Analysis of seismic waves crossing the Santa Clara Valley using the three-component MUSIQUE array algorithm

Manuel Hobiger, Cecile Cornou, Pierre Yves Bard, Nicolas Le Bihan, Walter Imperatori

Geophysical Journal International

Geophys. J. Int. (2016) **207,** 439–456
Advance Access publication 2016 August 1
GJI Seismology

doi: 10.1093/gji/ggw289

Analysis of seismic waves crossing the Santa Clara Valley using the three-component MUSIQUE array algorithm

Manuel Hobiger,[1,2] Cécile Cornou,[2] Pierre-Yves Bard,[2] Nicolas Le Bihan[3] and Walter Imperatori[1]

[1] *Swiss Seismological Service (SED), ETH Zurich, Zurich, Switzerland. E-mail: manuel.hobiger@sed.ethz.ch*
[2] *Institut des Sciences de la Terre (ISTerre), CNRS, IRD, IFSTTAR, Université de Grenoble, Grenoble, France*
[3] *GIPSA-Lab, CNRS, Grenoble, France*

Accepted 2016 July 28. Received 2016 July 28; in original form 2015 July 15

SUMMARY

We introduce the MUSIQUE algorithm and apply it to seismic wavefield recordings in California. The algorithm is designed to analyse seismic signals recorded by arrays of three-component seismic sensors. It is based on the MUSIC and the quaternion-MUSIC algorithms. In a first step, the MUSIC algorithm is applied in order to estimate the backazimuth and velocity of incident seismic waves and to discriminate between Love and possible Rayleigh waves. In a second step, the polarization parameters of possible Rayleigh waves are analysed using quaternion-MUSIC, distinguishing retrograde and prograde Rayleigh waves and determining their ellipticity. In this study, we apply the MUSIQUE algorithm to seismic wavefield recordings of the San Jose Dense Seismic Array. This array has been installed in 1999 in the Evergreen Basin, a sedimentary basin in the Eastern Santa Clara Valley. The analysis includes 22 regional earthquakes with epicentres between 40 and 600 km distant from the array and covering different backazimuths with respect to the array. The azimuthal distribution and the energy partition of the different surface wave types are analysed. Love waves dominate the wavefield for the vast majority of the events. For close events in the north, the wavefield is dominated by the first harmonic mode of Love waves, for farther events, the fundamental mode dominates. The energy distribution is different for earthquakes occurring northwest and southeast of the array. In both cases, the waves crossing the array are mostly arriving from the respective hemicycle. However, scattered Love waves arriving from the south can be seen for all earthquakes. Combining the information of all events, it is possible to retrieve the Love wave dispersion curves of the fundamental and the first harmonic mode. The particle motion of the fundamental mode of Rayleigh waves is retrograde and for the first harmonic mode, it is prograde. For both modes, we can also retrieve dispersion and ellipticity curves. Wave motion simulations for two earthquakes are in good agreement with the real data results and confirm the identification of the wave scattering formations to the south of the array, which generate the scattered Love waves visible for all earthquakes.

Key words: Earthquake ground motions; Surface waves and free oscillations; Site effects; Wave scattering and diffraction.

1 INTRODUCTION

Understanding the response of underground structures to earthquake excitations is an important task for the assessment of the seismic hazard. In deep sedimentary valleys the response can be complicated and also depend on the earthquake backazimuth. It can be studied either by analysing real earthquake recordings or by simulations.

In this article, we will investigate the seismological wavefield produced by regional earthquakes in the Santa Clara Basin in California both by analysing real earthquake recordings and by simulating two of the analysed earthquakes. The data of this study have been recorded between 1999 and 2003 by the San Jose Dense Seismic Array, a temporary acceleration sensor array in California. The array was located in the city of San Jose, near the eastern edge of the Santa Clara Valley, where deep sedimentary layers build the

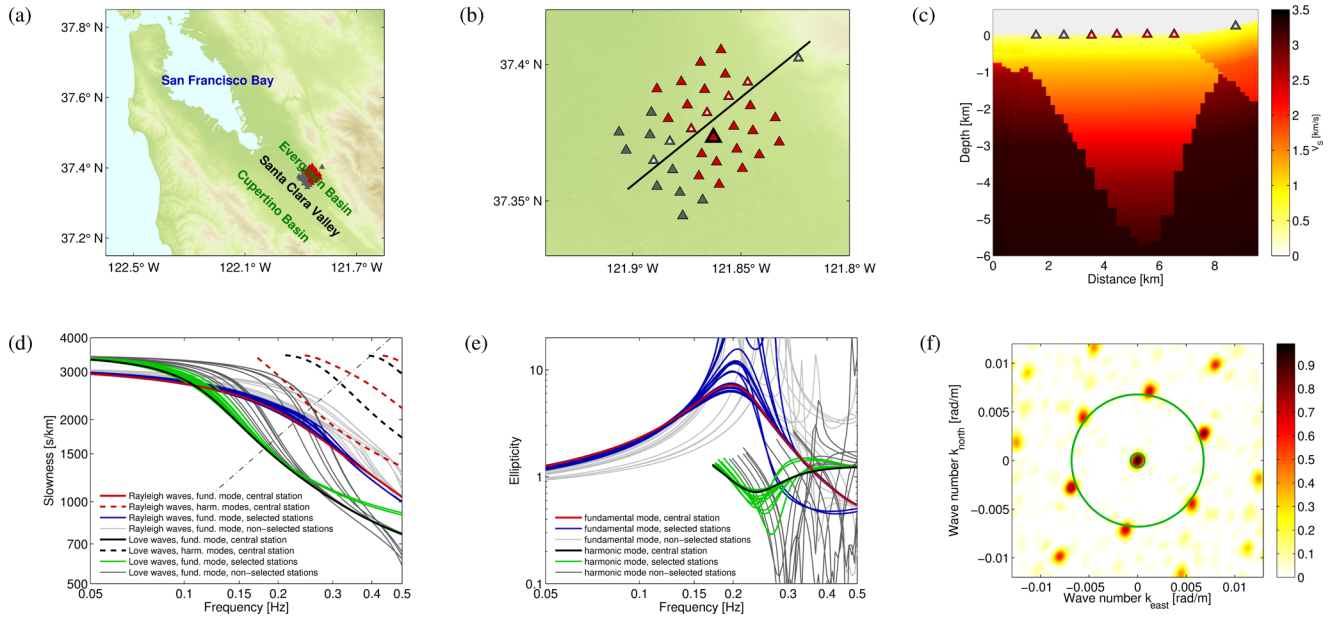

Figure 1. (a) Map of the location of the Santa Clara Valley to the southeast of San Francisco Bay. The western part of Santa Clara Valley is formed by the Cupertino Basin, the eastern part by the Evergreen Basin. The San Jose Dense Seismic Array is located in the Evergreen Basin. (b) Closer map of the array layout of the San Jose Dense Seismic Array. The red stations were used in this study, the grey stations are other stations of the array which were not used for this study. The locations for the stations with white interiors are included in the profile (c). The station with a large black border is used as a central station of the array.(c) Shear wave velocity profile along the black line indicated in (b) from southwest (left) to northeast (right). The profile corresponds to the Bay Area Model, the station symbols are indicated for comparison reasons. (d) Phase dispersion curves for Rayleigh and Love waves calculated for the different stations using the Bay Area Model. The light and dark grey curves correspond to the Rayleigh and Love wave dispersion curves of the stations not selected for the array analysis, that is, the dark grey stations in (b). The coloured lines correspond to the different stations of the array. The curves of the central station, which is marked in (b), are highlighted in different colours. For this station, also the harmonic modes are shown. The diagonal dashed line corresponds to the low-frequency resolution limit of the array. (e) Rayleigh wave ellipticity curve calculated for the Bay Area Model at the central station of the array. (f) Theoretical array response for the array used in this study. The green circles correspond to the array resolution limits of the array. The background topography of the maps in (a) and (b) is based on ASTER GDEM satellite data.

Evergreen Basin. The Evergreen Basin is separated by a subsurface ridge from the Cupertino Basin, the southwestern part of the Santa Clara Valley. Fig. 1 shows the location of the San Jose Dense Seismic Array southeast of the San Francisco Bay.

The data of the San Jose Dense Seismic Array resulted in many important scientific publications. Frankel *et al.* (2001) analysed the amplification factors in the basin and found that the amplification is strongest towards the southwestern edge of the Evergreen Basin. Furthermore, they observed the patterns of waves crossing the array. For earthquakes located east of the basin, the direct *S*-wave arrivals are followed by surface waves arriving from the south. Frankel *et al.* (2001) identified these waves as Love waves and attributed their generation to a large *S*-wave scattering structure at the southern border of the Santa Clara Valley.

Another study using the data of the San Jose Dense Seismic Array was performed by Hartzell *et al.* (2003), who investigated the site response for local and regional earthquakes and identified surface waves generated or scattered at the edges of the Santa Clara Valley and near the western edge of the Evergreen Basin. Hartzell *et al.* (2006) developed a detailed 3-D velocity model of the Santa Clara Valley and showed the importance of surface waves crossing the basin by modelling the ground motion.

In a further study, Hartzell *et al.* (2010) showed that the maximum amplification occurs at the southern edge of the basin for the horizontal motion, whereas it occurs at the deepest parts of the basin for the vertical motion. The authors of that study attribute this effect to different interactions with the basin shape for Love and Rayleigh waves or to a possibly different influence of the low shear wave velocity region at the southwestern edge of the basin on Love and Rayleigh waves.

Fletcher *et al.* (2003) analysed the data of a different array covering the whole Santa Clara Valley. They investigated the *P*- and *S*-wave arrival times from both local deep events and teleseismic events and found that the arrivals are delayed for stations located in the Evergreen or Cupertino Basins, but arrive earlier for stations located on the ridge between both basins. Furthermore, they determined that the attenuation is larger in the basins than on the ridge. Dolenc *et al.* (2005) used the same array recordings and showed that the *P*-wave arrival time delays and amplification parameters for teleseismic events are correlated. Another study by Dolenc & Dreger (2005) analysed microtremor records and indicated that the energy of the seismic noise recorded by the array is related to the wave height on the ocean and that the H/V ratio for the different sensors is stable in time.

In this study, we use the MUSIQUE algorithm (Hobiger 2011) to perform a systematic analysis of the propagation of surface waves in the basin for a set of earthquakes and investigate the correlation between surface wave type (Love, Rayleigh) and earthquake characteristics (location, magnitude). To this purpose, we selected 22 regional earthquakes in epicentral distances between 40 and 600 km and analysed the energy contributions of Love and Rayleigh waves and their azimuth directions. Combining the results of the different earthquakes, we finally obtain dispersion curves for Love and Rayleigh waves as well as Rayleigh wave ellipticity curves.

The MUSIQUE algorithm is an advanced three-component seismic array processing technique based on the classical MUSIC

algorithm (Schmidt 1986) and the quaternion-MUSIC algorithm (Miron *et al.* 2005, 2006). The combination of these two techniques allows us to estimate a larger number of parameters of the propagating waves. The azimuth and slowness, that is, the wave vector, of propagating waves are determined using the original MUSIC algorithm. Furthermore, Love and possible Rayleigh waves are identified using an energy criterion. In a second step, the polarization parameters of possible Rayleigh waves are estimated using the quaternion-MUSIC technique. These polarization parameters are the Rayleigh wave ellipticity and the retrograde or prograde sense of particle motion of the Rayleigh waves. The identification of the sense of particle motion is helpful in the separation of the fundamental and harmonic Rayleigh wave modes or in accurately determining the ellipticity peaks of the Rayleigh waves.

The MUSIC algorithm was first introduced in seismology by Goldstein & Archuleta (1987). Goldstein & Archuleta (1991a,b) used this algorithm to measure the 2-D earthquake rupture propagation during earthquakes in Taiwan ($M_L = 6.3$ in 1981 and $M_L = 7.0$ in 1986). Bokelmann & Baisch (1999) applied the MUSIC algorithm to seismic array recordings in Germany to locate the source of a narrow-band seismic signal.

In the following, we will first introduce the MUSIQUE technique and the data analysed in this study. Then we show the results of the application of the MUSIQUE algorithm to the data recordings of 22 regional earthquakes in the Santa Clara Valley. Finally, we will compare the findings with wavefield simulations for two of the analysed earthquakes.

2 THE MUSIQUE ALGORITHM

The MUSIQUE algorithm (Hobiger 2011; Hobiger *et al.* 2012) consists of two main processing steps and is performed at each frequency individually. To increase the performance of the algorithm, the signal is pre-filtered with a small-band Chebyshev filter around the desired frequency. Then, the signal is divided into time windows of several periods length (for example, three or five periods). Therefore, there are more time windows for large frequencies than for low ones. For each time window, the following processing is performed.

2.1 First step: MUSIC

The first main processing step is the original MUSIC algorithm (Schmidt 1986). It is based on the separation of signal and noise subspaces. For an array composed of N three-component seismic sensors, the data from these sensors at frequency f can be stored in three complex data vectors $\mathbf{X}_i(f)$ of size $N \times 1$, where $i = 1$ denotes the vertical, $i = 2$ the eastern and $i = 3$ the northern component. For each component, a covariance matrix $\mathbf{S}_i(f)$ of size $N \times N$ is computed by

$$\mathbf{S}_i(f) = E\left(\mathbf{X}_i(f)\mathbf{X}_i^{\dagger}(f)\right), \tag{1}$$

where E denotes the mathematical expectation and † denotes the conjugate transpose of a matrix. The mathematical expectation is realized by summing over a small number of frequencies around the central frequency (in our case 5 frequency values, that is, the central frequency, two frequencies above and two below; this corresponds to a bandwidth of less than 0.001 Hz in our case, which is much smaller than the analysed frequency range). The covariance matrices of the different components are summed to form a single covariance matrix $\mathbf{S}(f)$. The eigenvectors and eigenvalues of $\mathbf{S}(f)$ are calculated. The eigenvectors associated with the largest eigenvalues span the signal subspace and the remaining eigenvectors span the noise subspace $\mathbf{G}$. In principle, the MUSIC algorithm is capable of identifying multiple wave arrivals independently. Because of the signal projection to be performed in the second step, we only use the signal with the largest eigenvalue, that is, the most energetic wave in the signal. The propagation backazimuth ϑ and the slowness s of this dominant wave arrival are identified by maximizing the MUSIC functional P, which depends on the wave parameters of the assumed propagating wave (Schmidt 1986) and is given by

$$P(\mathbf{k}) = \frac{1}{\mathbf{a}^{\dagger}(\mathbf{k})\,\mathbf{G}\mathbf{G}^{\dagger}\,\mathbf{a}(\mathbf{k})}. \tag{2}$$

$\mathbf{a}(\mathbf{k})$ is the steering vector, that is, the vector which indicates the theoretical phase delays for the different stations of the array for the propagation of a wave with wave vector $\mathbf{k}$. The steering vector is given by

$$\mathbf{a}(\mathbf{k}) = \exp(-i\mathbf{R}\mathbf{k})/\sqrt{N}, \tag{3}$$

where $\mathbf{R}$ is the matrix of sensor positions and $\mathbf{k}$ the wave vector given by

$$\mathbf{k} = -2\pi f s(f) \cdot (\sin\vartheta, \cos\vartheta, 0)^T. \tag{4}$$

The parameters ϑ and s maximizing P are determined by a 2-D grid search.

