



HAL
open science

Tremor Waveform Extraction and Automatic Location With Neural Network Interpretation

Claudia Hulbert, Romain Jolivet, Blandine Gardonio, Paul A Johnson,
Christopher X Ren, Bertrand Rouet-Leduc

► **To cite this version:**

Claudia Hulbert, Romain Jolivet, Blandine Gardonio, Paul A Johnson, Christopher X Ren, et al.. Tremor Waveform Extraction and Automatic Location With Neural Network Interpretation. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60, pp.1 - 9. 10.1109/tgrs.2022.3156125 . hal-03872000

HAL Id: hal-03872000

<https://hal.science/hal-03872000v1>

Submitted on 25 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tremor Waveform Extraction and Automatic Location With Neural Network Interpretation

Claudia Hulbert¹, Romain Jolivet², Blandine Gardonio, Paul A. Johnson³,
Christopher X. Ren, and Bertrand Rouet-Leduc

Abstract—Active faults release tectonic stress imposed by plate motion through a spectrum of slip modes, from slow, aseismic slip, to dynamic, seismic events. Slow earthquakes are often associated with tectonic tremor, nonimpulsive signals that can easily be buried in seismic noise and go undetected. We present a new methodology aimed at improving the detection and location of tremors hidden within seismic noise. After identifying tremors with a classic convolutional neural network (CNN), we rely on neural network attribution to extract core tremor signatures. We observe that the signals resulting from the neural network attribution analysis correspond to a waveform traveling in the Earth’s crust and mantle at wavespeeds consistent with seismological estimates. We then use these waveforms signatures to locate the source of tremors with standard array-based techniques. We apply this method to the Cascadia subduction zone, where we identify tremor patches consistent with existing catalogs. This approach allows us to extract small signals hidden within the noise, and to locate more tremors than in existing catalogs.

Index Terms—Neural network attribution, tremor location, waveform extraction.

I. INTRODUCTION

GEODETIC and seismological observations suggest fault slip can take place over a large range of time scales, ranging from seconds or minutes during earthquakes to days, weeks, months or permanently as slow slip [1]–[4]. Among this spectrum of slip, slow displacements generated by slow slip and creep are often accompanied by characteristic seismic signals termed tremor [5]–[8]. The precise relationship between tremor and slip remains poorly understood; proposed explanations of tremor origins include the breakage of small asperities as the fault displaces, fluid-driven fractures or fluid migration, among others [9]–[11]. In general, tremor has been associated with slow slip, as the source mechanism is consistent with either a double couple mechanism [12] and/or with marked geodetic displacements [5].

Since its discovery in Japan in the early 2000s [6], tectonic tremor has been observed in many subduction zones, including Cascadia [5], [13], Alaska [14], Mexico [13], [15], [16], Southern Chile [13], [17], New Zealand [13], [18], and Costa Rica [19]. Observation of tremor activity extends to strike-slip fault systems [8], [20], [21]. Like earthquakes, tremor can occur spontaneously but can also be triggered, either by tidal loading [22], [23] or by quasi-static or dynamic loading from earthquakes [20], [24].

In contrast to earthquakes that are characterized by impulsive P- and S-wave arrivals, tremor waveforms are not associated with systematic, well-defined patterns. This absence of characteristic traits makes the detection and location of tremors challenging. While identifying an earthquake and picking its phases on a single seismic waveform can be a relatively easy task for an analyst, detecting tremor on a single station is extremely difficult. For this reason, tremor identification typically relies on a large network of seismometers: detecting a coherent signal across the network allows one to discriminate it from noise. Methods to locate tremors are usually based on waveform correlation across an array of seismic stations, through the cross correlation of tremor envelopes from which a differential travel time between stations can be estimated. These differential travel times can then be compared to theoretical ones through grid-search [6], [8]. In some other cases, the entire correlation functions are stacked and its maximum value is used to estimate the associated event location [25], [26].

