Geophysical Research Letters

Supporting Information for

Construction of Long-term Seismic Catalog with Deep Learning and Characterization of Preseismic Fault Behavior in the Ridgecrest-Coso Region (2008-2019)

Yijian Zhou[1*], Abhijit Ghosh[1], Lihua Fang[2], Han Yue[3], Shiyong Zhou[3]

[1]Department of Earth and Planetary Sciences, University of California, Riverside, USA

[2]Institute of Geophysics, China Earthquake Administration, Beijing, China

[3]Institute of Theoretical and Applied Geophysics, Peking University, Beijing, China

Contents of this file

Introduction

This supporting material provides additional information on the deep learning workflow, the location strategy and uncertainty analysis, and the b-value calculation.

Text S1.

Construction of the training set. The training set for the CERP model consists of positive and negative samples that is of roughly the same size: the CNN for event detection uses both positive and negative samples, and the RNN only uses the positive samples. The positive samples, i.e. earthquakes, are sliced randomly within a time range from *tp-win_len/2* to *tp+win_len*, where *tp* is the P arrival time, and *win_len* is the window length of each sample (Figure S6). The negative samples, i.e. noises, are sliced randomly with a time range from *ts* to *ts+2*win_len*, where *ts* is the S arrival time. Note the for very intense sequences, our code also provides another choice, which is to slice negative samples relative to *tp*. Such strategy will generate negative samples containing both tail waves and random noises, so that the CNN only detect windows with both P & S arrivals covered.

However, in the case of preseismic Ridgecrest-Coso, we need special treats on the data glitches (Figure S3), since they takes a large portion (Figure S6). We detect the data glitches by applying forward and backward STA/LTA, and a glitch is defined as both STA/LTA values larger than 50 within 1 s (Figure S6). If a station-date pair has more than 100 such glitch detections, we define it as a "glitch station-date", whose distribution is shown in Figure S3. We merge glitches into the negative set by slicing 50 samples randomly in the glitch station-date pairs, which finally gives 270,601 glitch samples. Besides, we exclude the PAL detections in station-date pairs from the positive set.

For the training of RNN, considering that the 1-channel data takes a great amount of volume, whose S picks are not as accurate as that of the 3-channel data, we only use the 3-channel picks to construct the training set for RNN. Specifically, using those picks on 3-channel data, we randomly choose 50% of them with only Z-channel (i.e. [Z,Z,Z]), and the remaining half the original shape (i.e. [E,N,Z]).

Finally, the samples are randomly divided into two groups: the training set and validation set. In our case, we group 90% of the samples for training, and 10% for validation.

Text S2.

More details on the CERP parameters. The CERP parameters include the parameters that determine the neural network architecture, that controlling the training process, and that controlling the picking behavior in applying to continuous data.

The users first need to determine the window length, which should be large enough to cover both P & S arrivals for stations within a target epicentral distance, but not too large to give reasonable data weight for signals. Note that the window length will also determine the complexity of neural networks. For CNN, the neural network architecture consists of the depth, kernel size, and the number of convolutional layers. For the case in Ridgecrest, since we have more than 400,000 training samples, we adopt 8 CNN layers, and each layer is composed of 32 size-3 kernels. For RNN, we adopt 2 forward-backward RNN layers, and each layer contains GRU unit with size 32. The RNN first divide the event window into multiple time steps, and we slice 0.8-s steps with 0.2-s striding. See Zhou et al. (2019) for more details on the algorithm.

The training of CERP is in a mini-batch manner. For both CNN and RNN, we feed 128 samples in each iteration, and a learning rate of 10^{-4} to adjust the learnable parameters. We train the CNN & RNN by at most 14 epochs, and use the early stop strategy to avoid over fitting (Figure S9). Usually, the training and validation of RNN will keep decreasing, and we only need to select the check point for the CNN.

In applying CERP to continuous data, we first need to select the step for the sliding windows, which should be ranging from ¼ to ½ of the window length, empirically. The same picks from different sliding windows are merged when the time differences are within 2 s. A modification of our new implementation of CERP is the repick process with PAL picker (Zhou et al., 2021), which further improve the picking precision of the sequence labelling-based RNN method (previously know as DetNet & PpkNet, Mousavi et al., 2020). We apply PAL picker within a time range of -1~1s relative to the original pick, considering the maximum picking uncertainty of RNN.

Text S3.

The details of *ph2dt_cc* algorithm. The *ph2dt_cc* is a modulus that convert phase file (*ph*) into high-resolution travel time differences (*dt*) with cross-correlation (CC). It first find all possible event pairs by comparing the location differences and common station picks. For each event, in our case, we select candidates from a lateral difference of 3 km and a vertical difference of 4 km, and there need to be at least 4 stations shared by the two events. To control the quality of *dt* measurement, we only use stations within an epicentral distance of 80 km, and events with M<0 cannot be linked to each other (but can be linked to M>0 events). To avoid too many measurements, we also limit the maximum number of stations to be 15 for each event pair (if exceed, select from the nearest ones), and each event can only be linked to up to 200 neighbors, as in the *ph2dt* strategy (Waldhauser, 2001).

The calculation of *dt* is based on 3-channel waveforms with 100 Hz sampling rate and filtered to 2-20 Hz. The *dt_p* is measured on Z channel, and *dt_s* on all three channels based on average CC traces. Specially, in our case, if one of the stations in an event pair is 1-channel data, we also measure *dt_s* on Z channel and lowers down the weighting in hypoDD to 0.35^2, instead of CC^2 as for other pairs. The window for P wave correlation is 2.5-s long starting from 0.5 s before P arrival, and that for S wave is 4-s long starting from 0.2 s before S arrival. After the calculation, we further select the *dt* measurements by dropping that with *dt_p*>0.5 s or *dt_s*>0.8 s, or if the CC value is below 0.35. Finally, only event pairs with at least 4 stations fulfilling those criteria are linked, and output to *dt.cc*.

Text S4.

The weighting scheme in the (re)location process and the location uncertainty. The relocation process of CERP detection experiences absolution location from hypoInverse, double-difference relocation with hypoDD (sequentially with *dt.ct* and *dt.cc* data).

In the hypoInverse process, the weighting for each phase is determined by its epicentral distance, travel time residual, and the type of phase (P or S), and the product of these three factors are the final; weight. We give full weight for stations within 30 km, and 0 beyond 90 km, and a cosine taper is applied in between. Similarly, for time residual, we assign full weight for <0.25-s residual, 0 for >0.75-s residual, and cosine taper for 0.25-0.75-s residual. We assign 1 & 0.6 for P and S picks, respectively, and 0.3 for S picks from 1-channel stations. We run hypoInverse with a velocity model adjusted from the SCSN model (Table S1), which avoid steep velocity change by adding two more layers. Finally, within the study region, 98,479 out of 104,335 events (94%) are well constraint, among which 4%, 41%, 47%, and 8% belongs to A, B, C, and D quality location, according to the criteria of hypoInverse (Klein, 2002). The average location error along lateral and vertical direction is 1.22 & 2.60 km.

