2023-02-06

Quantifying rupture characteristics of microearthquakes in the parkfield area using a high-resolution borehole network

Geophysical Journal International

Geophys. J. Int. (2023) **233,** 1772–1785
Advance Access publication 2023 January 25
GJI Seismology

https://doi.org/10.1093/gji/ggad023

Quantifying rupture characteristics of microearthquakes in the Parkfield Area using a high-resolution borehole network

Colin N. Pennington [1], Qimin Wu [2], Xiaowei Chen [3] and Rachel E. Abercrombie [4]

[1] *Lawrence Livermore National Laboratory, Livermore, CA* 94550*, USA. E-mail: pennington6@llnl.gov*
[2] *Lettis Consultants International, Concord, CA* 94520*, USA*
[3] *Department of Geology and Geophysics, Texas A&M University, College Station, TX* 77843*, USA*
[4] *Department of Earth and Environment, Boston University, Boston, MA* 02215*, USA*

Accepted 2023 January 13. Received 2023 January 4; in original form 2022 October 4

SUMMARY

It is well known that large earthquakes often exhibit significant rupture complexity such as well separated subevents. With improved recording and data processing techniques, small earthquakes have been found to exhibit rupture complexity as well. Studying these small earthquakes offers the opportunity to better understand the possible causes of rupture complexities. Specifically, if they are random or are related to fault properties. We examine microearthquakes ($M < 3$) in the Parkfield, California, area that are recorded by a high-resolution borehole network. We quantify earthquake complexity by the deviation of source time functions and source spectra from simple circular (omega-square) source models. We establish thresholds to declare complexity, and find that it can be detected in earthquakes larger than magnitude 2, with the best resolution above M2.5. Comparison between the two approaches reveals good agreement (>90 per cent), implying both methods are characterizing the same source complexity. For the two methods, 60–80 per cent (M 2.6–3) of the resolved events are complex depending on the method. The complex events we observe tend to cluster in areas of previously identified structural complexity; a larger fraction of the earthquakes exhibit complexity in the days following the M_w 6 2004 Parkfield earthquake. Ignoring the complexity of these small events can introduce artefacts or add uncertainty to stress drop measurements. Focusing only on simple events however could lead to systematic bias, scaling artefacts and the lack of measurements of stress in structurally complex regions.

Key words: Earthquake dynamics; Earthquake source observations; Rheology and friction of fault zones.

1 INTRODUCTION

Accurate characterization of earthquake rupture processes provides useful information for the understanding of not only the physics of earthquakes, but also fault properties and stress regime. For small to moderate earthquakes, one of the most common methods to characterize their ruptures is through the fitting of their source spectra with a theoretical model. For a simple circular rupture, the far field displacement spectrum can be explained by omega-square models with a single corner frequency (Brune 1970; Madariaga 1976). With a few assumptions, one can obtain estimates of rupture duration, rupture radius and stress drop based on the corner frequency measured from spectral fitting (Abercrombie 1995, 2014; Ide *et al.* 2003; Venkataraman *et al.* 2006). A well-known limitation of this method is that it cannot be accurately applied to events that are not simple circular ruptures such as many large earthquakes (Boatwright *et al.* 1991; Atkinson 1996). If a simple source model is applied to an event with a complex rupture, the estimated source parameters will be biased (Abercrombie 2014; Fan & McGuire 2018; McGuire & Kaneko 2018; Shearer *et al.* 2019) and interpretations of these estimates can obfuscate the true rupture processes occurring in a region (Madariaga 1979; Wu *et al.* 2019; Abercrombie 2021). But how much of a concern should such behaviour be to our analysis of small events?

With increasing station coverage and the closer proximity of stations to small and moderate earthquakes, there are an increasing number of observations indicating that some small events do not fit the assumption of a simple circular rupture (e.g. Yamada *et al.* 2005; Abercrombie 2014; Fan & McGuire 2018; Pennington *et al.* 2022). These events require more sophisticated methods of rupture characterization, such as second moment analysis (e.g. McGuire 2017; Meng *et al.* 2020) or line sources (Gallovič *et al.* 2009; Böse *et al.* 2012; Abercrombie *et al.* 2020). In the best case with enough data, a finite fault inversion can be used to estimate both

the temporal and spatial distribution of slip on a faults surface (e.g. Hartzell & Heaton 1983, 1986; Yamada *et al.* 2005; Mai & Thingbaijam 2014). In the time domain, the temporal evolution of the slip distribution (source–time function) will capture the complexity of the slip distribution to a certain degree by showing multiple pulses, each of which represent different episodes of the rupture (Kikuchi & Kanamori 1982; Vallée & Douet 2016; Danré *et al.* 2019). Due to the sparsity of some seismic networks and resolution limits caused by limited frequency bandwidth, these methods are usually only applied to large events. When these methods are applied to small magnitude earthquakes, they reveal levels of complexity similar to those observed for large magnitudes (Ide 2001; Fischer 2005; Wang *et al.* 2014; Pennington *et al.* 2022). This observation aligns with idea of self-similarity across earthquake scales as some studies suggest (e.g. Choy & Boatwright 1995; Mcgarr 1999; Ide & Beroza 2001; Allmann & Shearer 2009; Uchide & Ide 2010); if earthquakes are self-similar, then we would expect small earthquakes to exhibit similarly complex ruptures as large ones.

A large question remains, how prevalent is this complexity in smaller magnitude earthquakes ($M < 3$)? Does it vary in space or with time? What factors lead to heterogeneity? Current observations provide conflicting answers. At Parkfield, Abercrombie *et al.* (2020) found that repeating earthquakes had similar levels of complexity to other events in their repeating sequences. Elsewhere, temporal variation is seen in the size of the rupture areas of repeating earthquakes (Uchida *et al.* 2015; Chaves *et al.* 2020). To better understand the factors that influence rupture complexity and its prevalence, we need to find ways of quantifying it systematically for multiple events. If the goal of this quantification is simply to identify whether an event has a complex rupture or a simple one (e.g. one that aligns with the assumed models), such analysis would be simpler than trying to accurately represent that rupture through other techniques (e.g. finite-fault inversions, line sources, etc.). Complex source processes affect both the source time functions (STFs) and the amplitude spectra (e.g. Abercrombie 2014; Abercrombie *et al.* 2020), which means the examination of either of these could allow for the simple quantification of rupture behaviour as deviating from some assumed expected behaviour.

There are many examples of complex source processes being reflected in STF complexities (e.g. multiple pulse) (Kikuchi & Kanamori 1982; Vallée & Douet 2016; Danré *et al.* 2019). When the method of using an EGF to deconvolve a STF are also applied to small events they have revealed complexity as well (Fischer 2005; Wang *et al.* 2014), thereby offering a method of quantifying small earthquake complexity. This requires good quality Green's function to work of course, reducing the number of events that can be examined. This limitation and the desire to examine more events brings us to another observation method. This one is based on the many observations that complex source processes causes an earthquake source spectra to deviate from a simple circular model (i.e. omega-square spectral model) that is commonly used for spectral fitting (Schneider *et al.* 1993; Atkinson & Silva 1997; Silva *et al.* 1998; McGuire *et al.* 2002). However, only recently was this common occurrence systematically identified in smaller earthquakes; for example Uchide & Imanishi (2016) found events with magnitude ranging between 3.2 and 4.0 in the northern Ibaraki prefecture and the Fukushima prefecture of Japan, consistently deviated from the simple model, indicating that such a method could be used to broadly assess if earthquakes ruptures were complex or simple. Estimating STF's provides a second means of identifying complex ruptures.

These two methods should theoretically be able to identify the same events as being complex or simple. Modelling by Madariaga (1981) showed that when an event has two moment release pulses during its rupture, its spectra will reflect the width of the individual pulses and the delay time between them. This behaviour is reflected in real observations with Wu *et al.* (2019) observing and using both spectral complexity and source–time function complexity to characterize an M_w 4 earthquake in central Oklahoma, a rupture that showed significant directivity and multiple subevents.

To test these methods and see if they can be used to verify one another's observations, we analyse well-recorded earthquakes at Parkfield using both frequency and time-domain methods to quantify the degree of complexity. Both methods use small earthquakes as empirical Green's function events to correct for the path and site effects. We then compare the two methods and assess their resolution limitations. The methods chosen to characterize events are their source spectral deviation from a simple model (Uchide & Imanishi 2016) and by their relative source–time functions (e.g. Wang *et al.* 2014; Wu *et al.* 2019). We find good agreement of event classification (complex or simple) between the two methods. The majority of the $M > 2.6$ earthquakes are classified as complex from both methods. The agreement indicates that the two methods are observing the same complexity to a certain degree. This also means that either of the method could be used to quantify rupture characteristics, but using both validates the observations to a better degree.

2 DATA AND PARKFIELD STUDY AREA

We examine microearthquakes (M 0–3.0) that occurred between 2001 and 2011 along a 80 km long Parkfield fault segment of the San Andreas Fault centred on Parkfield (Fig. 1). They were recorded by the shallow borehole High-Resolution Seismic Network, HRSN (Northern California Earthquake Data Center 2014). The Parkfield area is chosen because of these high quality recordings, also because it is a densely studied area with well-documented structural and lithological features for comparison with our earthquake complexity observations. We identify target and EGF event pairs using the double difference relocated catalogue of Northern California (Waldhauser & Schaff 2008; Waldhauser 2009). We treat each of the 14,603 catalogued earthquake (M 0-3.5) as a potential target event, and search for suitable EGF events that are within a 400 m radius of the hypocentre and more than 1 magnitude unit smaller than the given target event. The magnitude we use in this study are from a spectral-calibrated moment magnitude catalogue (Zhang *et al.* 2022). We find 48,343 target and EGF event pairs to use in the analysis. For all the selected events, we download triggered waveforms from the 13 HRSN borehole stations with a sampling rate of 250 samples per sec from the Northern California Earthquake Data Center (NCEDC). Instrument gain corrections are applied according to the simple instrument response table from NCEDC data base; the constant instrument response functions cancel in the EGF analyses. P-wave arrival times are either from catalogue phase picks or an autopicker based on 1-D velocity model (Li & Peng 2016). By using multiple stations and EGFs, we can obtain multiple estimates of the spectral ratio and relative source time functions (RSTF) for each target event. For the spectral measurements we stack over multiple stations and EGFs to obtain a single measurement per target event and calculate a single estimate of complexity for that event. For the time domain deconvolution, we stack multiple EGF results for each station to obtain a single RSTF for each station. We then calculate an independent estimate of source complexity for each

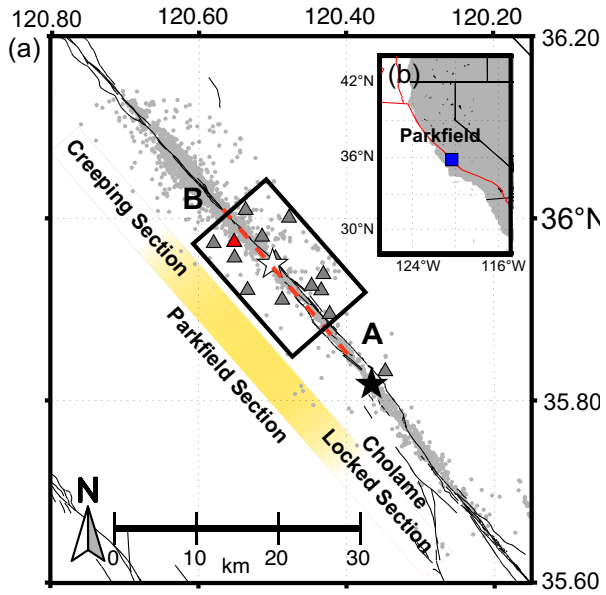

Figure 1. (a) Parkfield study area with (b) inset map of overall region with study area marked by blue square. All earthquakes in the relocated catalogue are shown as grey dots and thin black lines are Quaternary fault traces from the U.S. Geological Survey & California Geological Survey (2006). Full cross section length is shown as dashed red line between points A and B. The 1966 M6 earthquake (white star), 2004 M6 earthquake (black star) and stations used in the analysis are plotted (grey triangles) with SAFOD also being shown (red triangle). Area considered to be within the network is shown by black rectangle box. The interpreted locked, creeping and Parkfield section are from Perrin *et al.* (2019). The exact location of the transition point between sections is not known, and we therefore represent it by a gradual change in the colour of one section to the next.

station using each EGF and then take the average over all stations to obtain an estimate for the event.