Manuscript received June 16, 2021; revised October 25, 2021 and January 13, 2022; accepted February 14, 2022. Date of publication March 10, 2022; date of current version March 29, 2022. The work of Claudia Hulbert was supported in part by the Joint Research Laboratory Effort in the framework of the CEA-ENS Yves Rocard LRC, France; and in part by the European Research Council (ERC) through the European Union’s Horizon 2020 Research and Innovation Program (Geo-4D Project) under Grant 758210. The work of Romain Jolivet was supported by the ERC through the European Union’s Horizon 2020 Research and Innovation Program (Geo-4D Project) under Grant 758210. The work of Blandine Gardonio was supported by the Joint Research Laboratory Effort in the framework of the CEA-ENS Yves Rocard LRC. The work of Paul A. Johnson was supported in part by the Institutional Support [Laboratory Directed Research and Development (LDRD)], Los Alamos; and in part by the DOE Office of Science (Geoscience Program) under Grant 89233218CNA000001. The work of Christopher X. Ren was supported by LDRD and CSES, Los Alamos. The work of Bertrand Rouet-Leduc was supported by the LDRD under Grant LDRD 20200278ER. (Corresponding author: Claudia Hulbert.)

Claudia Hulbert is with the Laboratoire de Géologie, Département de Géoscience, École normale supérieure, PSL University, CNRS UMR 8538, 75231 Paris, France (e-mail: claudia.hulbert@ens.fr).

Romain Jolivet is with the Laboratoire de Géologie, Département de Géoscience, École normale supérieure, PSL University, CNRS UMR 8538, 75231 Paris, France, and also with the Institut Universitaire de France, 75005 Paris, France.

Blandine Gardonio is with UCBL, CNRS, LGL-TPE, Université Lyon, 69622 Villeurbanne, France.

Paul A. Johnson is with the Geophysics Group, Los Alamos National Laboratory, Los Alamos, NM 87545 USA.

Christopher X. Ren is with the Intelligence and Space Research Division, Los Alamos National Laboratory, Los Alamos, NM 87545 USA.

Bertrand Rouet-Leduc is with the Geophysics Group, Los Alamos National Laboratory, Los Alamos, NM 87545 USA, and also with the Disaster Prevention Research Institute, Kyoto University, Kyoto 611-001, Japan.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TGRS.2022.3156125>, provided by the authors.

