Cal Poly

Caltech

SCEC
AN NSF+USGS CENTER

UC Irvine

UCLA

UC Santa Barbara

USC

Usability of Ground Motions Recorded by Community Seismic Network

Jonathan P. Stewart, Shako A. Mohammad, Tristan E. Buckreis, and Yousef Bozorgnia
UCLA Samueli Engineering

Chukwuebuka C. Nweke and Rashid Shams
USC Civil & Environmental Engineering

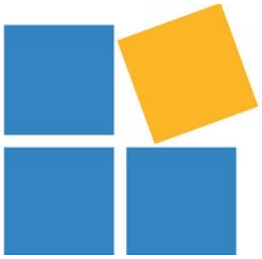
Monica D. Kohler
California Institute of Technology

A report on collaborative research by the University of California, Los Angeles, University of Southern California, and California Institute of Technology that was supported by the California Strong Motion Instrumentation Program under contract number 1021-006.

Report GIRS-2023-08

DOI: https://doi.org/10.34948/N36K5M

University of California, Los Angeles (headquarters)



Natural Hazards Risk & Resiliency Research Center

B. John Garrick Institute for the Risk Sciences

Usability of Ground Motions Recorded by Community Seismic Network

A report on collaborative research by the University of California, Los Angeles, University of Southern California, and California Institute of Technology that was supported by the California Strong Motion Instrumentation Program under contract number 1021-006.

Jonathan P. Stewart, PI, University of California, Los Angeles

Shako Mohammed, Doctoral Candidate, University of California, Los Angeles

Chukwuebuka C. Nweke, co-PI, University of Southern California

Rashid Shams, Graduate Student, University of Southern California

Tristan E. Buckreis, Postdoctoral Scholar, University of California, Los Angeles

Monica D. Kohler, co-PI, California Institute of Technology, Pasadena, CA

Yousef Bozorgnia, co-PI, University of California, Los Angeles

Natural Hazards Risk and Resilience Research Center

B. John Garrick Institute for Risk Sciences

UCLA (headquarters)

Report GIRS 2023-08

DOI: https://doi.org/10.34948/N36K5M

Original Version: October 13, 2023

Revised: February 20, 2024

ABSTRACT

Among the global regions that experience shallow crustal earthquakes in active tectonic regions, southern California is among the most seismically active and most densely instrumented. Networks currently operating in southern California with relatively modern (new, or with recently updated) sensors include the California Strong Motion Instrumentation Program (CSMIP), United States Geological Survey (USGS), and Southern California Seismic Network (SCSN). Research with substantial engineering impact, such as the NGA-West projects, has made extensive use of this data. A major data source in urban Los Angeles that has not been considered in these projects is the Community Seismic Network (CSN), which utilizes low-cost, Micro-electro-mechanical system (MEMS) sensors and cloud-based communications (Clayton et al. 2011, 2020). Because of their low cost, CSN stations are deployed at much higher density (i.e., smaller average station-to-station spacing) than stations for other networks. CSN data are publicly available for significant felt earthquakes in southern California. This dense network of ground motions provides enticing opportunities for investigations of spatially variable ground motions and variations in source-to-site path and site response effects over short length scales.

We have downloaded and processed CSN data for 29 earthquakes with **M** > 4 from 2012 to 2023. This processing includes automated and visual checkers for data usability – for example, records for which seismic wave arrivals cannot be readily distinguished from pre-event noise are, for practical purposes, not usable. Records that are usable are processed using an adaptation of standard procedures in NGA projects in the USGS open-source program gmprocess (Hearne et al., 2019; Ramos-Sepulveda et al. 2023). The adaptation allows for both signal-to-noise ratio and displacement wobble to be considered in the selection of high-pass corner frequencies (f_{cHP}). To a greater extent than is typical with other modern networks, low-pass filters (with corner frequencies denoted f_{cLP}) are also applied due relatively low signal-to-noise ratios at high frequencies. Based on this processing, each record is classified as one of the following: (1) Broadband Record (BBR) – relatively broad usable frequency range from $f_{cHP} < 0.5$ Hz to $f_{cLP} > 10$ Hz; (2) Narrowband Record (NBR) – limited usable frequency range relative to those for BBR; and (3) Rejected Record (REJ) – visual evidence suggests seismic waves cannot be distinguished from noise.

We examine distributions of component-specific peak accelerations for BBR and REJ records and find that the two data sets are distinguished at a PGA of 0.005g, which correctly classifies record components 93% of the time. The variations in this threshold acceleration by event are nominal. This implies the CSN recordings at lower acceleration levels are likely to be noise-dominated, where "noise" is a combination of environmental and anthropogenic ambient ground vibrations, as well as instrument sources. The BBR / REJ threshold shaking level significantly exceeds the electronic instrument noise for the MEMS sensors (estimated as 0.00028 g), suggesting that the effective noise is influenced by ambient ground vibrations. We also examine the threshold between usable signals and pre-event noise and find that the two data sets are distinguished at a PGA of

0.0013 g, which correctly classifies record components 79% of the time. We examined time- and location-dependence of noise levels from CSN data using REJ records and pre-event noise. The results indicate nominally higher noise levels in areas of high population density (≥ 20 people/100m^2), typical of urban areas, and lower noise levels by about a factor of 1.5 in low population density areas. We find inconclusive temporal variations, following an investigation of noise-dependence on the time of day of events. We had anticipated lower noise levels in early morning hours and higher noise levels in afternoon to early evening hours, but the limited data do not reveal clear trends. Nonetheless, spatial and temporal analyses, collectively, suggest that some refinements to the BBR / REJ threshold may be possible, depending mainly on location.

If we apply the 0.0015 g threshold, the limiting distances for medium-stiff soil site (V_{S30} = 400 m/s) based on the expected ground motion level at the 5th-percentile are 89, 210, 280, and 370 km for **M** 5, 6, 7, and 8 events, respectively. CSN records at distances beyond these thresholds have a potential for under-measurement bias and likely should not be applied in ground motion modeling studies.

We compare recordings from co-located (within about 10m) and proximate (within 3 km) CSN and non-CSN (generally SCSN or CSMIP) stations. The proximate pairs are screened to only include cases of similar surface geology. These results show that for commonly recorded events, high- to medium-frequencies of ground motions (i.e., PGA and Sa for T < 5 sec) are similar for CSN BBR and non-CSN ground motions, whereas CSN NBR ground motions are biased slightly low (approximately -0.1 natural log units) for a range of periods. This indicates that the relatively limited frequency ranges of NBRs produce slightly biased ground motions, at least as represented by conventional intensity measures.

Based on these findings, BBR and NBR ground motions from the CSN networks have been incorporated into the ground motion database being assembled for the ongoing NGA-West3 project. Frequency ranges for which the data accurately capture the seismic signals are reported. We encourage users to screen the data appropriately for their applications to ensure that data selected for a given application is usable over the frequency range of interest.

ACKNOWLEDGEMENTS

Funding for this study was provided by the California Strong Motion Instrumentation Program under contract number 1021-006. Partial support for the second and fourth authors was also provided by the UCLA and USC Civil & Environmental Engineering Departments, respectively. This support is gratefully acknowledged. The work presented here represents the views and opinions of the authors and does not reflect the policy, expressed or implied, of the State of California.

Helpful input was received during the project from Eric Thompson, Scott Brandenberg, and Maria Ramos-Sepulveda.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

The Community Seismic Network (CSN) is a network currently with over 800 three-component seismic stations, mainly in southern California (Clayton et al. 2011, 2020; http://csn.caltech.edu/), which are operated as a collaborative research effort between Caltech and UCLA. CSN utilizes low-cost, three-component, MEMS accelerometers capable of recording accelerations up to twice the level of gravity. The primary product of the network is measurements of shaking of the ground as well as upper floors in buildings, in the seconds during and following a major earthquake.

In terms of its layout and configuration, CSN differs from other seismic networks in two principal respects. First, the sensors are spatially concentrated in certain parts of southern California. The vast majority of CSN stations are in the San Fernando Valley, Pasadena, San Gabriel Valley, downtown Los Angeles, Hollywood, and South Los Angeles; many of these areas have high densities of population or industrial activity and hence are culturally noisy. This noise takes the form of background ambient vibrations that have been found to depend on time (i.e., time of day, day of week, and season; Clayton et al. 2020). As a consequence, as currently configured the CSN is relatively ineffective for some classical applications like earthquake location for distant events or recording motions over a wide distance range, but they are effective at capturing ground motion characteristics over relatively short length scales (i.e., in close proximity to each other). Second, the instruments have relatively high noise levels compared to broadband seismometers or modern accelerometers.

The effective noise levels of CSN instruments is important because recorded earthquake ground motions are subject to sampling errors for low amplitudes. In the case of triggered instruments, sampling errors occur when the ground shaking level at a site falls below the trigger threshold. In the case of continuously recording instruments, sampling errors occur when signal amplitudes are not stronger than the instrument noise threshold. This is typically the case at large distances, and is more pronounced for small magnitude events than large magnitude events. For a magnitude-distance condition where the mean ground motion amplitude is near the threshold, unusually strong motions that exceed trigger thresholds or that fall above the noise floor are recorded. However, weaker motions that do not exceed trigger thresholds or that fall near the noise floor are not available. Accordingly, the problem is not that no records are obtained for such conditions, but that the recorded ground motions, and GMMs derived from those motions, become biased at large distances towards larger values.

The objectives of this study were to evaluate the effective noise threshold of CSN data based on the currently available recordings, to validate the recordings against those from higher-resolution sensors, and to make available in a public database CSN data that is judged to be reliable along with its associated metadata. Two types of noise thresholds are considered: (1) threshold between

clear earthquake motions and signals recorded during earthquakes but for which no typical earthquake characteristics are visually apparent and (2) threshold between earthquake signals and pre-event ambient vibrations. The ambient ground vibrations considered in these analyses represent the combined effects of environmental and cultural sources (geological/geotechnical, traffic, construction, meteorological, seasonal, motors, human activity), and hardware-related sources (electronic, sensor and digitizer resolution, power). The thresholds presented here are averaged over all or portions of the network and may differ from the effective thresholds located in a specific, small geographical region, and for an earthquake occurring at a particular time.

Chapter 2 of this report provides further background information on the CSN, the data produced by the network, and the events considered in this study. Additional chapters reflect the scope of work that was undertaken to meet project objectives, including data processing and assignment of classes that indicate record quality (Chapter 3), comparisons of CSN data to data from other networks (Chapter 4), analysis of noise recordings from CSN sensors to evaluate spatial and temporal variations as well as the amplitude threshold separating usable from noise-dominated records (Chapter 5), and identification of usable distance ranges for CSN data (Chapter 6).

2 CSN Overview

2.1 CSN INSTRUMENTS AND HOUSING

2.1.1 Station Distribution

Over the duration of the current project, the Community Seismic Network (CSN) comprised 769 seismic station locations, most of which are in southern California (Clayton et al., 2020). In addition, there are 339 previously active but now decommissioned station locations, some of which produced data that are evaluated. Figure 2.1 shows the locations of CSN stations overlaid on a regional map that also shows stations from other regional networks (CSMIP, USGS, SCSN).

CSN utilizes low-cost, three-component, MEMS accelerometers. The primary product of the network is measurements of shaking of the ground as well as upper floors in buildings, in the seconds during and following a major earthquake. Each sensor uses a small, dedicated ARM processor computer running Linux, and analyzes time series data in real time at 250 samples per second (sps), which is then downsampled to 50 sps. Innovations in cloud computing for data processing, coupled with sensor developments for the video-gaming and automotive air bag industries, have helped form the technological basis of this network. Prior to ~2014, most CSN stations consisted of plug-in sensors that were attached to community hosts' laptops and desktop computers; the hosts determined the deployment location and coupling. Data from these deployments went into the early earthquake database, but this deployment type no longer exists. After 2014, all CSN sensors are stand-alone devices deployed by a CSN field engineer who determines location and physical coupling with the floor.

Some CSN station locations have multiple instruments. This occurs because of multiple instruments (referred to here as a "station") within a structure at different heights, and in some cases, different locations in plan at a given height. The number of three-component instruments is 1868, which includes 1250 ground stations, 27 basement stations, and 463 stations on floors of buildings above the ground line. The instrumented buildings have between 1 and 3 triaxial sensors deployed per floor. The sensing hardware and parameters are the same as for the free-field. There are no sensors on lifelines infrastructure at the present time.

In this project we have focused on ground and basement stations and do not consider above-ground stations. Each of the ground-level and basement stations has been assigned an instrument housing code using guidelines provided in Table 6 of COSMOS (2001). That table provides two main categories for classifying stations (free-field and structural or array stations); within each category a series of specific codes are provided. This information is provided as metadata accompanying the CSN sites in the ground motion database (Buckreis et al. 2023a). The applicable codes that were applied to CSN stations are as follows:

1. "04" - ground-floor in a 1-2 story building without a basement (1250 CSN stations)
2. "05" - ground-floor in a larger structure (118 CSN stations)
3. "09" - basement or underground in a large vault (27 CSN stations)
4. "10" - upper levels of a structure (463 CSN stations)

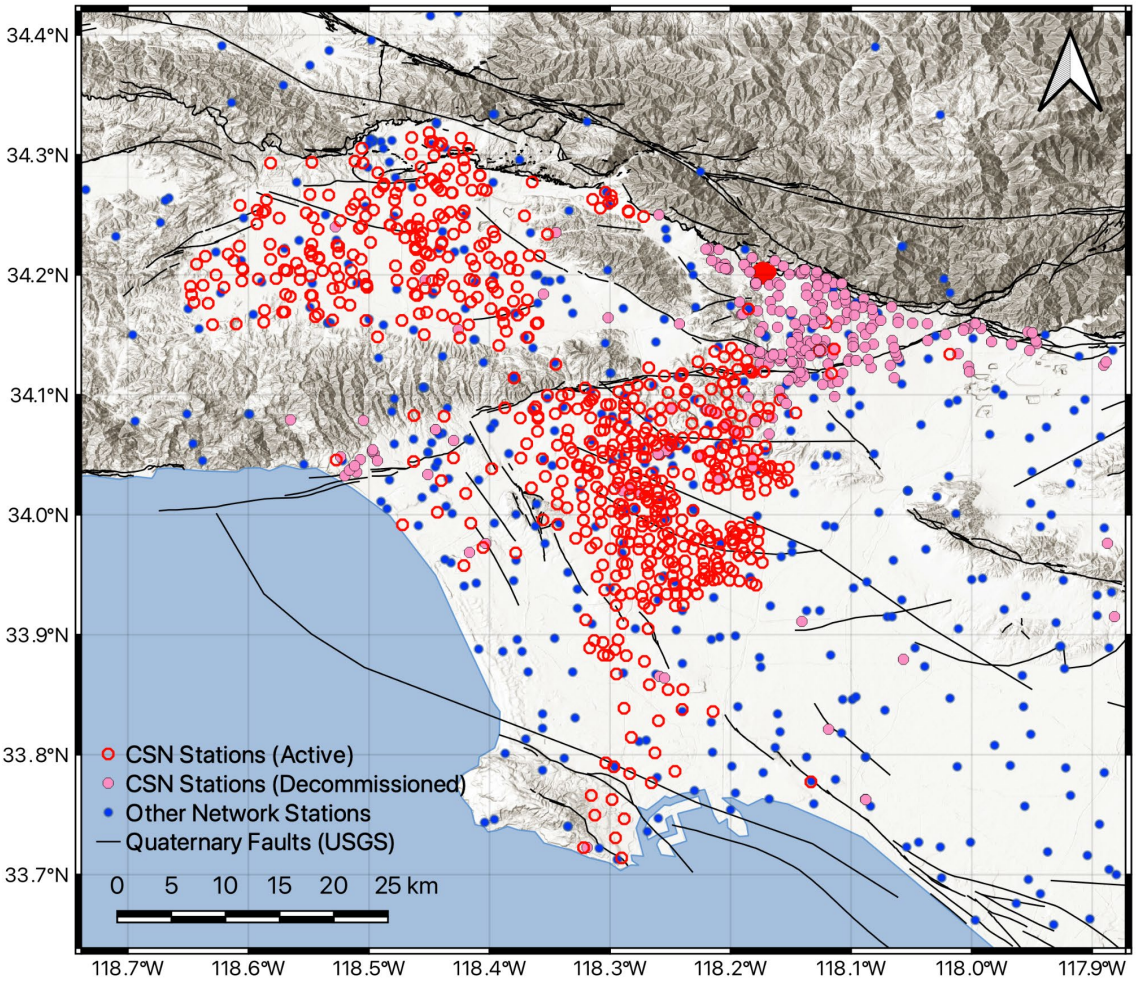


Figure 2.1. Map of southern California showing locations of ground motion stations considered in prior work (NGA-West2/Bozorgnia et al. 2014 & basin study by Nweke et al. 2022) (CSMIP, USGS, SCSN) and CSN stations (active and decommissioned) considered in this project.

Stations in group 04 can be considered "free-field." Stations in 05 and 09 might be approximated as free-field depending on the depth of embedment (for 09) and plan size of the structure (for 05). The difference between the 769 figure mentioned at the start of this section and the sum of 04, 05, and 09 is caused by the occurrence of multiple stations at a given site at the ground level or basement level.

Because this project is not only using the CSN in its current form, but also in the form it has had in the past, it is useful to examine the state of the network at different times. Figure 2.2 shows the

configuration of the network in 2013, 2015, and 2019. As shown in Figure 2.2, the initial network of sensors were deployed prior to 2015 and were primarily concentrated in the western-northwestern portion of the San Gabriel Basin (Pasadena & Altadena). All of these stations used early desktop "plug-in" versions of sensors. There were 191 instruments in operation during this period (2011 - 2015), with only 13 remaining in operation. All the other station locations from that period had subsequently been dropped from the network due to updates in sensor configuration or were misplaced due to homeowner turnover. Between 2016 and 2018, an additional 360 locations were instrumented in the Elysian Hills, Arroyo Seco, and downtown Los Angeles (DTLA). Of those 360 stations, 199 remain operational. Between 2018 and 2020, 309 locations were instrumented in the northern portion of the Los Angeles Basin (most of South Central Los Angeles and East Los Angeles via LAUSD), areas around Long beach, and parts of West Los Angeles (Marina Del Rey, Culver City and Westwood). All of these additional stations remain operational Current ongoing sensor deployments are expanding coverage to the San Fernando Basin while improving resolution in the Los Angeles Basin. At the beginning of the project the CSN comprised approximately 860 seismic stations (with both operational and decommissioned sensor locations across the existence of the CSN) in southern California (Clayton et al. 2020), however, the network has expanded and added 248 station locations to date, now exceeding 1100 seismic stations.

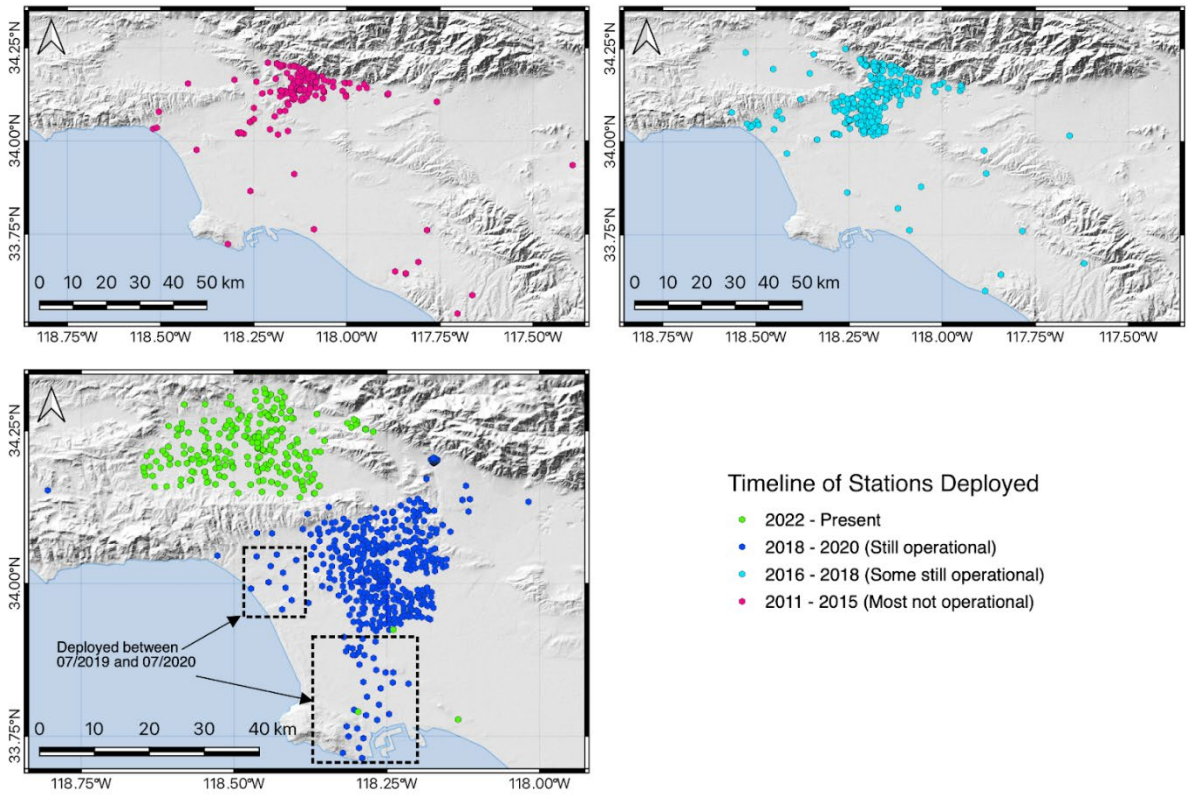


Figure 2.2. Maps showing the evolution and expansion of the CSN stations deployed in southern California.

2.1.2 MEMS Accelerometers

The MEMS accelerometers used by CSN provide 16 bits of resolution and are capable of recording accelerations up to ±2g. It is classified as a class-C accelerometer. The hardware components needed for the operation of one station includes a Phidget 1056-1 three-axis, class-C MEMS accelerometer integrated with a signal conditioner/digitizer on a single chip, a Linux micro-computer (Raspberry PI 4b or SheevaPlug), removable storage for the computer, an ethernet cord for internet connection, and power cables often combined with a small uninterruptible power source device to allow continuous operations during power outages. All the components are housed in a 10×10×23 cm orange box that is secured to the floor with two-sided tape pads.

Instrument responses for the accelerometers are technically flat to zero frequency (DC). Early measurements of noise levels recorded at a relatively quiet site in Pasadena (labeled "Phidget" in Figure 2.3) show a constant noise level from 0.01 to 1.0 Hz, with a slight upward ramp for higher frequencies. However, the actual usable frequency bandwidth is controlled by individual noise levels at each sensor location for the specific recorded earthquake. In the case of the USB-attached tri-axial accelerometers, the sensors sample at rates of order 1000 measurements per sec internally and present data at a rate of 250 triple floats over the USB. The computer then down-samples the data to the lower sample rates (50 samples per sec) as a configurable option. Additional hardware may include a dedicated, small-form-factor, "Cloudlets processor" that computes location-specific situational awareness products using location-specific data streams for a defined subset of sensors; this processor also serves as an on-site Network Time Protocol host.

The stand-alone processor frequently polls a Network Time Protocol (NTP) server and uses the replies from the server to maintain time synchronization; the system time offsets obtained from the NTP server are numerically fitted to a windowed series in order to provide an accurate estimated timestamp for each sample obtained from the connected sensors. The measurement of parameters such as the maximum waveform amplitude values with their associated timestamps, and the signal-to-noise measurements are made by the sensor-processor systems. This information and updates to it are sent over the network to the Cloud facility.

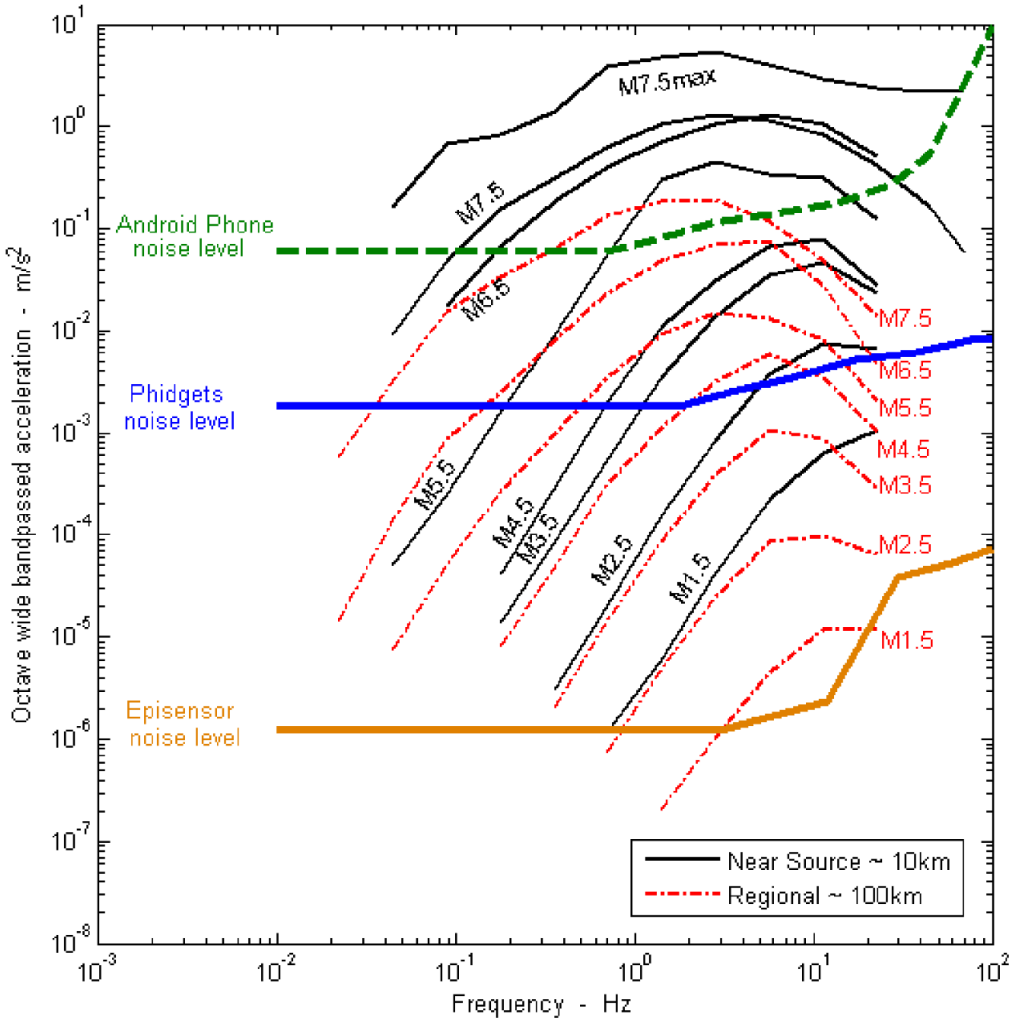


Figure 2.3. Noise level of CSN MEMS accelerometers (Phidgets) as compared to cell phone and episensor accelerometers. Fourier spectra for earthquakes of different magnitudes and for average (approximately firm soil) site conditions shown for comparison, which suggests usable bandwidths of 1-8 Hz, 0.2-15 Hz, and 0.1-20 Hz for earthquakes at regional distances for **M** 4.5, 5.5, and 6.5 earthquakes, respectively. Source: Clayton et al. (2011).