By projecting the horizontal signals in the identified direction ϑ, the radial and transverse particle motions of the wave are separated:

$$\mathbf{X}_{\text{radial}} = -\sin\vartheta \cdot \mathbf{X}_2 - \cos\vartheta \cdot \mathbf{X}_3,$$
$$\mathbf{X}_{\text{transverse}} = \cos\vartheta \cdot \mathbf{X}_2 - \sin\vartheta \cdot \mathbf{X}_3.$$

The total energies of the vertical, radial and transverse components in the respective time window are calculated by summing the squared signals of the respective components for all sensors of the array. As we are only interested in the relative energies, we do not need to take the proportionality factor for this energy conversion into account. The identification of Love and Rayleigh waves is based on these energies. If the transverse component carries more energy than the radial and vertical components together, the wave is identified as a Love wave and the processing is stopped here. In the opposite case, it is considered as a possible Rayleigh wave and its polarization parameters are determined by quaternion-MUSIC in the following processing step.

2.2 Second step: Quaternion-MUSIC

The polarization of possible Rayleigh waves is estimated by applying the quaternion-MUSIC algorithm (Miron *et al.* 2005, 2006) to the radial and vertical components. The processing is based on quaternions, hypercomplex numbers of dimension 4 (e.g. Ward 1997). The main properties of the quaternion algebra are given in Appendix A. The advantage of using quaternions is that we can store the two complex data matrices for the radial and vertical components in a single quaternion-valued data matrix. In this way, the polarization properties of the signal remain naturally preserved in the data representation. The quaternion data matrix $\mathbf{X}_q(f)$ is formed by

$$\mathbf{X}_q(f) = \text{Re}(\mathbf{X}_{\text{vertical}}(f)) + \text{i} \cdot \text{Re}(\mathbf{X}_{\text{radial}}) + \text{j} \cdot \text{Im}(\mathbf{X}_{\text{vertical}}) + \text{k} \cdot \text{Im}(\mathbf{X}_{\text{radial}}). \tag{5}$$

The quaternion-valued covariance matrix $\mathbf{S}_q(f)$ is then built according to eq. (1). The quaternion-MUSIC functional is analogue to eq. (2), but the expression of the steering vector is changed to

$$\mathbf{a}_q = [\cos\rho + \mathrm{i}\sin\rho\exp(\mathrm{j}\varphi)]\cdot\exp(-\mathrm{j}\mathbf{Rk})/\sqrt{N}, \quad (6)$$

where φ is the phase difference between the vertical and radial components and ρ the amplitude parameter. The ellipticity is then given by $\tan\rho$.

The original quaternion-MUSIC algorithm (Miron *et al.* 2005, 2006) includes a time-consuming 4-D grid search for the four parameters: azimuth ϑ, slowness s, phase difference φ and amplitude parameter ρ. However, azimuth and slowness have already been determined in the first step, and the two remaining parameters can be determined in an analytical way (Hobiger 2011). In theory, retrograde particle motion corresponds to a phase shift of exactly $\varphi = 90^\circ$ between vertical and radial component and prograde particle motion to exactly $\varphi = 270^\circ$. In a real experiment, there will always be differences from the theoretical values. Therefore, from a practical point of view, we identify waves with a phase difference in the range $[45^\circ, 135^\circ]$ as retrograde Rayleigh waves, while waves with a phase difference in the range $[225^\circ, 315^\circ]$ are interpreted as prograde waves. The intermediate values are left unclassified.

The total energy in every time window and for each frequency is calculated as the sum of the energies of all components and all stations of the array. If the dominant wave in the time window is identified as a Love wave, the sum of the transverse energies of all stations of the array is taken as the Love wave energy in the time window. If the dominant wave is identified as a Rayleigh wave, the sum of the horizontal and vertical energies of all stations is taken as Rayleigh wave energy.

3 ARRAY SETUP, BASIN MODEL AND DATABASE

The San Jose Dense Seismic Array was deployed by the United States Geological Survey (USGS) in early 1999 and recorded until May 2004. The array consisted of 36 Kinemetrics K2 accelerometers in total, which were operated in triggered mode, that is, did not record continuously, but only if the acceleration exceeded a predefined threshold.

The original array layout is shown in Figs 1(a) and (b). The USGS has a detailed seismic velocity model of the whole San Francisco Bay Area (USGS 2014), which is based on the works of Brocher (2005) and Phelps *et al.* (2008). The shear wave velocity profile based on this model along the black line of Fig. 1(b) is shown in Fig. 1(c). Along this profile, the maximum depth of the sedimentary filling of the basin reaches down to over 5 km. The horizontal resolution of the model is between 100 m at the surface and 800 m at depth, the vertical resolution between 25 and 200 m. Based on this model, we calculated the theoretical Rayleigh and Love wave dispersion curves for the structure beneath all stations of the array. They are shown in Fig. 1(d). The southwestern stations of the array, which are located on the border of the Evergreen Basin, have significantly different theoretical dispersion curves (the grey dispersion curves, in contrast with the coloured curves for the other stations). Therefore, we did not use these stations and used the subarray highlighted in Fig. 1(b) for our analysis. This subarray consists of 25 seismic stations which are all located close to the basin centre and have a rather homogeneous structure underneath them. The dispersion curves of the harmonic modes for the central station of the array are also indicated in Fig. 1(d). We should keep in mind that these theoretical dispersion curves are calculated under the assumption of a 1-D structure, which is certainly not the case here, and that the surface waves in the frequency of interest have wavelengths of few to several kilometres. Therefore, we expect the waves to have properties which are more or less averaged over the whole structure.

The theoretical Rayleigh wave ellipticity curves for the different stations are shown in Fig. 1(e). The theoretical curves for most of the selected stations are qualitatively similar and show a peak of the fundamental mode around 0.2 Hz. The ellipticity curves for the first harmonic mode, which is only present for frequencies above 0.17 Hz, are also similar. As the theoretical ellipticity curve for the fundamental mode does not exhibit a singularity because of the weak velocity contrast, the motion of this mode is expected to be retrograde at all frequencies. As the motion of the first harmonic mode is prograde, both modes should be distinguishable by their sense of particle motion.

The theoretical array response for our array is given in Fig. 1(f). From this figure, we can determine the array resolution limits of our array as $k_{\min} = 7.3 \times 10^{-4}\,\mathrm{rad\,m^{-1}}$ and $k_{\max} = 6.8 \times 10^{-3}\,\mathrm{rad\,m^{-1}}$. The resolution limit corresponding to the minimum wave vector is also represented in Fig. 1(d). The higher resolution limit is above the limits of the plot. We can see that the Love waves should be well resolved for all frequencies above 0.15 Hz, Rayleigh waves for all frequencies above 0.2 Hz. Anyhow, as the lower MUSIQUE resolution limit might be lower than the classical limits, we will analyse the data also for lower frequencies, down to 0.05 Hz, depending on the earthquakes. Actually, only eight of the 22 earthquakes have significant energy below 0.15 Hz and only one of them below 0.1 Hz. In these low frequency ranges, we risk to get bad estimations for the wave velocities, but we expect the wave propagation azimuths still to be quite well resolved. At least for Love waves, azimuth and energy should be reasonably determined for those frequencies.

Fig. 2 shows a geologic map of the study area based on Wentworth (1997). The surface geology in the Santa Clara Valley consists of Holocene and Pleistozene alluvial deposits. The areas adjacent to the basin along the eastern and western basin borders consist mainly of materials from the Mesozoic. Directly south and southeast of the array, there are structures of sheared serpentinite and from the Franciscan Assemblage. This formation is a complex mixture of different rock types and was formed in a subduction zone (Elder 2013).

Between May 1999 and April 2006, the San Jose array recorded a total of 199 earthquakes. However, some stations of the array were subsequently moved to different locations in the Santa Clara Valley, thus reducing the number of recording stations for earthquakes occurring in the late phase of the deployment. As the different seismic sensors were triggered, the number of stations actually recording for each event also depends on magnitude and distance of the earthquakes. We selected 22 earthquakes for our study for which we have a good signal-to-noise ratio and more than 12 recording seismic sensors for the array analysis. The properties of these earthquakes are given in Table 1. Fig. 3 shows the spatial distribution of these earthquakes with respect to the array.

Example accelerograms are shown in Fig. 4. The first example is for event 5, the Hector Mine earthquake, which is the strongest event in the database and for which the recording is the longest. The frequency spectrum of this earthquake shows energy between 0.04 and 3 Hz. The second example is from event 19, one close event out of two distinct clusters of events 13–16 and 18–21. Each of these earthquake clusters is composed of four events with extremely similar parameters, that is, the same location and depth, but different

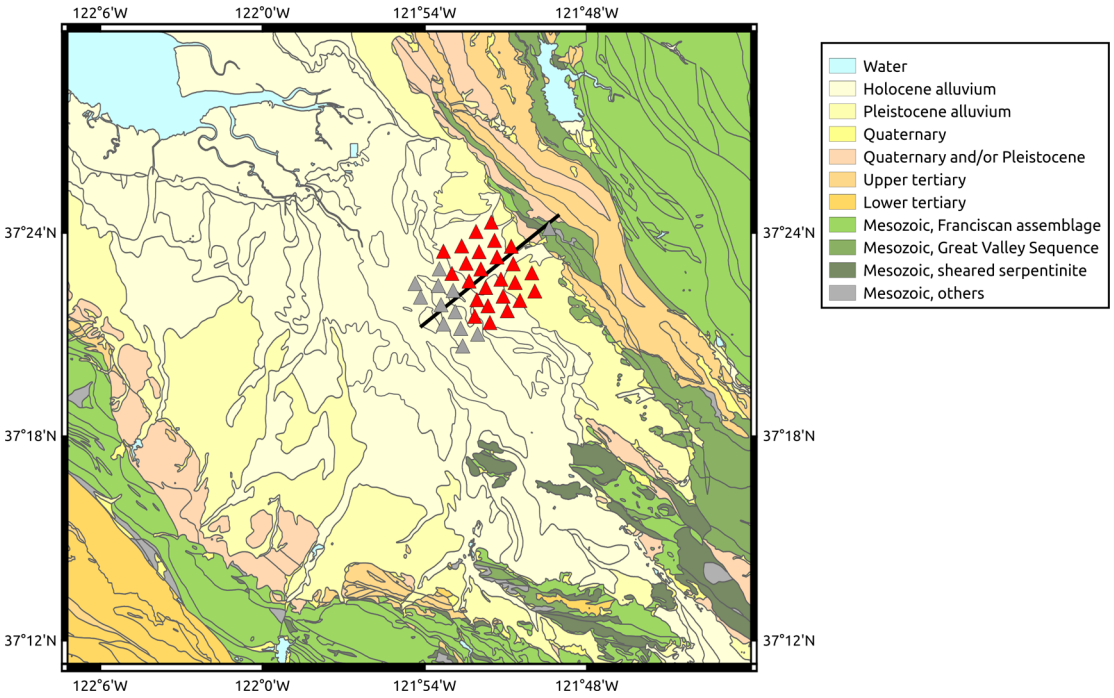

Figure 2. Map of the surface geology of the study area, based on Wentworth (1997). The red triangles mark the used array stations, the grey triangles the stations which were not used in this study. The black line indicates the location of the profile in Fig. 1(c).

Table 1. List of analysed earthquakes. For each earthquake, the origin date and time, location, depth and magnitude are indicated. Furthermore, the distances between the epicentre and the centre of the array of seismic sensors are given with the theoretical backazimuth under which the array should detect the seismic waves. Finally, the number of seismic stations contributing to the analysis of the given earthquake is indicated as well as the recorded signal length for which the indicated number of stations recorded simultaneously. We applied a high-pass filter for frequencies larger than the indicated high-pass corner frequency for all stations of the respective events. This frequency therefore also indicates the lowest analysed frequency for the respective events. The final column indicates the maximum peak ground acceleration recorded on the horizontal components of all stations.

Event ID	Date	Origin time (UTC)	Latitude [°N]	Longitude [°W]	Depth [km]	Magnitude		Distance to array [km]	Theoretical backazimuth	Number of stations	Signal length [s]	High-pass corner frequency [Hz]	Peak ground acceleration [cm s^{-2}]
1	05/15/1999	13:22	37.53	118.82	5.6	5.6	M_w	269	85°	23	436	0.1	0.79
2	05/15/1999	17:54	37.51	118.83	7.4	4.7	M_w	268	86°	23	136	0.1	0.28
3	08/01/1999	16:06	37.40	117.12	1.3	4.6	M_d	420	88°	24	540	0.1	1.00
4	08/18/1999	01:06	37.91	122.69	6.8	5.0	M_L	94	310°	19	203	0.1	2.07
5	10/16/1999	09:46	34.59	116.27	0.0	7.1	M_w	590	120°	23	1317	0.04	1.41
6	09/03/2000	08:36	38.38	122.41	10.1	5.2	M_L	122	337°	21	305	0.1	2.66
7	09/26/2000	07:20	38.66	119.53	9.2	4.2	M_L	250	54°	22	104	0.15	1.04
8	12/02/2000	15:34	39.38	120.45	14.3	4.9	M_L	255	28°	12	107	0.15	0.08
9	08/10/2001	20:19	39.81	120.62	4.0	5.5	M_L	292	21°	18	216	0.15	0.41
10	12/28/2001	21:14	36.64	121.25	6.8	4.7	M_L	98	146°	19	191	0.15	0.42
11	05/14/2002	05:00	36.97	121.60	7.2	4.9	M_L	51	153°	19	106	0.1	12.86
12	09/25/2002	07:08	36.59	121.20	8.0	3.9	M_L	105	146°	13	76	0.15	0.16
13	11/24/2002	14:54	37.76	121.95	11.0	3.9	M_L	44	350°	15	98	0.15	0.35
14	11/25/2002	03:57	37.76	121.94	10.0	3.8	M_L	44	350°	15	74	0.15	0.26
15	11/25/2002	18:22	37.76	121.94	11.0	3.5	M_L	44	350°	12	67	0.15	0.15
16	11/26/2002	12:38	37.76	121.95	10.6	3.6	M_L	44	350°	15	72	0.15	0.16
17	01/07/2003	22:29	36.81	121.39	8.9	4.3	M_w	76	146°	15	134	0.15	1.00
18	02/02/2003	16:22	37.75	121.95	16.8	3.6	M_L	42	350°	15	70	0.15	0.61
19	02/02/2003	18:22	37.74	121.94	16.7	4.2	M_L	42	350°	15	103	0.15	0.52
20	02/02/2003	18:47	37.75	121.95	16.9	4.0	M_L	42	350°	15	103	0.15	0.25
21	02/02/2003	19:02	37.75	121.95	17.1	3.5	M_L	43	350°	16	56	0.15	0.57
22	05/25/2003	07:09	38.46	122.70	5.4	4.3	M_L	141	329°	12	145	0.1	0.24

magnitudes. The signal recording for event 19 is much shorter than for event 5, and the energy content is shifted to higher frequencies. For this event, there is significant energy in the frequency range between 0.2 and 15 Hz.

