Supplement to "Deep learning detects uncataloged low-frequency earthquakes across regions"

Jannes Münchmeyer[1], Sophie Giffard-Roisin[1], Marielle Malfante[2], William Frank[3], Piero Poli[4], David Marsan[1], and Anne Socquet[1]

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, Grenoble, France

[2]Univ. Grenoble Alpes, CEA, List, Grenoble, France

[3]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

[4]Dipartimento di Geoscienze, Università di Padova, Padova, Italy

1 Data and methods

1.1 Reference LFE catalogs and seismic datasets

In this manuscript, we work on four study regions: Cascadia (Canada/USA) (Bostock et al., 2015), the central section of the San Andreas fault (USA) (Shelly, 2017), Guerrero (Mexico) (Frank et al., 2014), and Nankai (Japan) (Japan Meteorological Agency, 2023). For each of these study regions, detailed LFE catalogs are available. For Cascadia, San Andreas, and Guerrero, the catalogs are based on template matching on multiple seismic stations. For Nankai, LFEs are determined in routine processing of the Japanese Meterological Agency (JMA) using a traditional workflow of phase picking and phase association.

Machine learning based event detection and phase picking workflows for regular earthquakes have predominantly been trained on hand-labeled catalogs with traditional processing pipelines (Zhu and Beroza, 2019; Ross et al., 2018; Münchmeyer et al., 2022). This guarantees high quality training data with a very low fraction of false positive examples. At the same time, hand picked catalogs are usually incomplete, i.e., miss a substantial number of events. In contrast, template matching catalogs tend to be very complete, at least for events with corresponding templates, but come at the cost of substantially more false positive detections (Scotto di Uccio et al., 2023). Especially for events with low signal-to-noise ratio (SNR), such as LFEs, template matching is usually applied at multiple stations at once with events only accepted if their correlation is sufficient across multiple stations (Shelly, 2017; Bostock et al., 2015; Frank et al., 2014).

The different characteristics of classical and template matching catalogs have implications for training machine learning models. Template matching catalogs do not provide picks for individual events but instead times where a templated matched to the continuous waveforms. If picks on the templates are available, this allow inferring pick times on the continuous data. However, template matches for LFEs are often non-unique, i.e., multiple templates match at the same location. In our analysis, this lead to uncertainties in the inferred pick times > 1 s.

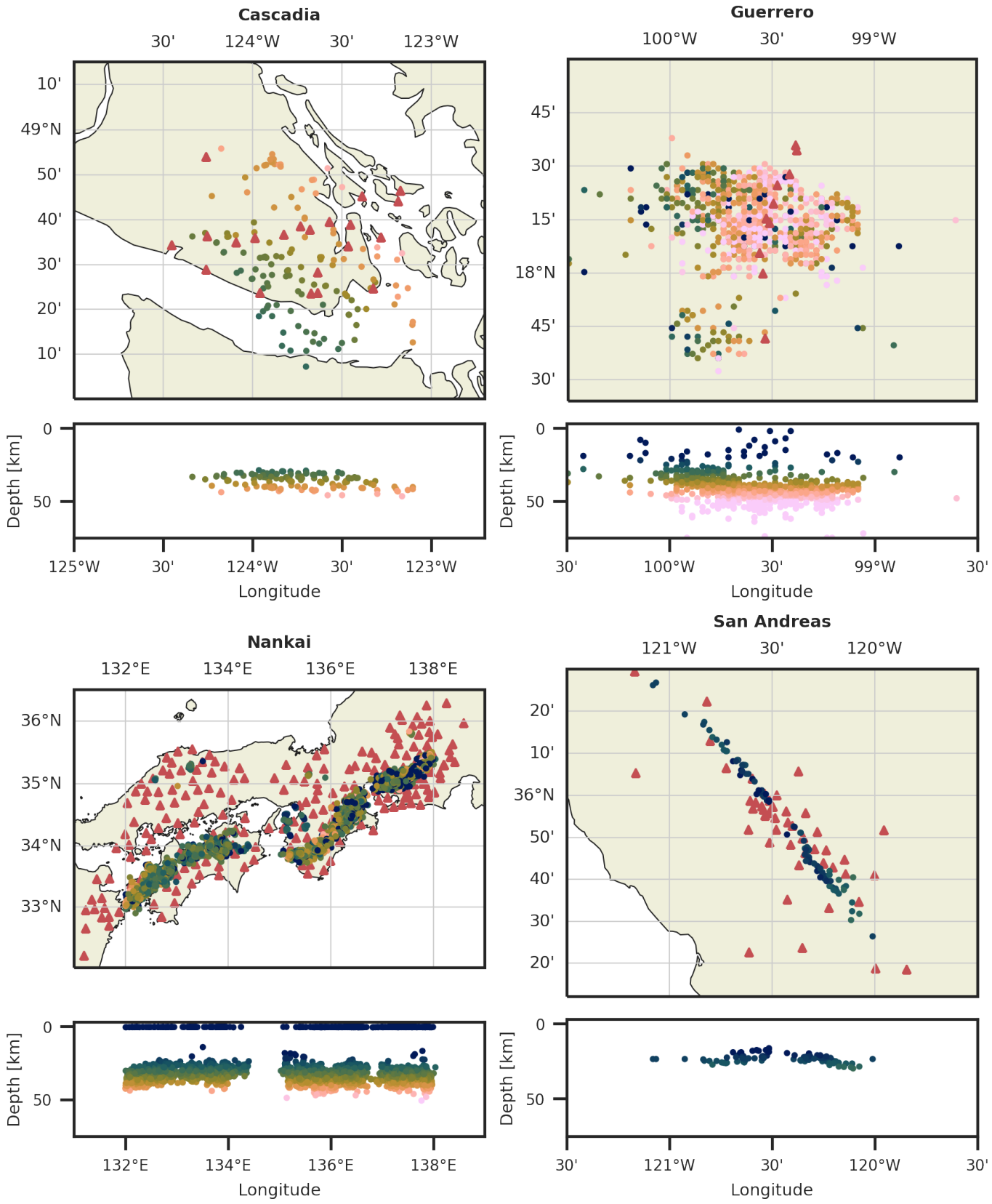

Figure S1: Spatial distribution of LFEs from the reference catalogs (dots) and stations (red triangles) in the four study regions. For Cascadia (Bostock et al., 2015), Guerrero (Frank et al., 2014), and San Andreas (Shelly, 2017), each dot represents an LFE family. For Nankai (Japan Meteorological Agency, 2023) individual LFEs are plotted. The event depth is encoded in color. We only show stations used for generating training data.

Another key difference between traditional earthquakes and LFEs is their temporal pattern. While EQs rarely occur with interevent times below tens of seconds, at least outside intense aftershock sequences, LFEs usually occur in burst

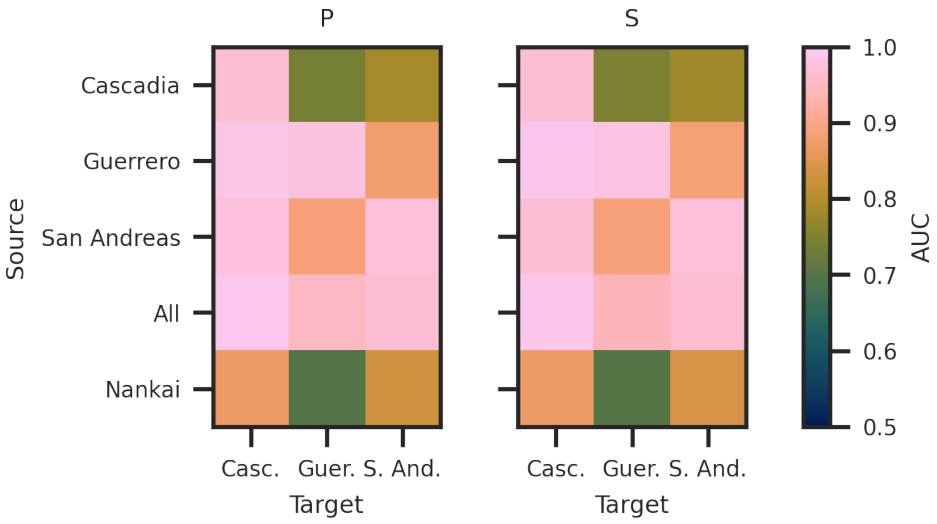

Figure S2: P and S phase detection performance within and across datasets at a signal-to-noise ratio of -2.5 dB. Each panel shows the area under the curve (AUC) of the receiver operating characteristics (ROC). The evaluation has been conducted on the synthetics generated from stacks and noise. No results are reported for evaluation on Nankai as the real data does not allow to define a gold standard.

with temporal spacings of only few seconds. Unfortunately, not all of these events can be correctly identified with template matching, leading to substantial numbers of missing annotations around existing annotations. This causes an issue for training typical detection models as these unidentified events are wrongly labeled as noise. With the dense spacing of events, this causes substantial amounts of label error and is detrimental to model performance. Therefore, we were not able to train LFE detection and picking models based on events from template matching catalogs.