2.1.3 Data Transmittal to Central Server

The stations are connected to the internet, typically via a local, on-site VPN network. They continuously stream data that is captured by a cloud-based facility (currently AWS S3). Clayton et al. (2020) describe two modes of data communication: (1) Continuous data are sent in 10 min intervals that include starting and ending time stamps, which in the cloud are resampled to 50 sample/sec.; (2) When an event is detected, the stations monitor the data stream for accelerations that exceed 0.005g, which triggers the system, causing it to send the maximum acceleration over

the next 1 sec to the cloud. This continues until the signal drops below the threshold. The data from the second communication are used to create an evolving map of shaking across the network.

2.2 DATA DISSEMINATION AND FORMAT

The data sent to the AWS cloud service are processed in a similar manner to SCSN data. The raw waveform segments are converted to Seismic Analysis Code (SAC) binary format. The archive grows rapidly with time and is backed up daily on a portable disk. Data can be retrieved from the archive by end-users using a variation of the Seismogram Transfer Program (STP) software (https://scedc.caltech.edu/data/stp). When an earthquake is detected, a window of the data is retrieved from the continuous waveforms with STP and placed on the CSN data webpage where it is publicly accessible: http://csn.caltech.edu/data/. The continuous ambient vibration data are not publicly disseminated due to privacy concerns.

As reported by Clayton et al. (2020) for a school site, maximum accelerations over a year-long period are generally in the range of 0.001 to 0.003g, but spikes of 0.01g or larger also occur when school is in session (students are present).

The data for this study was retrieved from the CSN data webpage: http://csn.caltech.edu/data/. The data were downloaded to servers as SAC files for subsequent processing, as described in Section 3.1. The data in the SAC files is in binary format and units of g.

2.3 EVENTS CONSIDERED

Figure 2.4 shows the locations of 29 events considered in this study. We include all events recorded by the network with **M** > 4. Table 2.1 lists the events and their key attributes for engineering studies. Per NGA protocols (e.g., Contreras et al. 2022), seismic moment is taken from the global centroid moment tensor catalog (Ekström et al. 2012; https://www.globalcmt.org/) as are other moment tensor attributes with the exception of hypocenter location, which is taken from USGS (https://www.usgs.gov/programs/earthquake-hazards/earthquakes). The number of CSN records is the total number of records considered in this study, even if the data was ultimately deemed unusable. The number of non-CSN records is the number of processed records in the ground motion database (Buckreis et al. 2023a).

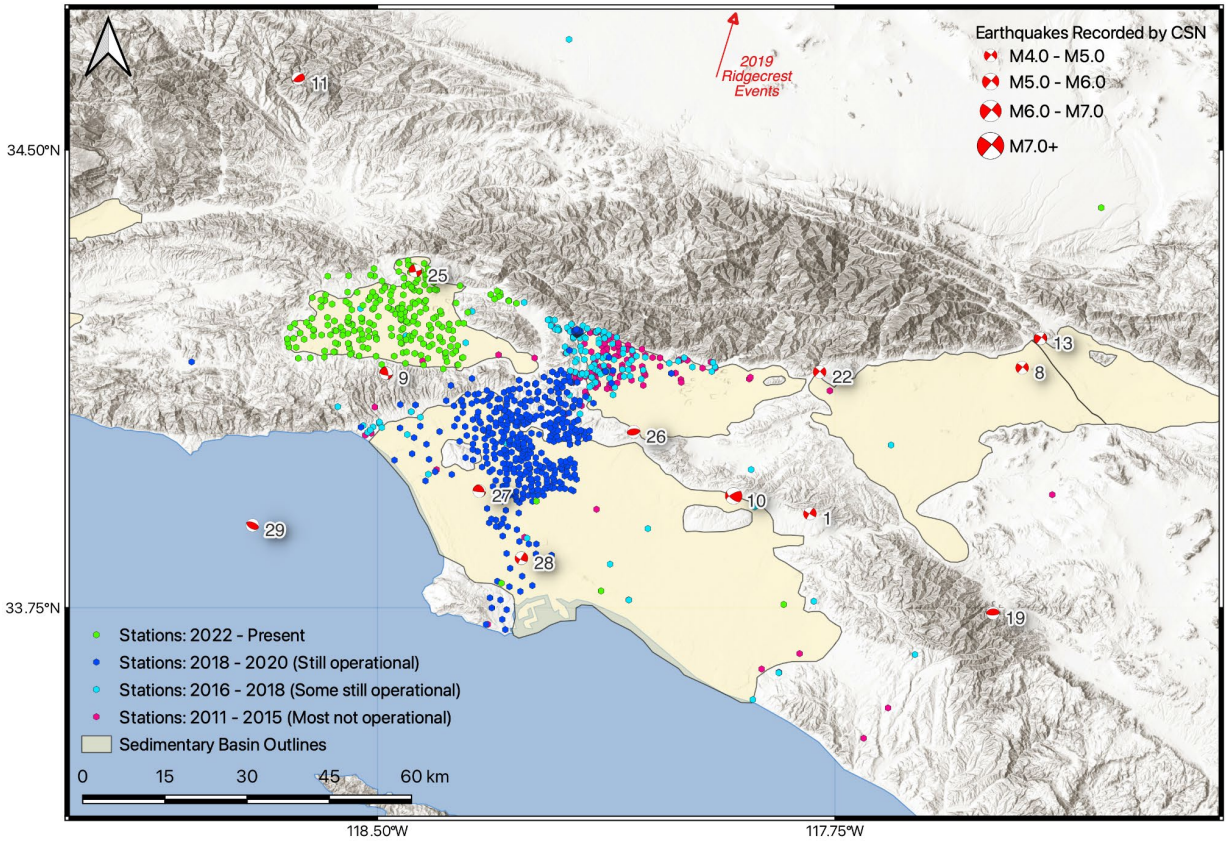


Figure 2.4. Map of CSN stations and the events they recorded.

Table 2.1. Earthquakes considered in present study

	Date	Name	M_0 (dyne-cm)	M	Hypo Lat	Hypo Long	Hypo depth (km)	#CSN recs	# Non-CSN recs
1	2012-08-08	Yorba Linda 2	1.90E+22	4.12	33.904	-117.791	10.2	561	831
2	2012-08-26	Brawley	1.64E+24	5.41	33.019	-115.54	8.3	609	198
3	2013-03-11	Anza	1.74E+23	4.79	33.501	-116.458	10.9	576	216
4	2013-05-29	Santa Barbara Channel	3.26E+23	4.97	34.406	-119.92	7.1	648	522
5	2013-07-24	Weldon		4.29[L]	35.486	-118.288	6.6	675	141
6	2013-08-25	Weldon 2	2.39E+22	4.19	35.48	-118.285	1.2	672	114
7	2013-10-06	Joshua Tree	1.23E+22	4.03	34.709	-116.294	0.8	597	420
8	2014-01-15	Fontana 3	5.59E+22	4.46	34.143	-117.443	2.9	423	747
9	2014-03-17	Westwood	5.03E+22	4.43	34.134	-118.486	9.2	585	957
10	2014-03-29	La Habra	6.29E+23	5.17	33.933	-117.916	5.1	525	1455
11	2015-01-04	Lake Castaic	2.95E+22	4.28	34.617	-118.63	7.8	702	783
12	2015-12-27	Johannesburg	2.96E+22	4.25	35.214	-117.282	3.2	1056	333
13	2015-12-30	Devore	4.94E+22	4.43	34.191	-117.413	7	1059	930
14	2016-01-06	Banning	4.80E+22	4.42	33.959	-116.888	16.7	1044	987
15	2016-02-20	Lucerne Valley	3.68E+22	4.34	34.61	-116.629	6.7	1104	816
16	2016-02-24	Wasco	3.32E+23	4.98	35.542	-119.373	22.1	1113	144
17	2016-06-10	Borrego Springs	1.03E+24	5.31	33.432	-116.443	12.3	996	408
18	2017-12-07	Julian	1.09E+22	3.96	34.148	-116.479	11.1	1488	321
19	2018-01-25	Trabuco Canyon	1.13E+22	3.97	33.741	-117.491	11.2	1479	471
20	2018-04-05	Santa Cruz Island	1.37E+24	5.39	33.82	-119.734	9.8	1197	471
21	2018-05-08	Cabazon	6.80E+22	4.49	34.016	-116.78	12.9	1155	363
22	2018-08-29	La Verne 3	4.66E+22	4.38	34.136	-117.775	5.5	975	456
23	2019-07-04	Searles Valley	5.95E+25	6.48	35.705	-117.504	9.78	1608	1488
24	2019-07-06	Ridgecrest	4.39E+26	7.06	35.77	-117.599	8	1602	2217
25	2020-07-30	Pacoima	2.25E+22	4.17	34.302	-118.438	8.9	2112	453
26	2020-09-19	El Monte	8.11E+22	4.54	34.038	-118.08	16.9	2115	183
27	2021-04-05	Lennox	1.25E+22	4.03	33.941	-118.333	19.3	2118	111
28	2021-09-18	Carson	3.29E+22	4.28	33.831	-118.264	11.9	2124	111
29	2023-01-25	Malibu	2.44E+22	4.19	33.885	-118.705	14.7	2058	387

[L]Local magnitude (M_L)

3 Data Processing and Classification

3.1 CSN DATA PROCESSING

3.1.1 Processing Steps

The Next Generation Attenuation (NGA) program has developed standard steps that are used to process earthquake ground motions. The aim of the steps is to minimize the effects of noise on recorded ground motions, while optimizing the dynamic range for which a given recording can be considered to accurately represent the ground shaking at the site. The most recent procedures are described by Goulet et al. (2021) and Kishida et al. (2020), although the main elements of the procedure were presented earlier by Boore (2005), Boore and Bommer (2005), and Douglas and Boore (2011). The steps are listed below with a brief description.

Screening of time series: Records are visually inspected to identify if a seismic signal is present. When records have indiscernible p- or s-waves, it indicates that noise likely prevails over the ground motion signal, and the record is rejected. Furthermore, any odd shapes like spurious spikes, multiple arrivals, and gradual intensity increase outside the p- or s-wave windows are rejected. No minimum number of passing components is applied during screening (e.g., when a station has only one or two passing components, that data is passed forward to the database). Figures 3.1 and 3.2 show examples of rejected records because of indiscernible seismic wave arrivals and spurious features, respectively.

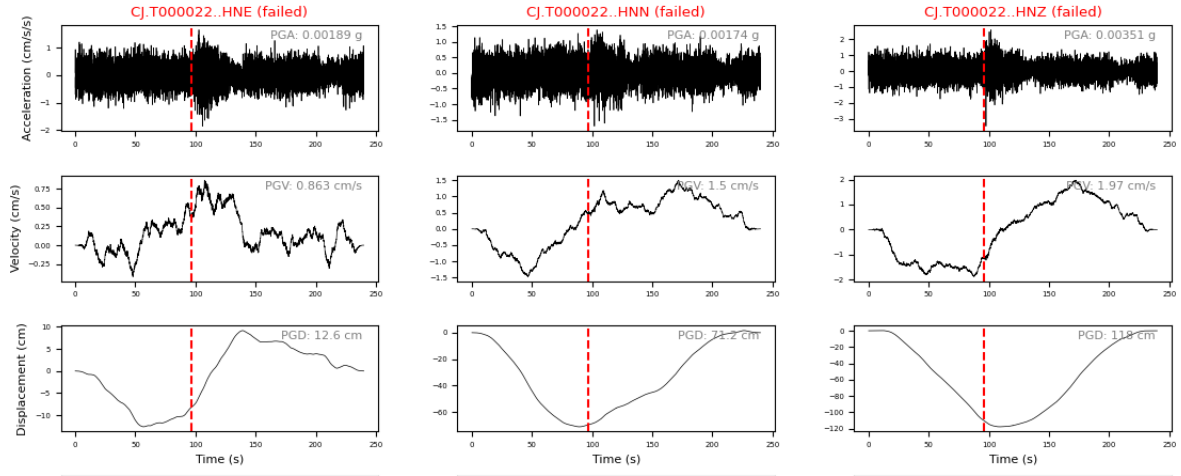


Figure 3.1. Time series of acceleration (top), velocity (middle), and displacement (bottom) that are rejected due to lack of clear seismic wave signal (the records appear to be noise-dominated). The red dashed line represents the p-wave arrival time (t_p). These are three components from station CJ.T000022 from the Pacoima M4.2 earthquake.

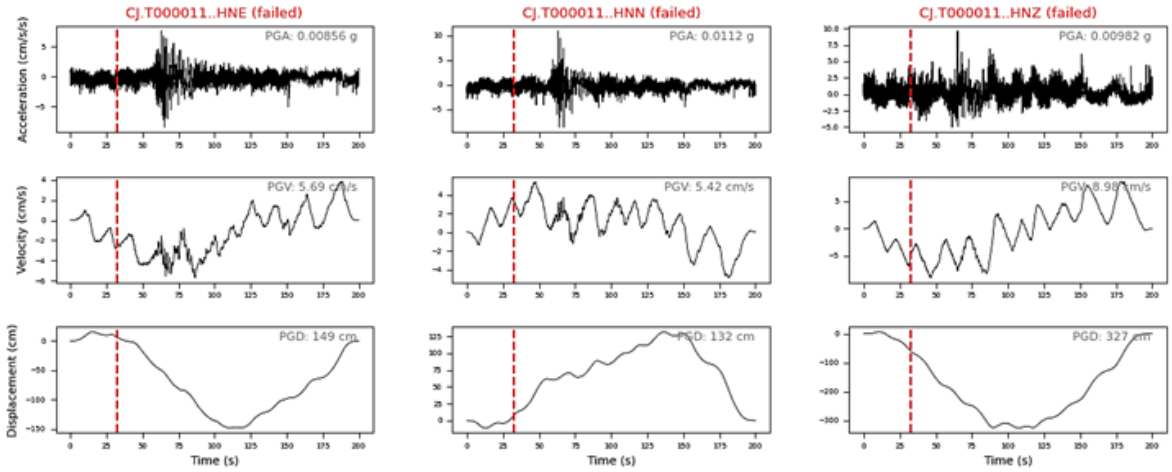


Figure 3.2. Time series of acceleration (top), velocity (middle), and displacement (bottom) that are rejected due to spurious features (long period oscillations, about 5 sec in period). The red dashed line represents the p-wave arrival time (t_p). These are three components from station T000011 from the 2019 Searles Valley earthquake.

Identify noise and signal windows: The p-wave arrival time (t_p) is identified. As shown in Figure 3.3, the time interval before the p-wave arrival is taken as the noise window, which is 60 sec in duration with CSN data (only a portion of this time window is shown in the figure). The time window containing the signal begins at t_p and extends to t_p+D_S, where D_S is the 5-95% significant duration, which is taken from available models (gmprocess: 95th percentile value from the Afshari and Stewart (2016) model, i.e. mean + 2 standard deviations; NGA: durations from older models as described in Section 3.2.2 of Kishida et al. 2020). A window function is applied to the noise and signal for subsequent steps of data processing.

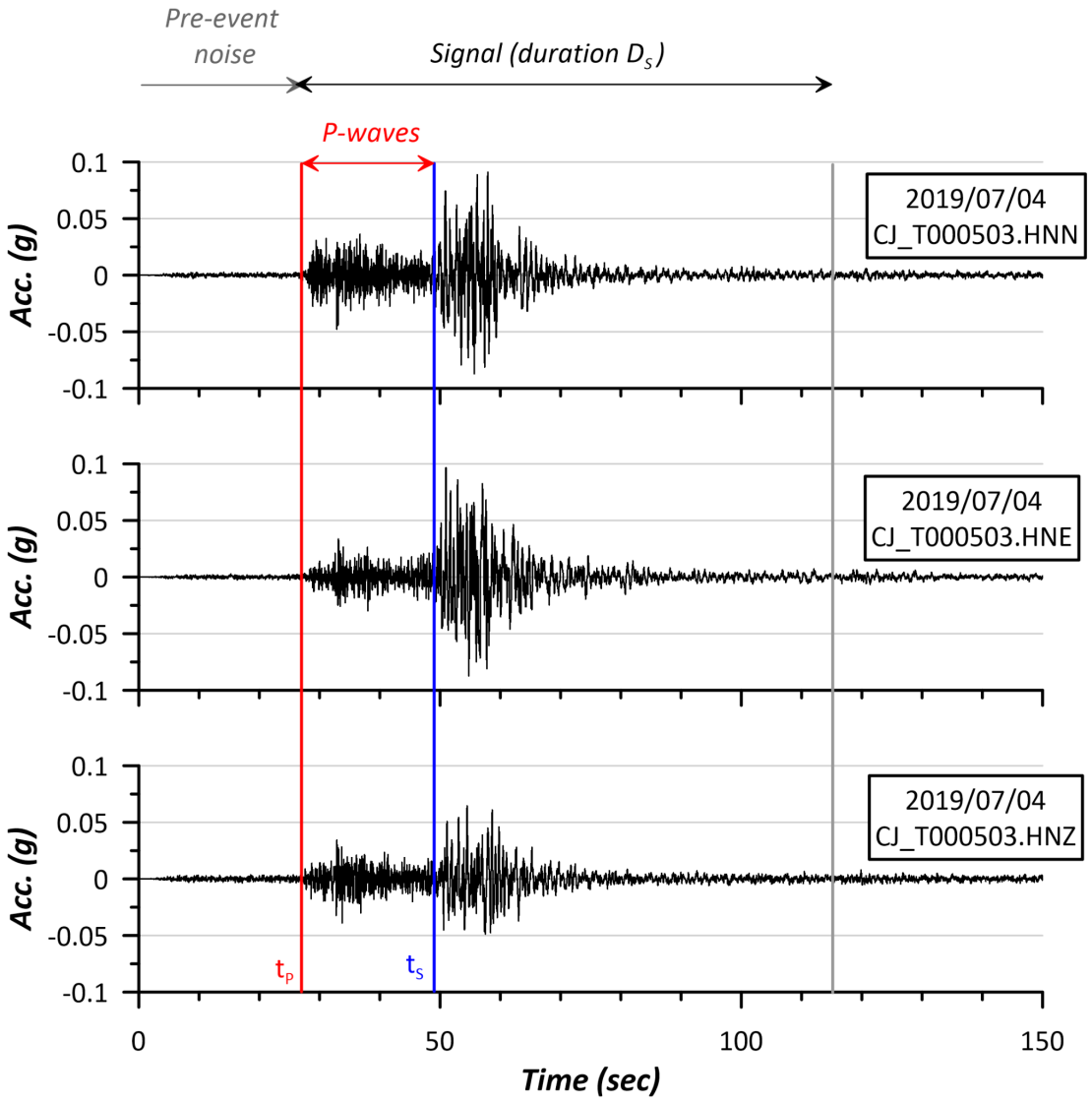


Figure 3.3. Three-component acceleration time series for station CJ_T000503 from the 2019 Searles Valley earthquake (**M**6.48) showing p-wave arrival time (t_p) as estimated from gmprocess, the preceding noise window, p-wave window, approximate s-wave arrival time (t_S), and signal duration (D_S), which is taken as the 95%tile value from the Afshari and Stewart 5-95% significant duration model. The start time for the s-window is indicated for illustrative purposes only and is not directly used in the signal processing.

Compute Fourier amplitude spectra (FAS): FAS are computed for each individual component using a standard set of 390 frequencies as used in the ground motion database (Buckreis et al. 2023a). FAS are computed for both the signal and noise windows, and prior to filtering, produce spectra such as those shown in Figure 3.4. Because standard Fourier spectra computation routines (e.g. Kramer 1996) produce amplitudes that increase with increasing signal duration, there is potential for bias when comparing noise and signal FAS if they have different durations. To correct for this, each FAS is normalized by the square-root of the signal duration, which produces unusual units like g×sqrt(sec) or $m/s^{1.5}$.

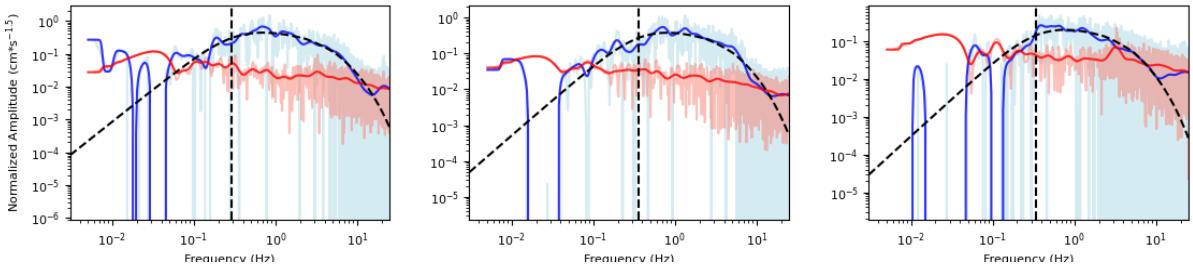


Figure 3.4. FAS for three-component acceleration time series' noise (red) and signal (blue) windows. Example is for CSN station CJ.T000107 in the 2019 Searles Valley earthquake. Both smoothed and unsmoothed FAS are shown.

High-pass filter: The selection of high-pass corner frequency (f_{cHP}) is informed by multiple criteria. The initial guess of the f_{cHP} is based on the theoretical acceleration decay at low frequency, f^2 model (Brune, 1970; Boore and Bommer, 2005); the initial selection is made where the shape of the signal FAS deviates from the theoretical decay (approximately 0.06 Hz for the example in Figure 3.4). There is some debate about whether this is an appropriate selection, as not all ground motions will necessarily follow this shape at low frequencies, and such selections may be unnecessarily conservative (i.e., filtering out too much low frequency signal).

Also considered is the ratio of FAS for the signal and noise windows (referred to as signal-to-noise ratio, or SNR). A typical threshold value of SNR used in many applications is 3 (e.g., Goulet et al. 2021, Kishida et al. 2020). It is important to recognize that the window referred to as "signal" is actually the sum of the true signal and noise, hence the ratio being computed is more properly expressed as (S+N)/N = S/N +1. Accordingly, the common threshold of SNR = 3 actually corresponds to a S/N ratio of 2. The example in Figure 3.4 reaches the SNR=3 threshold at approximately 0.1 Hz for the two horizontal components (left and center plots).

A final check is made to ensure that when the waveform is integrated to displacement, that no unusual long-period fluctuations (referred to as "wobble") are present (Boore and Bommer, 2005). If wobble is observed using an f_{cHP} initially estimated from shape or SNR considerations, the user goes back, chooses a higher f_{cHP} and evaluates the displacement again. The user continues this iterative process until the displacement waveform lacks wobble based on visual inspection or various numerical criteria. Figure 3.5 shows example FAS for the north-south-component of the record in Figure 3.4 before and after high-pass filtering. The effect of the filter is apparent from the decay in FAS at low frequencies.

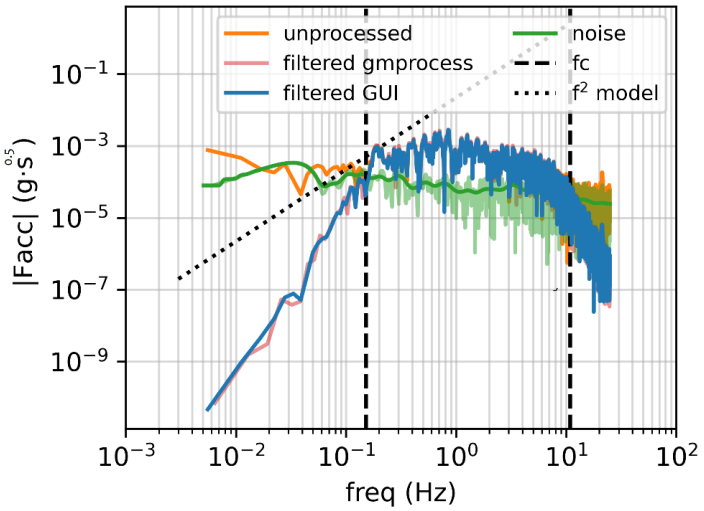


Figure 3.5. FAS for east-west component acceleration time series, showing unprocessed, auto-processed with two tools (labeled gmprocess and GUI, as discussed further below), along with the noise spectra (green) and signal prior to filtering (orange). Record is from 2019 M6.48 Searles Valley earthquake, Station CJ.T000107.