In the hypoDD process, we first utilize the *dt.ct* to constrain cluster-to-cluster location. We calculate *dt* for event pairs within a hypocentral distance of 10 km (i.e. WDCT=10), and requires at least 8 phases to link a pair. To account for 1 & 3-channel difference, we weight the P & S in 3-channel and S in 1-channel data as 1, 0.5, and 0.25, respectively. We perform 1 set of 4 iterations of inverse. Finally, the *dt.ct* relocation gives an uncertainty of 110 & 190 m in lateral and vertical direction. In the *dt.cc* relocation process, we calculate *dt.cc* following the procedure in Text S3, which gives 19,187,060 *dt* measurements. We utilize CC to refine the relative location within a 4-km separation (i.e. WDCC=4), and run 4 iterations of iterations. The final CC-relocated catalog has a location uncertainty of 60 & 80 m in the lateral and vertical direction, based on the least-square criteria.

Text S5.

The b-value calculation and its uncertainties. The b-value calculation is based on maximum-likelihood method (Aki, 1965):

$$b = \frac{\log_{10} e}{\bar{M} - M_C + \frac{\Delta M}{2}}, \quad (S1)$$

where $\bar{M}$, M_C, and ΔM is the mean magnitude, lower cut-off magnitude, and the magnitude bin, respectively. The b-value uncertainty is estimated with Shi and Bolt (1982):

$$\delta b = 2.3 \times b^2 \sum_{i=1}^{n} \frac{(M_i - \bar{M})^2}{n(n-1)}, \quad (S2)$$

where n is the number of events. In mapping the b-value distribution, we test two strategies: adopting a uniform M_C (Figure S10 & S12) and calculating M_C on each grid (Figure S11). We prefer the result with uniform M_C, since the b-value estimation tend to increase with the M_C (Zhou et al., 2018). However, whatever mapping strategy and slicing radius adopted, the basic patterns are the same.

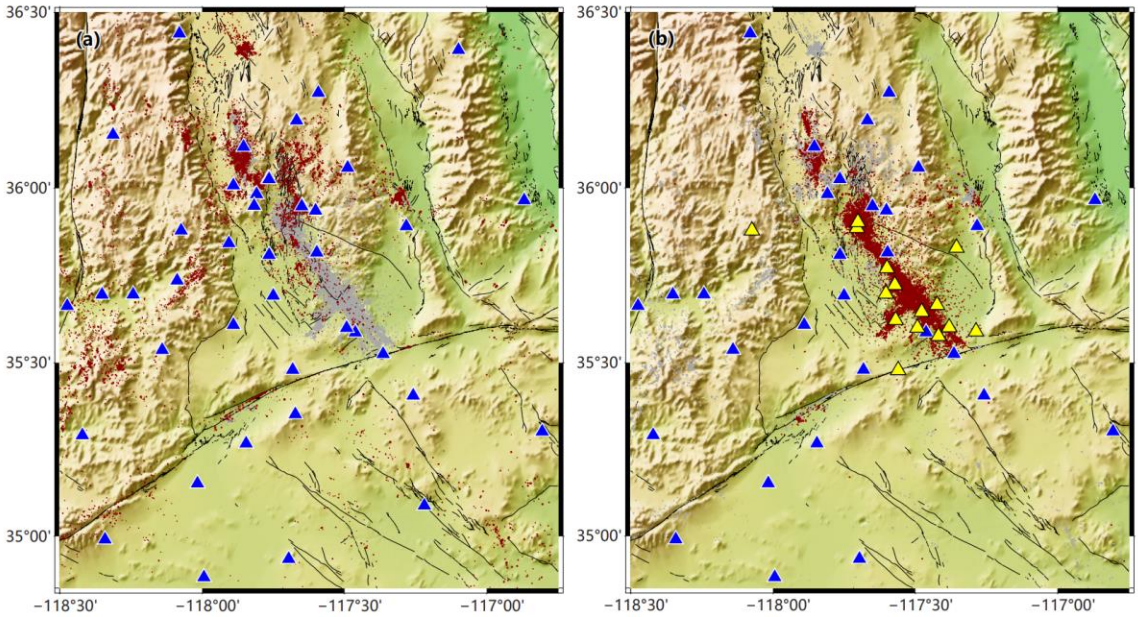

Figure S1. Distribution of selected stations for the (a) preseismic and (b) aftershock period. The blue triangles are broad-band and short-period seismometers, and yellow triangles are temporary stations deployed after the 2019 Ridgecrest earthquake (Hauksson et al., 2020).

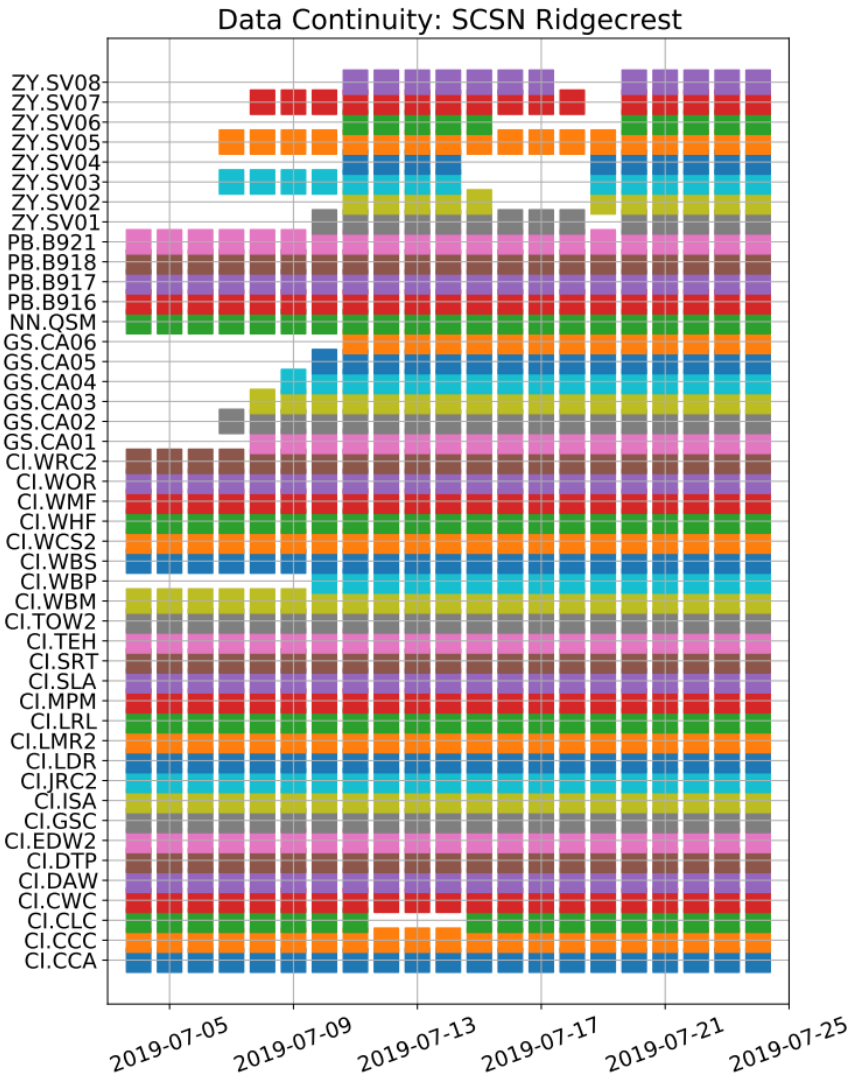

Figure S1. Data continuity for the aftershock period. The station-date pairs with data available are marked by rectangles.