3 METHODS

3.1 Multiple spectral ratio analysis

To determine if an earthquake is spectrally complex, we need to first obtain its source spectra and then compare it to a simple model. If an earthquake is a simple circular rupture, then its spectra should not significantly deviate from the omega-square source spectra model. To understand if such deviations are common for small magnitude earthquakes, we perform multiple spectral ratio analysis following Uchide & Imanishi (2016) and fit the spectral ratios of target to EGF event with an omega-square model (Boatwright 1978):

$$R_i(f) = \frac{M_0^{\text{Target}}}{M_0^{\text{EGF}[i]}} \left[\frac{1 + \left(f/f_c^{\text{EGF}[i]}\right)^4}{1 + \left(f/f_c^{\text{Target}}\right)^4} \right]^{1/2} \quad (1)$$

in which f is the frequency, and f_c and M_0 are the corner frequency and moment of the 'Target' or the ith 'EGF' event, respectively. We allow the f_c to vary across stations and event pairs and do not hold it to an average value. We do this with the goal of estimating a single spectral ratio averaged across stations and EGFs to estimate complexity from.

We calculate event displacement spectra for P waves using the vertical channel recordings. For the P waves, we calculate the spectra over a window that begins 0.1 s before the P arrival and extends to 0.5–1 s depending on the S-wave arrival time. A noise window of the same length is selected before the P wave to be used for an estimation of the signal-to-noise ratio (SNR). Only spectra with a minimum SNR ≥ 5 between 2 and 60 Hz are used to calculate spectral ratios, and only EGF and Target event pairs that have spectral ratios from 6 or more stations are used in the following analysis. We limit it to 60 Hz instead of 125 Hz (the maximum possible frequency), due to very few EGFs passing the SNR criteria if the range is set above 60 Hz. For each event pair, corner frequencies and moment ratios are estimated using a grid-search method to minimize the misfit between the observed median of the individual ratios across stations (R_i^{Obs}) and predicted (R_i^{Syn}) ratio obtained from eq. (1). This is defined as

$$\text{misfit}_i = \left[\log R_i^{\text{obs}}(f) - \log R_i^{\text{syn}}\left(f; M_0^{\text{Target}}/M_0^{\text{EGF}[i]}, f_c^{\text{Target}}, f_c^{\text{EGF}[i]}\right)\right]^2 \quad (2)$$

in which the superscripts 'obs' and 'syn' indicate the observed and synthetic spectral ratios, respectively. For each of the EGF events, we examine the variance reduction which is defined as

$$VR_i = \frac{\text{misfit}_i}{\text{Var}\left(\log R_i^{\text{obs}}(f)\right)} \quad (3)$$

in which 'Var' denotes variance with respect to the logarithmic mean over the frequency band of interest. We only include those event pairs that obtain variance reductions greater than 85 per cent. This is done to remove those pairs that deviate so strongly that it makes fitting a model to even estimate the deviation from that model difficult.

For target events with at least six EGFs with measurements at six or more stations (not necessarily common between the EGF Target pairs) that pass this criterion, we estimate the misfit from the omega-square model by taking the median of the residual between the (R_i^{Obs}) and predicted (R_i^{Syn}) of all event pairs (Figs 2a–d). The median misfit emphasizes the deviations from the predicted model that are common across all EGFs, which can be attributed to the target events source spectra, and suppresses the deviations due to a single EGF's source spectrum.

To quantify the misfit observed for each target event, we follow Uchide & Imanishi (2016) and calculate the peak-to-peak residual ratio, which is defined as ratio of the maximum to minimum residuals of the median misfit. The peak-to-peak ratio is calculated using an incrementally increasing frequency range to identify the frequency at which maximum deviation from the model occurs. This bandwidth starts at 2–3 Hz and increases by 1 Hz until it reaches a maximum bandwidth of 2–50 Hz (Fig. 2). It was limited to 50 Hz instead of the maximum SNR tested frequency of 60 Hz due to the high variability in spectral ratios across a target events spectral ratios observed above 50 Hz (e.g. Fig. 2b). This range fundamentally limits the complexity we can observe, since as the target earthquake magnitude decreases so does the moment of any subevents it contains. Eventually the effects of these subevents will extend into frequency ranges that exceed the instrument sampling rate or fall below a good SNR level.

We visually examine the results to determine what peak-to-peak ratio constitutes a significant deviation in the spectra. We find that events with peak-to-peak ratios ≤ 1.6 exhibit little deviation from the omega-square model (e.g. Figs 2g–h) and are classified as 'SR_{Dev} simple' (Fig. 3a). Events with peak-to-peak ratios ≥ 1.8 exhibit significant deviation from the model (e.g. Figs 2e–f) and are classified as 'SR_{Dev} complex' (Fig. 3a). Events with peak-to-peak ratios between 1.6 and 1.8 are difficult to attribute to either group and therefore are unclassified. The events excluded in this range account for roughly 20 per cent of all events and only 17 per cent of the larger magnitude (≥ 2.5) (see Supplemental Fig. S1). The absolute values of deviation above and below the estimated model depends on the model fit to the spectra, which is strongly controlled by f_c^{EGF}. This

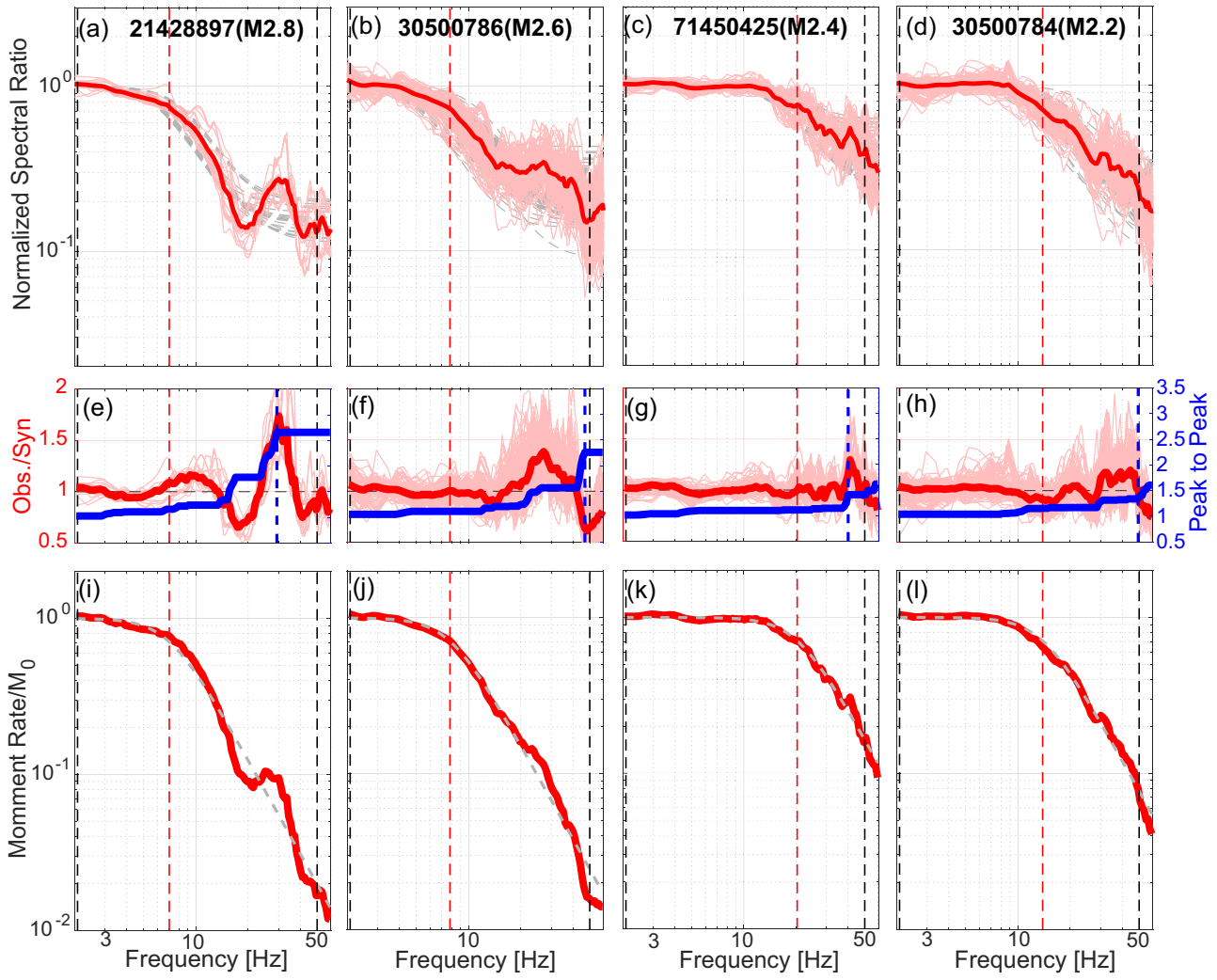

Figure 2. Results of the spectral ratio analysis for four selected earthquakes. (a)–(d) Spectral ratios of individual empirical Green's function events averaged over all stations (thin light red lines), and their median stack (dark red line), which are normalized by the moment ratio. Synthetic fits, to each EGF spectral ratio are plotted in grey. The estimated corner frequency for each event is denoted by the vertical dashed red line. (a)–(b) are considered complex events, while (c)–(d) are considered simple. (e)–(h) Ratios of the spectral ratios from the observed data to the synthetic curves (thin-light red lines) and their median stack (dark red line). The peak-to-peak ratio calculated from 2 Hz to a given frequency is plotted as a solid blue line, the minimum (2 Hz) and maximum (50 Hz) range of frequencies tested are denoted by the black dashed lines. The frequency where the maximum peak-to-peak ratio is first obtained is denoted by blue dashed line. (i)–(l) Grey line is the synthetic omega-square model for the target event based median fc estimate from all EGFs. The red line is the synthesized source spectra created by multiplying omega-square model of the target event by the median of the residual spectra in (e)–(h) and then normalizing it by the target events moment.

is reduced by using multiple EGF's that reduce the bias of one poor EGF biasing the entire result.