Low-pass filter: The processing of ground motion data from higher-resolution sensors (e.g., episensor seismometers as shown in Figure 2.3) often does not require the application of a low-pass filter (used to filter out high frequency noise). Low-pass filters are not needed if the ratio between the maximum FAS and the FAS at $0.75f_{Nyq} > 10$ (Douglas and Boore, 2011) (where f_{Nyq} is the Nyquist frequency). Where this condition is met, it is common that the response spectra saturate to PGA and there are no unrealistic high-frequency spikes. Because of the relatively high noise levels for data from the Phidgets instruments and environments associated with the station locations (Figure 2.3), low-pass filters were used more frequently than with other modern sensors. Low-pass filtering is performed with a corner frequency (f_{cLP}) typically taken as the smaller of $0.75 \times f_{Nyq}$ or the frequency where SNR falls below a threshold. The data in Figure 3.5 was low-pass filtered at a corner frequency of $f_{cLP} = 11.8$ Hz.

Baseline correction: The acceleration is integrated in the time domain (Nigam and Jennings 1969) to obtain velocity and integrated a second time to obtain displacement, with initial values assumed to be zero. The baseline drift that is sometimes encountered by assuming initial values of zero is later corrected with the baseline correction procedure explained in Boore et al. (2012).

3.1.2 Processing Tool

We used the U.S. Geological Survey (USGS) open-source ground motion processing tool, gmprocess (Hearne et al., 2019), to perform the processing steps described in Section 3.1.1. This tool was used in a semi-automated mode that was enabled by several revisions that were

undertaken in nearly coincident work at UCLA in collaboration with USGS (Ramos-Sepulveda et al. 2023). The default options when gmprocess is executed in a fully automated manner are as follows:

> Screening: Time series are first screened to ensure they meet the following configurable requirements: (1) free-field stations, (2) both horizontal components are available, (3) minimum sampling rate of 20 Hz, (4) minimum of 0.1 zero crossings per second after demeaning the raw time series, and (5) SNR must be greater than 3 from the Brune (1970) corner frequency to 5 Hz.
>
> Noise and signal windows: The p-wave arrival is estimated from a travel-time calculation using the velocity model of Kennet et al. (1991). The signal window duration is estimated as the mean plus two standard deviation 5-95% significant duration from Afshari and Stewart (2016), and 60 seconds of pre-event noise is included for the purpose of computing SNR.
>
> High- and low-pass filters: High-pass corner frequency f_{cHP} is set as the lowest frequency where SNR = 3. For low-pass filtering, the default is to not apply a filter.

We use an SNR=1 criterion for high-pass filtering as an initial guess to estimate f_{cHP}. This guess is then refined using a routine described in Ramos-Sepulveda et al. (2023) and implemented in gmprocess that removes displacement wobble in an iterative manner by fitting a polynomial to the displacement waveform that is subtracted from the data in the time-domain. Moreover, we do not apply the low-frequency spectral shape criterion discussed in Section 3.1.1. These steps are applied in an automated manner in gmprocess, which produces an h5 file that contains the original and processed records.

The data are then visualized using a GUI tool (Ramos-Sepulveda et al. 2023) to ensure that filtered records are reasonable with respect to FAS shape, Sa shape, and displacement wobble. These visual inspections check for the same features considered in standard procedures (Section 3.1.1). The GUI allows the salient features to be observed and for users to adjust f_{cHP} and f_{cLP} as needed. Once a record is approved, the GUI allows records to be saved in an .h5 file format. The final step that precedes uploading the data to the database is calculation of median and maximum component peak ground motions and response spectra (RotD50 and RotD100, respectively; Boore 2010) using the RCTC code by Wang et al. (2017), as well as calculation of effective amplitude spectra, which are an orientation-independent measure of FAS conceptually similar to RotD50 (Goulet et al. 2021).

The data processing work conducted as part of this project occurred coincident with several changes being implemented in gmprocess, including the revisions described by Ramos-Sepulveda et al. (2023) and a major update that corrected a prior error in Fourier amplitude calculations (in the earlier version, the data were not normalized by the square-root of duration). The 2nd and 1st

authors, along with Scott Brandenberg, were in contact with the gmprocess team regarding several issues that came about when applying the tool to CSN data. These issues include the following:

- Gmprocess was originally developed for compatibility with Mac and Linux systems. To enable a windows installation, prerequisite software that needs to be installed in the Windows system, including a C compiler (e.g. Windows SDK C/C++ compiler) and Git (or Anaconda prompt). Revisions were made to streamline windows installation of the tool.
- While the units in the sac file output from gmprocess were known to be in g, the results plotted in gmprocess were non-physical (motions too large). We developed fixes that apply to all sac file inputs in the configuration file; these changes are now integrated into the main gmprocess tool.

We used non-default (lower) values for SNR thresholds in the config.sys file when running gmprocess in automated mode (prior to GUI data inspections). This was done to pass records through to the GUI for visual inspection, corner frequency selections, and classification (Section 3.3). Otherwise, large numbers of CSN records were screened out and hence could not efficiently be inspected and evaluated. For similar reasons, we disabled the check noted above that SNR must be greater than 3 from the Brune (1970) corner frequency to 5 Hz.

3.2 NON-CSN DATA PROCESSING

All of the CSN data considered in the present work was processed using the procedures described in Section 3.1. Because a portion of this project includes comparison of CSN and non-CSN data for common events, we performed processing of non-CSN data for events that did not already have ground motions in the ground motion database. These events were 2013 **M**4.79 Anza, 2012 **M**5.41 Brawley, 2013 ML4.29 Weldon, 2013 **M**4.19 Weldon2, 2015 **M**4.25 Johannesburg, 2016 **M**4.98 Waso, 2016 **M**5.31 Borrego Springs, 2017 **M**3.96 Julian, 2018 **M**3.97 Trabuco Canyon, 2018 **M**5.39 Santa Cruz Island, 2018 **M**4.49 Cabazon, 2018 **M**4.38 La Verne 3, 2020 **M**4.17 Pacoima, 2020 **M**4.54 El Monte, 2021 **M**4.03 Lennox, 2021 **M**4.28 Carson, and 2023 **M**4.19 Malibu.

The non-CSN data were processed using an R code that has been used for record processing on multiple NGA projects including NGA-West2, NGA-East, and NGA-Subduction (Bozorgnia et al. 2014, Goulet et al. 2021, Bozorgnia et al. 2022, respectively). The code applies the procedures described in Section 3.1.1, but with the following features (details in Kishida et al. 2020):

Screening: Time series are screen visually to check for noise-dominance or spurious features.

Noise and signal windows: The p-wave arrival is identified manually, generally from the vertical-component ground motion. The p-to-s-wave arrival time difference is estimated based on hypocentral distance and adjusted based on visual inspection, as needed. The s-wave duration is evaluated from models for source and path duration.

High- and low-pass filters: High-pass corner frequency f_{cHP} is selected in consideration of SNR, removing displacement wobble (through iterative analysis), and often based on shape to avoid flat portions of the FAS at low frequencies. We did not apply the FAS shape criterion when selecting f_{cHP}. The use of a low-pass filter was generally not required.

As with the data saved from the GUI, the R output is used to compute intensity measures, including EAS, which were then uploaded to the database. Whereas the 2nd author performed most of the processing of CSN data, the 4th author performed most of the processing of non-CSN data. As a quality-control check, some data was jointly processed by both students, which served as a training exercise that improved the understanding of the relevant methods for both. After this exercise, the results from the two students for a common data set were similar.

3.3 CSN DATA CLASSIFICATION

In our evaluations of the CSN data, we observed three general categories of records. The "best" records (BBR) clearly reflect earthquake shaking, having waveforms where the different wave arrivals are evident and modest effects of noise. Records deemed unusable (REJ) appear to be noise dominated, generally based on visual inspection of time series, but sometimes also from similar levels of signal and noise FAS. The intermediate case (NBR) consists of records that have the visual appearance of earthquakes, but the signal is of modest strength in comparison to noise and the record bandwidths are relatively limited.

In order to achieve project objectives as outlined in Chapter 1, it was necessary to identify criteria distinguishing BBR and NBR. Specifically, assignment of BBR, NBR, or REJ to each record was required for threshold analyses (Section 3.4) and data comparisons (Chapter 4). After some trial and error, the category definitions we adopted were as follows:

1. BBR: relatively broad usable frequency range from $f_{cHP} < 0.5$ Hz to $f_{cLP} > 10$ Hz
2. NBR: limited usable frequency range because corner frequencies do not meet the criteria for BBR (i.e., $f_{cHP} > 0.5$ Hz or $f_{cLP} < 10$ Hz)
3. REJ: visual evidence suggests seismic waves cannot be distinguished from noise

Figures 3.6-3.8 show examples of records assigned to BBR, NBR, and REJ, respectively. In each case the figures are plots generated by gmprocess. At the top of each column an indication is given of whether the record "passed" or "failed" depending on criteria described in Section 3.1.2. Corner frequencies applied in the filtering are those from automated algorithms, and as a result, there are cases where displacement wobble occurs that would be removed in subsequent processing via the GUI. In Figure 3.6 the signal FAS exceeds that from noise over a wide frequency range. In Figure 3.7, the signal FAS exceeds that from noise over a narrower frequency range – in particular the values of f_{cHP} are > 0.5 Hz, which causes the NBR assignment. In Figure 3.8, the signal FAS is generally similar to, or in some cases below, the noise FAS. Those relative amplitudes, along with the obvious effects of noise in the time series, are the reason for the REJ assignment.

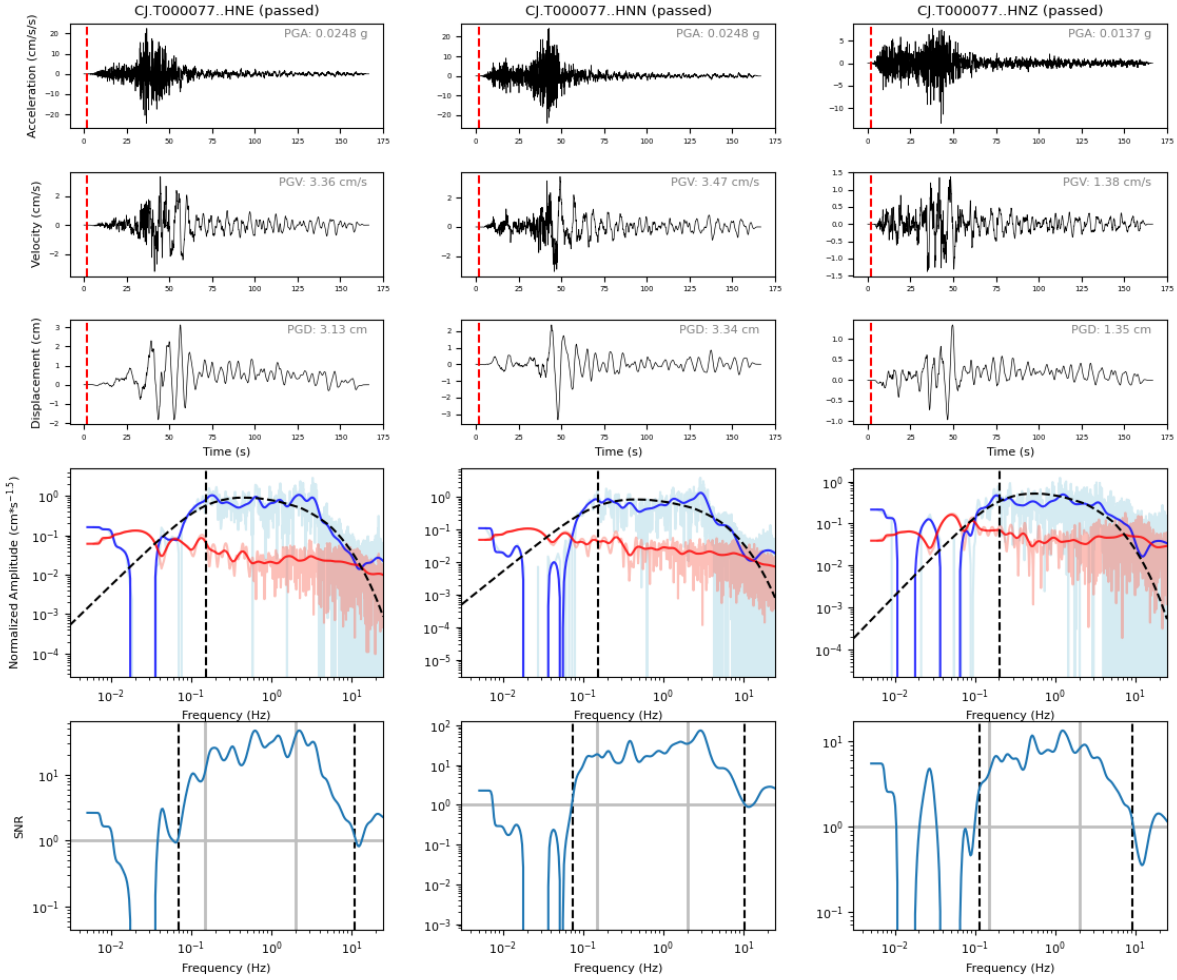


Figure 3.6. Plots from gmprocess showing three-component records from the 2019 Ridgecrest earthquake in which each component was assigned as BBR and the filter corners were automatically selected within gmprocess. The top three rows are time series, the fourth row is FAS, and the bottom row is SNR. In the time series plots, the vertical red line indicates p-wave arrival. In the FAS plots, raw and smoothed FAS are shown for the signal (blue) and noise (red). The dashed black curve indicates a Brune spectrum fit to the signal FAS with the corner frequency indicated by the vertical dashed line.

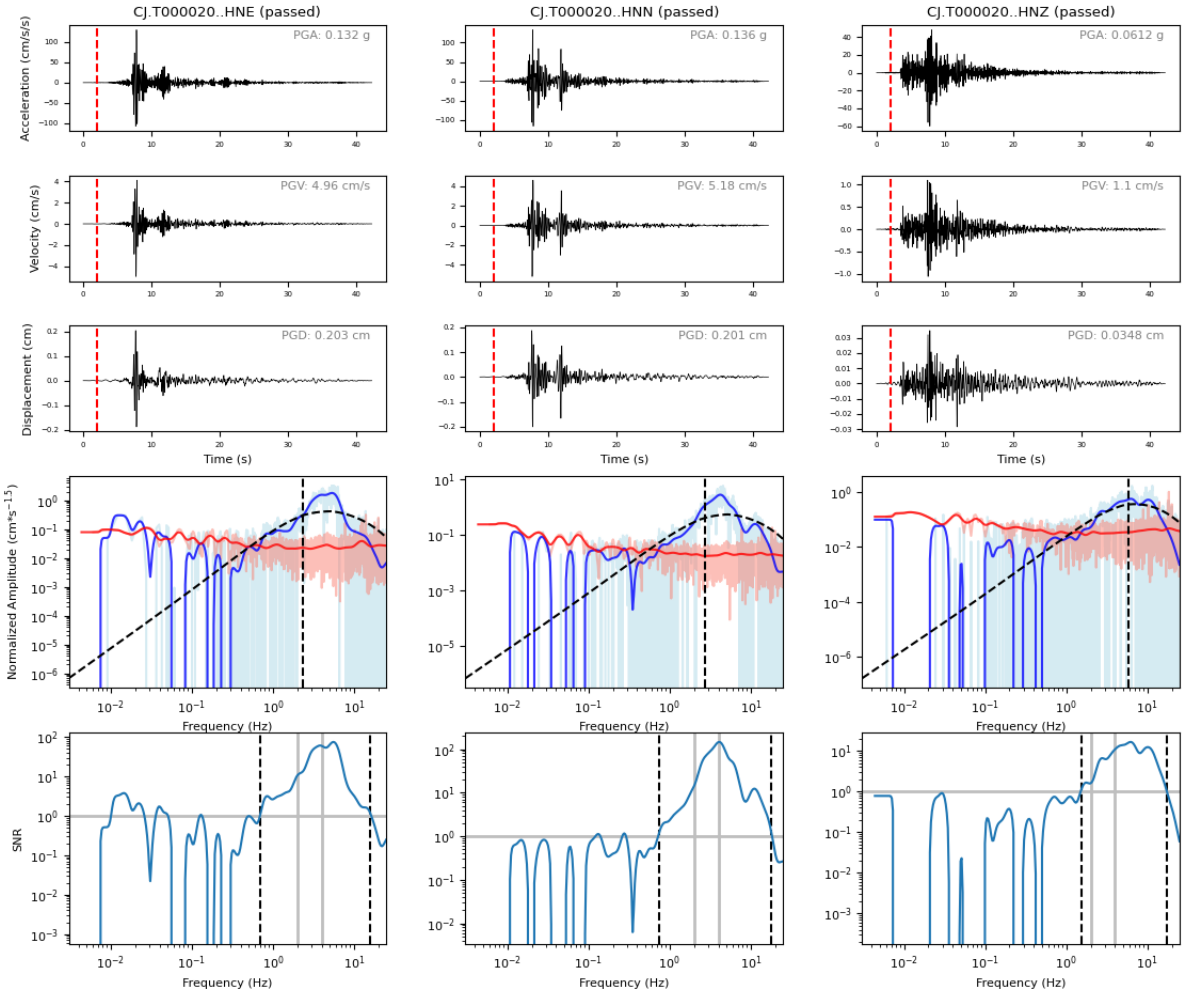


Figure 3.7. Plots from gmprocess showing three-component records from the 2020 El Monte earthquake in which each component was assigned as NBR and the filter corners were automatically selected within gmprocess. The top three rows are time series, the fourth row is FAS, and the bottom row is SNR. In the time series plots, the vertical red line indicates p-wave arrival. In the FAS plots, raw and smoothed FAS are shown for the signal (blue) and noise (red). The dashed black curve indicates a Brune spectrum fit to the signal FAS with the corner frequency indicated by the vertical dashed line.

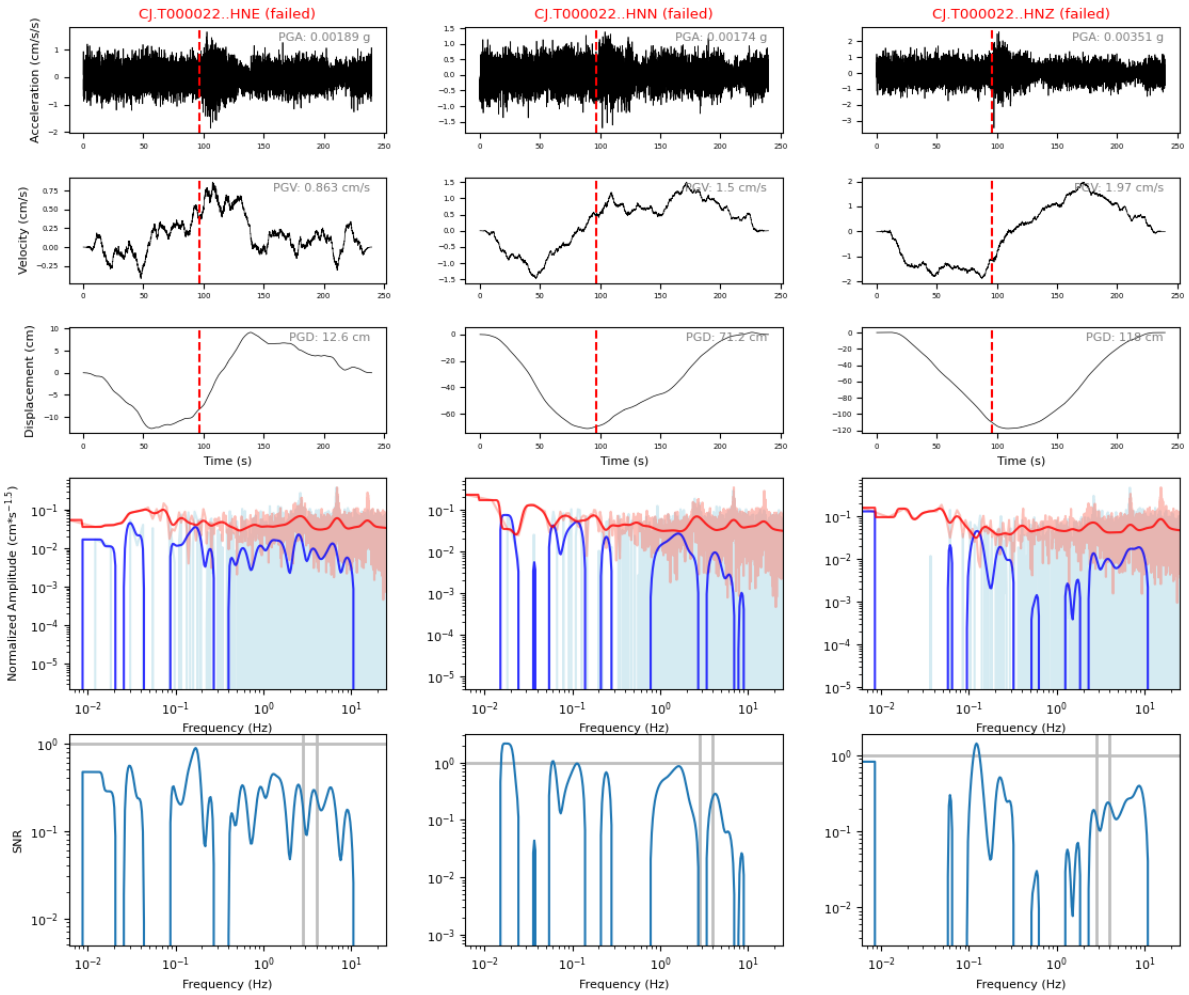


Figure 3.8. Plots from gmprocess showing three-component records from the 2020 Pacoima earthquake in which each component was assigned as REJ. No filtering was applied because the record failed screening criteria. The top three rows are time series, the fourth row is FAS, and the bottom row is SNR. In the time series plots, the vertical red line indicates p-wave arrival. In the FAS plots, raw and smoothed FAS are shown for the for the signal (blue) and noise (red). The dashed black curve indicates a Brune spectrum fit to the signal FAS with the corner frequency indicated by the vertical dashed line.

Table 3.1 indicates the number of CSN individual-component recordings in each category for each of the 29 considered events. The number of usable ground motions for modeling purposes, which combine horizontal components typically as RotD50, are approximately ½ to 1/3 of the numbers shown in Table 3.1.

