparing other stations at similar distances, and that metadata parameters appeared appropriate so outliers could be excluded from dissemination for further follow up with the operating network. The complete workflow currently in place is shown in Figure 1. We anticipate that in the future the review step will only be triggered for previously flagged stations, and stations on a "whitelist" will automatically make it to the CESMD without needing human review.

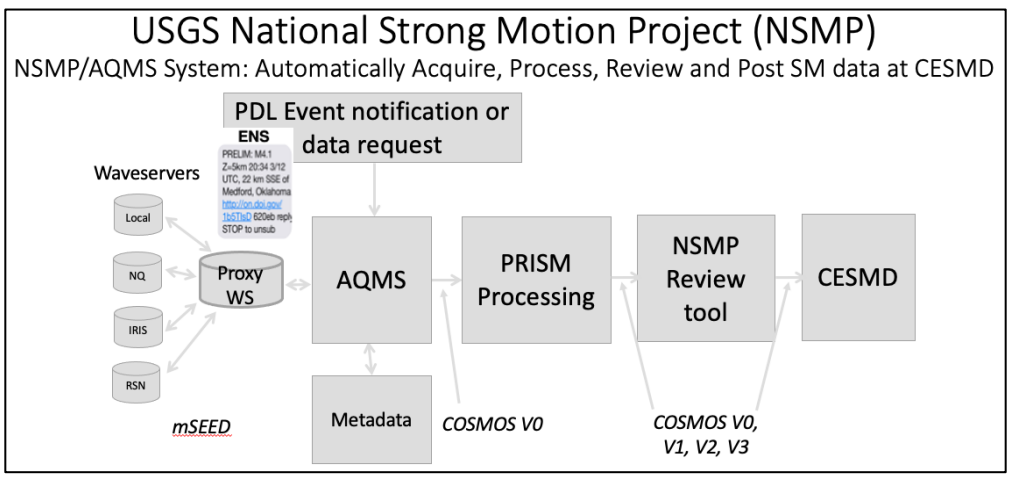


Figure 1. The USGS/NSMP system automatically collects ANSS strong motion data from multiple sources and processes it to produce COSMOS data products for distribution at the CESMD.

The web-based review page for each event has a header containing earthquake information at the top of the page along with the processing history, links to the processing log file, and links to NEIC ShakeMap event pages. This page provides a tabular view of the acceleration data for every channel and the ability to review critical metadata parameters. There are also links to plots of the acceleration, velocity, displacement, and response spectra, and at the bottom of the page is an interactive plot showing Peak Ground Acceleration (PGA) for every channel vs. distance. An example of this webpage is shown in Figure 2 for the M4.8 Maria Antonia, Puerto Rico earthquake on 18 July 2021.

The review webpage is customizable, so that the analyst can select the number of channels to display per page, and then step through the pages. In figure 2, the first 12 channels (rows) of 142 total channels of data are shown. The table is interactive, clicking on column headers enables sorting data on that column, while clicking on any of the metadata column parameters will produce a pop-up of the COSMOS file header (Figure 3).

COMCAT event: pr2021199004 [pt21199050,us7000en2d,pr2021199004] M 4.5 - 6 km WSW of La Parguera, Puerto Rico @2021-07-18 03:52:25
COSMOS event: pr2021199004 (M4.4 @2021/07/18 03:52:25 UTC La Parguera, 07 Earthquake of)