This challenge impacts even more significantly the evaluation of model results. Traditional picking and detection algorithms are evaluated based on their detection performance and picking accuracy (Münchmeyer et al., 2022). Estimating detection accuracy on template matching catalogs is not possible, as the missing detections make it impossible to guarantee that windows do not contain LFEs. One option would be to select windows far outside LFE burst, however, these are too easy to distinguish from windows around LFEs that do not contain arrivals and, therefore, do not offer an appropriate metrics. We observed similar incompleteness issues for the Japan catalog, making an evaluation of detection capability on this data impossible. For picking accuracy, the inaccuracy of the inferred picks on the continuous waveforms makes it impossible to infer performance measures on the template catalogs. Given these difficulties in both training and evaluating LFE detection and picking models, different strategies than for regular earthquakes are required. We outline these in the subsequent sections.

1.1.1 Cascadia

The northern Cascadia subduction zone, beneath Canada and the United States, has long been known for hosting subduction megathrust events, slow slip events, and LFEs (Satake et al., 1996; Rogers and Dragert, 2003). Due to its low seismicity rate, the area is often employed as a natural laboratory for studying the connection between aseismic deformation and their seismic signature (Rouet-Leduc et al., 2019).

For this study, we use the LFE catalog by Bostock et al. (2015). We use 21 seismic stations from the C8, CN and PO networks. The catalog contains 129 LFE families, located underneath Southern Vancouver island and offshore to the South. All families locate at depth between 35 and 50 km with a Northwest dipping distribution that is following the subduction interface. We note that tremor observations suggest that further LFEs occur to the South and North of

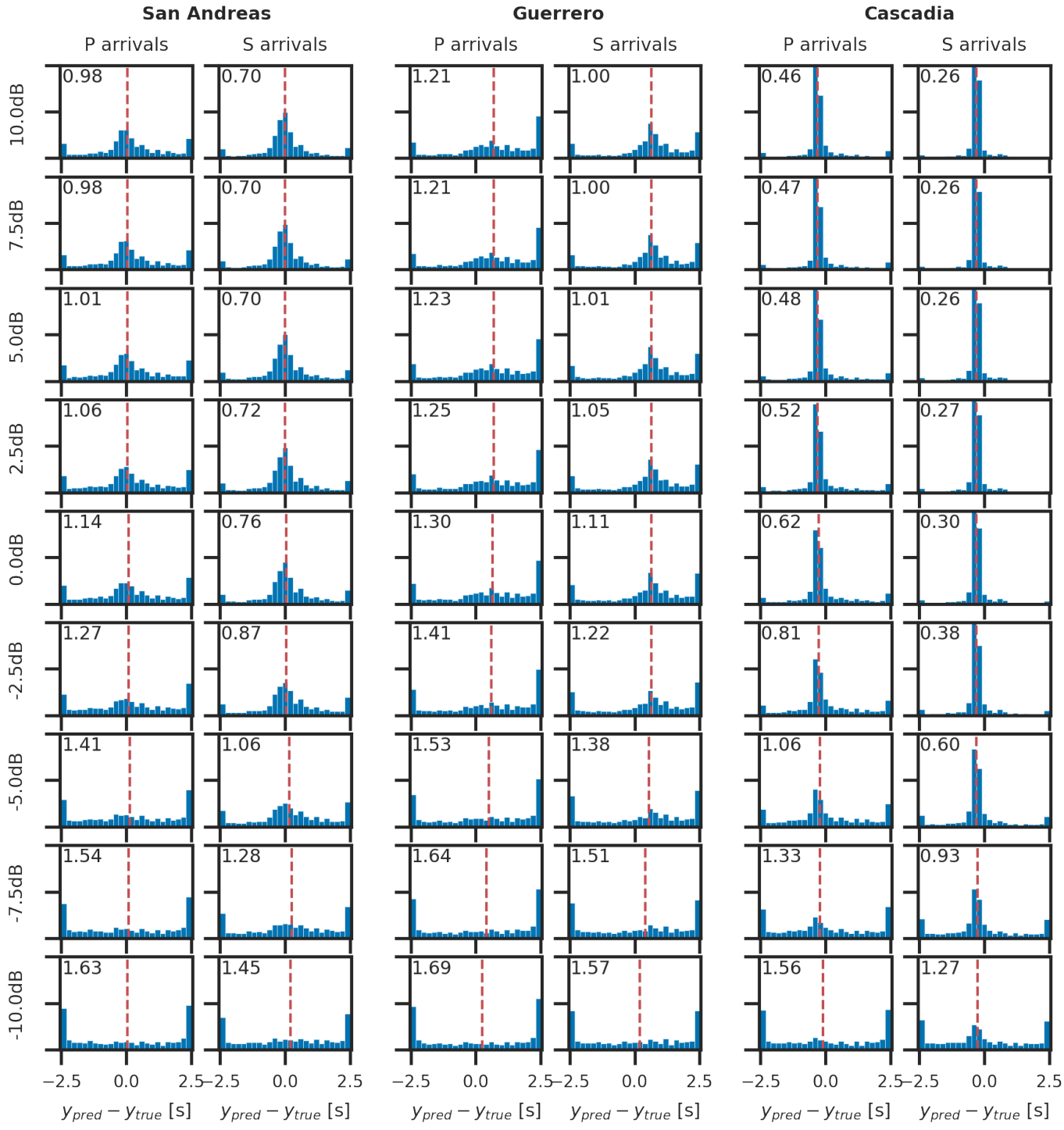

Figure S3: Pick accuracy at different signal-to-noise ratios. Number in the corner indicate mean absolute errors (MAE) in seconds around the median prediction. Red vertical lines indicate the median values. Larger bars on the sides are caused by windows in which no pick has been identified. For each 5 second window, we take the time of highest predicted probability as the pick time.

the cataloged LFEs (Wech, 2021). These could likely not be tracked using the limited coverage of the stations used by Bostock et al. (2015). The data spans the years from 2003 to 2012, with 80 % of the waveforms coming from the years 2003 to 2005, as these years have the best station coverage. The catalog only spans major slow slip events (M $>$ 6) and not the full duration from 2003 to 2012. In total, we extracted 2703 templates (1621 training / 271 development / 881 test) from the reference catalog.

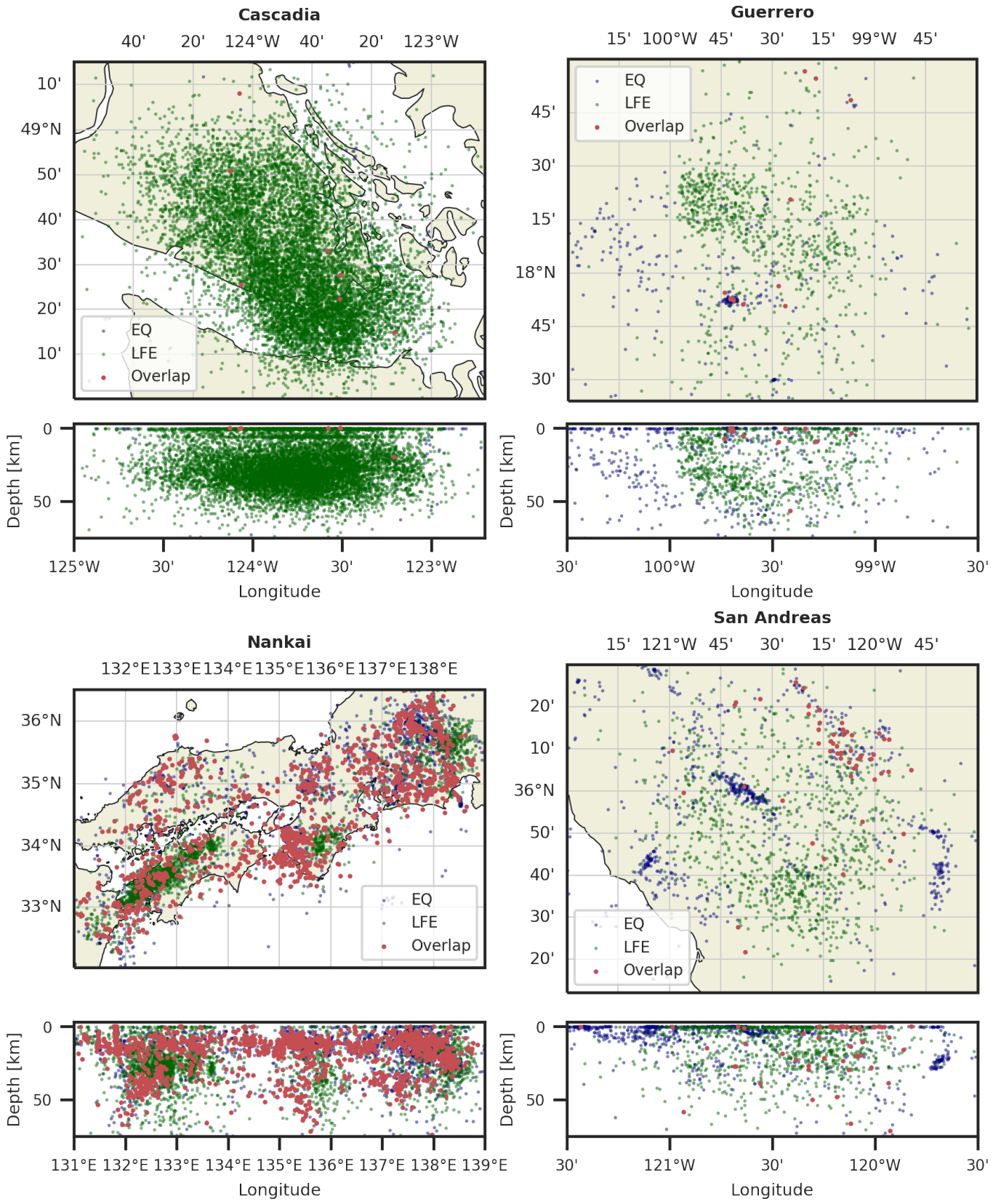

Figure S4: Spatial visualisation of the overlap between LFE and EQ detections. Blue dots denote exclusive EQ detections, green dots exclusive LFE detections, red dots joint detections. Joint detections are visualised at the location determined with the EQ picker as these locations are more accurate. The numbers of unique EQs/unique LFEs/overlapping events are: 89/10211/8 (Cascadia), 612/876/12 (Guerrero), 2329/2525/1787 (Nankai), and 992/975/57 (San Andreas).