For some events (e.g., 2019 Searles Valley, 2019 Ridgecrest, and 2020 El Monte), large pluralities of ground motions have usable bandwidth (BBR and NBR), whereas for others (2013 Weldon and Joshua Tree) all records are rejected based on the criteria presented in Section 3.1. Figure 3.9 illustrates the reason for these discrepancies by showing BBR (green), NBR (yellow), and REJ

(red) records in distance-magnitude space. In the upper-left portion of the plot (large magnitude or close distances for **M** < 5 events), most records are BBR, whereas the lower-right portions (**M** < 5 event and distances > 50-100 km) are REJ. Clearly the level of ground shaking strongly affects the classifications. This is also reflected in summary statistics for the data set. Among events since 2018, large-magnitude events and events generally closer than 70-80 km from the network (Malibu, Carson, Lennox, El Monte, Pacoima, Searles Valley, Ridgecrest, La Verne) have the following aggregate component record classifications:

- Usable records (BBR and NBR): 5784 (60%) (1122 BBR, 4662 NBR)
- Rejected records: 3858 (40%)

The database as a whole, which includes many events with small magnitude and large distances, breaks down as

- Usable records (BBR and NBR): 9948 (45.7%) (1446 BBR, 8502 NBR)
- Rejected records: 11801 (54.3%)

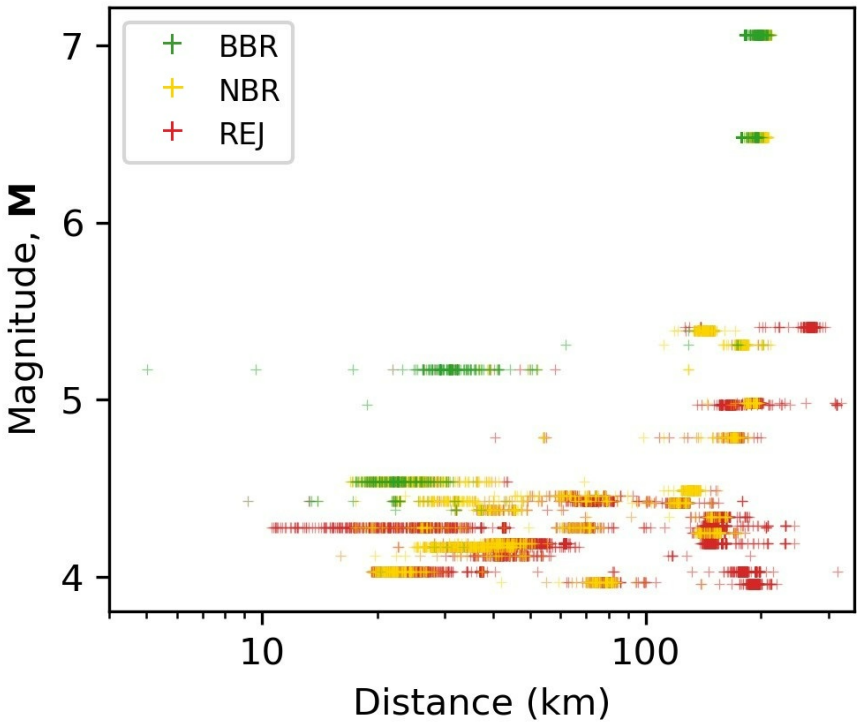


Figure 3.9. Record classification as function of distance and magnitude

Table 3.1. Numbers of BBR, NBR, and REJ ground motion components for the earthquakes considered in present study – horizontal components (vertical components)

Date	Name	**M**	Distance Range (km)	# CSN BBR comp.	# CSN NBR comp.	# CSN REJ components
2012/08/08	Yorba Linda 2	4.12	16 - 56	0 (0)	53 (6)	321 (181)
2012/08/26	Brawley	5.41	198 - 293	0 (0)	2 (0)	404 (203)
2013/03/11	Anza	4.79	98 - 199	0 (1)	145 (12)	238 (179)
2013/05/29	Santa Barbara Channel	4.97	136 - 238	2 (0)	2 (2)	428 (214)
2013/07/24	Weldon	3.84	139 - 208	0 (0)	0 (0)	450 (225)
2013/08/25	Weldon 2	4.19	138 - 208	0 (0)	0 (0)	448 (224)
2013/10/06	Joshua Tree	4.03	131 - 215	0 (0)	0 (0)	398 (199)
2014/01/15	Fontana 3	4.46	40 - 92	0 (0)	127 (5)	155(136)
2014/03/17	Westwood	4.43	9 - 79	38 (5)	295 (120)	22 (70)
2014/03/29	La Habra	5.17	10 - 129	270 (102)	56 (43)	22 (30)
2015/01/04	Lake Castaic	4.28	43 - 111	0 (0)	190 (4)	278 (230)
2015/12/27	Johannes-burg	4.25	132 - 181	0 (1)	242 (4)	462 (347)
2015/12/30	Devore	4.43	49 - 106	0 (1)	84 (2)	621 (350)
2016/01/06	Banning	4.42	92 - 155	0 (0)	175 (3)	520 (345)
2016/02/20	Lucerne Valley	4.34	131 - 187	0 (0)	94 (1)	642 (367)

2016/02/24	Wasco	4.98	171 - 202	0 (0)	201 (5)	541 (366)
2016/06/10	Borrego Springs	5.31	159 - 211	12 (0)	578 (101)	74 (231)
2017/12/07	Julian	3.96	155 - 219	0 (0)	1 (0)	991 (496)
2018/01/25	Trabuco Canyon	3.97	55 - 106	0 (0)	348 (17)	638 (476)
2018/04/05	Santa Cruz Island	5.39	119 - 171	2 (0)	582 (21)	214 (378)
2018/05/08	Cabazon	4.49	112 - 154	0 (0)	665 (0)	58 (385)
2018/08/29	La Verne 3	4.38	22 - 61	9 (1)	427 (72)	214 (252)
2019/07/04	Searles Valley	6.48	161 - 193	245 (40)	733 (316)	84 (180)
2019/07/06	Ridgecrest	7.06	167 - 205	599 (248)	446 (266)	11 (20)
2020/07/30	Pacoima	4.17	23 - 67	3 (0)	950 (43)	445 (661)
2020/09/19	El Monte	4.54	17 - 43	265 (3)	1022 (387)	20 (315)
2021/04/05	Lennox	4.03	19 - 40	1 (0)	434 (0)	964 (706)
2021/09/18	Carson	4.28	11 - 44	0 (0)	28 (0)	1388 (708)
2023/01/25	Malibu	4.19	28 - 63	0 (0)	622 (8)	750 (678)

[7]Southern California Earthquake Data Center (SCEDC)

3.4 CSN USABLE AMPLITUDE THRESHOLD

As shown in Figure 2.3 and in example ground motion plots (Figures 3.1 and 3.8), the effects of noise on CSN data are more significant than what many ground motion specialists are accustomed to from their analyses of data from relatively modern instruments. As such, it is important to establish a threshold level of ground motion that separates CSN earthquake signals (BBR or NBR) from noise-dominated data (REJ signals or pre-event noise). In this section, we address this

question on a network-wide level using the ground motion parameter of individual-component PGA. Individual components are used in lieu of combinations of components (e.g. RotD50) because for some stations individual components can have different classifications. Various intensity measures were considered for the derivation of this threshold and PGA was found to be the most effective intensity measure. Only data from ground level instruments (COSMOS codes 4 and 5) are considered.

3.4.1 Threshold Between BBR and REJ

The threshold ground motion level is evaluated to distinguish BBR from REJ components, and is denoted PGA_{th}. Figure 3.10 shows histograms of PGA for both groups (BBR and REJ) and the vertical red line shows the PGA_{th} derived from the data. This threshold was identified iteratively by computing cumulative relative likelihoods from the BBR and REJ histograms for various trial values of PGA_{th}. Denoting the cumulative relative likelihoods under the BBR curve with PGA > PGA_{th} as A and cumulative relative likelihood under the REJ curve with PGA < PGA_{th} as R, a confidence index (I_{conf}) is computed as:

$$I_{conf} = \frac{A+R}{2} \tag{3.1}$$

The trial value of PGA_{th} that maximizes I_{conf} was selected at the preferred threshold value, which is 0.005 g. The corresponding value of I_{conf} is 0.94, which indicates that only 6% of records are misclassified as either usable (BBR) or not usable (NBR) with this threshold.

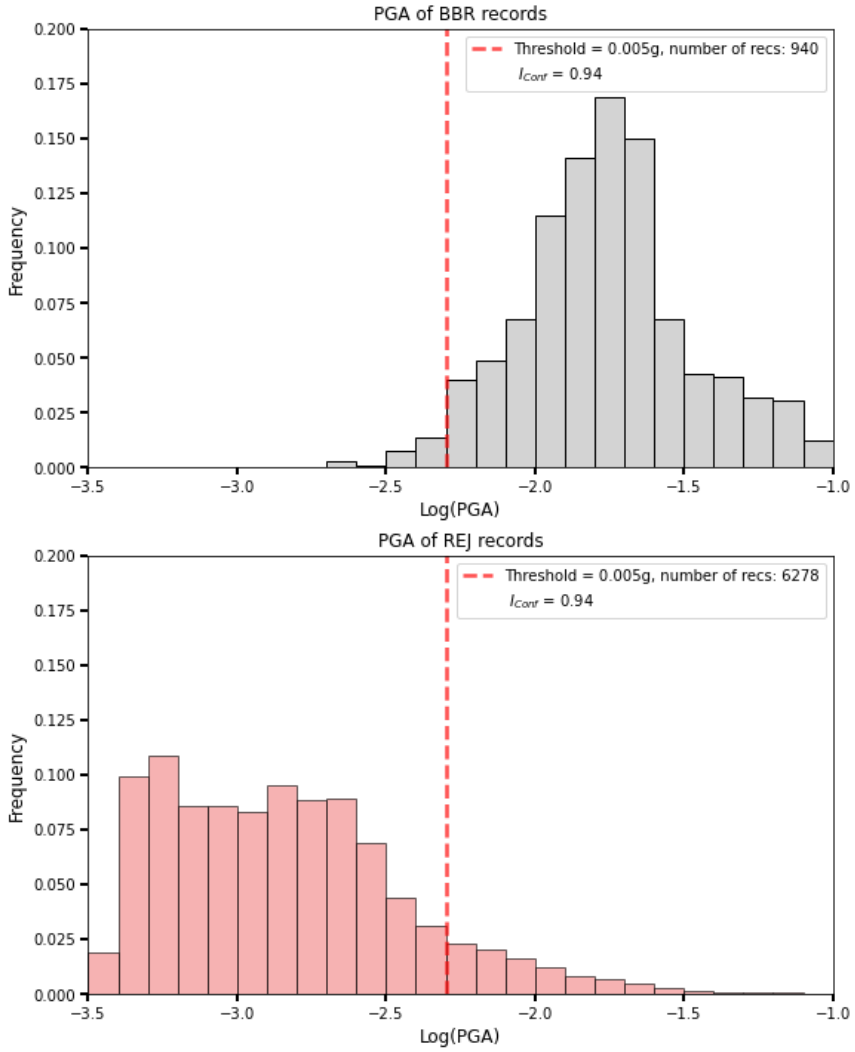


Figure 3.10. Histograms of individual-component log(PGA) for the BBR and REJ data as identified in this study. The threshold occurs at 0.005 g, for which only 6% of records are misclassified.

We separately evaluated PGA_{th} using binary logistic regression. Binary logistic regression is a method for developing a statistical model to estimate the probability of discrete binary outcomes of a process (e.g., will the outcome of a test be "pass" or "fail"?). In the present case, the data are the ground motions and the discrete binary outcomes are their assignment to BBR or REJ. The logistic function is:

$$p(x) = \frac{1}{1+exp(-(x-\mu)/s)} \tag{3.2}$$

where $p(x)$ is the probability of acceptance (BBR) for a record, x is the $\log_{10}$(PGA), and μ and s are the mean and standard deviation of a normal distribution estimated by the regression (i.e., μ is the value of x where $p(x)$=0.5 and s represents the distribution's dispersion). The calculations were performed using the logistic regression function in Python (linear_model.LogisticRegression). Figure 3.11 shows the outcome of the binary logistic regression. The x-axis is log(PGA) and the y-axis is $p(x)$. The symbols at $p(x)$ = 1.0 represent the BBR records (plotted at their respective PGAs) and the symbols at $p(x)$ = 0 represent the REJ records. The green line is a plot of Eq. (3.2) and the threshold identified in the figure is 10^{μ} = 0.011g (s = 0.29). If PGA_{th} is taken as 10^{μ}, this estimate of the threshold PGA is twice the previous value (Figure 3.10).

We anticipated that the higher threshold identified from binary logistic regression could have been influenced by the aforementioned significant imbalance between the numbers of BBR and REJ motions (many more in REJ). To test the importance of this data balance, we created an artificial data set that matches the frequencies in Figure 3.11 but equally sampled from each distribution (100 data points each). For example, a bin with a frequency of 0.04 would have 0.04×100 = 4 accelerations in the population at the bin-center PGA. Repeating the logistic regression for this data set produces the result in Figure 3.12, which provides a value of 10^{μ} = 0.0053g, matching the result by maximizing I_{conf}. As a result, we consider the appropriate threshold to be PGA_{th} = 0.005g.

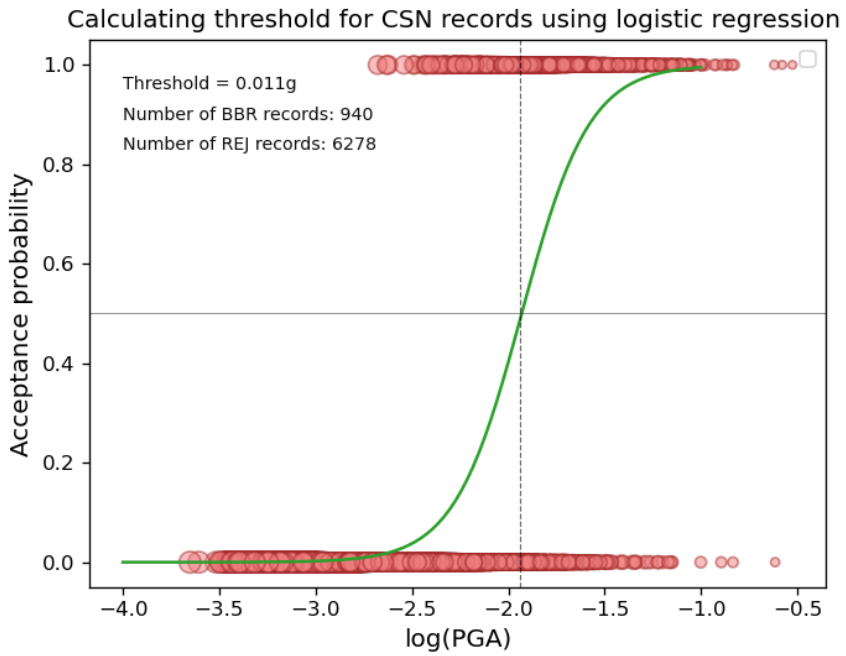


Figure 3.11. Data and binary logistic regression fit for acceptance (BBR) and rejection (REJ) of CSN data as function of log(PGA). The logistic fit produces parameters μ = -1.96 and s = 0.29.

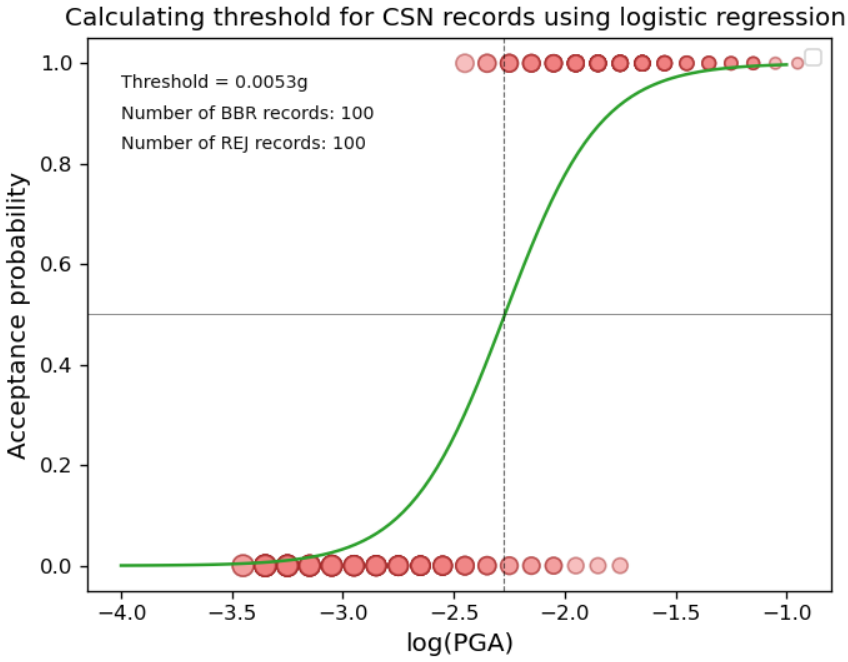


Figure 3.12. Artificial data set created to match BBR and REJ histograms (Figure 3.9) but with even sampling, and binary logistic regression fit to that data. The logistic fit produces parameters μ = -2.28 and s = 0.36.

It is noteworthy that the preferred value of PGA_{th} is nearly 20 times the noise level identified for the Phidgets sensors in Figure 2.3 of 0.00028 g (Clayton et al. 2020). There are several possible reasons for the differences. The effective noise levels of the sensors installed in the field are expected to be higher than what is reflected in Figure 2.3, which is a relatively quiet site with limited ambient vibrations. The vast majority of CSN sites are in the culturally noisy San Fernando, Los Angeles and San Gabriel basins; thus, noise levels at those sites will almost always be higher due to environmental conditions associated with the station location. PGA_{th} is an average for the entire network and therefore incorporates the effects of noise in these environments.

3.4.2 Threshold Between NBR and REJ

The BBR to REJ threshold arguably has the greatest practical importance because it establishes a CSN-wide threshold to distinguish clearly non-usable records from the records of good quality. The NBR / REJ threshold can also be identified by applying the same procedures described in Section 3.4.1, and is of potential interest to distinguish records that have visible earthquake signals from those that are noise-dominated. Here we present the results of the I_{conf} threshold analysis and interpret the results.

Figure 3.13 shows histograms of PGA for all three groups (BBR, NBR, and REJ). The NBR histogram spans a wide range of PGA from 0.0006 to 0.05g (approximate 5-95% range), with significant overlap with BBR at the high PGA end and with REJ at the low PGA end. As in Section 3.4.1, I_{conf} was computed using Eq. 3.1 for different values of PGA_{th} and the threshold was identified as 0.0025g by minimizing I_{conf}. This threshold is shown in Figure 3.13 as a vertical yellow line and the minimized I_{conf} value is 0.73. This relatively low value of I_{conf} indicates a high rate (27%) of misclassifications (REJ records above the threshold and NBR records below it).

The lower threshold for NBR / REJ (0.0025g) than for BBR / REJ (0.005g) is expected because a higher level of record quality is required for BBR than for NBR. Moreover, the relatively high rate of misclassifications with the NBR / REJ threshold is expected given the significant overlap of the two distributions. Interestingly, even this lower threshold still exceeds (by nearly a factor of 10) the Phidgets sensor noise level of 0.00028 g, which indicates that the Phidget noise level appears to have limited applicability to typical field conditions across the network.

For forward applications, the BBR / REJ threshold is useful for preliminary identification of usable records, but it is not sufficient because, as shown in Figure 3.13, a significant number of records with PGA > 0.006 g could be NBR. Hence, only record processing (Section 3.1) can definitively confirm record reliability for ground motion prediction applications. The NBR / REJ threshold has limited practical application, but could be used to identify records that have visual characteristics of earthquake shaking, albeit with limited confidence.

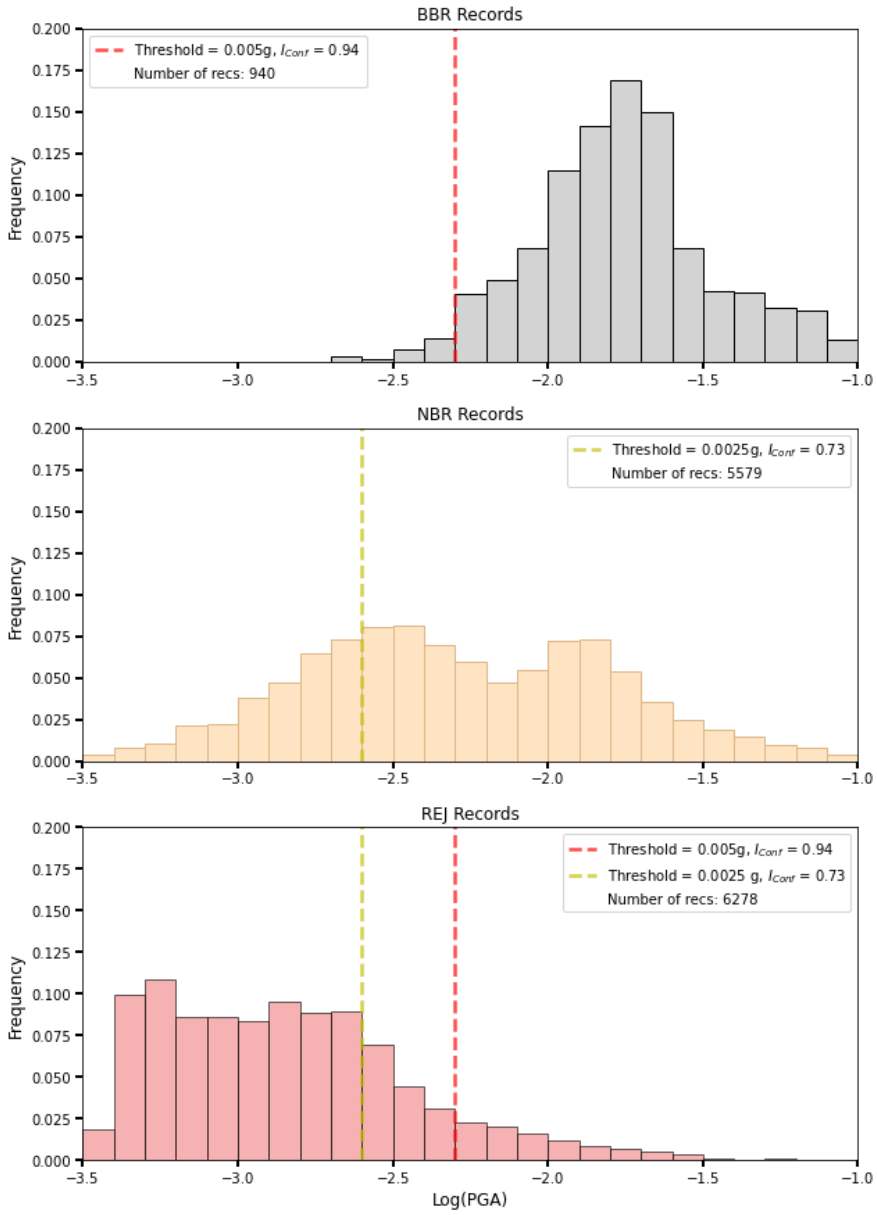


Figure 3.13. Histograms of individual-component log(PGA) for the BBR, NBR and REJ data as identified in this study. BBR / REJ and NBR / REJ thresholds are shown; the first matches the prior result from Section 3.4.1 and the second occurs at 0.0025 g.

3.4.3 Threshold Between Pre-Event Noise and Usable Records

The REJ signals considered in threshold analyses in Sections 3.4.1-3.4.2 consist of both ambient noise and earthquake shaking. The relative contributions of the two sources are generally unknown. The thresholds identified previously therefore serve to distinguish between signals recorded during earthquakes (and thus containing energy from earthquake shaking) that do or do not contain clearly evident seismic features.

An alternative objective for threshold identification is to distinguish ambient noise lacking a seismic signal from earthquake signals. This is the subject of this section, with the non-earthquake signal taken from pre-event noise windows from BBR and NBR signals, and the earthquake signal taken from the subsequent windows in those same records. The noise / seismic signal threshold is identified by applying the I_{conf} threshold analysis.

Figure 3.14 shows histograms of PGA for the two groups. The top histogram contains component PGAs for earthquake signals (both BBR and NBR). The bottom histogram contains component PGAs for pre-event noise from BBR and NBR signals. The earthquake signals range from 0.00063 to 0.06g (approximate 5-95% interval) and have some overlap at the low end of the distribution with pre-event noise PGAs, which range from 0.0004 to 0.006g. As in Section 3.4.1, I_{conf} was computed using Eq. 3.1 for different values of PGA_{th} and the threshold was identified as 0.0013g by minimizing I_{conf}. This threshold is shown in Figure 3.13 as a vertical red line and the minimized I_{conf} value is 0.79.

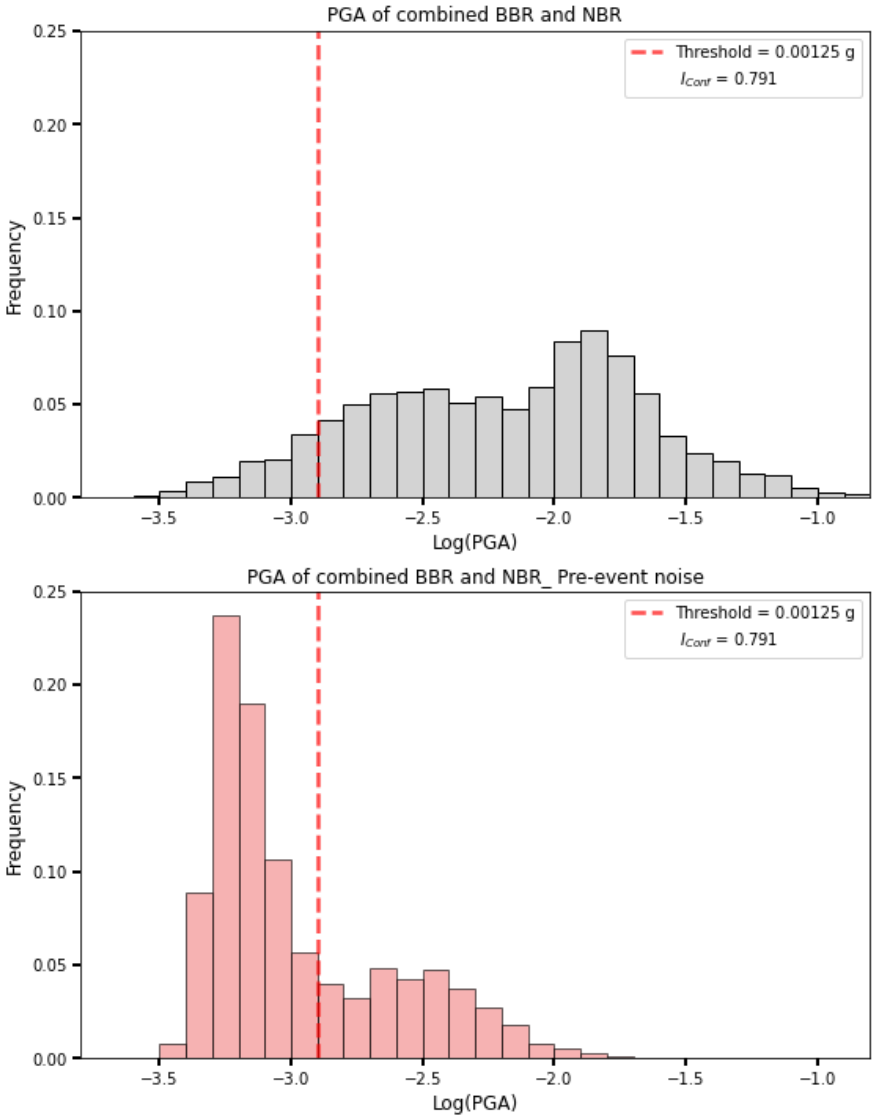


Figure 3.14. Histograms of individual-component log(PGA) for the seismic signal (BBR and NBR) and pre-event noise data as identified in this study. The threshold occurs at 0.0013 g, for which 21% of records are misclassified.

4 Data Comparisons

4.1 CO-LOCATED SENSOR COMPARISONS

An important step in the evaluation of the usability of ground motions recorded by CSN stations is to compare with ground motions recorded by non-CSN/traditional network sensors that have been used in previous studies (i.e., NGA projects). Such comparisons are most robust when sensors from both networks share the same location and both record a given event. Such co-located sensors were deliberately placed by CSN system operators (including the 6th author) to be verified against stations in the state-wide ground motion database (GMDB) (Buckreis et al. 2023a). These locations were identified by using the nearest neighbor tool in QGIS. This list was further refined by removing sites with sensors located in the upper floors of buildings. The resulting list included three sites with co-located CSN and non-CSN sensors. After searching through the event catalog, it was discovered that all three co-located station sites had ground motion recordings from at least two of the same events (the 2019 **M**6.48 Searles Valley and **M**7.06 Ridgecrest). In addition, two other earthquakes (2020 **M**4.17 Pacoima and **M**4.54 El Monte) provided ground motion records for two out of the three co-located station sites. More details are provided in Table 4.1.