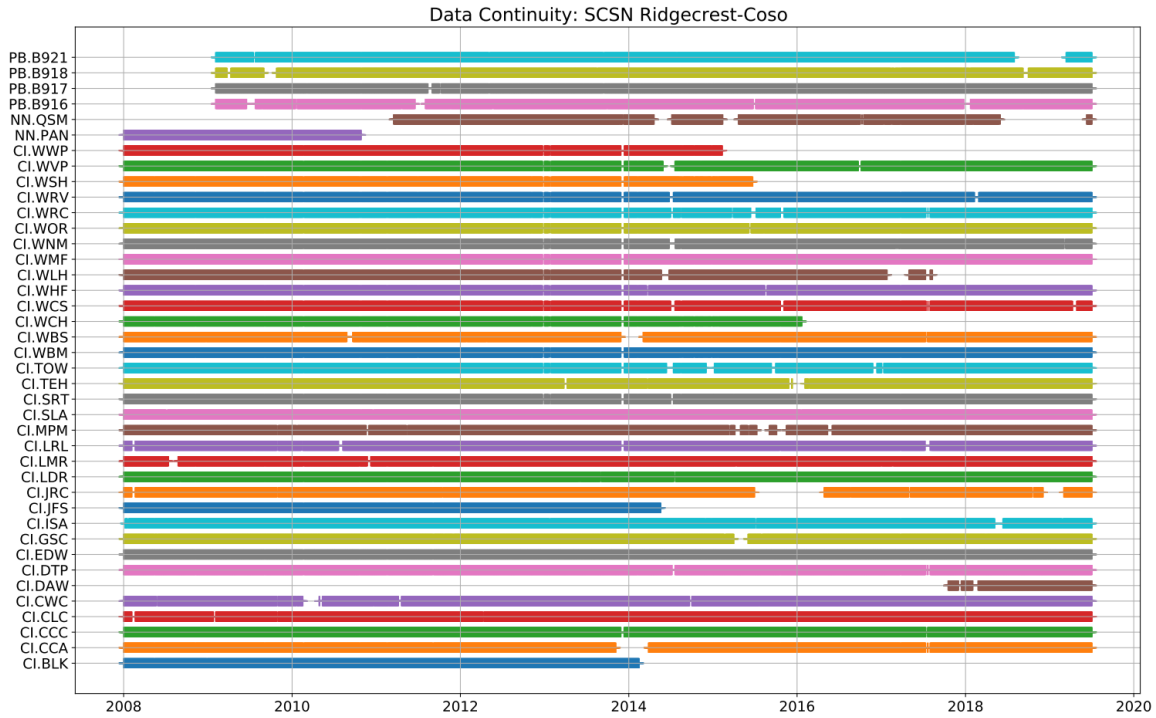

Figure S2. Data continuity for the preseismic period. The station-date pairs with data available are marked by "+".

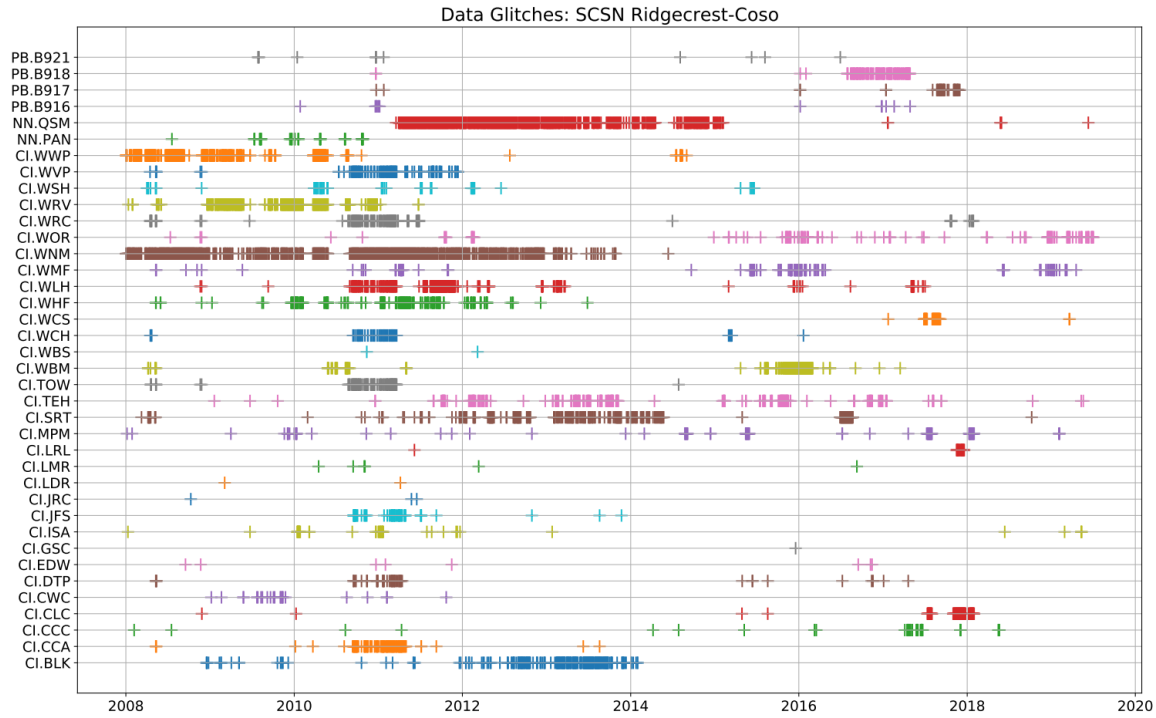

Figure S3. Distribution of data glitches. The station-date pairs with data available are marked by "+".