Digital Object Identifier 10.1109/TGRS.2022.3156125

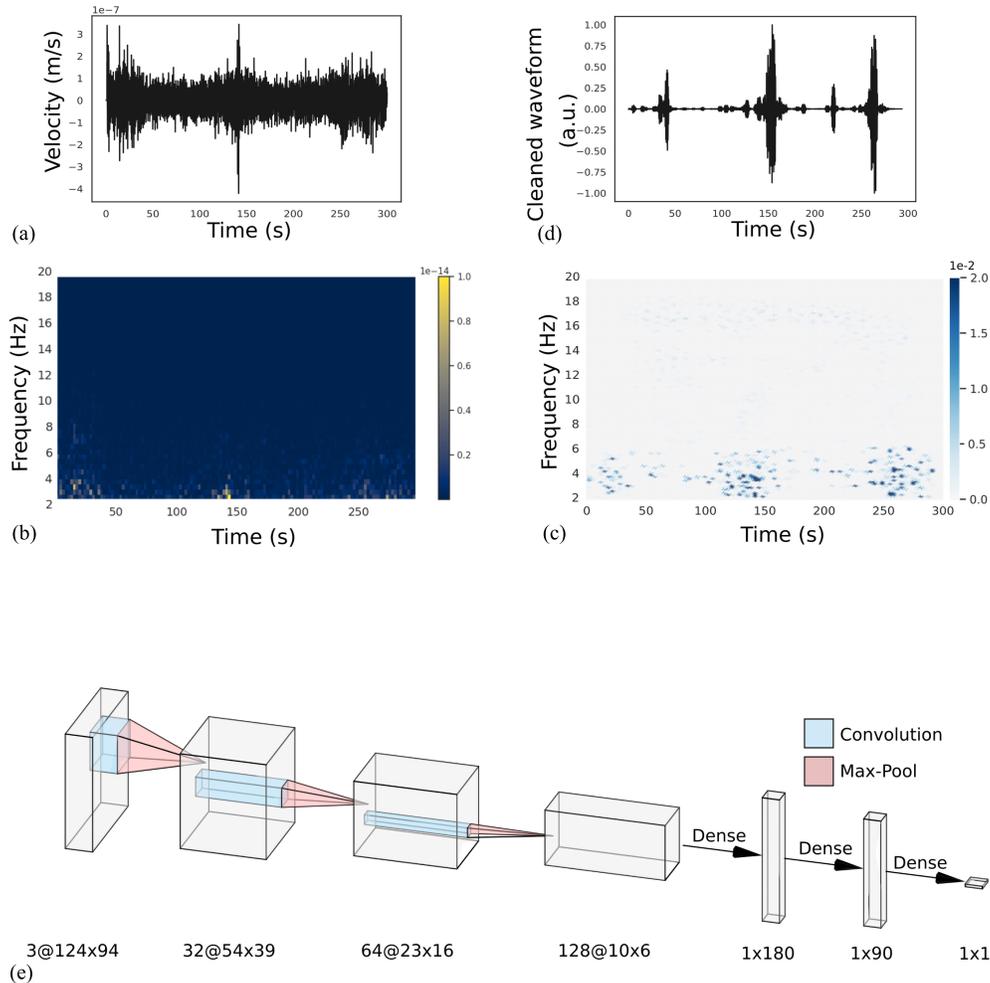


Fig. 1. **Extraction of tremor waveforms with neural network attribution.** Raw waveforms (a) are turned into spectrograms (b) and fed to a classic CNN. We rely on a standard network (three convolution+max pool and three dense layers). From these spectrograms, the CNN is tasked with classifying tremor from nontremor. Once the model is trained (with tremors identified from the PNSN catalog), we rely on Integrated Gradients [41] attribution to interpret its results. (c) Result of the positive attribution for the same waveform; the blue areas correspond to the parts of the spectrogram that carry core tremor information according to the attribution analysis. We can then select exclusively these areas to inverse-Fourier transform, to get a clean tremor signal. (d) These cleaned waveforms correspond to the core tremor signals, as seen by our model, and have clear structure compared to the original waveforms; our goal here is to rely on these clean waveforms to locate tremors. (e) Architecture of the network.

In addition to envelope-based location, approaches described in the literature include template matching [12] and the backpropagation of seismic signals [27]. Refinements have also been introduced to improve the precision of tremor location, especially regarding depth, which is often poorly resolved. These approaches include double-difference location techniques [21], or the estimation of phase lags between P- and S-wave timings [28].

Array-based tremor detections and locations are unlikely to work well for low-amplitude tremors, that may appear only at one or two stations and not be identified as coherent signal throughout the array. This type of approach is also particularly challenging in the case of sparse seismic networks because local sources of noise prohibit high correlations between station pairs. In what follows, we propose a new methodology for tremor location, based upon neural network attribution to extract tremor waveforms. We find that our approach gives results coherent with existing catalogs, and allows the location of tremors below the noise level. This method should prove a

valuable tool for the location of small tremors, or in presence of sparse seismic arrays.

II. METHOD

A. Extracting Tremor Waveforms With Neural Network Attribution

A trained neural network can easily detect tremor at a single seismic station, vastly increasing the number of detections when compared to the catalog it has been trained on [29], [30]. Our goal here is to leverage the learned representations of a deep neural network to extract seismic waveforms, revealing the signals that triggered the detection and locating the corresponding tremors.

The network we rely upon is a standard CNN [see Fig. 1(b)]. Our CNN attempts to classify tremor from seismic noise. It is trained on three components, 5-min spectrograms of approximately 165 000 examples of tremors identified from the Pacific Northwest Seismic Network (PNSN) catalog [25], and 250 000 examples of seismic noise. The noise examples are drawn

randomly from times when no tremor has been detected by the PNSN. Examples of noise and tremor are drawn from several seismic stations in Cascadia, located along the Western U.S. coast (nearly all the stations are outside Vancouver Island, our region of interest). The CNN attempts to classify tremor from single, three-component examples of spectrograms, thereby building a single-station detector. We train the classifier on examples from August 2009 to June 2016. Examples for June 2016 to December 2018 are used for validation and test, with the first half corresponding to the validation set, and the second half to the test set. The performance of our classifier in testing, as measured by the receiver operator characteristic-area under curve (ROC-AUC) score, is of 0.913. We leverage adversarial training to further improve the generalization of our model and the interpretability of its gradients [31]. Details regarding the database used for training, the training and testing procedures, and the performance of the model can be found in the Supplementary Material.