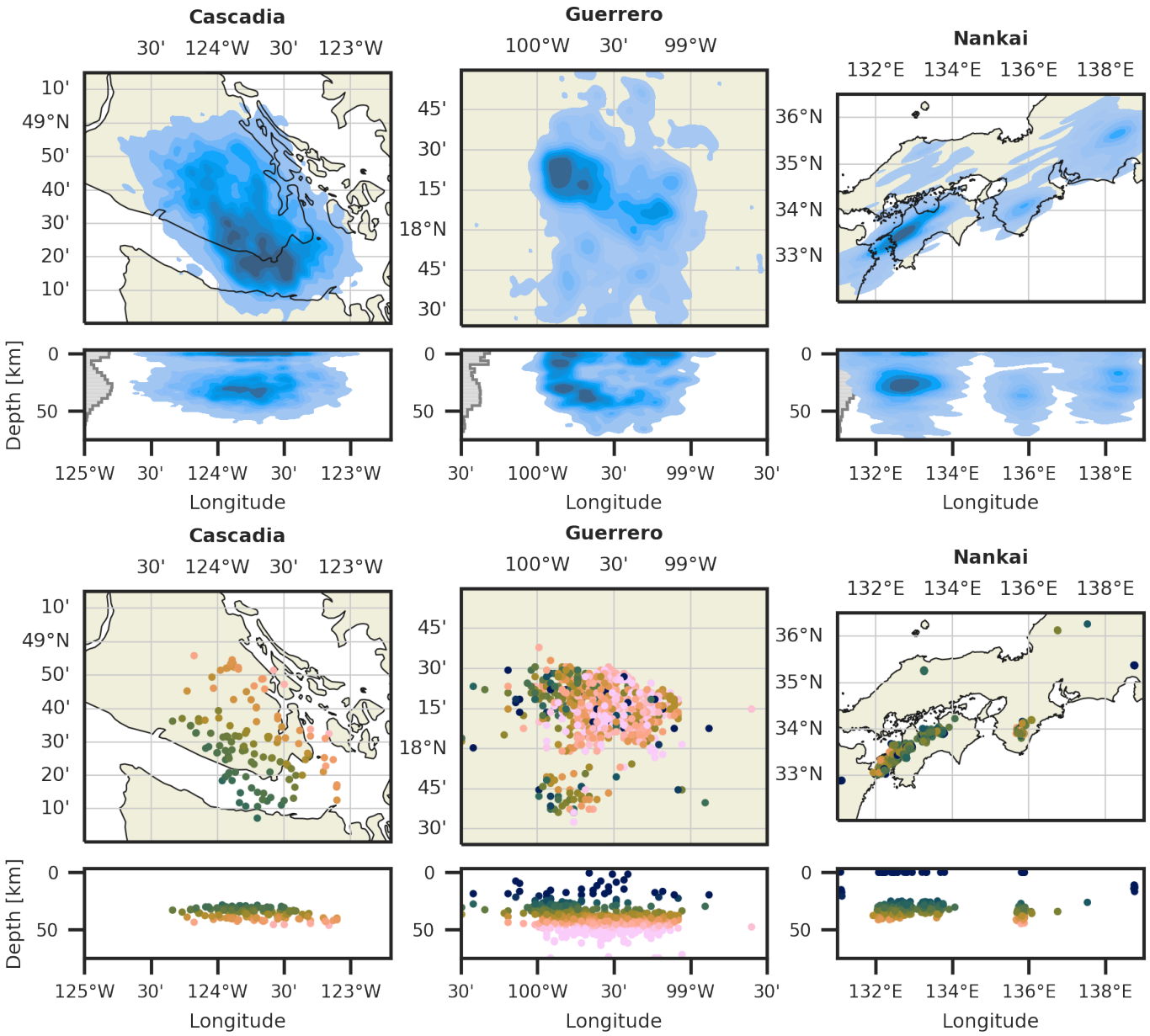

Figure S5: LFE density plots (top) and reference catalogs (bottom). For both the horizontal and vertical cross sections, the remaining axis is ignored.

1.1.2 San Andreas

The San Andreas fault runs along the West Coast of the United States and regularly causes major earthquakes (Peng and Zhao, 2009; Ross et al., 2019). The fault exposes a strong segmentation along strike with intermittent locked and creeping sections. The region stands in contrast to the other regions in this study, as the LFEs occur along a strike-slip fault and not in a subduction zone as in the other examples. We still include this dataset to increase the diversity of training data available to our models.

Shelly (2017) studied the central San Andreas fault and created a 15 year catalog of low frequency earthquakes. Using templates for 88 individual LFE families, they detected more than one million individual LFEs. All LFE families are located in close proximity to the fault line in a narrow depth band between 20 and 30 km. They employed data from 11 stations in the High-Resolution Seismic network (HRSN), a borehole network installed in the Parkfield area. Additional stations have been employed on an ad hoc basis for locating LFE families but not for detection.

For our study, we created waveform templates by stacking the waveforms according to the families and event times

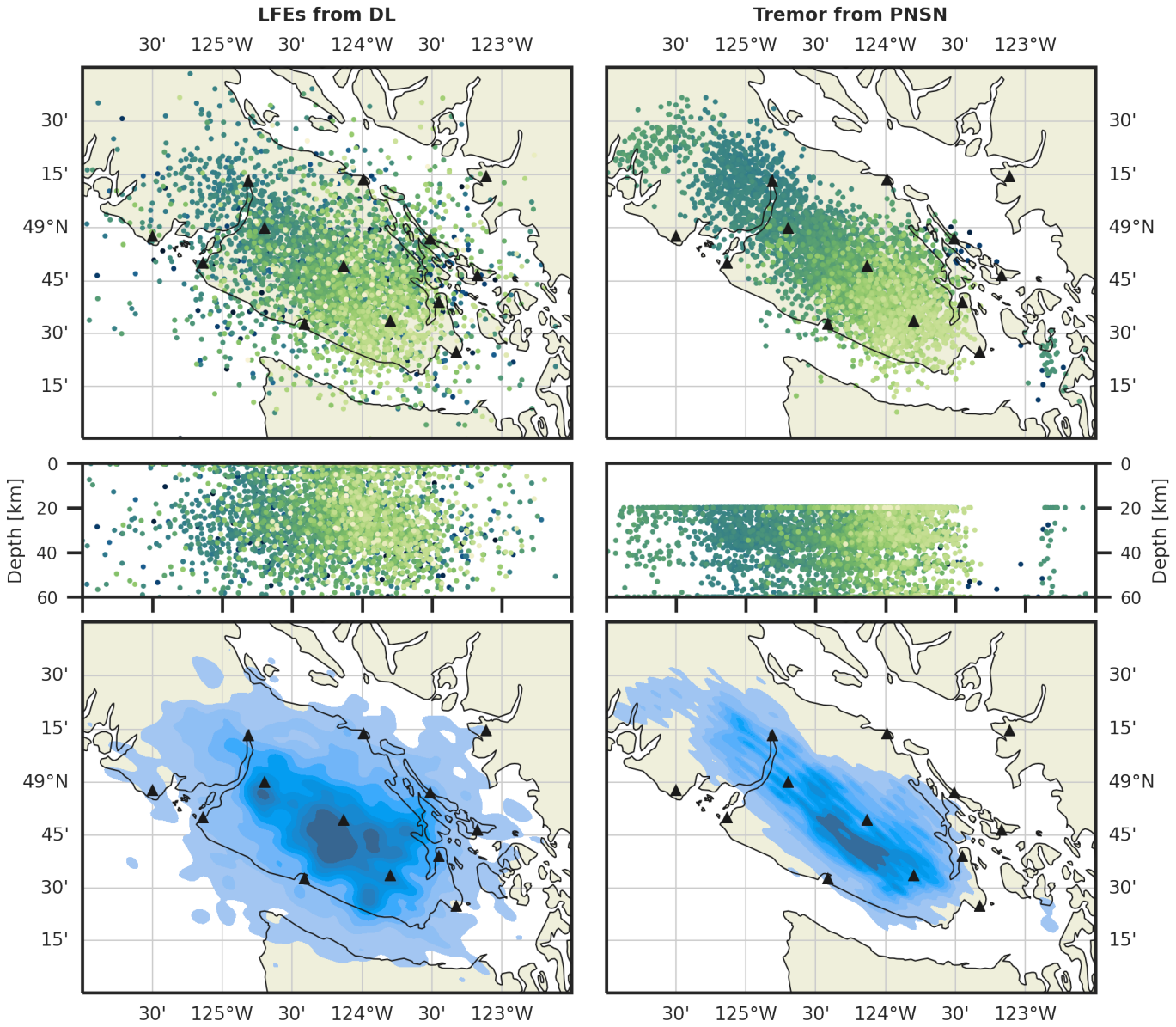

Figure S6: Maps of LFEs detected with deep learning (left) and tremors from the PNSN catalog (right) in the period from 2021-01-15 to 2021-02-15. In the top plots, color indicates time within the sequence, from dark blue (early) to light green (late). Black triangles indicate the stations used for generating the LFE catalog. All stations are from the CN network and are used for the PNSN tremor catalog as well. We note that tremor and LFE catalogs are only partially comparable because they map distinct, even though related, phenomena. In particular, LFEs are essentially point processes, while tremors are extended in time. This makes tremor detection and location easier. Nonetheless, even high-quality tremor catalogs tend to show artifacts, such as the gridding visible in the figure in both depth and latitude/longitude (rotated).

indicated in the catalog. For each families we created templates at the 11 HRSN stations used for template matching. In addition, we included further stations on a family basis as employed for location by Shelly (2017). Templates with insufficient data have been discarded. We use the original phase picks on the stack and manually picked missing arrivals. In total, we extracted 3889 templates (2333 training / 389 development / 1167 test).

