Final report for SCEC award #19154

Title: Understanding the Seismic Ground Motion Spatial Variability Using Network Analysis Community Detection

PI: Qingkai Kong
   Affiliation: Previously at Berkeley Seismology Lab, UC Berkeley, till Dec. 2020
               Present: Lawrence Livermore National Laboratory (Jan. 2021 - )

co-PI: Gregory Beroza
Student: Yixiao Sheng

Publication:

Yixiao Sheng, Qingkai Kong, Gregory Beroza. “Network Analysis of Earthquake Ground Motion Spatial Correlation: A Case Study with the San Jacinto Seismic Nodal Array”
Geophysical Journal International, Volume 225, Issue 3, June 2021, Pages 1704-1713,
https://doi.org/10.1093/gji/ggab058

Acknowledgement:

This research was supported by the Southern California Earthquake Center
(Contribution No.19154). SCEC is funded by NSF Cooperative Agreement
EAR-1600087 and USGS Cooperative Agreement G17AC00047. This work was
performed under the auspices of the U.S. Department of Energy by Lawrence
Livermore National Laboratory under Contract Number DE-AC52-07NA27344. This is
LLNL Contribution Number LLNL-TR-832758.
Project objectives

This project is to explore ground motion spatial distribution using a new approach graph-based network analysis. In this study, we combine a large-N seismic array and graph analytics to explore spatial variability and correlation at a local scale using small local and regional earthquakes. In this method, each seismic station is modeled as a node and the similarities of the waveforms that represent ground motions between two stations are modeled as edges. By analyzing this graph network using the similarity matrices and community detection algorithm, we can group the stations spatially with similar patterns. A random forest algorithm is used to reveal the important features that affect the spatial grouping. The result suggests site conditions, and how they interact with the incident seismic wavefield, strongly condition the spatial correlation of ground motion. Future progress in characterizing ground motion spatial variability will require dense wavefield measurements, either through nodal deployments, or perhaps distributed acoustic sensing measurements of seismic wavefields.

Background information

Quantifying the spatial correlation of earthquake ground shaking is critical for estimating the seismic risk of extended infrastructure, such as bridges, utility pipelines, and highway networks (e.g. Lee & Kiremidjian 2007; Jayaram & Baker 2010; DeBock et al. 2013). It is also critical to understand the vulnerability of cities as aggregations of structures. Spatial coherence has previously been estimated through the residual of seismic intensity measures, such as spectral acceleration (e.g. Boore et al. 2003; Baker & Jayaram 2008; Jayaram & Baker 2009), acceleration spectrum intensity (Bradley 2014), Arias intensity (Foulser-Piggott & Stafford 2012), or a combination thereof (Loth & Baker 2013). The residual is defined as the difference between the observation and the prediction based on an empirical ground motion prediction equation (e.g. Chiou & Youngs 2014). Due to the lack of repeated strong-motion records at the same site, spatial correlation models are based on the assumption of stationarity and isotropy, that is, the correlation varies only with the interstation distance. With an increasing number of earthquake records, however, considering the contributions of individual site conditions in ground motion spatial correlation (GMSC) has become practical (Chen & Baker 2019; Kuehn & Abrahamson 2020; Wang et al. 2021).

Dense seismic array deployments have become increasingly common in recent years. Pioneering studies have demonstrated the feasibility of using these dense networks to extract shallow crustal structure at very fine scales (e.g. Lin et al. 2013; Nakata et al. 2015; Hillers et al. 2016; Mordret et al. 2018). Small interstation distances should lead to similar path effects from earthquakes, which will manifest in similar waveforms. It is reasonable to expect that local site conditions will also exert a similar influence on the incident seismic wavefield, such that waveform similarity, in addition to ground motion intensity measures, could be used to quantify the correlation characteristics of earthquake ground motions. Unlike traditional measures, GMSC based on waveform similarity is naturally carried out between pairs of individual sites.
Moreover, it does not require a priori empirical ground-motion models, because the influences from different factors, such as site effects, are embedded in the waveform itself.

