DISCLAIMER

Final Report DOE Award DE-SC0016527 (170455-00001)

Collaborative Research: Properties and Dynamics of the Shallow Crust

Debi Kilb[1] & Frank Vernon[1]

1. Scripps Institution of Oceanography, University of California,

San Diego, La Jolla, California, 92093-0225

USA

Keywords: Ground motion, PGV, PGA, southern California, small aperture arrays, data quality, ternary plots, Sage Brush Flat linear array, Pinyon Flat 2D array, Blackburn linear array, Jackass Flat linear array, Transect linear array, Ramona linear array.

Appendix A: Mapped spatial distributions of the seven small aperture arrays.

Appendix B: Ternary plots showing relative scaling of the three ground motion components (north-south, east-west, and vertical).

ABSTRACT

Ground motions recorded at one location are often extrapolated to nearby regions within a given radius. Here, we explore the appropriateness of spatial extrapolation using data from seven small aperture seismic network deployments in southern California. Six of these deployments are linear arrays of 4-13 stations, and one is a 2D array of 13 stations at Pinyon Flats Observatory. The spatial footprint array diameters are 3 km or less, and each array was operational for a year or more. From our base catalog (M2.5+ earthquakes; 4038 events; September 2010 - June 2023), automated methods remove temporally overprinted waveforms from nearby events (< 5 km) in quick succession (< 5 min) and data with non-viable waveforms. These 200 samples per second data are filtered at 0.5-25 Hz and must have signal-to-noise ratios (SNRs) of 2.5+. Peak ground acceleration (PGA) and peak ground velocity (PGV) are derived individually from the maximum absolute values of each of the 3-component waveforms (vertical, north-south, and east-west). Five of the seven arrays traverse the San Jacinto fault, and two do not. Ground motion observations are compared with theoretical estimates from Abrahamson et al., 2014. On average, arrays deployed within and across fault zones consistently record ground motions above theoretical expectations, whereas off-fault arrays record ground motions at or slightly below theoretical expectations. We attribute these differences to site conditions because these trends prevail for the full data suites. For each network and each individual channel, the coefficient of variation indicates that the standard deviations are ~30±6% of the mean. Exploring relative ground motion contributions from all three channels (ternary plots), as expected, most data show that vertical ground motions are attenuated compared to horizontal ones. However, this is not always the case for RA array data, where vertical motions can be ~2-3 times larger than horizontal motions for select events near Cahuilla, CA. These anomalously high vertical motions are focal mechanism-related. These results suggest that ternary plots created using only a small amount of data can be used as a data quality metric and a tool to find anomalous features in three-component data.

PLAIN LANGUAGE SUMMARY

Monitoring ground motion is a well-known way to track how the earth behaves to various stimuli such as earthquakes and drought. Here, we use data from seven small aperture array networks (4-13 seismic stations within each array) deployed in dense configurations (spatial footprint diameters < 3 km) in southern California. Five of the seven arrays span the active San Jacinto fault zones, and two are off-fault. All seven network arrays were operational for at least one year. For each array, we create sub-catalogs of all M2.5+ earthquakes (September 2010 - June 2023) within 150 km of each array network centroid. For quality control, we require all seismic waveform data to have a signal-to-noise ratio of 2.5 or above. On average, we find a 30% variability in the peak ground acceleration (PGA) and peak ground velocity (PGV) measurements compared with the mean, which is on par with other small aperture array research findings. We evaluate how the ground motions align with theoretical estimates. We find peak ground motions recorded within and near fault zones tend to be higher than theoretically predicted, and off-fault ground motions are either as expected or slightly lower. As expected, differences within the three component ground motions (north-south, east-west, and vertical) show lower vertical values than the horizontal counterparts, except for select data recorded by the RA array. Vertical motions are larger than horizontal for select earthquakes near Cahuilla, CA, and are dictated by a focal mechanism effect. These results suggest that ternary plots can be used as a data quality metric, revealing key differences in peak ground motion distributions from a suite of earthquakes.

Key Findings

1. Consistent with other studies, for each channel in each network, deviations from the mean PGV and PGA are ~30±6%.
2. Compared to theoretically estimated PGV and PGA values, ground motions within fault zones are accentuated, whereas off-fault motions are more consistent with the theoretical values.
3. For select events within a narrow zone near Cahuilla, CA, data recorded by the RA array show elevated PGA and PGV vertical motions compared with the corresponding horizontal motions.

ABSTRACT

Ground motions recorded at one location are often extrapolated to nearby regions within a given radius. Here, we explore the appropriateness of spatial extrapolation using data from seven small aperture seismic network deployments in southern California. Six of these deployments are linear arrays of 4-13 stations, and one is a 2D array of 13 stations at Pinyon Flats Observatory. The spatial footprint array diameters are 3 km or less, and each array was operational for a year or more. From our base catalog (M2.5+ earthquakes; 4038 events; September 2010 - June 2023), automated methods remove temporally overprinted waveforms from nearby events (< 5 km) in quick succession (< 5 min) and data with non-viable waveforms. These 200 samples per second data are filtered at 0.5-25 Hz and must have signal-to-noise ratios (SNRs) of 2.5+. Peak ground acceleration (PGA) and peak ground velocity (PGV) are derived individually from the maximum absolute values of each of the 3-component waveforms (vertical, north-south, and east-west). Five of the seven arrays traverse the San Jacinto fault, and two do not. Ground motion observations are compared with theoretical estimates from Abrahamson et al., 2014. On average, arrays deployed within and across fault zones consistently record ground motions above theoretical expectations, whereas off-fault arrays record ground motions at or slightly below theoretical expectations. We attribute these differences to site conditions because these trends prevail for the full data suites. For each network and each individual channel, the coefficient of variation indicates that the standard deviations are ~30±6% of the mean. Exploring relative ground motion contributions from all three channels (ternary plots), as expected, most data show that vertical ground motions are attenuated compared to horizontal ones. However, this is not always the case for RA array data, where vertical motions can be ~2-3 times larger than horizontal motions for select events near Cahuilla, CA. These anomalously high vertical motions are focal mechanism related. These results suggest that ternary plots created using only a small amount of data can be used as a data quality metric and a tool to find anomalous features in three-component data.

PLAIN LANGUAGE SUMMARY

Monitoring ground motion is a well-known way to track how the earth behaves to various stimuli such as earthquakes and drought. Here, we use data from seven small aperture array networks (4-13 seismic stations within each array) deployed in dense configurations (spatial footprint diameters < 3 km) in southern California. Five of the seven arrays span the active San Jacinto fault zones, and two are off-fault. All seven network arrays were operational for at least one year. For each array, we create sub-catalogs of all M2.5+ earthquakes (September 2010 - June 2023) within 150 km of each array network centroid. For quality control, we require all seismic waveform data to have a signal-to-noise ratio of 2.5 or above. On average, we find a 30% variability in the peak ground acceleration (PGA) and peak ground velocity (PGV) measurements compared with the mean, which is on par with other small aperture array research findings. We evaluate how the ground motions align with theoretical estimates. We find peak ground motions recorded within and near fault zones tend to be higher than theoretically predicted, and off-fault ground motions are either as expected or slightly lower. As expected, differences within the three component ground motions (north-south, east-west, and vertical) show lower vertical values than the horizontal counterparts, except for select data recorded by the RA array. Vertical motions are larger than horizontal for select earthquakes near Cahuilla, CA, and are dictated by a focal mechanism effect. These results suggest that ternary plots can be used as a data quality metric, revealing key differences in peak ground motion distributions from a suite of earthquakes.

Key Findings

1. Consistent with other studies, for each channel in each network, deviations from the mean PGV and PGA are ~30±6%.
2. Compared to theoretically estimated PGV and PGA values, ground motions within fault zones are accentuated, whereas off-fault motions are more consistent with the theoretical values.
3. For select events within a narrow zone near Cahuilla, CA, data recorded by the RA array show elevated PGA and PGV vertical motions compared with the corresponding horizontal motions.

INTRODUCTION

Ground motion data are essential for a range of geoscience topics, including but not limited to earthquake early warning (Allen & Stogaitis, 2022), seismic hazard estimates (Lervolino, 2022), and identification of earthquake source directivity features (Kurzon et al., 2014; Hirakawa et al., 2023). In regions of sparse station coverage, researchers typically rely on extrapolation techniques, assuming that the ground motions within a given radius of a station can be estimated from ground motions recorded at the station. There are various ways to extrapolate measurements from one location to nearby regions. One option is to assume that the ground motions will be similar when extended to a given radius from the station or alternatively apply a damping factor that reduces the amplitude by a factor that scales with distance (Tamhidi et al., 2023). These simplified extrapolations can be misguided in regions of variable geological features, complex fault geometry, or pore pressure variability (Graves et al., 2008).

Ground motions within 3D sedimentary basin structures can generate surface waves at the basin edge, creating reverberations and significantly increasing ground motions (Vidale & Helmberger, 1988; Olsen *et al.,*1995a,b). Theoretical simulations of ground motions within the Los Angeles basin found that the scale factors for ground motion amplification depend on basin depth. For regions near Los Angeles, California, where the basin depth extends to 9km, an amplification factor as large as six can be produced (Olsen, 2000). These types of basin resonance are well known, and current work is exploring how amplification, or alternative attenuation, occurs within and around fault zones, which is the aim of this work.

How ground motions are elevated or diminished can be categorized by three terms: source, path, and site terms (Douglas & Edwards, 2016). An example of a source term effect is when an earthquake source has a strong directivity, producing more seismic energy in one direction than another (Kilb et al., 2002; Kurzon et al., 2014; Hirakawa et al., 2023). Path terms, on the other hand, are generated along the path

from the earthquake source to the station. If the seismic wave path traverses only hard rock materials, the path term is typically minimal and does not substantially increase or decrease ground motions. In contrast, the ground motion typically increases if the path traverses primarily unconsolidated material or variable material strengths (Sahakian et al., 2018). Influences near the recording station create site terms. Examples of site terms are small basin resonances, traffic from a nearby roadway, or wind-generated signals created by the movement of trees in response to high winds (Johnson et al., 2019).

For research focusing on a generalized examination of ground motions (i.e., no interest in fine-scale ground motion variability), a scale factor called a 'site term' can be applied to unify the findings. Site terms are specific to the region near the recording station. Site terms often rely on a factor called Vs30, which is the time-averaged shear-wave velocity within the top 30 meters of material near the site. Low values of Vs30 suggest the site lacks a large sediment layer, and the site is assumed to have low amplification factors. In comparison, high Vs30 values indicate thick sediment layers and, in turn, high amplification factors.

When there is a lack of additional information, topographic slopes can be used as a proxy for Vs30 values (Wills et al., 2015; Yong, 2016). This pairing broadly assumes that minimal topography is more common near thick basin sediments, and steep topography is more common near hard-rock regions with minimal sediment coverage (Heath et al., 2020). There are concerns that assuming a linear relationship between Vs30 and topography can be misleading, and instead, additional information should be incorporated when estimating site terms (Li et al., 2022). This concern is justified, especially for engineering purposes, because these simplified site terms do not account for near-site soil conditions that can experience non-linear and permanent deformations (Trifunac, 2016).

Increasing the number of seismic stations in a region reduces the inter-station spacing and, in turn, allows for an increased understanding of ground motion variability. EarthScope's USArray deployment aimed to

have station spacings of 70 km (IRIS Transportable Array, 2003), and the ShakeAlert earthquake early warning system deployed within western USA (California, Washington, and Oregon) strives to have ~10 km inter-station spacing (Given et al., 2014). Small aperture arrays, a network of stations deployed 10's of meters apart, can explore nuances in ground motion variability within even smaller spatial footprints with radial distances of a few kilometers (e.g., Vernon et al., 1991; Ben Zion et al., 2015; Share et al., 2017; Qiu et al., 2017; Share et al., 2019; Zigone et al, 2019; Anthony et al., 2020; Qiu et al., 2021).

Using a large amount of data to explore geoscience concepts is a growing field, of which small aperture seismic arrays are a growing subset of the new projects (Ben-Zion et al., 2015; Li et al., 2018; Mordret et al., 2019; Qi et al., 2021; Arrowsmith et al., 2022). A study of small-aperture arrays in Iceland found that in regions of large geologic and topographic variability, ~57% of ground motion variability can be attributed to site terms alone, and even in regions where the geologic structure is similar and topography variations minimal that site terms can still produce ~13% of ground motion variations (Rahpeyma et al., 2019). A study from the Sage Brush Flats region in southern California compared surface ground motions with ground motions from a 148-meter deep borehole and found that surface recordings can be 3-10 times larger than borehole recordings (Johnson et al., 2020).

Here, we explore data from seven small aperture arrays with 3 km or less spatial footprint diameters. The aim is to understand better when ground motions recorded at one location can be reliably extrapolated to nearby regions and when they cannot. This work will allow us to assess how variable ground motions can be within a well-defined small area sufficiently instrumented within and outside fault zones. We find a dichotomy in the ground motion observations, such that ground motions within the fault zones are larger than expected, and off-fault locations record ground motions are more consistent with the theoretical estimates.

SEISMIC NETWORKS AND DATA SELECTION

The seven small aperture network arrays were selected because they have 3-component data and were active for at least one year (Figures 1-3; Figures A1-A7; Table 1). Two of these seven networks (PY and TR) are located away from the primary fault system, and the other five span the San Jacinto fault (BB, DW, JF, RA, and SG). All arrays are linear except for the Pinon array, which is two-dimensional (Appendix A; Figures A1-A7).

We start with a base catalog of reviewed earthquakes spanning a 13.5-year period (2010-09-20 through 2023-06-30) (Figure 1; ComCat catalog; see data and resources), including 4038 M2.5+ events. From this base catalog, we create sub-catalogs limited to events within 150 km of the network centroid for each of the seven networks. Because of the spatial differences in the network centroids, these sub-catalogs differ from network to network (Table 1; Figure 1), but this is not a concern because the catalogs are vast (each contains over ~5000 waveforms that have passed the data quality restrictions) and can adequately provide general ground motion trends.

DATA PREPROCESSING

Seismic data pre-processing includes four steps. First, to eliminate waveform overprinting, which could result in misleading elevated ground motion measurements, we remove all earthquake pairs in the base catalog that are spatially (<5 km) and temporally (<5 minutes) close. We purposely process data in the base catalog, not the individual catalogs, to avoid introducing errors at the edges of the sub-catalog spatial boundaries. Using the entire base catalog is also time efficient as this processing is applied only once to the base catalog (i.e., not individually for each of the seven sub-catalogs). The overprinting removal process iteratively repeats until no additional data are flagged for removal, indicating all remaining data are at least 5 km and 5 minutes apart.

We access seismic waveform data through the IrisFetch Matlab interface (see Data and Resources). In some cases, the returned data contains multiple data segments. When this occurs, we remove the mean

from each sub-trace before combining the sub-traces into a final trace. This processing is essential, especially for the older data, because the DC offset can differ from sub-trace to sub-trace.

In the second data pre-processing step, we select robust waveforms requiring a signal-to-noise ratio (SNR) of 2.5 or above (Figure 4). To measure SNR, we rely on theoretical P-wave (Vp; 6.5 km/s) and S-wave (Vs; 2 km/s) speeds and estimate theoretical P-wave and S-wave arrival times at each station using a linear relationship:

$$Parrival = Disthypo/Vp \quad (1a)$$

$$Sarrival = Disthypo/Vs \quad (1b)$$

Parrival and Sarrrival are the theoretical P-wave and S-wave arrival times, respectively. Disthypo is the hypocentral distance in kilometers between the earthquake and the station, calculated using the GRS80 reference ellipsoid. The simplicity of these estimated seismic wave arrival times might be in error because they do not account for the earth's curvature or velocity depth differences. However, these estimates are appropriate for our purposes as the aim is simply to select the data of interest, a noise window, and to estimate a signal-to-noise ratio (SNR), so exact values are not required.

A ten-second noise window is selected from data prior to Parrival. This window, however, is shifted, so the window ends two seconds before the theoretical Parrival. This two-second shift is precautionary and is applied because the estimated Parrival times are not exact, and this shift will allow for errors in the estimates. The data window encompasses data from two seconds before Parrival through 30 seconds after Sarrival. In this way, the duration of each data window is variable, resulting in longer durations for more distant events when the P- and S-wave separation times are larger (Figure 5). We calculate SNRs by dividing the maximum absolute amplitude of the data window by the maximum absolute amplitude of the noise window.

The SNR criterion removes most of the bad data, but it is possible that additional non-viable signals not related to earthquakes may still be retained. Given this, in our third data quality test, we qualitatively evaluate signals with large amplitudes. Instead of using a raw count amplitude threshold to separate good and bad signals, we take a broader approach and qualitatively examine data with the five highest absolute raw count amplitudes for the acceleration (HN* channels) and velocity (HH* channels) data. These qualitative checks examine both time-domain and spectral-domain information (Figure 6).

The final data pre-processing step is applying an instrument response correction, removing the mean from the data traces, and band-pass filtering. For these 200 samples per second data, we apply 4th-order bandpass Butterworth filters of 0.05 - 25 Hz. The low-pass level is set to avoid low-frequency artifacts, and the high-pass is set to avoid spurious high-frequency signals (Figure 7).

METHODS

For each data trace (each channel of each station in each network), peak ground velocity (PGV) and peak ground acceleration (PGA) are computed. Acceleration (velocity) data are converted to velocity (acceleration) using integration (differentiation). All findings and associated metadata (i.e., SNR, epicentral and hypocentral distance, station, channel, filtering, etc.) are stored in a data structure for later use.

We assume the GMPE of Abrahamson et al. (2014) is the theoretical ground truth and compare our results with these estimates. The Abrahamson et al. (2014) GMPE was derived from horizontal motions of M3+ events, so it is possible this GMPE might not correctly represent vertical motion findings and results from smaller events (<M3). However, because we find good agreement between our results and the GMPE, this suggests that the Abrahamson et al. (2014) estimates are appropriate for our use.

We identify outliers within these data using the log of the ratio of the theoretically predicted ground motions to observed values. Ratios below (above) zero indicate the observed ground motions were greater than (less than) expected. We assess these ratios using Matlab's *isoutlier* routine, using the 'mean' method that flags elements that are three or more standard deviations from the mean. This method is appropriate for these data as these data distributions are symmetric, not skewed.

Many studies do not look at ground motions from each of the three individual components (north-south, east-west, and vertical) and instead opt to use an average of all three components (Kilb et al., 2020) or maintain only the vertical information and an average of the two horizontal components (Beyer & Bommer, 2006). This type of averaging can result in the loss of key information about how the ground motions of one component differ from the other two. Here, we maintain and assess data from all three components using ternary plots, visual depictions of data triads often used to assess focal mechanism variability, i.e., differences in strike, dip, and rake (Frohlich, 1992). In this way, we can assess ground motion variations within and between all three components of data (i.e., north-south, east-west, and vertical).