1.1.3 Guerrero

The Mexican subduction zone has long been known as hosting slow slip, tremor, and LFEs (Larson et al., 2007; Payero et al., 2008; Frank et al., 2013). With the trench located close to the shore and an almost 200 km long, flat subduction

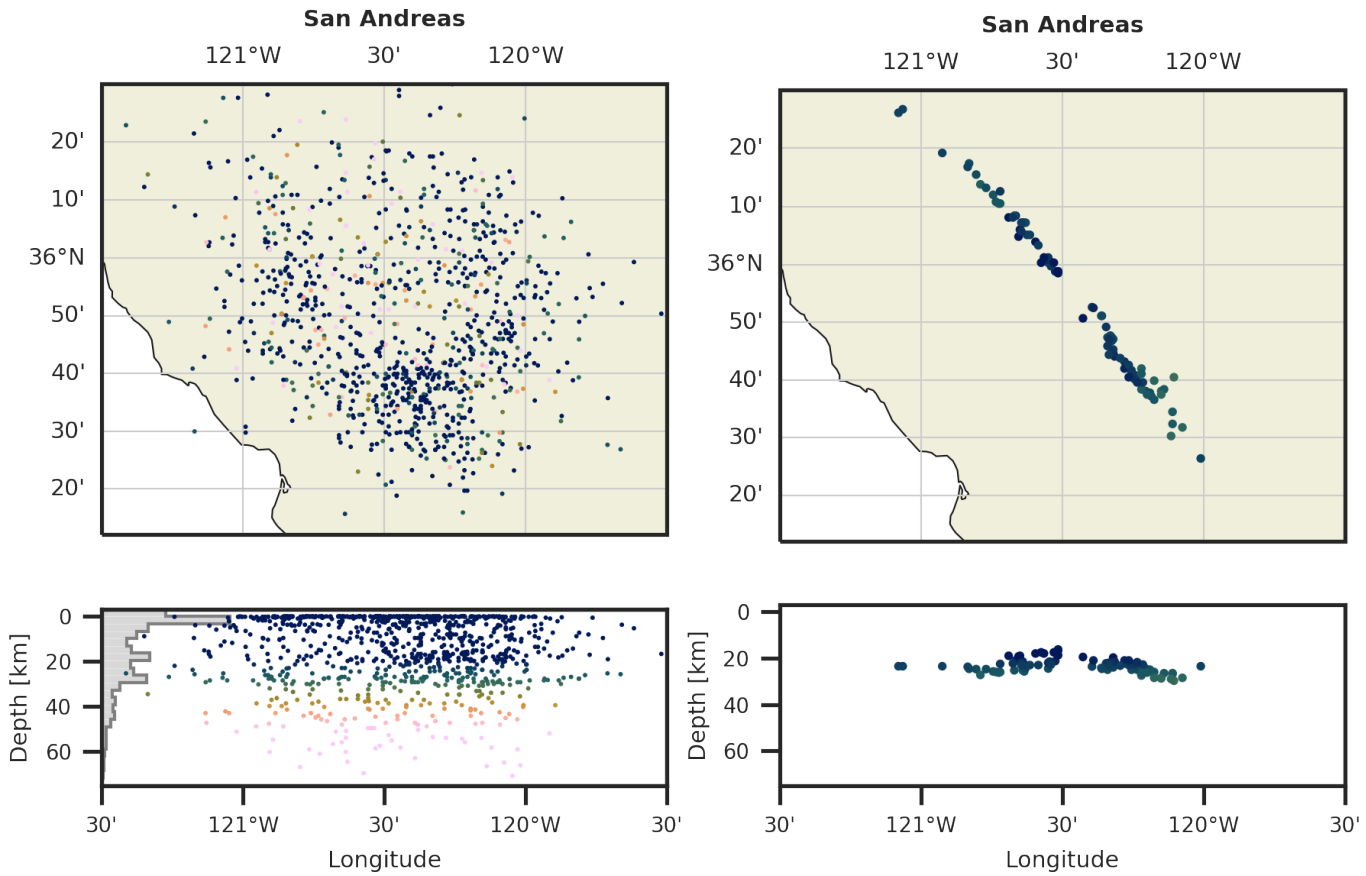

Figure S7: LFE catalog obtained using the deep learning (left) and reference catalog (right) for San Andreas. For the deep learning catalog, each dot represents one LFE. In the reference catalog (Shelly, 2017), each dot represents a family. Color encodes event depth. The bottom panels show depth cross-sections, showing longitude and depth of events. The histogram on the left of the cross-section shows the depth distribution of the detected events.

interface at a depth of around 45 km, the region is favourable for the observation of such phenomena. In this study, we use the catalog of Frank et al. (2014). The catalog contains 1120 LFE families located around the subduction interface. The families cover a range of roughly 0.5 by 1.0 degrees with the majority of families locating between 30 and 50 km depth. Based on these families, Frank et al. (2014) identified ∼1.85 million individual LFE occurrences. However, many events overlap, suggesting that there is overlap between the spatial coverage of individual template events.

For training, we use the same 10 stations from the MASE deployment of the TO network as Frank et al. (2014). We note that the MASE deployment is a line array, which limits the location precision perpendicular to the array. In total, we extracted 11200 templates (6720 training / 1119 development / 3361 test).

1.1.4 Nankai

The Nankai trough offshore Southern Japan is the first region worldwide where NVTs and LFEs have been described (Shelly et al., 2006). The region is known for hosting the full range of slow earthquakes, including VLFEs, LFEs and SSEs (Araki et al., 2017; Nakano et al., 2018). Along two elongated bands of roughly 200 and 300 km extent along strike, intense LFE activity has been observed. The events are distributed between 30 and 50 km depth and dip towards the North-West in accordance with the subduction. The majority of events occurs underneath land, leading to an exceptionally good station coverage and small azimuthal gaps.

We use the LFE catalog from the JMA, including all events between 32.5° and 36° North and 132° and 138.5° East (Japan Meteorological Agency, 2023). We use all events from 05/2008 to 04/2010. For each event, we include waveforms

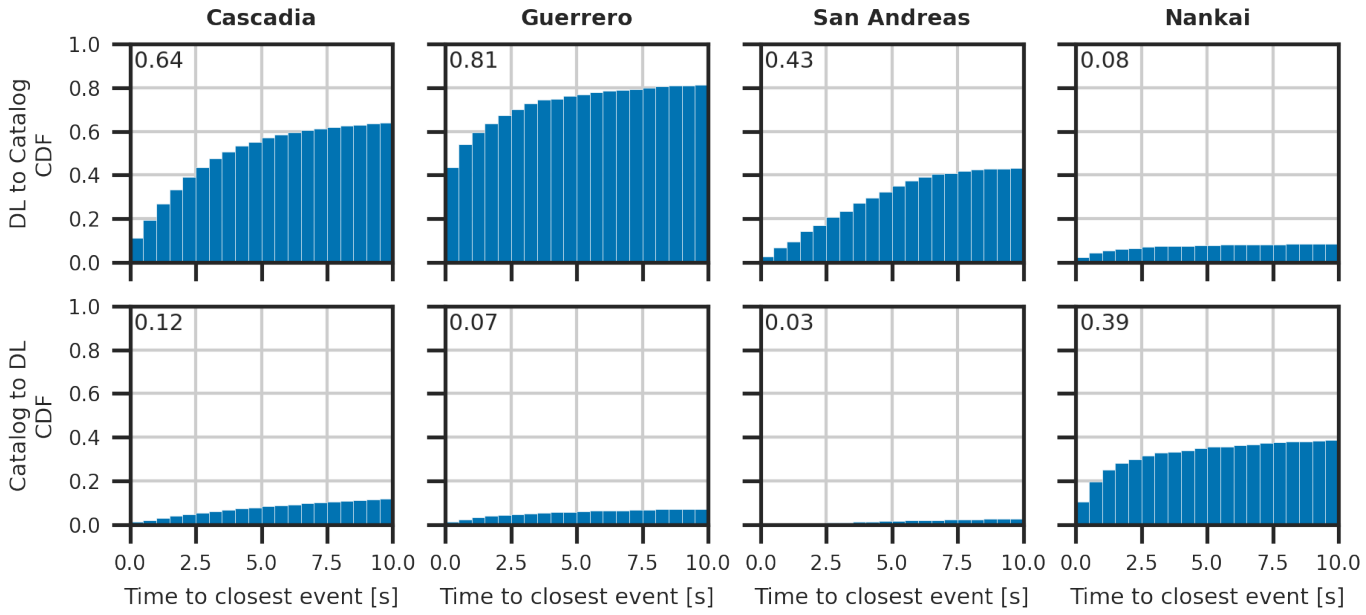

Figure S8: Temporal matching between events from the deep learning and the reference catalogs. Each panel shows the cumulative distribution function of the time difference between an event in one catalog and the closest catalog in the other catalog. The top row matches events from the deep learning catalog to the reference catalog, the bottom row the other way around. The numbers in the top left corners indicate the fraction of events with a reference event within 10 s. Note that these statistics are influenced by the total size of the catalogs. As the template matching catalogs usually do not report origin times but instead the time the template aligns, origin times for these catalogs have been estimated and carry higher uncertainties.