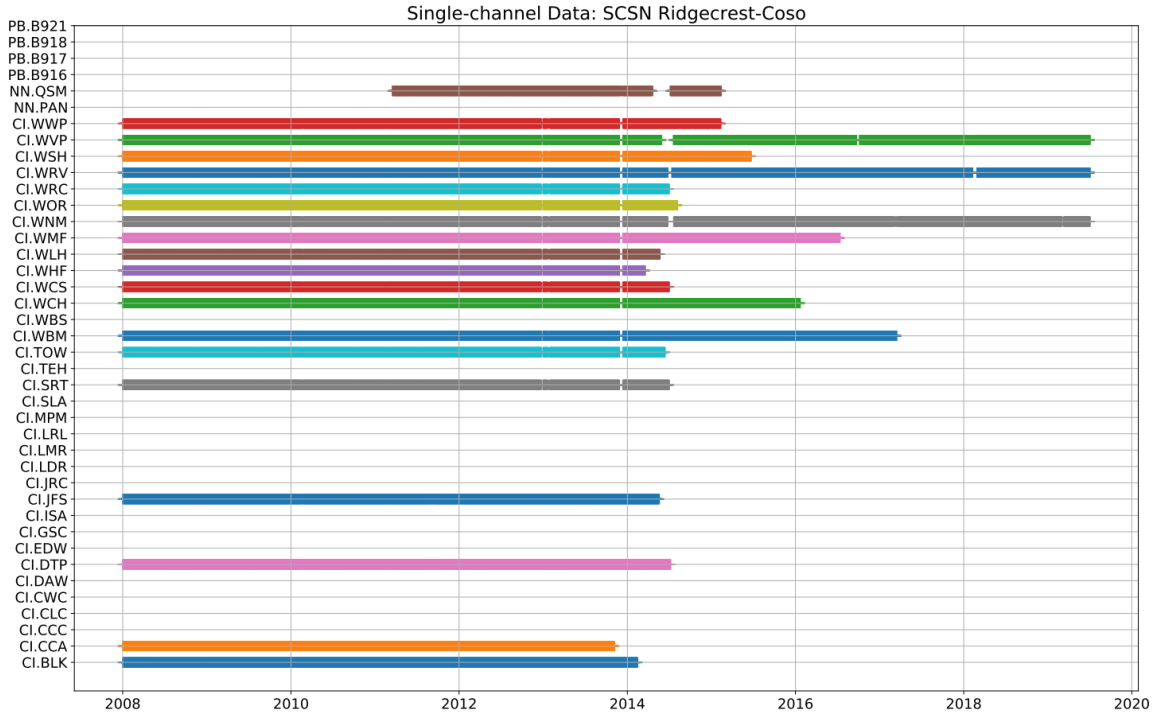

Figure S4. Distribution of single-channel data. The station-date pairs with data available are marked by "+".

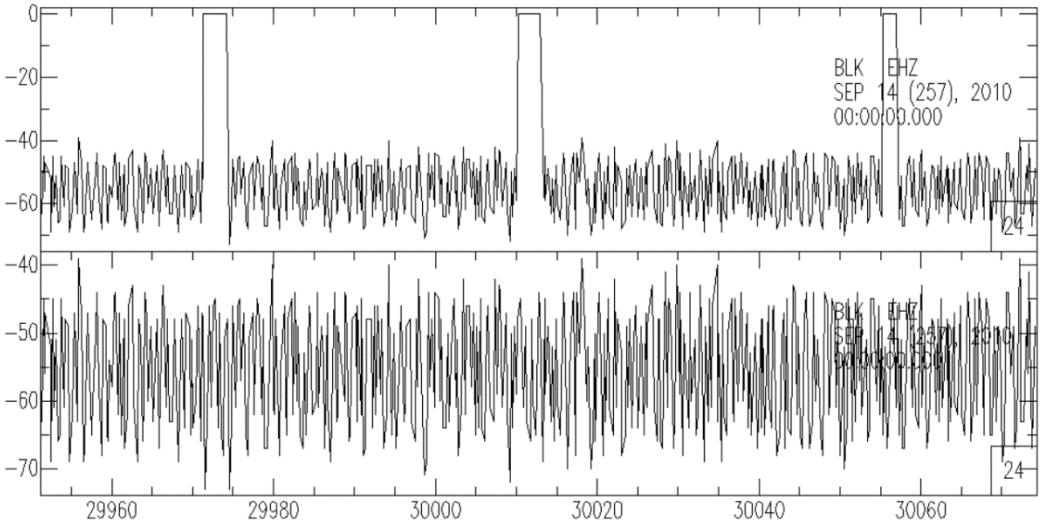

Figure S5. Example data gap and filling method.

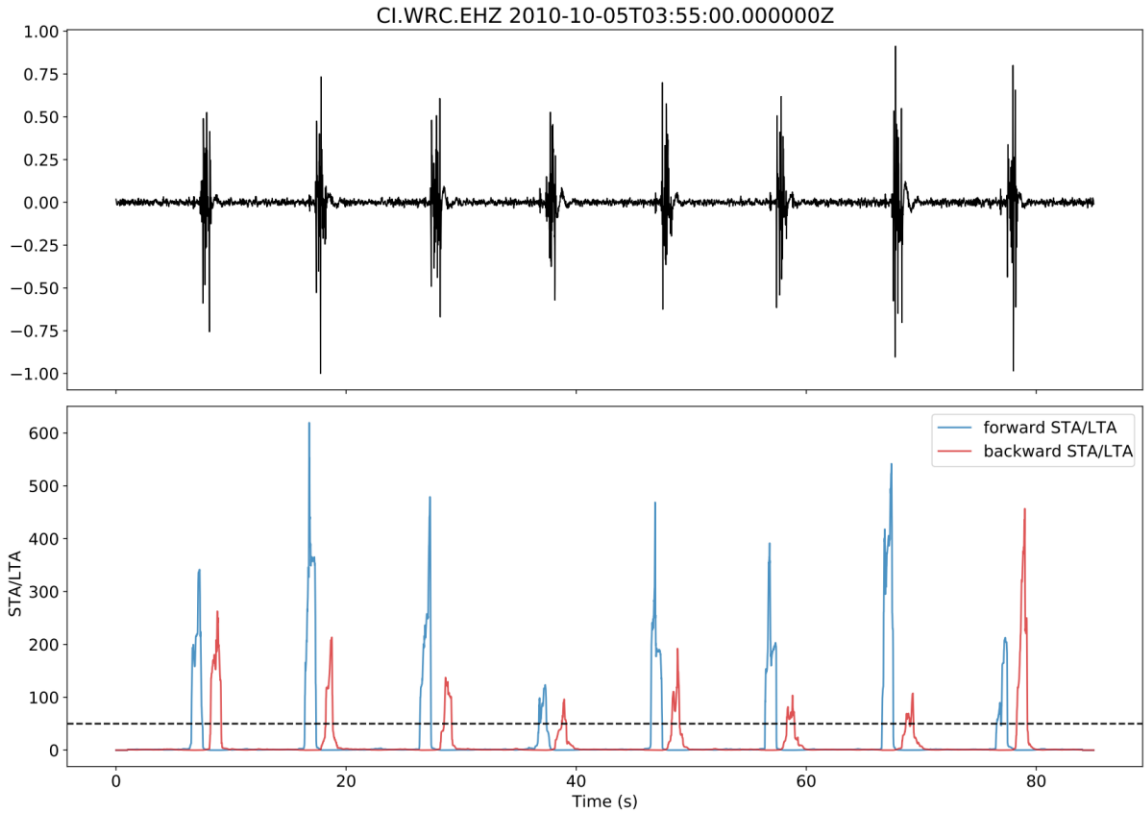

Figure S6. Example data glitch and detection method. Waveform is band-pass filtered to 1-20 Hz.