3.2 Relative source time functions

Deconvolution of waveforms in the time domain results in the relative moment rate functions of the target event relative to the EGF event at each seismic station, and are commonly referred to as Relative Source Time Functions (RSTF). An event with a simple rupture should have a single pulse RSTF for all stations, with possible pulse width variations due to directivity. If an event contains more than one pulse across multiple stations, this indicates that the rupture could contain multiple subevents and can be referred to as complex (referred to as compound in Wang *et al.* 2014). Due to this possible variability across stations caused by directivity it's difficult to estimate a mean RSTF for the event as a whole so our complexity estimates will be made per station and from those an average will be estimated for the entire event. To estimate RSTF for each target event at each station, we use the EGF based time domain forward modelling approach outlined in Wu *et al.* (2019). This method, based on the work of Kikuchi & Kanamori (1982), assumes that the source time function of the target event can be represented as the superposition of multiple Gaussian pulses that can have varying amplitudes and time lags. We use the method of Wu *et al.* (2019) because it does not limit the number of pulses as opposed to other approaches (e.g. Tan & Helmberger 2010; Wang *et al.* 2014). Instead, it fits multiple pulses iteratively and evaluates their inclusion by the variance reduction in waveform fitting. It also allows for the examination of the moment and need for an individual pulse to replicate the target events waveform. We cannot resolve subevents that are shorter than the sampling resolution of the recording (0.02 s) or EGF duration. This means that for smaller

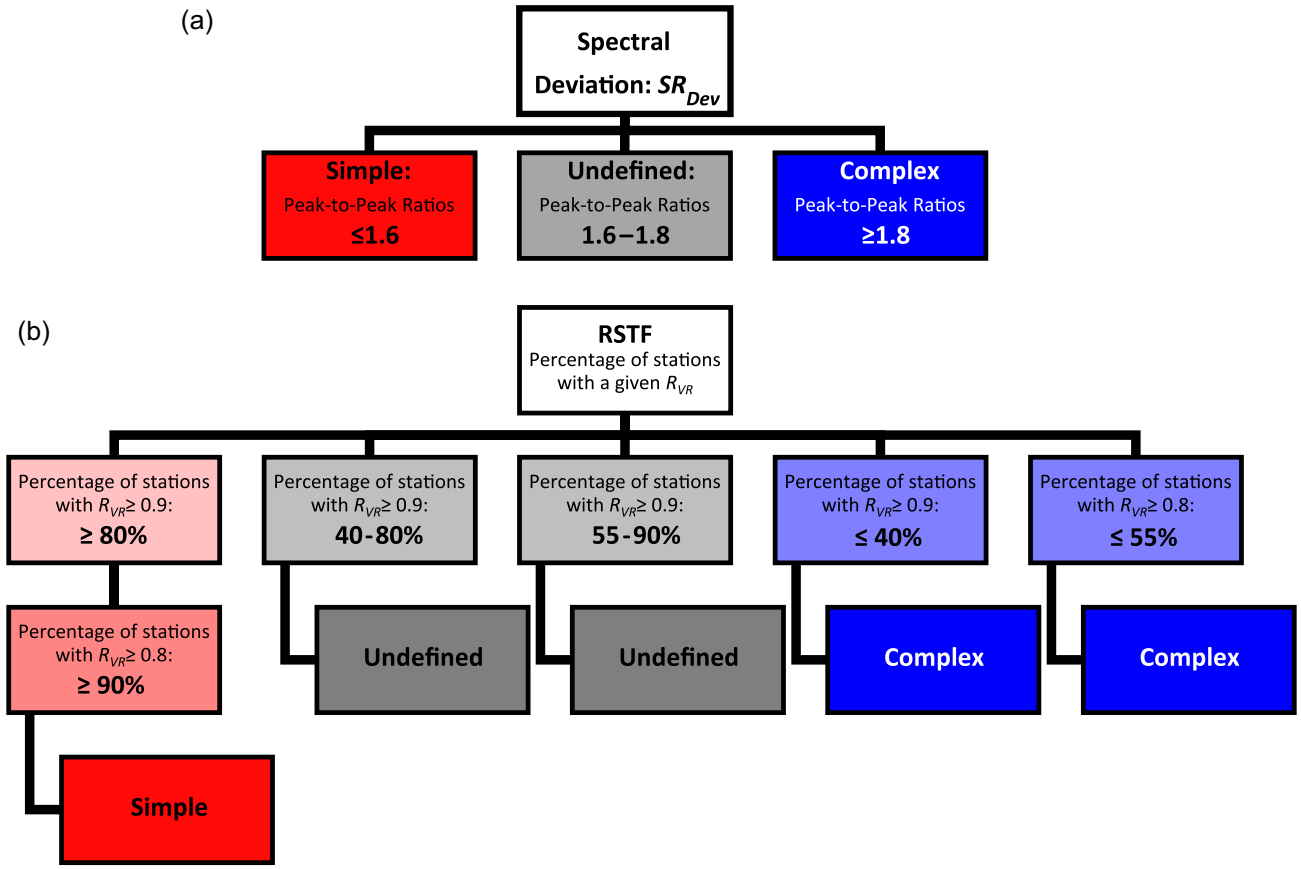

Figure 3. A decision tree showing the criteria for a events classification for the two types of analysis spectral ratios or RSTF. (a) Classification criteria's for spectral ratios, which have been averaged over all stations and EGFs to create a single spectral ratio, using spectral deviation (SR_{Dev}) from the omega-square model measured using peak-to-peak ratios. (b) Classification decision tree for RSTFs per event using the measure of percentage of stations with high (0.9 or 0.8) variance reduction ratio (R_{VR}).

Table 1. Parameters used in the RSTF inversion process.

Magnitude	Filter (Hz)	Window length (s)	Pulse width (s)
>3.5	0.5–20	1.2	0.02–0.4
3–3.5	0.5–20	1	0.02–0.4
2.5–3	0.5–50	0.8	0.02–0.2
<2.5	0.5–50	0.6	0.02–0.2

magnitudes, our ability to resolve subevents is more limited than for larger events.

To avoid overfitting the observed waveforms, we place constraints on the frequency range examined through bandpass filtering, the window length of data and the duration of the resulting RSTF. Due to the varying magnitude range examined in this study, we use different constraints for different magnitude ranges (see Table 1). We estimate the RSTF at each station (referred to as $RSTF_{st}$) for the event pairs examined in the spectral analysis section. During the deconvolution process, pulses are iteratively added with a single pulse being added with each iteration. We use a grid search to determine the optimal timing, amplitude and duration of the subevents at each station. In addition to the constraints in Table 1, we further constrain the pulse widths to a smaller range to prohibit unrealistically broad or narrow pulses. We base this constraint on a theoretical corner frequency estimate from the circular crack model (Eshelby 1957; Brune 1970):

$$\Delta\sigma = \frac{7M_0}{16}\left(\frac{f_c}{k\beta}\right)^3 \quad (4)$$

and using the events seismic moment M_0, a constant shear wave velocity (β) of 3 km/s, constant stress drop ($\Delta\sigma$) of 2 MPa and a k value of 0.38 for P waves. We set the maximum possible pulse width to double the estimated source duration, which is based on the theoretical corner frequency of the target event. We then set the smallest possible pulse width to be half the theoretical source duration estimated from the theoretical corner frequency of the EGF. If either of these estimated values surpass the width allowed by the bandpass filter, we constrain it to that value instead. Due to bandpass filter being used as the minimum constraint the Gaussian pulses are truncated at 99.7 per cent, to avoid contamination with the filter. Therefore, pulse width (pw) is $pw = 3/(\pi * f_c)$. These ranges still can be large, so to stabilize the deconvolution process and reduce the non-uniqueness, we only allow the pulse width to vary within the full range for the first three iterations; for later iterations we constrain it to the range of pulse widths observed in the first three iterations. We find a maximum of 20 iterations to be sufficient to reproduce any events RSTF, and so we limit the number of iterations to 20. Many events converge towards a solution well before this number of iterations and we consider convergence to have been reached if any of the following criteria are meet: (1) The first pulse produces a

variance reduction of ≥80 per cent. (2) The next iteration does not improve the overall variance reduction by more than 1.5 per cent. (3) The relative variance reduction improvement is <2 per cent. We define relative variance reduction as $(VR_n^{\mathrm{stf}} - VR_{n-1}^{\mathrm{stf}})/VR_{n-1}^{\mathrm{stf}} \times 100$ per cent with VR_n^{stf} representing the variance reduction of the nth iteration in the source time function inversion process.

If no limit is placed on the resulting RSTF duration, then the iterative approach can continue adding subevents to model anomalous pulses resulting from noise, or later phase in the waveform. To prevent this, we determine an optimal maximum duration for the $\mathrm{RSTF_{st}}$ for each station for use in the iterative subevent fitting by performing a grid search of durations between 20 per cent and 80 per cent of the time window selected for the data. The $\mathrm{RSTF_{st}}$ is then calculated for each of these possible durations. The longest duration that improves the overall variance reduction by more than 5 per cent is selected as the optimal RSTF duration. The $\mathrm{RSTF_{st}}$ obtained for this duration for that specific station is then used. In this analysis, we only examine those RSTFs that obtain a variance reduction $VR^{\mathrm{stf}} \geq 70$ per cent between the synthetic and observed waveforms.

To measure the complexity of a $\mathrm{RSTF_{st}}$, we use the variance reduction ratio (R_{VR}) defined as:

$$R_{VR} = VR_{1\mathrm{st}}^{\mathrm{stf}}/VR_{\mathrm{final}}^{\mathrm{stf}}, \tag{5}$$

where $VR_{1\mathrm{st}}$ is the variance reduction of the first iteration and VR_{final} is the variance reduction of the final iteration. This measure provides an overall quantitative estimate of whether an $\mathrm{RSTF_{st}}$ can be represented by a single pulse, or whether multiple pulses are needed. We measure R_{VR} for each $\mathrm{RSTF_{st}}$ meaning that for each target event, we have independent measurements for each EGF event at each station that meets the SNR and fitting criteria.

For each target event we have multiple EGF events, and we perform the deconvolution process separately on each pair at each individual station to obtain multiple estimates of the RSTF. However, due to emergent arrivals or opposite polarities, some EGF events will perform well on a few stations but not all—an issue that does not affect the spectra examination but affects the STF analysis. The inclusion of these poor to okay EGF events can significantly alter the averaged $\mathrm{RSTF_{st}}$ across EGFs for a given station, and ultimately how its rupture is characterized. To reduce these effects, we must identify and use EGFs that we consider to best resolve the rupture of the target event and do so across as many stations as possible. Therefore, we first select only those EGFs that have the most $\mathrm{RSTF_{st}}$ (minimum ≥ 5) to ensure a matching radiation pattern to the target event. This selection is further reduced to only by those EGFs whose mean $\mathrm{RSTF_{st}}$ variance reduction is within 5 per cent of the best variance reduction EGF. This is to ensure that we only use the EGFs whose path and site effects are similar to those of the target. We allow target events that have a single EGF that pass selection to be analysed for complexity.