4 DATA PROCESSING

The signals recorded for each of the events shown in Table 1 have been analysed in the following way: The signals of those of the 25 stations which recorded the event (i.e. for which the signal was

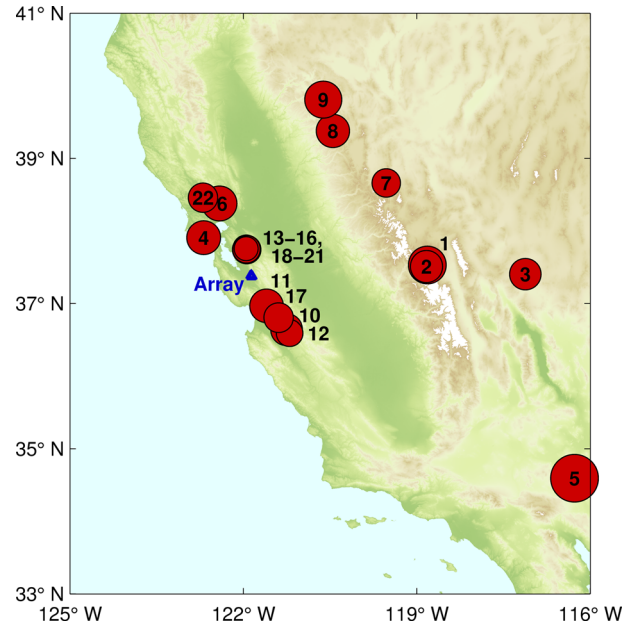

Figure 3. Map showing the epicentres of the analysed earthquakes with respect to the location of the San Jose Dense Seismic Array. The diameter of the circles is proportional to the magnitude of the respective earthquakes. The background topography is based on ASTER GDEM satellite data.

strong enough to trigger the recording) form the array recording (see Table 1 for the number of array stations which recorded the respective event). As the signals are accelerogram recordings, we first transformed the signals into velocity space. In order to do so, the first step was to high-pass filter the signals. Event 5 showed significant energy above 0.04 Hz, the other events only showed significant energies above 0.1 or 0.15 Hz. These frequencies were taken as corner frequencies for the high-pass filter (Table 1). After this filtering, the signals were integrated to calculate the velocity values. The velocity signals of events 5 and 19 are also shown in Fig. 4.

The next step was then to analyse the signals with MUSIQUE for 50 different frequency values between 0.05 and 0.5 Hz, using a logarithmic frequency scale, restricting the analysis for the respective earthquakes to frequencies above the respective high-pass corner frequencies (Table 1).

5 RESULTS FOR THE EARTHQUAKE RECORDINGS

5.1 Azimuthal energy distribution

For all events, we analysed the wavefield between the respective high-pass corner frequency and 0.5 Hz and classified the results in six different frequency ranges (0.05–0.1 Hz, 0.1–0.15 Hz, 0.15–0.2 Hz, 0.2–0.3 Hz, 0.3–0.4 Hz, 0.4–0.5 Hz). In the following, we will show exemplary results for event 4 (the Bolinas earthquake), event 5 (the Hector Mine earthquake), event 11 (the Gilroy earthquake) and the clusters of events 13–16 and 18–21. The results for the other events are given in the Supporting Information.

5.1.1 Event 4

Event 4 (Fig. 5), the Bolinas earthquake, occurred in the extension of the Santa Clara Valley to the northwest. The polar plots in Fig. 5 show how much energy arrives from which backazimuths for the different wave types. The plots for the different frequency ranges are normalized to their respective maxima. The absolute proportion of energy for the different frequency ranges is given by the top right plot in Fig. 5. The bottom right plot of Fig. 5 shows the energy proportion of the identified waves. In all frequency ranges, around 40 per cent or more of the total signal energy could be attributed to the different wave types. The remaining energy is carried by unclassified waves.

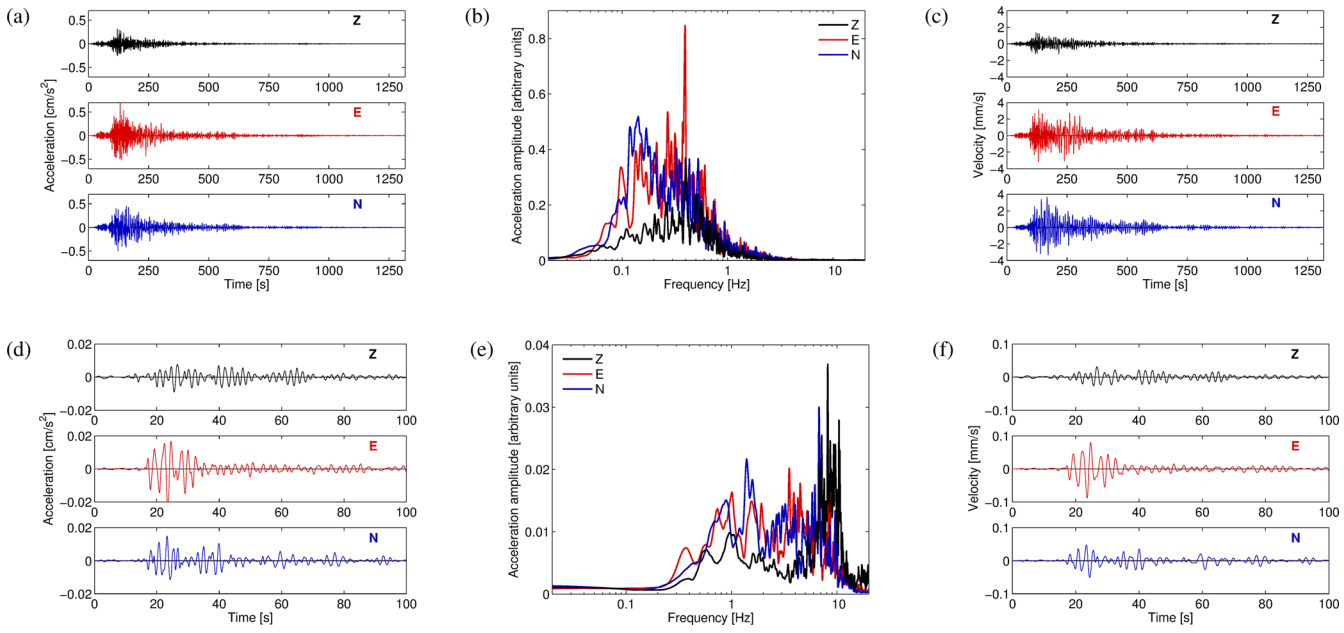

Figure 4. (a) Accelerogram for the three components of the central station for event 5, the Hector Mine earthquake, filtered between 0.05 and 0.5 Hz. (b) The respective spectra of the signals of (a). (c) Velocity signal of the central station for event 5, obtained by integration of the accelerogram of (a). The plotted signal has been filtered between 0.05 and 0.5 Hz. (d–f) The same as (a–c), but for event 19.

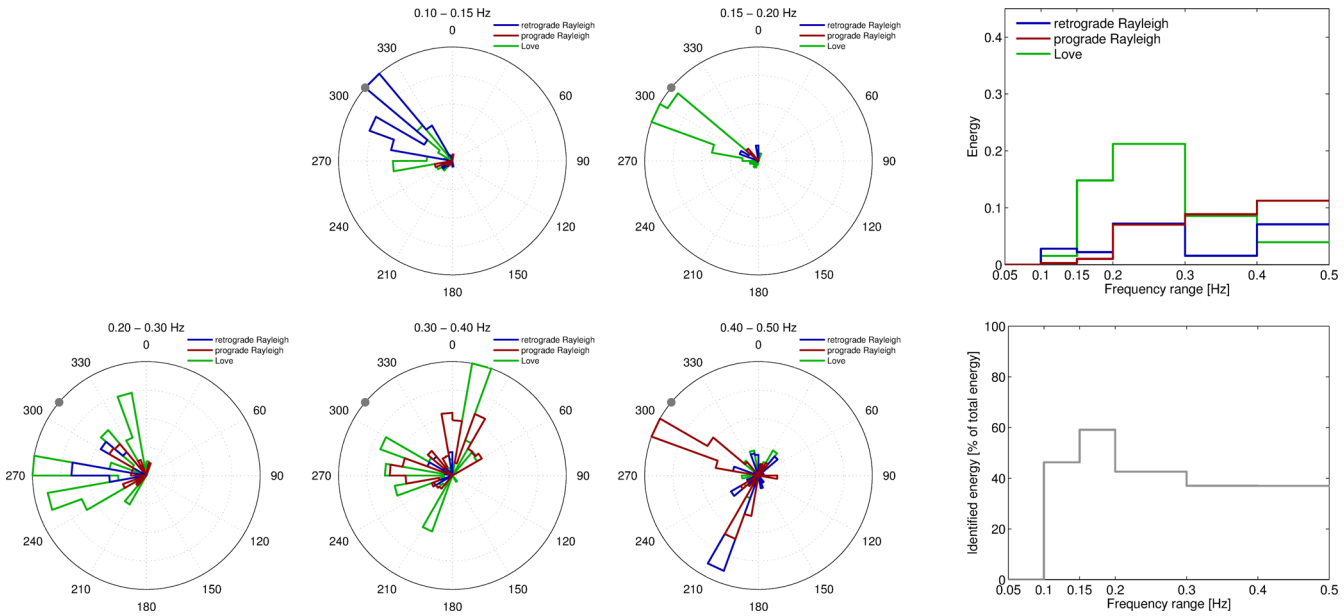

Figure 5. The polar plots show the azimuthal energy distribution in different frequency ranges for event 4, the Bolinas earthquake. The grey dot indicates the theoretical backazimuth of the event. For clarity reasons, the energies for Rayleigh and Love waves are normalized by the overall maximum in each frequency range. Top right: relative energy distribution of the wave types in the different frequency bands. The sum of the relative energies of all wave types and all frequency ranges equals 1. Bottom right: percentage of the total energy in the respective frequency range carried by identified Love and Rayleigh waves, i.e., the proportion of wave energy characterized in the respective frequency range.

Below 0.15 Hz, all waves come from the theoretical backazimuth and the largest contribution consists in retrograde Rayleigh waves. Love waves dominate the signal between 0.15 and 0.3 Hz. Between 0.15 and 0.2 Hz, they come from the direct backazimuth. Between 0.2 and 0.3 Hz, they arrive from a multitude of backazimuths, mainly diffracted by around 40° to the north and west from the theoretical backazimuth. These directions most probably correspond to waves reflected at the basin borders before detection. Between 0.3 and 0.4 Hz, waves arrive under all azimuths. Prograde Rayleigh waves and Love waves contribute equal amounts of energy in this frequency range. The largest single energy contribution comes from Love waves from the north. In the highest analysed frequency range, Rayleigh waves dominate the signal. The main arrivals are prograde Rayleigh waves from the theoretical backazimuth, but there are also significant retrograde Rayleigh waves from a southern direction. In any case, it is striking that all identified wave arrivals are from the northwestern hemicycle (arrivals between 200° and 20°), the wave energy arriving from the other hemicycle is negligible.

5.1.2 Event 5

Event 5 (Fig. 6), the Hector Mine earthquake, is the largest earthquake in the data set. It had a magnitude of 7.1. For this event, Love waves dominate the wavefield for each frequency range. In the three lowest analysed frequency ranges below 0.2 Hz, these waves come mainly from the theoretical backazimuth. At higher frequencies, they are more diffracted at the basin borders. Anyhow, for all frequency ranges, significant Love waves arriving from the south can be detected, but they carry less energy than the direct waves. The Rayleigh waves, which are of much lower energy than the Love waves, arrive mainly from the direct backazimuth. For this event, which is located on the southeast, all waves arrive from the southeastern hemicycle (wave arrivals between 60° and 200°), there is no significant energy from the opposite hemicycle.

An analysis of the temporal evolution of the backazimuth of Love waves can give interesting insights on the wave scatterers in the Evergreen Basin. As the Love wave energy for this event is very low above 0.3 Hz, we restrained this analysis to the four frequency ranges below this frequency.

Fig. 7 shows the backazimuth and energy of all waves identified as Love waves, with the radius axis of the polar plots corresponding to time. In every frequency range, the early signal with the most energetic wave arrivals comes from the theoretical backazimuth. In this early signal, there is no scattered wave energy arriving from other backazimuths.

In the lowest frequency range, Love waves arrive mainly from the theoretical backazimuth and the south during the first 500 seconds. Only after that time, waves start to arrive from different directions, including some from the northwestern quadrant.

In the higher frequency ranges, that is, above 0.1 Hz, scattered waves start to arrive earlier. Between 0.1 and 0.15 Hz, the wave arrivals are limited to the southeastern hemicycle for the first 500 s, some waves also arrive from the west after that time. However, there is a gap in the north from where no waves arrive. Between 0.15 and 0.2 Hz, the pattern is similar. The first waves are limited to the southeastern hemicycle in the first 500 s. Later, waves arrive from all different directions.

For the last frequency range, between 0.2 and 0.3 Hz, the pattern looks different. Only the very early signal is restricted to the southeastern hemicycle, but already after 250 s waves arriving from the north are observed. From that time on, waves from very different backazimuths arrive, but there is a gap in the southwest, from where no waves arrive at all.

