

Detecting Repeating Earthquakes on the San Andreas Fault with Unsupervised Machine Learning of Spectrograms

Theresa Sawi^{*1}, Felix Waldhauser¹, Benjamin K. Holtzman¹, and Nate Groebner²

Abstract

Repeating earthquakes—sequences of colocated, quasi-periodic earthquakes of similar size—are widespread along California’s San Andreas fault (SAF) system. Catalogs of repeating earthquakes are vital for studying earthquake source processes, fault properties, and improving seismic hazard models. Here, we introduce an unsupervised machine learning-based method for detecting repeating earthquake sequences (RES) to expand existing RES catalogs or to perform initial, exploratory searches. We implement the “SpecUFEx” algorithm (Holtzman *et al.*, 2018) to reduce earthquake spectrograms into low-dimensional, characteristic fingerprints, and apply hierarchical clustering to group similar fingerprints together independent of location, allowing for a global search for potential RES throughout the data set. We then relocate the potential RES and subject them to the same detection criteria as Waldhauser and Schaff (2021). We apply our method to ~4000 small (M_L 0–3.5) earthquakes located on a 10 km long segment of the creeping SAF and double the number of detected RES, allowing for greater spatial coverage of slip-rate estimations at seismogenic depths. Our method is novel in its ability to detect RES independent of initial locations and is complimentary to existing cross-correlation-based methods, leading to more complete RES catalogs and a better understanding of slip rates at depth.

Cite this article as Sawi, T., F. Waldhauser, B. K. Holtzman, and N. Groebner (2023). Detecting Repeating Earthquakes on the San Andreas Fault with Unsupervised Machine Learning of Spectrograms, *The Seismic Record*, **3**(4), 376–384, doi: [10.1785/0320230033](https://doi.org/10.1785/0320230033).

Supplemental Material

Introduction

Groups of small magnitude ($M_L \leq 3.5$) earthquakes with highly similar waveforms, locations, and magnitudes that occur in a periodic or quasi-periodic fashion were first observed on the San Andreas fault (SAF) decades ago (Geller and Mueller, 1980; Poupinet *et al.*, 1984; Vidale *et al.*, 1994; Nadeau *et al.*, 1995). Each repeating earthquake sequence (RES) is presumably generated by near-identical failures on an asperity embedded within a region of aseismic creep (Nadeau and Johnson, 1998; Dreger *et al.*, 2007; Uchida, 2019). Some RES are included in long lived, horizontal streaks of seismicity, suggesting that they represent asperities from different rock strata intersecting the fault zone, or that they delineate regions of creep from the locked sections of the fault where larger magnitude earthquakes occur (Rubin *et al.*, 1999; Schaff *et al.*, 2002; Waldhauser *et al.*,

2004). Highly periodic RES can be used to estimate source parameters and fault properties for small earthquakes (Vidale *et al.*, 1994; Nadeau and Johnson, 1998; Peng *et al.*, 2005), and infer slip rates at depth (Nadeau and McEvilly, 1999; Bürgmann *et al.*, 2000; Templeton *et al.*, 2008; Shakibay and Funning, 2019; Waldhauser and Schaff, 2021, hereafter referred to as WS2021).

1. Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, U.S.A., <https://orcid.org/0000-0001-5938-6743> (TS); <https://orcid.org/0000-0002-1286-9737> (FW); <https://orcid.org/0000-0002-6226-7689> (BKH); 2. Strabo Analytics, Inc., New York, New York, U.S.A., <https://orcid.org/0000-0002-6535-1739> (NG)

*Corresponding author: tsawi@ldeo.columbia.edu

© 2023. The Authors. This is an open access article distributed under the terms of the CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

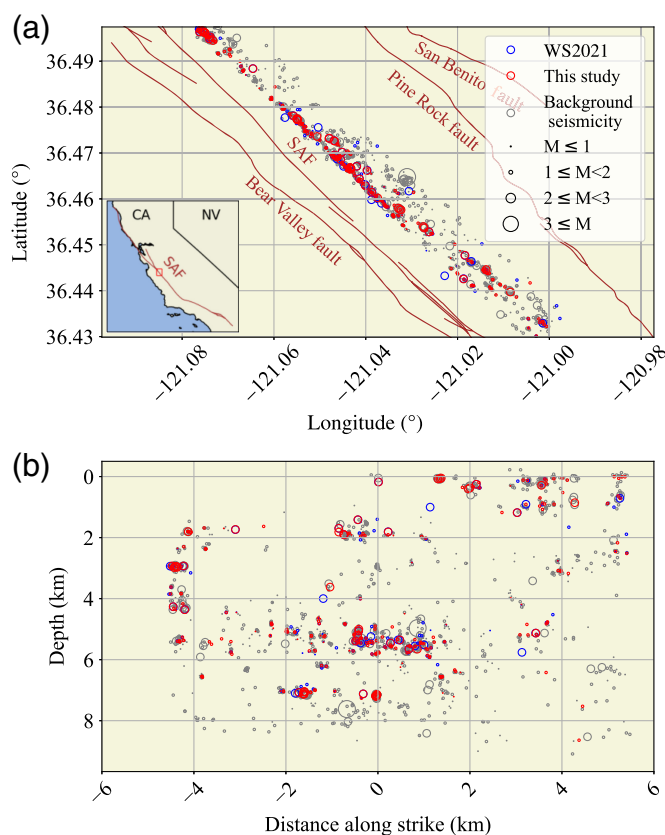


Figure 1. (a) Map view of repeating earthquake sequences (RES) from Waldhauser and Schaff (2021) (blue circles) and newly detected or updated RES from this study (red circles) along a 10 km segment of the San Andreas fault (SAF) in central California (inset). The gray circles are background seismicity (1984–2019). The brown lines are fault surface traces. (b) Depths of earthquakes along strike, colors the same as in panel (a). Each RES sequence is represented by one circle.