If we have multiple EGFs, we align and stack their respective $\mathrm{RSTF_{st}}$ at each station. We restrict this so that only those $\mathrm{RSTF_{st}}$ have a cross-correlation coefficient ≥ 0.85 (in reference to the highest variance reduction result that meet the previous criteria) are aligned and then averaged to make a single $\mathrm{RSTF_{st}}$ at the station. This is done to avoid the inclusion of an anomalous RSTF. The R_{VR} of the averaged $\mathrm{RSTF_{st}}$ at a station is the mean R_{VR} from the EGFs used to create the averaged $\mathrm{RSTF_{st}}$. Using these selection criteria, we obtain more RSTF complexity estimates than the spectral deviation analysis that required a minimum of six EGFs, as we only require a single EGF that passes the above criteria. This lower threshold for the RSTF analysis is because the variance reduction estimated for RSTFs is directly linked to how well the observed and synthetic waveforms match, and provides a quantitative measure of the quality of RSTFs. In the RSTFs analysis, a complex EGF would violate the point source assumption of EGF selection and leads to poorly fitted synthetic waveforms using the variance reduction criteria. In the spectral analysis, it is difficult to separate if the complexity in spectra is due to the chosen EGF or is a feature of the target event. The only way to be sure is if the complexity is featured across multiple EGFs, which is why we require more EGFs for the spectral analysis.

Now that we have estimates of complexity and we have averaged those results with multiple EGFs, we classify a target event as simple or complex. To do this, we obtain the percentage of $\mathrm{RSTF_{st}}$ s for a target event with $R_{\mathrm{VR}} \geq 0.9$ (R_{VR90}) and/or $R_{\mathrm{VR}} \geq 0.8$ (R_{VR80}) out of all $\mathrm{RSTF_{st}}$s (stations). We use Fig. 4 as an example. Event 21428897 and 30500786 (Figs 4a and b) require multiple pulses in their $\mathrm{RSTF_{st}}$s to replicate the target events waveforms for many of their stations. Due to this, only two stations SMNB and EADB for event 21428897 and MMNB and CCRB for 30500786 have a rating of R_{VR90} or R_{VR80}. In contrast to these two events, for the events 71450425 and 30500784 (Figs 4c and d), almost all of their stations require only a single pulse to represent the source in their RSTFs. For this reason, almost all of their stations have ratings of R_{VR90}, with only JCNB of event 30500784 having a rating of R_{VR80}.

We use the following criteria that was refined through the visual inspection of results to classify the event as simple or complex (a visual representation of this is shown in Fig. 3b):

(i) Simple events: the percentage of $R_{\mathrm{VR90}} \geq 80$ per cent and $R_{\mathrm{VR80}} \geq 90$ per cent

(ii) Complex events: the percentage of $R_{\mathrm{VR90}} \leq 40$ per cent or $R_{\mathrm{VR80}} \leq 55$ per cent

Events that do not meet these criteria are considered unclassified, which represent ∼31 per cent of events and ∼41 per cent of large magnitude events (Supplemental Fig. S2 shows the distribution of classified events). The results of the RSTF analysis for the same four events from Fig. 2 are plotted in Fig. 4. Events 21428897 (Fig. 4a) and 71450425 (Fig. 4c) have also been examined by Abercrombie *et al.* (2020), who found similar azimuth variations in duration for the groups of events these two earthquakes occur in which they term Sequence A1 (contains 21428897) and Sequence A3 (contains 71450425). For these events, we obtain the same directivity direction and observations of complex and simple STFs as Abercrombie *et al.* (2020) supporting our procedures.

4 RESULTS AND RESOLUTION

From our spectral ratio analysis, we obtain results for 337 events ($>M$ 1.1), among them, 186 events are simple, 85 events are complex and 66 events are unclassified. From the RSTF analysis, we obtain results for 480 events ($>M$ 1.1), among them, 231 are simple, 100 are complex and 149 are unclassified. 290 common events are found between the two methods. Of these common events the 119 simple events from RSTF analysis, 100 are classified as spectrally simple and 19 are classified spectrally complex. Of the 49 complex events from RSTF analysis, 28 are classified spectrally complex. This disagreement is reduced when only the better resolved larger events ($M > 2$) that are within the network (Fig. 1 black box), and so recorded from a full azimuthal range are considered. Restricting the analysis to these events within the network significantly reduces the disagreement, so that only 3 out of 20 RSTF complex events are

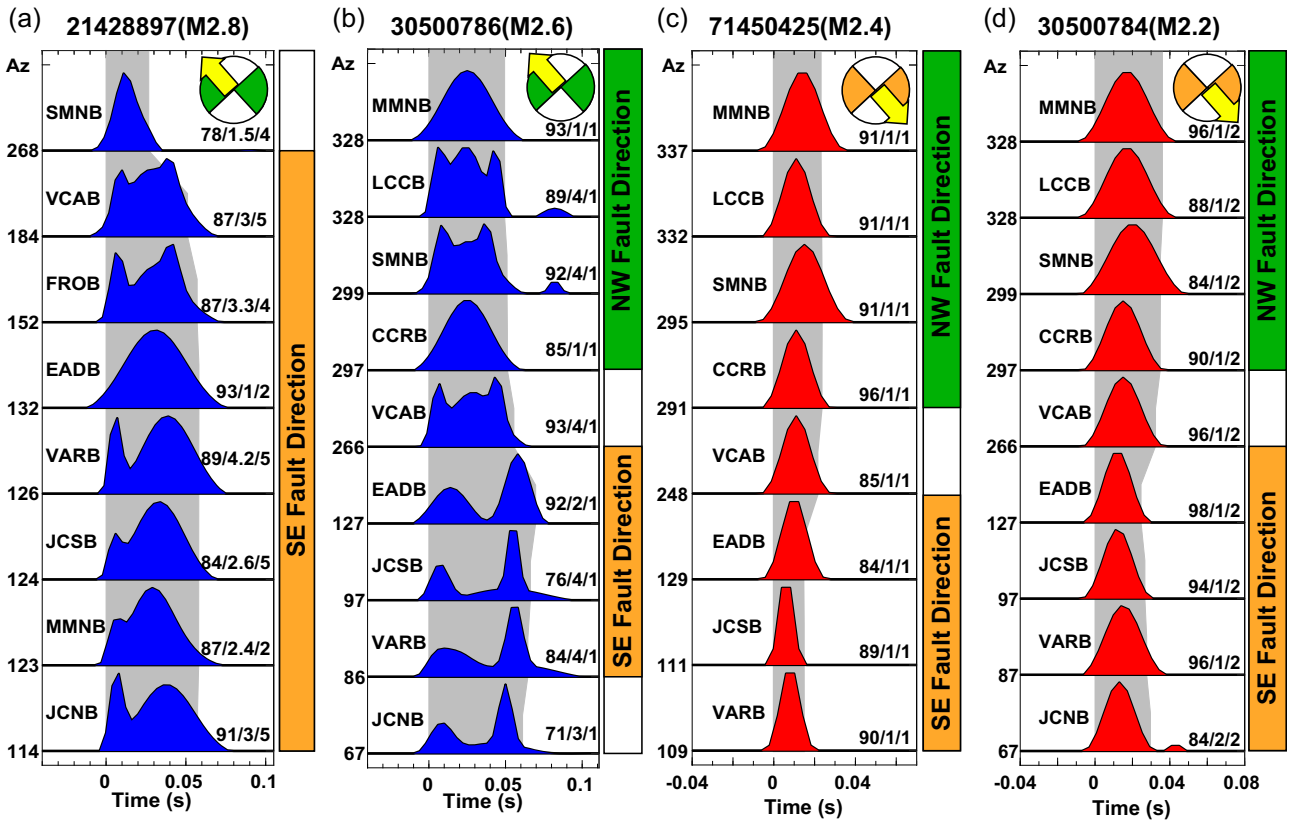

Figure 4. RSTF's sorted by azimuth that were obtained for the same four events shown in Fig. 2 with RSTF to the SE and NW of hypocentre denoted and rupture directivity denoted on the focal mechanism of the fault plane that best aligns with the San Andreas Fault. (a)–(b) are defined as complex and (c)–(d) defined as simple. Values to the right of each RSTF is the average variance reduction/average number of pulses/number of EGFs stacked in the RSTF for the stacked event pairs. The text to the left of the RSTF in each box denotes the station. The right vertical axis denotes how many individual event pairs were stacked to create the final RSTF for that station. The left vertical axis is the azimuth of the RSTF. The grey region outlines the duration of the RSTFs to emphasize azimuthal variations and are based upon manually picked RSTF durations.

classified as spectrally simple; and only 2 out of the 48 RSTF simple are classified as spectrally complex. The reason for this stark change with magnitude has to do with the resolution of the two methods.

The primary factor that influences our ability to resolve complexity at different magnitudes is the available frequency bandwidth. The higher frequency the data, the smaller scale features we can observe. This effect can be clearly seen when examining the frequency where the peak-to-peak ratio of deviation reaches 1.8 ($PtP_{1.8f}$), which increases as the target events magnitude decreases (Fig. 5a). At magnitudes of 2.1, $PtP_{1.8f}$ reaches the limit of our observable frequency range, and then remains close to this boundary. The frequency where peak-to-peak ratio reaches 1.4 ($PtP_{1.4f}$) follows a similar trend as $PtP_{1.8f}$, but remains lower in frequency than the frequency limit. This indicates that $PtP_{1.8f}$ for lower magnitudes is beyond our frequency limit and our resolution range, so the capability to classify an event as complex decreases with decreasing magnitude for spectral analysis. This limit of resolution around M 2.1 events has been observed by other spectral analysis studies (e.g. Abercrombie 2014; Abercrombie *et al.* 2020) and finite-fault inversions (e.g. Dreger *et al.* 2007) for events in this area.

Our resolvable pulse width range has a similar resolution problem. The theoretical pulse width of the EGF event falls below our observable range of 0.02 s due to the requirement of five samples needed to represent a Gaussian pulse (Fig. 5b). We are still able to resolve complexity for the target event, until we pass M 2.0–2.2 range where the target events pulse width is now just 0.03 s larger in width than our observable range. In Fig. 5(b) we assume that all of our events have a constant stress drop of 2 MPa. If we assume a higher stress drop of 10 MPa (Zhang *et al.* 2022), then the predicted f_c increases and the predicted width of the target events pulses would be even narrower.

For the events that meet the resolution thresholds ($M > 2$), we obtain agreement of the classification of the two methods of 85 per cent for the complex events and 95 per cent for the simple events. This high level of agreement implies that the two methods are identifying the same source of complexity. The source of the possible small disagreement could be that for an individual subevent to be resolved in a RSTF, its radiated energy must be responsible for over 10–20 per cent of the waveform on a majority of the stations for it to be classified as complex. If it does not produce observable complexity in the waveform, then it will not be present in the RSTF model. Due to these reasons, the RSTF will only capture the complexity of events with strong abrupt changes in slip history. For these reasons, the classification of complex events can be considered conservative, since the complexity we are identifying in this paper is solely related to multiple-phase failures. For this reason, we do not consider those events with single slip pulses that show directivity as complex events.