We present synthetic examples to understand how to interpret ternary plots. The simplest case to understand is when all three components have similar peak ground motions, which nets a cluster of the event markers at the centroid of the triangle (Figure 8a). If the vertical component is reduced compared with the horizontals, the data will cluster toward the bottom of the triangle (Figure 8b). And, if one component has larger amplitudes than the other two components, ternary plots will present a cluster of points within the triangle corners. For large Z, NS, and EW cases, the cluster of points will locate at the triangle's top, right, and left sides, respectively (Figure 8c-d). Data that is not clustered and has spurious outliers could indicate that the measurements are in error.

For each ternary plot, a circular guide is placed at the center of the triangle and can be used as a reference. This reference circle can be compared with the data circle centered at the median of each station's data suite. Generally, if the data circle is below the reference circle (Figure 8b and d), the vertical ground motions are diminished compared to the horizontal components. If the data circle is located to the right of the reference circle, the east-west component is elevated compared to the other two components (Figure 8d). For crustal earthquakes, the vertical component is typically ⅔ of the horizontal component, so having the thin line circle fall below the thick line reference circle is the correct behavior.

RESULTS

We remove events within 5 km and 5 mins from the base catalog. This data vetting is relatively harsh as it removes 16% of the base catalog (649 of 4038; 3389 retained) and all six M5+ events in our base catalog, resulting in a maximum event magnitude of 4.6 (Table 1). This requirement, however, is essential to ensure our derived ground motion values are measuring the induced ground motions for a single earthquake, not multiple earthquakes. Overall, this process primarily flags events during the Salton Sea swarms in 2012, 2016, 2020, and 2021 (Kilb et al., 2021). Because our catalog includes only M2.5+ events, some smaller event overprinting will likely remain in these data.

After applying the SNR restriction (SNR 2.5+), the amount of data retained is high, ranging from 92.7% at SG to 99% or above at BB, DW, and PY. These high percentages suggest that we are using quality data. The SNR histograms from the different arrays, however, show variable distributions (Figure 9). The PY SNR histogram distribution differs from the others, showing a relatively similar number of values for each SNR data bin. This is likely because PY is an off-fault array and is not subject to fault-zone-related noise sources. The SNR distributions for arrays TR, JF, BB, and DW arrays are similar in that the number of values in the 0-2.5 SNR bin (data deemed too noisy to be useful) is at least twice as large as the numbers for the 2.5-5 bin. Our interpretation of these findings is that our SNR threshold of 2.5 is working well to identify bad data. The RA and SG distributions show a general tapering trend, with

substantially more small SNR measurements than larger
ones.

We also investigate the SNR distributions for each individual station in each network. When looking at the SNR histogram data distributions of the seven arrays, four (BB, DW, JF, and PY) show similar SNR ratio distributions at all stations in the array, as expected for dense station deployments. Three arrays (RA, TR, and SG), however, show large differences in the SNR ratio distributions. The SNR distribution differences for the SG stations are the most pronounced, showing a prevalence of high noise floors at stations SGBFA, SGBN1, and SGBS3 (Figure 10). These high noise floors resulted from instrumental problems. These issues arose for stations SGBFA and SGBN1 in late 2011 and early 2012 and were subsequently corrected. The SGBS3 issues began in ~2021 and are less problematic than the other two stations. For the RA and TR networks, stations RA09 and TR01 have a much higher noise floor than the other RA and TR station reports (Figure 11-12). These low SNR values at RA09 stem from a relatively short-lived instrumental issue at the end of 2015 extending through a portion of 2016, which was identified and correct. The low SNR values at TR01 were also caused by instrumental issues, primarily during portions of 2015 and portions of 2017. In all cases, the SNR threshold tests successfully flagged these bad data and removed them from the catalog.

Comparing our results with the theoretical estimates described in the methods section, we find systematically that arrays deployed within the fault zone recorded larger PGV and PGA than expected. Alternatively, arrays not deployed within the fault zone show PGV and PGA values as expected or slightly less than expected (Figure 13).

Our outlier method successfully identifies ground motions that are either larger or smaller than theoretically expected. We find no dependency between outliers and depth regardless of the network. More outliers were flagged during the Salton Sea swarms in 2012, 2016, 2020, and 2021, showing larger

ground motions than expected. These elevated ground motions are not a surprise, as these swarms contained many events below M2.5, which were not accounted for in our overprinted removal processing. The outlier identification method identifies all but one of the 28 non-viable signals. The outlier method did not flag a non-viable signal during the magnitude 3.7 earthquake on 30 May 2015, which included substantial instrumental noise at station TR01.

Our next goal is determining how much peak ground motions differ across a single array. Using our refined data, we compute the mean and standard deviations of the PGV and PGA values individually for each earthquake and each channel. From these values, we compute the coefficient of variance, which is the standard deviation divided by the mean (Figure 14). We find consistent results across all arrays, showing coefficients of variances of 26% to 36%. Generally, the standard deviations are typically ~30±6% of the mean.

We create ternary plots using the refined catalogs (overprinting, SNR, and non-viable). We want the cleanest data possible, so in processing these data, we increase the SNR threshold to seven to ensure we are examining quality data. We expect the ternary results to exhibit similar behavior collectively. We also expect the median circle placement to be lower in the triangle than the reference circle, a difference dictated by the fact that for crustal earthquakes, vertical PGA and PGV values are expected to be ⅔ of their horizontal counterparts. We find these expected behaviors for networks BB, DW, JF, PY, TR, and SG (Figures in Appendix B; Figures B1, B2, B3, B4, B6, and B7; except TR01). But this is not the case for the RA stations (Figure B5 in Appendix B). Data from stations in the RA array tend to exhibit more vertical than horizontal motion (i.e., data clusters extend toward the top of the triangles in the ternary plots). This systematic behavior observed at all stations indicates that this region's elevated vertical motions are a true feature, not caused by recording malfunctions.

Waveforms from data exhibiting the most vertical exaggeration in the ternary plots for RA stations confirm unusually large P-wave motion dictating the behavior. Events elevating the vertical motions the most are primarily north and south of the RA array. The largest vertical amplifications are strictly from events in the Cahuilla region, suggesting a dependence on earthquake focal mechanisms.

We track the largest ground motions recorded by each station (Table 4), listing PGV values for velocity sensors and PGA values for accelerometers. Recall that the overprinted waveform removal process removed all M5+ earthquakes, and an M4.6 earthquake is the largest event in the base catalog (Table 1). The expectation is that because of the small spatial footprints of these networks, stations within a given array would likely report the same earthquake as producing the largest ground motions. This is not what we found.

For the TR array, all four stations reported different earthquakes as generating the highest PGV. The spatial extent of the TR array is the longest of the arrays (3km compared with less than 0.2 km for some of the other arrays; see Table 2), so a wide span of results might be expected. The SG and BB arrays have the smallest footprints, with linear spans of 0.2 km or less, but even for these small footprint networks, different events are flagged as producing the largest ground motions. For the BB array, three different events are listed as generating the maximum PGV at the seven BB stations. Similarly, for the SG array, three different events are listed as generating the maximum PGA at the seven SG stations. Similar differences are found in data for the other arrays (see Table 4). Results from PY show more consistency, showing that 11 of the 13 stations reported that the same M4.3 2023-04-01 earthquake produced the largest PGV at this array.

DISCUSSION

For each network, differences in the log of the observed to theoretical ground motion ratios are not random or skewed, primarily showing a normal distribution of values (see Figure 13). These normal

distributions suggest that the site term is the dominant property dictating ground motions, not the source or path terms. This is expected for two reasons. First, we do not expect source terms to dominate because the catalog contains primarily small-magnitude events, and in turn, most events can be considered point sources. The second reason is the small footprints of these arrays that, in theory, should be measuring very similar ground motions (i.e., Meng and Ben-Zion, 2018; Johnson et al., 2020).

As listed in Table 4, for each array, we do not find that a single designated earthquake produces the largest ground motion at each station across the array. Instead, the earthquake responsible for the largest ground motion produced at each individual station varies. This means that even for these small footprint arrays, a single earthquake can produce ground motions across the array that are highly variable. For arrays within the fault zone, this variability could stem from fault zone structural properties, velocity heterogeneities, the extent and shape of the damage zone, or basin resonance from unconsolidated sedimentary basins within the region (Qiu et al., 2017; Share et al., 2017; Qin et al., 2018; Share et al., 2019; Zigone et al., 2019; Share et al., 2020; Share et al., 2023). This conjecture that fault zone properties are responsible for the ground motion variability is supported by the off-fault PY results, which show less variability than the fault zone arrays.

Typically, for near-field crustal earthquakes, vertical peak ground motions are smaller than the horizontal peak ground motion, where the peak ground motion vertical-to-horizontal ratios are expected to be ⅔ (Bozorgnia et al., 1995; Strasser et al., 2004). Previous work indicates that differences in vertical and horizontal motions can be influenced by site conditions (Bozorgnia and Campbell, 2016), and in particular, the presence of a 10–20 m soil layer can significantly increase vertical motions (Elgamal and He, 2004), and a study of data from Italy finds that the vertical-to-horizontal ratios can change with the seasons (Vassallo et al., 2022).

In this work, ternary plots show that five of the seven arrays (BB, DW, JF, SG, and TR) are consistent with the expectation that vertical ground motions are smaller than horizontal motions (Appendix B; Figures B1-B7). The PY array, however, shows that the horizontal and vertical motions are similar, with

the vertical motions only slightly diminished compared to the horizontal counterparts. However, the most notable variations in the ternary plots are for the RA array. Select events near Cahuilla, CA, produce elevated vertical motions at the RA array compared with horizontal motions, and these differences can be 2-3 fold (Figure 15 and Figure B5). This is a focal mechanism effect, as large amplitude P-waves drive this finding.

On average, the coefficient of variations for each channel's PGV and PGA measurements ranges from 26% to 36%. Although these percentages seem high, these deviations align with other studies of dense networks. Johnson et al., 2020 found deviations of 22% for the Sage Brush Flat 2D array that studied only vertical motion data. This 2D array is co-located with the SG linear array examined in this work. Our findings for the SG linear array show variations of 36% for both the PGV and PGA data. Compared with the Johson et al. (2020) study, the 14% difference likely results from the duration of the two datasets. The Johson et al. (2020) data spanned only ~1 month and used only vertical data, whereas the SG data used in this study encompassed ~13 years and had three components (see Table 1).

CONCLUSIONS

Ground motion measurements can be misleadingly elevated when the signals from temporal and spatial close events overprint. We attempt to avoid this problem by removing all earthquake pairs within 5 km and 5 minutes of each other, which removes 16% of the data in our base catalog. This process removes all M5+ events from our dataset but is an essential step, as we do not want to be misguided by elevated signals produced by multiple earthquakes. Our catalog includes only M2.5+ events. Thus, overprinted events consisting of earthquakes below magnitude 2.5 have not been accounted for, but SNR threshold requirements can flag some of these.

All data must have SNRs of 2.5 or higher. This criterion removed many events during the Salton Sea swarms (2012, 2016, 2020, and 2021), especially the very rigorous 2012 swarm. We compare the SNR histogram distributions of all networks and leverage the SNR distributions within each network to

identify anomalies at individual stations. Given the small apertures of the networks (spatial diameters within 3 km), the base expectation is that the SNR distributions would be similar for each station in each network. Indeed, similar SNR distributions are found for DW, JF, BB, and PY networks. But, for the RA, SG, and TR networks, the distribution differences reveal that the ground motions recorded within these small aperture arrays are non-uniform. Specifically, stations RA09, TR01, SGBFA, and SGBN1 show an elevated noise floor compared to the other stations in their respective networks. These differences stem from instrumentation problems over relatively short durations that were identified and corrected by the network operators. Although not shown here, we found that examining the SNR as a function of time can highlight instrumentation issues. Our SNR restriction successfully removes these spurious signals from our catalog.

Data quality vetting was essential for this work. In general, of the original waveform catalogs, only ~84% were retained after the overprinting, SNR, and non-viable signals were removed. Without quality vetting, incorrect results could ensue. In this work, we found it essential to study the data in aggregate and, if anomalous behavior was found, to examine the associated waveforms as the individual waveforms provided insight into the cause of the ground motions (i.e., car traffic, elevated noise floors, instrumental issues) that might otherwise go unnoticed.

The peak ground motions (velocity and acceleration) observed at each station in the seven networks have a consistent trend compared to the theoretical estimates. Observed ground motions recorded at networks within the fault system (BB, DW, JF, RA, and SG) produce larger-than-expected motions. In contrast, off-fault networks produce ground motions as expected (TR) or slightly lower than expected (PY). These observations indicate that ground motions within fault zones tend to be higher than expected, and ground motions outside faults tend to be as expected or lower.

We use ternary plots (Appendix B; Figures B1-B7) to examine the relative ground motions of these three-component data (i.e., comparing NS, EW, and Z motions). We generally find the ternary plot distributions as expected (i.e., lower vertical motions than horizontal motions and roughly similar horizontal motions) for networks BB, DW, JF, SG, and TR (except for station TR01). But the ternary plot distributions for PY and RA differ from the expected distributions. For the PY network, the three components of motion are roughly similar, although there is a slight tendency for reduced vertical motions compared to horizontal motions (Figure B4). Ternary plots for RA array data are more striking, showing consistently elevated vertical amplitudes (Figure 15 and Figure B5). This is particularly true for events near Cahuilla recorded by station RA01 (the westernmost station in the RA network; see Figure A5). As many stations in the RA network exhibit elevated vertical motions, we can confirm that the origin is not instrumental and instead is mapping the general response behavior of this region. We attribute this vertical motion elevation to focal mechanism effects. These ternary results highlight the importance of three-component networks that can reveal more information than single-component networks and that ternary plots can be a useful data quality metric.

In summary, we return to our original question – to what extent can data recorded at one station be extrapolated to nearby regions? This work indicates that even within the small spatial footprints of these arrays (spatial footprint diameters of 3 km or less; Table 2), the variability in PGV and PGA measurements for individual channels (NS, EW, and vertical) has a variation coefficient between 30±6%. This means that what is recorded at one location in the array can be ±30% different from what is recorded at another location in the array for the same channel.

DATA AND RESOURCES

Seismic catalogs were obtained from the USGS ComCat interface (last accessed June 2023). Waveform data and seismic station information were obtained from IRIS via the irisFetch.m matlab interface

(http://www.iris.edu/dms/nodes/dmc/software/downloads/irisFetch.m), last accessed November 2023.

ACKNOWLEDGMENTS

We thank those who planned and deployed these networks and collected and cataloged these data for public distribution. The project benefited from multiple email discussions with Robert Weekly about obtaining these data from IRIS. We also thank Yehuda Ben-Zion for helping frame the research project proposal and Katherine Puttroff for assisting with data acquisition speeds. This research was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Geosciences program under Award Number DE-SC0016527 (170455-00001).

REFERENCES

Allen, R. M., & Stogaitis, M. (2022). Global growth of earthquake early warning. Science, 375(6582), 717-718.

Anthony, R. E., Ringler, A. T., Wilson, D. C., Zebulon Maharrey, J., Gyure, G., Pepiot, A., ... & Voss, N. (2020). Installation and performance of the Albuquerque Seismological Laboratory small-aperture posthole array. Seismological Research Letters, 91(4), 2425-2437.

Arrowsmith, S. J., Trugman, D. T., MacCarthy, J., Bergen, K. J., Lumley, D., & Magnani, M. B.(2022). Big Data Seismology. Reviews of Geophysics, 60(2), e2021RG000769.https://doi.org/10.1029/2021RG000769.

Ben-Zion, Y., Vernon, F. L., Ozakin, Y., Zigone, D., Ross, Z. E., Meng, H., ... & Barklage, M. (2015). Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone. *Geophysical Journal International*, *202*(1), 370-380.

Beyer, K., & Bommer, J. J. (2006). Relationships between median values and between aleatory variabilities for different definitions of the horizontal component of motion. *Bulletin of the Seismological Society of America*, *96*(4A), 1512-1522.

Bozorgnia, Y., Niazi, M., & Campbell, K. W. (1995). Characteristics of free-field vertical ground motion during the Northridge earthquake. *Earthquake spectra*, *11*(4), 515-525.

Bozorgnia, Y., & Campbell, K. W. (2016). Ground motion model for the vertical-to-horizontal (V/H) ratios of PGA, PGV, and response spectra. *Earthquake Spectra*, *32*(2), 951-978.

Douglas, J., & Edwards, B. (2016). Recent and future developments in earthquake ground motion estimation. *Earth-Science Reviews*, *160*, 203-219.

Elgamal, A., & He, L. (2004). Vertical earthquake ground motion records: an overview. *Journal of Earthquake Engineering*, *8*(05), 663-697.

Frohlich, C. (1992). Triangle diagrams: ternary graphs to display similarity and diversity of earthquake focal mechanisms. *Physics of the Earth and Planetary interiors*, *75*(1-3), 193-198.

Given, D.D., Cochran, E.S., Heaton, T., Hauksson, E., Allen, R., Hellweg, P., Vidale, J., and Bodin, P., 2014, Technical implementation plan for the ShakeAlert production system—An Earthquake Early Warning system for the West Coast of the United States: U.S. Geological Survey Open-File Report 2014–1097, 25 p., https://dx.doi.org/10.3133/ofr20141097.

Graves, R. W., Aagaard, B. T., Hudnut, K. W., Star, L. M., Stewart, J. P., & Jordan, T. H. (2008). Broadband simulations for Mw 7.8 southern San Andreas earthquakes: Ground motion sensitivity to rupture speed. Geophysical research letters, 35(22).

Haoran Meng, Yehuda Ben-Zion, Christopher W. Johnson; Analysis of Seismic Signals Generated by Vehicle Traffic with Application to Derivation of Subsurface Q-Values. Seismological Research Letters 2021; 92 (4): 2354–2363. doi: https://doi.org/10.1785/0220200457.

Heath, D. C., Wald, D. J., Worden, C. B., Thompson, E. M., & Smoczyk, G. M. (2020). A global hybrid VS 30 map with a topographic slope–based default and regional map insets. *Earthquake Spectra*, *36*(3), 1570-1584.

Hirakawa, Evan, et al. "Rupture Directivity of the 25 October 2022 M w 5.1 Alum Rock Earthquake." The Seismic Record 3.2 (2023): 144-155.

Iervolino, I. (2022). Ground-Motion Observations and Probabilistic Seismic Hazard: Frequently Asked Questions. Seismological Society of America, 93(4), 2360-2366.