For all frequencies, we can observe that the main wave energy arrives from the theoretical backazimuth of the earthquake and from the south. Scattered waves from different directions only arrive later in time, but their energy is much smaller than the energy of the early signal.

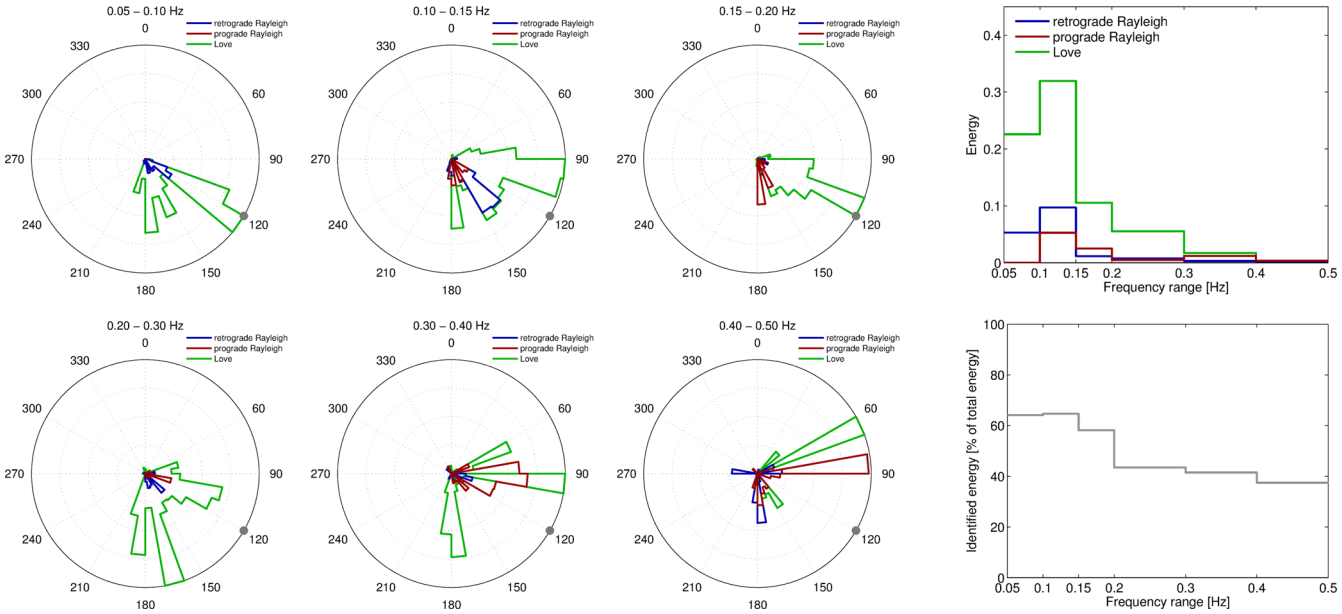

Figure 6. The same as Fig. 5, but for event 5, the Hector Mine earthquake.

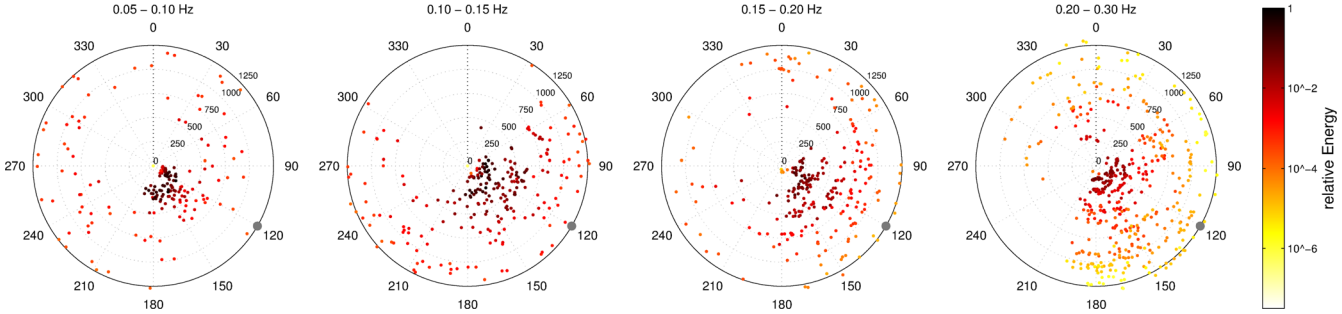

Figure 7. Evolution of the backazimuth of Love waves with time for different frequency ranges for event 5, the Hector Mine earthquake. The backazimuth for all waves identified as Love waves is shown. The radius of the polar plot is the time from the beginning of the recording in seconds. The grey dot indicates the theoretical backazimuth for the Hector Mine earthquake. The point colour indicates the energy of the respective time window on a relative logarithmic scale.

5.1.3 Event 11

Event 11, the Gilroy earthquake, occurred in the southeastern direction, along the axis of the basin. The energy of waves below 0.2 Hz is negligible compared with the higher frequencies (Fig. 8). Between 0.2 and 0.3 Hz, the wavefield is dominated by Love waves, which come mainly from the direct southern direction and from southeastern directions, but only to a very small part from the direct backazimuth. At higher frequencies, Rayleigh waves carry more energy than Love waves, but they are also deviated from the direct backazimuth. Above 0.4 Hz, both retrograde and prograde Rayleigh waves arrive mainly from the east.

5.1.4 Events 13–16 and 18–21

The results for the two clusters of events 13–16 and 18–21 are shown in Fig. 9. These events have very similar epicentres and times of occurrence, suggesting that each cluster occurred on the same fault segment. As all these events show very similar results, we grouped them together. The most energetic component of these earthquakes are Love waves in the frequency range between 0.3 and 0.4 Hz. Those waves arrive, like the Love waves at the other frequencies, from the theoretical backazimuth. Only above 0.4 Hz, the Love waves are deviated about 30° to the east. The main contribution of Rayleigh waves are prograde waves above 0.4 Hz. One part of these waves comes from the theoretical backazimuth, the other part are diffracted waves under different directions. Also for these events, which are located in the north, all significant energy arrivals are restricted to northern directions (arrivals between 270° and 40°).

5.1.5 Summary

For all other events, the figures corresponding to Figs 5, 6, 8 and 9 can be found in the Supporting Information. The events discussed in detail here are characteristic for all events.

The events can be divided into two principal groups, events arriving from the north (backazimuths between 310° and 28°) and events arriving from the southeast (backazimuths between 85° and 153°). For events arriving from the north (i.e. events 4, 6, 8 and 9, 13–16, 18–22), the wave energy is confined to the northwestern hemicycle and no significant energy is arriving from southeastern directions. For this group of events, most of the wave energy is arriving from the theoretical backazimuth of the respective earthquake, but for some of the events, however, Love wave arrivals from the south can be seen. Furthermore, especially for earthquakes occurring in the prolongation of the Santa Clara Valley, a lot of waves are reflected on the basin borders and arrive therefore from directions deviated from the direct backazimuth (e.g. events 6 and 22).

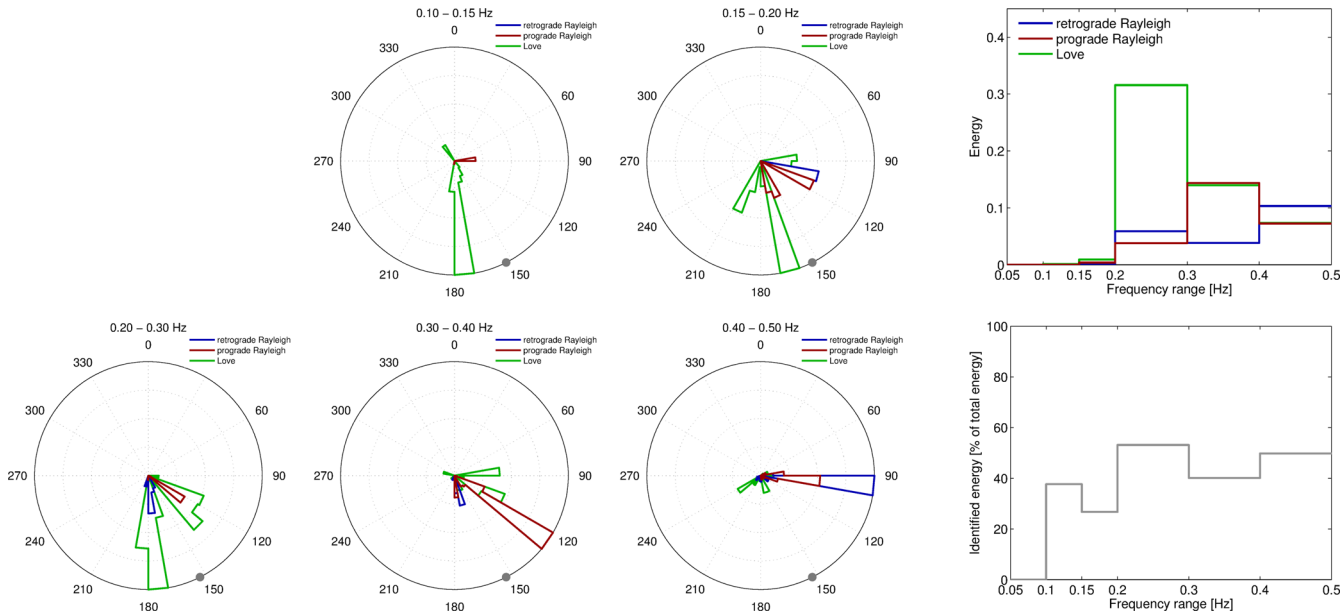

Figure 8. The same as Figs 5 and 6, but for event 11, the Gilroy earthquake.

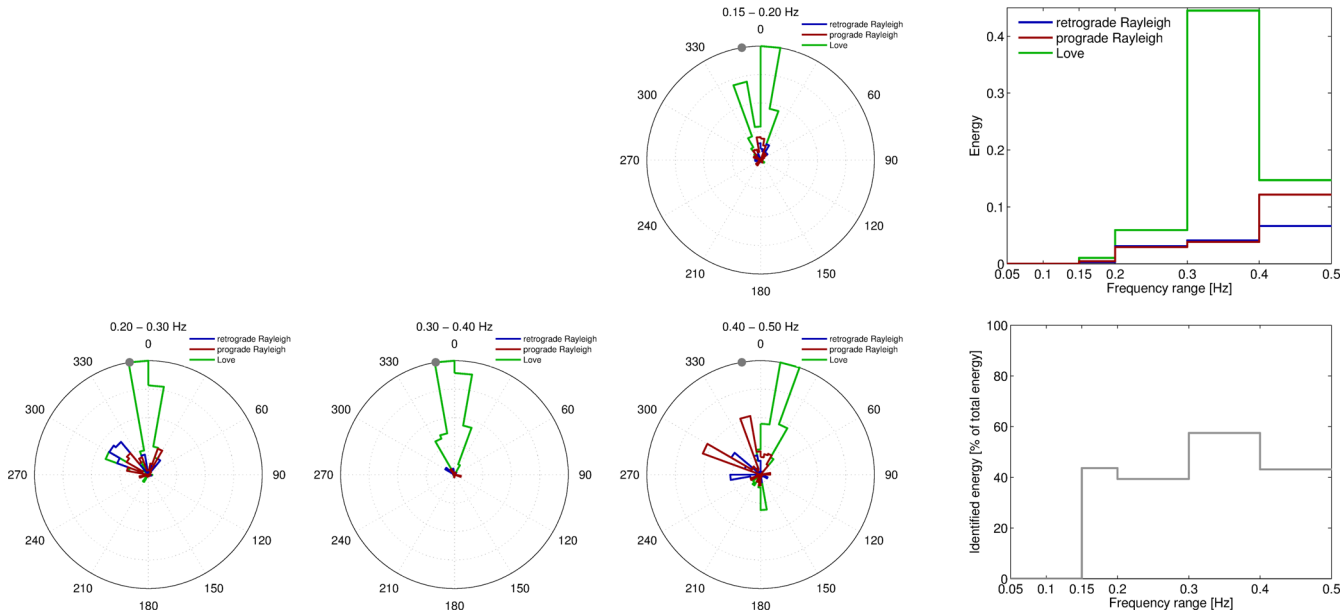

Figure 9. The same as Figs 5, 6 and 8, but for the combination of the events 13–16 and 18–21.

Event 7 is a special case because its theoretical backazimuth is perpendicular to the eastern basin border. For this event, the largest part of energy arrives from that theoretical backazimuth. However, in the frequency range between 0.3 and 0.4 Hz, the entire Love wave energy comes from the south.

For events arriving from the southeast (i.e. events 1, 2, 3, 5, 10, 11, 12 and 17), the main wave arrivals are only slightly deviated from the theoretical azimuth. For events occurring in the prolongation of the valley in the southeastern direction, waves reflected on the inner basin borders can be seen (e.g. events 10, 11, 12, 17). For all events, Love waves arriving from the south can be seen. The maximum contribution of these southern Love waves is obtained between 0.2 and 0.4 Hz (e.g. events 1, 2 and 11).

For most of the events (exceptions are events 8, 9, 12 and 17), the wavefield is dominated by Love waves. The energy repartition between the different wave types will be discussed later in this paper.