Based upon the resolution limitations of each method, we restrict our analysis of spatiotemporal patterns to just $M > 2$ events that fall within the network (outlined by the black box in Fig. 1).

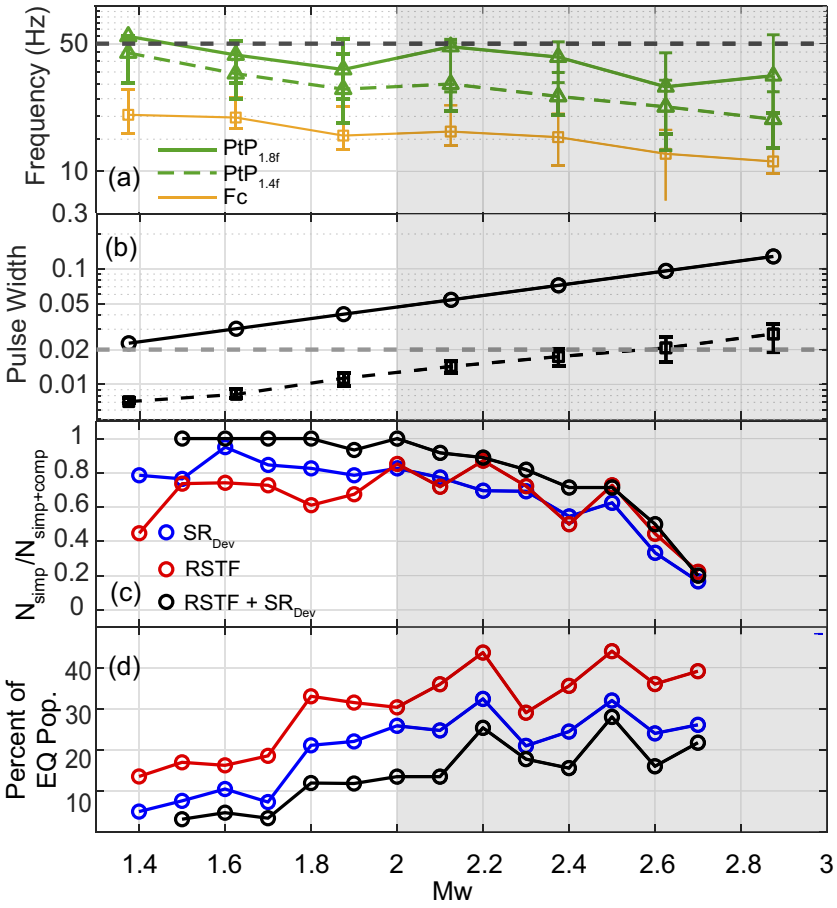

Figure 5. The limits of our methods resolution is roughly M 2–2.2 and above this magnitude (grey region) is where we observe a majority of the complex event. (a) Resolution of detecting complexity in spectra for different magnitude bins. Corner frequency estimates obtained from spectral ratios (yellow line), frequency where peak-to-peak ratio reaches 1.4 (green dashed line), and 1.8 (green line) in relation to the limit of fitted frequency range at 50 Hz (black dashed line). 80th and 20th percentiles of each magnitude bin are denoted by error bars. The figure shows that at magnitude 2.1 we can no longer observe accurately where a peak-to-peak ratio of 1.4 occurs. (b) Pulse width resolution limits. Theoretical pulse width of target events in magnitude bin (black line) and of EGFs (dashed black line) used for those target events based upon theoretical corner frequency obtained from event magnitude, assuming a stress drop of 2 MPa and rupture velocity of 3 km/s. The smallest pulse we use is 0.02 (grey dashed line), but theoretically we could be using a smaller width if noise level allowed for it. However, it does not mean that theoretically we are missing complexity that the EGF could deconvolve to a certain degree below magnitudes 2.4. (c) Ratio of simple to simple + complex estimates for the different methods between magnitudes 1.4 to 2.8 with a interval of 0.1 (see legend for method). (d) Percentage of earthquake population within the network (Fig. 1 black box) that have a estimate of complexity (simple or complex) for each 0.1 magnitude bin between the magnitudes of 1.4 to 2.8.

5 DISCUSSION

5.1 Complexity from spectral and RSTF analyses

When we examine the spatial distributions of the earthquakes classified as simple or complex, the patterns resulting from the frequency and time domain methods appear to show different styles of event occurring more often in certain segments of the fault system (Figs 6a–b). The non-common events between the two methods are the cause of the most divergence between results. This can be due to an event being unclassified by either the time domain (RSTF) or frequency domain (spectral) method (e.g. Fig. 7). One reason for this lack of classification could be rupture directivity. Both the duration of the RSTF and any delay between separate pulses of moment release will be shorter, at stations in the forward direction of rupture, potentially allowing them to be represented by a single Gaussian pulse within our resolution limits. For the stations in the direction opposite to the rupture direction, the reverse will be true making it easier to resolve complexity in the RSTF on those stations. Depending on the network distribution and the event location within it, there could be bias in the observations of complexity, making it hard to accurately classify it, and therefore it might remain unclassified. An example of this can be found in Fig. 7, with only three out of the eight stations showing complexity between the azimuth ranges of 210–90 deg. This is the prime example of why we require the events we examine to be within the network. These stations indicate there might be multiple subevents of this event, but there are simply not enough stations showing such behaviour to definitively classify it, so our chosen criterion leaves it as unclassified. Also, if the pulses are overlapped or not well separated in time, the deconvolution process will result in a single large pulse depending on the resolution of the deconvolution, which could be occurring for some of stations in Fig. 7.

The spectral deviations we observe in our frequency domain results can have multiple causes, but their agreement with those obtained by the RSTF results for $M > 2.0$ implies that the variation

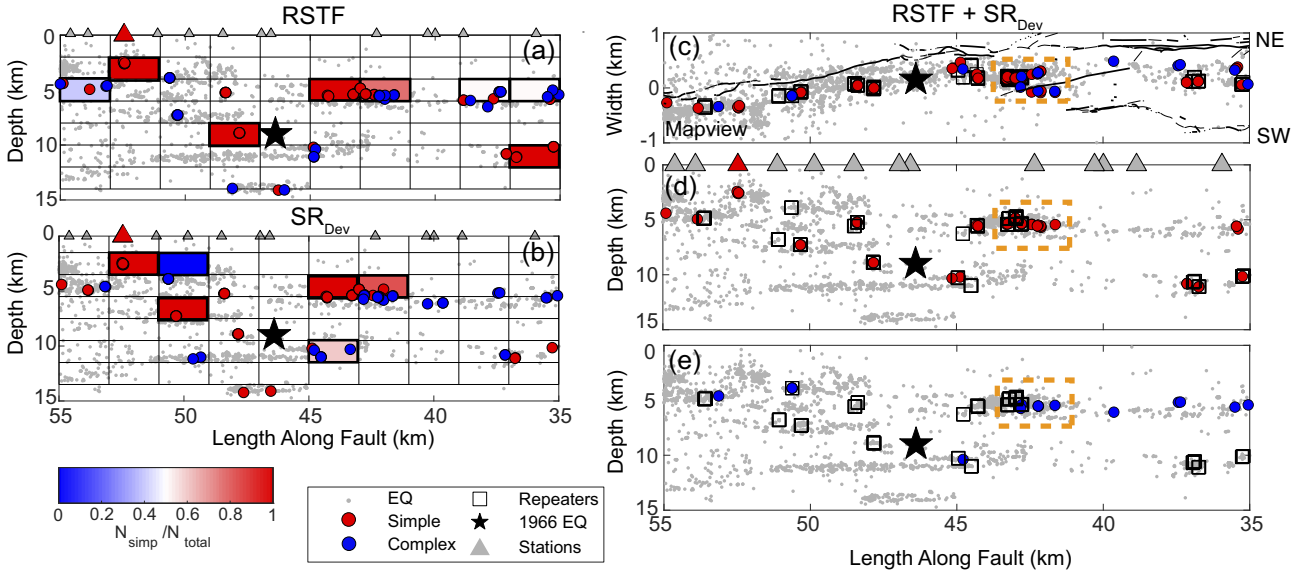

Figure 6. Cross-section view across A–B shown in Fig. 1 of classified earthquakes from (a) RSTF ($M > 2.0$), SRDev ($M > 2.0$) (b), and those whose classification agrees across both (c)–(e). In (a)–(b) in order to better represent areas with many events grids with five or more classified events are filled in according to the ratio of simple events to the total number of classified events (e.g. if there are more complex events the grid will be blue or if there are more simple events the grid will be red), and individual events are plotted as coloured dots. The common results across both methods are shown in (c)–(e). Both simple and complex classified events are present in map view in (c), (d) Cross-section A–B with common across method simple classified events and (e) the same as (d) but for complex classified events. All earthquakes in the relocated catalogue are shown as grey dots. The 1966 M6 earthquake (black star), and stations used in the analysis are plotted (grey triangles) with SAFOD also being shown (red triangle) repeating events from Rubinstein *et al.* (2012) are shown as squares. The cluster of complex events discussed in the paper is outlined by the orange dashed box.

from the omega-square model represents variation in the rupture, probably indicating heterogeneities in fault properties and applied stress (e.g. Koyama 1994). Our findings support those of the previous observations of similar spectral deviations found for larger events M 3.2–4.0 by Uchide & Imanishi (2016) in the Hamadori region of Japan. The RSTF results also support the findings of events with multiple subevents by Wang *et al.* (2014) and Fischer (2005). The observations of these other studies support the conclusion that this behaviour is not unique to the Parkfield region, or even the San Andreas Fault system.

At magnitude ranges higher than M 2.6, we are able to resolve 20–40 per cent of all catalogue events within the network, and complex events make up 60–80 per cent of these resolved events depending on method (Figs 5c–d). This high percentage of complex events indicates that small magnitude events that are routinely modelled as circular ruptures exhibit significant levels of complexity. This percentage of simple to complex events is similar to those resolved by Wang *et al.* (2014) for parts of northern San Andreas and part of the Calaveras faults. Fischer (2005) observed 61 per cent events being complex for $M > 2.2$ in an intraplate swarm in Central Europe. For that same magnitude range, we only observe 15 per cent of events being complex, due to a lot of the smaller magnitudes being predominately simple. This could indicate a difference between interplate and intraplate regions ruptures. It's important to note that how Wang *et al.* (2014) and Fischer (2005) define complex events differs from ours. Wang *et al.* (2014) defines an event as complex (referred to as 'compound' in their paper) if two Gaussian pulses produce a significantly large variance reduction compared to one pulse. Fischer (2005) examined RSTFs and identified complex events (referred to as 'Multiple-events' in their paper) as those which showed multiple pulses in their RSTFs.