Figure 4.1 shows the locations of the three sites where CSN sensors are co-located with non-CSN sensors. The non-CSN sensor at the site in the basin, 14403, has been a 200Hz Kinemetrics Episensor accelerometer since 2021. Prior to that it was a 200Hz Kinemetric FBA-11 accelerometer. WNS, located in the Hollywood Hills, has 100Hz Kinemetrics Episensor ES-DH and ES-T accelerometers. PASC, located in the San Rafael Hills, has a 100Hz MBB-2 Velocity Transducer seismometer.

Figures 4.2 - 4.4 compare the response spectra derived from the shared events recorded at the three co-located sites. Figure 4.2 shows slightly lower response spectra at the CSN sensor (T000337) compared to the non-CSN sensor (PASC) for the two large magnitude events recorded at long distance (2019 **M**6.48 Searles Valley and **M**7.06 Ridgecrest; rupture distances R_{rup} = 170.5 km and 177.1 km). This is more pronounced at shorter periods (approx. $T < 0.15$s) but the difference is still minimal. At longer periods ($T > 0.5$s) the differences are very small. Overall, there is good agreement between the data recorded by CSN and non-CSN sensors at this site, which are both housed in a vault located on bedrock in the San Rafael Hills.

Table 4.1. List of co-located CSN and non-CSN sensor sites and the recorded Earthquake Events with low-pass and high-pass corner frequencies. A row is shown for an event and site if the CSN records are BBR or NBR and a non-CSN record is available.

CSN Station	Non-CSN Station	Event	Magnitude	Class	RotD50 PGA(g) (CSN)	RotD50 PGV (cm/s) (CSN)	RotD50 PGA(g) (Non-CSN)	RotD50 PGV (cm/s) (Non-CSN	High pass fc Avg., Hz (CSN)	Low pass fc Avg., Hz (CSN)	High Pass fc Avg., Hz (Non-CSN)	Low Pass fc Avg., Hz (Non-CSN)
T000337	PASC	La Habra (2014)	5.17	BBR	0.002	0.152	0.003	0.186	0.446	11.526	0.042	41.664
		Searles valley (2019)	6.48	NBR	0.004	0.729	0.004	0.915	0.161	8.000	0.030	30.069
		Ridgecrest (2019)	7.06	BBR	0.006	1.502	0.007	1.889	0.111	10.643	0.027	38.575
T000890	WNS	Searles valley (2019)	6.48	NBR	0.004	0.644	0.007	1.548	0.149	6.982	0.032	29.178
		Ridgecrest (2019)	7.06	NBR	0.005	1.931	0.008	3.965	0.099	8.349	0.023	35.505
		Pacoima (2020)	4.17	NBR	0.025	0.648	0.024	0.808	2.646	12.500	0.085	38.171
		El Monte 2020	4.54	REJ	N/A							
T001250	14403	Searles valley (2019)	6.48	NBR	0.011	1.876	0.012	1.913	0.065	8.849	0.060	17.860
		Ridgecrest (2019)	7.06	BBR	0.015	4.848	0.016	5.196	0.072	10.049	0.050	22.124
		El Monte 2020	4.54	BBR	0.052	2.124	0.064	2.556	0.530	18.000	0.132	81.833

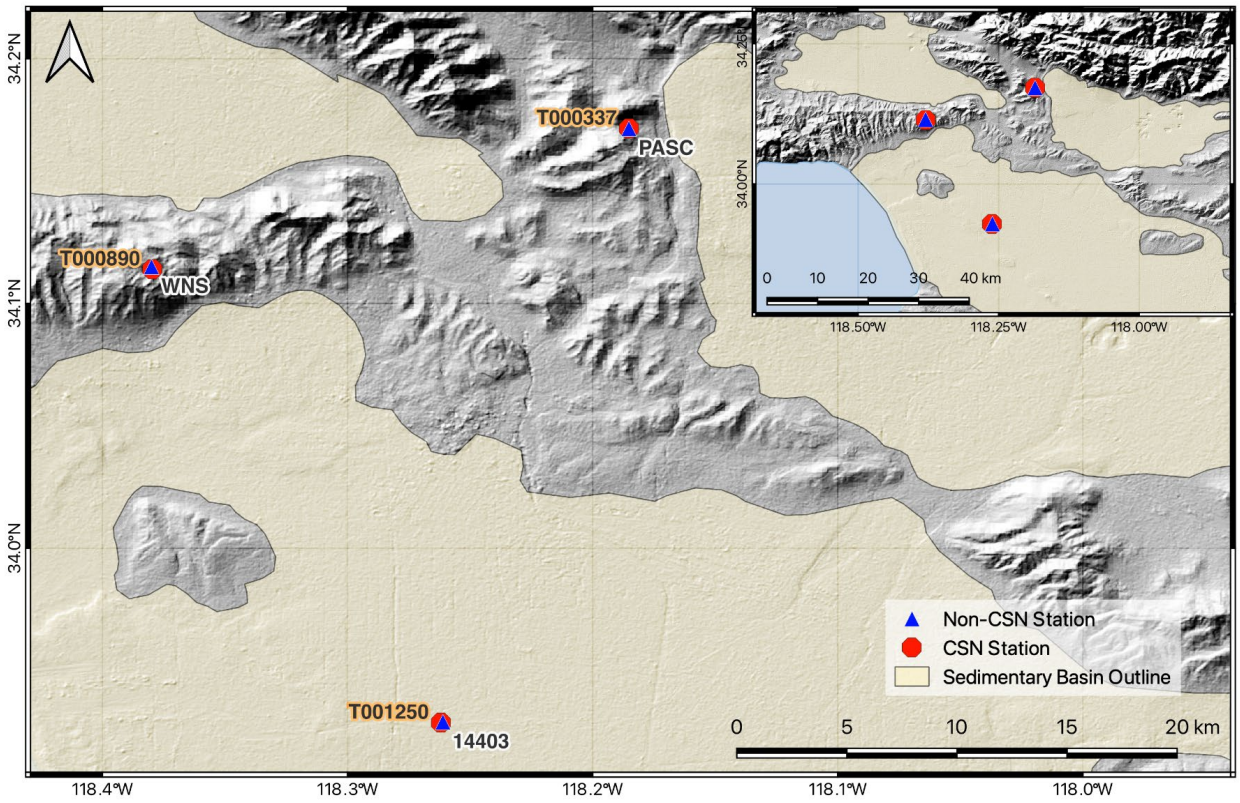


Figure 4.1. Map of co-located CSN and non-CSN stations.

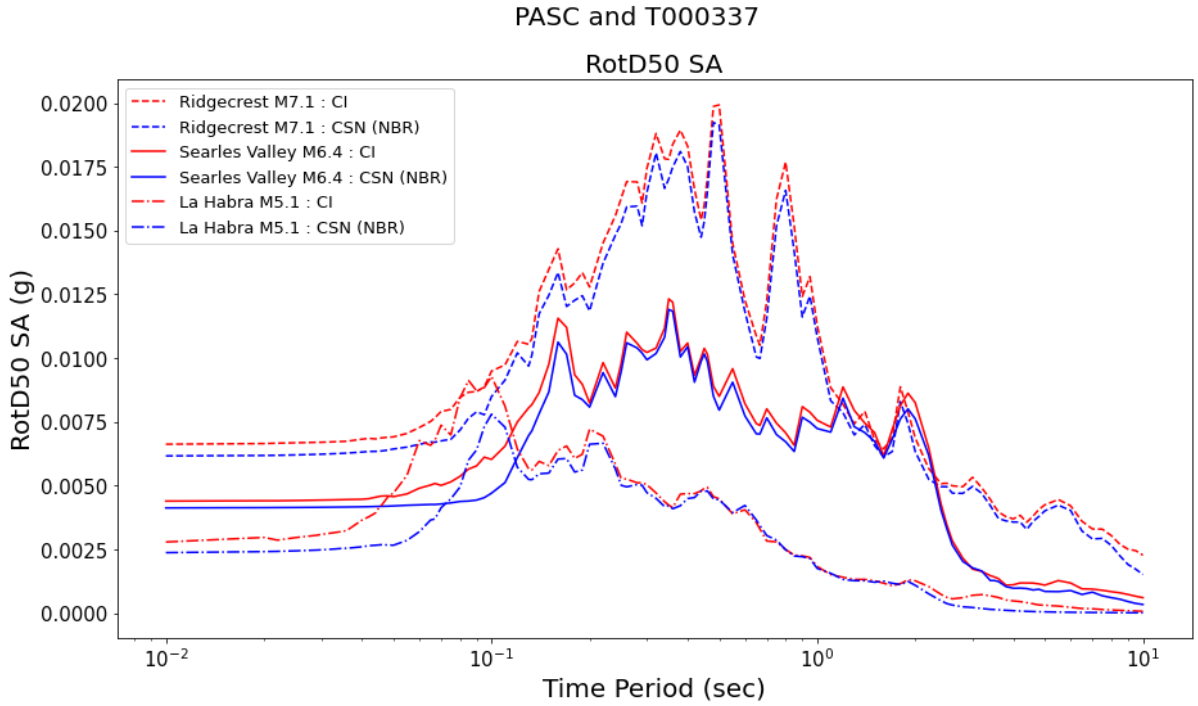


Figure 4.2. Comparison of the response spectra between PASC (non-CSN station) and T000337 (CSN station)

Figure 4.3 compares response spectra between a CSN sensor (T000890) and a non-CSN sensor (WNS) for the two large magnitude and long distance events as the prior pair, but also a smaller magnitude and near distance event (2020 **M**4.17 Pacoima). For the large magnitude events recorded at long distance (R_{rup} = 180.0 km and 186.5 km), the CSN sensor produces response spectra that are much lower than the CSN counterpart for all periods. However, for the smaller magnitude event recorded at a short distance (R_{rup} = 22.7 km), the CSN sensor produces larger Sa at short periods (T < 0.2 sec) and smaller Sa at longer periods up to 2.0 sec, where the spectra match. At this site, both the CSN and non-CSN sensors are located a few inches apart on a slab at the bottom of a three-story RC structure in the Hollywood Hills on a granitic formation (measured V_{S30} = 1043 m/s).

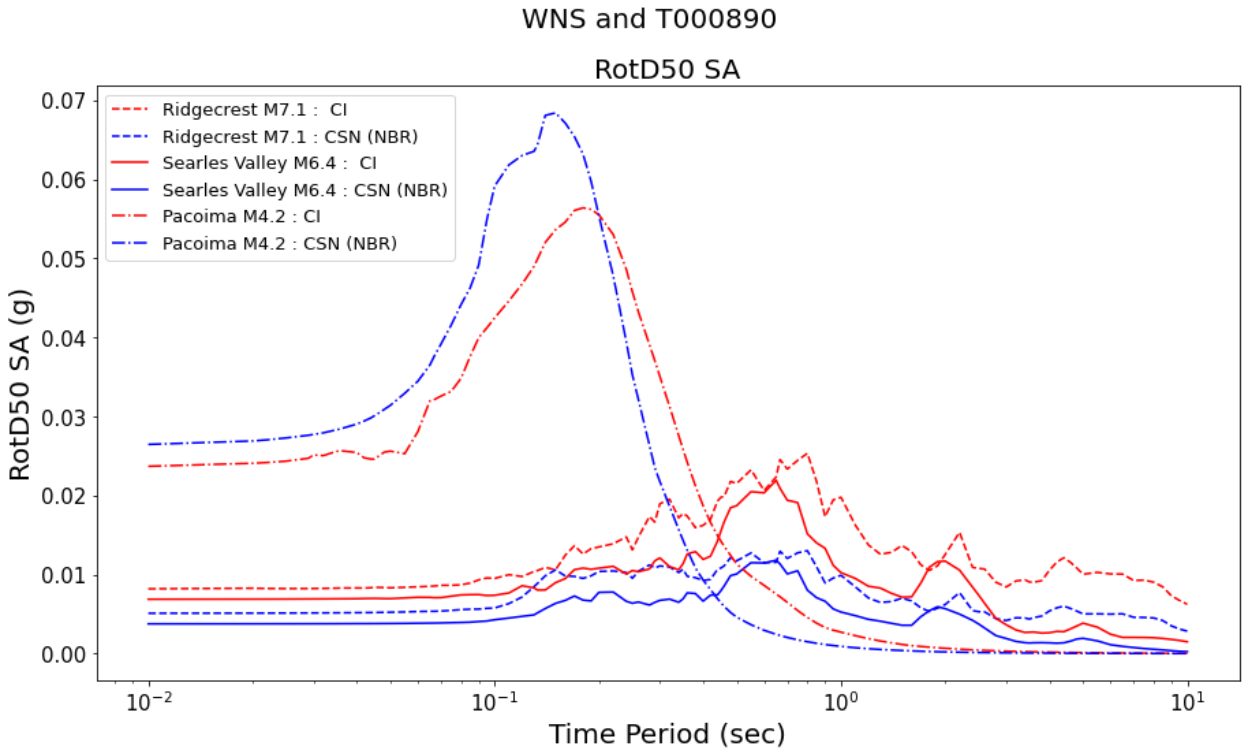


Figure 4.3. Comparison of the response spectra between WNS (non-CSN station) and T000890 (CSN station)

Figure 4.4 similarly compares response spectra for a close-distance, smaller magnitude event (2020 **M**4.54 El Monte earthquake, R_{rup} = 25.1 km) and the two larger, distant events. The spectral ordinates are very close for the larger events, but Sa from the CSN sensors are smaller for T < 0.15s and similar at longer periods. This sensor pair, 14403 and T001250, are located in the middle of the Los Angeles basin (measured V_{S30} = 317 m/s) in distinct structures. Both structures are at ground level (no basements). The CSN instrument is a school building (approximate dimensions 46 × 11 m) while the non-CSN instrument is in a separate building on the same campus (46 × 21 m) (on the CESMD web site, this instrument is indicated as being in a shed next to the school

building, but it is actually in the school building; R. Guy, personal communication, 9/17/23). The two instruments are approximately 60 m apart. Given the different foundation dimensions, different levels of kinematic soil-structure interaction are expected from base-slab averaging (NIST 2012) that would suppress the high-frequency (short-T) ordinates for the sensor on the larger, building foundation. This is consistent with the smaller ground motion amplitudes at short periods for the CSN instrument.

Overall, one of the three site pairs shows very similar results (Figure 4.2), another shows slightly different results at short period with the differences explainable based on kinematic soil-structure interaction principles (Figure 4.4), and the third has relatively large differences that may indicate a calibration problem for one or both sensors (Figure 4.3). Given the mixed results and the small number of site pairs, these analyses of co-located sensors are not conclusive. To more conclusively evaluate CSN station performance, larger separation distances must be considered, which is the subject of the next section.

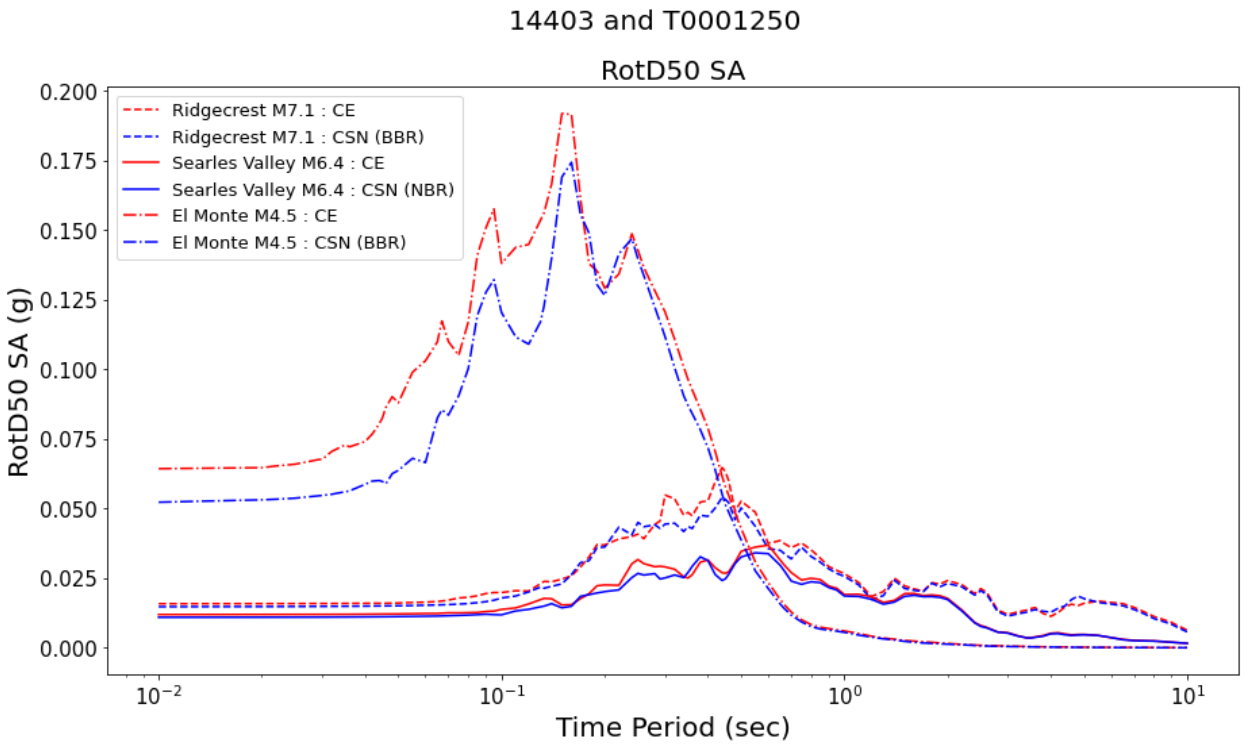


Figure 4.4. Comparison of the response spectra between 14403 (non-CSN station) and T001250 (CSN station)

4.2 PROXIMATE SENSOR COMPARISONS

The co-located sensor analysis in Section 4.1 involves a small number (three) of free-field station pairs. To broaden the comparison, in this section we consider "proximate" CSN and non-CSN

sensors, for which we consider stations that meet two criteria: (1) the stations are separated by ≤ 3 km and (2) the stations have the same surface geology, based on the statewide map by Wills et al. (2015). Station pairs that meet these criteria are mapped in Figure 4.5 (arrows are drawn between paired stations).

For each station pair, a differential ground motion IM is computed as:

$$\delta(lnIM) = ln(IM_{csn}) - ln(IM_{net}) \tag{4.1}$$

where the 'csn' subscript indicates the IM is from the CSN station and the 'net' subscript indicates the IM is from the non-CSN station. Both IMs are taken from individual as-recorded components of ground motion (generally north-south and east-west). The average value of $\delta(lnIM)$ is denoted μ_δ.

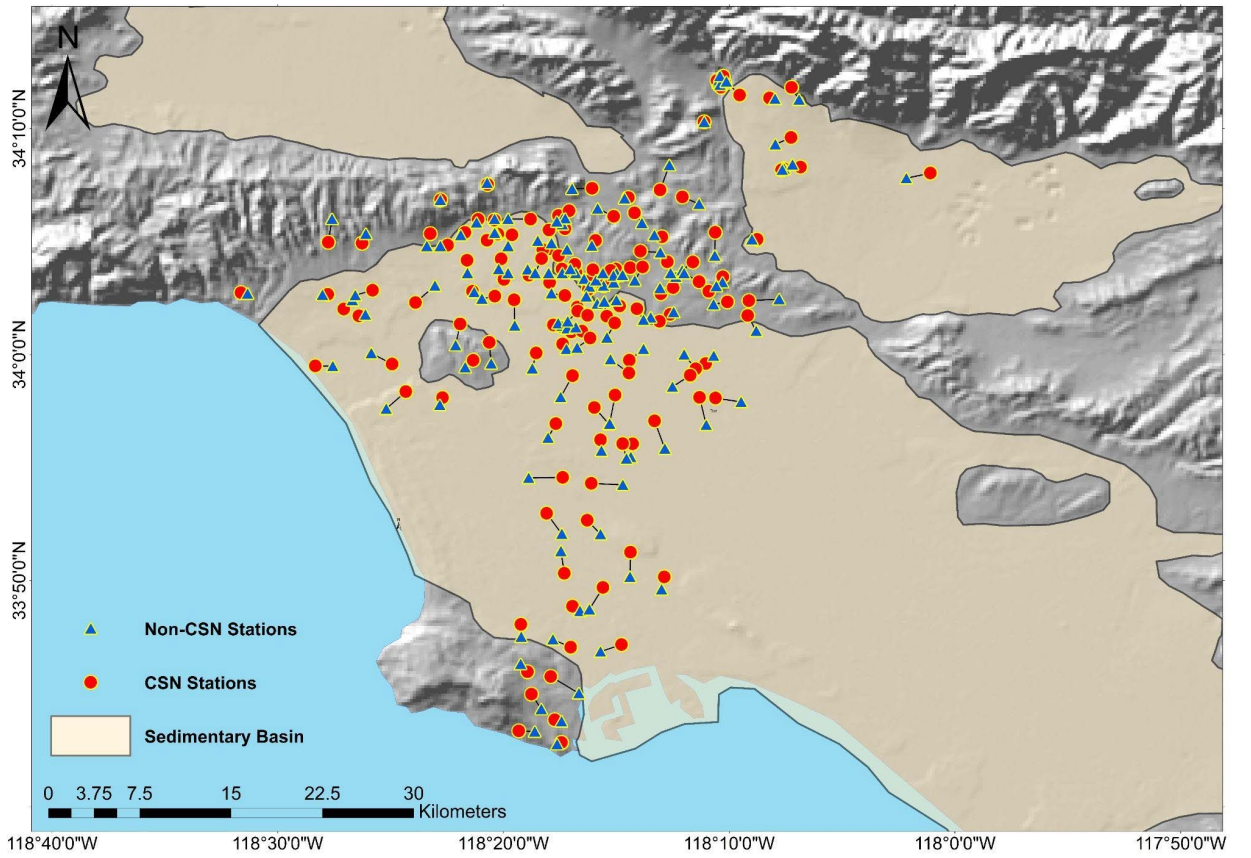


Figure 4.5. Map showing proximate CSN and non-CSN stations (160 pairs), defined by separation distances ≤ 3 km and matching surface geologies as provided by Wills et al. (2015)

Figure 4.6 plots $\delta(lnIM)$ vs separation distance for cases in which the CSN records are BBR and the IM is PGA. The mean difference in this case is μ_δ = -0.017 with a standard error of the mean of 0.071. These results show that the CSN PGAs are on average slightly smaller than the non-

CSN PGAs, but that the differences are small and within the margin of error. Figure 4.7 shows the variation of $\delta(lnIM)$ with period for Sa over the period range of 0.01 to 10 sec. Data are only considered in the calculation of the binned means when both the CSN and non-CSN Sa values are within their usable ranges given the data filtering (i.e., the oscillator period $T < 0.8/f_{cHP}$ for both instruments). The results in Figure 4.7 show a negative bias (CSN lower) for periods near 1.0 sec (~$0.6 < T < 2.0$ sec) and for $T >$ ~5 sec, but otherwise the two sets of IMs essentially match. The bias near 1.0 sec is about 10-15% (-0.1 to -0.15 ln units).

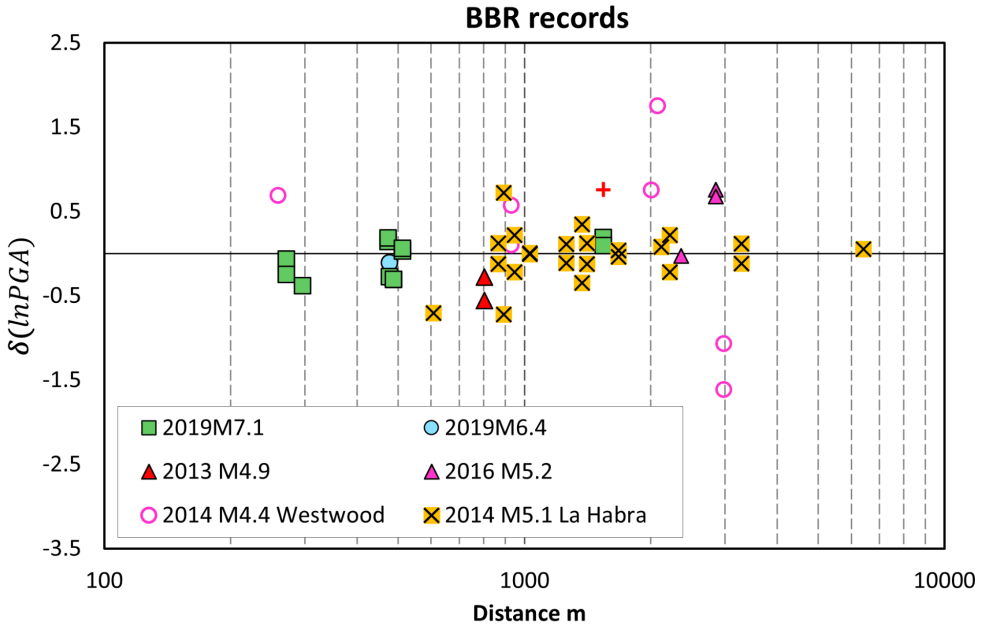


Figure 4.6. Variation of differential PGA with station separation distance for BBR CSN recordings. The mean and standard deviation of the data are μ_δ = -0.017± 0.071

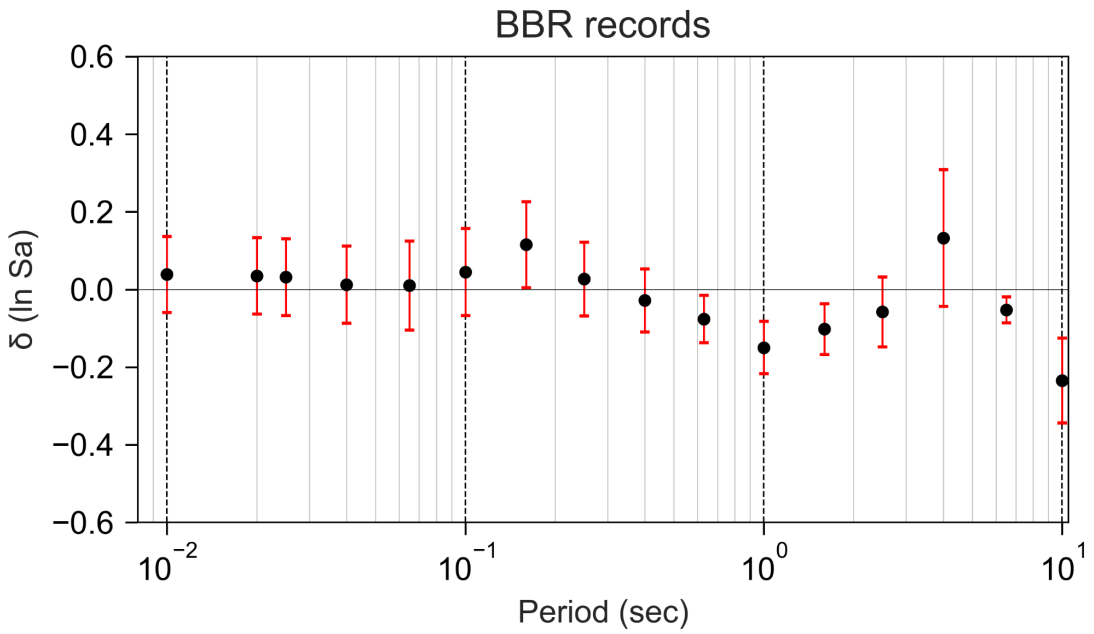


Figure 4.7. Variation of mean differential Sa with period for BBR CSN recordings.