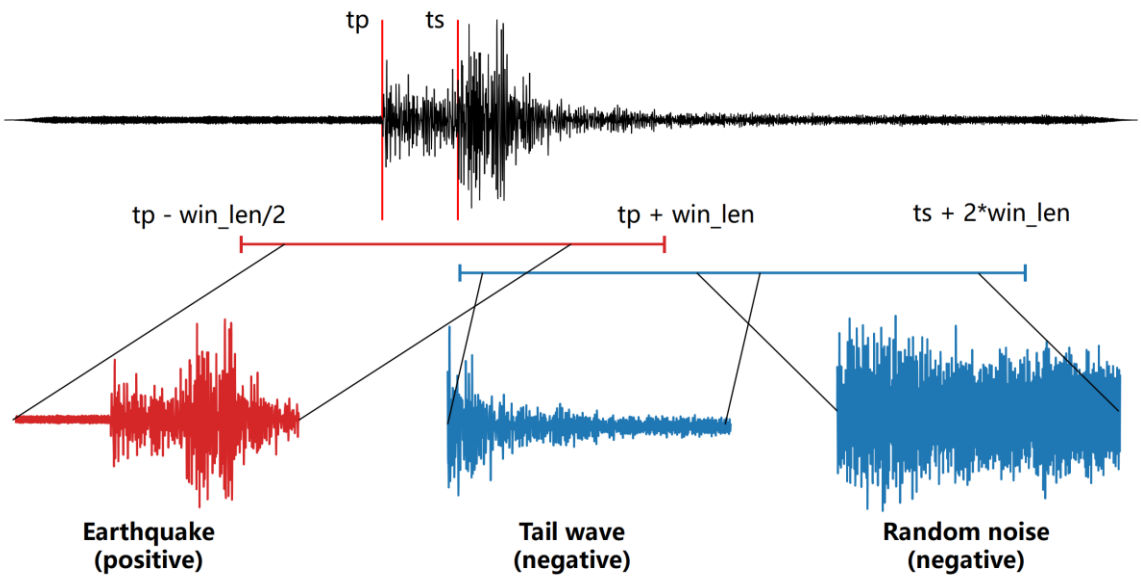

Figure S7. Demonstration of training data slicing strategy. The red and blue horizontal line mark the time range of randomly sliced positive / negative samples.

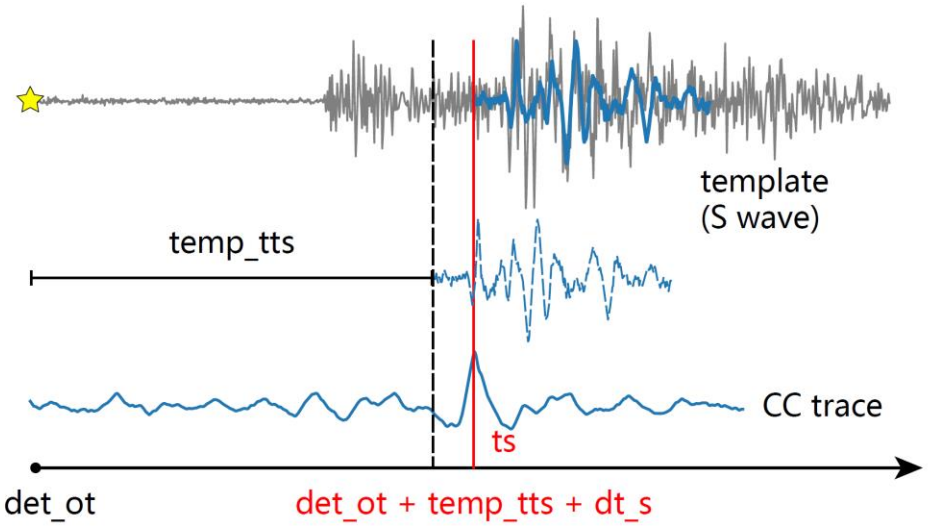

Figure S8. Demonstration of differential travel time measurement with cross-correlation. The two events are named as "detection" and "template", travel time and differential travel time are abbreviated as "tt" and "dt", respectively.

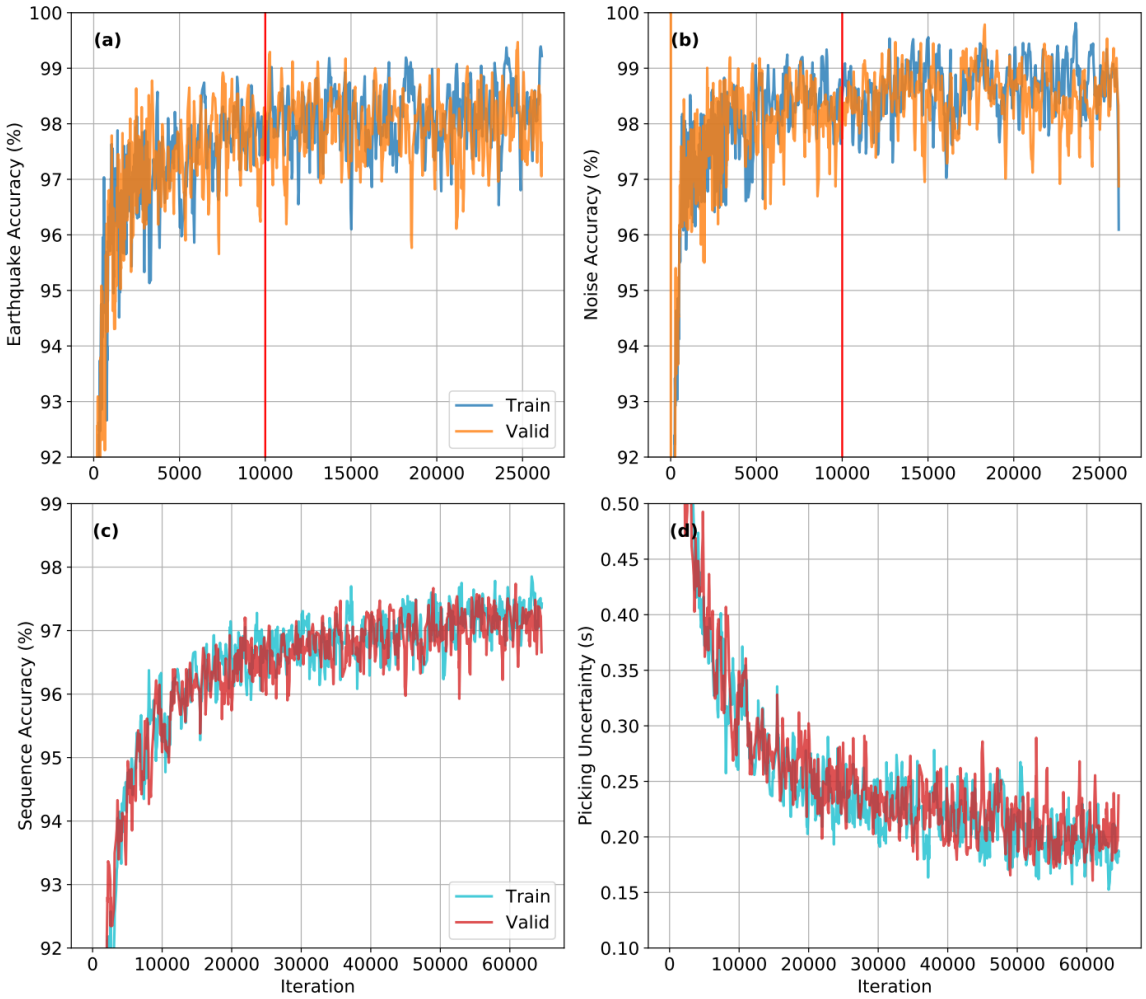

Figure S9. Training curve and selected check point. (a-b) and (c-d) is the training curve for the CNN and RNN, respectively. (a) and (b) plot the training and validation accuracy for positive and negative samples, respectively. The vertical red line marks the selected checkpoint using the early stop strategy. (c) plots the sequence accuracy which is defined as the correctly classified time step over the total number of time step. (d) is the converted effective picking uncertainty.