Comparing our results to large events the result of 60–80 per cent of events M 2.6–3.5 at Parkfield being complex is around the level observed for ($M > 8$) events around the world (Danré *et al.* 2019). The observation of high percentage of complex events for M 2–3 events strongly suggest that the assumption that small earthquakes can be described with a kinematic model for a circular fault (Brune 1970) or the quasi-dynamic model (Madariaga 1976; Kaneko & Shearer 2014) is not always a reliable assumption. The complexity of these small events can introduce artefacts and increase uncertainty and bias in stress drop measurements (e.g. Abercrombie 2021), and could explain the broad range of stress drops found between studies for the same sequence (e.g. Pennington *et al.* 2020). One way to address this problem is to remove these events and focus on simple events, well-fit by circular source models (Ruhl *et al.* 2017; Yoshimitsu *et al.* 2019), however, this too could lead to selection bias and scaling artefacts.

5.2 Possible causes of complexity

The complexity that manifests both in earthquake source spectra and in the time domain RSTFs may be caused by a number of factors, including whether the rupture is assumed to be continuous or segmented. Due to the small magnitude of these events, it could be assumed that the ruptures are continuous. Dynamic modelling of closely located repeating events at Parkfield performed by Wen *et al.* (2018) found that simple spectra and RSTFs are associated with self-arresting ruptures and the complex ones with runaway ruptures. In these types of ruptures, the factors that control the complexity in the STF at each station are the directivity effect and rupture propagation (Wen *et al.* 2018). For the complexity in spectra certain factors have been shown to affect the first large deviation (e.g. hole) in the spectra. These deviations in spectra are related to the interference of stopping phases (Madariaga 1976). Its frequency is a function of the radius of the circular fault/fault

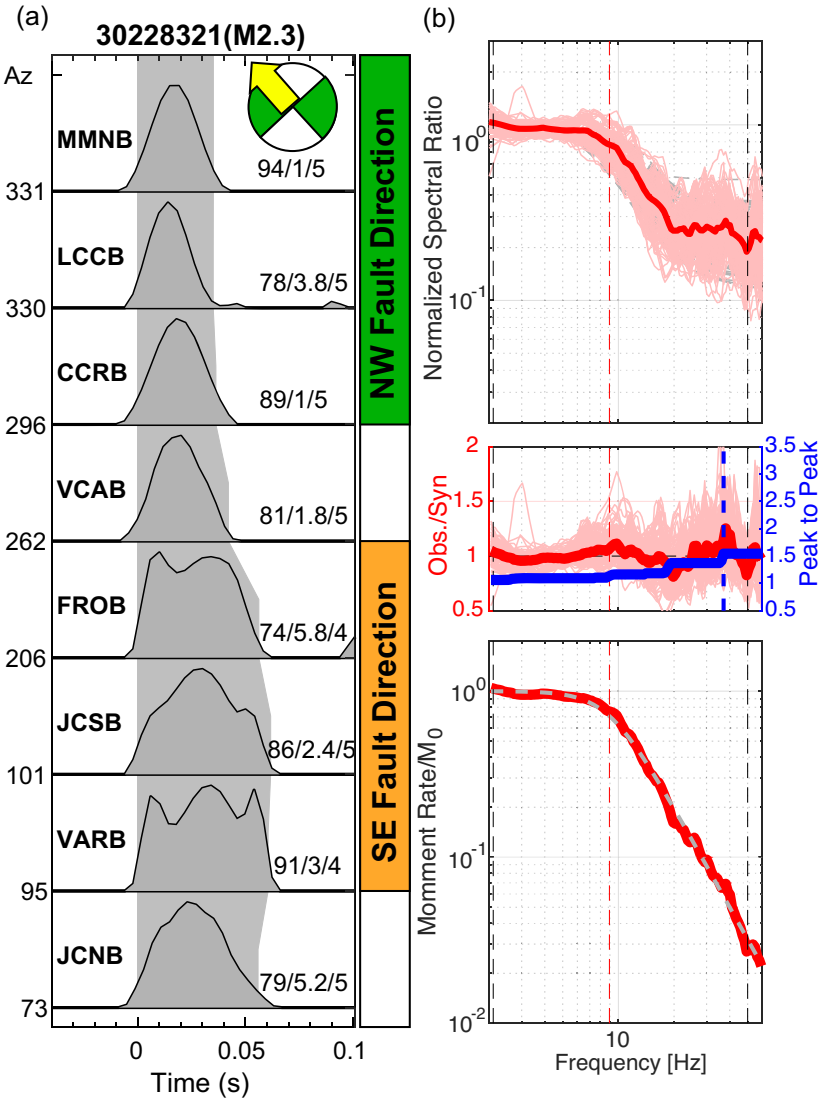

Figure 7. Example of an unclassified RSTF event which has three RSTFs with multiple pulses and four with a single pulse. One azimuth being dominant by complex RSTFs indicates it could be complex, but since the majority of stations are simple our criteria cannot place it in either. (a) RSTFs for the event. The text to the left of the RSTF in each box denotes the station. Values to the right of each RSTF is the average variance reduction/average number of pulses/number of EGFs stacked in the RSTF for the stacked event pairs. The left vertical axis is the azimuth of the RSTF. The grey region outlines the duration of the RSTFs to emphasize azimuthal variations and are based upon manually picked RSTF durations. (b) Spectral analysis for the event (see description in Fig. 2). The spectral analysis has the event classified as simple with a peak to peak of 1.5, which is just under the threshold for being placed in the unclassified group for the spectral results.

length, the direction of rupture propagation and the rupture velocity (Ben-Menahem 1961; Madariaga 1976; Wen *et al.* 2018).

Alternatively, if we assume the ruptures are not continuous, then there could be other possible causes. This is also possible since some of the complex events occur in structurally complex regions of the fault, making it possible that their ruptures could be occurring in areas of heterogeneous stress or strength distributions. This type of complexity is explained by a number of models such as the asperity model (Rudnicki & Kanamori 1981), a model in which stress drop is very heterogeneous with patches of high stress drop (asperities), and patches which were already broken during previous events and have zero stress drop. Another model is the Das & Aki (1977) barrier model in which the rupture leaves unbroken patches on the fault plane. These unbroken patches are interpreted to be high strength and do not slip during the rupture and stress drop is negative. Fundamentally all these models agree that the rupture is heterogeneous spatially. Depending on when slip occurs during these heterogeneous ruptures, they can produce the multiple peaks in the STF. In these cases, the corner frequency will no long reflect the total duration, but the widths of each of the slip pulses with the spectra being dominated by the simplest common pulse (Das & Aki 1977; Madariaga 1979). The frequency of the spectral hole is controlled by the delay between the pulses, and their durations, which vary between stations depending on the directivity and rupture velocity (Madariaga 1981).

5.3 Spatial patterns

The common events with the same classification from the two methods that meet our resolution criteria provide the most stable set of results to interpret (Figs 6c–e). As we are prevented by resolution

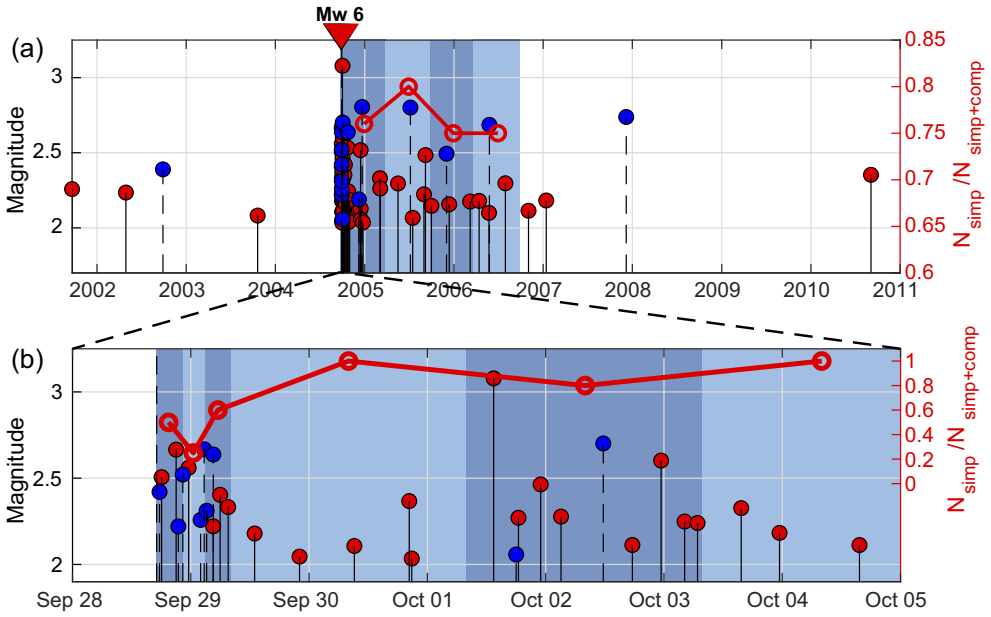

Figure 8. Temporal distribution of common simple and classified events that occur within the study region and the ratio of simple to all classified events for specified time bins. Stems show the magnitude of each event. (a) Temporal patterns of entire study time period. Stem plot with events coloured red if classified as simple and blue if complex. The solid red line is the ratio of simple events to all classified events with dots located at the centre of half year time period bins. Time bins are coloured in alternating blue and light blue. (b) Time period that focuses on 1 day before to 7 day after the 2006 M_w 6 earthquake. Time period bins of classified event ratio vary with those in the hours following the Mainshock being 4 hours in duration and later bins being 2 days in length.

of the data from including the smallest events, we are only able to obtain a very limited picture of the spatial complexity distribution, and consequently we keep our interpretations appropriately limited. A region that is dominated by simple events is the segment of fault north of the 1966 *M* 6 earthquake (area left of black star Figs 6d–e). This region of the fault is observed to have high amounts of creep compared to the segment of the fault to the south of it. The simple events in this section primarily occur in repeating earthquake groups (Rubinstein *et al.* 2012); this consistent style of rupture suggests that these events are triggering the same fault patches, and represent single asperity failures. The consistent nature of the RSTFs indicates that the asperities are relatively isolated and do not trigger any other fault patches nearby, supporting interpretations of these asperities as velocity weakening fault patches embedded in a velocity strengthening matrix suggested by past studies (e.g. Vidale *et al.* 1994; Chen & Lapusta 2009).

The region of the fault with the highest concentration of complex events is a cluster 2.5 km to the south of the 1966 M6 earthquake (Fig. 6 orange box c–e). Previous examinations of this area interpreted high levels of shear and normal stress concentration within this region due to the fault orientation and relatively complex structure (Perrin *et al.* 2019). This cluster also contains a large number of repeating earthquakes, indicating that some of these events are consistently occurring on the same fault patches, so a portion of them could represent single asperity failures as in the northern segment (Rubinstein *et al.* 2012). The complex events within this cluster occur abundantly within the hours following the 2004 M_w 6 earthquake (Fig. 8). A possible cause of this could be the concentration of a large amount of stress in this structurally complex area, which then could allow for the triggering of multiple slip patches or the changing of self-arresting ruptures to runaway ruptures. This concentration of stresses after 2004 M_w 6 is supported by the finding that this cluster borders a region of high post-seismic creep of 70 cm to its SW, but the cluster itself has a creep rate of 20 cm (Barbot *et al.* 2009).