Figure 4.8 plots $\delta(lnIM)$ vs separation distance for cases in which the CSN records are NBR and the IM is PGA. The mean difference in this case is μ_δ = -0.023 with a standard error of the mean of 0.056. These results show that the CSN PGAs are on average smaller than the non-CSN PGAs, but as with BBR data, the differences are small enough that the bias can be considered to be statistically insignificant. Figure 4.9 shows the variation of $\delta(lnIM)$ with period for Sa over the period range of 0.01 to 10 sec. The results in Figure 4.9 show a negative bias (CSN lower) over multiple period intervals including 0.05-0.1 sec, 0.4-1.0 sec, and > 3 sec. Within these period intervals, the levels of bias are small (~ -0.1 ln units) but are repeatable and statistically significant.

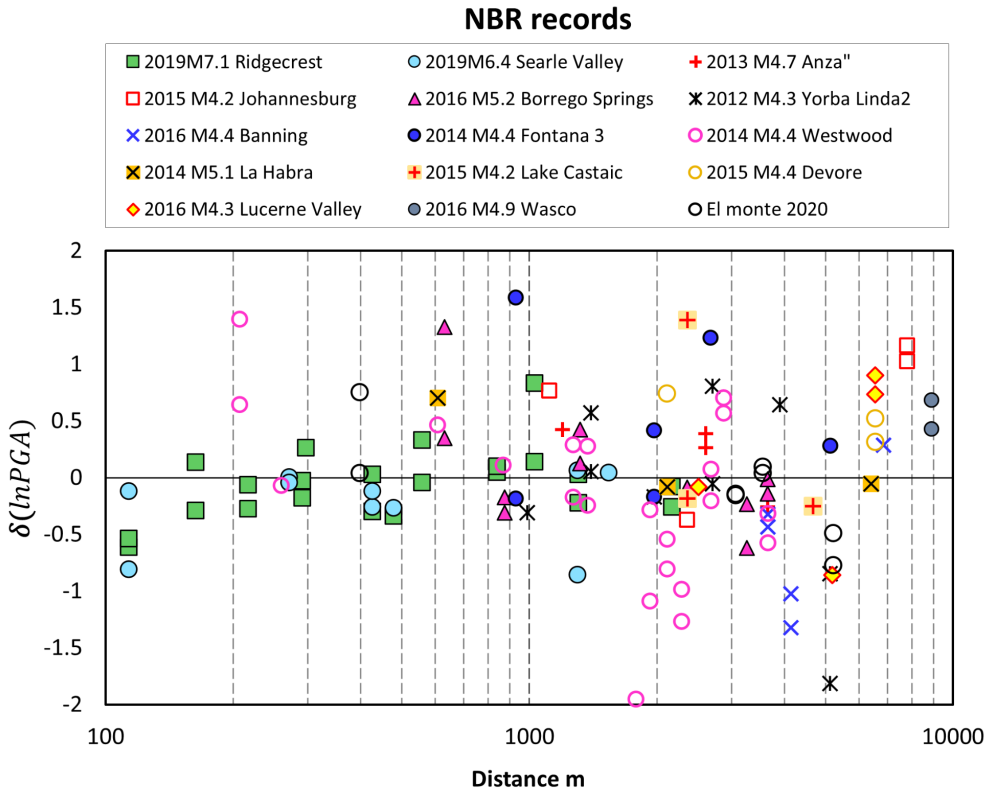


Figure 4.8. Variation of differential PGA with station separation distance for NBR CSN recordings. The mean and standard deviation of the data are μ_δ = -0.023± 0.056

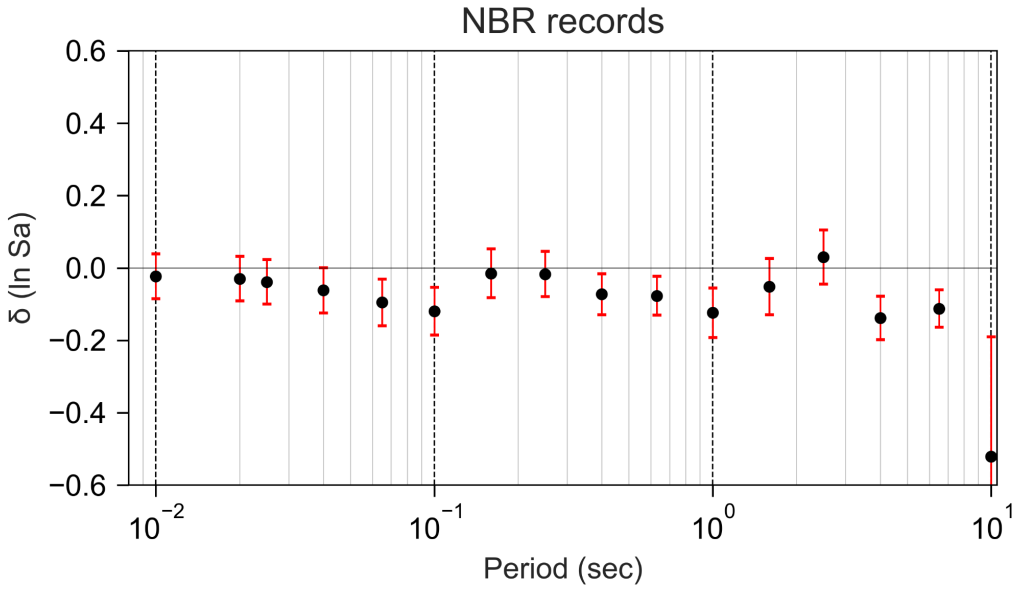


Figure 4.9. Variation of mean differential Sa with period for NBR CSN recordings.

4.3 SUMMARY OF FINDINGS

The results presented in Sections 4.1-4.2 show that BBR CSN and non-CSN records are similar within the typical usable period range of PGA to ~5 sec, with the exception of low CSN ground motions near 1.0 sec. The CSN NBR records are also unbiased for PGA, but these records have lower ground motions than the non-CSN records over a range of periods, which is expected because by definition these records have a relatively limited frequency range and hence are missing significant portions of the seismic signal at low and high frequencies. As a result, we suggest that the criteria used to define BBR recordings be used to identify usable CSN data for ground motion applications.

5 Spatial and Temporal Noise Analysis

Our review and processing of CSN data from the 29 earthquakes in Table 2.1 revealed that 53% of the downloaded records from CSN servers are REJ, meaning that the signals have the visual appearance of noise. As shown in Figure 3.9, 95% of the component PGAs for these records are between 0.0004 and 0.01 g, which exceed the nominal instrument noise level (Figure 2.3) of 0.00028g. This comprises a wide range of apparent noise levels across the data set. The aim of the work presented in this chapter is to examine whether these different noise levels have systematic trends in space and time.

Wilson et al. (2002) evaluated sources of seismic noise signals across different frequency bands, finding that low frequency noise (< 0.1 Hz) is dominated by thermal or atmosphere-driven local slab tilt effects, mid-frequency (0.1-0.3 Hz) noise levels are dominated by naturally occurring microseismic noise, and high-frequency (0.3-8 Hz) ambient vibrations are derived primarily from cultural sources. Lecocq et al. (2020) identified trains, airplanes, and industrial processes as contributing sources of cultural noise, and documented their relationships to human activities, which decreased during the Covid-19 pandemic. Diaz et al. (2017) describe how impactful cultural events (e.g. rock concerts, fireworks or football games) intensify noise signals. Whereas few of these prior studies report noise data in acceleration units that can be compared to CSN data, Clayton et al (2020) report such information for CSN data recorded at a Los Angeles area school, which generally range from 0.0015 to 0.003 g. The lowest noise amplitudes occurred in evening hours and higher amplitudes occurred in mid-afternoon.

5.1 AMPLITUDE METRICS FOR NOISE SIGNALS

We consider both REJ records and the pre-event noise portions of BBRs and NBRs to evaluate attributes of noise signals. Amplitudes of noise signals are considered using PGA and an average amplitude of the Fourier Amplitude spectrum. Individual Fourier amplitudes are computed as follows:

$$X(\omega_n) = \frac{\Delta t}{\sqrt{N\Delta t}} \sum_{k=1}^{N} x(t_k) e^{-i\omega_n t_k} \tag{5.1}$$

where $x(t_k)$ is the discretely sampled time series, N is the number of data points in the time series, Δt is the time step, n is an index for the discrete frequencies where Fourier coefficients are computed, $\Delta\omega$ is the frequency step in rad/sec, and $\omega_n = n\Delta\omega = 2\pi n/(N\Delta t)$. The mean noise amplitude $\bar{X}^{No}$ is then computed as,

$$\bar{X}^{No} = \frac{\sum_{f=1}^{N_f} X(\omega_n)}{N_f - 1} \tag{5.2}$$

where f is a frequency index that starts at 2 Hz = 4π rad /sec and N_f is the number of frequency steps to the maximum considered frequency of 20 Hz = 40 π rad / sec. The 2-20 Hz frequency range was selected because for most signals, the Fourier amplitudes of noise signals are relatively flat across that range (e.g., Figures 3.4 and 3.5).

The subsections below will examine the characteristics of noise metrics when extracted from pre-event noise (for BBR and NBR records) and the main signal (REJ records), and will examine how noise characteristics vary across the network for different times of day.

5.2 COMPARISONS OF ALTERNATE NOISE SIGNALS

Figure 5.1 shows the stations that recorded the 2019 Searles Valley earthquake, color-coded by the classification of the horizontal signals. There are appreciable numbers of stations with earthquake-like signals (BBR or NBR) and noise-like signals (REJ). Using the data for this earthquake, Figure 5.2 shows histograms of PGA and $\bar{X}^{No}$ for two groups of data:

- Group 1: pre-event noise signals for BBR and NBR recordings
- Group 2: REJ recordings

The results show a wide range of amplitudes for BBR group 1 data (PGA ≈ $10^{-3.0}$ to $10^{-2.0}$ or 0.001 to 0.01g) but a narrow range of higher amplitudes for group 2 data (PGA ≈ $10^{-2.2}$ to $10^{-1.4}$ or 0.006 to 0.04g). The narrow range of amplitudes from REJ records reflects the approximate amplitudes of ground motions from this event within the distance range of the CSN sensors (161 to 193 km). This example suggests that the pre-event noise signals may represent preferred representations of noise from CSN signals.

Figure 5.3 shows histograms for the group 1 and 2 data using both noise metrics, but now for the full data set of 29 earthquakes. These results show a wider range of amplitudes for BBR group 1 data (PGA ≈ $10^{-3.5}$ to $10^{-1.7}$ or 0.00032 to 0.02g). The group 2 data ranges from PGA ≈ $10^{-3.4}$ to $10^{-1.1}$ or 0.0004 to 0.08g, which is similar to group 1 but with a stronger tail at high amplitudes. Regarding the two noise intensity metrics (PGA and $\bar{X}^{No}$), the latter has less dispersion, although the levels of skewness are similar.

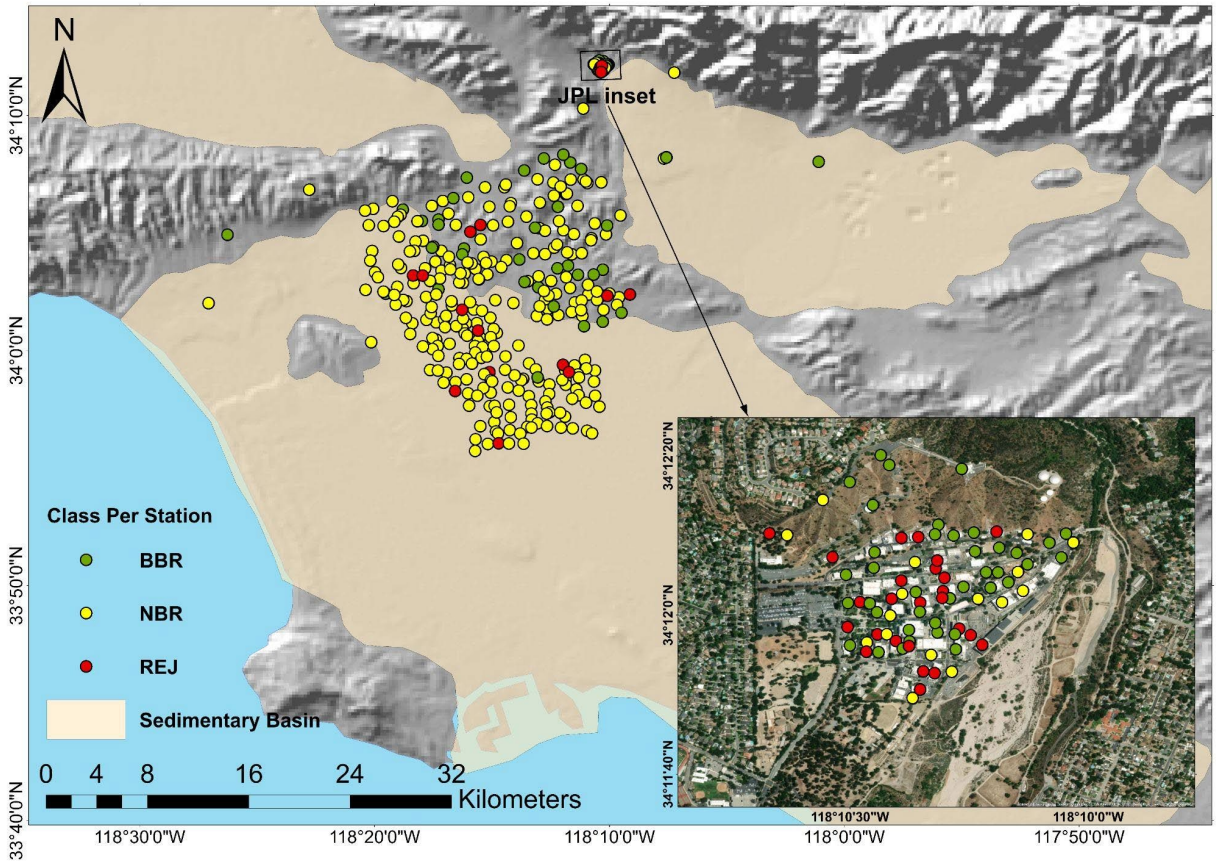


Figure 5.1. CSN stations that provide data for the 2019 **M** 6.4 Searles Valley earthquake color coded by record category. Inset shows stations on the JPL campus

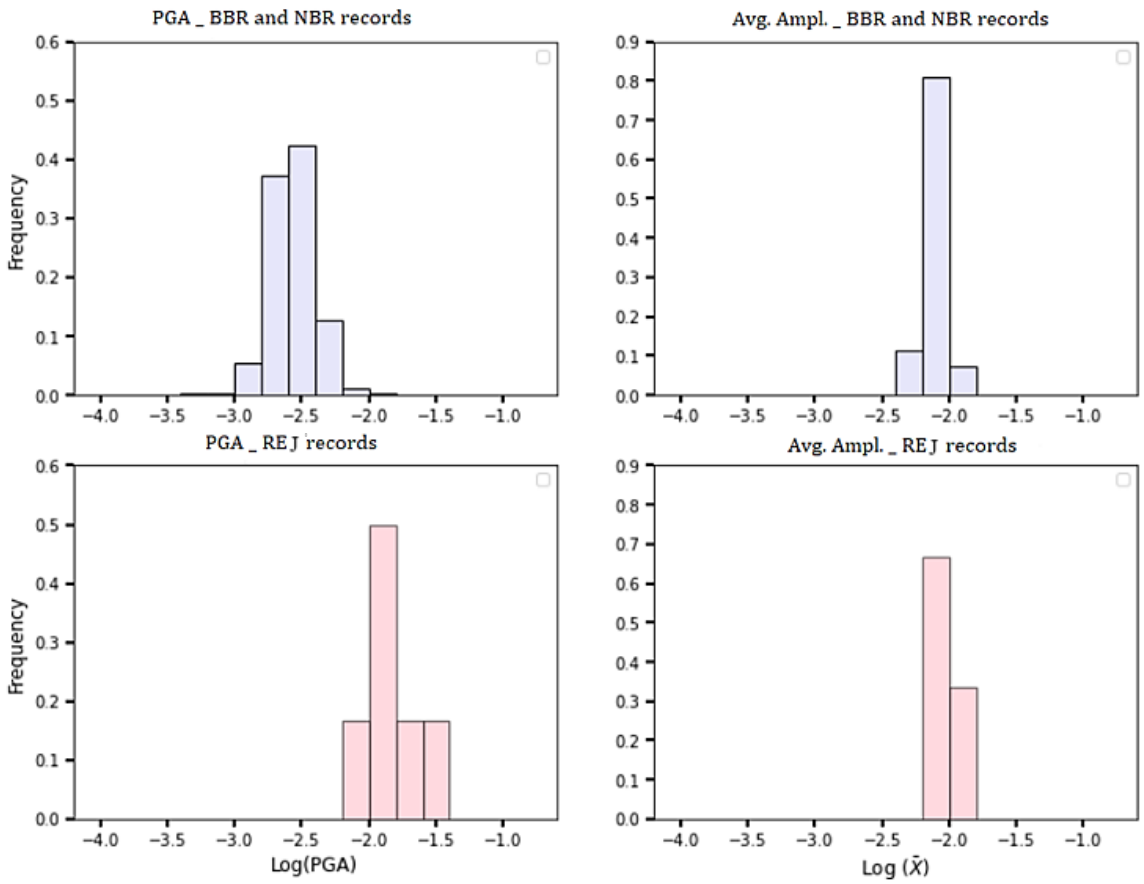


Figure 5.2. Noise amplitudes from CSN sensors for the 2019 **M** 6.4 Searles Valley earthquake as measured by pre-event noise from BBR and NBR signals (top) and REJ signals (bottom) and as represented by PGA (left; units log g) and averaged Fourier amplitude (right; units log $g \cdot s^{0.5}$).

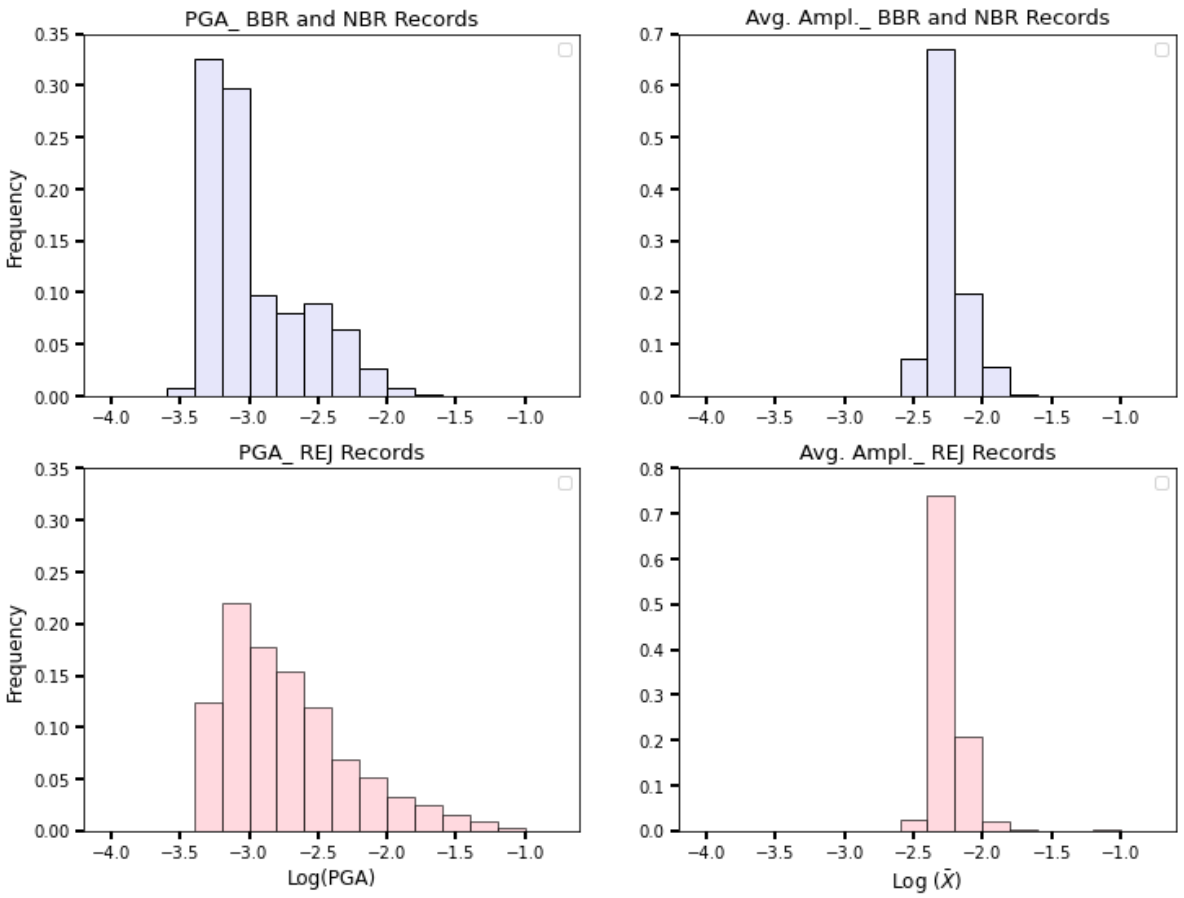


Figure 5.3. Noise amplitudes from CSN sensors for the 29 events in the database as measured by pre-event noise from BBR and NBR signals (top) and REJ signals (bottom) and as represented by PGA (left; units $\log_{10}$ g) and averaged Fourier amplitude (right; units $\log_{10}$ g·s$^{0.5}$).

5.3 TEMPORAL AND SPATIAL VARIATIONS OF NOISE

The noise from CSN instruments measured from pre-event signals and REJ records may have several contributing factors. One factor is instrument noise, which should nominally match the Phidget noise level of 0.00028g (Figure 2.3). Another factor is ambient noise (i.e., microtremor ground vibrations) at the instrument locations. Ambient noise might be expected to have location- and temporal-dependencies. Stronger ambient vibrations might be expected in urban areas of high population density such as Hollywood, whereas relatively suburban or exurban areas (such as Pasadena or the JPL campus) may have lower ambient noise levels. Moreover, for any given location, noise may have temporal variations, being stronger during daytime than evening hours.

Population density is considered as an index against which to compare noise levels. Figure 5.4 shows a map of the portion of the greater Los Angeles region where the CSN sensors are concentrated. The map is shaded by population density (number of people per 100m^2), which ranges from 0 to 2000 (data from Depsky et al. 2022), and is based on 2020 census data.