In this study, we examine the feasibility of using a dense geophone array to study GMSC at sub-km scale lengths. We are interested in exploring the use of waveform similarity as a proxy measure of the GMSC and comparing it with the spatial correlation estimated through the waveform peak amplitude. The resolved spatial correlation pattern shows an influence from local geological structure, but with a strong dependence on the direction of the incoming seismic waves. This observation suggests that current models of path and site response may be gross simplifications of the true behavior.

Data

We choose the trifurcation region of the San Jacinto Fault Zone (SJFZ), southeast of Anza, California, as the study area. The SJFZ is considered the most seismically active component of the plate boundary system in southern California (Hauksson et al. 2012), and provides sufficient local earthquakes for our investigation. A dense seismic array, with the code ZG (Vernon & Ben-Zion 2010), was deployed across the Clark Fault segment from 2014 May 7 to June 13 (Ben-Zion et al. 2015). Studies of the local structure (e.g. Hillers et al. 2016; Qin et al. 2017; Mordret et al. 2018) provide an excellent foundation for our interpretations. Notably Qin et al. (2017) reported seismic waveform changes and local velocity contrast across the fault. They explained the phenomenon with the occurrence of large bimaterial ruptures to the northwest, causing more shallow damage on the northeast side of the fault. The seismic array is shown in Fig. 1, along with the earthquakes used in this study. The array is rectangular in shape with dimensions of 600 m × 600 m, containing 1108 vertical (10 Hz) ZLand geophones, recording with a sampling rate of 500 Hz (Ben-Zion et al. 2015). It has 55 columns and 20 rows, and the interstation spacing is approximately 10 to 30 m. The seismicity is obtained from Southern California Earthquake Data Center (SCEDC) catalog (SCEDC 2013), with a 180 km limit on the epicentral distance from the array center. We discard events that were recorded by less than 75 percent of the entire array. We select a total of 51 events with magnitudes ranging from ML 1.60 to 3.84. This data set is clearly in the range of weak ground motion, which is not ideal for studying earthquake strong ground motion; however, our goal is to examine what information can be extracted and be used to further prepare for and motivate future experiments.

We convert the data to acceleration by removing the instrument response, and band pass filter the seismograms between 2 and 12 Hz to suppress high-frequency noise. After bandpass filtering, we time the P-wave arrival with the deep learning phase picking package PhaseNet (Zhu and Beroza, 2019) on all the traces. We extract a 40-second-long time window from the processed seismograms. Each time window starts 0.5 s before the P-wave arrival, and the window length is long enough to contain both P and S signals. We perform cross-correlation for fine-tuning the alignment of the P wave and use the absolute value of the Pearson correlation coefficient as the similarity metric.
Figure 1. (a) The dense geophone array. Colour indicates the station elevation. (b) Bayesian Information Criterion (BIC) scores for earthquake classification varying with different numbers of clusters. The classification procedure is explained in Section 3. The knee point, corresponding to eight clusters, is highlighted by the red circle. (c) reference map of the study area. Each earthquake is represented by its focal mechanism or by a solid circle if the focal mechanism is not available. The sizes of the focal mechanisms correspond to the hypocentral depths. Different colors represent different groups, classified based on waveform similarity across the array (Section 3). Three black arrows point to three sample earthquakes and the corresponding seismograms are shown in Fig. 5. The black triangle in the center of the plot shows the location of the dense array and the gray reversed triangles mark some nearby broad-band stations.

For each earthquake, the assembled correlation coefficient between each pair of nodes forms a similarity matrix, in which each row or column represents how similar the waveform on one geophone is compared to the others. Note that we refer to a geophone with contaminated recordings and exclude the corresponding row and column in a similarity matrix if more than 70% of the row (column) elements have absolute value smaller than 0.1. By comparing the similarity matrices from different events, we are able to separate the earthquakes into different clusters. Clustering earthquakes provides insight into what contributes to the observed waveform-similarity structure across the array. We take the Frobenius norm of the difference between two similarity matrices as their distance and apply a K-centroids classifier. We use the Bayesian Information Criterion (BIC), which considers a trade-off between accuracy and efficiency, to determine the proper number of clusters. The knee point corresponds to 8 clusters and the classification result is shown in Figure 1.