The successes of many of these studies depend on the relative completeness of RES catalogs, because missing repeating earthquakes in a given sequence can erroneously extend recurrence intervals, and missing sequences can degrade observational spatial resolution. To detect RES, most studies analyze cross-correlation (CC) coefficients between seismograms of earthquakes observed at common stations (Uchida and Bürgmann, 2019). However, at times, repeating earthquakes may not generate waveforms with high CC coefficients. Colocated events, for example, may rupture in different directions (Abercrombie *et al.*, 2020) and generate dissimilar waveforms. In addition, the introduction of noise to seismograms can reduce the similarity between waveforms. Multiple studies have found that the range of CC coefficient thresholds used to detect RES is highly inconsistent and, as a

sole criterion, is often insufficient to determine true RES (Uchida, 2019; Gao *et al.*, 2021; WS2021).

Overcoming these obstacles requires reliable, high-resolution earthquake catalogs, and strict detection criteria regarding colocation, size, and waveform similarity of events, such as those used in the recent studies. WS2021 developed a two-step procedure to compile a comprehensive RES catalog for northern California. In the first step, using the double-difference earthquake catalog of Waldhauser and Schaff (2008), and the underlying database of billions of *P*- and *S*-wave CC coefficients between pairs of seismograms, clusters of potentially repeating earthquakes with high-waveform similarities are formed. In a second step, actual RES are confirmed if a detailed analysis of the relative location and size of earthquakes in each cluster shows that they rupture the same (inferred) asperity. WS2021 is the most comprehensive catalog of RES in California to date, and yet, through this study, we discover that it most likely underestimates the number of RES in the area.

Here, we introduce a novel method to contribute to step one—the identification of potentially repeating earthquakes. Rather than relying on a database of CC coefficients, which is expensive to compute on a large scale, we compare spectral characteristics of the seismic waves using unsupervised machine learning. Many seismological studies have benefited from the recent advances in supervised machine learning that require labeled data sets (e.g., reviews by Kong *et al.*, 2018; Mousavi and Beroza, 2022). In this study, we use an unsupervised machine learning method applied to spectrograms (Holtzman *et al.*, 2018) to detect potentially repeating earthquakes without identifying waveform templates or calculating waveform correlation coefficients beforehand. When applied to a portion of the SAF north of Parkfield, California, we find about twice as many, when missing only 20%, of the repeaters compared with the WS2021. We demonstrate the advantages of complementing RES detection and selection methods with those aided by unsupervised machine learning, and discuss the resultant increased spatial resolution in RES-based slip-rate estimates.

Data

Our study region is a 10 km long segment of the SAF about 80 km north of Parkfield positioned between the locked portion of the fault to the northwest and the creeping section in the southeast (Fig. 1). We select 3968 events from the updated (1984–2019) double-difference (DD) catalog of Waldhauser and Schaff (2008) that are within 500 m from the main lineation of seismicity associated with the SAF. The local magnitudes

(M_L) of the events are between 0 and 3.5, and relative location errors are 10 m on average. Of these events, 3481 occurred between 1984 and 2014—the period covered by WS2021, which identifies 1024 repeating earthquakes divided among 145 sequences at depths up to 8 km (blue circles in Fig. 1).

To find potentially repeating earthquakes (step one), we choose the seven seismic stations that have been continuously operating since 1984 and are located within 25 km of the study area. We acquire 20 s long vertical-component waveforms from the Northern California Earthquake Data Center centered on the P -wave arrival for the 3968 events recorded at each station when available.

Methods

The workflow (Fig. 2) involves converting waveforms to spectrograms, to which we apply unsupervised feature extraction and then hierarchical clustering, followed by analysis of several selection criteria to determine true RES.

Spectral Unsupervised Feature Extraction (SpecUFEx)

We convert the waveforms into spectrograms by calculating short-time Fourier transforms (STFTs) applied to 1 s Tukey windows with a shape parameter of 0.25 and 99% overlap between adjacent windows. We trim the spectrograms to 10 s by keeping 1 s before the P -wave arrival and 9 s after it, long enough to capture both the P and S waves while minimizing the effect of coda waves. We apply a boxcar filter from 2 to 17 Hz to each spectrogram to reduce noise, and then, following Holtzman *et al.* (2018), we demean the spectrograms by dividing them by their median amplitude. We convert the spectrograms to decibels by multiplying the base-ten logarithm of the spectrogram by 20 and setting all negative values to zero to comply with the nonnegativity constraint of nonnegative matrix factorization (NMF).

We reduce the spectrograms to low-dimensionality “fingerprints” that retain the time-varying spectral information using the SpecUFEx algorithm (Fig. 2a; Holtzman *et al.*, 2018). The first step of SpecUFEx is NMF, wherein each spectrogram is decomposed into the inner product of an activation coefficient matrix (ACM) that is unique for each spectrogram, and a dictionary shared by all spectrograms. The ACM is composed of spectral patterns through time, creating a lower dimensionality representation of the spectrogram. The ACMs are then used as the inputs for the hidden Markov model (HMM), which learns how spectral patterns in the ACMs cooccur in time. Concurrent

spectral patterns are expressed as time-varying states in the state emissions matrix, then, transitions between states are counted to form a fingerprint matrix of dimension $\dim(T, T)$ in which T , the number of states in the HMM, is a user-defined parameter. As in Holtzman *et al.* (2018), we choose $T = 15$ states, although slight deviations from this number do not significantly change overall results. Unlike waveforms and spectrograms, fingerprints are not time dependent, because they are based on counts of state transitions over one time step. Consequently, fingerprints need not be aligned in time prior to computing similarity. For a more thorough description of the algorithm, see Holtzman *et al.* (2018).