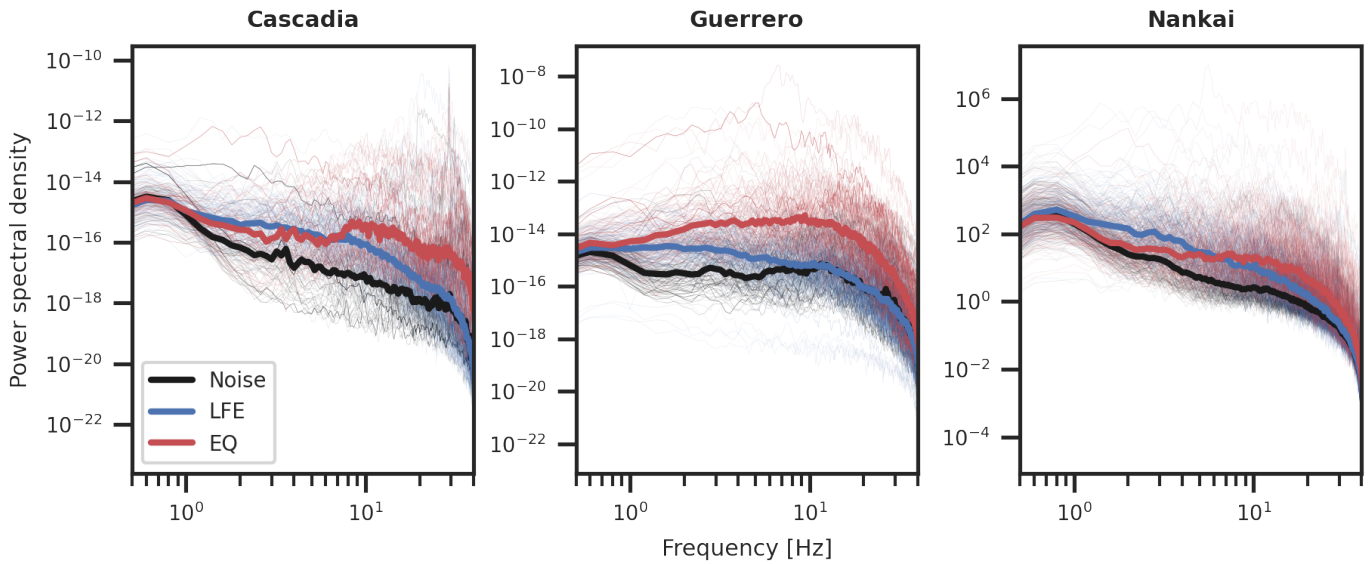

Figure S9: Velocity power spectral density for noise, earthquakes, and LFEs detected using deep learning. In contrast to Figure 4, spectra have been extracted from all stations available. See the caption of Figure 4 for further details.

from all FNet stations with at least one pick for the event. The resulting dataset consists of 57,702 traces of 7,996 events. There are 1,550 labeled P phases and 57,564 labeled S phases. We discuss mitigation strategies for the strong label incompleteness for P waves in the subsequent description of the model training. We split the dataset into training, development and test set between events, i.e., all traces for one event go into the same subset. We employ a chronological split with ratios 0.6 / 0.1 / 0.3 (training / development / test).

1.2 LFE picking model

For detecting LFEs and determining their phase arrival times, we build a deep learning network. Our network is a minor modification of PhaseNet (Zhu and Beroza, 2019). PhaseNet has a simple architecture and showed excellent performance in a recent benchmark on earthquake picking (Münchmeyer et al., 2022). We retain the basic original architecture of PhaseNet: a U-Net consisting of an encoder branch, a decoder branch, and residual connections. Compared to the original model, we modify the input shapes and the number of layers. Our input trace consists of 60 s of waveforms at a sampling rate of 20 Hz for a total of 1200 samples. Input waveforms are bandpass-filtered between 1 to 8 Hz as this is traditionally the main frequency band for LFE detection. Our output consists of 1200 samples as well, covering the same time range and representing P and S arrival probabilities. In contrast to the original PhaseNet, we remove the noise probability trace in the output and replace the final softmax layer by a sigmoid layer. This enables the labelling of overlapping P and S phases. We base our implementation on SeisBench (Woollam et al., 2022) and make our datasets, model and trained weights available through the software.

For training the model, we encode the phase arrivals as Gaussian peaks in the probability curves with a standard deviation of 0.5 s. We train the model using a binary-cross entropy loss independently for P and S phase. In the Japanese data, substantially fewer P than S picks were annotated. To reduce the impact on the probability curves, we down-weight the loss for the P wave curve to 20 % for the time from 10 s to 1.5 s before the S wave arrival, if no P wave has been annotated. This incentivises the model to pick more P arrivals based on the annotated ones and leads to higher confidence scores. Keeping the loss at a lower weight is beneficial over zeroing out the loss, as zeroing leads to high numbers of incorrect P picks.

We use two different data generation procedures. For Nankai, we select a random window of 60 s as input and label all contained P and S phases. To increase the diversity and train the model for lower SNRs, we mix the traces with noise traces. Noise weights are drawn from a Gamma distribution with shape parameter $a = 0.3$. The Gamma distribution will produces a similar level of noise for many examples, while including high noise examples according to its exponential tail. For the template matching datasets, we create synthetic examples from the waveform stacks. First, we randomly take up to three stacks. The stacks can originiate from different stations and even regions. Second, we sum these stacks with random time offsets and select a random window of 60 s duration. Third, similar to Nankai we mix the stacks with noise. All noise traces are taken from the INSTANCE noise dataset (Michelini et al., 2021). We use these waveforms from Italy as there are no known tectonic tremors or LFEs in Italy, making labelling errors unlikely. Noise scales are drawn from a Gamma distribution, the shape parameter a is increased from 0.1 to 5 during training with a geometric spacing of intermediate values. This means that the model training starts with easier high SNR examples before training on harder examples with low SNR. In addition to these two data generation procedures, for 20 % of the examples we provide the network with pure noise traces. We use the same data generation procedure for the validation dataset but keep the scale parameter of noise distribution constant at $a = 0.3$ for Nankai and $a = 2$ for the stack datasets. This makes the validation loss values comparable across epochs. Figure 1 shows an example of the combined stacks, the noise, the mixed stack and noise, and the labels. Even though we use a low noise level of 0.5, the LFE arrivals are already hardly identifiable by eye.

To further increase the diversity of our data, we use data symmetries. Phase picks should be invariant to the orientation of the sensor's horizontal components, i.e., to rotation around the Z axis. Similarly, they should be invariant to sign changes of the horizontal components. We therefore randomly rotate each training example around the Z axis. We further create five copies of the data: rotated by $\frac{2}{3}\pi$ and by $\frac{4}{3}\pi$, and mirrored versions of the three rotated copies. We feed all copies to the model at once, increasing the effective batch size by factor 6. This has the side benefit of substantially improving GPU utilisation and thereby reducing training times. As it is not the main focus of this paper, we do not study the effect of these augmentations in detail.

We train the model jointly on all four dataset for 250 epochs. We use the model weights from the epoch with lowest validation loss. We use the Adam optimizer with a learning rate of 10^{-2} or 10^{-3} and a batch size of 1024. We optimized the hyperparameters using a grid search and selected the parameters with minimum validation loss. While validation loss is a good proxy for model performance, it is usually not the first choice for model selection. However, in this case we do not have access to a more suitable performance metric due to the data issues discussed above and, therefore, resort back to the validation loss. As we discovered systematic picking offsets between the datasets, for the joint model, we introduce constant shifts for all picks to homogenize the arrival times.

1.3 LFE catalog workflow

To build LFEs catalogs from continuous data, we integrate the picking model into a pipeline. First, we use the model to generate candidate picks. To apply the model to continuous data, we use a sliding window approach. We use an overlap of 600 samples, i.e., each sample is covered by exactly two windows. The windowed predictions are merged using the average prediction for each sample.