This kind of single station tremor classification tends to generalize well to other areas [29]: a model trained on waveform spectrograms from a single seismic station in Vancouver Island, Canada, generalizes to other nearby seismic stations, and even to data from Japan—suggesting that tremor waveforms carry some kind of universal signature.

To probe the classifier for tremor signature patterns, we rely on neural network attribution tools. Neural networks are often considered as black-box models; however, there has been considerable effort over the recent years to develop their interpretability. Among these approaches, several attempts at interpretation focused on visualization throughout the hidden layers of the network: 1) either by considering all neurons in a given layer as a whole [32], or 2) by considering each neuron individually, often based on the activation value of each unit [33]–[35]. A parallel area of research is to analyze network attribution, i.e., the connection between input features and the prediction of the network [36]–[42].

We rely on a recent approach for network attribution, using the integrated gradients methodology [41]. Gradients are often used as a basis of attribution techniques, as they can be considered as the coefficients associated with a network’s features, and their analysis can therefore inform on important feature contributions. Consider a deep network $N: \mathbb{R}^n \rightarrow [0, 1]$, and an input datapoint $\mathbf{x} = (x_1, \dots, x_n)$. The attribution problem consists in analyzing the contribution of each individual x_i to the prediction of the network, $N(\mathbf{x})$. For example in image analysis, it would correspond to analyzing the contribution of individual pixels to the prediction of the model.

The integrated gradients correspond to a path integral along a straight line connecting a baseline vector to the input vector \mathbf{x}

$$\text{IG}_i(\mathbf{x}) = (x_i - x'_i) \int_{\alpha=0}^1 \frac{\partial N(\mathbf{x}' + \alpha(\mathbf{x} - \mathbf{x}'))}{\partial x_i} d\alpha$$

where \mathbf{x}' corresponds to the baseline vector, and $(\partial N(\mathbf{x})/\partial x_i)$ correspond to the gradient of the network along the i th dimension. The baseline vector \mathbf{x}' can be taken for example as a black image or a zero embedding vector. Because gradients are integrated in this way, the method is robust to plateaus in

the network’s response and areas of local gradient instability such as saddle points.

The integrated gradient approach solves two main issues encountered by several other attribution techniques proposed in the literature: 1) it avoids flat, zero gradient regions (plateaus in the output of the network with respect to a region in the input space) that can cause the attribution method to focus on nonrelevant features and 2) functionally equivalent networks (that associate the same inputs to the same outputs) return the same attribution, avoiding the attribution to focus on unimportant characteristics of the classifier. Furthermore, this approach is fast to compute.

We analyze our input spectrograms with the integrated gradient approach, identifying which parts of the spectrograms carry the core of the tremor information. Positive attribution results for a sample waveform is shown in Fig. 1(a), third panel. In this plot, the blue areas correspond to the parts of the spectrogram that were crucial for the algorithm to label a waveform as tremor. As expected, the dominant frequency band lies between 2 and 7 Hz, with some smaller contribution from higher frequencies. Note that this analysis is very different from a simple bandpass filter within tremor frequency bands (see Fig. 3), because of the temporal distribution of seismic energy in the attribution spectrogram. Indeed, the waveform extraction analysis emphasizes times where tremor signals are likely to occur, and is therefore akin to a time-dependent filter; in contrast, a bandpass filter treats all times in an equal manner. We then multiply the original amplitude values by the attribution spectrogram, and perform the associated inverse Fourier transform. This analysis yields waveforms with clear structure compared to the original signals [see Fig. 1(a)] (right). It is likely that the attribution analysis identifies some of the individual low-frequency earthquakes (LFEs) contained within the tremor signals.

An enlarged figure showing the original and extracted waveforms can be found in the Supplementary Material. Note that our algorithm identifies core tremor patterns in the raw spectrograms, that are backprojected in the time domain. These core patterns do not necessarily correspond to the entire tremor signals, and therefore the extracted waveforms may not capture the entire energy of the underlying tremor. This explains why we choose to normalize the amplitudes in the extracted waveforms, such as shown in Fig. 1 (right). However, as long as the same patterns are identified independently at multiple stations, we show in the following that these can be used to locate the origin of these signals.