Hierarchical Clustering to Find Potential RES

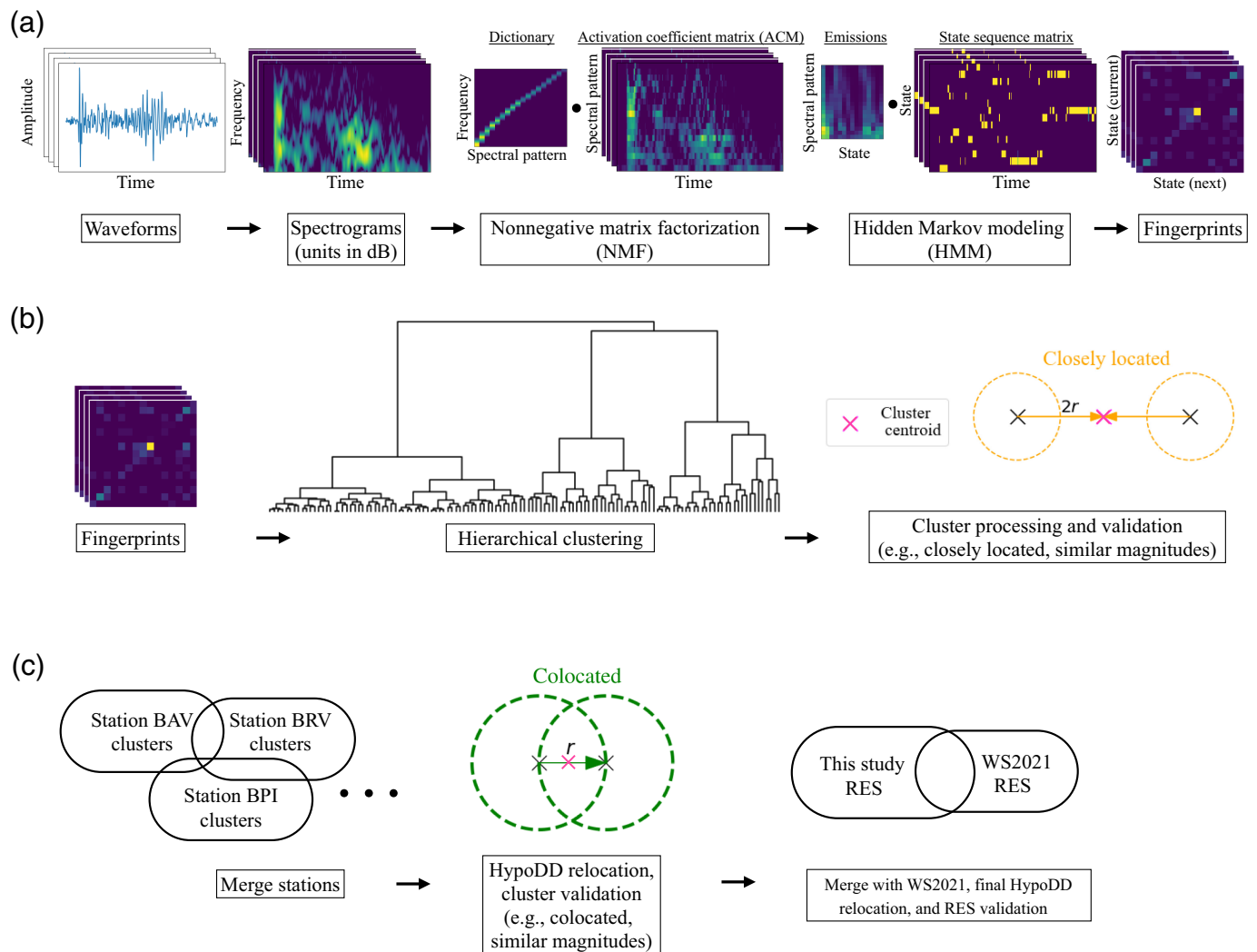
For each station, we cluster the fingerprints using hierarchical clustering applied to an $N \times N$ Euclidean distance matrix between N fingerprints (Fig. 2b). Initially, each fingerprint is assumed to be its own singleton cluster. Then, the nearest singleton clusters are merged such that a linkage function (in this case Ward linkage, the variance between clusters) is minimized. This process is repeated until all the clusters are merged, or, in our case, until a user-defined depth threshold is reached. We choose a depth threshold that maximizes the number of potentially repeating earthquakes.

Although the hierarchical clustering is distance independent, we screen potentially repeating earthquake clusters by their locations. We search for clusters of potentially repeating earthquakes with hypocenters separated from the cluster centroid by not more than two times the estimated rupture radius of the largest event. We assume the circular rupture model of Brune (1970) with a radius (R , in km) defined as follows:

$$R = 0.01 \left(\frac{7 \times 10^{1.5M_{\max}}}{16\Delta\sigma} \right)^{1/3}, \quad (1)$$

in which $M_{\max}$ is the maximum magnitude of all clustered events and assuming a stress drop, $\Delta\sigma$, of 3 MPa. We also impose a magnitude similarity constraint that all events in a cluster must be within 0.3 magnitude units of the cluster’s median magnitude (Waldhauser and Ellsworth, 2002). Uncertainties in relative locations and magnitudes are accounted for when the selection parameters are applied to each cluster.

We apply two processing steps to the potential RES that reduce the sensitivity of results to the hierarchical clustering depth threshold, and empirically lead to more complete and robust potential RES. First, if an earthquake is closely located



to a potential RES, is within 0.3-unit magnitude of the median magnitude, and is not already in another RES, we cross-correlate the 10 s waveform with that of the largest magnitude earthquake in the RES (since the largest magnitude earthquake should have the highest signal-to-noise ratio). If the pair achieves a CC coefficient higher than 0.9, the event is added to the potential RES. In addition, we calculate the median of the CC coefficients between all earthquakes in the RES. If the median increases when we add the event, then we add it to the RES. If it does not increase, or the largest magnitude-pair CC coefficient is less than 0.9, then the event is not added to the RES.