For declaring picks, we use fixed thresholds and use the local maximum of the pick probability curve as pick time. To identify LFEs, we associate the picks using the PyOcto associator (Münchmeyer, 2023). We conducted additional experiments with the REAL (Zhang et al., 2019) and GaMMA (Zhu et al., 2022) associators that gave viable catalogs as well but showed overall worse performance. We set region-specific requirements on the minimum number of picks. All events are relocated using NonLinLoc (Lomax et al., 2000) with 1D velocity models. Events with too high RMSE are discarded. We use station travel time residuals for PyOcto and NonLinLoc to improve the association and location. To infer the station residuals, we calculated earthquake catalogs using the same pipeline but with a PhaseNet picker (Zhu and Beroza, 2019) from SeisBench (Woollam et al., 2022) trained on INSTANCE (Michelini et al., 2021). The region-specific parameter settings for the catalog generation and quality control are detailed in Table S1. We parallelise the pipeline using the dask workflow management system (Dask Development Team, 2016) and execute it on a high performance cluster.

2 Performance evaluation on synthetics

For a quantitative evaluation, we test our model on synthetic data compiled from stacks and noise. For each dataset, we generated a test set of 20480 examples that we kept fixed across all evaluations. These example were generated using a set of stacks not employed in training. All results presented are from these independent test sets, while parameters have been selected on another independent development set with 20480 traces based on distinct stacks. As no stacks are available for the Nankai model and the data therefore lacks a reliable ground truth, we do not report results for Nankai here.

We design our synthetic evaluation after Münchmeyer et al. (2022). In addition to the traces, we define a collection of target windows. For each window, we extract the maximum predicted probability for the window being a P or S wave. For true examples, i.e., examples containing a P or S wave, we use a window of 5 s centered around the arrival. We use two types of noise examples. First, we select 5 s windows consisting exclusively of noise from INSTANCE. Second, we select 5 s windows of the synthetic stack plus noise waveforms that do not contain any arrival. These are substantially more challenging than the pure INSTANCE noise windows as they are usually in the coda of other LFEs. At the same time, these windows need to be included for a realistic evaluation. We analyse our results using receiver operating characteristics (ROC) and the area under the ROC curve (AUC). These metrics are threshold-independent and therefore do not require a choice of detection threshold for the evaluation.

We conduct evaluations at different noise levels. We define the signal-to-noise ratio (SNR) as the ratio between the

standard deviation of the stacks without noise to the standard deviation of the noise example. We use SNR values from +10 dB (signal is factor 10 above the noise) to -10 dB (noise is factor 10 above the signal).

Figure 1 shows the ROC curves for P and S waves in all regions using the model trained jointly on all 4 datasets. At high SNR values, all regions have SNRs of 0.98 to 1.00, indicating excellent performance. However, we note that the dataset for evaluation is balanced between signal and noise classes, while in continuous waveforms, noise is substantially more common. In this light the absolute values of the AUC should be interpreted with caution. The detection performance for both P and S waves stays almost constant until roughly -2.5 dB and degrades afterwards to values between 0.88 (San Andreas) and 0.93 (Cascadia) at -10 dB. This means, that event at low signal to noise conditions, the model still performs substantially above chance. We do not observe a systematic difference between P and S phase detection capabilities.

Seismic phase pickers for regular earthquake transfer well across regions (Münchmeyer et al., 2022). To identify if a similar transferability exists for LFE pickers, Figure S2 shows the AUC values in a cross-domain setting, i.e., with different training and test regions. Overall, while the in-domain models generally perform best, the off diagonal elements show good performance too. In particular, the models trained on Guererro and San Andreas show performance similar to the in-domain model when applied to Cascadia. In contrast, the models trained on Cascadia perform worse when applied to the other regions. The joined model, trained on all four datasets performs for every region similarly good, sometimes better than the models from single regions. We further suspect that this model has better generalisation performance, which can, however, not be analysed based on the templates. For reference, we include the performance of the model trained exclusively on real LFEs waveforms from Nankai. The model performs substantially inferior, however, this is likely caused by the model not being trained towards the specificities of the synthetic data.

In addition to the sensitivity of the phase picker, we are interested in the picking accuracy, i.e., whether the automatically picked times match the manual pick times. Figure S3 shows the pick timing of the joint model on the three stack datasets at different noise levels. Performance varies substantially between the datasets, with the lowest residuals for Cascadia, followed by the other two datasets. In all cases, residuals for S waves are lower than for P waves by a margin of about 20 to 50 %. This is consistent with the traditional understanding that LFE P arrivals are harder to pick than LFE S arrivals due to their lower SNR. Arrival time precision degrades noticable with the SNR, with substantial degradation setting in around -2.5 dB, similar to the observations for detection. Nonetheless, across all noise levels and datasets, the median prediction is always close to zero, i.e., the picks are always unbiased or only have little bias. Comparing the mean absolute error (MAE) of the LFE picker for LFEs to those of a deep learning earthquake picker for regular earthquakes, the MAE is higher by a factor of at least three (compared to INSTANCE from Münchmeyer et al. (2022)), but can even go up to more than 10 (compared to STEAD). Therefore, any downstream workflow needs to be tolerant to substantially higher pick uncertainties.

Table S1: Parameters for the LFE detection workflow. The parameters are grouped into the velocity model, picking parameters, and association parameters. Parameter names follow the naming in the underlying tools SeisBench and PyOcto.

Parameter	Cascadia	Guerrero	Nankai	San Andreas
Velocity model	Bostock et al. (2015)	Domínguez et al. (2006)	JMA2001	Small et al. (2017)
P_threshold	0.1	0.1	0.15	0.1
S_threshold	0.1	0.1	0.15	0.1
overlap	600	600	600	600
minimum number picks	13	14	12	10
minimum number p and s picks	2	3	3	2
maximum std nonlinloc	0.85	1.0	0.75	0.85
pick match tolerance	2.5	2.0	2.0	2.0
velocity model tolerance	2.0	2.0	2.0	2.0
association cutoff distance	-	-	150	-
min pick fraction	0.25	0.25	0.4	0.25

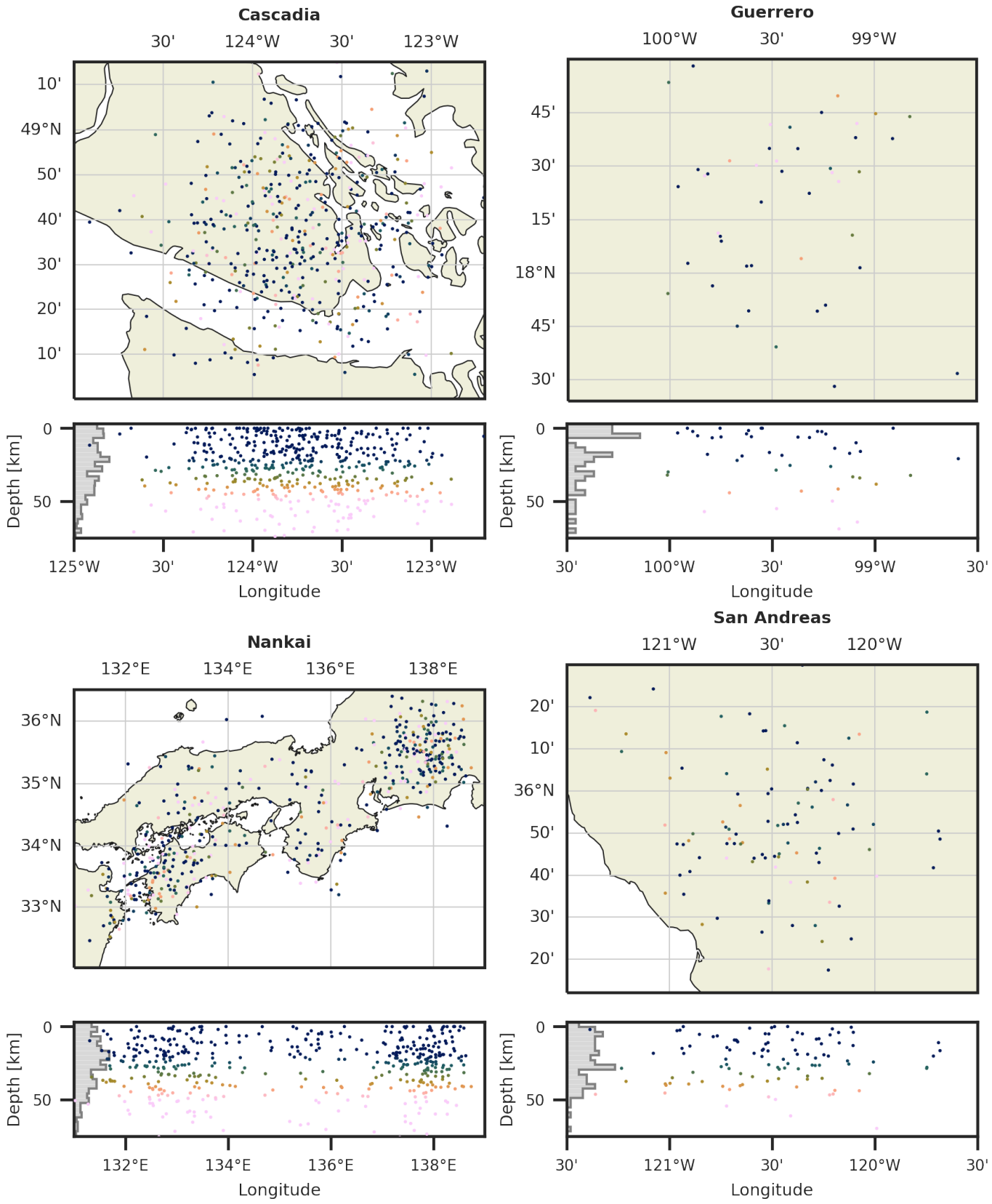

Figure S10: Example catalogs with scrambled pick times. For each region, the original raw picks from the LFE picking model have been scrambled. For this, all picks at a station within one hour are shifted by the **same** offset. Offsets differ between hours and stations. Offsets are drawn randomly from a Gaussian distribution with 30 s standard deviation. This strategy ensures that P to S times, the number of picks per station, and higher numbers of picks within tremor sequences are preserved. The scambled picks are processed using the same association and location pipeline as before, including the same quality control criteria. The higher number of detections in Japan is related to the very high station number, making associations even after scrambling likely.