5.2 Identification of wave scatterer directions

Fig. 10 shows a combination of all events, where all wave arrivals within $\pm 45^{\circ}$ from the theoretical backazimuth have been removed. In this way, it is possible to determine the main directions of arrival of scattered waves. For frequencies below 0.3 Hz, almost all scattered Love waves arrive from the south. It is only above 0.3 Hz that some scattered Love waves also arrive from the eastern basin

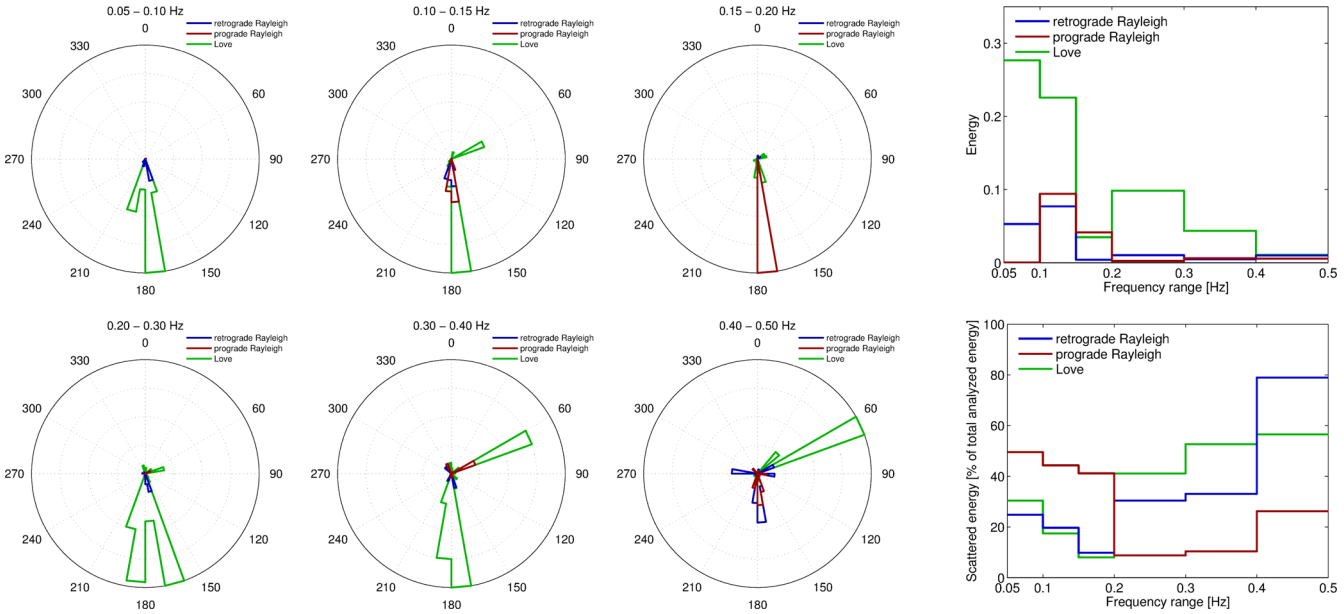

Figure 10. The polar plots show the azimuthal energy distribution of the scattered waves in different frequency ranges for the combination of all 22 earthquakes. All waves with backazimuths within ±45° from the theoretical backazimuth for the respective events are excluded so that only scattered waves are included. For clarity reasons, the energies for Rayleigh and Love waves are normalized by the overall maximum in each frequency range. Top right: relative energy distribution of the wave types in the different frequency bands. The sum of the relative energies of all wave types and all frequency ranges equals 1. Bottom right: percentage of the total energy in the respective frequency range carried by identified Love and Rayleigh waves, i.e., the proportion of wave energy characterized in the respective frequency range.

border. This result is in good agreement with the evidence for a large Love wave scatterer that Frankel *et al.* (2001) identified in the south. If we have another look on the geologic map of the area, it is easy to find possible candidates for the wave scatterer in the south. According to the Bay Area Model dispersion curve, the wavelength of Love waves should be around 10 km at 0.2 Hz and about 1.5 km at 0.5 Hz. The wave scattering structures must have a similar size. Directly to the south of the array is a prominent hill of sheared serpentinite of about three kilometres length at the surface. If we take into account that the buried structure is probably larger, this could be a good candidate for the wave scatterer. Directly east of this hill is another, even larger structure of the same material. For waves scattered at this point, however, we would expect arrival directions to be more southeastern than observed. Directly south of the first hill is the southern border of the valley, which is formed by material of the Franciscan assemblage, which is potentially a very good wave scatterer. One or several of these structures certainly act as wave scatterers in the Santa Clara Valley.

Most of the scattered wave energy is carried by Love waves in almost all frequency ranges, especially below 0.15 Hz and between 0.2 and 0.4 Hz. Significant scattered Rayleigh waves are identified below 0.2 Hz. In those frequency ranges, over 40 per cent of all identified prograde Rayleigh waves were scattered. Anyhow, according to the basin model (see Fig. 1), prograde Rayleigh waves should not exist below 0.18 Hz. As the energy of identified prograde Rayleigh waves is very small below 0.15 Hz for all earthquakes, it is not surprising that small amounts of waves misidentified as prograde Rayleigh waves can have a virtually large contribution of scattered waves below 0.15 Hz. In any case, the scattering directions for prograde Rayleigh waves are very similar to retrograde Rayleigh waves and to Love waves in all frequency ranges below 0.2 Hz. Above 0.2 Hz, less than 25 per cent of all prograde Rayleigh waves are scattered. The partition of scattered retrograde Rayleigh waves in those frequency ranges is much larger, reaching up to 80 per cent above 0.4 Hz. The majority of these waves also arrives from southern directions.

5.3 Energy repartition between Love and Rayleigh waves

The energy repartition between the different surface wave types is given for each event in Table 2. The indicated energies are the total energies for all 50 analysed frequencies in the range between 0.05 and 0.50 Hz. In the MUSIQUE code, the signal of a time window is identified as a Love wave if the transverse energy is dominant in the respective time window. The Love wave energy indicated in the table corresponds to the summed transverse energies of all time windows which have been attributed to Love waves. The radial and vertical energies of these windows are also included in the table. These latter energies correspond to the less energetic waves in the signal, which can include Love or Rayleigh waves arriving under different azimuths. The energy of these ‘wrong’ components is, by construction of the method, always smaller than the energy of the components corresponding to the identified wave. The columns for the retrograde and prograde Rayleigh waves are analogue to the Love wave columns. However, it should be mentioned that the MUSIQUE code does not directly identify a wave as a Rayleigh wave if the sum of the radial and vertical energies is larger than the transverse energy. It is also necessary that the polarization angle is within certain limits (between 45° and 135° for retrograde and between 225° and 315° for prograde Rayleigh waves). For all other polarization angles, the wave is left unclassified.

The table shows that the energy of the classified components represents between 39.1 and 60.1 per cent of the total energy for the different events, with an average of 47 per cent. Overall, the wavefield is dominated by Love waves for most of the events. Retrograde Rayleigh waves do not contribute much in terms of energy, and prograde Rayleigh waves are more pronounced.

Table 2. List of the energy partition for all analysed earthquakes, sorted by increasing azimuths starting in the northwest. The columns indicate the transverse and combined radial and vertical energies of all time windows identified as Love waves and retrograde or prograde Rayleigh waves, plus the energy attributed to time windows which have not been classified. The next column indicates the sum of the transverse energy of windows identified as Love waves and the radial and vertical energies of the windows identified as Rayleigh waves. The values are indicated as percentages of the total energy content in the frequency range between 0.05 and 0.50 Hz. The next columns indicates the energy repartition between Love, retrograde and prograde Rayleigh wave energy in the classified time windows. In the last column, the energy proportion of Love waves arriving from southern directions (with azimuths between 150° and 210°) compared to the total Love wave energy is indicated.

		Energies (in per cent of the total energy) of the windows identified as								Classified components [per cent]			
Theoretical backazimuth [°]	Event ID	Love waves E_t	Love waves E_r+E_v	Retrograde Rayleigh waves E_r+E_v	Retrograde Rayleigh waves E_t	Prograde Rayleigh waves E_r+E_v	Prograde Rayleigh waves E_t	Unclassified $E_t+E_r+E_v$	Sum of all classified components	Love waves	Retrograde Rayleigh waves	Prograde Rayleigh waves	Percentage of Love waves arriving from the south
310	4	21.1	14.7	9.0	5.9	12.1	7.8	29.5	42.2	50.0	21.3	28.7	3.9
329	22	40.1	23.0	6.6	3.9	5.6	3.4	17.3	52.3	76.7	12.6	10.7	2.7
337	6	36.4	21.5	6.6	4.3	6.6	4.0	20.6	49.6	73.4	13.3	13.3	0.6
350	13	32.5	22.3	5.8	3.6	10.9	7.3	17.6	49.2	66.1	11.8	22.2	3.0
350	14	23.2	12.1	10.1	4.9	13.9	8.3	27.4	47.2	49.2	21.4	29.4	0.0
350	15	29.2	14.9	8.2	4.6	9.8	5.6	27.7	47.2	61.9	17.4	20.8	0.9
350	16	28.7	16.5	6.7	4.1	9.3	6.1	28.6	44.7	64.2	15.0	20.8	0.0
350	18	29.8	18.2	11.3	6.6	9.0	6.1	19.0	50.1	59.5	22.6	18.0	3.6
350	19	36.6	20.2	6.6	3.8	8.0	4.5	20.2	51.2	71.5	12.9	15.6	6.3
350	20	30.0	18.3	6.6	3.5	10.3	5.8	25.5	46.9	64.0	14.1	22.0	0.7
350	21	31.5	18.5	9.7	6.0	7.1	4.4	22.8	48.3	65.2	20.1	14.7	3.4
21	9	18.3	12.8	10.7	6.1	12.7	7.5	31.9	41.7	43.9	25.7	30.5	4.0
28	8	8.6	6.6	5.2	3.3	26.2	11.4	38.8	40.0	21.5	13.0	65.5	20.6
54	7	21.8	14.5	2.8	2.0	14.8	8.1	36.0	39.4	55.3	7.1	37.6	38.3
85	1	43.3	20.8	6.4	3.3	6.7	3.7	15.8	56.4	76.8	11.3	11.9	20.6
86	2	24.0	15.1	7.5	4.8	14.2	7.6	26.8	45.7	52.5	16.4	31.1	14.7
88	3	27.6	18.2	3.6	2.7	20.6	10.1	17.2	51.8	53.3	6.9	39.8	27.0
120	5	43.7	19.1	10.5	4.4	5.9	2.2	14.2	60.1	72.7	17.5	9.8	30.1
146	10	20.6	14.9	9.7	5.1	11.7	7.1	30.9	42.0	49.0	23.1	27.9	21.1
146	12	14.6	10.8	7.6	4.7	16.9	11.7	33.8	39.1	37.3	19.4	43.2	11.3
146	17	16.4	12.3	6.3	3.8	22.9	13.4	24.8	45.6	36.0	13.8	50.2	40.1
153	11	25.2	17.2	9.5	5.9	12.1	7.9	22.2	46.8	53.8	20.3	25.9	42.3

The energy repartition between the different wave types varies largely for the different events. The proportion of Love waves ranges from 8.6 per cent (event 8) to 43.7 per cent (event 5), with an average of 27.4 per cent. The retrograde Rayleigh wave proportion is much smaller, taking values from 2.8 to 11.3 per cent for the different events, whereas the prograde Rayleigh wave proportion is larger on average and accounts for 5.6–26.2 per cent of the energy. The unclassified waves contribute between 14.2 and 38.8 per cent to the total energy, but, in fact, also the contributions of the components not corresponding to the identified polarization of each time window are not attributed to a wave type. The energy repartition between classified Love and Rayleigh waves (retrograde and prograde) is indicated in the table as well.

The last column of the table indicates the proportion of Love wave energy arriving from the south (with azimuths between 150° and 210°). For events located north of the array (with azimuths smaller than 21°), this proportion is very small (less than 6.3 per cent). For the other events, which are located east of the array, the proportion of Love waves from the south is much larger and ranges from 11.3 per cent (event 12) to 42.3 per cent (event 11).

In Fig. 11, the energy repartition between Love, retrograde and prograde Rayleigh waves is shown in different frequency ranges for all analysed events. In each frequency range, the sum of the Love, retrograde and prograde Rayleigh waves is normalized to 1. A relative peak of the Love wave energy can be seen for frequencies below 0.20 Hz for most events. For event 5, the strongest, but also farthest event in the data set, the Love wave energy dominates in the whole frequency range. The two groups of events 13–16 and 18–21 both show very energetic Love waves between 0.3 and 0.4 Hz. With distances of 44 km and 42 km, respectively, these events are the closest to the array.

The energy proportion of retrograde Rayleigh waves is rather small for most events. Only some events exhibit significant contributions of retrograde Rayleigh waves (e.g. events 4, 14, 15, 17 and 18).

Prograde Rayleigh waves are stronger at higher frequencies for most events. According to the basin model, the fundamental Rayleigh wave mode is retrograde at all frequencies. Therefore, prograde Rayleigh waves correspond to the first harmonic mode which does not exist below 0.18 Hz. Consequently, prograde Rayleigh waves found at lower frequencies are probably misidentified.

5.4 Dispersion curves

As the different earthquake signals carry wave energy in different frequency ranges, they can also give us insight on different parts of the Love and Rayleigh wave dispersion curves. The dispersion curves are retrieved in the following way: For each frequency, a velocity histogram with 50 different velocity values is calculated, where only waves identified as the respective wave type are considered and their respective energies are summed at each histogram point. The histograms for the different frequencies are then normalized individually and plotted one next to the other. In the resulting plot, the dispersion curves can be picked in a reasonable way.