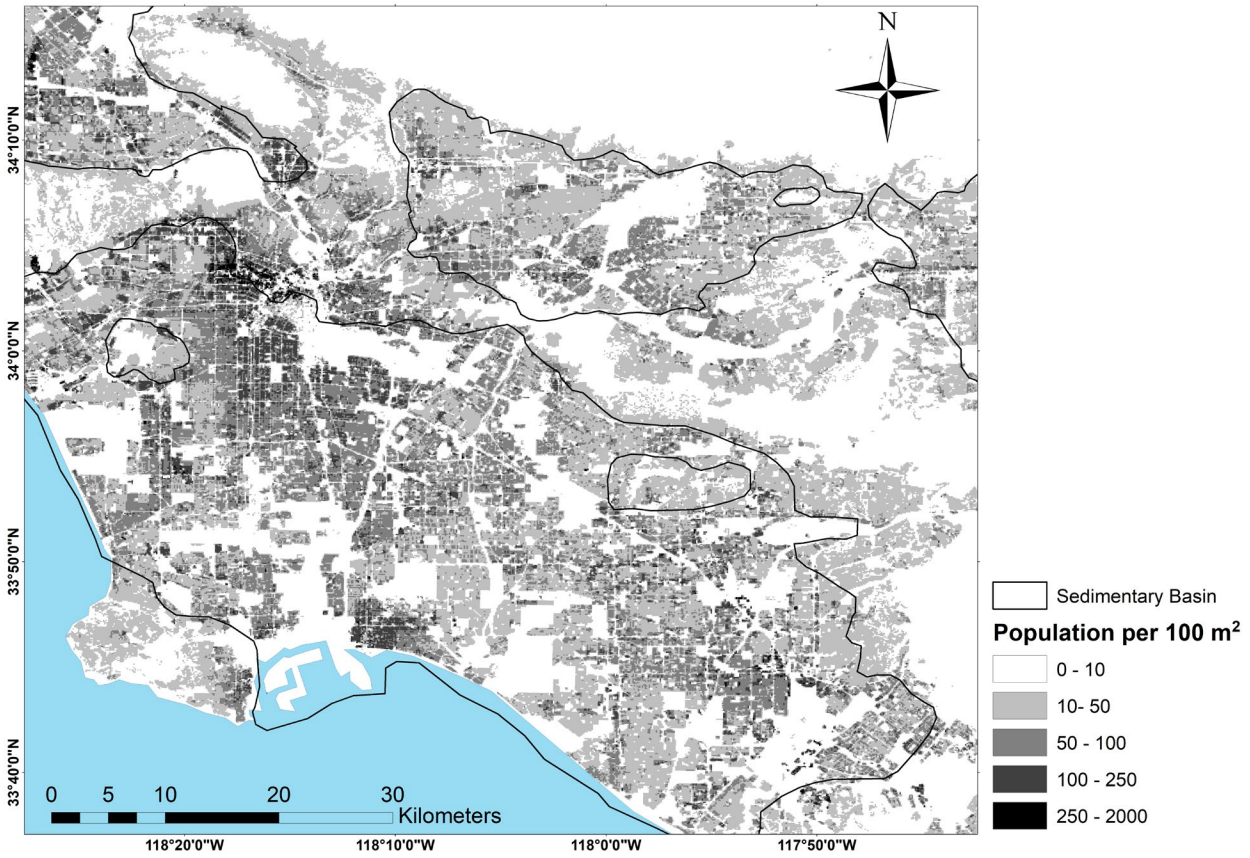


Figure 5.4. Population density from 2020 census within the portions of the Los Angeles area with the greatest concentration of CSN stations (Depsky et al., 2022).

The pattern of noise level with population density can be visualized by plotting the dependency, which is provided in Figure 5.5 for PGA using the two signal types (pre-event and REJ) and in Figure 5.6 for the $\bar{X}^{No}$ noise metric. There is no significant trend in the PGA pre-event noise data for densities of 0 to 100 people/100m^2, although a local high occurs near 50 people/100m^2. The average noise level is 0.0016 g. These levels are lower than those reported in Section 3.4, but consistent with the findings of Clayton et al. (2020). For the REJ signals, the population density trend is flat from 0 to 20 people/100m^2, but at a higher average level of 0.0045g that increases to 0.006-0.007 g at higher densities, which is relatively consistent with the thresholds identified in Section 3.4 and notably higher than reported by Clayton et al. (2020). We have not yet identified individual areas having systematically high or low noise levels. These results show that the pre-event noise signals produce lower average noise levels than do the REJ signals. Figure 5.6 shows similar patterns of the data with respect to population density for the two signal types.

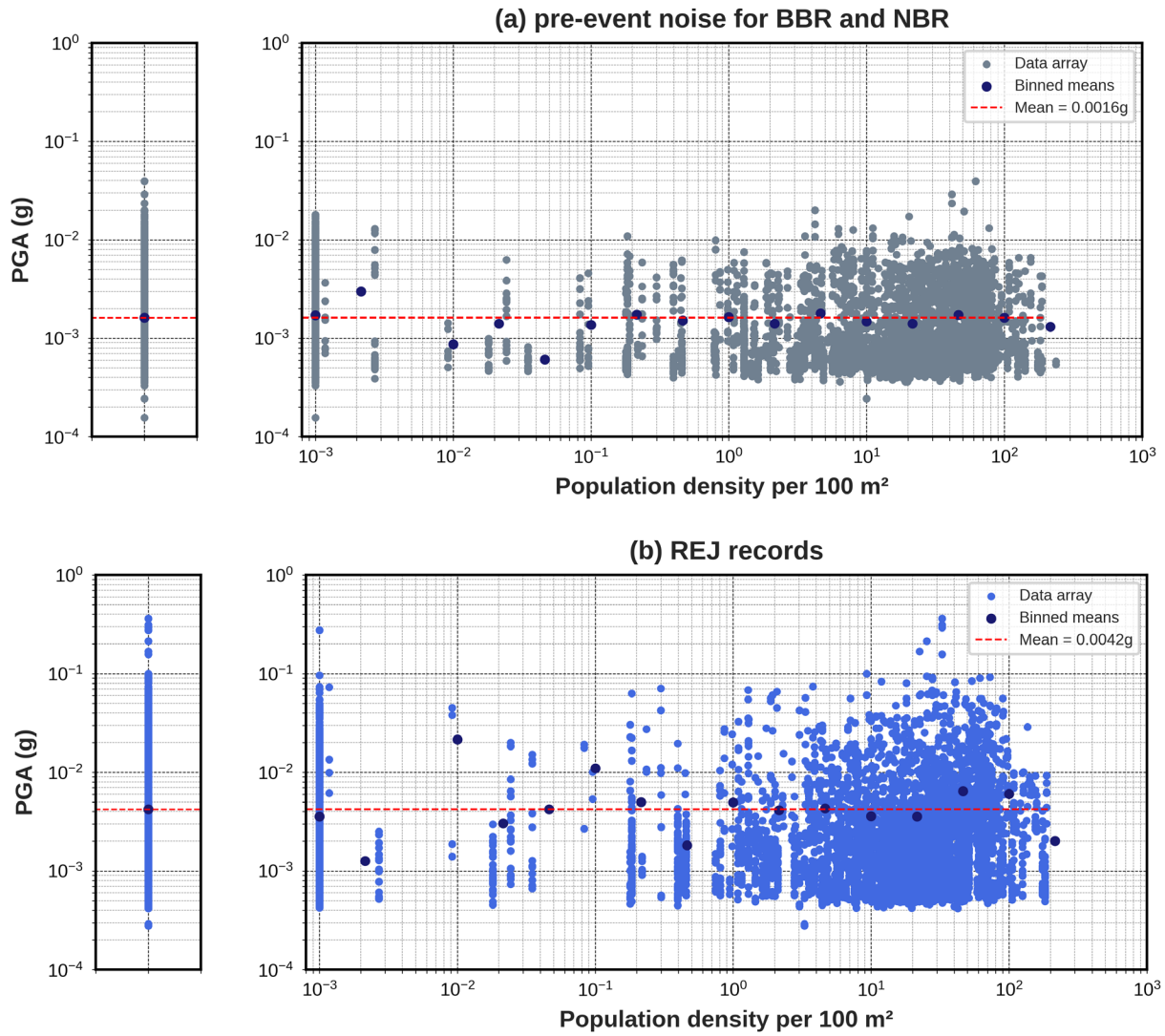


Figure 5.5. Variation with population density of (a) pre-event noise and (b) REJ signals as quantified by PGA across all events. Distribution on the left side of the figure represents the full data set. Dark blue symbols are binned means and red line is the mean for the portion of the plot at low population densities where the trend is flat.

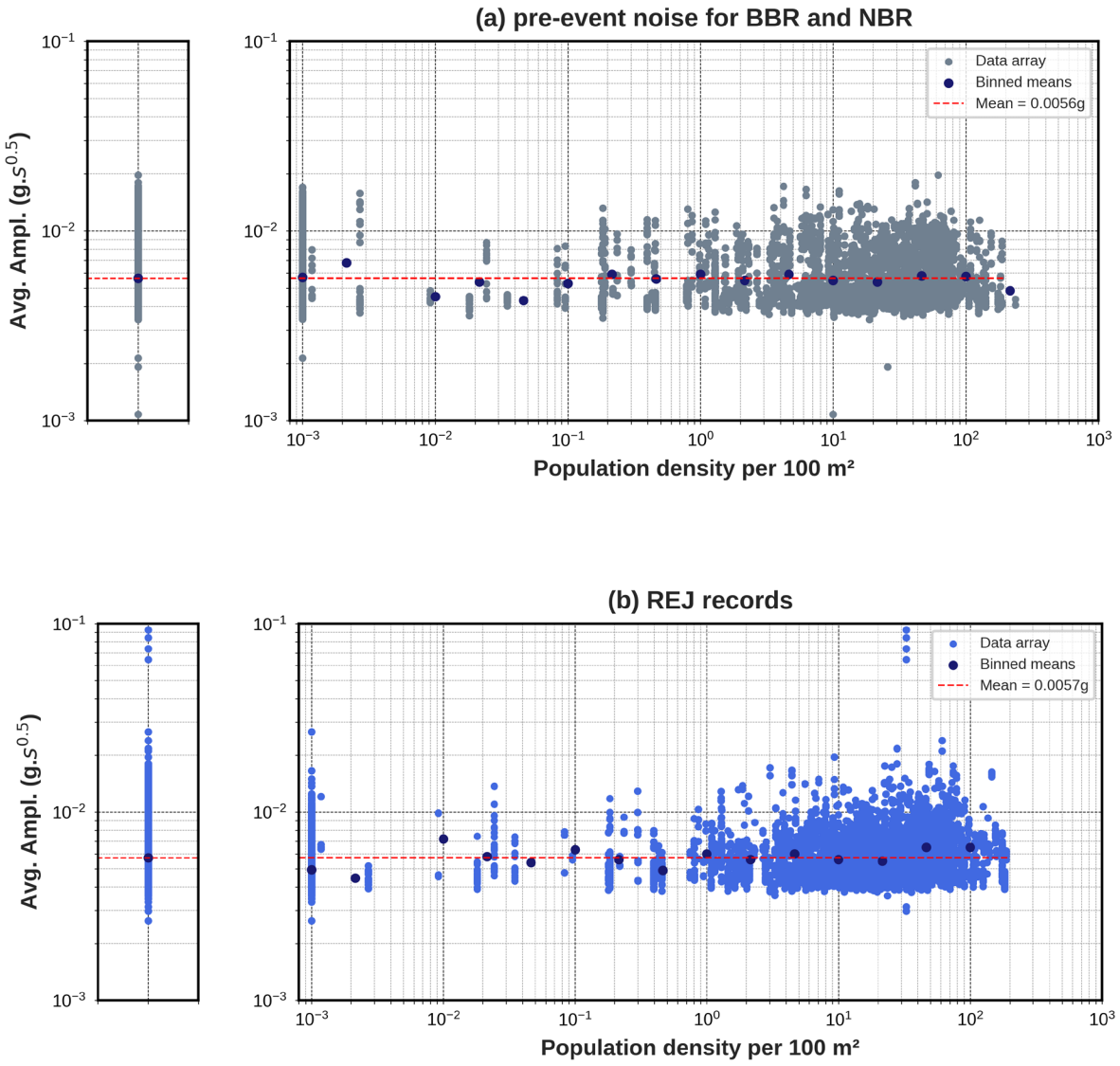


Figure 5.6. Variation with population density of (a) pre-event noise and (b) REJ signals as quantified by $\bar{X}^{No}$ across all events. Distribution on the left side of the figure represents the full data set. Dark blue symbols are binned means and red line is the mean for the portion of the plot at low population densities where the trend is flat.

Next we turn to temporal variations of noise metrics. Figure 5.7 shows a histogram of earthquake times for the 29 considered events (Table 2.1) using the local origin time for the event (Pacific daylight or standard time, depending on which was in effect on the event date). The results show that 17 events occur during time periods when human activity would be expected to be significant (roughly 9 am to 10 pm) and 8 events occur during relatively quiet time periods (roughly 12 pm to 6 am).

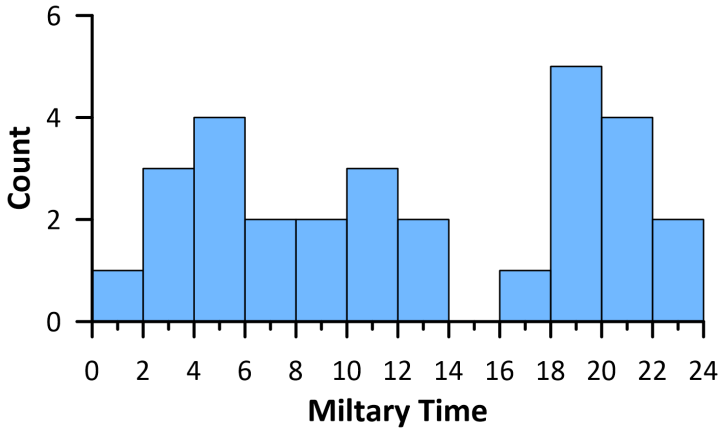


Figure 5.7. Histogram of event time by time of day using military time (0 = midnight, 12 = noon). During evening hours (0 to 6 hr) there are 8 events. During peak daytime hours (9 to 22 hr) there are 17 events.

Figure 5.8 and 5.9 show the time variations of noise levels using the PGA and $\bar{X}^{No}$ noise metrics, respectively, for all events. These results show that events with the highest noise levels tend to occur in evening hours, although this is not repeated across all events at that time of day. Overall, the trend of the data with time is inconclusive.

Based on the results in Figures 5.6-5.7 and 5.8-5.9, population density has more consistent and repeatable impact on noise levels than does event time of day. Relatively quiet conditions occur where population densities are low (< 20 people/100m^2). These results suggest that refinement of the noise thresholds identified in Section 3.4 could be made based on location.

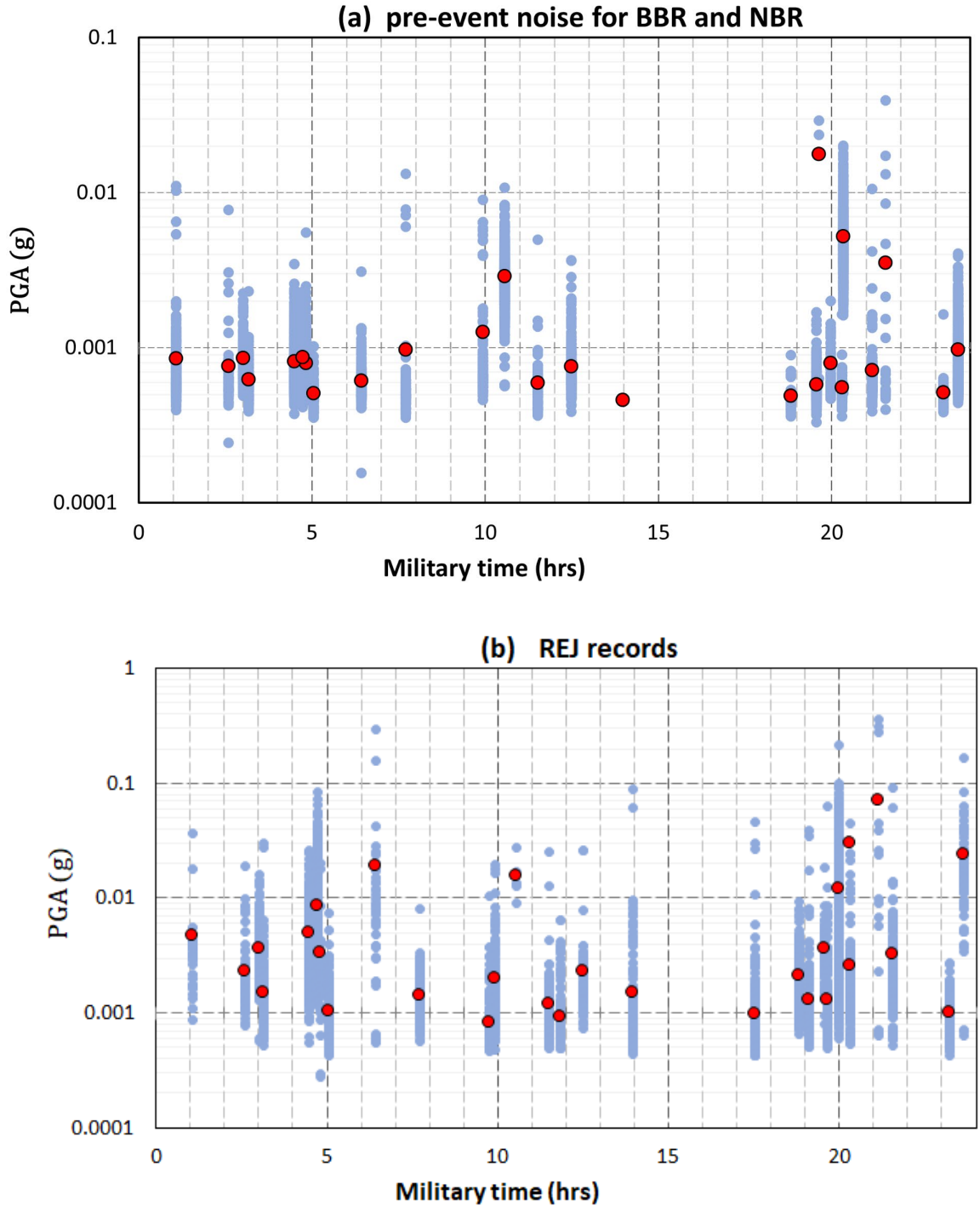


Figure 5.8. Variation with time of day of (a) pre-event noise and (b) REJ signals as quantified by PGA across all events. Red symbols are event means. Time zone is Pacific (0 = midnight, 12 = noon)

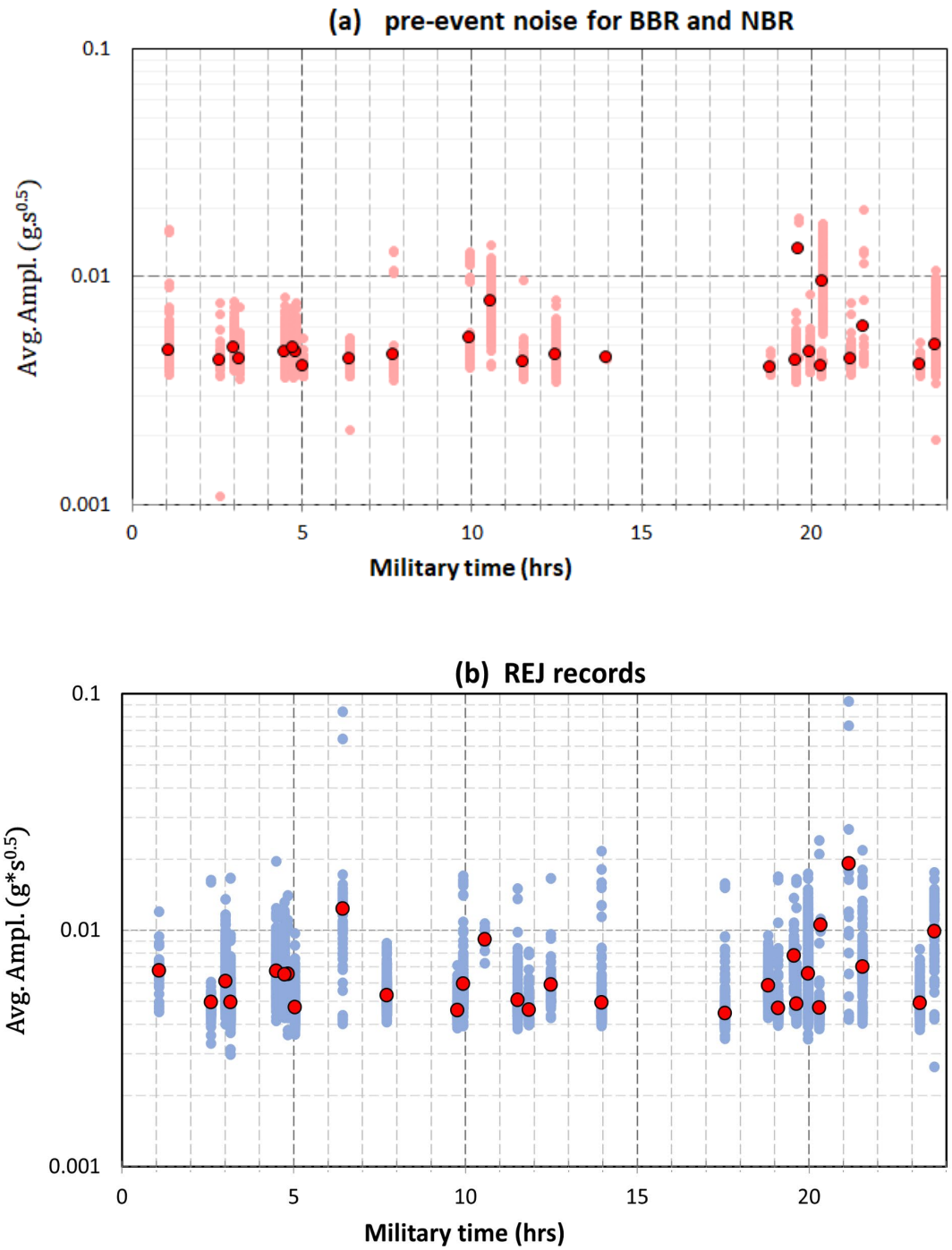


Figure 5.9. Variation with time of day of (a) pre-event noise and (b) REJ signals as quantified by $\bar{X}^{No}$ across all events. Red symbols are event means. Time zone is Pacific (0 = midnight, 12 = noon)

6 Usable Distance Ranges for CSN Data

A typical step in ground motion model (GMM) development is to apply data screening criteria, which are used to identify the data that will be considered in model development. One aspect of data screening is to minimize the effects of sampling errors from low-amplitude ground motions.

In the case of triggered instruments, sampling errors occur when the ground shaking level at a site falls below the trigger threshold. In the case of continuously recording instruments, which is the case for CSN stations, sampling errors occur when signal amplitudes are not stronger than the instrument noise threshold. This is typically the case at large distances, and is more pronounced for small magnitude events than large magnitude events. For a magnitude-distance condition where the mean ground motion amplitude is near the threshold, unusually strong motions that exceed trigger thresholds or that fall above the noise floor are recorded. However, weaker motions that do not exceed trigger thresholds or that fall near the noise floor are not available. Accordingly, the problem is not that no records are obtained for such conditions, but that the recorded ground motions become biased as a population towards larger values.

To overcome this problem, data selection criteria are applied during GMM development so as to screen out data that is potentially subject to sampling bias. In effect, for a given sensor network, these criteria provide the maximum source-to-site distance (R_{max}) that can be used with confidence. In the NGA-West2 project (Bozorgnia et al. 2014), conservative criteria were used consisting of magnitude-R_{max} relations that depended on instrument type (analogue; older, low-resolution digital; modern, high-resolution digital), as shown in Figure 6.1. These relations were established primarily on the basis of conjugate expert judgment (among model developers).

Here we establish a predictive model for R_{max} for CSN data. This is done by establishing the distance for which few (< 5%) of ground motions from a given event would be expected to fall below a threshold that is applicable to the network. Ground motion recordings with distances smaller than R_{max} could be used with confidence in GMM development because essentially the full statistical distribution of ground motions would be sampled within the data set.

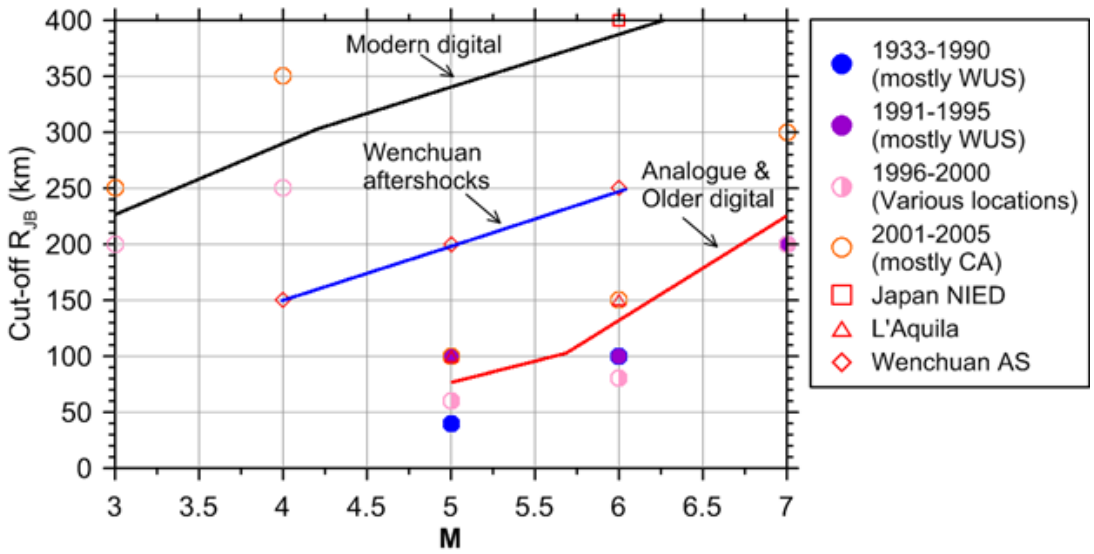


Figure 6.1. Variation of limiting distance (referred to as R_{max} in this report) with magnitude for different instrument types as used in NGA-West2 project. From Boore et al. (2014).