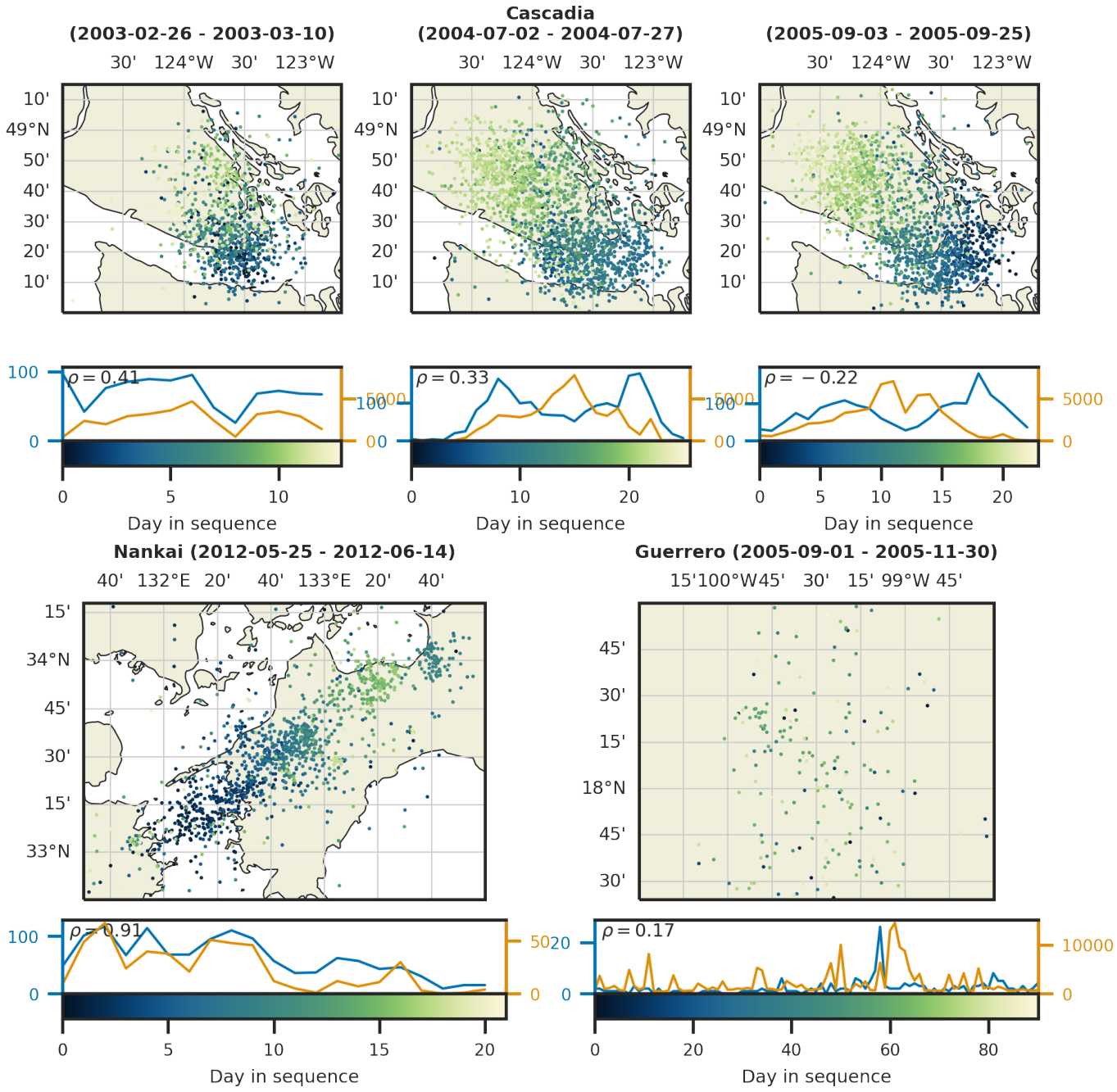

Figure S11: Spatial migration patterns of all LFEs without a match in the reference catalogs. A match to the reference catalog is defined as a reference event occuring within 10 s of the detected event. Note that the substantially lower correlation values between the number of events here and the reference catalogs are expected: here we only count events not in the catalog and it is likely that the fraction of such event changes over time with the migration of the events.

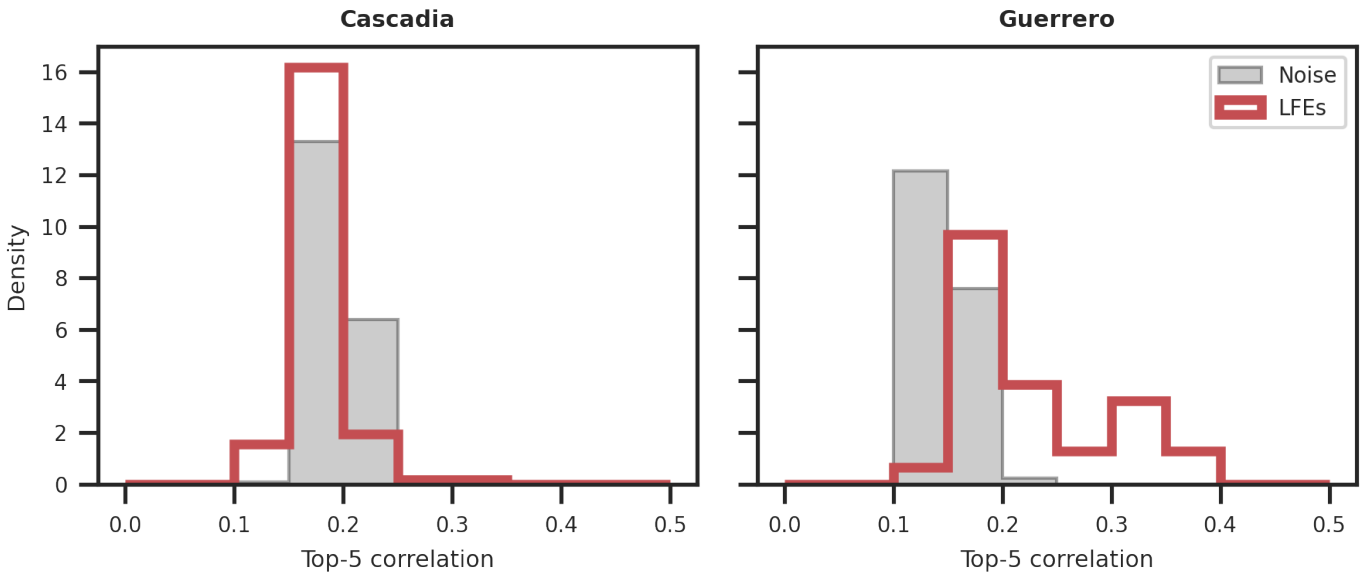

Figure S12: Cross correlation distribution between the family waveform stacks from the reference catalog and short times around the new detections from the catalog. Only detections without a temporal match to the reference catalog are shown, i.e., only uncataloged events. We correlate each detection to each template from the reference catalog, reporting the correlation with the best matching template for each detection. We correlate 6 s windows around the picked S waves on the template. We consider the mean of the correlations for the two horizontal components. Each correlation value is the mean of the five stations with highest correlation. We allow time offsets of up to 25 seconds between templates and new detections but keep the moveout fixed to the values from the reference catalog. The background shows the correlation for noise examples using the same procedure as for the LFEs. Noise waveforms were selected from days with low LFE activity. As the JMA catalog is not based on templates, this analysis is not applicable here.