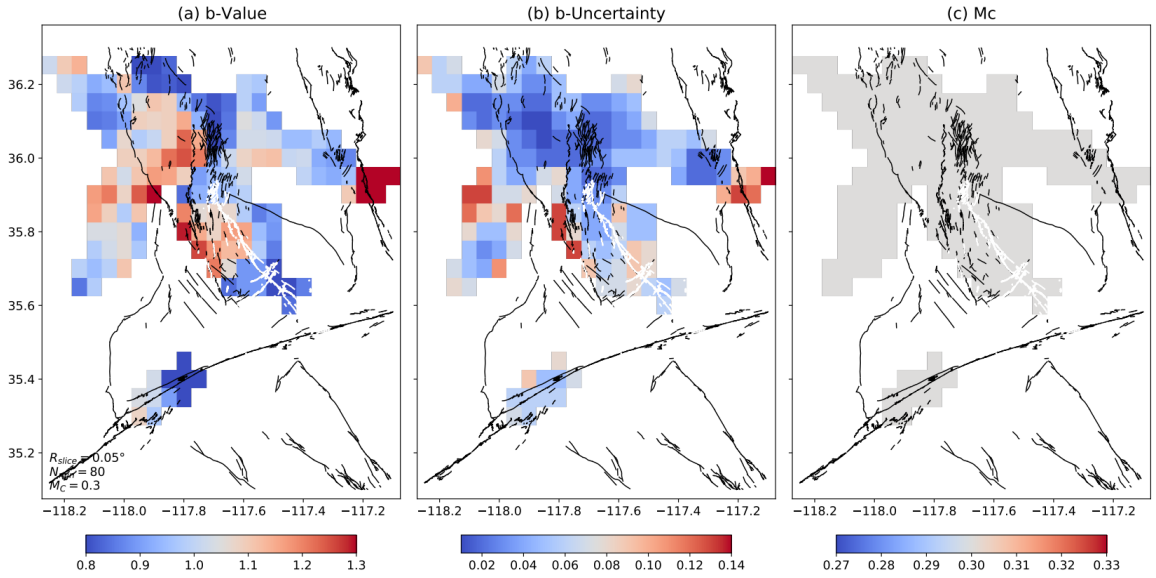

Figure S10. B-value distribution with fixed Mc (that in the main text).

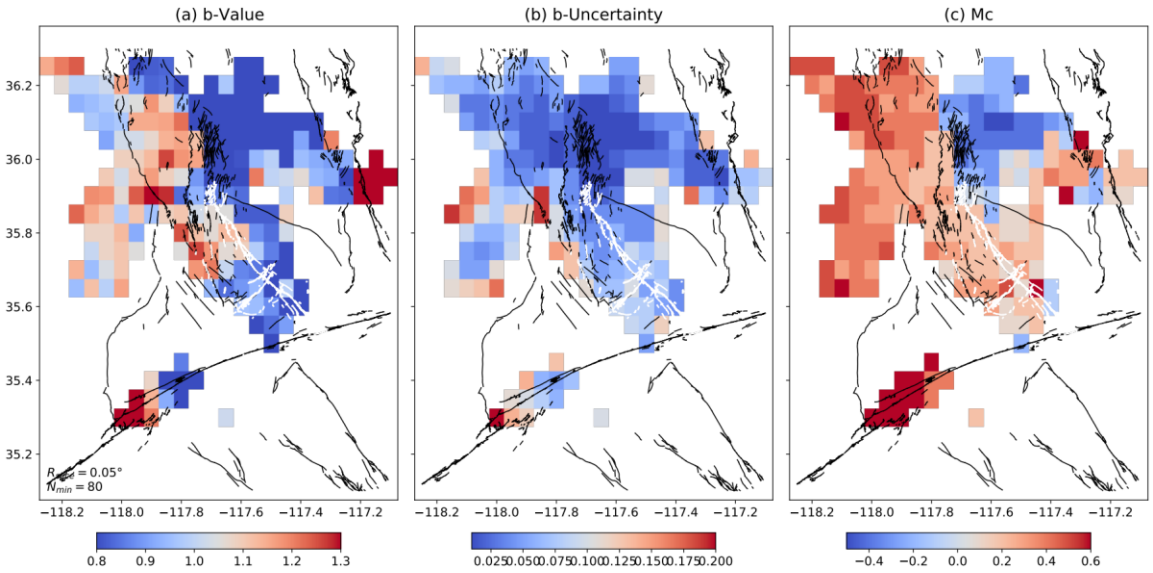

Figure S11. B-value distribution with Mc calculated on each grid. Other parameters are the same with the main text.

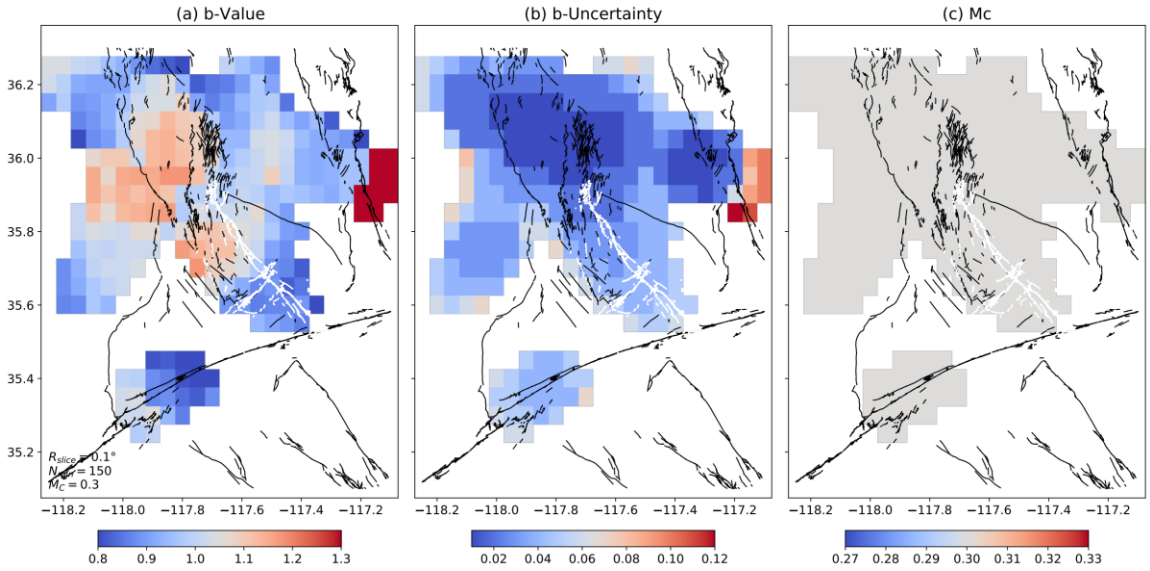

Figure S12. B-value distribution with slicing radius set to 0.1°.