This increase in stress being the driver behind the increase in complex events is supported by the observations that a change in loading rates in lab experiments has been found to effect the amount of coupling along a fault (Mclaskey & Yamashita 2017). These variations in seismic coupling due to differing loading conditions can affect the moment (Chen *et al.* 2010; Uchida *et al.* 2015) and stress drop (McLaskey *et al.* 2012; Abercrombie 2014), which were found by studies of repeating earthquake sequences in response to afterslip. This change in coupling could also change the number of contacts that have the potential to be triggered and therefore the asperities triggered in a rupture (Chaves *et al.* 2020). These changes in loading rate can cause sequences to spontaneously begin and end, and would explain the sudden abundance and then absence of complex events in the hours following the 2004 M_w 6 earthquakes.

Based upon these observations, the clearest conclusions that we can draw are that structurally complex regions (e.g. intersection of faults) of the fault system appear to host more complex events. This is likely due to these regions having a higher degree of heterogeneity in both stress and strength which would facilitate complex source processes to be more common.

6 CONCLUSIONS

Analysis of 480 earthquakes at Parkfield using time and frequency domain approaches confirms that rupture complexity occurs in roughly 60–80 per cent (*M* 2.6–3) of these resolved events depending on the method choosen.

(i) We find that a combination of methods is best for confirming and quantifying complexity and that these methods find the same complexity.

(ii) The methods are fundamentally limited by the EGF event duration and instrument limitations. This limitation means that the best resolved events are $M > 2.0$.

(iii) The complex events we observe predominantly are located in areas of structural complexity where heterogeneity in stress and strength are likely to be observed.

This complexity can be observed both in the slip history of the event and in its source spectra. This has important implications for the study of source processes of earthquakes, since it reveals that use of simple circular rupture model for small events might not be an accurate assumption. What controls the spatial pattern of this complexity appears to be related to geologic and structural features observed by past studies. As well as changes in loading rates due to the 2006 M_w 6 earthquake. Future research should focus on better linking frequency and time domain analysis of complexity and understanding how geologic or specific rupture processes manifest in each of them.

ACKNOWLEDGMENTS

Funding for this work was provided by Oklahoma state funding for induced earthquake research, and funding from NSF Award EAR-1547071 and EAR-1547083 and SCECAwards 18086 and 19163; SCEC is funded by NSF Cooperative AGREEMENT EAR-1033462 and USGS Cooperative Agreement G12AC20038. Portions of this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This is LLNL Contribution Number LLNL-JRNL-840866.

DATA AVAILABILITY

The Berkeley Parkfield High Resolution Seismic Network (HRSN) seismograms are archived at the Northern California Earthquake Data Center (NCEDC) under network code BP (Northern California Earthquake Data Center 2014) and can be accessed at http://service.ncedc.org/fdsnws/dataselect/1/. The Seismic Analysis Code (SAC) v. 101.6 (http://ds.iris.edu/files/sac-manual) and ObsPy v. 1.0 (Beyreuther *et al.* 2010) were used to process the data. Figures were made using Matplotlib v. 2.0.0 (Hunter 2007), and M_MAP (Pawlowicz 2020). All websites were last accessed in 2021 October unless otherwise noted.

REFERENCES

Abercrombie, R.E., 1995. Earthquake source scaling relationships from -1 to 5 M L using seismograms recorded at 2.5-km depth. *J. geophys. Res. Solid Earth,* **100,** 24 015–24 036.

Abercrombie, R.E., 2014. Stress drops of repeating earthquakes on the San Andreas Fault at Parkfield. *Geophys. Res. Lett.,* **41,** 8784–8791.

Abercrombie, R.E., 2021. Resolution and uncertainties in estimates of earthquake stress drop and energy release. *Phil. Trans. R. Soc. A Math. Phys. Eng. Sci.,* **379**. doi:10.1098/rsta.2020.0131

Abercrombie, R.E., Chen, X. & Zhang, J., 2020. Repeating earthquakes with remarkably repeatable ruptures on the San Andreas Fault at Parkfield. *Geophys. Res. Lett.,* **47**. doi:10.1029/2020GL089820

Allmann, B.P. & Shearer, P.M., 2009. Global variations of stress drop for moderate to large earthquakes. *J. geophys. Res.,* **114,** B01310. doi:10.1029/2008JB005821

Atkinson, G.M., 1996. The high-frequency shape of the source spectrum for earthquakes in eastern and western Canada. *Bull. seism. Soc. Am.,* **86,** 106–112.

Atkinson, G.M. & Silva, W., 1997. An empirical study of earthquake source spectra for California earthquakes. *Bull. seism. Soc. Am.,* **87,** 97–113.

Barbot, S., Fialko, Y. & Bock, Y., 2009. Postseismic deformation due to the *M* w 6.0 2004 Parkfield earthquake: stress-driven creep on a fault with spatially variable rate-and-state friction parameters. *J. geophys. Res.,* **114,** B07405. doi:10.1029/2008JB005748

Ben-Menahem, A., 1961. Radiation of seismic surface-waves from finite moving sources. *Bull. seism. Soc. Am.,* **51,** 401–435.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T. & Behr, Y., 2010. *Seismol. Res. Lett.,* **81,** 530–533.

Boatwright, J., 1978. Detailed spectral analysis of two small New York State earthquakes. *Bull. seism. Soc. Am.,* **68,** 1117–1131.

Boatwright, J., Fletcher, J.B. & Fumal, T.E., 1991. AA general inversion scheme for source, site, and propagation characteristics using multiply recorded sets of moderate-sized earthquakes. *Bull. seism. Soc. Am.,* **81,** 1754–1782.

Böse, M., Heaton, T.H. & Hauksson, E., 2012. Real-time finite fault rupture detector (FinDer) for large earthquakes. *Geophys. J. Int.,* **191,** 803–812.

Brune, J.N., 1970. Tectonic stress and the spectra of seismic shear waves from earthquakes. *J. geophys. Res.,* **75,** 4997–5009.

Chaves, E.J., Schwartz, S.Y. & Abercrombie, R.E., 2020. Repeating earthquakes record fault weakening and healing in areas of megathrust postseismic slip. *Sci. Adv.,* **6,** 2–10.

Chen, K.H., Bürgmann, R., Nadeau, R.M., Chen, T. & Lapusta, N., 2010. Postseismic variations in seismic moment and recurrence interval of repeating earthquakes. *Earth planet. Sci. Lett.,* **299,** 118–125

Chen, T. & Lapusta, N., 2009. Scaling of small repeating earthquakes explained by interaction of seismic and aseismic slip in a rate and state fault model. *J. Geophys. Res. Earth,* **114,** 12. doi:B01311 10.1029/2008jb005749

Choy, G.L. & Boatwright, J., 1995. Global patterns of radiated seismic energy and apparent stress. *J. geophys. Res.,* **100,** 18 205–18 228.

Danré, P., Yin, J., Lipovsky, B.P. & Denolle, M.A., 2019. Earthquakes Within Earthquakes: patterns in Rupture Complexity. *Geophys. Res. Lett.,* **46,** 7352–7360.

Das, S. & Aki, K., 1977. Fault plane with barriers: a versatile earthquake model. *J. geophys. Res.,* **82,** 5658–5670.

Dreger, D., Nadeau, R.M. & Chung, A., 2007. Repeating earthquake finite source models: strong asperities revealed on the San Andreas Fault. *Geophys. Res. Lett.,* **34**. doi:10.1029/2007GL031353

Eshelby, J.D., 1957. The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Phil. Trans. R. Soc. Lond. A.,* **241,** 376–396.

Fan, W. & McGuire, J.J., 2018. Investigating microearthquake finite source attributes with IRIS community wavefield demonstration experiment in Oklahoma. *Geophys. J. Int.,* **214,** 1072–1087.

Fischer, T., 2005. Modelling of multiple events using empirical Green's functions: method, application to swarm earthquakes and implications for their rupture propagation. *Geophys. J. Int.,* **163,** 991–1005.

Gallovič, F., Zahradník, J., Křížová, D., Plicka, V., Sokos, E., Serpetsidaki, A. & Tselentis, G.-A., 2009. From earthquake centroid to spatial-temporal rupture evolution: mw 6.3 Movri Mountain earthquake, June 8, 2008, Greece. *Geophys. Res. Lett.,* **36,** L21310. doi:10.1029/2009GL040283

Hartzell, S.H. & Heaton, T.H., 1983. Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake. *Bull. seism. Soc. Am.,* **73,** 1553–1583.

Hartzell, S.H. & Heaton, T.H., 1986. Rupture history of the 1984 Morgan Hill, California, earthquake from the inversion of strong motion records. *Bull. seism. Soc. Am.,* **76,** 649–674.

Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.,* **9,** 90–95, doi:10.1109/MCSE.2007.55

Ide, S, Beroza, G., Prejean, S. & Ellsworth, W., 2003. Apparent break in earthquake scaling due to path and site effects on deep borehole recordings. *J. geophys. Res.,* **108,** 2271. doi:10.1029/2001JB001617

Ide, S. & Beroza, G.C., 2001. Does apparent stress vary with earthquake size? *Geophys. Res. Lett.,* **28,** 3349–3352.

Ide, Satoshi., 2001. Complex source processes and the interaction of moderate earthquakes during the earthquake swarm in the Hida-Mountains, Japan, 1998. *Tectonophysics,* **334,** 35–54.

Kaneko, Y., Shearer P. M., 2014. Seismic source spectra and estimated stress drop derived from cohesive-zone models of circular subshear rupture, *Geophysical Journal International,* **197**(), 1002–1015. https://doi.org/10.1093/gji/ggu030.

Kikuchi, M. & Kanamori, H., 1982. Inversion of complex body waves. *Bull. seism. Soc. Am.,* **72,** 491–506.

Koyama, J., 1994. General description of the complex faulting process and some empirical relations in seismology. *J. Phys. Earth,* **42,** 103–148.

Li, Z. & Peng, Z., 2016. An automatic phase picker for local earthquakes with predetermined locations: combining a signal-to-noise ratio detector with 1D velocity model inversion. *Seism. Res. Lett.,* **87,** 1397–1405. doi:10.1785/0220160027

Madariaga, R., 1976. Dynamics of an expanding circular fault. *Bull. seism. Soc. Am.,* **66,** 639–666.

Madariaga, R., 1979. On the relation between seismic moment and stress drop in the presence of stress and strength heterogeneity. *J. geophys. Res.,* **84,** 2243. doi:10.1029/JB084iB05p02243

Madariaga, R., 1981. Dynamics of seismic sources, in *Identification of Seismic Sources—Earthquake or Underground Explosion,* pp. 71–96, eds, Husebye, E. S. & Mykkeltveit, S., Springer Netherlands. doi:10.1007/978-94-009-8531-5_3

Mai, P.M. & Thingbaijam, K.K.S., 2014. SRCMOD: an online database of finite-fault rupture models. *Seismol. Res. Lett.,* **85,** 1348–1357.

Mcgarr, A., 1999. On relating apparent stress to the stress causing earthquake fault slip. *J. geophys. Res.,* **104,** 3003–3011.

McGuire, J.J., 2017. A MATLAB toolbox for estimating the second moments of earthquake ruptures. *Seismol. Res. Lett.,* **88,** 371–378.