We will not show the dispersion curves for all events, but limit this study to showing some examples in Fig. 12. The Love wave dispersion curve for event 4 exhibits some clear data points which are in good agreement with the theoretical dispersion curve of the Bay Area Model at least for frequencies above 0.25 Hz. Below 0.2 Hz, the velocity values are outside the array resolution limits but still reasonable with respect to the Bay Area Model, even if the

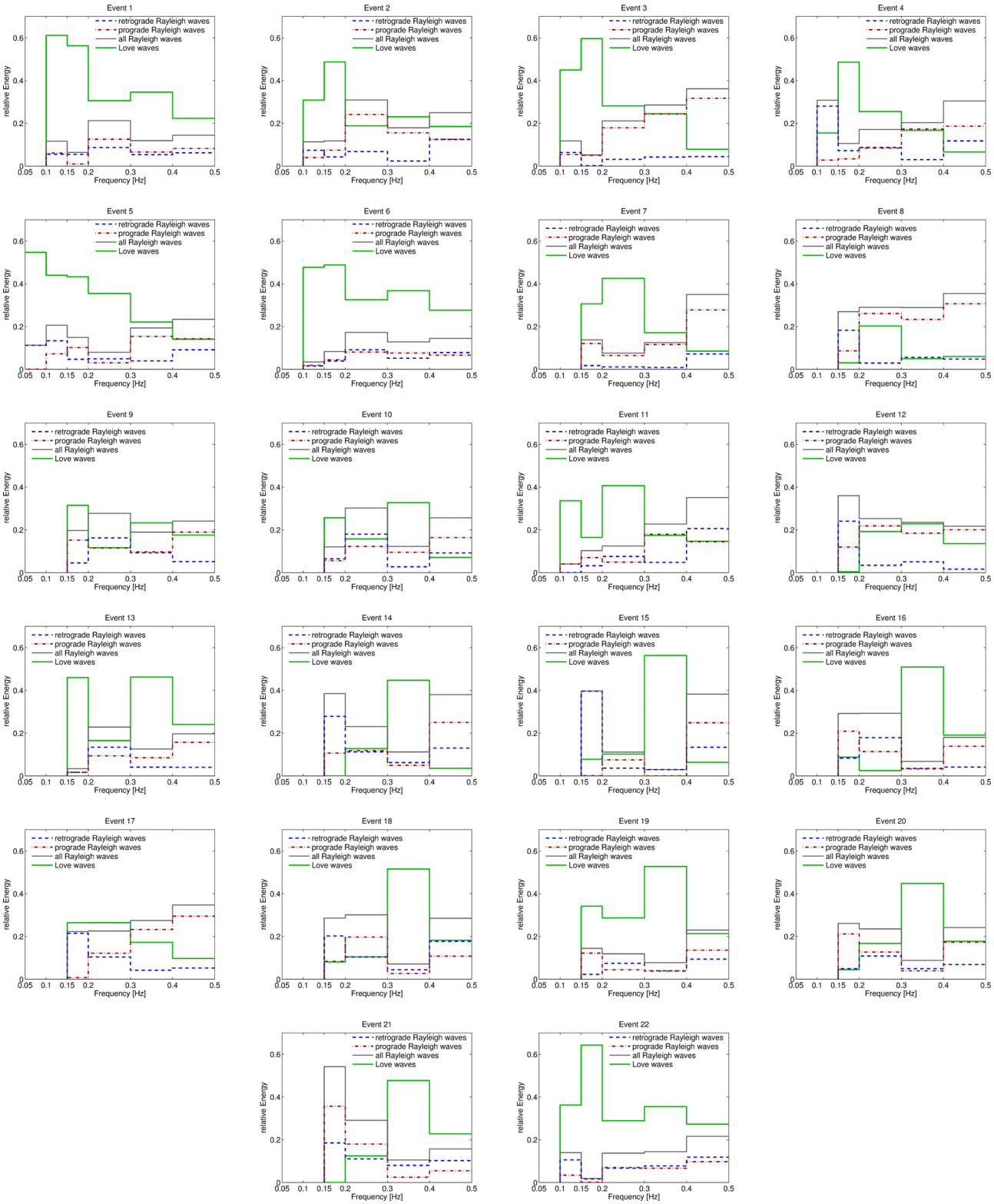

Figure 11. Repartition of energy between Love waves, retrograde and prograde Rayleigh waves in different frequency ranges for all events. The total energy in every frequency range is normalized to 1.

earthquake did not produce much energy in this frequency range, which questions the reliability of these data points.

The Love wave dispersion curve determined for event 5 provides much lower frequencies. For the fundamental mode of Love waves, the theoretical dispersion curve fits rather well. At lower frequencies, around 0.1 Hz, the measurements show larger velocities than the Bay Area Model predicts for the 1-D model of the central valley.

The Love wave dispersion curve for the clusters of events 13–16 and 18–21 is very different from all the other earthquakes. The

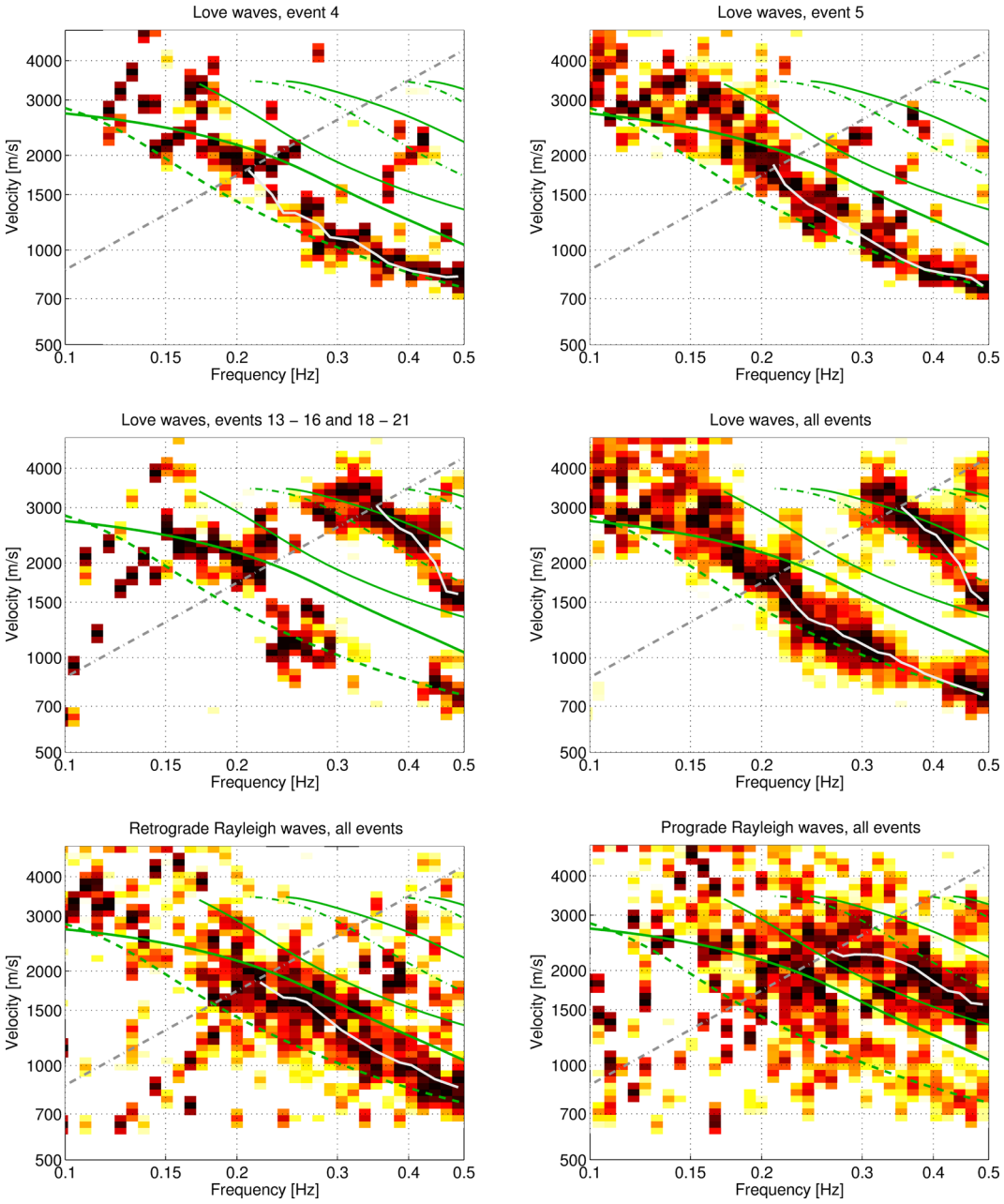

Figure 12. Measured dispersion curves for Love waves and Rayleigh waves. The velocity values for all time windows attributed to the respective wave type are included in the figures, weighted by the respective energy. The darkest colours indicate high energy concentrations. The solid green lines represent the fundamental and harmonic Rayleigh wave dispersion curves corresponding to the basin model for the central station of the array, the dashed and dash-dotted green lines represent the fundamental and harmonic Love wave dispersion curves. The dash-dotted grey line corresponds to the lower array resolution limit of the array. The solid white lines are manually picked. The figures show the Love wave results of event 4 (top left), event 5 (top right), the combination of events 13–16 and 18–21 (centre left) as well as the Love (centre right), retrograde (bottom left) and prograde (bottom right) Rayleigh waves for the combination of all events.

associated wave signals showed much more Love wave energy at higher frequencies (between 0.3 and 0.4 Hz) than the other events (Fig. 9). The dispersion curve shows much faster waves in this frequency range which are very close to the theoretical dispersion curve of the first harmonic Love wave mode according to the Bay Area Model. However, the fundamental Love wave mode is not well resolvable with these clusters of events.

A combination of all 22 analysed earthquakes, as shown in the centre right plot of Fig. 12, actually identifies both the fundamental and the first harmonic Love wave modes quite well. This figure has been obtained by summing the results for the different events after normalization by their respective total energies. In this way, the information of less energetic events also contributes to the results. The curves outside of the array resolution limits also seem reasonable.

The Rayleigh wave dispersion curves for the single earthquakes are badly retrieved. Even for the combination of all events, the actual dispersion curve is less evident than for the Love waves. A dispersion curve for retrograde Rayleigh waves can be identified, but the scattering is stronger than for the Love waves. The identified

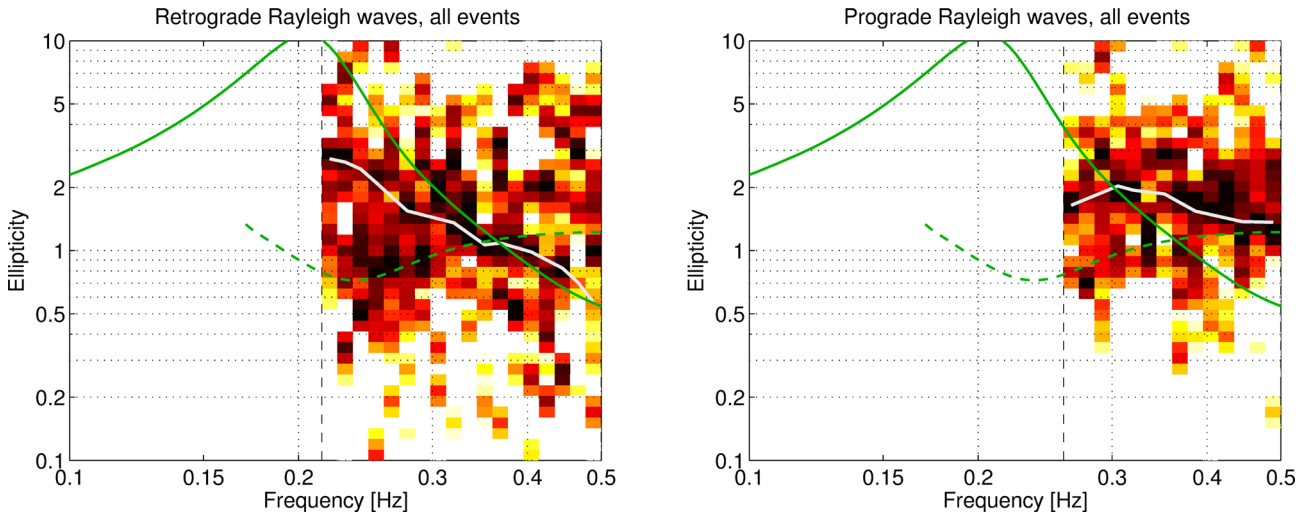

Figure 13. Ellipticity curves for retrograde (left) and prograde (right) Rayleigh waves. The figures are based on the combination of all events. The ellipticity values for all time windows identified for the respective wave types are included in the figure, weighted by their energy. The darkest colours indicate high-energy concentrations. The solid green line represents the ellipticity curve of the fundamental mode and the dashed green line the curve of the first harmonic mode of Rayleigh waves calculated for the basin model at the central station of the array. The solid white lines are manually picked.

dispersion curve is slower than the prediction of the Bay Area Model.

For the prograde Rayleigh waves, we see a cluster of data points at high frequencies which may well correspond to the first harmonic mode of Rayleigh waves. These data points come from different events compared to those of the first harmonic Love wave mode and are therefore not misidentified Love waves. Anyhow, the data points are quite scattered, but we can pick a dispersion curve which is actually not too far from the Bay Area Model.

5.5 Ellipticity curves

The ellipticity curves shown in Fig. 13 are of worse quality than the dispersion curves. They were obtained by only taking into account the data points contributing to the respective dispersion curves of Fig. 12 and neglecting all other waves. The curves are only shown in the frequency range that is inside the array resolution limits. For the retrograde Rayleigh waves, the data points are rather scattered, but we can define an ellipticity curve which is not too far from the theoretical ellipticity curve above 0.3 Hz. Around the theoretical ellipticity peak, however, the disagreement is rather large. For the prograde Rayleigh waves, the image is less scattered and an ellipticity curve can be determined which is rather flat. Compared to the identified dispersion curves, the uncertainty of the retrieved ellipticity curves is much larger.

6 SIMULATION RESULTS

The wave scattering in the basin which we measured for the different earthquakes is rather complex. The question arises if these scattering properties can also be retrieved by simulating selected earthquakes using the Bay Area Model. We selected two earthquakes from our data set for our simulations. The Bolinas earthquake (event 4) serves as an example for an earthquake from the northwest. The Gilroy earthquake (event 11) serves as an example for an earthquake from the southeast.

Ground-motion simulations were performed using SW4, a fourth-order finite-difference code based on the summation-by-parts principle (Petersson & Sjögreen 2014). The subsurface structure follows the USGS 3-D seismic velocity model, version 8.3 (USGS 2014), although we limited the minimum shear wave speed to 500 m s^{-1}. We set the grid spacing to 125 m, resulting in a maximum resolvable frequency of 0.5 Hz. Both topography and intrinsic attenuation were included in the numerical simulations. The seismic source was approximated by a point-source characterized by a Brune slip function with corner frequency of 1.0 Hz, compatible with the size of both events. Locations and mechanisms were retrieved from the online BSL moment tensor catalogue (Berkeley Seismological Laboratory 2016).

The data resulting from the simulations were analysed in the same way as the earthquake data. The resulting azimuthal energy distributions in the different frequency ranges for event 4 are shown in Fig. 14. The energy repartition between the different wave types and frequency ranges differs from the real earthquake data, but the distribution of the wave azimuths in the different frequency ranges is similar. Below 0.15 Hz, the wavefield is dominated by retrograde Rayleigh waves from the source direction in both cases, between 0.15 and 0.2 Hz, Love waves from the same direction dominate. Between 0.2 and 0.3 Hz, Love waves come mainly from the west for both the simulation and the real data, whereas the amount of prograde Rayleigh waves from the direct backazimuth is much larger in the simulation than for the real data. For higher frequencies, there is a complex pattern of different backazimuths both in the real data and the simulation cases, but they do not agree too much.

For event 11, the simulation results are shown in Fig. 15. Also here, the energy distribution between the different wave types is different from the earthquake data, but the results are qualitatively in good agreement. Below 0.2 Hz, most energy is carried by Love waves from southern directions in both cases. Some prograde Rayleigh waves arrive from the east in both cases. Between 0.2 and 0.3 Hz, the pattern for Love and Rayleigh waves is similar for the simulation and the real data. They arrive directly from the south and from the southeast, but not from the direct azimuth. Between 0.3 and 0.4 Hz, the majority of prograde Rayleigh waves comes from around 120 to 140° in both cases. Above 0.4 Hz, the simulation differs from the real data case.