Second, we merge potential RES if the distance between the centroids of two sequences is less than the radius of the circular rupture area (equation 1) based on the median magnitude of all events in both the sequences. And, if the difference between the median magnitudes between the two sequences is less than 0.3

Figure 2. Data processing and analysis workflows. (a) At each station, waveforms are converted to spectrograms and serve as input to the Spectral Unsupervised Feature Extraction (SpecUFEx) program. SpecUFEx uses two layers of unsupervised feature extraction (nonnegative matrix factorization [NMF] and hidden Markov model [HMM]) to generate fingerprints for each spectrogram. (b) Hierarchical clustering is applied to the fingerprints, with depth of clustering (i.e., the number of clusters) chosen to optimize the number of potential RES. Sequence processing involves adding closely located earthquakes (within two rupture radii from cluster centroid) to potential RES and merging closely located RES waveforms for which are highly correlated (cross correlation [CC] > 0.9) and with magnitudes within 0.3 units of the median magnitude. (c) Potential RES catalogs for each station are merged, combining sequences with shared events. Each potential RES is relocated, and evaluated for colocation and similar magnitudes, resulting in an actual RES catalog. This actual RES catalog is merged with that from WS2021 to produce an updated RES catalog after a final analysis of each merged sequence to ensure colocation (all hypocenters within half rupture radii).

magnitude units, then we calculate CC on the 10 s waveforms of the largest amplitude events in each potential RES. If the CC coefficient is greater than 0.95, then we combine the potential sequences. We repeat this process for the seven stations and then concatenate all catalogs, merging potential RES from different stations if they share at least one similar event (Fig. 2c).

Identification of Actual RES

Each potential RES found by our unsupervised machine learning approach is individually analyzed to confirm true RES following WS2021. We relocate the events in each potential RES using *P*- and *S*-wave correlation delay times based on 1 s windows and CC coefficients >0.7 between all events and all stations that recorded the events and compute least-squares relative location errors using the double-difference algorithm *HypoDD* (Waldhauser and Ellsworth, 2000; Waldhauser, 2001). We only keep colocated events in a sequence, that is, those that are located within each other's circular rupture area, accounting for both uncertainty in the magnitudes (± 0.1 magnitude unit) and uncertainty in the relative location estimated from the covariance matrix, choosing a lower bound of 10 m to acknowledge possible error underestimation. The resulting sequences are then combined with the WS2021 catalog and after a final relocation analysis and manual inspection of each potential RES a catalog of actual RES is established (see [Data and Resources](#)). Note that, to combine this study's RES with those of WS2021, we had to adjust the depths in the WS2021 catalog to the Geoid because the U.S. Geological Survey started reporting Geoid-referenced depths after the WS2021 study. This reanalysis resulted in removing 10 RES (mostly small-magnitude doublets) of the 144 original WS2021 RES, because they no longer pass the colocation criteria following relocation.

Slip-Rate Estimation

When selecting RES for slip-rate estimation, we follow WS2021 and discard "burst-type" events, which occur too closely in time, that is, less than 30 days apart, assuming that they are aftershocks and not actually reflective of the background creep rate. In addition, we only keep RES start and end dates, which are before or after a time span of three times the average recurrence time of the sequence from the beginning and end of the study period, respectively (WS2021). For these quasi-periodically repeating sequences (RESp) with three or more events, we calculate slip rates by first obtaining the slip (*S*) for each event (Nadeau and Johnson, 1998) with the modified static correction term from WS2021:

$$S = 10^{-2.46} M_0^{0.17}, \quad (2)$$

in which earthquake moment, M_0 , is calculated from the earthquake magnitude (M_L) by the following equation:

$$M_0 = 10^{1.2M_L + 17}. \quad (3)$$

We then calculate the average slip rate for each sequence containing *N* events with the following equation:

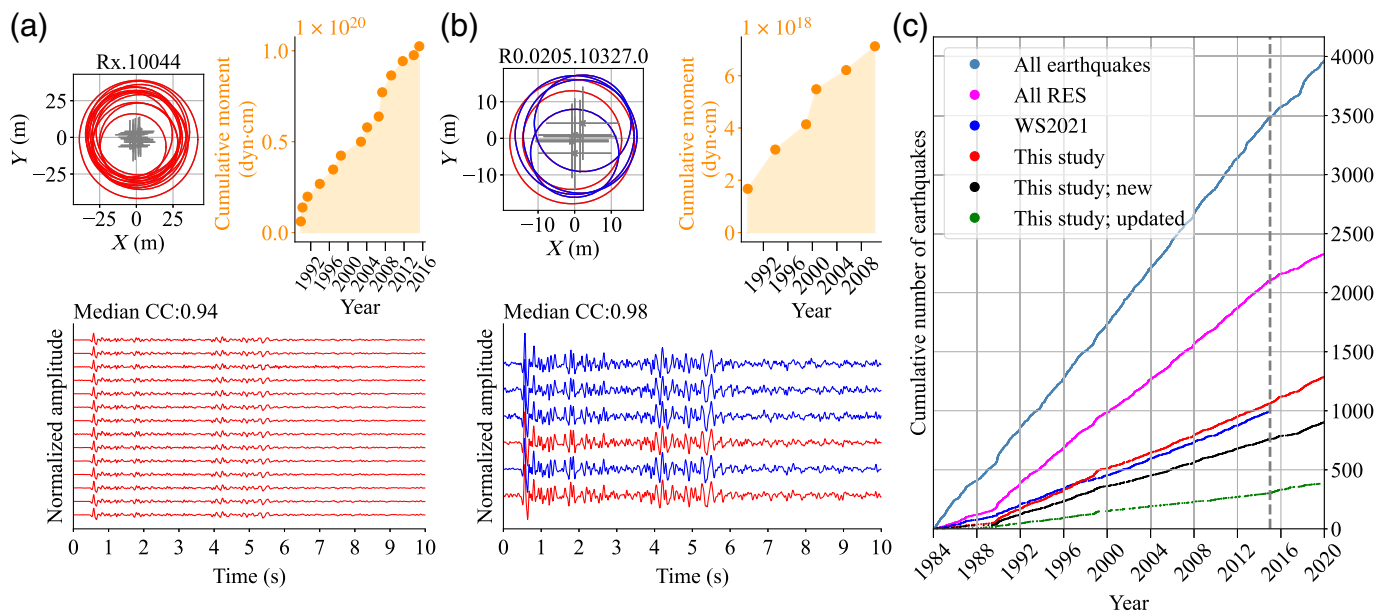
$$SR = \left(\frac{\sum_{i=1}^N S^i}{N} \right) / \left(\frac{\sum_{i=2}^N T_R^i}{N-1} \right). \quad (4)$$