ShakeMap links: *PGA Shake Map* *Analysis Shake Map*

Process history: `CosReview: 142 V2, 0 skipped @Mon, 19 Jul 2021 10:46:20 -0700`

Prism/CosReview Log

Channel Table Control

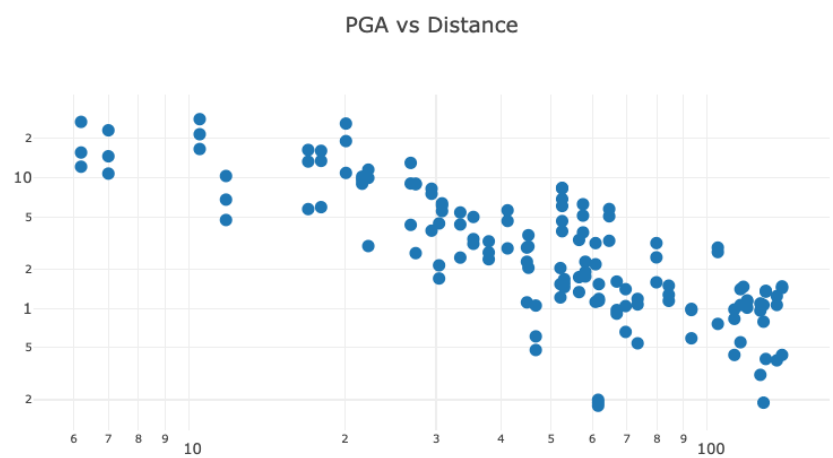
Enable REGEX SCNL filter

Page « 1 » of 12, with 12 rows per page

Review-user Control

Username: Password: Login

SCNL	Station_name	pgav2	orient	daq	sensor	v2filter	v2pga	epidist	Exclude	Reviewed	cesmd
MLPR.HN1.PR.00	MLPR	-0.015816	360deg	--- 2.50uV 20.97V	5.00v/g -999.00g	0.30->40.0Hz	-15.51	6.2			x
MLPR.HN2.PR.00	MLPR	0.027115	90deg	--- 2.50uV 20.97V	5.00v/g -999.00g	0.30->40.0Hz	26.59	6.2			x
MLPR.HNZ.PR.00	MLPR	-0.012317	Up	--- 2.50uV 20.97V	5.00v/g -999.00g	0.30->40.0Hz	-12.08	6.2			x
CRPR.HN1.PR.00	CRPR	-0.010907	14deg	--- 2.50uV 20.97V	5.00v/g -999.00g	0.30->40.0Hz	-10.7	7			x
CRPR.HN2.PR.00	CRPR	0.023383	104deg	--- 2.50uV 20.97V	5.00v/g -999.00g	0.30->40.0Hz	22.93	7			x
CRPR.HNZ.PR.00	CRPR	-0.014801	Up	--- 2.50uV 20.97V	5.00v/g -999.00g	0.30->40.0Hz	-14.52	7			x
PR01.HN1.GS.20	PR01, Lajas	0.016819	360deg	--- 1.59uV 13.33V	4.96v/g -999.00g	0.30->40.0Hz	16.49	10.5			x
PR01.HN2.GS.20	PR01, Lajas	0.028429	90deg	--- 1.59uV 13.33V	4.96v/g -999.00g	0.30->40.0Hz	27.88	10.5			x
PR01.HNZ.GS.20	PR01, Lajas	-0.021808	Up	--- 1.59uV 13.33V	4.96v/g -999.00g	0.30->40.0Hz	-21.39	10.5			x
LJS1.HNE.PR.--	Lajas OMME	0.010462	90deg	--- 0.30uV 2.50V	0.62v/g -999.00g	0.30->40.0Hz	10.26	11.8			x
LJS1.HNN.PR.--	Lajas OMME	-0.006899	360deg	--- 0.30uV 2.50V	0.62v/g -999.00g	0.30->40.0Hz	-6.77	11.8			x
LJS1.HNZ.PR.--	Lajas OMME	-0.004823	Up	--- 0.30uV 2.50V	0.62v/g -999.00g	0.30->40.0Hz	-4.73	11.8			x



Notes:

- 'daq' column is the one-sided full-scale input of the datalogger (rel#23), bit-weight, one-sided full-scale assuming 24bit
- 'sensor' column is the sensor sensitivity (rel#42), one-sided sensor max output (rel#44)
- The v2PGA column is the PGA vale from the V2 real-hdr#64, the units are probably cm/s/s but could be something else (int-hdr#3)
- Values in the 'orient' column with the "ref" instead of "deg" means the COSMOS header had a Relative rather than True azimith value
- The column sort on any numerical colum is based on absolute value
- I beliveve the apktable PGA values are in 'G'
- The PGAV2 column value is followed by (refOK) if its within 5% of 'reference' APKtable values
- The PGAV2 column value is followed by (vs <ref value>) if its not within 5% of previous APKtable file values
- The PGAV2 column value is followed by nothing if there was no entry in previous APKtable file values

- CESMD column: the "x" shows up If/when data is available at cesmd, click box to exclude from cesmdZip, purple shade means someone else (CGS) uploaded

Figure 2. Example of the AQMS/NSMP event review webpage for the M4.5 Maria Antonia, Puerto Rico earthquake of 7/18/2021.

Clicking on the station "SNCL" value will produce a pop-up plot of the full waveform and clicking on the station name will provide a downloadable 2-page pdf of the three-component plot produced for the CESMD website. This pdf plot shows zoomed in Acceleration, Velocity, Displacement, and Fourier acceleration spectra. The detailed information shown on this review page provides the analyst with the ability to verify that

the data looks acceptable, its ground motions are reasonable when compared to the ensemble, and the metadata values for the sensors and dataloggers make sense given the known sensor and datalogger information provided in the COSMOS header. Once an RSN's data and metadata, such as the PR and temporary GS stations in Puerto Rico, have been carefully reviewed, review of future events from the RSN is much faster.