To estimate R_{max}, we plot in Figure 6.2-6.4 the distance variation of the 5th percentile PGA (mean minus two within-event standard deviations) for five magnitudes (**M** 4, 5, 6, 7, 8) and two site conditions (V_{S30} = 300 and 760 m/s). The R_{max} value is the distance at which the PGA curve intersects the threshold PGA. Three thresholds are considered: (1) 0.005g, which is based on the threshold analyses presented in Section 3.4.1 (Figure 6.2); (2) 0.003g, which is loosely based on the average PGA for low populate density areas described in Section 5.2 (Figure 6.3); and (3) 0.0015g, which is based on the noise / seismic signal threshold analysis in Section 3.4.3 (Figure 6.4). Additional refinements of the limiting distances are possible if the event term for an earthquake is known (used to adjust the GMM prediction across all distances and site conditions up or down) or if a more refined noise threshold for a given subregion or event time can be determined.

In past work, R_{max} was estimated using truncated regression of data from a specific region (e.g., Contreras et al. 2022). The truncated regression was performed to develop a regional GMM that was used to develop intensity measure - distance curves like those shown in Figures 6.2(a)-6.4(a). Truncated regression of regional data was not needed in this case because the Boore et al. (2014) path model has been demonstrated to perform well for southern California data from arrays with sensitive instruments for which truncation issues are not at issue across the distance range considered (Nweke et al. 2022; Buckreis et al. 2023b). Accordingly, the Boore et al. (2014) model is used without modification to produce the distance attenuation curves shown in Figures 6.2(a)-6.4(a). The resulting values of R_{max} are plotted in Figure 6.2(b)-6.4(b).

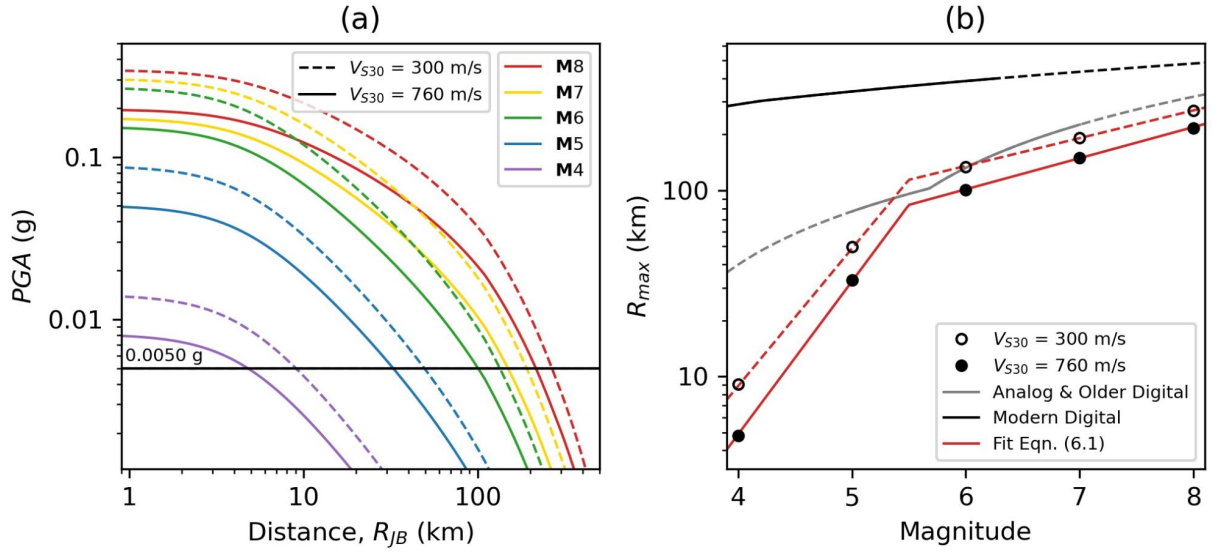


Figure 6.2. (a) R_{max} evaluation for events of magnitude 4 to 8 and V_{S30} = 300 and 760 m/s; the R_{max} value corresponds to the intersection of the noise threshold (0.005g) with the mean minus two within-event standard deviation PGAs. (b) Variations of R_{max} with magnitude for the two V_{S30} values (including a fit relationship), with prior recommendations from NGA-West2 shown for comparison.

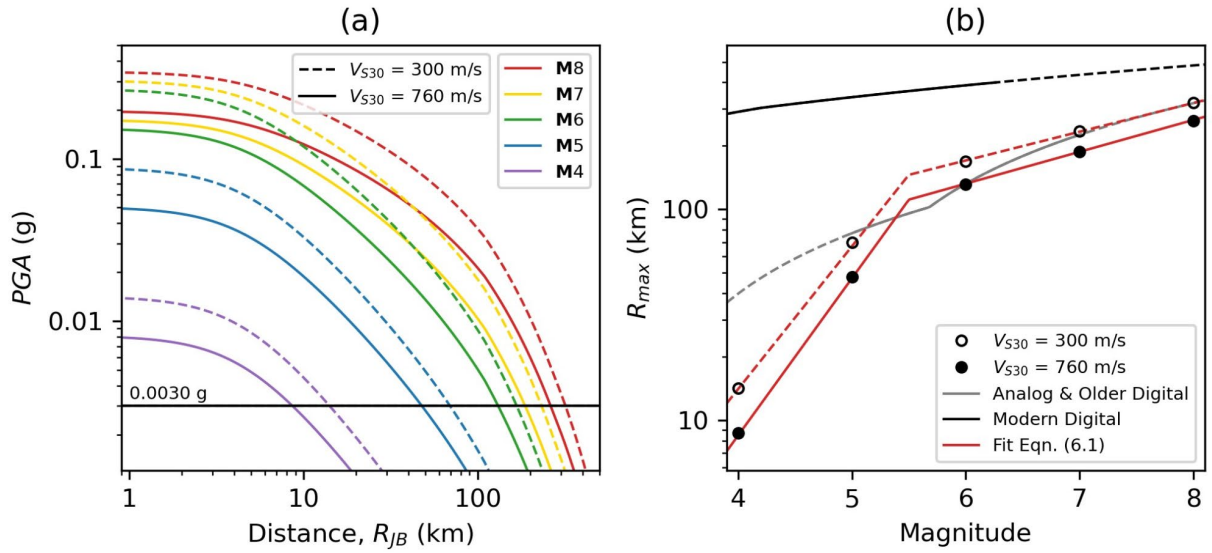


Figure 6.3. (a) R_{max} evaluation for events of magnitude 4 to 8 and V_{S30} = 300 and 760 m/s; the R_{max} value corresponds to the intersection of the noise threshold (0.003g) with the mean minus two within-event standard deviation PGAs. (b) Variations of R_{max} with magnitude for the two V_{S30} values (including a fit relationship), with prior recommendations from NGA-West2 shown for comparison.

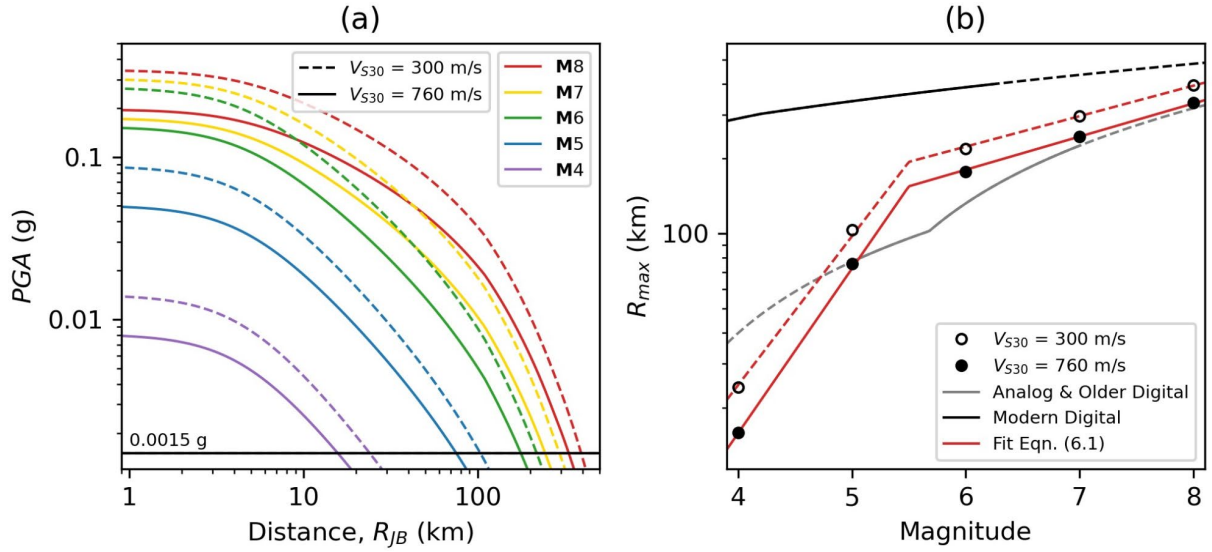


Figure 6.4. (a) R_{max} evaluation for events of magnitude 4 to 8 and V_{S30} = 300 and 760 m/s; the R_{max} value corresponds to the intersection of the noise threshold (0.0015 g) with the mean minus two within-event standard deviation PGAs. (b) Variations of R_{max} with magnitude for the two V_{S30} values (including a fit relationship), with prior recommendations from NGA-West2 shown for comparison.

As shown in Figures 6.2(b)-6.4(b), the resulting CSN R_{max} values are smaller than the recommendations from Boore et al. (2014) for modern digital instruments and are similar to the prior recommendations for analog and older digital events (for $\mathbf{M} > 5$ events). This is expected given the relatively large noise levels associated with CSN sensors (Figure 2.3).

Fits to the R_{max} - $\mathbf{M}$ data are provided using the following expression,

$$\ln(R_{max}) = \begin{cases} c_1\mathbf{M} + c_2 & \mathbf{M} \leq M_h \\ c_3\mathbf{M} + c_2 + M_h(c_1 - c_3) & \mathbf{M} > M_h \end{cases} \quad (6.1)$$

Where M_h = 5.5 and c_1, c_2 and c_3 are coefficients estimated using least-squares regression, and whose values are provided in Table 6.1. Coefficients are provided in Table 6.1 for the V_{S30} = 300 m/s and 760 m/s site conditions considered above, and also 400 m/s, which is the median condition across the CSN station network. The fits for the V_{S30} = 300 m/s and 760 m/s site conditions are shown in Figures 6.2b - 6.4b.

Table 6.1. Coefficients for limiting distance relationship in Eqn. (6.1).

Noise Threshold (g)	V_{S30} (m/s)	c_1	c_2	c_3
0.0015	300	1.3762 ± 0.0044	-2.3002 ± 0.0216	0.2819 ± 0.0024
	400	1.4210 ± 0.0042	-2.6128 ± 0.0206	0.2887 ± 0.0023
	760	1.5223 ± 0.0036	-3.3297 ± 0.0177	0.3066 ± 0.0020
0.0030	300	1.5584 ± 0.0034	-3.5889 ± 0.0166	0.3138 ± 0.0019
	400	1.6062 ± 0.0030	-3.9305 ± 0.0148	0.3229 ± 0.0017
	760	1.7156 ± 0.0024	-4.7240 ± 0.0121	0.3478 ± 0.0014
0.0050	300	1.7018 ± 0.0025	-4.6233 ± 0.0124	0.3439 ± 0.0014
	400	1.7536 ± 0.0022	-5.000 ± 0.0111	0.3570 ± 0.0012
	760	1.8939 ± 0.0026	-5.9905 ± 0.0128	0.3852 ± 0.0014

Figure 6.5 shows the variation of PGA with distance for five earthquakes (2014 La Habra and Westwood, 2019 Searles Valley and Ridgecrest, 2020 El Monte) along with the predictions of the Boore et al. (2014) GMM (median ± two standard deviations). CSN data shown on the plot are BBR motions only. Recordings from all other networks for these events from the GMDB are plotted in the background with gray symbols. The three threshold PGA levels defined earlier in this report are shown with horizontal lines. Figure 6.5 shows that the data generally lie above the thresholds. The GMM mean -2 standard deviations curve intersects the thresholds at a distance beyond the maximum distance of BBR stations for the lowest threshold but would cause the data to be screened out for the upper two thresholds. This suggests that the 0.005 and 0.003g threshold may be too restrictive and that the 0.0015g threshold may be preferred.

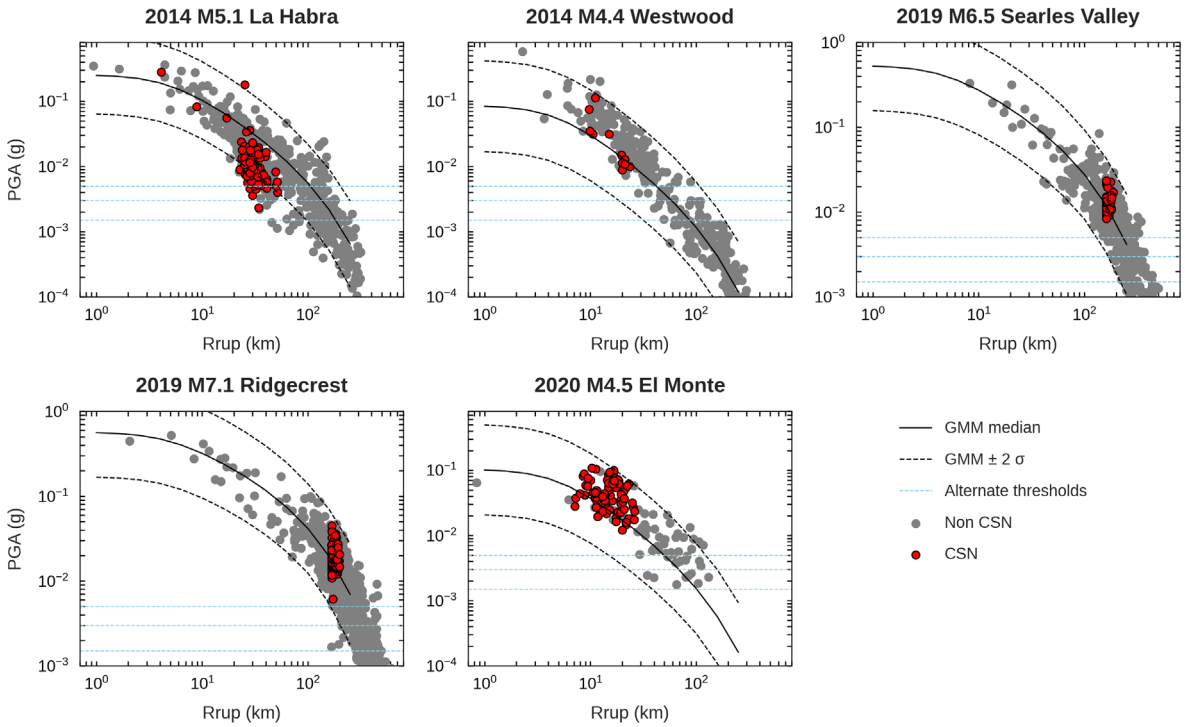


Figure 6.5. PGA versus distance data points for five earthquakes from Table 2.1 with substantial numbers of BBR CSN motions, overlain on data points for motions from other sensor networks and attenuation plots from the Boore et al. (2014) GMM using V_{S30} = 400 m/s (network median site condition). Horizontal lines indicate alternate PGA thresholds as defined in this report.

7 Summary and Conclusions

This study has undertaken a series of tasks that collectively aim to provide insight into the performance of CSN ground-level sensors during southern California earthquakes, provide processed data in an accessible form for users, and provide recommendations on the range of conditions for which the data can be used with confidence in ground motion modeling projects.

CSN data from 29 earthquakes has been uniformly processed using NGA-type procedures. For events where data from other networks was already available, the CSN data has been added to a national database for ground motion research applications (Buckreis et al. 2023a) to supplement the previously available data. For events not previously in the database, CSN and non-CSN data has been processed and added to the database. Relevant site and event metadata has been compiled and added so that this data is available for public use.

Among the events considered, approximately 50% of the recordings were judged to be not usable because they are noise-dominated based on visual inspection or have unusual features. These are referred to as REJ records in this report. However, this rate is potentially misleading as an indicator of network performance, because 27 of the 29 events are small magnitude (< 5.5) and often occurred at considerable distances from the network. Two large events (2019 Searles Valley and 2019 Ridgecrest) were successfully recorded by over 95% of sensor horizontal components, despite being located at distances > 150 km. This rate of data recovery is considered more representative of the performance that can be expected in future impactful earthquakes in the greater Los Angeles area.

Among the remaining (non-REJ) recordings, we distinguished records with relatively broad bandwidth (usable Fourier frequency range of at least 0.5 - 10 Hz) (denoted BBR) from those with relatively limited bandwidth (narrower than that for BBR at one or both ends of the frequency range; denoted NBR). Comparisons of BBR and NBR signals with signals from non-CSN proximate sensors (separation distance < 3 km and same geology) shows that PGA levels are not statistically distinguishable. Spectral accelerations from BBR CSN data appear to be unbiased over the oscillator period range of 0.01 to 5 sec based on these comparisons with the exception of lower CSN motions near 1.0 sec, whereas NBR CSN data have lower spectral accelerations for multiple period intervals < 5 sec (amount of the bias is generally < 10-15%). This is not surprising given the limited bandwidth of NBR signals.

Comparisons of BBR and REJ data indicate that 0.005g is a reasonable average threshold acceleration across the network, whereas comparisons of pre-event noise signals with usable signals (BBR and NBR) indicates a threshold of 0.0013g. These thresholds represent averages over broad areas and multiple times; we could anticipate that specific locations and times could produce

higher or lower thresholds. Using these threshold accelerations with a calibrated GMM for southern California, limiting distances have been provided as a function of magnitude and site condition. For locations beyond these limiting distances, the problem is not that no recordings are available, but that the overall population of recorded ground motions are biased towards larger values. The limiting distances derived for CSN are notably lower than those for more sensitive instruments as used in prior research, but consistent with those for analog or low-resolution digital sensors.

These results show that CSN data is useful for research and engineering applications, but its range of applicability is more limited than data from more sensitive instruments. Within its application range, the CSN data have advantageous features, including relatively small between-sensor spacings that facilitate site response or ground motion variability studies at short length scales.

REFERENCES

Aagaard B, Catchings R, Graymer R, Macheling P, Rodgers A, Taira T, Thurber C, 2021. San Francisco Bay Region community seismic velocity model development, https://baagaard-usgs.github.io/sfcvm-website/. Last Accessed 5/2021.

Afshari K, Stewart JP, 2016. Physically parameterized prediction equations for significant duration in active crustal regions, *Earthquake Spectra*, 32(4), 2057-2081.

Boore DM, 2005. On pads and filters: processing strong-motion data, *Bull. Seism. Soc. Am.*, 95, 745-750.

Boore DM, 2010. Orientation-independent, non geometric-mean measures of seismic intensity from two horizontal components of motion, *Bull. Seismol. Soc. Am.* 100, 1830–1835.

Boore DM and Bommer JJ, 2005. Processing of strong-motion accelerograms: Needs, options and consequences, *Soil Dynamics & Earthquake Engineering*, 25, 93-115

Bozorgnia Y, Abrahamson NA, Al Atik L, Ancheta TD, Atkinson GM, Baker JW, Baltay A, Boore DM, Campbell KW, Chiou BS-J, Darragh RB, Day S, Donahue J, Graves RW, Gregor N, Hanks T, Idriss IM, Kamai R, Kishida T, Kottke A, Mahin SA, Rezaeian S, Rowshandel B, Seyhan E, Shahi S, Shantz T, Silva WJ, Spudich P, Stewart JP, Watson-Lamprey J, Wooddell K, Youngs RR, 2014. NGA-West2 research project, *Earthquake Spectra*, 30, 973-987.

Bozorgnia Y, NA Abrahamson, SK Ahdi, TD Ancheta, L Al Atik, RJ Archuleta, GM Atkinson, DM Boore, KW Campbell, BSJ Chiou, V Contreras, RB Darragh, S Derakhshan, JL Donahue, N Gregor, Z Gulerce, IM Idriss, C Ji, T Kishida, AR Kottke, N Kuehn, DY Kwak, AOL Kwok, P Lin, J Macedo, S Mazzoni, S Midorikawa, S Muin, GA Parker, S Rezaeian, H Si, WJ Silva, JP Stewart, M Walling, K Wooddell, RR Youngs, 2022. NGA-Subduction research program, *Earthquake Spectra*, **38**(2), 783-798.

Brune, JN, 1970. Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophysical Res.*, 75, 4997-5009.

Buckreis, TE, Nweke CC, Wang P, Brandenberg SJ, Mazzoni S, Stewart JP, 2023a. Relational database for California strong ground motions, Geo-Congress 2023: Geotechnical Data Analysis and Computation, Los Angeles, CA, March 2023, *Geotechnical Special Publication No. 342*, EM Rathje, B Montoya, and MH Wayne (Eds.), 461-470, ASCE Geo-Institute

Buckreis, TE, JP Stewart, SJ Brandenberg, P Wang (2023b). Subregional anelastic attenuation model for California, *Bull. Seismol. Soc. Am.*, 113(6), 2292-2310.

Clayton R, Heaton T, Chandy M, Krause A, Kohler M, Bunn J, Guy R, Olson M, Faulkner M, Cheng MH, Strand L, Chandy R, Obenshain D, Liu A, and Aivazis M, 2011. Community Seismic Network, *Annals of Geophysics*, 54 (6), doi: 10.4401/ag-5269.

Clayton R, Kohler M, Guy R, Bunn J, Heaton T, and Chandy M, 2020. CSN/LAUSD network: A dense accelerometer network in Los Angeles schools, *Seis. Res. Lett*., 91(2A), 622-630, doi:10.1785/0220190200.

Contreras V, Stewart JP, Kishida T, Darragh RB, Chiou BSJ, Mazzoni S, Youngs RR, Kuehn NM, Ahdi SK, Wooddell K, Boroschek R, Rojas F, Ordenes J, 2022. NGA-Sub source and path database, *Earthquake Spectra*, 38(2), 799-840.

Consortium of Organizations for Strong-Motion Observation Systems (COSMOS), 2001 COSMOS strong motion data format. https://www.strongmotioncenter.org/vdc/cosmos_format_1_20.pdf (last accessed 9/15/2023).

Depsky NJ, Cushing L, Morello-Frosch R, 2022. High-resolution gridded estimates of population sociodemographics from the 2020 census in California, *Plos One*, https://doi.org/10.1371/journal.pone.0270746

Díaz J, M Ruiz, PS Sánchez-Pastor, et al., 2017. Urban seismology: on the origin of earth vibrations within a city. *Nature Scientific Reports* 7, 15296. https://doi.org/10.1038/s41598-017-15499-y

Douglas J and Boore DM, 2011. High-frequency filtering of strong-motion records, *Bull. Eqk. Eng.*, 9, 395–409.

Ekström G, Nettles M, Dziewonski AM, 2012. The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics the Earth and Planetary Interiors* 200–201: 1–9.

Falcone JA, 2016, U.S. block-level population density rasters for 1990, 2000, and 2010: U.S. Geological Survey data release, http://dx.doi.org/10.5066/F74J0C6M.

Goulet, CA, Kishida T, Ancheta TD, Cramer CH, Darragh RB, Silva WJ, Hashash YMA, Harmon J, Parker GA, Stewart JP, Youngs RR, 2021. PEER NGA-East database, *Earthquake Spectra*, 37(S1), 1331-1353

Hearne M, Thompson EM, Schovanec H, Rekoske J, Aagaard BT and Worden CB, 2019. USGS automated ground-motion processing software, USGS Software Release, doi:10.5066/P9ANQXN3.

Kishida T, Darragh RB, Chiou BSJ, Bozorgnia Y, Mazzoni S, Contreras V, Boroschek R, Rojas F, Stewart JP, 2020. *Chapter 3: Ground motions and intensity measures (Data resources for NGA-subduction project)* (ed JP Stewart). PEER report 2020/02. Berkeley, CA: Pacific Earthquake Engineering Research Center, UC Berkeley.

Kwak DY, Ahdi SK, Wang P, Zimmaro P, Brandenberg SJ, Stewart JP, 2021. Web portal for shear wave velocity and HVSR databases in support of site response research and applications. UCLA Geotechnical Engineering Group. DOI:10.21222/C27H0V

Kramer SL, 1996. *Geotechnical Earthquake Engineering*, Prentice Hall, Upper Saddle River, NJ.

Lecocq T, and 66 other authors, 2020. Global quieting of high-frequency seismic noise due to COVID-19 pandemic lockdown measures. *Science* 369,1338-1343.

Nweke, CC, Stewart JP, Wang P, Brandenberg SJ, 2022. Site response of sedimentary basins and other geomorphic provinces in southern California, *Earthquake Spectra*, 38(4), 2341-2370.

Ramos-Sepulveda ME, Parker GA, Thompson EM, Brandenberg SJ, Li M, Ilhan O, Hashash YMA, Rathje EM and Stewart JP, 2023. High-pass corner frequency selection for implementation in the USGS automated ground-motion processing tool, Geo-Congress 2023: Geotechnical Data Analysis and Computation, Los Angeles, CA, March 2023, Geotechnical Special Publication No. 342, EM Rathje, B Montoya, and MH Wayne (Eds.), 327-335, ASCE Geo-Institute

Wang P, Stewart JP, Bozorgnia Y, Boore DM, Kishida T, 2017. 'R' Package for computation of earthquake ground-motion response spectra, *Report No. 2017/09*, Pacific Earthquake Engineering Research Center, UC Berkeley.

Wills CJ, Gutierrez CI, Perez FG, Branum DM, 2015. A next generation VS30 map for California based on geology and topography, *Bulletin of the Seismological Society of America* 105(6), 3083–3091.

Wilson D, J Leon, R Aster, J Ni, J Schlue, S Grand, S Semken, S Baldridge, W Gao, 2002. Broadband seismic background noise at temporary seismic stations observed on a regional scale in the southwestern United States. *Bulletin of the Seismological Society of America* 92 (8), 3335–3342.