IRIS Transportable Array. (2003). *USArray Transportable Array*[Data set]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/TA (http://www.usarray.org/files/docs/pubs/usarray_5_years-final-lores.pdf),

Johnson, C. W., Meng, H., Vernon, F., & Ben-Zion, Y. (2019). Characteristics of ground motion generated by wind interaction with trees, structures, and other surface obstacles. *Journal of Geophysical Research: Solid Earth*, *124*(8), 8519-8539.

Johnson, C. W., Kilb, D., Baltay, A., & Vernon, F. (2020). Peak ground velocity spatial variability revealed by dense seismic array in southern California. *Journal of Geophysical Research: Solid Earth*, *125*(6), e2019JB019157.

Kilb, D., Gomberg, J., & Bodin, P. (2002). Aftershock triggering by complete Coulomb stress changes. *Journal of Geophysical Research: Solid Earth*, *107*(B4), ESE-2.

Kilb, D., Fan, W., Laske, G., 2021, Semantics at the Salton Sea: When does a swarm become a sequence?, Temblor, http://doi.org/10.32858/temblor.181

Kilb, D., Bunn, J. J., Saunders, J. K., Cochran, E. S., Minson, S. E., Baltay, A., ... & Kodera, Y. (2021). The PLUM earthquake early warning algorithm: A retrospective case study of West Coast, USA, data. *Journal of Geophysical Research: Solid Earth*, *126*(7), e2020JB021053.

Kurzon, I., F.L. Vernon, Y. Ben-Zion, and G. Atkinson (2014). Ground Motion Prediction Equations in the San Jacinto Fault Zone – Significant Effects of Rupture Directivity and Fault Zone Amplification, *Pure Appl. Geophys*., **171**, doi: 10.1007/s00024-014-0855-2.

Lecocq, T., Hicks, S. P., Van Noten, K., Van Wijk, K., Koelemeijer, P., De Plaen, R. S., ... & Xiao, H. (2020). Global quieting of high-frequency seismic noise due to COVID-19 pandemic lockdown measures. *Science*, *369*(6509), 1338-1343.

Li, Z., Peng, Z., Hollis, D., Zhu, L., & McClellan, J. (2018). High-resolution seismic event detection using local similarity for Large-N arrays. *Scientific reports*, *8*(1), 1646.

Li M, Rathje EM, Cox BR, Yust M. A Texas-specific VS30 map incorporating geology and VS30 observations. *Earthquake Spectra*. 2022;38(1):521-542. doi:10.1177/87552930211033622

Meng, H., & Ben-Zion, Y. (2018). Detection of small earthquakes with dense array data: Example from the San Jacinto fault zone, southern California. *Geophysical Journal International*, *212*(1), 442-457.

Mordret, A., Roux, P., Boué, P., & Ben-Zion, Y. (2019). Shallow three-dimensional structure of the San Jacinto fault zone revealed from ambient noise imaging with a dense seismic array. *Geophysical Journal International*, *216*(2), 896-905.

Olsen, K. B., J. C. Pechmann, and G. T. Schuster (1995a). Simulation of 3-D elastic wave propagation in the Salt Lake Basin, *Bull. Seism. Soc. Am* . 85, 1688–1710.

Olsen, K. B., R. J., Archuleta, and J. R. Matarese (1995b). Three-dimensional simulation of a magnitude 7.75 earthquake on the San Andreas fault in southern California, *Science* 270, 1628–1632.

Olsen, K. B. (2000). Site amplification in the Los Angeles basin from three-dimensional modeling of ground motion. *Bulletin of the Seismological Society of America*, *90*(6B), S77-S94.

Qiu, H., Ben-Zion, Y., Catchings, R., Goldman, Trifunac, M. D. (2016). Site conditions and earthquake ground motion–A review. Soil dynamics and earthquake engineering, 90, 88-100. M. R., Allam, A. A., & Steidl, J. (2021). Seismic imaging of the Mw 7.1 Ridgecrest earthquake rupture zone from data recorded by dense linear arrays. Journal of Geophysical Research: Solid Earth, 126(7), e2021JB022043.

Qiu, H., Ben-Zion, Y., Ross, Z. E., Share, P. E., & Vernon, F. L. (2017). Internal structure of the San Jacinto fault zone at Jackass Flat from data recorded by a dense linear array. *Geophysical Journal International*, *209*(3), 1369-1388.

Qin, L., Share, P. E., Qiu, H., Allam, A. A., Vernon, F. L., & Ben-Zion, Y. (2021). Internal structure of the San Jacinto fault zone at the Ramona Reservation, north of Anza, California, from dense array seismic data. *Geophysical Journal International*, *224*(2), 1225-1241.

Qin, L., Ben-Zion, Y., Qiu, H., Share, P. E., Ross, Z. E., & Vernon, F. L. (2018). Internal structure of the San Jacinto fault zone in the trifurcation area southeast of Anza, California, from data of dense seismic arrays. *Geophysical Journal International*, *213*(1), 98-114.

Rahpeyma, S., Halldorsson, B., Hrafnkelsson, B., Green, R. A., & Jónsson, S. (2019). Site effect estimation on two Icelandic strong-motion arrays using a Bayesian hierarchical model for the spatial distribution of earthquake peak ground acceleration. Soil Dynamics and Earthquake Engineering, 120, 369-385.

Sahakian, V., Baltay, A., Hanks, T., Buehler, J., Vernon, F., Kilb, D., & Abrahamson, N. (2018). Decomposing leftovers: Event, path, and site residuals for a small-magnitude Anza region GMPE. *Bulletin of the Seismological Society of America*, *108*(5A), 2478-2492.

Share, P. E., Ben-Zion, Y., Ross, Z. E., Qiu, H., & Vernon, F. L. (2017). Internal structure of the San Jacinto fault zone at Blackburn Saddle from seismic data of a linear array. *Geophysical Journal International*, *210*(2), 819-832.

Share, P. E., Allam, A. A., Ben-Zion, Y., Lin, F. C., & Vernon, F. L. (2019). Structural properties of the San Jacinto fault zone at Blackburn Saddle from seismic data of a dense linear array. *Pure and Applied Geophysics*, *176*, 1169-1191.

Share, P. E., Tábořík, P., Štěpančíková, P., Stemberk, J., Rockwell, T. K., Wade, A., ... & Ben-Zion, Y. (2020). Characterizing the uppermost 100 m structure of the San Jacinto fault zone southeast of Anza, California, through joint analysis of geological, topographic, seismic and resistivity data. Geophysical Journal International, 222(2), 781-794.

Share, P. E., Vernon, F. L., & Ben-Zion, Y. (2023). The variable continuous bimaterial interface in the San Jacinto fault zone revealed by dense seismic array analysis of fault zone head waves. Journal of Geophysical Research: Solid Earth, 128(2), e2022JB025070.

Strasser, F. O., Bommer, J. J., & Abrahamson, N. A. (2004). The need for upper bounds on seismic ground motion. In *Proceedings of 13th world conference on earthquake engineering, Vancouver, BC, Canada, paper* (No. 3361).

Tamhidi, A., Kuehn, N. M., & Bozorgnia, Y. (2023). Uncertainty quantification of ground motion time series generated at uninstrumented sites. Earthquake Spectra, 39(1), 551-576.

Trifunac, M. D. (2016). Site conditions and earthquake ground motion–A review. Soil dynamics and earthquake engineering, 90, 88-100.

Vassallo, M., Cultrera, G., Di Giulio, G., Cara, F., & Milana, G. (2022). Peak frequency changes from HV spectral ratios in central Italy: Effects of strong motions and seasonality over 12 years of observations. Journal of Geophysical Research: Solid Earth, 127(5), e2021JB023848.

Vernon, F. L., Fletcher, J., Carroll, L., Chave, A., & Sembera, E. (1991). Coherence of seismic body waves from local events as measured by a small-aperture array. Journal of Geophysical Research: Solid Earth, 96(B7), 11981-11996.

Vidale, J. E., and D. V. Helmberger (1988). Elastic finite-difference modeling of the 1971 San Fernando, California, earthquake, *Bull. Seism. Soc. Am* . 78122–141.

Wills, C. J., Gutierrez, C. I., Perez, F. G., & Branum, D. M. (2015). A next generation VS30 map for California based on geology and topography. *Bulletin of the Seismological Society of America*, *105*(6), 3083-3091.

Yong A. Comparison of Measured and Proxy-Based VS30 Values in California. *Earthquake Spectra*. 2016;32(1):171-192. doi:10.1193/013114EQS025M

Zigone, D., Ben-Zion, Y., Lehujeur, M., Campillo, M., Hillers, G., & Vernon, F. L. (2019). Imaging subsurface structures in the San Jacinto fault zone with high-frequency noise recorded by dense linear arrays. *Geophysical Journal International*, *217*(2), 879-893.

Tables.

Table 1. Data used in this study. Station codes stand for Jackass Flat (JF), Ramona (RA), Transect (TR), Pinyon Flat (YN), Blackburn (BB), Sage Brush Flat (SG), and Dry Wash (DW). Three of these networks are still operational (YA.RA, PY.BH, and YN.SG), but we only consider data through 30 June 2023 for this project. The number of waveforms includes only quakes that have passed data quality tests (i.e., overprinting, SNR, and non-viable).

Net.Sta	Deployment Start Date	Deployment End Date	Duration (years)	Number of waveforms	Magnitude Range	Depth Range (min - max km)
PY.BP*	2014-04-11	2021-06-06	7.2	31460	2.5-4.5	0-25
YN.BB*	2012-11-22	2014-07-10	1.6	4977	2.5-4.4	-1-19
YN.DW*	2012-03-26	2017-10-31	5.6	28557	2.5-4.5	-1-24
YN.JF*	2011-11-12	2016-11-15	5	20314	2.5-4.5	-1-24
YN.RA*	2012-08-11	2023-06-26	11	55880	2.5-4.5	-1-26
YN.SG*	2010-10-23	2023-06-26	13	41127	2.5-4.6	-1-25
YN.TR*	2011-11-12	2017-10-31	6	11015	2.5-4.5	-1-24

Table 2. Network sizes, sampling rates, geology, and elevation. Sensor codes HH* and HN* are velocity and acceleration, respectively.

Network.Station (number of stations in array)	Sensor Type	Sample Rate (Hz)	Geology	Length or aperture of the array (km)	Elevation Range (meters)	Elevation differential (meters)
YN.RA* (Nsta = 12)	HN*	200	Spanning the Clark Branch of the San Jacinto Fault Zone. Middle of the M7.3 1800 rupture. Termination of the M6.7 1918 rupture.	0.46	1467-1502	35
YN.SG* (Nsta = 7)	HN*	200	Spanning the Clark Branch of the San Jacinto Fault Zone. Middle of the M7.3 1800 rupture.	0.15	1430-1439	9
YN.JF* (Nsta = 9)	HH*	200	Spanning the Clark Branch of the San Jacinto Fault Zone. Middle of the M7.3 1800 rupture.	0.37	657-667	10
YN.DW* (Nsta = 13)	HH*	200	Spanning the Clark Branch of the San Jacinto Fault Zone. Middle of the M7.3 1800 rupture.	1.10	1085-1280	195 (Linear only 37)
YN.TR* (Nsta = 4)	HH*	200	Off-fault, between the Clark and Coyote Creek branches of the San Jacinto Fault Zone, southwest of the JF array	2.90	739-774	35
YN.BB* (Nsta = 7)	HH*	200	Spanning the Clark Branch of the San Jacinto Fault Zone. Middle of the M7.3 1800 rupture. Middle of the M6.7 1918 rupture.	0.18	1160-1110	21
PY.BP* (Nsta = 13)	HH*	200	Off-fault 2D array at the Pinyon Flats Observatory.	1.10	1251-1302	51

Table 3. Waveforms identified as non-seismic and deemed non-viable signals. These waveforms are removed from the catalogs and not considered in the analysis. Of the seven arrays, only data from three were flagged as problematic (JF, RA, and TR). Data from BB, DW, PY, and SG arrays were not flagged.

ID	NET.STA	Chan	Date:Time	Lat	Lon	Depth (km)	Mag	Dist (km)	Az (degrees)
ci37390968	YN.JFS3	HHE	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci37390968	YN.JFS2	HHN	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci37390968	YN.JFS1	HHN	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci37390968	YN.JFS1	HHZ	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci37390968	YN.JFN1	HHZ	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci37390968	YN.JF00	HHZ	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci38007544	YN.TR01	HHN	20170921T231047	32.38	-115.4	12.1	3.00	150	140
ci37766567	YN.TR01	HHZ	20170919T081359	33.08	-116.05	9.9	2.63	52	139
ci37766567	YN.TR01	HHN	20170919T081359	33.08	-116.05	9.9	2.63	52	139
ci37764391	YN.TR01	HHN	20170916T094550	33.17	-115.62	8.65	3.37	80	110
ci37764391	YN.TR01	HHE	20170916T094550	33.17	-115.62	8.65	3.37	80	110
ci37762367	YN.TR01	HHN	20170913T222127	32.84	-115.62	15.2	3.47	99	131
ci37390168	YN.TR01	HHZ	20150530T052356	33.88	-116.14	7.04	3.70	57	26
ci37390968	YN.JFN1	HHN	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci37390968	YN.JF00	HHN	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci37390968	YN.JFS3	HHE	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci37390968	YN.JFS1	HHN	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci37390968	YN.JFN1	HHZ	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci37390968	YN.JFS2	HHZ	20150531T130256	33.31	-116.28	12.8	3.59	21	137
ci11371154	YN.RA01	HNE	20130929T084131	32.9	-116.28	1.96	3.58	88	153
ci11371154	YN.RA01	HNN	20130929T084131	32.9	-116.28	1.96	3.58	88	153
ci11371154	YN.RA01	HNZ	20130929T084131	32.9	-116.28	1.96	3.58	88	153
ci38007544	YN.TR01	HHN	20170921T231047	32.38	-115.4	12.1	3.00	150	140
ci37766567	YN.TR01	HHZ	20170919T081359	33.08	-116.05	9.9	2.63	52	139
ci37764391	YN.TR01	HHN	20170916T094550	33.17	-115.62	8.65	3.37	80	110
ci37766567	YN.TR01	HHN	20170919T081359	33.08	-116.05	9.9	2.63	52	139
ci37390168	YN.TR01	HHZ	20150530T052356	33.88	-116.14	7.04	3.7	57	26
ci37762367	YN.TR01	HHN	20170913T222127	32.84	-115.62	15.2	3.47	99	131

Table 4. For each station in each network, we track the largest ground motions. All M5+ events were removed from the catalog after removing signals with potentially overprinted waveforms (close space/time earthquake pairs). Thus, a M4.6 earthquake was the largest earthquake in the base catalog.

NET.STA	PGV	Date Time	Lat	Lon	Depth (km)	Magnitude	Distance (km)
YN.BB01	0.22 (cm/s)	20140710T204143	33.5	-116.51	13.9	3.22	35
YN.BB02	0.22 (cm/s)	20131230T234421	33.7	-116.73	18.2	3.63	19
YN.BB03	0.21 (cm/s)	20131230T234421	33.7	-116.73	18.2	3.63	19
YN.BB04	0.25 (cm/s)	20130425T185943	33.65	-116.74	15.5	3.08	17
YN.BB05	0.23 (cm/s)	20130425T185943	33.65	-116.74	15.5	3.08	17
YN.BB06	0.41 (cm/s)	20130425T185943	33.65	-116.74	15.5	3.08	17
YN.BB07	0.3 (cm/s)	20130425T185943	33.65	-116.74	15.5	3.08	17
YN.JF00	0.84 (cm/s)	20160106T144234	33.96	-116.89	16.7	4.39	76
YN.JFN1	0.7 (cm/s)	20160106T144234	33.96	-116.89	16.7	4.39	76
YN.JFN2	0.7 (cm/s)	20160106T144234	33.96	-116.89	16.7	4.39	76
YN.JFN3	0.56 (cm/s)	20160106T144234	33.96	-116.89	16.7	4.39	76
YN.JFN4	0.56 (cm/s)	20140717T142434	33.43	-116.43	12.8	3.37	13
YN.JFS1	0.67 (cm/s)	20160106T144234	33.96	-116.89	16.7	4.39	76
YN.JFS2	0.63 (cm/s)	20160106T144234	33.96	-116.89	16.7	4.39	76
YN.JFS3	0.55 (cm/s)	20160106T144234	33.96	-116.89	16.7	4.39	76
YN.JFS4	0.41 (cm/s)	20160106T144234	33.96	-116.89	16.7	4.39	76
YN.DW01	0.45 (cm/s)	20160610T111411	33.46	-116.42	11.2	3.46	12
YN.DW02	0.37 (cm/s)	20160610T111411	33.46	-116.42	11.2	3.46	12
YN.DW03	0.42 (cm/s)	20160610T111411	33.46	-116.42	11.2	3.46	12
YN.DW04	0.53 (cm/s)	20160610T111411	33.46	-116.42	11.2	3.46	12
YN.DW05	0.33 (cm/s)	20140717T142434	33.43	-116.43	12.8	3.37	14
YN.DW06	0.32 (cm/s)	20140717T142434	33.43	-116.43	12.8	3.37	14
YN.DW07	0.63 (cm/s)	20160610T111411	33.46	-116.42	11.2	3.46	12
YN.DW08	0.67 (cm/s)	20160610T111411	33.46	-116.42	11.2	3.46	12
YN.DW09	0.53 (cm/s)	20140717T142434	33.43	-116.43	12.8	3.37	14
YN.DW10	0.65 (cm/s)	20140717T142434	33.43	-116.43	12.8	3.37	14
YN.DW11	0.51 (cm/s)	20160610T111411	33.46	-116.42	11.2	3.46	12
YN.DW12	0.2 (cm/s)	20150531T130256	33.31	-116.28	12.8	3.59	28
YN.DWRPT	0.67 (cm/s)	20160610T111411	33.46	-116.42	11.2	3.46	12
YN.TR01	0.49 (cm/s)	20160106T144234	33.96	-116.89	16.7	4.39	76
YN.TR02	0.28 (cm/s)	20130920T061043	33.34	-116.39	12.8	3.8	16
YN.TR03	0.25 (cm/s)	20150531T130256	33.31	-116.28	12.8	3.59	21
YN.TR04	0.54 (cm/s)	20120518T103712	33.32	-116.4	6.34	3.61	11
PY.BPH01	0.26(cm/s)	20230401T011607	33.38	-116.91	13	4.15	51
PY.BPH02	0.29(cm/s)	20230401T011607	33.38	-116.91	13	4.15	51
PY.BPH03	0.27(cm/s)	20230401T011607	33.38	-116.91	13	4.15	51
PY.BPH04	0.26(cm/s)	20230401T011607	33.38	-116.91	13	4.15	51
PY.BPH05	0.12(cm/s)	20180226T184441	33.48	-116.5	12.6	3.66	20
PY.BPH06	0.23(cm/s)	20230401T011607	33.38	-116.91	13	4.15	51
PY.BPH07	0.27(cm/s)	20230401T011607	33.38	-116.91	13	4.15	51

PY.BPH08	0.22(cm/s)	20230401T011607	33.38	-116.91	13	4.15	51
PY.BPH09	0.2(cm/s)	20230401T011607	33.38	-116.91	13	4.15	51
PY.BPH10	0.23(cm/s)	20230401T011607	33.38	-116.91	13	4.15	51
PY.BPH11	0.23(cm/s)	20220930T115011	33.48	-116.51	9.3	3.41	18
PY.BPH12	0.24(cm/s)	20230401T011607	33.38	-116.91	13	4.15	50
PY.BPH13	0.22(cm/s)	20230401T011607	33.38	-116.91	13	4.15	51

NET.STA	PGA	Date Time	Lat	Lon	Depth (km)	Mag	Distance (km)
YN.RA01	0.02(g)	20200202T005950	33.58	-116.82	5	3.64	12
YN.RA02	0.05(g)	20230401T011607	33.38	-116.91	13	4.15	34
YN.RA03	0.045(g)	20230401T011607	33.38	-116.91	13	4.15	34
YN.RA04	0.043(g)	20230401T011607	33.38	-116.91	13	4.15	34
YN.RA05	0.036(g)	20230401T011607	33.38	-116.91	13	4.15	34
YN.RA06	0.065(g)	20230401T011607	33.38	-116.91	13	4.15	34
YN.RA07	0.017(g)	20160109T114310	33.66	-116.77	14	3.3	16
YN.RA08	0.062(g)	20230401T011607	33.38	-116.91	13	4.15	34
YN.RA09	0.065(g)	20230401T011607	33.38	-116.91	13	4.15	34
YN.RA10	0.05(g)	20230401T011607	33.38	-116.91	13	4.15	34
YN.RA11	0.036(g)	20180815T012426	33.48	-116.8	2	4.41	18
YN.RA12	0.039(g)	20230401T011607	33.38	-116.91	13	4.15	35
YN.SGBF0	0.058(g)	20190318T085542	33.52	-116.57	2	3.11	3
YN.SGBFA	0.062(g)	20230410T145100	33.48	-116.49	14	2.72	18
YN.SGBN1	0.066(g)	20190318T085542	33.52	-116.57	2	3.11	3
YN.SGBN2	0.048(g)	20190318T085542	33.52	-116.57	2	3.11	3
YN.SGBS1	0.044(g)	20190318T085542	33.52	-116.57	2	3.11	3
YN.SGBS2	0.02(g)	20220113T031908	33.47	-116.44	14	3.86	21
YN.SGBS3	0.037(g)	20190318T085542	33.52	-116.57	2	3.11	3

Figures.