The similarity matrices capture the intrinsic relationship among geophones in response to the seismic wavefields. Geophones with strong similarities cluster and form relatively dense groups
often called communities. Such clustering process reveals a complementary spatial correlation pattern, differing from the analysis through traditional intensity measures. The first step in our analysis is to construct a graph that is composed of a group of nodes connected by edges. In our case, the nodes are the geophones and the edges are the averaged similarity measures. We discard the edge between two nodes if the corresponding correlation coefficient is smaller than 0.4. This ensures the resulting graph is sparse, yet captures the most important similarity structure. After constructing the graph, we separate the nodes into different clusters, so that within each cluster the nodes are densely connected, i.e. they contain geophones recording similar waveforms. This process is called community detection. We use the Louvain algorithm for this task. The Louvain method is an iterative heuristic for modularity optimization. Modularity quantifies how much more densely connected the nodes within a community are, compared to how connected they would be in a random network. We choose this algorithm because it is effective, widely used, and easy to implement.

Waveform Similarity

We convert the data to acceleration by removing the instrument response, and bandpass filter the seismograms between 2 and 12 Hz to suppress high-frequency noise. Given a 500 m/s shear wave velocity near the surface (Mordret et al. 2018), the 2–12 Hz frequency band corresponds to a wavelength between 40 and 250 m. This range is smaller than the array aperture but larger than the interstation distance, giving a reasonable range for estimating the waveform similarity and studying the corresponding spatial variation. After bandpass filtering, we time the P-wave arrival with the deep learning phase picking method PhaseNet (Zhu & Beroza 2018) on all the traces. We extract a 40-s time window from the processed seismograms. Each time window starts 0.5 s before the P-wave arrival, and the window length is long enough to contain both P and S signals. We perform cross-correlation for fine-tuning the alignment of the P wave and use the absolute value of the Pearson correlation coefficient, r, as the similarity metric. For each earthquake, the assembled correlation coefficient between each pair of nodes forms a similarity matrix, in which each row or column represents how similar the waveform on one geophone is compared to the others. To exclude geophones with contaminated recordings we exclude the corresponding row and column in the similarity matrix if more than 70 per cent of the row (column) elements have absolute value smaller than 0.1. By comparing the similarity matrices from different events, we are able to separate the earthquakes into different clusters.

Clustering earthquakes provides insight into what factors influence waveform-similarity structure across the array. We take the Frobenius norm of the difference between two similarity matrices as their distance and apply a K-centroids classifier. We use the Bayesian Information Criterion (BIC, Schwarz 1978), which considers a trade-off between accuracy and efficiency, to determine the proper number of clusters. The BIC curve is shown in Fig. 1(b). We follow the method proposed in Satopaa et al. (2011) to find the knee point on the BIC curve for the optimal number of clusters. The classification result for eight clusters is shown in Fig. 1(c). Most of the earthquakes are strike-slip events so it is difficult to determine how the focal mechanism contributes to the waveform similarity. Excepting some outliers, earthquake clusters show
geographic proximity. For earthquakes very close to the array, the waveform-similarity structure is possibly dominated by the S–P traveltime; for earthquakes far from the array, the wave propagation path is similar for all geophones, and waveform similarity should largely be determined by how the seismic wavefields interact with the local structure.

![Figure 2](image)

**Figure 2.** Average similarity matrices. From left to right the result is from Group S, Group N and Group W. The black rows/columns are due to missing data or contaminated recordings. They are more numerous in Group W because there are only four events in this group.