McGuire, J.J. & Kaneko, Y., 2018. Directly estimating earthquake rupture area using second moments to reduce the uncertainty in stress drop. *Geophys. J. Int.,* **214,** 2224–2235.

McGuire, J.L., Zhao, L. & Jordan, T.H., 2002. Predominance of unilateral rupture for a global catalog of large earthquake. *Bull. seism. Soc. Am.,* **92,** 3309–3317.

McLaskey, G.C., Thomas, A.M., Glaser, S.D. & Nadeau, R.M., 2012. Fault healing promotes high-frequency earthquakes in laboratory experiments and on natural faults. *Nature,* **491,** 101–104.

Mclaskey, G.C. & Yamashita, F., 2017. Slow and fast ruptures on a laboratory fault controlled by loading characteristics. *J. geophys. Res. Solid Earth,* **122,** 3719–3738.

Meng, H., McGuire, J.J. & Ben-Zion, Y., 2020. Semiautomated estimates of directivity and related source properties of small to moderate Southern California earthquakes using second seismic moments. *J. geophys. Res. Solid Earth,* **125,** 1–21.

Northern California Earthquake Data Center., 2014. *High Resolution Seismic Network (HRSN),* Northern California Earthquake Data Center. doi:10.7932/HRSN

Pawlowicz, R., 2020. M_Map: a mapping package for MATLAB. Last Accessed October 1, 2022, Retrieved from https://www.eoas.ubc.ca/~rich/map.html.

Pennington, C., Chen, X., Abercrombie, R.E. & Wu, Q., 2020. Cross validation of stress drop estimates and interpretations for the 2011 Prague, OK, earthquake sequence using multiple methods. *J. geophys. Res. Solid Earth.,* **126,** doi: 10.1029/2020JB020888

Pennington, C.N., Chang, H., Rubinstein, J.L., Abercrombie, R.E., Nakata, N., Uchide, T. & Cochran, E.S., 2022. Quantifying the sensitivity of microearthquake slip inversions to station distribution using a dense nodal array. *Bull. seism. Soc. Am.,* **112,** 1252–1270. doi:10.1785/0120210279

Perrin, C., Waldhauser, F., Choi, E. & Scholz, C.H., 2019. Persistent fine-scale fault structure and rupture development: a new twist in the Parkfield, California, story. *Earth planet. Sci. Lett.,* **521,** 128–138.

Rubinstein, J.L., Ellsworth, W.L., Chen, K.H. & Uchida, N., 2012. Fixed recurrence and slip models better predict earthquake behaviour than the time-and slip-predictable models: 1. Repeating earthquakes. *J. geophys. Res. Solid Earth,* **117,** 1–23.

Rudnicki, J.W. & Kanamori, H., 1981. Effects of fault interaction on moment, stress drop, and strain energy release. *J. geophys. Res. Solid Earth,* **86,** 1785–1793.

Ruhl, C., Abercrombie, R.E. & Smith, K.D., 2017. Spatiotemporal variation of stress drop during the 2008 Mogul, Nevada, Earthquake Swarm. *J. geophys. Res. Solid Earth,* **122,** 8163–8180. doi:10.1002/2017JB014601

Schneider, J.F., Silva, W.J. & Stark, C., 1993. Ground motion model for the 1989 *M* 6.9 Loma Prieta earthquake including effects of source, path, and site. *Earthq. Spectra.,* **9,** 251–287, doi:10.1193/1.1585715

Shearer, P.M., Abercrombie, R.E., Trugman, D.T. & Wang, W., 2019. Comparing EGF methods for estimating corner frequency and stress drop from *P* wave spectra. *J. geophys. Res. Solid Earth,* **124,** 3966–3986.

Silva, W.J., Wong, I.G. & Darragh, R.B., 1998, Engineering characterization of earthquake strong ground motions in the Pacific Northwest, in *US Geol. Surv. Prof. Pap,* U.S. Geological Survey.

Tan, Y. & Helmberger, D., 2010. Rupture directivity characteristics of the 2003 big bear sequence. *Bull. seism. Soc. Am.,* **100,** 1089–1106.

U.S. Geological Survey and California Geological Survey., U.S. Geological Survey 2006. Quaternary Fault and Fold Database of the United States. https://doi.org/10.5066/F7S75FJM.

Uchida, N., Shimamura, K., Matsuzawa, T. & Okada, T., 2015. Postseismic response of repeating earthquakes around the 2011 Tohoku-oki earthquake: moment increases due to the fast loading rate. *J. geophys. Res. Solid Earth,* **120,** 259–274.

Uchide, T. & Ide, S., 2010. Scaling of earthquake rupture growth in the Parkfield area: self-similar growth and suppression by the finite seismogenic layer. *J. geophys. Res. Solid Earth,* **115,** 1–15.

Uchide, T. & Imanishi, K., 2016. Small earthquakes deviate from the omega-square model as revealed by multiple spectral ratio analysis. *Bull. seism. Soc. Am.,* **106,** 1357–1363.

Vallée, M. & Douet, V., 2016. A new database of source time functions (STFs) extracted from the SCARDEC method. *Phys. Earth planet. Inter.,* **257,** 149–157.

Venkataraman, A., Beroza, G., Ide, S., Imanishi, K., Ito, H. & Iio, Y., 2006. Measurements of spectral similarity for microearthquakes in western Nagano, Japan , *J. geophys. Res.,* **111**. doi:10.1029/2005JB003834

Vidale, J.E., Ellsworth, W.L., Cole, A. & Marone, C., 1994. Variations in rupture process with recurrence interval in a repeated small earthquake. *Nature,* **368,** 624–626.

Waldhauser, F., 2009. Near-real-time double-difference event location using long-term seismic archives, with application to northern California. *Bull. seism. Soc. Am.,* **99,** 2736–2748, doi:10.1785/0120080294

Waldhauser, F. & Schaff, D.P., 2008. Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods. *J. geophys. Res. Solid Earth.,* **113,** doi:10.1029/2007JB005479

Wang, E., Rubin, A.M. & Ampuero, J.-P., 2014. Compound earthquakes on a bimaterial interface and implications for rupture mechanics. *Geophys. J. Int.,* **197,** 1138–1153.

Wen, J., Chen, X. & Xu, J., 2018. A dynamic explanation for the ruptures of repeating earthquakes on the San Andreas fault at Parkfield. *Geophys. Res. Lett.,* **45,** 116–122.

Wu, Q., Chen, X. & Abercrombie, R.E., 2019. Source complexity of the 2015 *M*w 4.0 Guthrie, Oklahoma earthquake. *Geophys. Res. Lett.,* **46,** 4674–4684.

Yamada, T., Mori, J.J., Ide, S., Kawakata, H., Iio, Y. & Ogasawara, H., 2005. Radiation efficiency and apparent stress of small earthquakes in a South African gold mine. *J. geophys. Res. Solid Earth,* **110,** 1–18.

Yoshimitsu, N., Ellsworth, W.L. & Beroza, G.C., 2019. Robust stress drop estimates of potentially induced earthquakes in Oklahoma: evaluation of empirical Green's function. *J. geophys. Res. Solid Earth,* **124,** 5854–5866.

Zhang, J., Chen, X. & Abercrombie, R.E., 2022. Spatiotemporal variability of earthquake source parameters at Parkfield, California, and their rela-

tionship with the 2004 M6 earthquake. *J. geophys. Res. Solid Earth,* **127,** 1–22.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Peak-to-peak ratio estimates and the break down of simple and complex events. (a) Histogram of peak-to-peak ratio estimates for all earthquakes and those for the large events. Estimates indicating simple events fall in the red region ($\leq$1.6), undefined in the grey region (1.6–1.8), and complex events in the blue region ($\geq$1.8). (b) Magnitude versus peak-to-peak ratio with regions that denote simple (red), undefined (grey) and complex (blue) coloured similar to (a). With increasing magnitude the number of complex events increases mainly due to improving resolution that captures complexity. A trend of increasing peak-to-peak ratios can be seen with magnitudes $>$ 2.2. (c) Max peak-to-peak ratio versus the frequency that the ratio is recorded at. Most of the observed complex events observe their peak-to-peak ratio before reaching the resolution limit of 50 Hz.

Figure S2. Ratio of stations with variance reduction ratios (R_{VR}) $\geq$ 0.9 (R_{VR90}) or $\geq$ 0.8 (R_{VR80}) estimates and the break down of simple and complex events. (a) Histogram of the ratio of stations with variance reduction ratios $\geq$ 0.9 (R_{VR90}) for all earthquakes and large earthquakes. Estimates indicating simple events fall in the red region ($\geq$ 80 per cent of stations), undefined in the grey region (40–80 per cent), and complex events in the blue region ($\leq$40 per cent of stations). For all earthquakes simple events dominate, but this is due to resolution limits, since when examining just large events we see a fairly uniform distribution. (b) Histogram of the ratio of stations with variance reduction ratios $\geq$ 0.8 (R_{VR80}) for all earthquakes and large earthquakes. Estimates indicating simple events fall in the red region ($\geq$90 per cent of stations), undefined in the grey region (55–90 per cent), and complex events in the red region ($\leq$55 per cent of stations). (c) Magnitude versus the ratio of stations with variance reduction ratios $\geq$ 0.9 (R_{VR90}). You can observe a increasing number of simple events with decreasing magnitude, which is an artefact of resolution limitations.

Figure S3. Peak-to-peak ratio estimates compared to corner frequencies for different magnitude ranges. Simple events fall in the red region ($\leq$1.6), undefined in the grey region (1.6–1.8), and complex events in the blue region ($\geq$1.8). The error bars for each measurement is the 90th and 10th percentiles of distribution of corner frequencies estimated for all EGFs for each target event. The estimate is centred at the mean of all the EGF fc estimates. The main observation to draw from this is that corner frequency decreasing with increasing complexity. This is likely due to the fitted model fitting the large deviation in the spectra which causes it to find a lower corner frequency. Examples of this phenomenon can be found in Supplemental Fig. S4. Another observation is that the complex events have smaller amounts of error. This could be due to the corner frequency being lower and the spectra being sampled logarithmically, which would increase the error for higher corner frequencies.

Figure S4. (left-hand panel) RSTF's sorted by azimuth that were obtained for two events with RSTF to the SE and NW of hypocentre denoted and rupture directivity denoted on the focal mechanism of the fault plane that best aligns with the San Andreas Fault. Values to the right of each RSTF is the average variance reduction/average number of pulses/number of EGFs stacked in the RSTF for the stacked event pairs. The text to the left of the RSTF in each box denotes the station. The left vertical axis is the azimuth of the station that the RSTF is from. The grey region outlines the duration of the RSTFs to emphasize azimuthal variations and starts at the 25th percentile and end at the 75th percentile of the RSTFs distribution. (Right-hand panel) The spectral ratio for the same EGF event used to make the RSTF for the same station. (a) Is defined as complex and (b) defined as simple. The two important observations to be made from this figure is that for (a) the spectral 'hole' decrease in frequency with increasing delay time between pulses. The second major observation is for the simple event (b). The high frequency content increases (increasing fc) with decreasing pulse width of the RSTF. Both these observations show that the spectra is reflecting the observations of the RSTF.

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.