Results

Machine learning spectral similarity discovers an additional 1067 repeating events during the WS2021 period 1984–2014 (an increase of 107%), including 759 new repeating earthquakes in 158 new sequences (Table S1, in the supplemental material available to this article, and example in Fig. 3a). About 79 RES from the WS2021 study were updated with 308 repeating events that were missed in the earlier study (example in Fig. 3b). However, 210 of the 995 repeating earthquakes in WS2021 (41 RES; 21%) were missed via our screening method but were subsequently appended for our final RES catalog.

Figure 3c shows the cumulative number of repeating earthquakes detected over time in our study region. The steel blue dots show the total number of earthquakes in the DD catalog of Waldhauser and Schaff (2008), while the fuchsia dots are the repeating earthquakes determined by this study. The repeating earthquakes of WS2021 are shown in blue dots. Throughout the study period (1984–2019), we find that ~60% of all seismicity in the region is RES, compared with ~25% in WS2021 for the period 1984–2014 (Table S1). Map and along-fault depth views of the updated RES catalog are given in Figure 1a,b (red circles). Overall, the spatial distribution of the new RES is similar to those of WS2021 (blue circles).

A comparison of RESp locations and inferred slip rates between WS2021 and this study (Fig. 4) shows an increase in spatial coverage, especially at shallow (<0.5 km) and relatively deep (>7 km) depths. Note also how three seismic gaps (black ellipses in Fig. 4b) become more clearly demarcated as the WS2021 catalog is enhanced. The median RESp-derived slip rate at shallow depths (<0.5 km) is 1.14 ± 0.52 cm/yr (12 observed RESp). The WS2021 catalog with six observed RESp in that depth range shows slightly higher slip rates



(1.79 ± 0.37 cm/yr). At the deepest extent of our observations (>7 km, 15 RESp observed), we estimate a slip rate of 1.18 ± 0.55 cm/yr, compared with WS2021 with four RESp and a median slip rate of 2.11 ± 0.76 cm/yr. The median slip rate along the entire fault segment is consistent between the two studies: 1.42 ± 0.66 cm/yr (157 RESp) for this study compared with 1.77 ± 0.83 cm/yr (97 RESp) for WS2021, indicating the robustness of fault-slip estimates derived from repeating earthquakes.