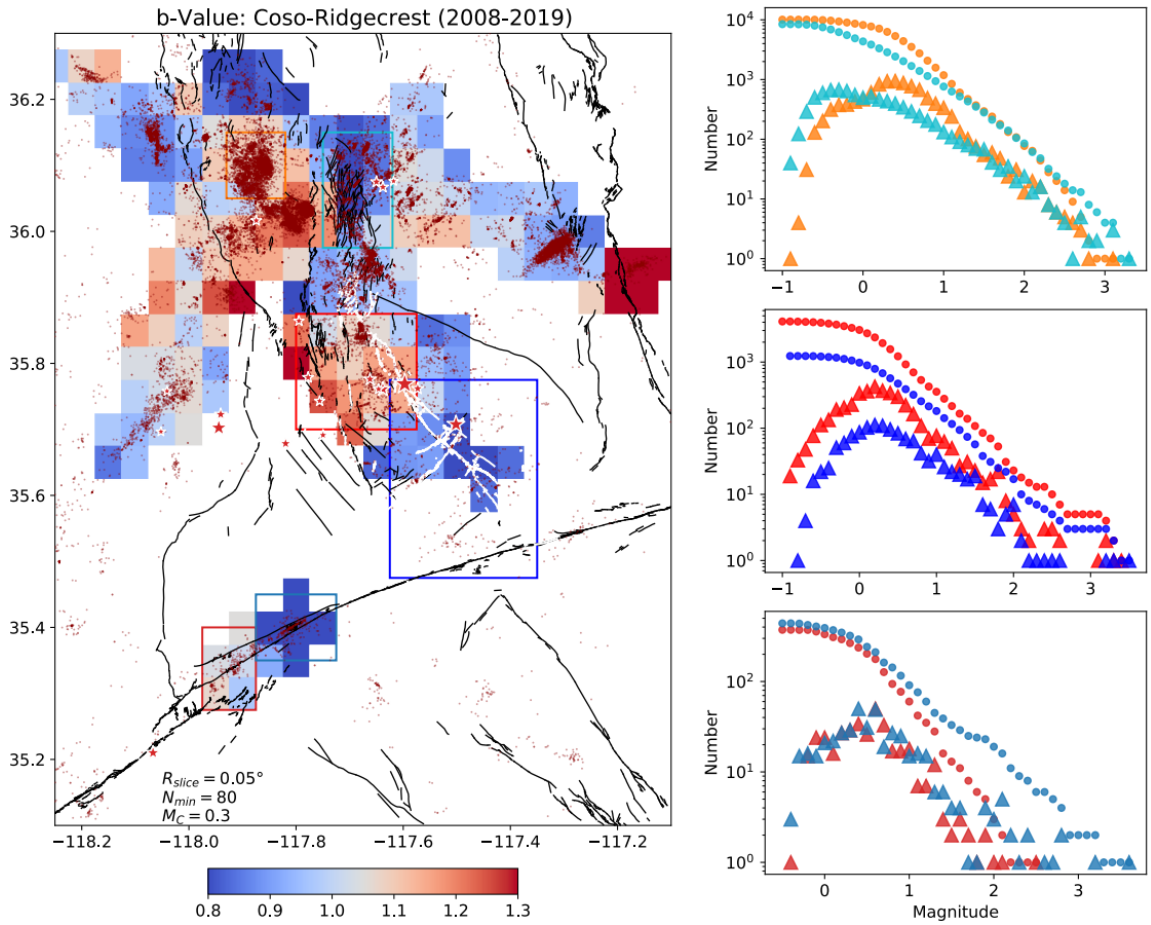

Figure S13. Example frequency-magnitude distribution. Three regions with contrasting b-values are compared. The dots and triangles mark the cumulative and non-cumulative distribution, respectively.

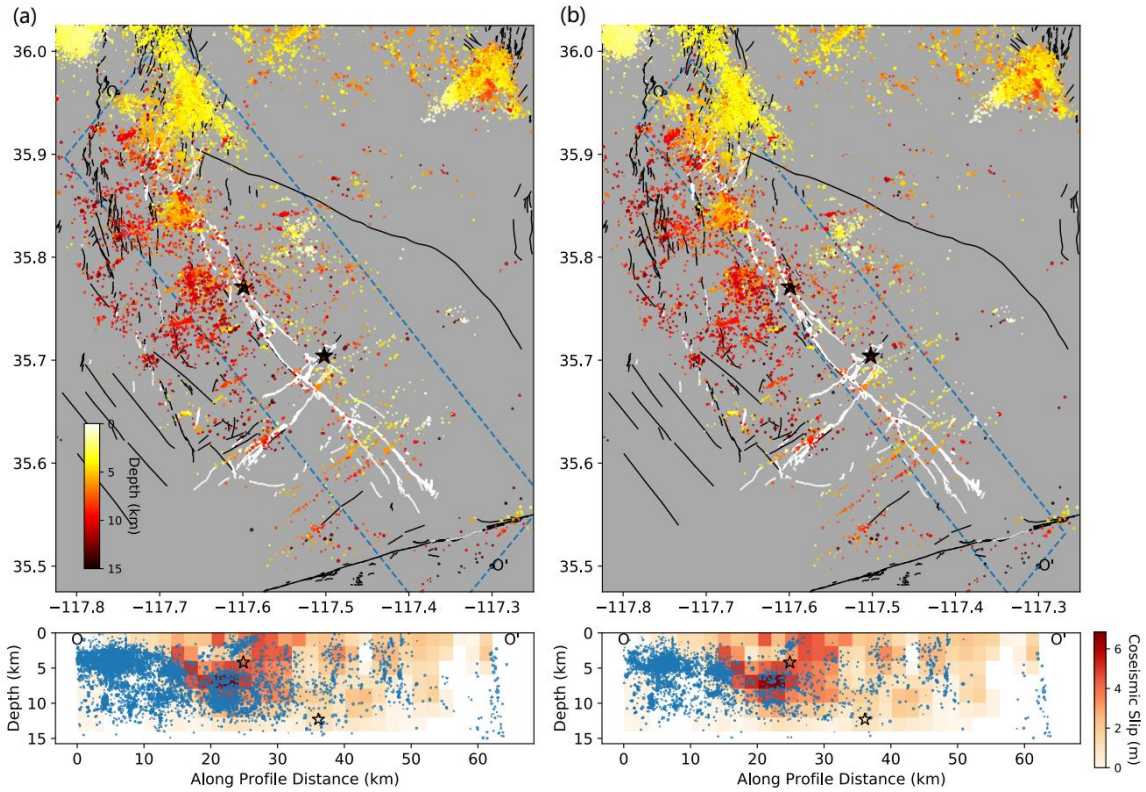

Figure S14. Clear version of Ridgecrest profile, with different profile widths.