As the main characteristics of the simulated wavefield are in good agreement with the real data, we can suppose that the simulations

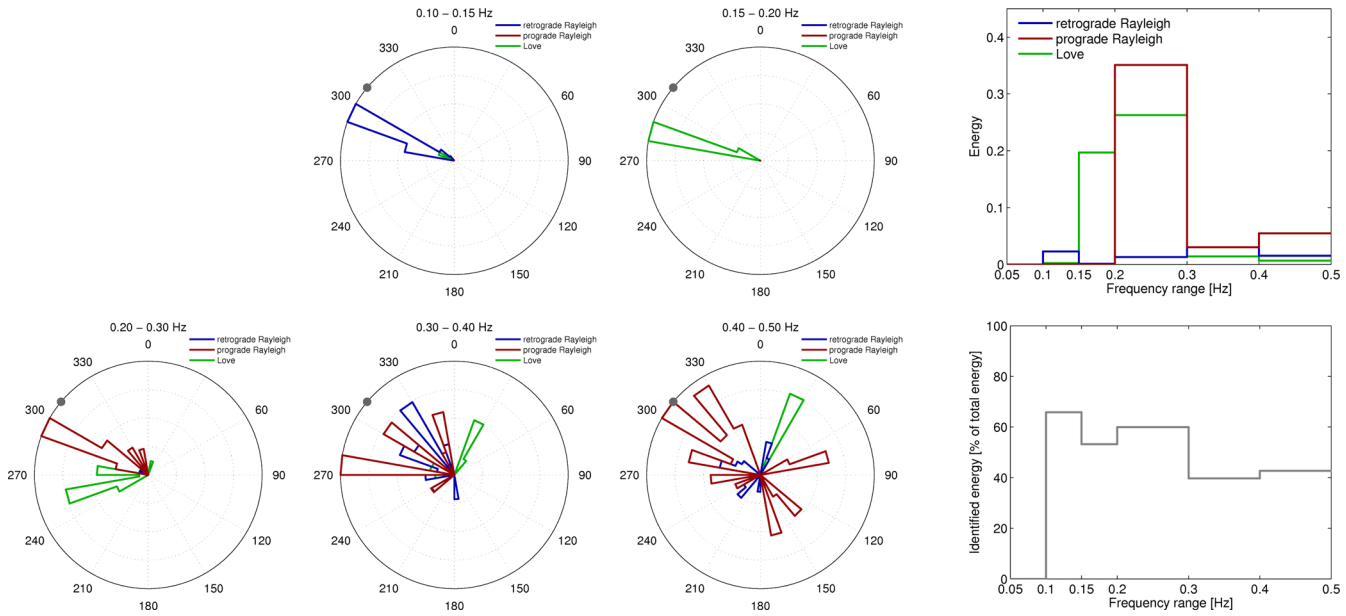

Figure 14. Azimuthal energy distribution in different frequency ranges, relative energy distribution and percentage of identified waves (similar to Fig. 5) as determined by the MUSIQUE analysis of the simulated data for the Bolinas earthquake (event 4).

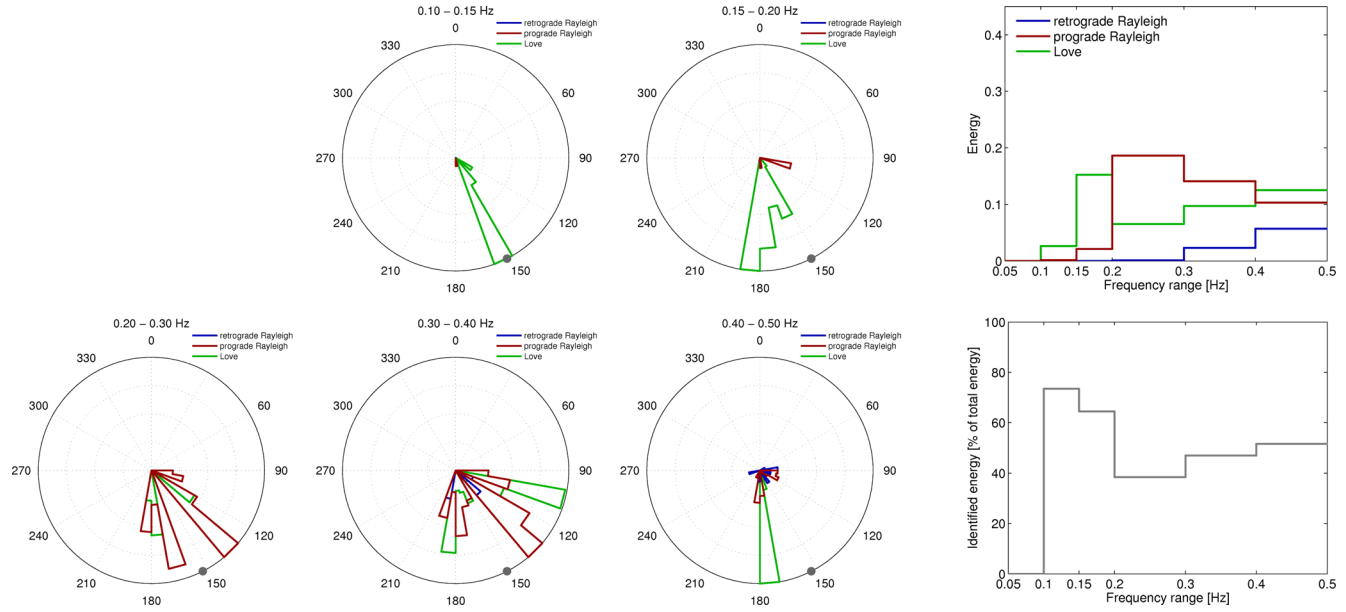

Figure 15. The same as Fig. 14 for the MUSIQUE analysis of the simulation results for the Gilroy earthquake (event 11). The real data analysis for this earthquake was shown in Fig. 5.

can help us in identifying the wave scatterers in the basin. Movies for the simulations of both events can be found in the Supporting Information. For event 4, the movie shows that the wavefield is mainly moving along the eastern basin edge. For event 11, we see that the rock formation directly south of the array acts as a wave scatterer and deviates the propagation direction in such a way that the waves cross the array directly from the south. Furthermore, the southern basin border also diffracts the waves crossing it. This confirms our earlier hypothesis of potential wave scattering geologic formations in the south of the basin. The differences of the wave energies between the simulations and the real data are certainly caused on one side by the simplified source model used for the simulation and on the other side by the complexity of the basin which cannot be completely included in the Bay Area Model.

7 DISCUSSION

The analysis of different earthquakes showed that the wave propagation in the Santa Clara Valley differs for the different earthquakes and that it is possible to assemble the results of the different earthquakes to get a better overview of the wave propagation properties in the valley.

The array was not optimal for our analysis. First, the inter-sensor spacing of the array limits the accessible frequency range. Furthermore, the structure of the basin is not homogeneous under the entire array. Additionally, the use of accelerometers limits the usage for this type of study. However, even with these limitations, it was possible to retrieve valuable information about the behaviour of seismic waves inside the Santa Clara Valley.

We have found a clear difference in the valley response for earthquakes occurring in the northwest compared with earthquakes occurring in the southeast. In both cases, the energy remains in the hemicycle where the earthquake occurs. The only exception to this rule are Love waves which arrive from the south and which are also generated by earthquakes in the north. Such a large Love wave scatterer effect was already seen by Frankel *et al.* (2001), who attributed it to either a large rock formation in the south of the Santa Clara Valley, exactly south of the San Jose Dense Seismic Array, or to the southern basin border. Our analysis and simulations confirm this result. We can identify both the large rock formation south of the array and the southern basin edge as wave scatterers. Waves with small deviations from the theoretical backazimuth can be attributed to the refraction of incident waves at the lateral basin borders.

In the analysis of the energy partition between Love and Rayleigh waves, we have found large differences between the different earthquakes. In general, between 40 and 50 per cent of the total wave energy were analysed for the single events. Overall, Love waves dominate the wavefield for most events (19 out of 22). The contribution of retrograde Rayleigh waves is rather weak, contributing to less than 26 per cent for all earthquakes. For three events, prograde Rayleigh waves represent the dominant wave type.

Combining all analysed earthquakes, we were able to determine dispersion curves for the fundamental and the first harmonic Love wave modes, where different earthquakes contributed to different parts of the dispersion curves. The first harmonic Love wave mode is only visible for the closest events. For the other events, no such waves of significant amplitude can be identified. They might be as well produced during these earthquakes, but attenuated on their way to the array or this might be a feature attributed to the special location of the close earthquakes. However, for retrograde Rayleigh waves, the determination of the dispersion curve is less clear than for Love waves. For prograde Rayleigh waves, we can also identify a dispersion curve and attribute it to the first harmonic Rayleigh wave mode. The determination of ellipticity curves was possible, but with a lot of uncertainty. Our results show differences between the measured dispersion and ellipticity curves and the theoretical curves of the Bay Area Model. However, our theoretical curves are based on a 1-D assumption for the central part of the array, which is certainly not true because the basin represents a more complex structure. There was also large scattering between the theoretical curves for the different stations (Fig. 1). Therefore, differences between these theoretical curves and the measured ones are not surprising. In any case, the Bay Area Model predicts that there should not be a singularity in the ellipticity of the fundamental Rayleigh wave. Consequently, this mode should be exclusively retrograde and the first harmonic mode prograde. Therefore, we identify the prograde Rayleigh waves as the first harmonic mode.

8 CONCLUSIONS

Using the MUSIQUE algorithm to analyse seismological data, we showed that this technique is capable to separate the contributions of Love and Rayleigh waves. The energies of the different wave types are quantified.

Our analysis showed differences in the basin response for different earthquakes. The azimuthal energy distribution is clearly different for waves incident from northwestern or southeastern directions. In both cases, the energy at the array is almost exclusively arriving from the respective hemicycle. However, even for waves coming from the north, there is some Love wave energy arriving from the south. Diffracted waves from different azimuths can also be seen, but they carry less energy. The wave types incident in the basin and their azimuthal distribution clearly depend on the earthquake source location. Apart from the large part of Love waves scattered in the south, there is no systematic basin response. Other studies in smaller arrays have shown such systematic patterns. Numerical simulations based on the detailed model of the San Francisco Bay Area were in agreement with the earthquake measurements and helped to identify the most important wave scatterers in the valley.

We identify prograde Rayleigh waves as the first harmonic mode. For the events that are the closest to the array (about 40 km), the first harmonic Love wave mode can be found.

By combining the information of all earthquakes, we succeeded in retrieving the Love wave dispersion curves for the fundamental and the first higher mode. For retrograde and prograde Rayleigh waves, it was also possible to identify dispersion and ellipticity curves, but they were less clear.

The array was not optimal for the type of analysis we did, but it was possible to gain important information on the seismic behaviour of the Santa Clara Valley. A smaller, but denser array of velocimeters located in the central part of the valley would allow us to study the higher-frequency part of incident waves, mainly for local earthquakes. This might help to further increase the understanding of the mechanisms in the Santa Clara Valley.

ACKNOWLEDGEMENTS

We would like to thank Arthur Frankel and David Carver for kindly providing the data and Arthur Rodgers for help with the San Francisco Bay Area model and personal discussions. The editor Eiichi Fukuyama and an anonymous reviewer gave precious suggestions to improve the manuscript. ASTER GDEM is a product of METI and NASA. Walter Imperatori is partly funded by ENSI.

REFERENCES

Berkeley Seismological Laboratory, 2016. 'Moment tensor homepage'. Available at: http://seismo.berkeley.edu/mt/, last accessed 8 April 2016.

Bokelmann, G.H.R. & Baisch, S., 1999. Nature of narrow-band signals at 2.083 Hz, *Bull. seism. Soc. Am.,* **89,** 156–164.

Brocher, T.M., 2005. Compressional and shear wave velocity versus depth in the San Francisco Bay Area, California: rules for USGS Bay Area velocity model 05.0.0, U.S. Geological Survey Open-File Report 2005-1317 Version 1.0.

Dolenc, D. & Dreger, D., 2005. Microseisms observations in the Santa Clara Valley, California, *Bull. seism Soc. Am.,* **95,** 1137–1149.

Dolenc, D., Dreger, D. & Larsen, S., 2005. Basin structure influences on the propagation of teleseismic waves in the Santa Clara Valley, California, *Bull. seism. Soc. Am.,* **95,** 1120–1136.

Elder, W.P., 2013. Bedrock geology of the San Francisco Bay Area: a local sediment source for bay and coastal systems, *Mar. Geol.,* **345,** 18–30.

Fletcher, J.B., Boatwright, J. & Lindh, A.G., 2003. Wave propagation and site response in the Santa Clara Valley, *Bull. seism. Soc. Am.,* **93,** 480–500.

Frankel, A., Carver, D., Cranswick, E., Bice, T., Sell, R. & Hanson, S., 2001. Observations of basin ground motions from a dense seismic array in San Jose, California, *Bull. seism. Soc. Am.,* **91,** 1–12.

Goldstein, P. & Archuleta, R.J., 1987. Array analysis of seismic signals, *Geophys. Res. Lett.,* **14,** 13–16.

Goldstein, P. & Archuleta, R.J., 1991a. Deterministic frequency-wavenumber methods and direct measurements of rupture propagation during earthquakes using a dense array: Theory and methods, *J. geophys. Res.,* **96,** 6173–6185.

Goldstein, P. & Archuleta, R.J., 1991b. Deterministic frequency-wavenumber methods and direct measurements of rupture propagation during earthquakes using a dense array: data analysis, *J. geophys. Res.,* **96,** 6187–6198.

Hamilton, W.R., 1847. On quaternions, *Proc. R. Irish Acad.,* **3,** 1–16.

Hartzell, S., Carver, D., Williams, R.A., Harmsen, S. & Zerva, A., 2003. Site response, shallow shear-wave velocity, and wave propagation at the San Jose, California, Dense Seismic Array, *Bull. seism. Soc. Am.,* **93,** 443–464.

Hartzell, S. *et al.*, 2006. Modeling and validation of a 3D velocity structure for the Santa Clara Valley, California, for seismic-wave simulations, *Bull. seism. Soc. Am.,* **96,** 1851–1881.

Hartzell, S., Ramirez-Guzman, L., Carver, D. & Liu, P., 2010. Short baseline variations in site response and wave-propagation effects and their structural causes: four examples in and around the Santa Clara Valley, California, *Bull. seism. Soc. Am.,* **100,** 2264–2286.

Hobiger, M., 2011. Polarization of surface waves: Characterization, inversion and application to seismic hazard assessment, *PhD thesis,* Université de Grenoble.