Fig. 2 shows the average similarity matrices from three groups of earthquakes in Fig. 1: the red cluster to the south of the array (group S), the orange cluster to the north (group N) and the four events (two of them are colocated) in the Salton Trough area with no focal mechanism reported (group W). In each matrix, geophones are aligned column by column, from column 1 to 55, and inside each column starting from row A to T (see Fig. 1a). With this arrangement, geophones on the southwestern side of the fault comes first in the matrices. As expected, we observe high similarity along the diagonal of the matrices, indicating that the closer two nodes are, the more similar the seismograms will be. Compared to group N, the similarity matrices in the other two cases show larger values across a wider area. This difference shows the effect of local structure on the seismic waves from different directions. For earthquakes with large epicentral distance and greater hypocentral depth, for example events in Group W, the incident angles of the seismic waves are steep. This leads to strong P-wave motions and simpler waveforms, contributing to higher waveform similarity between geophones with large interstation distance. Though subtle, the upper left corners of the similarity matrices extend to larger distances than do the lower right corners in Group S and Group N. This suggests the seismic wavefield on the east side of the fault (larger column numbers in Fig. 1a) is more complex than the west side, reflecting different local velocity structures (Qin et al. 2017).

**Spatial correlation from similarity matrix**

The similarity matrices capture the relationship among geophones for the seismic wavefields. Geophones with strong similarities cluster and form relatively dense groups - often called communities - revealing a spatial correlation pattern. We adopt graph analytics for this analysis. The details are as follows: The first step in our analysis is to construct a graph composed of a group of nodes connected by edges. In our case, the nodes are the geophones and the edges
are the average similarity measures. We discard the edge between two nodes if the corresponding correlation coefficient is smaller than 0.4. This ensures the resulting graph is sparse, yet captures the most important similarity structure. Figs 3(a) - (c) helps visualize the graphs by projecting them onto the geophone array. For the purpose of clarity, rather than plot all the edges, we only retain those for which the correlation coefficient is higher than 0.5 and the interstation distance is between 100 and 200 m. After imposing these selection filters, the plot is not overwhelmed by the edges between nodes with very high similarity due to their close proximity. As discussed previously, in Group S and W, high waveform similarity spans a wider area.

Figure 3. Waveform similarity analysis results from three groups of earthquakes. Panel (a) shows the selected similarity edges projected onto the geophone array from earthquakes in Group S. Panels (b) and (c) show the same results from earthquakes in Group N and Group W, respectively. Panels (d), (e) and (f) present the community detection results for Group S, N and W. Each community is color-coded for clarity. Communities with less than 10 nodes are excluded for visualization purposes.

It is also reflected in the density of edges shown in Fig. 3. In all three cases, either the well-connected nodes or the isolated ones, outline regions of geological interest. For instance, on the northeastern edge of the array where the elevation is higher, the nodes are loosely connected. This area is shown to have strong anisotropy in Rayleigh wave speed revealed by the focal spot imaging, an imaging technique based on the spatial distribution of the amplitude of the zero-lag noise correlations (Hillers et al. 2016). In Group N, there is a clear gap on the west side of the fault zone. In the same area, ambient noise tomography (Mordret et al. 2018) shows a high velocity anomaly at a depth of 200 m (Fig. 5d).

After constructing the graph, we separate the nodes into different clusters, so that within each cluster the nodes are densely connected. This process is called community detection. We use the Louvain algorithm (Bruns et al. 2014) for this task. The Louvain method is an iterative
heuristic method for modularity optimization. Modularity quantifies how much more densely connected the nodes within a community are, compared to how connected they would be in a random network. We choose this algorithm because it is effective, widely used, and straightforward to implement (Bruns et al. 2014; Lu et al. 2015). We show the community detection results in Figs 3(d) - (f). The spatial pattern of the communities echoes the geological structure: the fault zone is separated from the high elevation area on two sides of the geophone array and the northern segment of the fault is detached from the southern segment. Despite the consistency among all the cases, the community size in Group N is much smaller compared to the other two groups. Note the Louvain method yields the partitions for which it is guaranteed that no communities would be merged. Therefore, the small-sized communities in Group N are consistent with its similarity matrix, which exhibits high similarity concentrating along the diagonal, as only nearby geophones record similar waveforms. Recall that the minimum correlation coefficient is 0.4 in the presented community detection analysis. We raise the minimum coefficient to 0.5, compared to Fig. 3, the general spatial patterns are preserved, however, the size of the communities are much smaller. These observations demonstrate that waveform similarity not only captures the general site response to earthquake strong motions but also reveals more subtle details imprinted by complex interactions between the wavefields and the local structure. It, however, adds complexity and poses challenges to interpret the resolved spatial pattern.