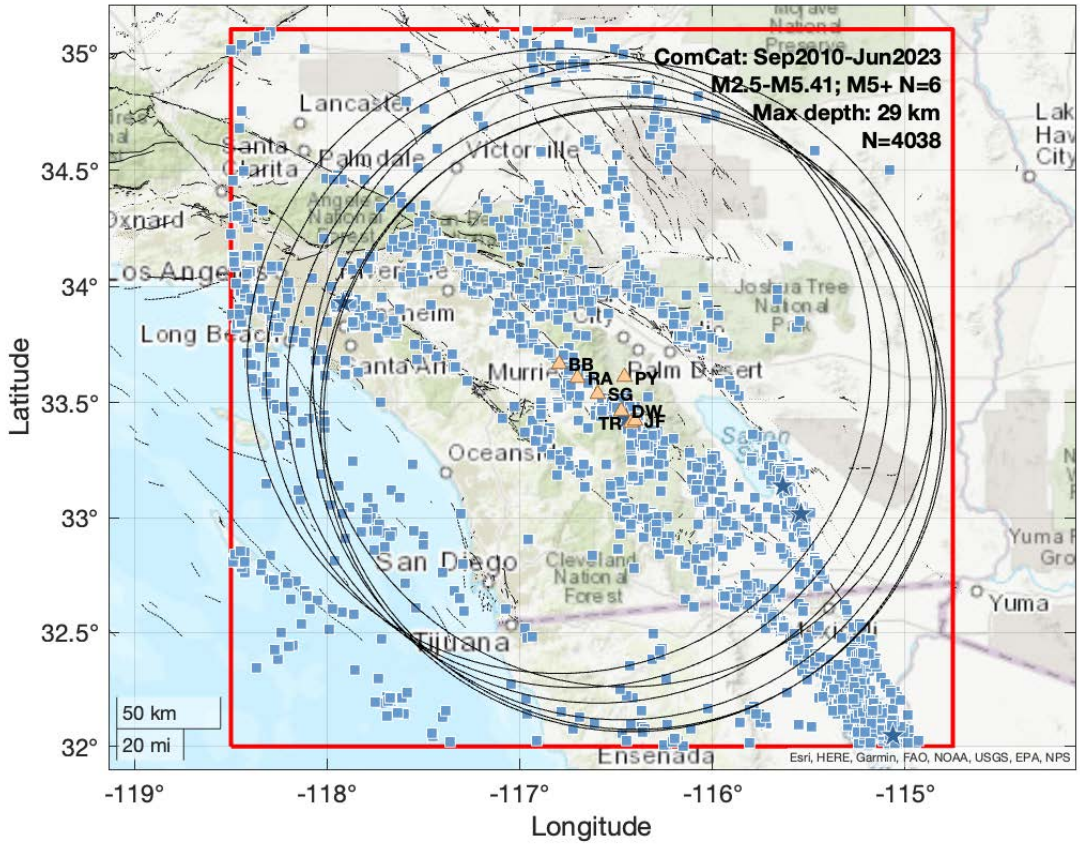

Figure 1. Study region in southern California. A base catalog of 4038 events (red square region; M2.5+; 2010-09-20 through 2023-06-30; latitudes 32.0 to 35.1; longitudes -118.5 to -114.75) is used to identify and remove overprinted events (events within 5 km and 5 mins) that have the potential to elevate ground motions misleadingly. The final data of interest are sub-catalogs that contain events within 150 km of each network centroid (circular regions).

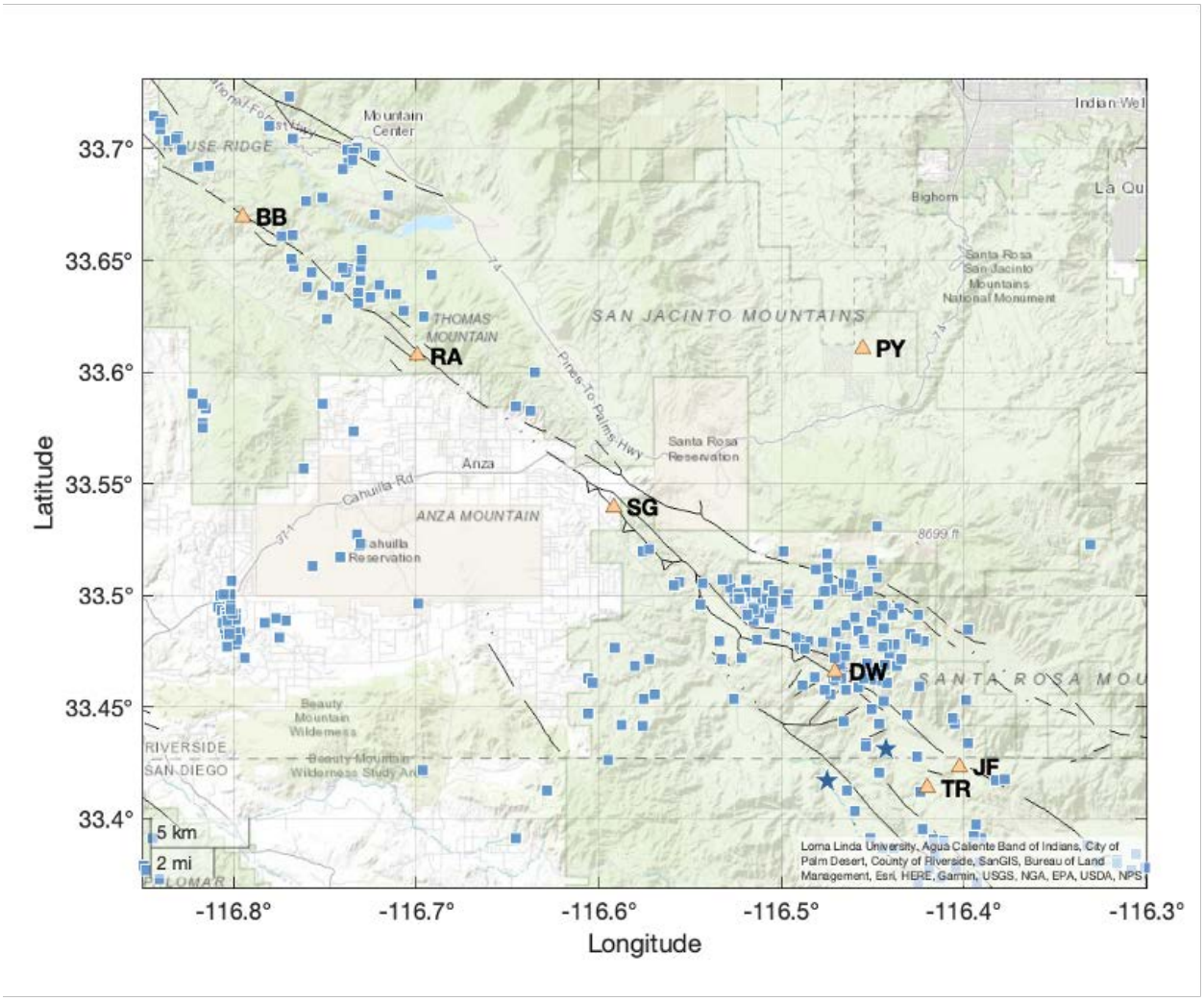

Figure 2. Zoom of the network locations showing the spatial extent of the seven network centroids (BB, RA, SG, PY, DW, JF, and TR). Five networks are within the San Jacinto fault zone (BB, RA, SG, DW, and JF), and two are not (PY and TR).

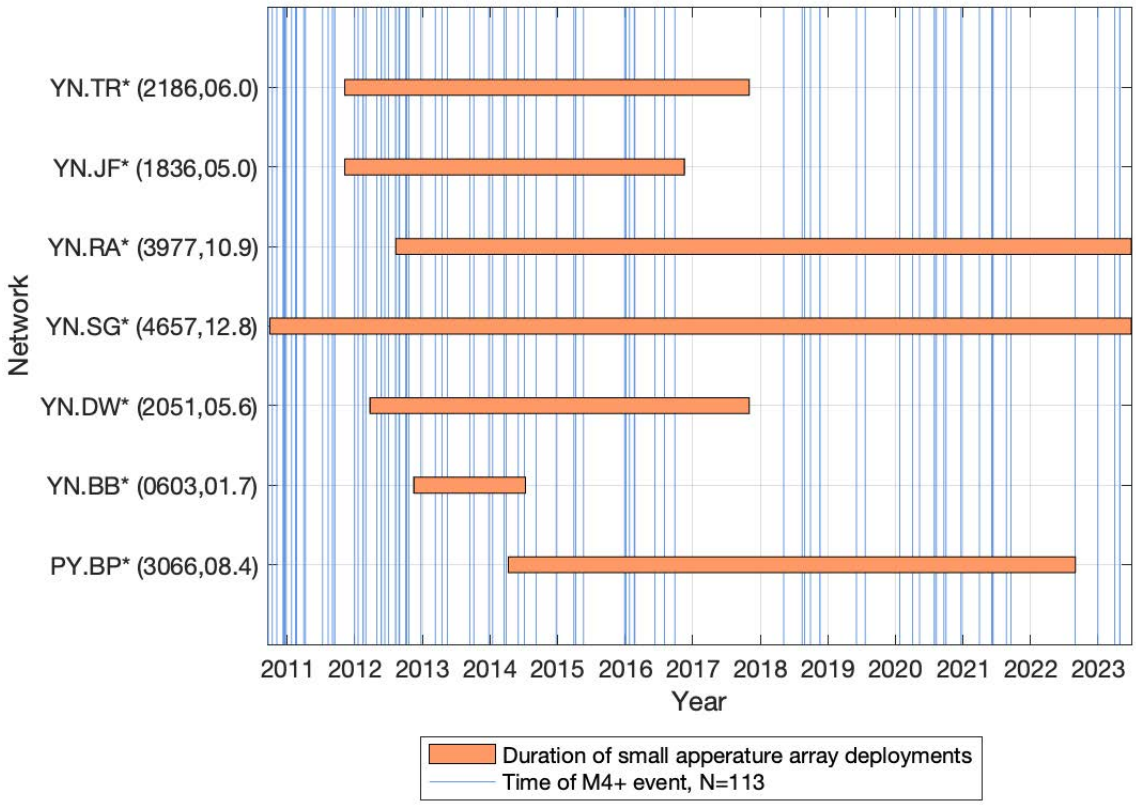

Figure 3. Deployment durations for this study's seven small aperture arrays (temporal durations indicated by orange bar lengths; numbers in parenthesis are duration days and years, respectively). All networks have a deployment span of a year or more. The collective temporal extent of these deployments spans 20 September 2010 through 30 June 2023, although some networks are still running. Blue vertical lines indicate times of M4+ earthquakes within the study region. Interestingly, within our study region in 2017, there were no M4+ earthquakes.

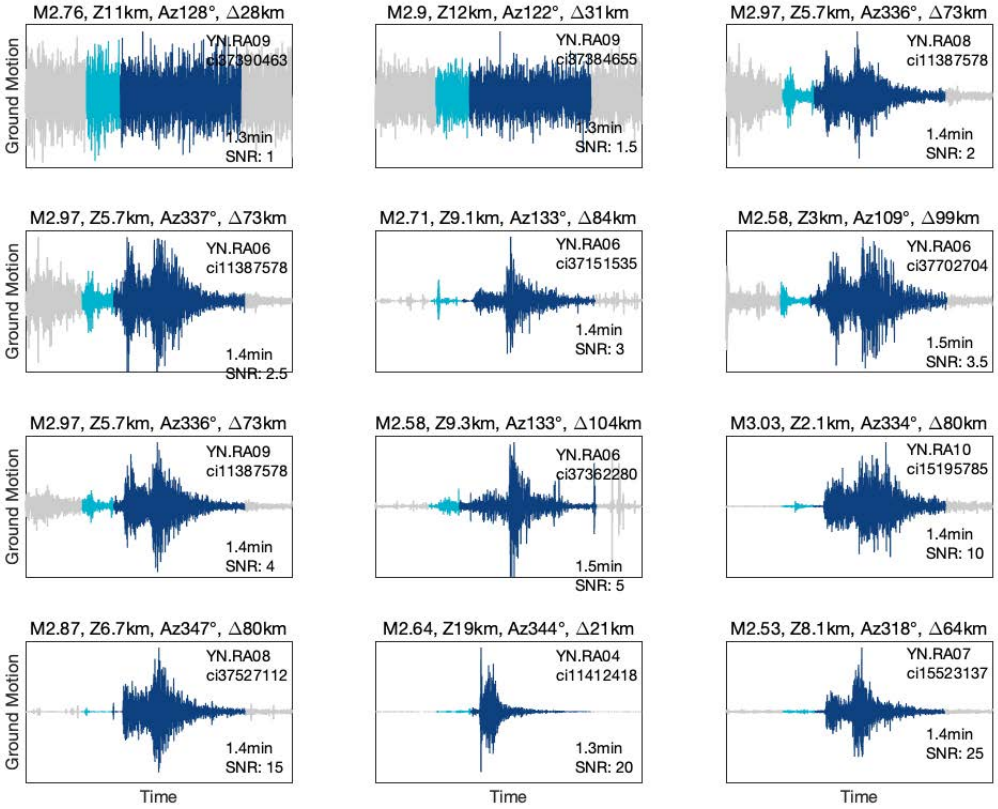

Figure 4. SNR level examples, ordered from smallest (top left) to largest (bottom right). The SNR is computed by taking the ratio of the largest absolute value of the data waveform (dark blue) divided by the largest absolute value of the noise waveform (light blue). Subtitles include the earthquake magnitude, depth, azimuth, and hypocentral distance values. Additional information (network, station, earthquake ID, waveform duration, and SNR) is listed within the subfigures. Only data with SNRs 2.5+ are retained. The intermittent spike behavior is from local traffic on a nearby roadway.

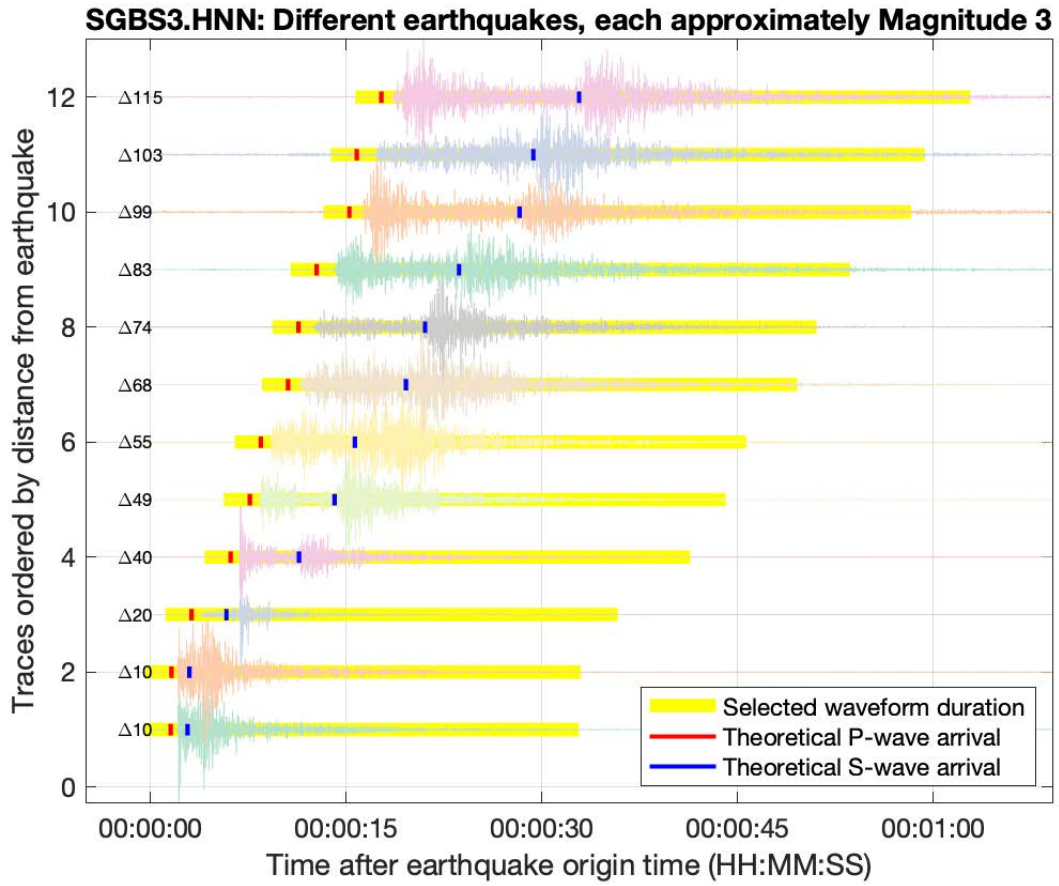

Figure 5. Examples of the variability of seismogram duration selection for ~M3 earthquakes recorded at station SGBS3.HNN. These ~M3 earthquakes are located at different distances (Δ values listed) from the recording station. Using theoretical estimates of P-wave and S-wave arrival times (red and blue small vertical bars, respectively), we select data from 2 seconds prior to the theoretical p-wave arrival (red vertical bar) through 30 seconds after the theoretical s-wave arrival (blue vertical bar). These estimated arrival times are not exact but are sufficient for our data snippet selection purposes. Importantly, data snippet durations (yellow bar lengths) differ to account for the larger time separation between the P- and S-wave arrival times for more distant recordings.