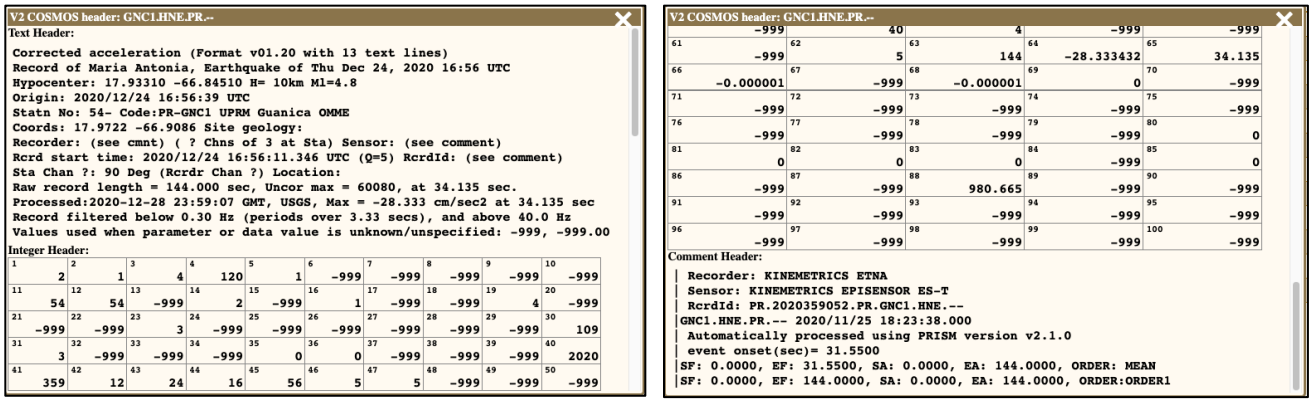


Figure 3. Review page pop-up window when clicking on metadata values. The left image shows the text and top of the integer header values in COSMOS V2 file. Scrolling down to the bottom of the pop-up window (right image) shows the bottom of the real header, and the comment header, which provide some of the processing history details as well as instrumentation types. Holding the cursor over the header number produces a pop-up description of the header field.

Increase in Data Availability at CESMD

The result of the development work on the review system described above has been a significant increase in the amount of data at the CESMD. In particular, for recent larger events and their aftershock sequences, in addition to the mainshock data, it's now common to find aftershock data down to M4.0-4.5. If one looks at the number of records available at the Engineering Data Center (EDC) search engine at CESMD over the time period from 2010 through 2021 [2], the effect of this new workflow and web-based review tools is clear. While the Anchorage and Ridgecrest sequences did produce many records and are partially responsible for this increase, the trend continues into 2021. By mid-2020 the review tools reached a steady state and the processing and review has now become routine. The NSMP is now able to keep up with the rate at which the AQMS is pushing event data through the PRISM software, significantly increasing the speed at which data is available and the amount of data uploaded due to the increase in the number of events now being processed. The addition of RSN data that had not previously been included in CESMD is also significant. A highlight of some of this newly available data is the inclusion of many processed records of aftershocks that in the past would not be analyzed due to the time needed in manual review. These aftershock records contain data from rapid aftershock mobilization projects (RAMP) that in many cases are now integrated into the RSN's in real-time. There have been many near-source large PGA recordings from aftershocks recorded at RAMP stations now available to the NSMP/AQMS data processing flow. These records include motions greater than 0.5 g from the Puerto Rico and Magna, UT, events, and some as large as 1g at stations within 10 km of the Monte Cristo and Mina, NV, events. Additional planned efforts include incorporating lower magnitude event data at CESMD [4] and incorporating NSMP's compilation of near-surface characteristics, including surface geology and the time-averaged shear-wave velocity in the upper 30 m, or V_{S30}, at all ANSS accelerometer sites [5-6].

Conclusions

The NSMP's modernized workflow has significantly increased the speed and the number of waveform records available at CESMD in recent years since the 11NCEE. All waveform products are publicly available for download at strongmotioncenter.org and we encourage their use by the earthquake engineering community.

References

1. Jones, J., Kalkan, E. and Stephens, C. (2017). Processing and Review Interface for Strong Motion Data (PRISM)—Methodology and Automated Processing, Version 1.0.0: U.S. Geological Survey Open-File Report, 2017–1008, 81 p.
2. COSMOS Strong Motion Programs Board (2001). COSMOS Strong Motion Data Format, v. 1.20, https://www.strongmotioncenter.org/NCESMD/reports/cosmos_format_1_20.pdf
3. Hagos L., Haddadi H, Schleicher L., Steidl J, Gee L, Dhar, M. Updates on the Center for Engineering Strong-Motion Data (CESMD) *Proceedings of the 12th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Salt Lake City, UT. 2022.
4. Aagaard, B.T., Wald, D.J., Thompson, E.M., Hearne, M., and Schleicher, L.S. Improving the Development Pipelines for USGS Earthquake Hazards Program Real-Time and Scenario Products *Proceedings of the 12th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Salt Lake City, UT. 2022.
5. Huddleston, G.J., Schleicher, L.S., Knudsen, K., Gee, L., Steidl, J.E. Compilation of Geologic and Seismic Velocity Characteristics at Advanced National Seismic System Strong Motion Accelerometer Sites *U.S. Geological Survey Data release*, https://doi.org/10.5066/P9YKIUW2.
6. Huddleston, G.H., Schleicher, L.S., Blair, L., Knudsen, K., Gee, L., and Steidl, J.H. Interfacing the National Strong Motion Project (NSMP) Near-surface Site Characteristic Data Compilation at Advanced National Seismic System Accelerometer Sites with Strong-Motion Waveform Datasets, *Proceedings of the 12th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Salt Lake City, UT. 2022.