**Figure 4.** Feature importance scores for different features, including elevation and shear wave velocity at different depths in determining node communities.
We perform a feature importance analysis to gain insight into how the communities are influenced by the local structure. Feature importance analysis assigns a score to input features based on how useful they are at predicting a target variable. We treat the geophone elevation and the S-wave velocity at different depths (Mordret et al. 2018) as the input features and the node communities as the target. We fit the data with a decision tree classifier (Steinberg & Colla 2009) and obtained the feature importance score presented in Fig. 4. For all three groups, very shallow structure plays a critical role in determining the observed waveform similarity, however, the relative significance among these features vary. For example, shear wave velocity at 100 m depth (Vs100) is more important for forming the communities in Group N than in the other two groups and Vs500 has slightly more influence in Group W. These differences possibly result from the complex 3-D structure as well as the azimuth and incident angles of the seismic wavefields, as previously pointed out. High-frequency numerical simulations (Roten et al. 2016) can advance the understanding of how the 3-D structure interacts with seismic waves from varying directions.

Figure 5. Examples of seismic waveforms recorded by the geophone array. Panels (a), (b) and (c) each shows the waveforms from one earthquake sample in Group S, N and W, respectively. Magnitude, depth, back azimuth and distance of the earthquake to the array center are listed in each plot. Panel (e) zoomed in on the P-wave arrival in (c). Geophones with amplified waveforms are marked by the red arrow on the right and their locations are highlighted in (d) with red triangles. The background color in (d) shows the S-wave velocity at 200 m depth adapted from Mordret et al. (2018). The large blue triangle marks geophone R3417. Sample seismograms recorded on this station are presented in Fig. 8.
Figs 5(a) - (c) shows the recorded waveforms from three earthquakes, each of which belongs to one of the three groups discussed previously. These earthquakes are marked by arrows in Fig. 1(c). We arrange the geophone recordings in the same order as before, that is starting from column 1 to column 55 and inside each column, starting from row A to row T. The seismograms in panel (c) are distinct from the others: the P-wave energy is stronger than the S wave. As discussed previously, this is likely because this earthquake is deep and far from the array. In the same panel, there are two P phases, which perhaps represent the combination of Pg (direct) and Pn (refracted) waves and will be discussed in the next section. Panel (e) focuses on the P-wave arrivals. From station number 280 to 480, the seismograms are significantly amplified. Panel (d) highlights these nodes, which are located on the west side of the fault zone. This region overlaps with the gap of the similarity edges in Fig. 3(b) and collocates with a shallow high-velocity anomaly (Mordret et al. 2018). This emphasizes that variations of local site conditions are reflected in both seismic intensity and waveform similarity measures. It also indicates a connection between the GMSC estimated from these two types of data. In the following section, we present further analysis of the amplitude variation.

Spatial correlation of peak amplitude

The data we have to work with are for small, local events, for which the signal-to-noise ratio (SNR) is low. To compare the GMSC estimated from waveform similarity and seismic intensity, we use the peak amplitude of the filtered seismograms as the metric to quantify the scale of ground motion. Despite the fact that our data are not as broadband as we would like and our peak amplitude measurement differs from the commonly used PGA or SA, this measure represents a reasonable proxy for our analysis. Following a typical ground motion model (e.g. Atik et al. 2010), we have:

$$\ln PA_{i,j} = \mu_{\ln PA(Site_j)} + \delta B_i + \delta W_{i,j}$$

$$\epsilon_{i,j} = \delta B_i + \delta W_{i,j}$$

in which $PA_{i,j}$ is the peak amplitude measured at site j caused by the ith earthquake; $\mu_{\ln PA(Site_j)}$ represents the natural logarithmic mean of the peak amplitude intensity, estimated as the sample mean of $\ln PA$ at site j over all the earthquakes; $\delta B_i$ is between-event residual for the ith earthquake; $\delta W_{i,j}$ is the within-event residual for site j from event i and $\epsilon_{i,j}$ represents the total residual. We present $\mu_{\ln PA}$, which is approximated as the site response in Fig. 6(a). A clear high intensity area emerges on the west side of the fault zone, which matches well with our previous observations (Figs 5d and e). The southern segment of the fault shows smaller intensity than the northern segment, suggesting complex fault structure. This variability within the fault zone is also captured by the similarity analysis (Fig. 3). Both observations correlate well with the fact that fault zone trapped waves, typically appearing with strong oscillations, are only observed in the northern fault segment (Qin et al. 2017).
With the assumption that the within-event residual is a normal random variable with zero mean, we follow Chen & Baker (2019) and estimate the between-event residual as:

\[
\hat{B}_i = \frac{1}{N_s} \sum_{j=1}^{N_s} \epsilon_{i,j} = \frac{1}{N_s} \sum_{j=1}^{N_s} \ln PA_{i,j} - \mu \ln PA(Site_j),
\]

where \( \hat{\cdot} \) denotes an estimate and \( N_s \) is the total number. The within-event residual can be calculated through:

\[
\delta \hat{W}_{i,j} = \ln PA_{i,j} - \mu \ln PA(Site_j) - \hat{B}_i.
\]

**Figure 6.** Panel (a) shows the average peak amplitude in the natural logarithmic scale from all the earthquakes considered. (b), (c) and (d) show the within-event residual for the sample earthquakes in Group S, N and W, respectively.
Figure 7. Spatial correlation of peak amplitude within-event residual. From left- to right-hand side the panels show the result for Group S, Group N and Group W. The white stripes are due to missing data or contaminated recordings.

Figs 6(b) - (d) shows the within-event residual for three earthquakes discussed previously. We find spatially localized areas with large residuals in both panels (b) and (c), however, in (d), the southwest side of the array shows systematically larger residuals compared to the northeast side of the array. This pattern shows a great resemblance to the waveform-similarity structure shown in Fig. 3(c). This again supports that waveform similarity and some intensity measures carry consistent information on the GMSC. The sample earthquake in Group W shows a distinct pattern because its seismic energy is dominated by P waves at steep incidence, while the others are dominated by S waves and scattered arrivals. P waves oscillate in the direction of wave propagation while S waves oscillate perpendicular to it. These different particle motions yield different responses from the local structure, particularly for steep angles of incidence for which particle motion is better aligned with the vertical sensors for P waves than for S waves.

Fig. 7 shows the spatial correlation between the peak amplitude within-event residual from three groups of earthquakes. We calculate the residual correlation $\rho$ between site $j$ and site $k$ through the following formula:

$$
\rho(j, k) = \frac{\sum_{i=1}^{N_e} \delta W_{i,j} \delta W_{i,k}}{\sqrt{\sum_{i=1}^{N_e} \delta W_{i,j}^2} \sqrt{\sum_{i=1}^{N_e} \delta W_{i,k}^2}}
$$

in which $N_e$ represents the total number of earthquakes considered. The residual correlation for Group S and N show random variations with small correlation value, while Group W shows a blocky structure, which arises because the within-event residual has opposite signs on two sides of the fault (Fig. 6d). Fig. 8 further presents the seismograms of earthquakes in Group W recorded on a geophone in the array (marked by the blue triangle in Fig. 5d). We also compare the seismograms from the other two sample earthquakes in Group S and Group N. A secondary phase after the first P-wave arrival for all four events in Group W is observed. To support our hypothesis that this second phase is Pg and the first arrival is Pn, in Fig. 9 we plot the vertical
seismograms of the sample earthquake in Group W recorded by several broad-band stations close to the ZG array (Fig. 1). Each seismogram is bandpass filtered between 1 and 10 Hz and normalized by its peak P-wave amplitude. We also plot the predicted travel time curves for different phases assuming a 1-D model with one crust layer overlaying the mantle. The P-wave velocity is set to 6 km/s in the crust and 7.8 km/s in the mantle while the S-wave velocity is calculated as the corresponding P-wave velocity divided by $\sqrt{3}$. The depth of the earthquake is 14.2 km in the SCEDC catalog and the Moho depth is set to 25 km (Shaw et al. 2015). The predicted arrival time for each phase aligns remarkably well with the data in support of our interpretation. Note that for all earthquakes in Group W, P waves, including both Pn and Pg, show stronger energy than S waves.