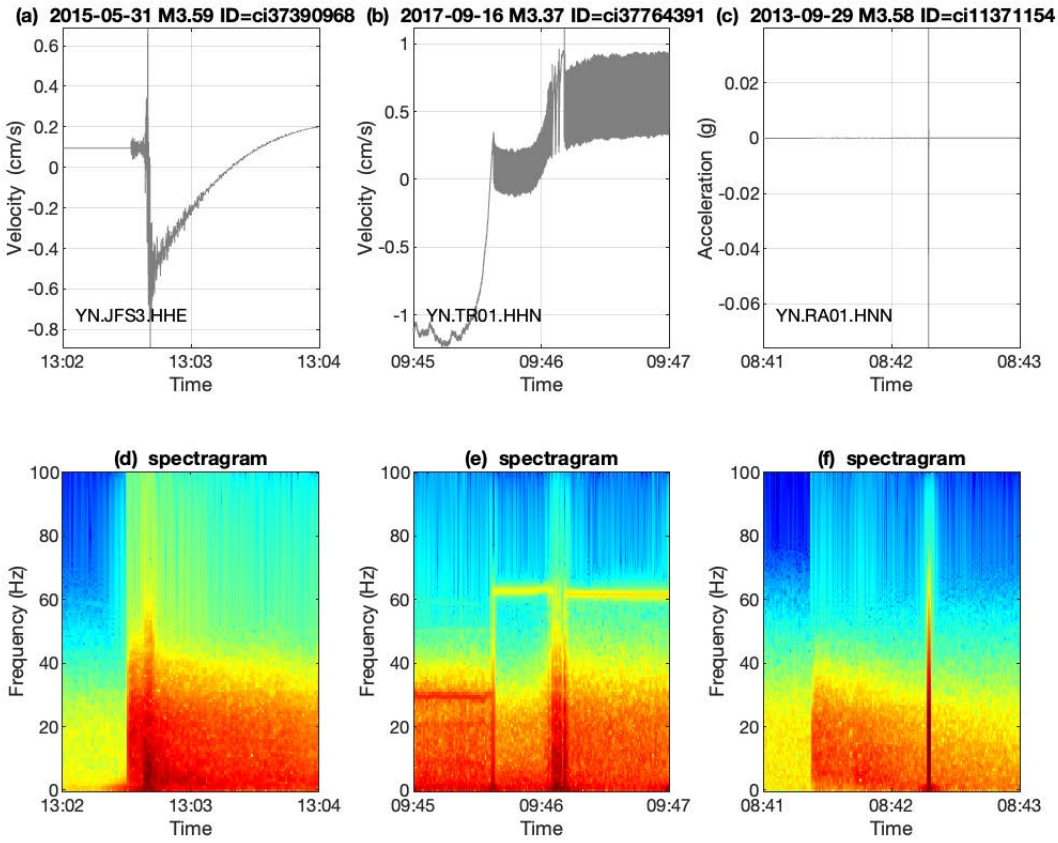

Figure 6. Examples of signals flagged as non-seismic and removed from the catalogs. These include time-domain signals (a-c; top row) and the associated frequency-domain signals (d-f; bottom row). These examples include near-clipped signals (a and d), instrumental noise (b and e), and a car passing along a nearby road (c and f).

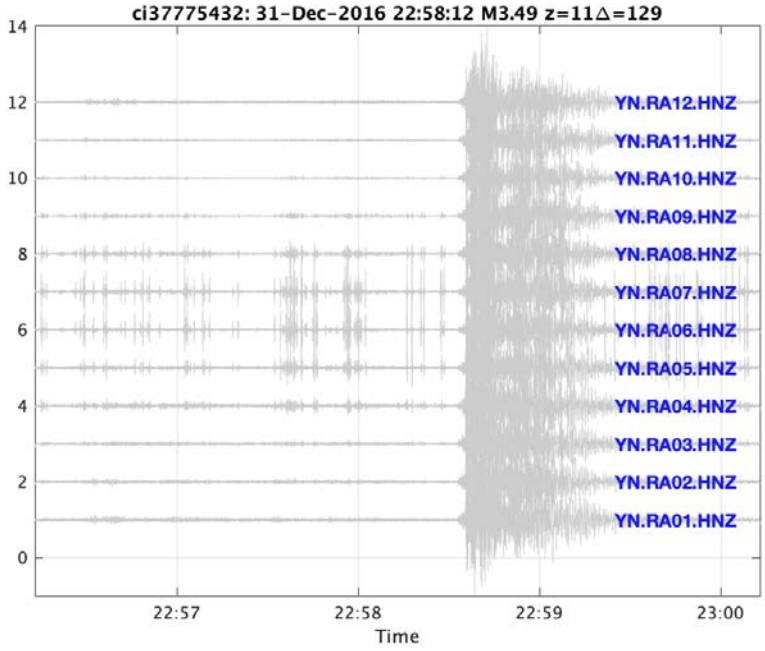

Figure 7. An example of car traffic signals recorded by the RA array, which traverses the Table Mountain Truck Trail road along the San Jacinto Fault Zone. The prevalent spike signals recorded by stations closest to the road/fault (RA08, RA07, RA06, and RA05) are the signature of passing cars. This example is from 31 December 2016, when traffic was perhaps elevated because of New Year's Eve celebration preparations.

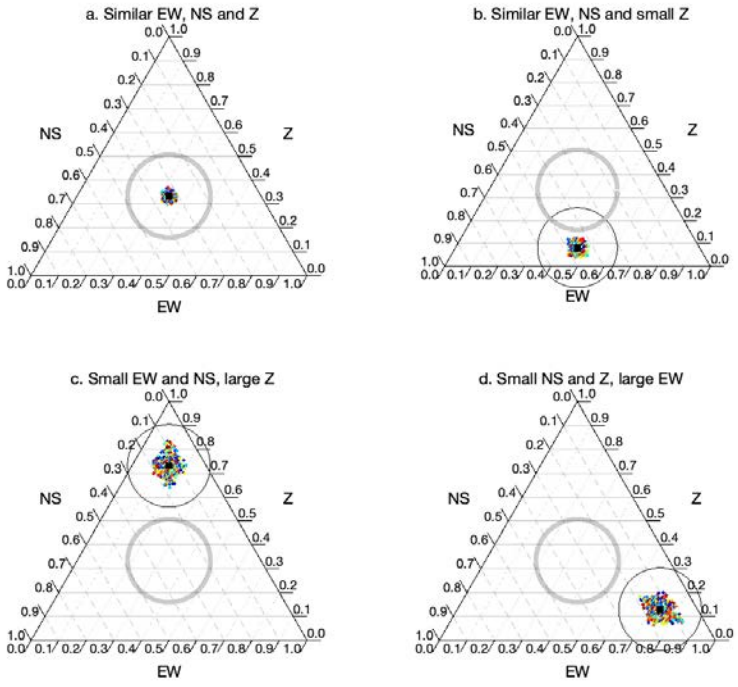

Figure 8. Guide to interpretation of ternary plots. The EW, NS, and Z grid lines are dotted, dashed, and solid, respectively. Titles explain the synthetic sample data distributions. A reference circle (thick line) is placed at the center of the triangle, encompassing the region where the points would plot if all components had equal values. Thin-line circles are centered at the median of the suite of values. Both circles have a radius of 0.1 units. (a) When three components (Z, NS, EW) have similar maximum absolute amplitudes, the ternary plot points will be located in the middle of the triangle. (b) When the horizontal components are approximately equal and the vertical amplitudes small, the points will be located toward the bottom middle of the ternary diagram. (c) When the vertical component amplitudes are larger than the horizontal component amplitudes, the points will be located at the top of the triangle. (d) When the EW component is larger than the NS and Z components, the points will be located at the bottom right of the triangle.

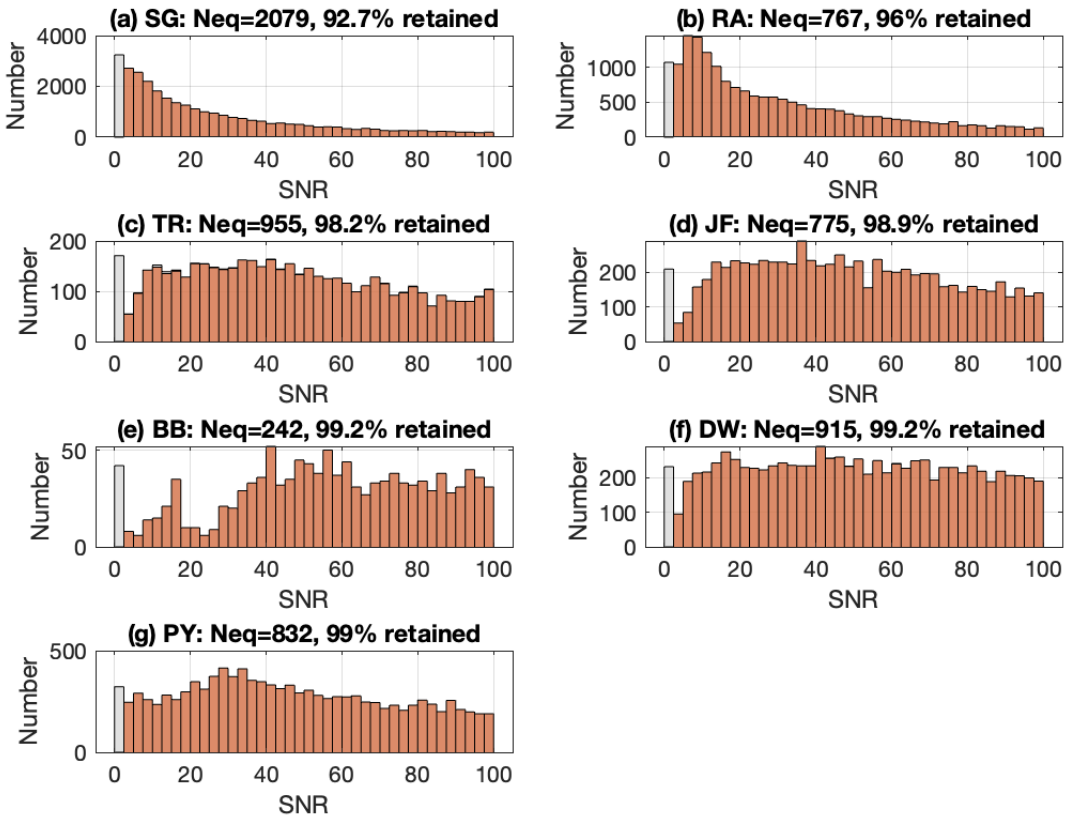

Figure 9. Summary SNR histograms for each network. Data with SNR<2.5 (grey) are removed. The number of earthquakes tested, and the number of earthquakes retained after the SNR restrictions are listed in the sub-plot titles. The SG and RA networks are the most prone to SNR issues, retaining only 92.7% and 96% of the data, respectively.

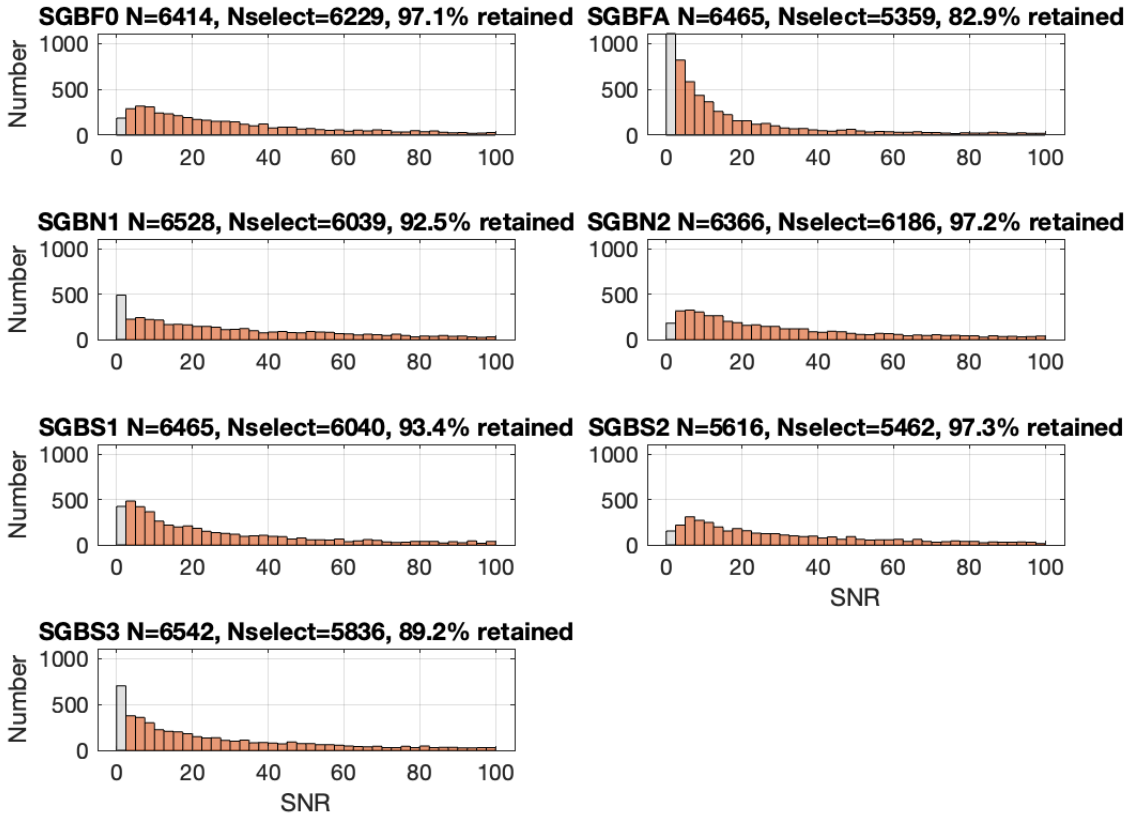

Figure 10. As in Figure 9, but for the SG network data. SNR for each station in the SG network, y-axis scales are identical in all subplots. Stations SGBFA, SGBN1, and SGBS3 net the most SNR below the threshold value (SNR<2.5; elevated grey bars for SGBFA, SGBN1, and SGBS3 distributions), retaining 82.9%, 92.5%, and 89.2% of the data, respectively.

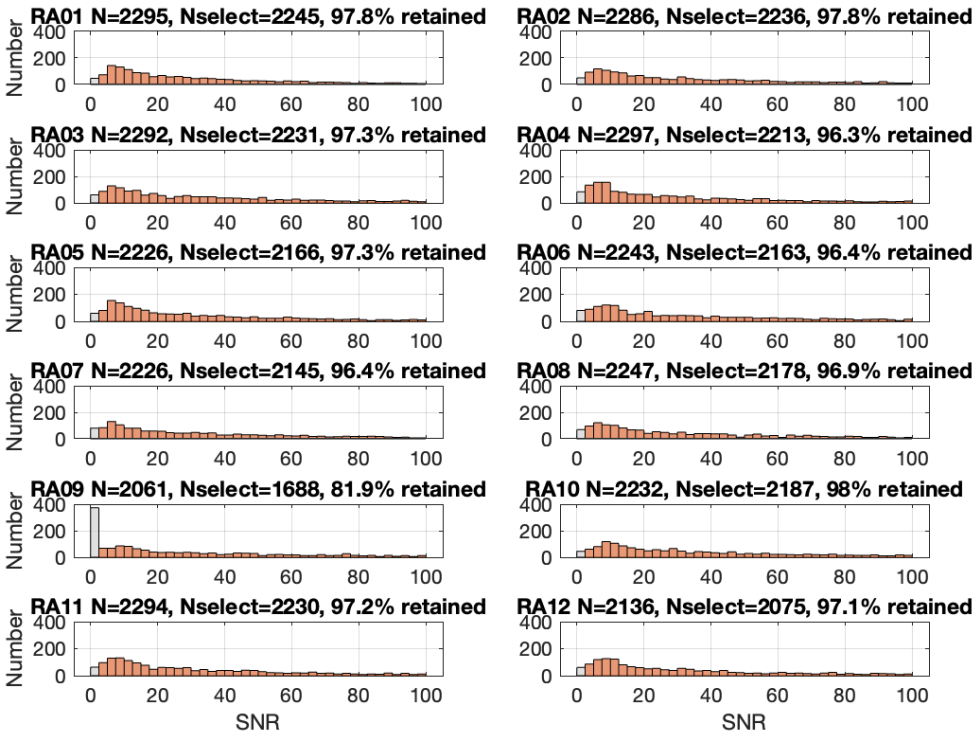

Figure 11. As in Figures 9-10, but for the RA data. Note that station RA09 has substantially more data that do not meet the SNR restriction (i.e., larger gray bar in RA09 histogram). Only 81.9% of the RA09 data is retained.