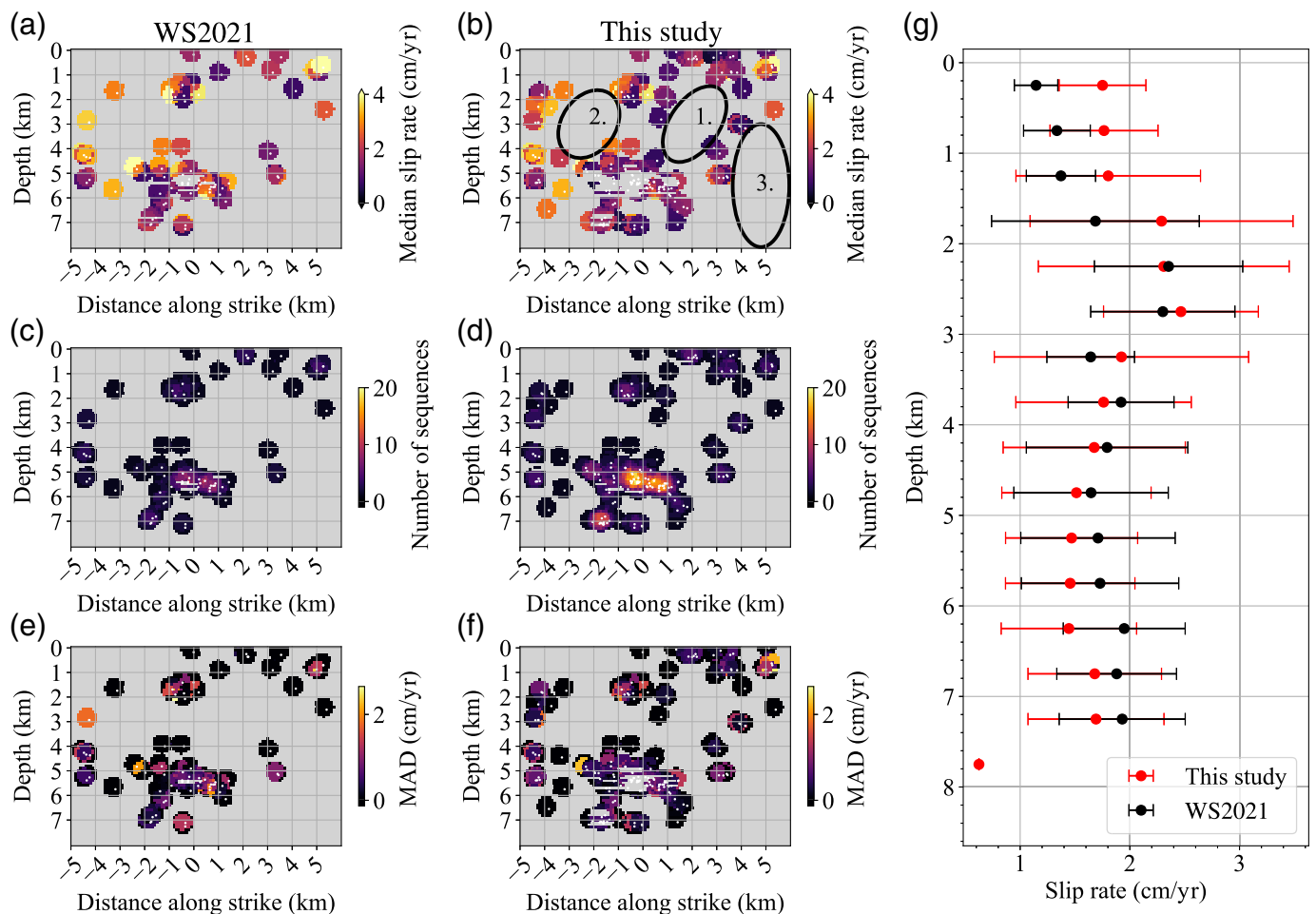
Discussion

Earthquake waveforms contain information about the earthquake source as well as properties of the geologic medium through which the waves pass. Our approach harnesses the similarity of the time-varying spectral content of waveforms via the SpecUFEx fingerprint, in contrast to most RES detection studies that only depend on time-domain similarities evaluated through waveform CC. Overall, similar waveforms (those with high CC coefficients) tend to result in similar fingerprints (those with low Euclidean distances), as shown in Figure 5, suggesting that both metrics measure similar attributes between events. However, this study finds over 100% more RES with a spectral-based approach, implying that CC and fingerprint distances have different sensitivities. One reason for this could be subtle shifts in the phase of the earthquake coda, possibly caused by changes in the geological medium over the years during which the RES occurs. Because the spectrograms' 1 s STFT windows average over several cycles

Figure 3. (a) An example of a new sequence found in this study and (b) an updated sequence showing repeating earthquake (RE) locations, location errors (gray crosses), cumulative number of RE through time, and waveforms recorded at station BAV along with median CC coefficient between all waveforms. The red and blue colors indicate newly identified repeating earthquakes from this study and from WS2021, respectively. (c) Plot showing the cumulative numbers of earthquakes and RE over time. The blue-gray dots are all earthquakes in the study area from 1984 to 2019; and blue dots are RE from the WS2021 catalog, which ends in 2014. The red dots are newly detected RE (not in WS2021) from this study; fuchsia dots are all actual RE from this study and WS2021 combined; black and green dots are entirely new RE and updated RE from this study, respectively.

of the wavetrain, the fingerprint will not be sensitive to such phase shifts within a given window. In addition, noise in the waveforms, especially for small earthquakes, can lead to low-waveform CC coefficients, whereas the noise issue is alleviated through the NMF component of SpecUFEx. Further studies are required to investigate the degree to which noise influences fingerprint results.

Other factors contributed to the significant increase in detected repeating events. For example, the focus in WS2021 was on finding isolated RES with no background seismicity nearby, which deselected some potential RES during the initial screening process. They also used a hypocentral separation threshold to search for potential RES that may have been too strict to catch all potential repeaters. This is especially apparent for shallow events where the uncertainty in depths appears to have been larger than the hypocentral separation



criteria implemented in WS2021. Some of these missed RES indeed have high CC coefficients (red circles in Fig. 5b). These issues highlight the advantage of using fingerprints as an efficient way of comparing the similarity of seismic signals independent of the distance between earthquakes. That is especially beneficial when searching for potential repeating events in large catalogs with millions of earthquakes and routine network locations that typically have large errors. Once the potential RESs are found in such catalogs, then step two in the process of detecting repeating events, the high-resolution analysis of each potential RES as outlined in WS2021 and implemented here, can be restricted to only the catalog of potential RES and does not have to involve the entire earthquake catalog.

The increase in RES detections helps expand the spatial coverage of on-fault slip-rate estimates within the seismogenic fault zone (Fig. 4). Newly found RESp improve the fine-scale slip distribution near the surface along the southeastern portion of the SAF in our study region (south of $\sim 36.46^\circ$), whereas the northern

Figure 4. (a,b) Median slip rates, (c,d) number of earthquakes, and (e,f) median absolute deviation (MAD) of slip rates as a function of depth and distance along strike of the SAF for WS2021 in panels (a,c,e) and this study in panels (b,d,f). The white dots represent quasi-periodically repeating sequences (RESp) locations, and gray areas indicate no observations. The numbered black ellipses in panel (b) demarcate the approximate boundaries of aseismic regions on the fault surface. Data are interpolated on a 100×100 m grid with 400 m search radius and 50% overlap. (g) Slip rates at depth in 500 m bins for this study (red) and WS2021 (black). The dots show the median slip rate and error bars are the median absolute deviation.

portion is devoid of repeaters (Fig. 1b). This abrupt change is consistent with the transition in surface geology from moderately consolidated marine sediments in the southeast to unconsolidated sediments in the northwest (Saucedo *et al.*, 2000). Creep meter measurements in the region (Schulz *et al.*, 1982) are consistent with our slip estimations, with a creep rate near zero (0.025 cm/yr) in the northwest where no surface RESp are