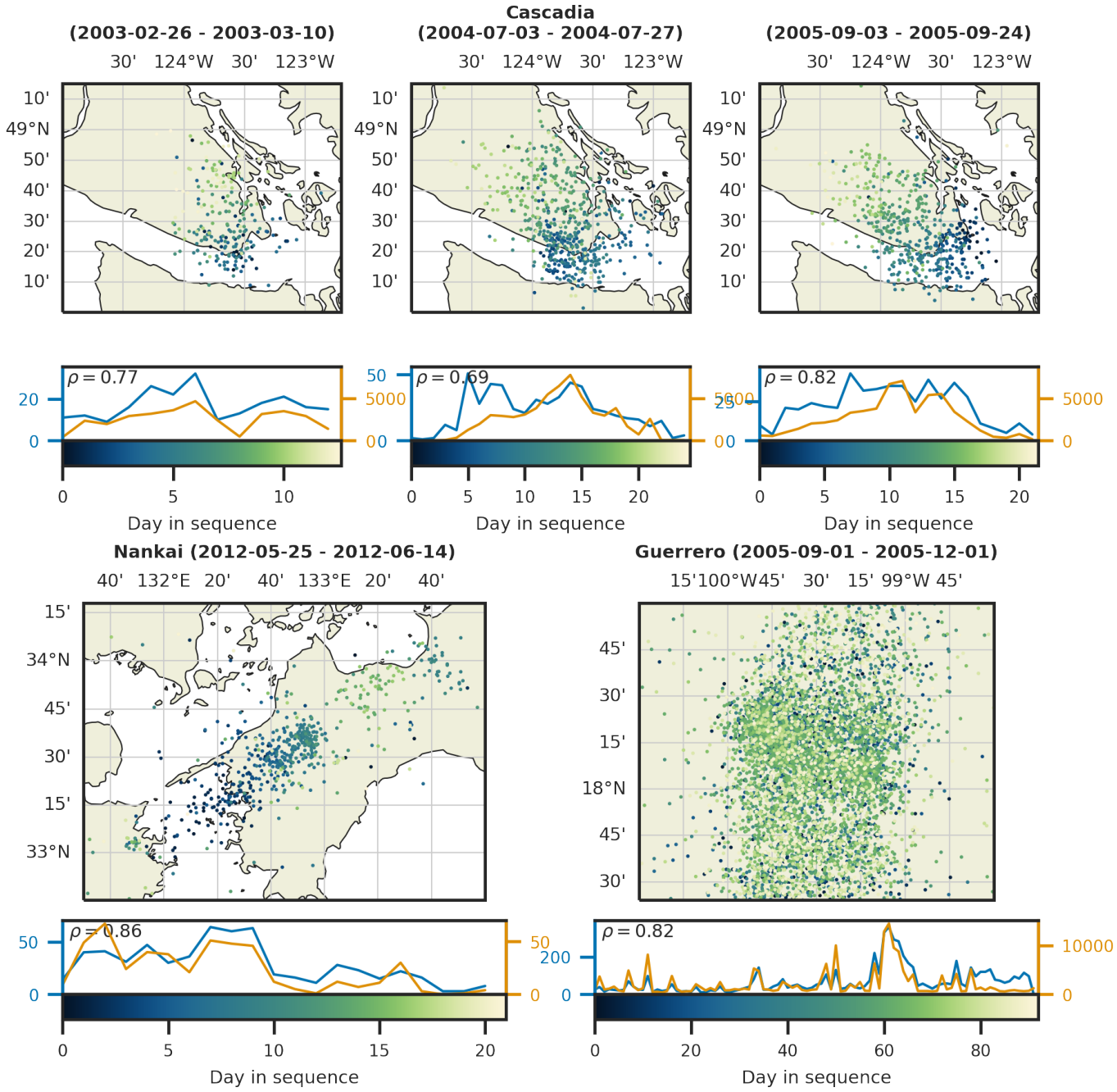

Figure S13: Spatial migration patterns of LFEs detected using a cross-domain approach. For each region we trained the model using a leave-one-out approach, i.e., on all regions except the one we later applied it to. The substantially higher number of events for Mexico is related to variabilities in the absolute confidences values and could be adjusted using fine-tuned thresholds.

References

Araki, E., Saffer, D. M., Kopf, A. J., Wallace, L. M., Kimura, T., Machida, Y., Ide, S., Davis, E., and Scientists, I. E. . S. Recurring and triggered slow-slip events near the trench at the Nankai Trough subduction megathrust. *Science*, 356(6343):1157–1160, 2017.

Bostock, M. G., Thomas, A. M., Savard, G., Chuang, L., and Rubin, A. M. Magnitudes and Moment-Duration Scaling of Low-Frequency Earthquakes beneath Southern Vancouver Island. *Journal of Geophysical Research: Solid Earth*, 120 (9):6329–6350, 2015. doi: 10.1002/2015JB012195.

Dask Development Team. *Dask: Library for dynamic task scheduling*, 2016.

Domínguez, J., Suárez, G., Comte, D., and Quintanar, L. Seismic velocity structure of the Guerrero gap, Mexico. *Geofísica internacional*, 45(2):129–139, 2006.

Frank, W. B., Shapiro, N. M., Kostoglodov, V., Husker, A. L., Campillo, M., Payero, J. S., and Prieto, G. A. Low-Frequency Earthquakes in the Mexican Sweet Spot. *Geophysical Research Letters*, 40(11):2661–2666, 2013. doi: 10.1002/grl.50561.

Frank, W. B., Shapiro, N. M., Husker, A. L., Kostoglodov, V., Romanenko, A., and Campillo, M. Using Systematically Characterized Low-Frequency Earthquakes as a Fault Probe in Guerrero, Mexico. *Journal of Geophysical Research: Solid Earth*, 119(10):7686–7700, 2014. doi: 10.1002/2014JB011457.

Japan Meteorological Agency. The Seismological Bulletin of Japan, 2023.

Larson, K. M., Kostoglodov, V., Miyazaki, S., and Santiago, J. A. S. The 2006 aseismic slow slip event in Guerrero, Mexico: New results from GPS. *Geophysical Research Letters*, 34(13), 2007.

Lomax, A., Virieux, J., Volant, P., and Berge-Thierry, C. Probabilistic earthquake location in 3D and layered models: Introduction of a Metropolis-Gibbs method and comparison with linear locations. *Advances in seismic event location*, pages 101–134, 2000.

Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., and Lauciani, V. INSTANCE–the Italian seismic dataset for machine learning. *Earth System Science Data*, 13(12):5509–5544, 2021.

Münchmeyer, J. PyOcto: A high-throughput seismic phase associator. *arXiv preprint arXiv:2310.11157*, 2023.

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., et al. Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. *Journal of Geophysical Research: Solid Earth*, 127(1):e2021JB023499, 2022.

Nakano, M., Hori, T., Araki, E., Kodaira, S., and Ide, S. Shallow very-low-frequency earthquakes accompany slow slip events in the Nankai subduction zone. Nat Commun 9: 984, 2018.

Payero, J. S., Kostoglodov, V., Shapiro, N., Mikumo, T., Iglesias, A., Pérez-Campos, X., and Clayton, R. W. Nonvolcanic tremor observed in the Mexican subduction zone. *Geophysical Research Letters*, 35(7), 2008.

Peng, Z. and Zhao, P. Migration of early aftershocks following the 2004 Parkfield earthquake. *Nature Geoscience*, 2(12): 877–881, 2009.

Rogers, G. and Dragert, H. Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science*, 300(5627):1942–1943, 2003.

Ross, Z. E., Meier, M.-A., Hauksson, E., and Heaton, T. H. Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, 108(5A):2894–2901, 2018.

Ross, Z. E., Idini, B., Jia, Z., Stephenson, O. L., Zhong, M., Wang, X., Zhan, Z., Simons, M., Fielding, E. J., Yun, S.-H., et al. Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence. *Science*, 366(6463): 346–351, 2019.

Rouet-Leduc, B., Hulbert, C., and Johnson, P. A. Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nature Geoscience*, 12(1):75–79, 2019.

Satake, K., Shimazaki, K., Tsuji, Y., and Ueda, K. Time and size of a giant earthquake in Cascadia inferred from Japanese tsunami records of January 1700. *Nature*, 379(6562):246–249, 1996.

Scotto di Uccio, F., Scala, A., Festa, G., Picozzi, M., and Beroza, G. C. Comparing and integrating artificial intelligence and similarity search detection techniques: application to seismic sequences in Southern Italy. *Geophysical Journal International*, 233(2):861–874, 2023.

Shelly, D. R. A 15 Year Catalog of More than 1 Million Low-Frequency Earthquakes: Tracking Tremor and Slip along the Deep San Andreas Fault. *Journal of Geophysical Research: Solid Earth*, 122(5):3739–3753, 2017. doi: 10.1002/2017JB014047.

Shelly, D. R., Beroza, G. C., Ide, S., and Nakamula, S. Low-Frequency Earthquakes in Shikoku, Japan, and Their Relationship to Episodic Tremor and Slip. *Nature*, 442(7099):188–191, July 2006. doi: 10.1038/nature04931.

Small, P., Gill, D., Maechling, P. J., Taborda, R., Callaghan, S., Jordan, T. H., Olsen, K. B., Ely, G. P., and Goulet, C. The SCEC unified community velocity model software framework. *Seismological Research Letters*, 88(6):1539–1552, 2017.

Wech, A. G. Cataloging Tectonic Tremor Energy Radiation in the Cascadia Subduction Zone. *Journal of Geophysical Research: Solid Earth*, 126(10):e2021JB022523, 2021. doi: 10.1029/2021JB022523.

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., et al. SeisBench—A toolbox for machine learning in seismology. *Seismological Research Letters*, 93 (3):1695–1709, 2022.

Zhang, M., Ellsworth, W. L., and Beroza, G. C. Rapid earthquake association and location. *Seismological Research Letters*, 90(6):2276–2284, 2019.

Zhu, W. and Beroza, G. C. PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273, 2019.

Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L., and Beroza, G. C. Earthquake phase association using a Bayesian Gaussian mixture model. *Journal of Geophysical Research: Solid Earth*, 127(5):e2021JB023249, 2022.