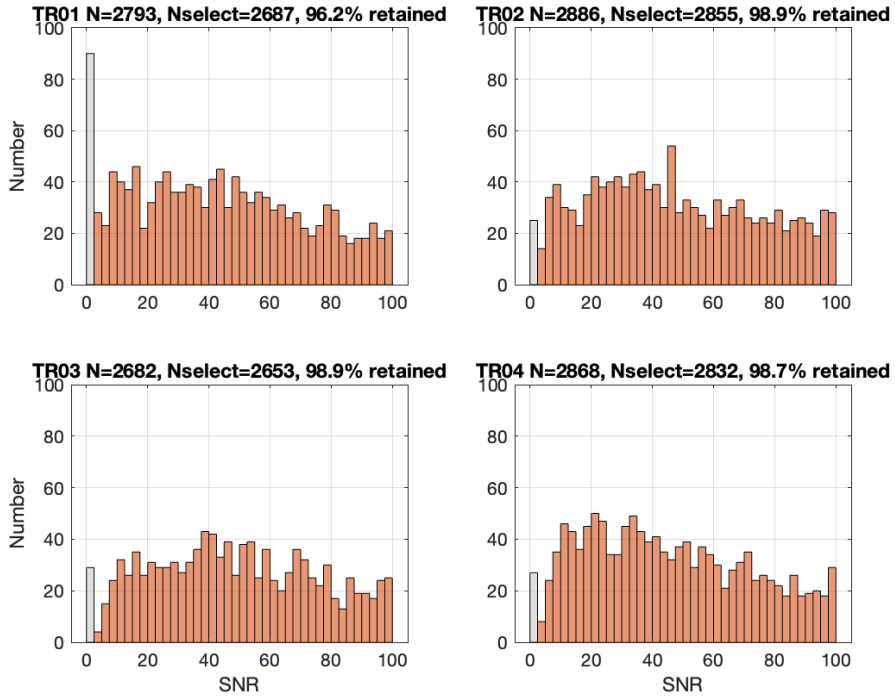

Figure 12. As in Figures 9-11, but for the TR data. Station TR01 has substantially more data that do not meet the SNR restriction (i.e., larger grey bar in TR01 histogram).

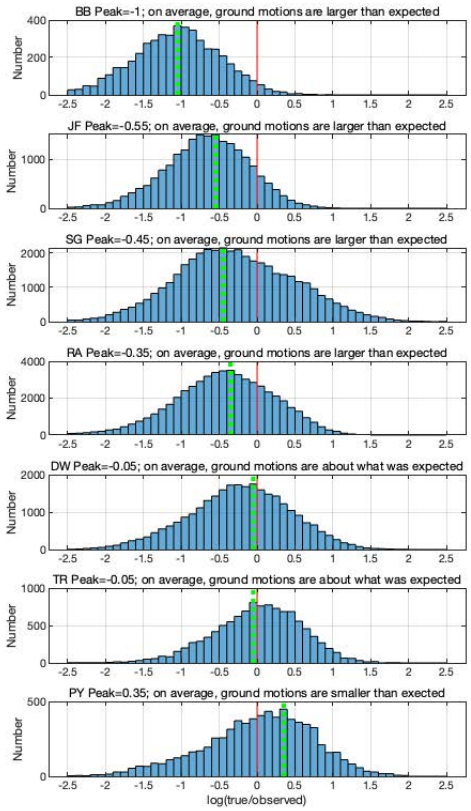

Figure 13. Summary of overall results. Histograms of log(true/observed) ground motions for each network ordered from smaller than expected motions (top; BB array) to larger than expected motions (bottom; PY array). Notice that the two bottom histograms show the least deviation from the expected value (dashed vertical green line close to the vertical red line), and both these arrays are off-fault. The top five histograms are from arrays within the fault zone and show more deviation than the off-fault arrays.

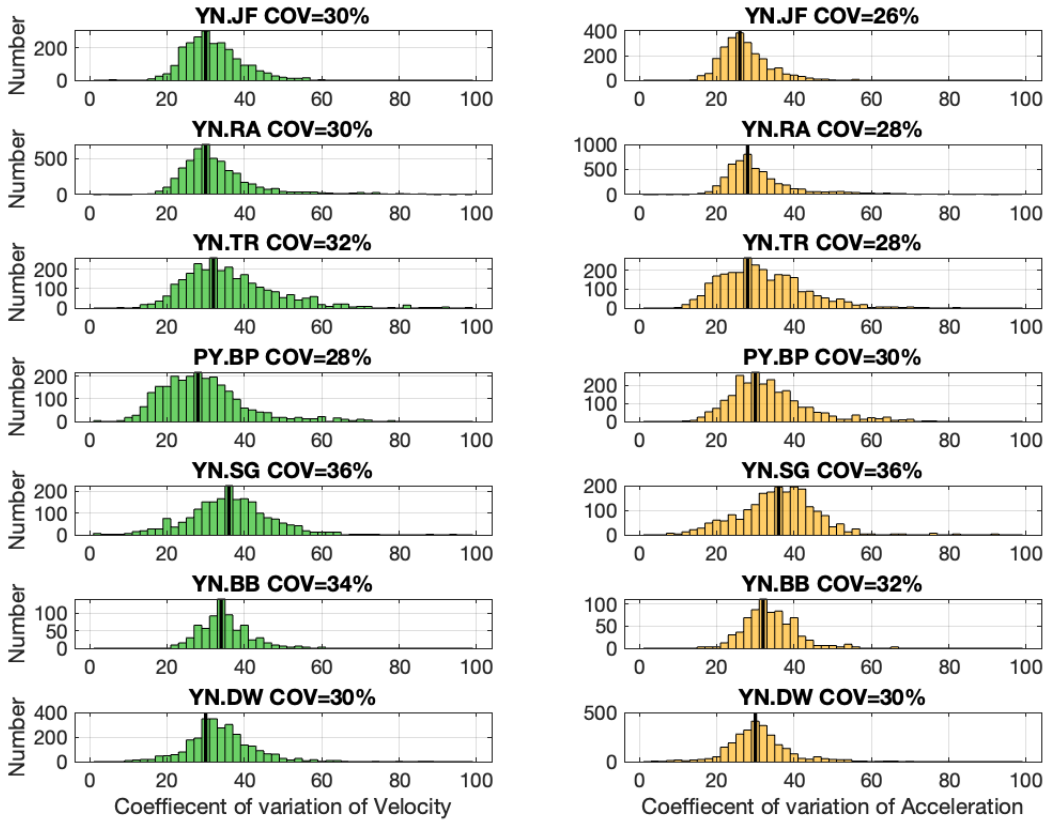

Figure 14. Coefficient of variations (standard deviation/mean) percentages show consistent values for PGV (left, green) and PGA (right, orange), deviating by no more than 4% between the PGV and PGA estimates for each network (i.e., compare values in the titles for each row). These results indicate that, on average, there is a ~30% deviation in the PGV and PGA values across the arrays. This means that the peak ground motion measured at one station in the array can be ~30% different from that measured at another station in the same array.

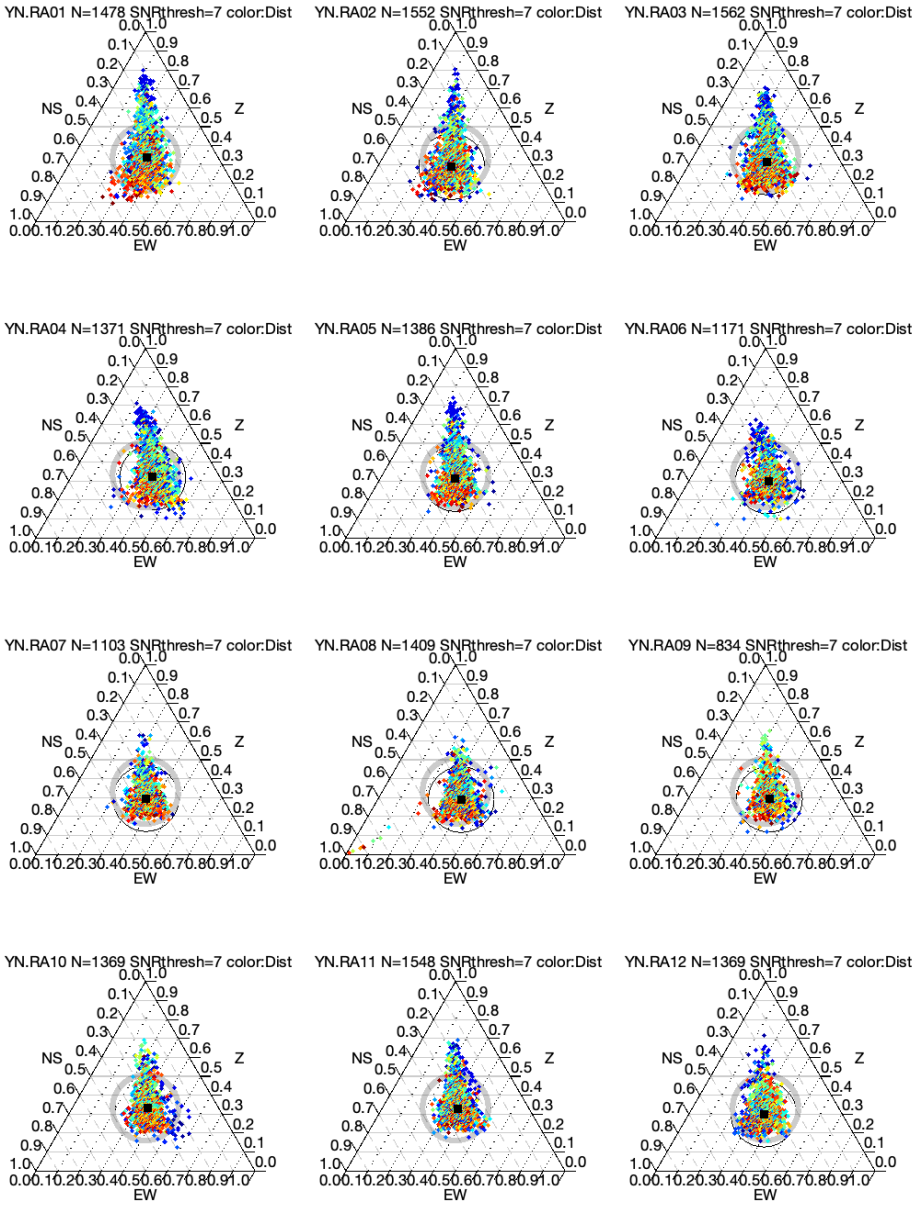

Figure 15. Ternary plots for the RA array using a SNR threshold of seven. Although the median values remain similar at all stations (i.e., alignment of the reference circle (thick line) and median (thin line)), there is a notable tendency for some of the data to show elevated vertical motion (i.e., points located toward the top of the triangle). Points are color-coded by distance, where cool colors represent events closer to the array than warm colors. Most blue points at the top of the triangle are events from Cahuilla, California. A map of the RA array can be found in Figure A6.

Appendix A: Mapped Spatial Distributions of the Seven Small Aperture Arrays.

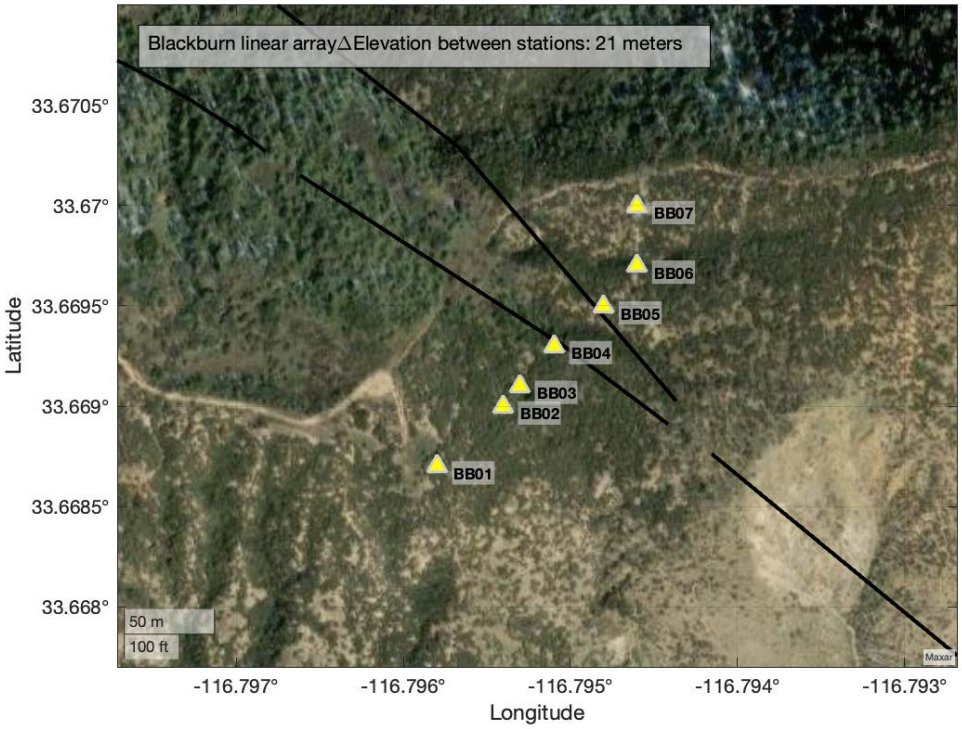

Figure A1. Mapped distribution of the Blackburn networks seismic stations.

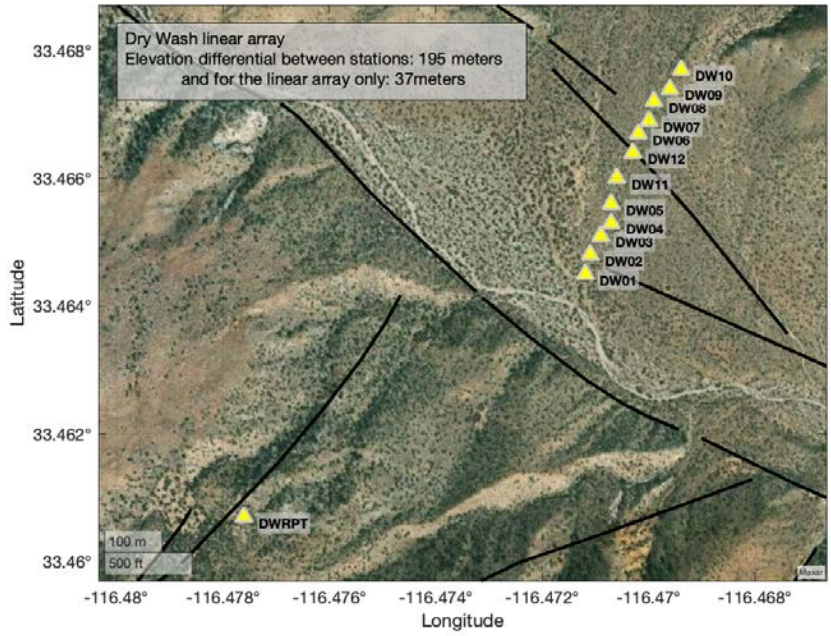

Figure A2. Mapped distribution of the Dry Wash networks seismic stations.

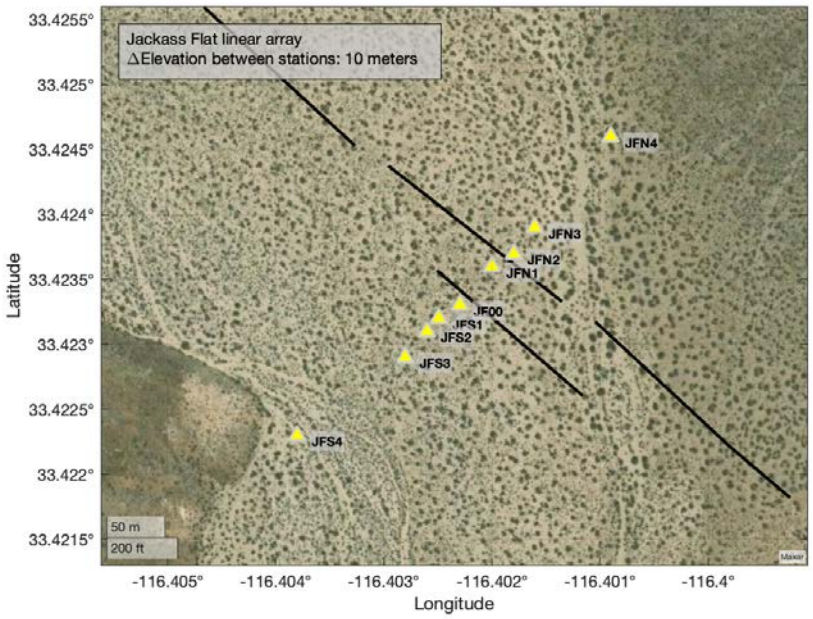

Figure A3. Mapped distribution of the Jackass Flat networks seismic stations.

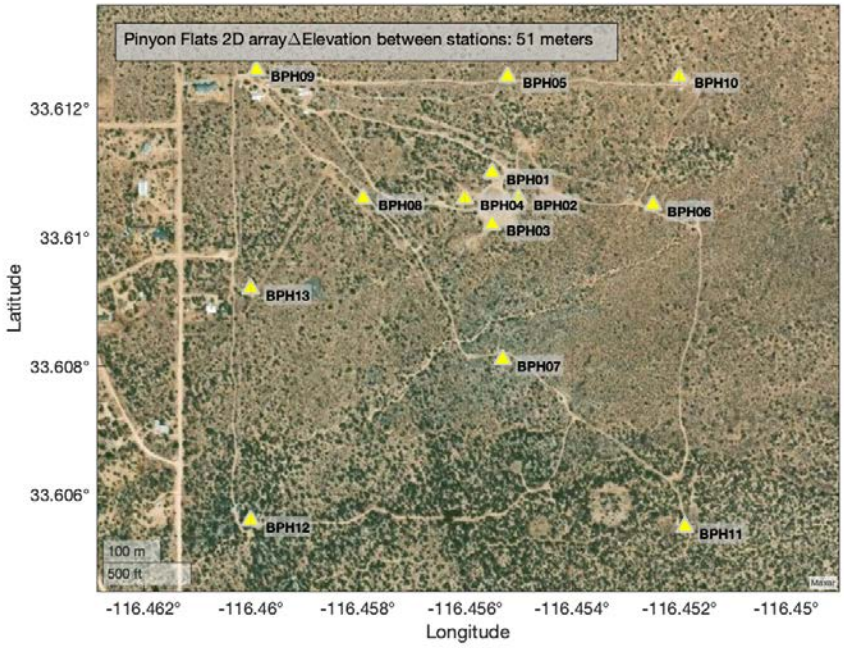

Figure A4. Mapped distribution of the Pinyon array network seismic stations.