![Figure 8](image.png)

Figure 8. Sample seismograms recorded on geophone R3417 (marked by the enlarged blue triangle in Fig. 5c). Seismograms from earthquakes in Group W are in black and the others in gray. The waveforms are aligned based on the first arrival and P and S waves are labeled.
Discussion

We examine the feasibility of using waveform similarity to quantify the GMSC with a dense geophone array. Our analyses suggest that the waveform-similarity structure across the array has a strong dependence on the direction of incidence of the seismic energy. We interpret this as the interaction between seismic waves and the local structure varies with the direction of incidence. Previous studies have explored similar effects, but based on the particle motion and spectral analysis measured from three-component data on individual stations. Bonamassa & Vidale (1991) found that in a certain frequency range, some sites have preferred directions of ground motion polarization and amplification, independent of earthquake focal mechanism. Spudich et al. (1996) analyzed the main shock and aftershocks of the 1994 Northridge earthquake. They confirmed the observation of Bonamassa & Vidale (1991) and associated such directional resonance with local topography. Del Gaudio & Wasowski (2007) examined small to moderate earthquakes in Italy and concluded that directional site effects are caused by the redistribution of the spectral energy modulated by both topographic and geological features. Despite the intrinsic connection, these observations differ from ours. These previous studies examine the directional preference of a given site while we focus on the spatial correlation across a wide area and how the correlation is affected by the direction of the incoming seismic energy. Further studies are necessary to gain more insights into this important issue.
We show that waveform similarity and seismic intensity are related as they sample the same seismic wavefields. Such a connection is reflected in our data set as earthquakes with steep incidence angles have not only high waveform similarity across a broader area but also peak amplitude dominated by P-wave motion as opposed to S wave. This suggests that waveform similarity and seismic intensity could either be used to describe the GMSC but from different perspectives. It may not be straightforward to compare the results from waveform similarity and seismic intensity correlation, as the former is a direct measure on the seismic observation while the latter is performed on the residual, which is designed to capture the spatial correlation omitted by the predicting model. Although a more quantitative relationship remains to be established, our results support the notion that waveform similarity is a useful alternative under circumstances where amplitude information is either difficult to obtain or unreliable, for instance, due to weak coupling or bad instrumental calibration. Other advantages of using waveform similarity include no requirement of a priori ground-motion models and less bias from comparing different wave packages, for example the peak amplitude could be estimated from either the S wave or the P wave, as shown in this study. One promising application of waveform-similarity based GMSC could be to urban regions using measurements from fiber-optic telecommunication infrastructure as seismic sensors. Using dark fiber for seismic monitoring is a rapid growing study area. Numerous studies (e.g. Lindsey et al. 2017; Li & Zhan 2018; Wang et al. 2018; Ajo-Franklin et al. 2019; Lellouch et al. 2019) have demonstrated their feasibility for recording reliable earthquake waveforms, and Spica et al. (2019) has shown that it can be used for quantifying site amplification. Such advances in data acquisition will enhance understanding of the seismic response of cities and add value to earthquake hazard analysis.

There are many limitations to our study, including the short duration of operation; however, we demonstrate the feasibility of using dense geophone arrays to investigate the spatial correlation of earthquake ground motion. The array aperture is less than 1 km, which favors the estimate of waveform similarity but limits the spatial extent. It would be interesting to extend such analysis to larger arrays. The findings in our study should be explored in other regions and for more events. We analyze 51 earthquakes and most of them distribute along the northwest-southeast direction of the array. The majority of the events are smaller than ML 2, with relatively low SNR. More earthquakes with wider azimuth coverage and moderate to large magnitude would be ideal for future analysis. An aftershock sequence following a significant main shock would provide ample data for this purpose. The vertical component is normally assumed to be less affected by local site conditions, compared to the horizontal components, which are of greater interest for seismic hazard assessment. Our results suggest the vertical component also provides valuable information, especially within topographically and geologically complex locales. The availability of multicomponent data would be extremely useful, because it would enable a complete picture of the ground motion distribution.

References