Hobiger, M., Le Bihan, N., Cornou, C. & Bard, P.-Y., 2012. Multicomponent signal processing for Rayleigh wave ellipticity estimation, *IEEE Signal Process. Mag.,* **29,** 29–39.

Kantor, I.L. & Solodovnikov, A.S., 1989. *Hypercomplex Numbers,* Springer-Verlag.

Miron, S., 2005. Méthodes multilinéaires et hypercomplexes en traitement d'antenne multicomposante haute résolution, *PhD thesis,* INP Grenoble.

Miron, S., Le Bihan, N. & Mars, J.I., 2005. High resolution vector-sensor array processing using quaternions, in *IEEE International Conference on Statistical Signal Processing, Bordeaux, France, 2005,* Novosibirsk, pp. 918–923.

Miron, S., Le Bihan, N. & Mars, J., 2006. Quaternion-MUSIC for vector-sensor array processing, *IEEE Trans. Signal Process.,* **54,** 1218–1229.

Petersson, N.A. & Sjögreen, B., 2014. Super-grid modeling of the elastic wave equation in semi-bounded domains, *Commun. Comput. Phys.,* **16,** 913–955.

Phelps, G.A., Graymer, R.W., Jachens, R.C., Ponce, D.A., Simpson, R.W. & Wentworth, C.M., 2008. Three-dimensional geologic map of the Hayward Fault zone, San Francisco Bay region, California, U.S. Geological Survey Scientific Investigations Map 3045. Available at: http://pubs.usgs.gov/sim/3045/.

Schmidt, R.O., 1986. Multiple emitter location and signal parameter estimation, *IEEE Trans. Antennas Propag.,* **34,** 276–280.

USGS 2014. '3-D geologic and seismic velocity models of the San Francisco Bay Region'. Available at: http://earthquake.usgs.gov/data/3dgeologic/, last accessed 8 April 2016.

Ward, J.P., 1997. *Quaternions and Cayley Numbers: Algebra and Applications,* Kluwer.

Wentworth, C.M., 1997. Geologic materials of the San Francisco Bay region, U.S. Geological Survey Open-File Report 97-744, part 5.

Zhang, F., 1997. Quaternions and matrices of quaternions, *Linear Algebr. Appl.,* **251,** 21–57.

APPENDIX A: QUATERNIONS

Quaternions are an extension of the complex numbers and have been discovered by the Irish mathematician Hamilton (Hamilton 1847). For a good overview of quaternion algebra, we refer the interested reader to Kantor & Solodovnikov (1989). Here, we will only give an overview of the basic rules of quaternion algebra.

A1 Definition

A quaternion is defined as

$$q = q_0 + q_1\,\mathtt{i} + q_2\,\mathtt{j} + q_3\,\mathtt{k}, \tag{A1}$$

where $q_0, \ldots, q_3$ are real numbers and $\mathtt{i}$, $\mathtt{j}$ and $\mathtt{k}$ are imaginary units defined by

$$\mathtt{i}^2 = \mathtt{j}^2 = \mathtt{k}^2 = \mathtt{ijk} = -1. \tag{A2}$$

q_0 is called scalar part of the quaternion and $q_1\,\mathtt{i} + q_2\,\mathtt{j} + q_3\,\mathtt{k}$ is the vector part. Based on eq. (A2), it follows that

$$\begin{aligned} \mathtt{ij} &= \mathtt{k}, \quad \mathtt{ji} = -\mathtt{k} \\ \mathtt{jk} &= \mathtt{i}, \quad \mathtt{kj} = -\mathtt{i} \\ \mathtt{ki} &= \mathtt{j}, \quad \mathtt{ik} = -\mathtt{j}. \end{aligned} \tag{A3}$$

The multiplication of quaternions is not commutative.

A2 Properties

Be $p = p_0 + p_1\,\mathtt{i} + p_2\,\mathtt{j} + p_3\,\mathtt{k}$ and $q = q_0 + q_1\,\mathtt{i} + q_2\,\mathtt{j} + q_3\,\mathtt{k}$ quaternions, then their addition is defined as

$$p + q = (p_0 + q_0) + (p_1 + q_1)\,\mathtt{i} + (p_2 + q_2)\,\mathtt{j} + (p_3 + q_3)\,\mathtt{k}. \tag{A4}$$

The multiplication of p and q yields

$$\begin{aligned} pq = {} & (p_0q_0 - p_1q_1 - p_2q_2 - p_3q_3) \\ & + (p_0q_1 + p_1q_0 + p_2q_3 - p_3q_2)\,\mathtt{i} \\ & + (p_0q_2 - p_1q_3 + p_2q_0 + p_3q_1)\,\mathtt{j} \\ & + (p_0q_3 + p_1q_2 - p_2q_1 + p_3q_0)\,\mathtt{k}. \end{aligned} \tag{A5}$$

It follows from eqs (A4) and (A5) that the addition of quaternions is commutative, whereas the multiplication, in general is not: $pq \neq qp$. However, quaternion multiplication is still associative, that is, $(pq)r = p(qr)$ for any quaternions p, q and r.

In analogy to the complex numbers, the conjugate of a quaternion $q = q_0 + q_1\,\mathtt{i} + q_2\,\mathtt{j} + q_3\,\mathtt{k}$ is defined by

$$q^* = q_0 - q_1\,\mathtt{i} - q_2\,\mathtt{j} - q_3\,\mathtt{k}. \tag{A6}$$

Using eq. (A5), it can be shown that the product of q with its conjugate q^* is given by

$$qq^* = q^*q = q_0^2 + q_1^2 + q_2^2 + q_3^2. \tag{A7}$$

In this way, the norm or absolute value of a quaternion can be defined by

$$|q| = \sqrt{q^*q} = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}. \tag{A8}$$

The inverse value of a quaternion $q \neq 0$ is given by

$$q^{-1} = \frac{q^*}{|q|^2}. \tag{A9}$$

In this way, the quaternion division is also defined.

A3 Quaternion vectors

A quaternion vector $\mathbf{q}$ is an N-dimensional vector whose elements are quaternions. For two quaternion vectors $\mathbf{q}$ and $\mathbf{p}$, the inner product is given by $\mathbf{q}^\dagger\mathbf{p}$, (e.g. Miron 2005) where $\mathbf{q}^\dagger = \left(\mathbf{q}^T\right)^*$ denotes the conjugate transpose of a quaternion vector.

The norm of a quaternion vector can be defined by $\|\mathbf{q}\| = \sqrt{\mathbf{q}^\dagger\mathbf{q}}$. The set of quaternion-valued vectors forms a Hilbert space. Two quaternion vectors $\mathbf{q}$ and $\mathbf{p}$ are orthogonal if their inner product $\mathbf{q}^\dagger\mathbf{p}$ equals zero.

A4 Quaternion matrices

The properties of quaternion matrices differ from real or complex matrices because quaternions are not commutative. A good overview of quaternion matrices is given by Zhang (1997), so we will only give a short overview of the properties of quaternion matrices here. Considering a quaternion-valued matrix $\mathbf{A}$ of size $M \times N$, we will denote by $\mathbf{A}^T$ the transpose, by $\mathbf{A}^*$ the conjugate and by $\mathbf{A}^\dagger$ the conjugate transpose of $\mathbf{A}$. Considering two quaternion-valued matrices $\mathbf{A}$ (of size $M \times N$) and $\mathbf{B}$ (of size $N \times O$), the following properties hold (Zhang 1997):

(i) $(\mathbf{A}^*)^T = (\mathbf{A}^T)^*$
(ii) $(\mathbf{A}\,\mathbf{B})^\dagger = \mathbf{B}^\dagger\,\mathbf{A}^\dagger$
(iii) $(\mathbf{A}\,\mathbf{B})^* \neq \mathbf{A}^*\,\mathbf{B}^*$, in general
(iv) $(\mathbf{A}\,\mathbf{B})^T \neq \mathbf{B}^T\,\mathbf{A}^T$, in general
(v) $(\mathbf{A}\,\mathbf{B})^{-1} = \mathbf{B}^{-1}\,\mathbf{A}^{-1}$, if $\mathbf{A}$ and $\mathbf{B}$ are invertible
(vi) $(\mathbf{A}^\dagger)^{-1} = (\mathbf{A}^{-1})^\dagger$, if $\mathbf{A}$ is invertible
(vii) $(\mathbf{A}^*)^{-1} \neq (\mathbf{A}^{-1})^*$, in general
(viii) $(\mathbf{A}^T)^{-1} \neq (\mathbf{A}^{-1})^T$, in general

Due to the non-commutativity of quaternions, a quaternion-valued matrix $\mathbf{A}$ of size $N \times N$ has right and left eigenvalues, satisfying the equations

$$\mathbf{A}\mathbf{u}_r = \mathbf{u}_r\lambda_r, \tag{A10}$$

$$\mathbf{A}\mathbf{u}_l = \lambda_l\mathbf{u}_l, \tag{A11}$$

respectively. In general, both eigenvalues are different. The theory of left eigenvalues is not yet well understood in literature, but the properties of right eigenvalues are well established (Zhang 1997). Therefore, we use exclusively right eigenvalues and eigenvectors.

The quaternion-matrix $\mathbf{A}$ has N linearly independent eigenvectors. In the MUSIQUE algorithm, the quaternion-valued covariance matrix $\mathbf{S}_q(f)$ is built according to eq. (1) and therefore Hermitian, that is, $\mathbf{S}^\dagger = \mathbf{S}$. Let λ be a complex-valued eigenvalue of a Hermitian quaternion matrix $\mathbf{A}$ (of size $N \times N$) and $\mathbf{u}$ the associated normed quaternion-valued eigenvector of size $N \times 1$ ($\mathbf{u}^\dagger\mathbf{u} = 1$). Then the eigenvalue λ is real-valued, as can be shown by the short calculation:

$$(\mathbf{A}\mathbf{u})^\dagger\mathbf{u} = (\mathbf{u}\lambda)^\dagger\mathbf{u} = \lambda^*\mathbf{u}^\dagger\mathbf{u} = \lambda^* \tag{A12}$$

$$(\mathbf{A}\mathbf{u})^\dagger\mathbf{u} = \mathbf{u}^\dagger\mathbf{A}\mathbf{u} = \mathbf{u}^\dagger\mathbf{u}\lambda = \lambda. \tag{A13}$$

Therefore, $\lambda^* = \lambda$ and λ is real-valued.

Different eigenvectors of a Hermitian quaternion-valued matrix $\mathbf{A}$ are orthogonal and the set of eigenvectors can be used to form an orthonormal basis of the Hilbert space. This can be proved in the following way: We consider two different eigenvalues $\lambda_1 \neq \lambda_2$ of a Hermitian matrix $\mathbf{A}$ with associated eigenvectors $\mathbf{u_1}$ and $\mathbf{u_2}$. Then, we can calculate the following equations:

$$(\mathbf{A}\mathbf{u}_1)^\dagger\mathbf{u}_2 = (\mathbf{u}_1\lambda_1)^\dagger\mathbf{u}_2 = \lambda_1(\mathbf{u}_1^\dagger\mathbf{u}_2), \tag{A14}$$

$$(\mathbf{A}\mathbf{u}_1)^\dagger\mathbf{u}_2 = \mathbf{u}_1^\dagger\mathbf{A}\mathbf{u}_2 = (\mathbf{u}_1^\dagger\mathbf{u}_2)\lambda_2, \tag{A15}$$

which can only be true if $\mathbf{u}_1^\dagger\mathbf{u}_2 = 0$. Therefore, different eigenvectors of $\mathbf{A}$ are orthogonal.

A Hermitian quaternion-valued matrix $\mathbf{A}$ (of size $N \times N$) can be decomposed in the following way:

$$\mathbf{A} = \mathbf{U}\Delta\mathbf{U}^\dagger, \tag{A16}$$

where $\mathbf{U}$ is a unitary matrix ($\mathbf{U}\mathbf{U}^\dagger = \mathbf{U}^\dagger\mathbf{U} = \mathbf{I}$) and Δ is a diagonal real-valued matrix. Therefore, we can write:

$$\mathbf{A} = \sum_{i=1}^{r-k} \mathbf{u}_i \otimes \mathbf{u}_i^\dagger\lambda_i + \sum_{i=r-k+1}^{r} \mathbf{u}_i \otimes \mathbf{u}_i^\dagger\lambda_i. \tag{A17}$$

Here, r is the rank of matrix $\mathbf{A}$. k is the dimension of the signal space and $r - k$ the dimension of the noise subspace. k is the number of 'large' eigenvalues, the signal and noise subspaces are orthogonal. This separation of signal and noise subspaces is crucial for the MUSIC algorithm.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this paper:

Figure S1. Azimuthal energy distribution for different frequency ranges for event 1. The grey dot indicates the theoretical backazimuth of the event. In each frequency range, all curves were normalized to the maximum. For clarity reasons, the curves for Rayleigh and Love waves are normalized by their respective maxima. The figure on top at the right side indicates the relative energy distribution of the wave types in the respective frequency bands. The figure below indicates the percentage of the total energy in the respective frequency range which is carried by identified Love and Rayleigh waves.
Figure S2. Azimuthal energy distribution for event 2 (similar to Fig. S1).
Figure S3. Azimuthal energy distribution for event 3 (similar to Fig. S1).
Figure S4. Azimuthal energy distribution for event 6 (similar to Fig. S1).
Figure S5. Azimuthal energy distribution for event 7 (similar to Fig. S1).
Figure S6. Azimuthal energy distribution for event 8 (similar to Fig. S1).
Figure S7. Azimuthal energy distribution for event 9 (similar to Fig. S1).
Figure S8. Azimuthal energy distribution for event 10 (similar to Fig. S1).
Figure S9. Azimuthal energy distribution for event 12 (similar to Fig. S1).
Figure S10. Azimuthal energy distribution for event 17 (similar to Fig. S1).
Figure S11. Azimuthal energy distribution for event 22 (similar to Fig. S1).
Movie S12. Movie of the simulation of wave propagation in the San Francisco Bay Area for the Bolinas earthquake (event 4). The brown-shaded background indicates the bedrock depth, which is defined as the depth at which the shear wave velocity reaches 3000 m s^{-1}. The simulation indicates the magnitude of the ground velocity, which is the length of the 3-D velocity vector. The time of the video is in seconds, the north is towards the top. The locations of the seismic array stations are indicated by dark brown triangles.
Movie S13. Movie of the simulation of wave propagation for the Gilroy earthquake (event 4), similar to Movie S12. (http://gji.oxfordjournals.org/lookup/suppl/doi:10.1093/gji/ggw289/-/DC1)

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.