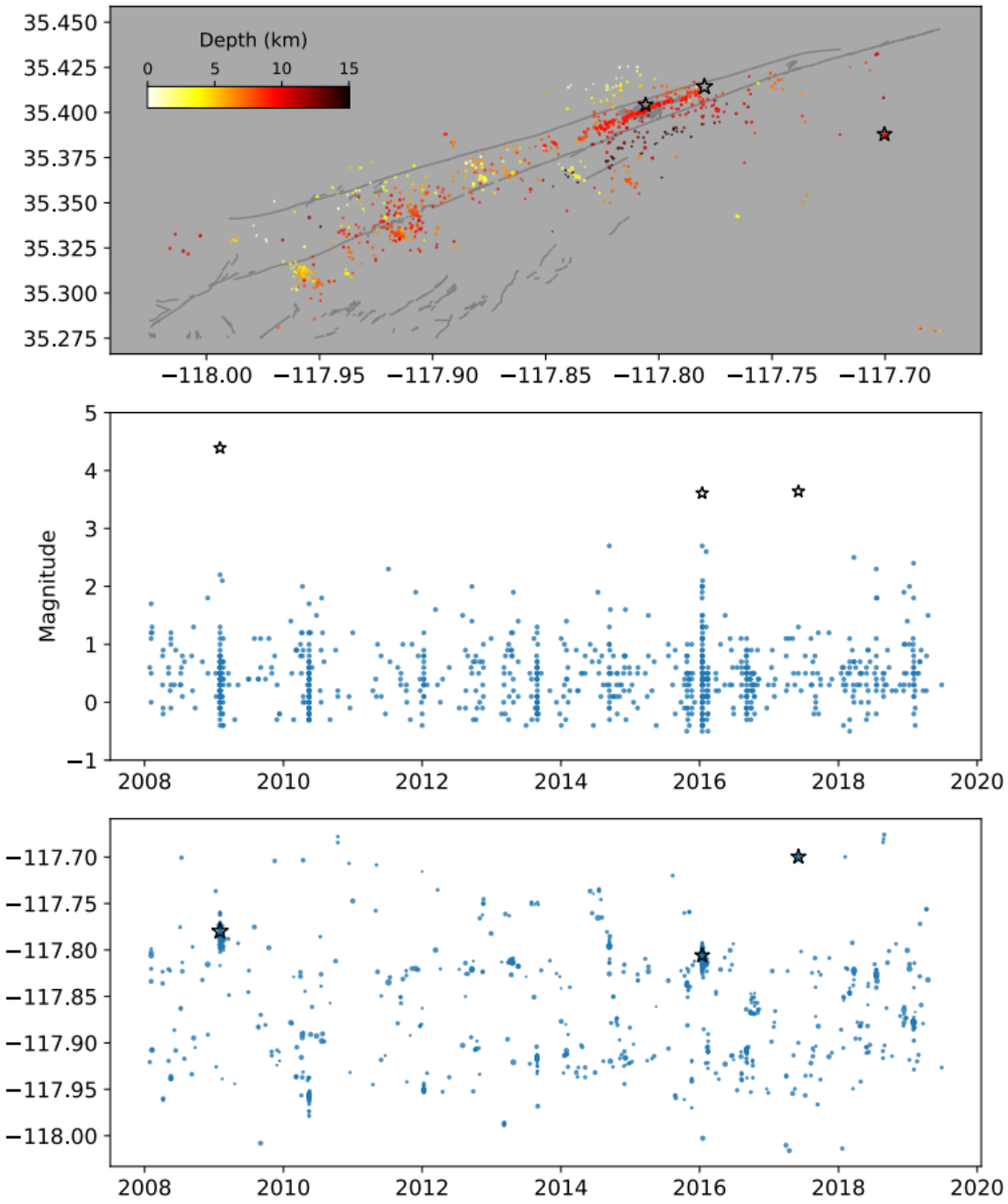

Figure S15. Spatiotemporal seismicity pattern on the Garlock fault. Starts mark the M>3.5 events.

Table S1. Velocity model used for location, modified from Hutton et al. (2010).

Top of layer (km)	P-wave velocity (km/s)
0	5.5
4	5.9
7	6.3
16	6.7
32	7.8

Note: Vp/Vs = 1.73; depth relative to ground elevation (1500 m)

Data Set S1. Seismic catalog for the Ridgecrest-Coso region during the preseismic period (20080101-20190703).

Data Set S2. Early aftershock catalog for the 2019 Ridgecrest sequence (20190704-20190724).

References

Aki, K. (1965). Maximum likelihood estimate of b in the formula log N= a-bM and its confidence limits, *Bull. Earthq. Res. Inst., Tokyo Univ.*, *43*, 237-239.

Hauksson, E., C. Yoon, E. Yu, J. R. Andrews, M. Alvarez, R. Bhadha, and V. Thomas (2020). Caltech/USGS Southern California Seismic Network (SCSN) and Southern California Earthquake Data Center (SCEDC): Data Availability for the 2019 Ridgecrest Sequence, *Seismological Research Letters*, *91*(4), 1961-1970, doi:10.1785/0220190290.

Hutton, K., J. Woessner, and E. Hauksson (2010). Earthquake Monitoring in Southern California for Seventy-Seven Years (1932–2008), *Bulletin of the Seismological Society of America*, *100*(2), 423-446, doi:10.1785/0120090130.

Klein, F. W. (2002). User's guide to HYPOINVERSE-2000, a Fortran program to solve for earthquake locations and magnitudes*Rep. 2331-1258*, US Geological Survey.

Mousavi, S. M., W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking, *Nature Communications*, *11*(1), 3952, doi:10.1038/s41467-020-17591-w.

Shi, Y., and B. A. Bolt (1982). The standard error of the magnitude-frequency b value, *Bulletin of the Seismological Society of America*, *72*(5), 1677-1687.

Waldhauser, F. (2001). hypoDD--A program to compute double-difference hypocenter locations, *U.S. Geological Survey Open-File Report 01-113*, 25 pp.

Zhou, Y., S. Zhou, and J. Zhuang (2018). A test on methods for MC estimation based on earthquake catalog, *Earth and Planetary Physics*, *2*(2), 150-162, doi:10.26464/epp2018015.

Zhou, Y., H. Yue, Q. Kong, and S. Zhou (2019). Hybrid Event Detection and Phase-Picking Algorithm Using Convolutional and Recurrent Neural Networks, *Seismological Research Letters*, *90*(3), 1079-1087, doi:10.1785/0220180319.

Zhou, Y., H. Yue, L. Fang, S. Zhou, L. Zhao, and A. Ghosh (2021). An Earthquake Detection and Location Architecture for Continuous Seismograms: Phase Picking, Association, Location, and Matched Filter (PALM), *Seismological Research Letters*, *93*(1), 413-425, doi:10.1785/0220210111.