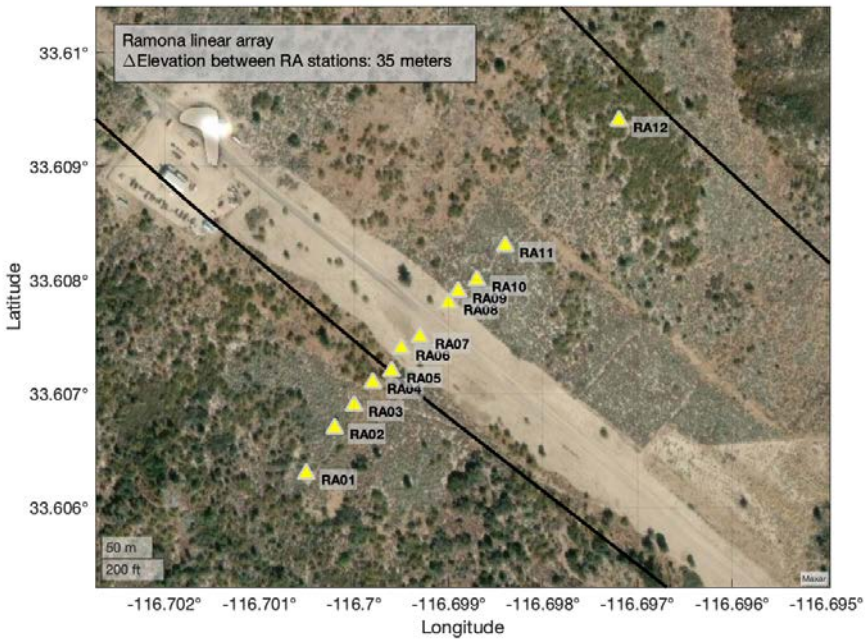

Figure A5. Mapped distribution of the Ramona array network seismic stations.

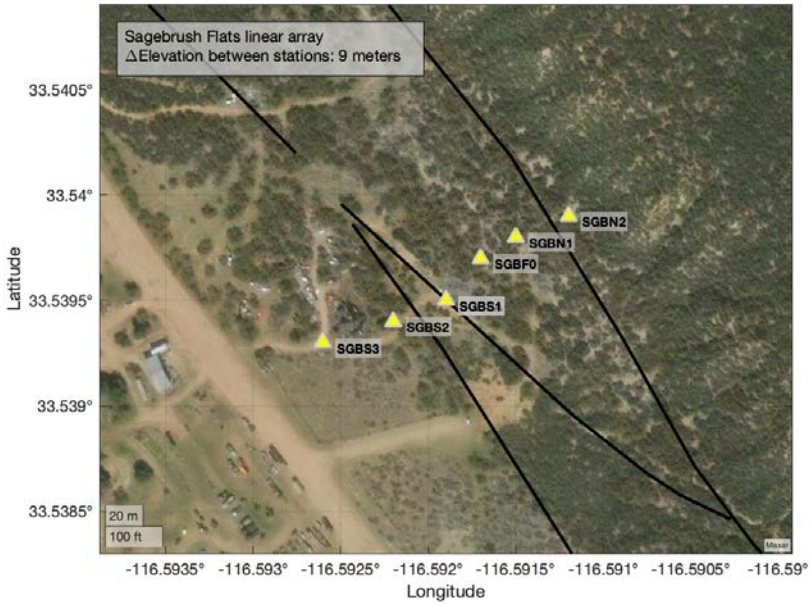

Figure A6. Mapped distribution of the Sagebrush Flat array network seismic stations. Note that stations SGBN1 and SGBFA (not labeled) are co-located but at different depths (SGBF0 is at an elevation of 1434; SGBFA is at an elevation of 1432 meters).

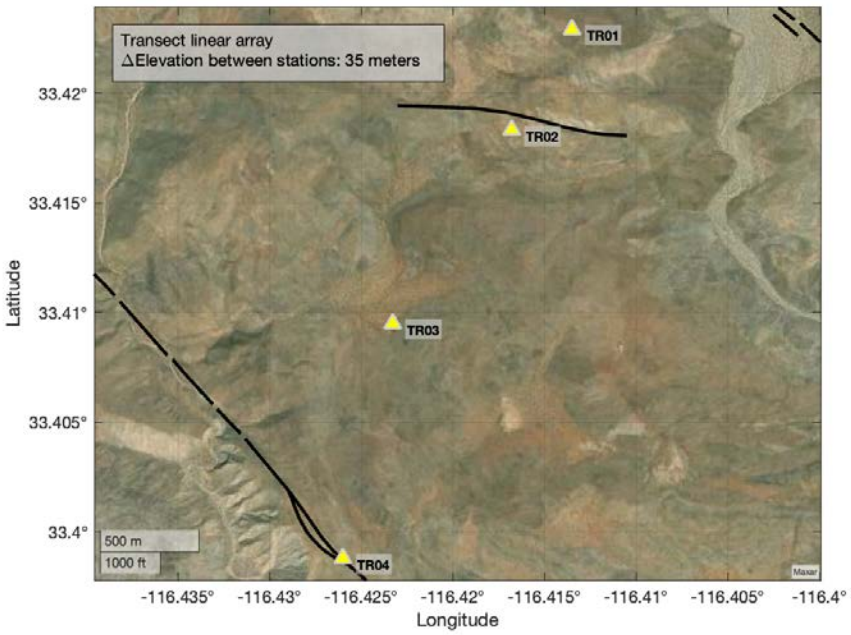

Figure A7. Mapped distribution of the Transect array network seismic stations.

Appendix B: Ternary Plots For Each Array

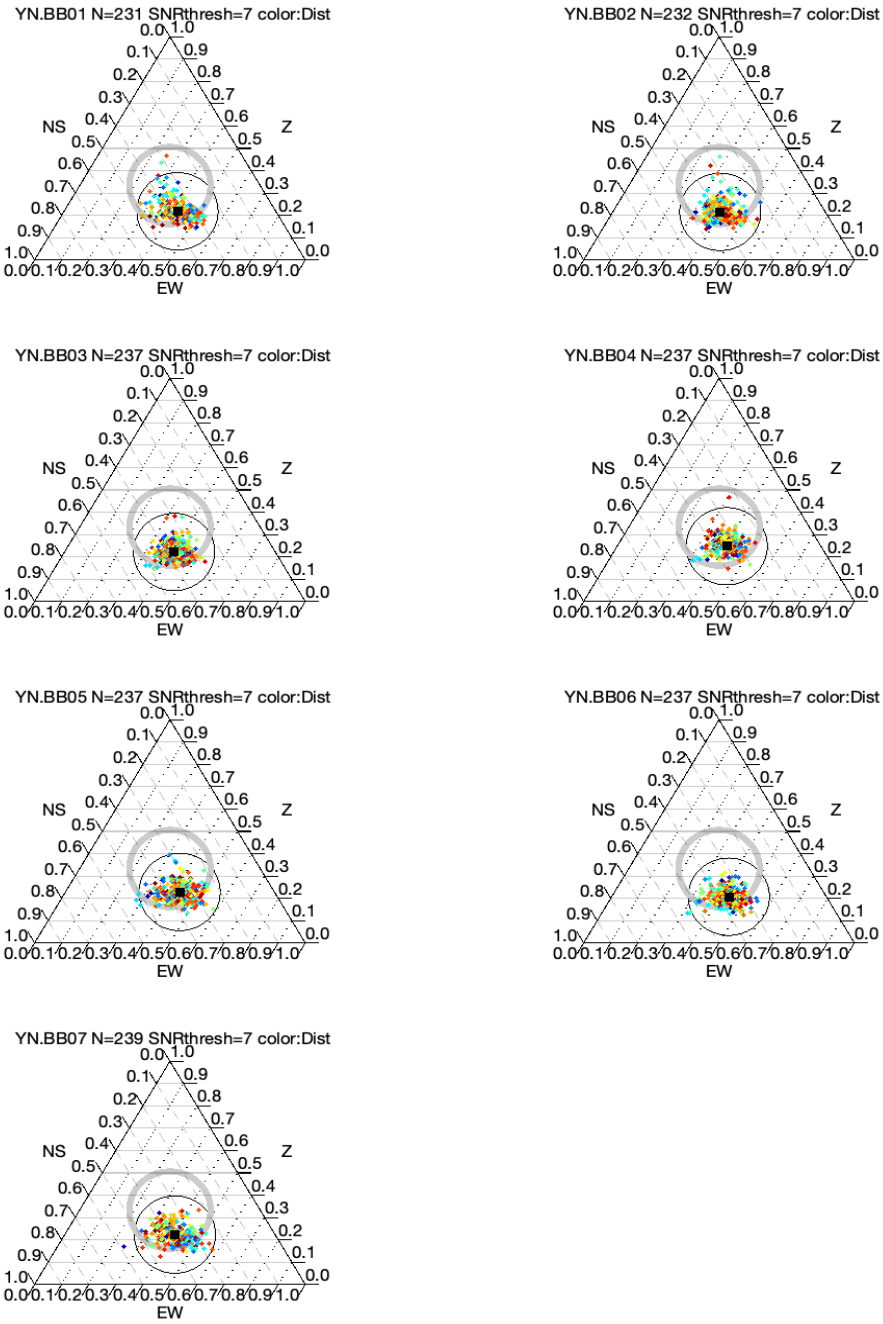

Figure B1. Ternary plots for Blackburn linear array, using a SNR threshold of seven.

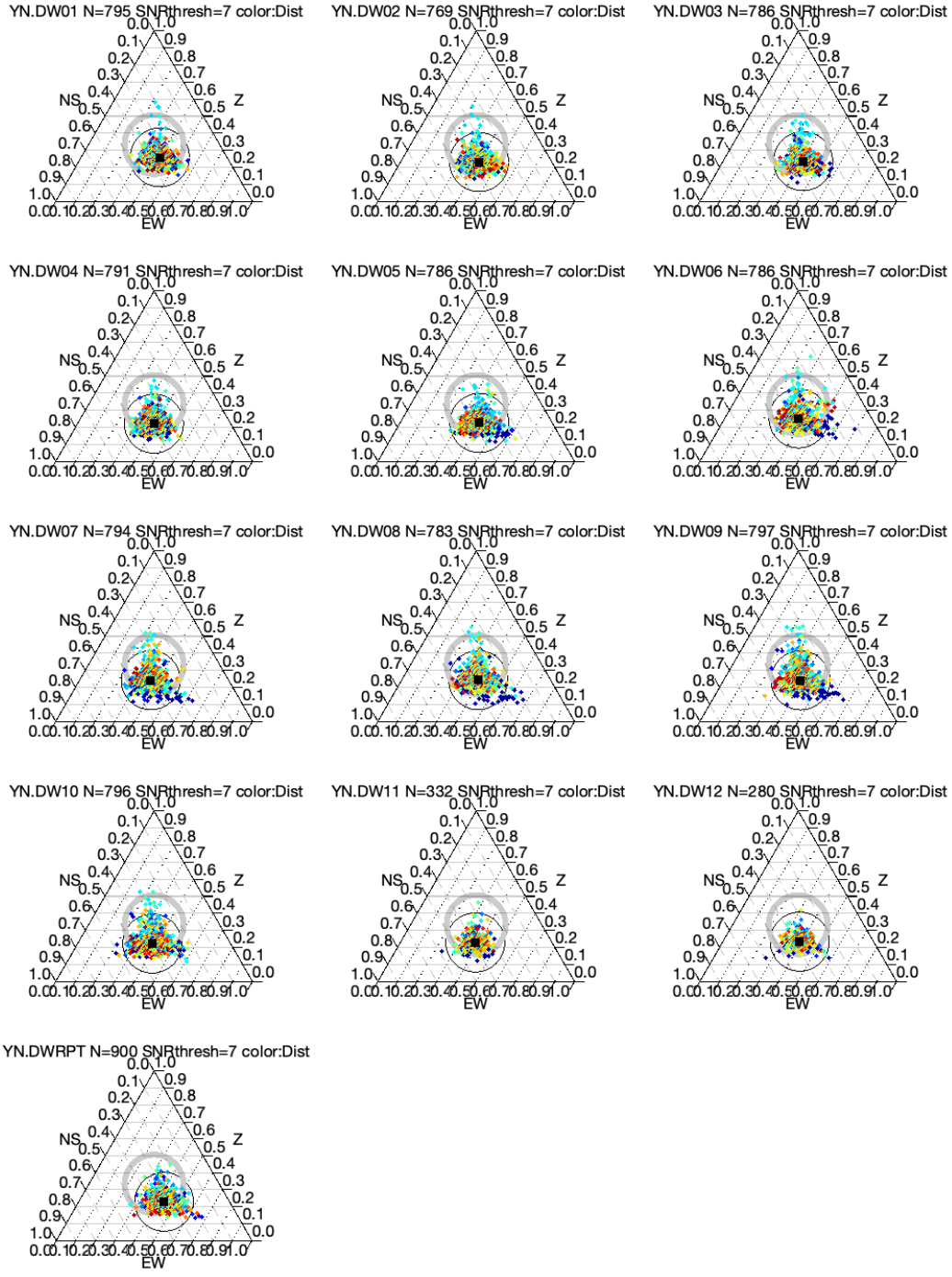

Figure B2. Ternary plots for Drywash linear array, using a SNR threshold of seven.

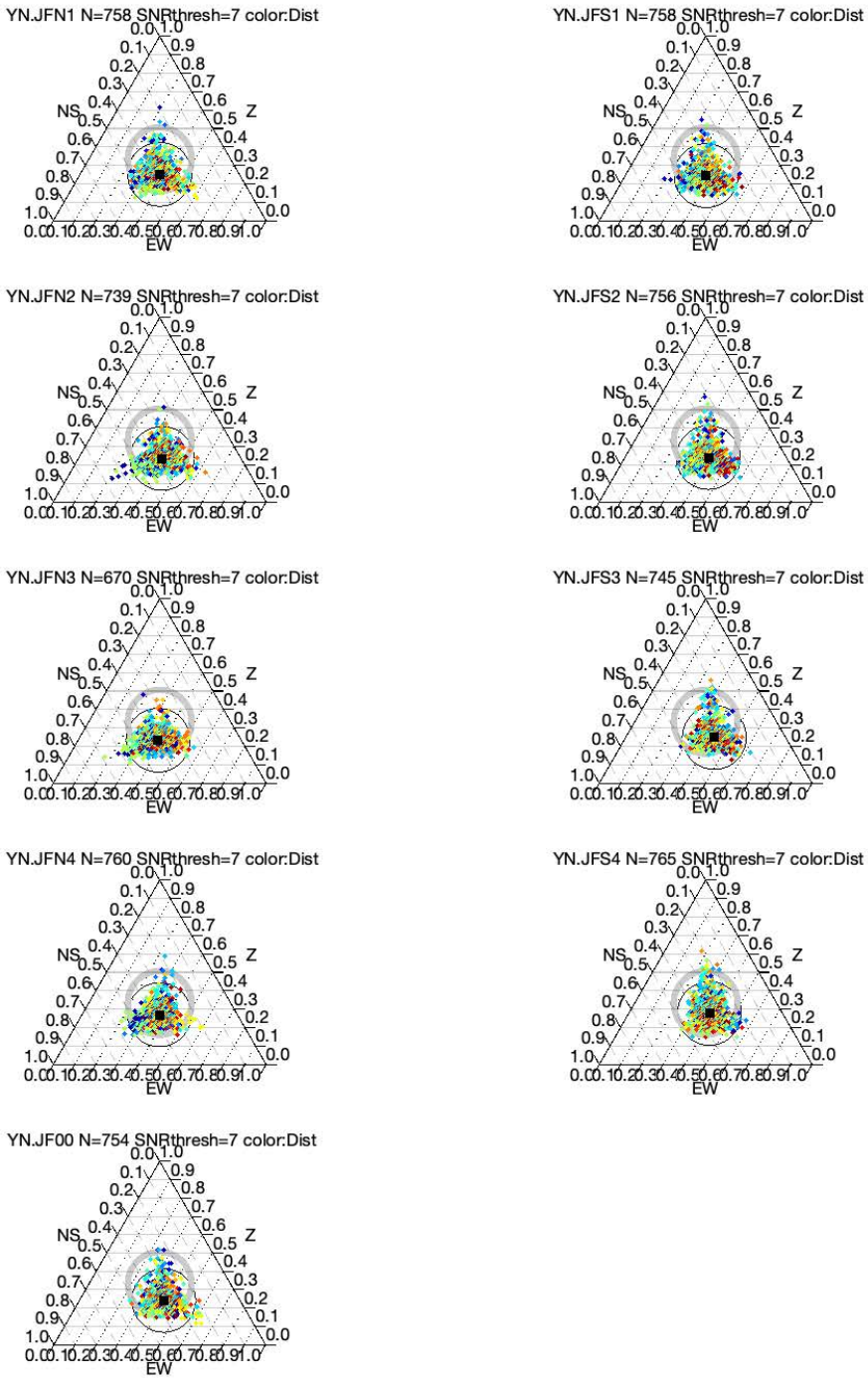

Figure B3. Ternary plots for Jackass Flat linear array, for a SNR of seven.

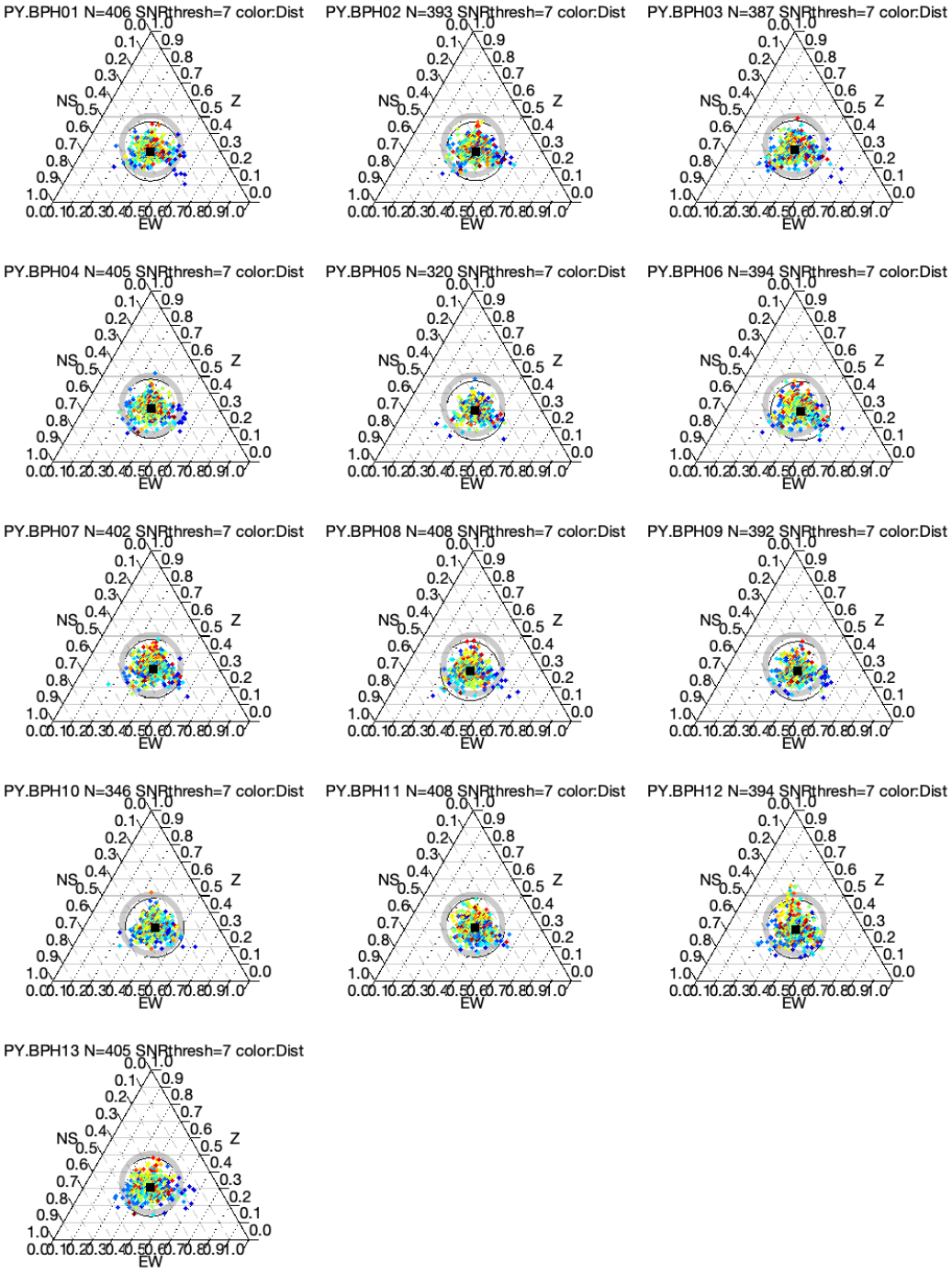

Figure B4. Ternary plots for PY 2D array, using a SNR threshold of seven.