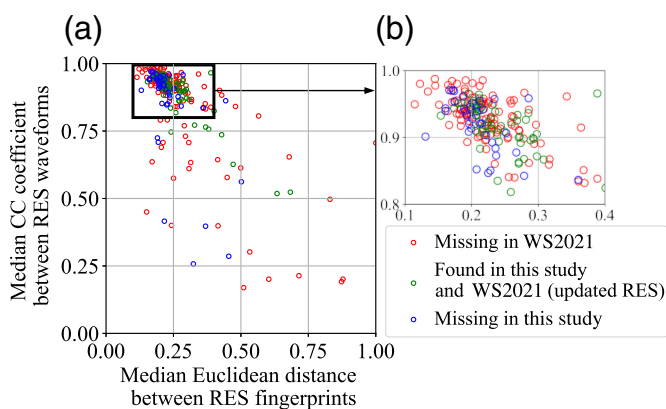


Figure 5. Comparison of similarity metrics of waveforms and fingerprints within repeating sequences that were found in this study but missing in WS2021 (red dots), sequences that were found in WS2021 but omitted in this study (blue dots) and sequences that were found in WS2021 and updated by this study (green dots). Note the general inverse relationship between waveform CC coefficients and the (normalized) Euclidean fingerprint distances, demonstrating that similar waveforms tend to result in similar fingerprints. Panel (b) is a zoomed-in view on the inset boxed data in panel (a). Data come from station BPL.

observed, and 1.70 cm/yr in the southeast where our median near-surface slip-rate estimate is 1.14 ± 0.52 cm/yr. Alignment array (AA) measurements report higher slip rates, with 2.19 cm/yr in the northwest and 3.13 cm/yr in the southeast (Burford and Harsh, 1980). These higher values are likely due to slip partitioning in the region as AA data capture slip across multiple, subparallel fault strands. This is supported by repeating events that have been found on the subparallel Pine Rock and San Benito faults (WS2021).

Slip-rate variations across the fault interface have a median average deviation of 0.5 cm/yr (Fig. 4e,f), which we interpret as an expression of the natural heterogeneity of frictional properties of the asperities. The updated RESp catalog reveals three distinct RESp gaps along the fault interface (black ellipses in Fig. 4b), which also contain little background seismicity (Fig. 1b). These gaps are visible in the catalog of WS2021, but their boundaries clarify as RESp are supplemented by this study. Each gap has an area of ~ 25 km², which, if we assume that they represent frictionally locked fault areas, can generate a M_w 5 earthquake if ruptured simultaneously.

Conclusion

In this study, we introduce a novel, unsupervised machine learning method for detecting potential RES based on spectrograms. We find more than twice as many repeating

earthquakes as traditional methods based on comparing waveforms on a 10 km creeping segment of the SAF. The performance enhancement is in part, because this method does not rely on initial high-precision earthquake catalogs, although such catalogs and their associated CC measurements help build a more complete RES catalog. Slip rates estimated from these new RES indicate fault heterogeneity from the kilometer scale down to the size of individual asperities. We plan to apply our detection technique to more regions in California to gain a better understanding of the nature of repeating earthquakes, and fault behavior along the SAF system and beyond.

Data and Resources

The catalog of repeating earthquakes from this study is archived at [zenodo.org](https://zenodo.org/doi/10.5281/zenodo.10251244) (doi: 10.5281/zenodo.10251244). The Python implementation of Spectral Unsupervised Feature Extraction (SpecuFEx) can be found at github.com/specufex. Other codes used in this study are at github.com/tsawi/Repeating_Earthquakes_SpecUFEx. For hierarchical clustering, we use the “agglomerative clustering” package from scikit-learn, and we implement HypoDD for relocations (Waldhauser and Ellsworth, 2000). All seismic data used in this study are available from ddrt.ldeo.columbia.edu and the Northern California Earthquake Data Center (NCEDC) at <https://ncedc.org/>. All the websites were last accessed in March 2023. The supplemental material includes a table summarizing results and a catalog of RES.

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

Acknowledgments

This article greatly benefitted from the thoughtful feedback given by the Editor-in-Chief Keith Koper, Associate editor Steve Gibbons, and two anonymous reviewers. Theresa Sawi was supported by the National Science Foundation Graduate Research Fellowship Program (NSF-GRFP) Grant Number 2036197 (2020–2023) with additional support from U.S. Geological Survey National Earthquake Hazards Reduction Program (US-NEHRP G20AP00025) and National Science Foundation Office of Advanced Computing (NSF-OAC 2103741).

References

- Abercrombie, R. E., X. Chen, and J. Zhang (2020). Repeating earthquakes with remarkably repeatable ruptures on the SAF at Parkfield, *Geophys. Res. Lett.* **47**, no. 23, 1–10, doi: 10.1029/2020GL089820.

- Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.* **75**, no. 26, 4997–5009, doi: [10.1029/JB075i026p04997](https://doi.org/10.1029/JB075i026p04997).
- Burford, R. O., and P. W. Harsh (1980). Slip on the San Andreas fault in central California from alignment array surveys, *Bull. Seismol. Soc. Am.* **70**, 1233–1261.
- Bürgmann, R., D. Schmidt, R. M. Nadeau, M. d'Alessio, E. Fielding, D. Manaker, T. V. McEvilly, and M. H. Murray (2000). Earthquake potential along the northern Hayward fault, California, *Science* **289**, no. 5482, 1178–1182, doi: [10.1126/science.289.5482.1178](https://doi.org/10.1126/science.289.5482.1178).
- Dreger, D., R. M. Nadeau, and A. Chung (2007). Repeating earthquake finite source models: Strong asperities revealed on the SAF, *Geophys. Res. Lett.* **34**, no. 23, 1–5.
- Gao, D., H. Kao, and B. Wang (2021). Misconception of waveform similarity in the identification of repeating earthquakes, *Geophys. Res. Lett.* **48**, 13, doi: [10.1029/2021GL092815](https://doi.org/10.1029/2021GL092815).
- Geller, R. J., and C. S. Mueller (1980). Four similar earthquakes in central California, *Geophys. Res. Lett.* **7**, 821–824, doi: [10.1029/GL007i010p00821](https://doi.org/10.1029/GL007i010p00821).
- Holtzman, B. K., A. Paté, J. Paisley, F. Waldhauser, and D. Repetto (2018). Machine learning reveals cyclic changes in seismic source spectra in Geysers geothermal field, *Sci. Adv.* **4**, 5, doi: [10.1126/sciadv.aao2929](https://doi.org/10.1126/sciadv.aao2929).
- Kong, Q., D. Trugman, Z. Ross, M. Bianco, B. Meade, and P. Gerstoft (2018). Machine learning in seismology: Turning data into insights, *Seismol. Res. Lett.* **90**, doi: [10.1785/0220180259](https://doi.org/10.1785/0220180259).
- Mousavi, S. M., and G. C. Beroza (2022) Deep-learning seismology, *Science* **377**, doi: [10.1126/science.abm4470](https://doi.org/10.1126/science.abm4470).
- Nadeau, R. M., and L. R. Johnson (1998). Seismological studies at Parkfield VI: Moment release rates and estimates of source parameters for small repeating earthquakes, *Bull. Seismol. Soc. Am.* **88**, no. 3, 790–814.
- Nadeau, R. M., and T. V. McEvilly (1999). Fault slip-rates at depth from recurrence intervals of repeating microearthquakes, *Science* **285**, no. 5428, 718–721, doi: [10.1126/science.285.5428.718](https://doi.org/10.1126/science.285.5428.718).
- Nadeau, R. M., W. Foxall, and T. V. McEvilly (1995). Clustering and periodic recurrence of microearthquakes on the San Andreas fault at Parkfield, California, *Science* **267**, no. 5197, 503–7, doi: [10.1126/science.267.5197.503](https://doi.org/10.1126/science.267.5197.503).
- Peng, Z., J. E. Vidale, and A. Rubin (2005). Systematic variations in recurrence interval and moment of repeating aftershocks, *Geophys. Res. Lett.* **32**, L15301, doi: [10.1029/2005GL022626](https://doi.org/10.1029/2005GL022626).
- Poupinet, G., W. L. Ellsworth, and J. Frechet (1984). Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras fault, California, *J. Geophys. Res.* **89**, no. B7, 5719–5731.
- Rubin, A. M., D. Gillard, and J. L. Got (1999). Streaks of microearthquakes along creeping faults, *Nature* **400**, no. 6745, 635–641, doi: [10.1038/23196](https://doi.org/10.1038/23196).
- Saucedo, G. J., D. R. Bedford, G. L. Raines, R. J. Miller, and C. M. Wentworth (2000). *GIS Data for the Geologic Map of California, CD 2000-007*, California Department of Conservation, California Geological Survey, Sacramento, California, U.S.A.
- Schaff, D. P., G. H. R. Bokelmann, G. C. Beroza, F. Waldhauser, and W. L. Ellsworth (2002). High-resolution image of Calaveras fault seismicity, *J. Geophys. Res.* **107**, no. B9, ESE 5-1–ESE 5-16, doi: [10.1029/2001jb000633](https://doi.org/10.1029/2001jb000633).
- Schulz, S. S., G. M. Mavko, R. O. Burford, and W. D. Stuart (1982). Long-term fault creep observations in central California, *J. Geophys. Res.* no. B8, 6977–6982, doi: [10.1029/JB087iB08p06977](https://doi.org/10.1029/JB087iB08p06977).
- Shakibay, S. N., and G. J. Funning (2019). Widespread fault creep in the northern San Francisco Bay area revealed by multistation cluster detection of repeating earthquakes, *Geophys. Res. Lett.* **46**, no. 12, 6425–6434, doi: [10.1029/2019GL082766](https://doi.org/10.1029/2019GL082766).
- Templeton, D. C., R. M. Nadeau, and R. Bürgmann (2008). Behavior of repeating earthquake sequences in central California and the implications for subsurface fault creep, *Bull. Seismol. Soc. Am.* **98**, no. 1, 52–65, doi: [10.1785/0120070026](https://doi.org/10.1785/0120070026).
- Uchida, N. (2019). Detection of repeating earthquakes and their application in characterizing slow fault slip, *Progr. Earth Planet. Sci.* **6**, no. 1, doi: [10.1186/s40645-019-0284-z](https://doi.org/10.1186/s40645-019-0284-z).
- Uchida, N., and R. Bürgmann (2019). Repeating earthquakes, *Ann. Rev. Earth Planet. Sci.* **47**, doi: [10.1146/annurev-earth-053018-060119](https://doi.org/10.1146/annurev-earth-053018-060119).
- Vidale, J. E., W. L. Ellsworth, A. Cole, and C. Marone (1994). Variations in rupture process with recurrence interval in a repeated small earthquake, *Nature* **368**, no. 6472, 624–626, doi: [10.1038/368624a0](https://doi.org/10.1038/368624a0).
- Waldhauser, F. (2001). hypoDD—A program to compute double-difference hypocenter locations, *U.S. Geol. Surv. Open-File Rept. 01-113*, 25, available at <https://pubs.usgs.gov/of/2001/0113/> (last accessed March 2023).
- Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California, *Bull. Seismol. Soc. Am.* **90**, no. 6, 1353–1368.
- Waldhauser, F., and W. L. Ellsworth (2002). Fault structure and mechanics of the Hayward fault, California, from double-difference earthquake locations, *J. Geophys. Res.* **107**, no. B3, 2054.
- Waldhauser, F., and D. P. Schaff (2008). Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.* **113**, no. 8, 1–15, doi: [10.1029/2007JB005479](https://doi.org/10.1029/2007JB005479).
- Waldhauser, F., and D. P. Schaff (2021). A comprehensive search for repeating earthquakes in northern California: Implications for fault creep, slip-rates, slip partitioning, and transient StRES, *J. Geophys. Res.* **126**, no. 11, 1–22, doi: [10.1029/2021JB022495](https://doi.org/10.1029/2021JB022495).
- Waldhauser, F., W. L. Ellsworth, D. P. Schaff, and A. Cole (2004). Streaks, multiplets, and holes: High-resolution spatio-temporal behavior of Parkfield seismicity, *Geophys. Res. Lett.* **31**, no. 18, 2–5, doi: [10.1029/2004GL020649](https://doi.org/10.1029/2004GL020649).