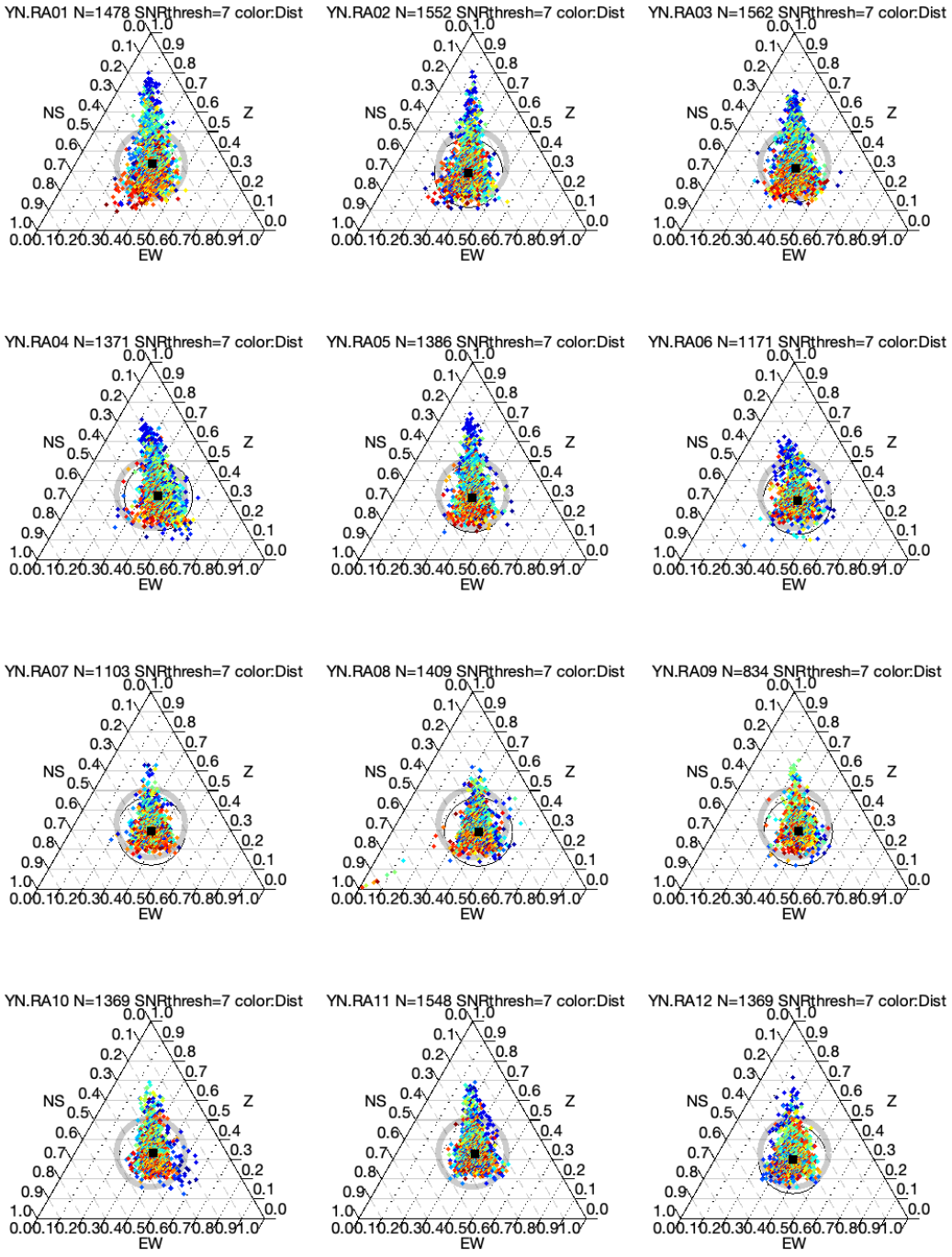

Figure B5. Ternary plots for Ramona linear array, using a SNR threshold of seven.

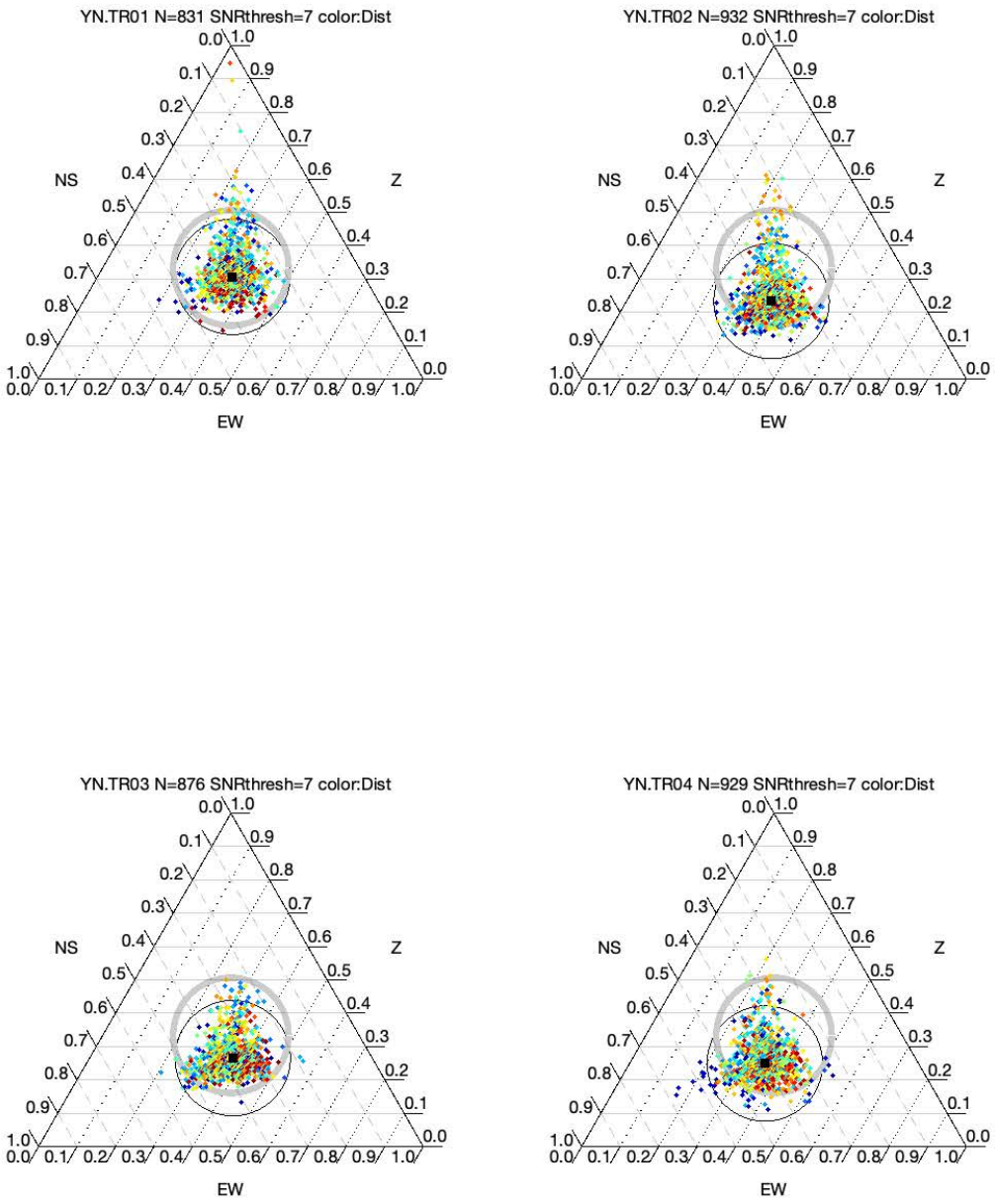

Figure B6. Ternary plots for Transect linear array, using a SNR threshold of seven.

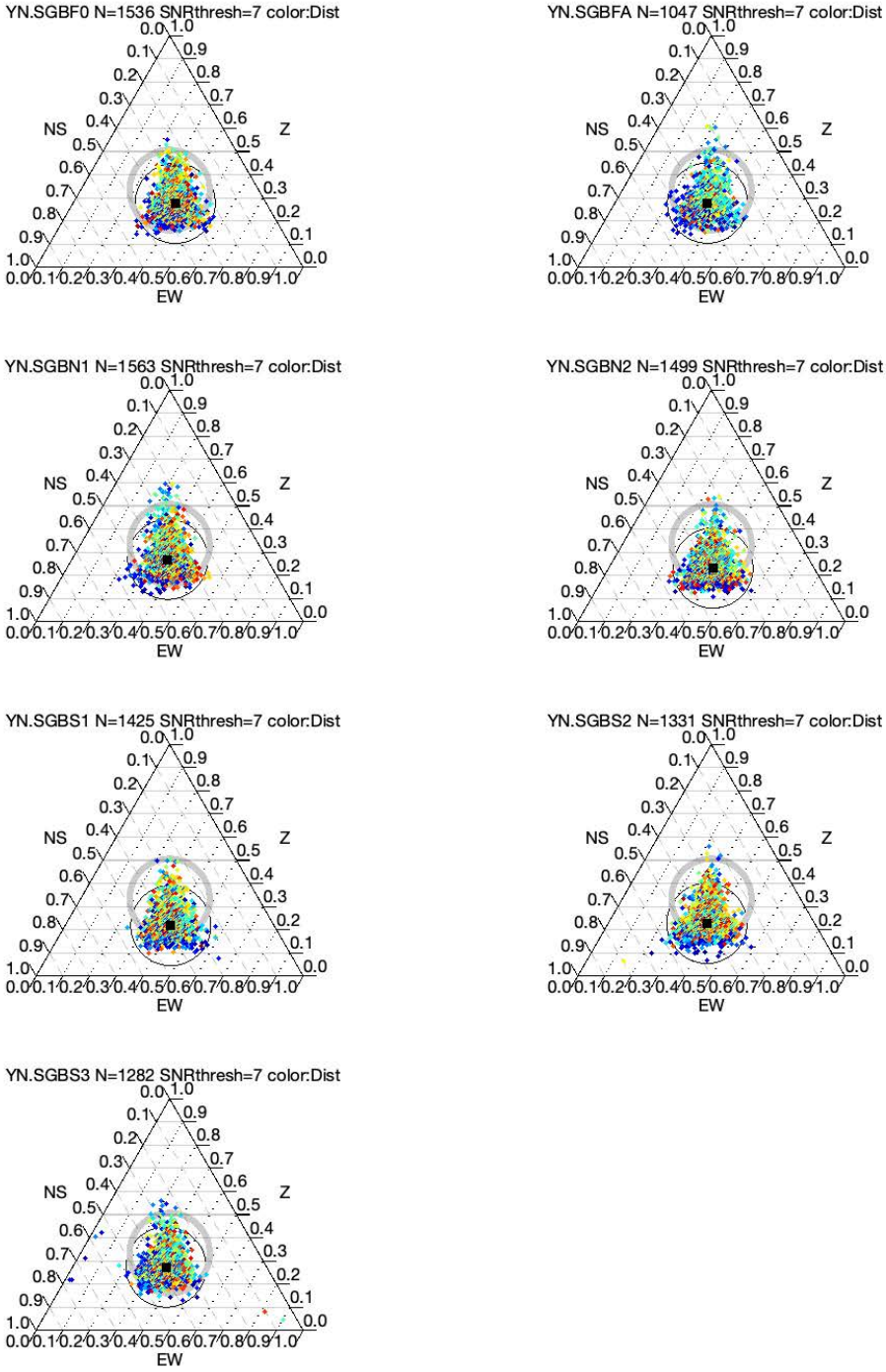

Figure B7. Ternary plots for Sagebrush Flat linear array, using a SNR threshold of seven.

Bibliography of Papers Published under DOE DE-SC0016527

Brenguier, F., P. Boué, Y. Ben-Zion, F. Vernon, C. Johnson, A. Mordret, O. Coutant, P.-E. Share, E. Beaucé, D. Hollis and T. Lecocq, 2019. Turning Vehicle Traffic into a Powerful Seismic Source for Monitoring Active Faults, Geophys., Res. Lett., 46, 9529-9536, doi:10.1029/2019GL083438.

Cheng, Yifang, Yehuda Ben-Zion, Florent Brenguier, Christopher W. Johnson, Zefeng Li, Pieter-Ewald Share, Aurélien Mordret, Pierre Boué, Frank Vernon; An Automated Method for Developing a Catalog of Small Earthquakes Using Data of a Dense Seismic Array and Nearby Stations. *Seismological Research Letters* 2020;; 91 (5): 2862–2871. doi: https://doi.org/10.1785/0220200134

Johnson C, Meng H, Vernon F, Ben-Zion Y. Characteristics of Ground Motion Generated by Wind Interaction With Trees, Structures, and Other Surface Obstacles. Journal of Geophysical Research: Solid Earth. 2019 August 14; 124(8):8519-8539. DOI: 10.1029/2018JB017151

Johnson C, Ben-Zion Y, Meng H, Vernon F. Identifying Different Classes of Seismic Noise Signals Using Unsupervised Learning. Geophysical Research Letters. 2020 August 03; 47(15):-.DOI: 10.1029/2020GL088353

Pinzon-Rincon L, Lavoué F, Mordret A, Boué P, Brenguier F, Dales P, Ben-Zion Y, **Vernon F**, Bean C, Hollis D. Humming Trains in Seismology: An Opportune Source for Probing the Shallow Crust. Seismological Research Letters. 2021 January 20; 92(2A):623-635. DOI: 10.1785/0220200248

Qin, L., Y. Ben-Zion, H. Qiu, P.-E. Share, Z. E. Ross and F. L. Vernon, 2018. Internal structure of the San Jacinto fault zone in the trifurcation area southeast of Anza, California, from data of dense seismic arrays, *Geophys. J. Int.*, 213, 98-114**,** doi: 10.1093/gji/ggx540.

Qiu, H., Y. Ben-Zion, Z.E. Ross, P.-E. Share and F. L. Vernon, 2017. Internal structure of the San Jacinto fault zone at Jackass Flat from data recorded by a dense linear array, Geophys. J. Int., 209, 1369-1388, doi: 10.1093/gji/ggx096.

Qin, L., F. L. Vernon, C. W. Johnson and Y. Ben-Zion, 2019. Spectral characteristics of daily to seasonal ground motion at the Piñon Flats Observatory from coherence of seismic data, Bull. Seism. Soc. Am., 109, 1948–1967, doi: 10.1785/0120190070.

Qin, Lei, Pieter-Ewald Share, Hongrui Qiu, Amir A Allam, Frank L Vernon, Yehuda Ben-Zion, Internal structure of the San Jacinto fault zone at the Ramona Reservation, north of Anza, California, from dense array seismic data, Geophysical Journal International, Volume 224, Issue 2, February 2021, Pages 1225–1241, https://doi.org/10.1093/gji/ggaa482

Share, P.-E., Y. Ben-Zion, Z.E. Ross, Qiu, H and F. L. Vernon, 2017. Internal structure of the San Jacinto fault zone at Blackburn Saddle from seismic data of a linear array, Geophys. J. Int., 210, 819–832, doi: 10.1093/gji/ggx191.

Share, P., Allam, A.A., Ben-Zion, Y., Lin, F.C., Vernon, F.L. (2019). Structural Properties of the San Jacinto Fault Zone at Blackburn Saddle from Seismic Data of a Dense Linear Array. Pure Appl. Geophys. 176, 1169–1191. https://doi.org/10.1007/s00024-018-1988-5

Share, Pieter-Ewald, Petr Táborík, Petra Štěpančíková, Jakub Stemberk, Thomas K Rockwell, Adam Wade, J Ramon Arrowsmith, Andrea Donnellan, Frank L Vernon, Yehuda Ben-Zion(2020). Characterizing the uppermost 100 m structure of the San Jacinto fault zone southeast of Anza, California, through joint analysis of geological, topographic, seismic and resistivity data, Geophysical Journal International, Volume 222, Issue 2, August 2020, Pages 781–794, https://doi.org/10.1093/gji/ggaa204

Sheng Y, Mordret A, Brenguier F, Boué P, Vernon F, Takeda T, Aoki Y, Taira T, Ben-Zion Y. Seeking Repeating Anthropogenic Seismic Sources: Implications for Seismic Velocity Monitoring at Fault Zones. Journal of Geophysical Research: Solid Earth. 2022 December 30; 128(1):-. Available from: https://onlinelibrary.wiley.com/doi/10.1029/2022JB024725 DOI: 10.1029/2022JB024725

Sheng, Yixiao, Aurélien Mordret, Korbinian Sager, Florent Brenguier, Pierre Boué, Baptiste Rousset, Frank Vernon, Q. Higueret, and Y. Ben-Zion. "Monitoring seismic velocity changes across the San Jacinto Fault using train-generated seismic tremors." Geophysical Research Letters 49, no. 19 (2022): e2022GL098509.

Zigone, D., Y. Ben-Zion, M. Lehujeur, M. Campillo, G. Hillers, F. L. Vernon (2019). Imaging subsurface structures in the San Jacinto fault zone with high-frequency noise recorded by dense linear arrays, Geophysical Journal International, V 217, Pages 879–893, https://doi.org/10.1093/gji/ggz069