B. Reconstructed Waveforms Are the Signature of Traveling Waves

Because seismic energy in the attribution spectrograms is intermittent in time, extracted waveforms have a much more impulsive structure than the original signals. As illustrated by the example in Fig. 2, raw waveforms (even filtered between 1 and 8 Hz) often exhibit weak and emergent signals, which appear much more impulsive once extracted as described above. Importantly, the reconstructed waveforms show impulsive signals that are coherent between stations.

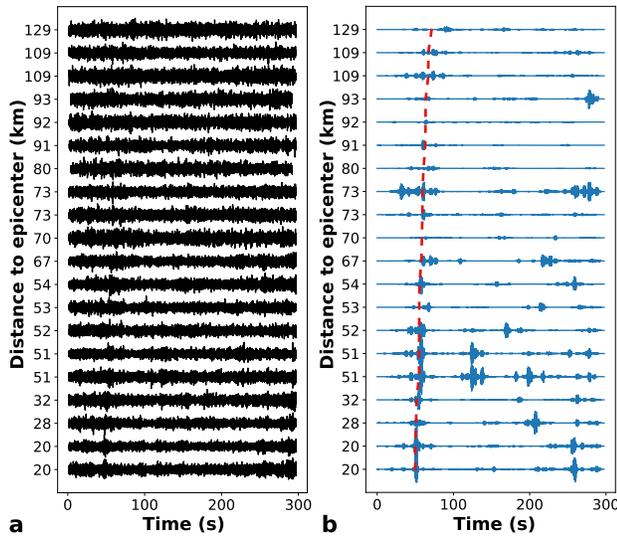


Fig. 2. Application to an array of seismometers. (a) Raw waveforms bandpassed between 1 and 8 Hz. (b) Result of the neural network attribution analysis to denoise the waveforms (in blue). The extraction procedure is performed on each station independently. Clean move-out patterns can be seen across the network, which can be compared to shear wave theoretical travel times (red line). The move-outs match theoretical wave speed to first order. Note that in order to visualize the moveout, waveforms are ordered according to the epicenter of one particular tremor (occurring at 50 s).

An enlarged figure showing this moveout can be found in the Supplementary Material.

We can compare the travel times of the reconstructed signals to the arrival time predictions from a simple 1-D velocity model for S-waves. When performing the attribution analysis independently on each station, the observed travel times match the theoretical ones to first order, suggesting envelope cross correlation location should be possible. The quality of the reconstructed waveforms decays with distance to the source, which is expected and consistent with seismic source radiation patterns. This spatio-temporal propagation provides a strong validation that the reconstructed waveforms indeed correspond to signals of seismological interest (here tectonic tremors), extracted from noisy time series. Because the attribution analysis does not return a waveform corresponding to the full tremor energy, the identified feature representations may vary at different stations. In such cases, the extracted waveforms would not reach the correlation threshold, and the correlation functions would not coalesce into a single point in space, leading us to miss some tremor occurrences. While this is unlikely for nearby stations, it could occur for networks spread out over large distances.

The conservation of travel times in the reconstructed waveforms is also supported by the fact that binary classification results from a CNN are not station dependent. Indeed, applying the same model to other stations, even in different tectonic regions, yields good performance [29]. Therefore, we expect that tremor signatures are shared across stations, and, if detected, these should consequentially align across a seismic network according to true move-out patterns.

C. Location Procedure

If indeed our reconstructed waveforms preserve the tremors' move-out patterns, it should be possible to use them to locate the source of the tremor. Once all waveforms of interest are reconstructed as described above, we rely on traditional array-based methods to locate tremors. We begin by running a classic short term average/long term average (STA/LTA) picker through the cleaned waveforms. The fact that these waveforms are more impulsive and structured than the original signals makes the picking procedure more effective, resulting in many picks identified. These waveforms are then sliced around each pick, using a window size based upon the maximum theoretical travel time that can be observed across the array.

We then measure the correlation of the envelopes of the sliced waveforms, for each pair of stations within the array. The envelopes are computed as the smoothed rms of the denoised waveforms, in a manner similar to [43]. Each envelope correlation value is associated with a time lag between the two station waveforms, which can be compared with shear wave theoretical time differences between station pairs. A regional 1-D velocity model is used to estimate travel times [44]. For each station, both horizontal components are used for the analysis. We keep the entire correlation functions to locate the events [25], [26]; these correlation functions are stacked for all pairs of stations that display a high envelope correlation (see below). The point in the spatial grid corresponding to the maximum stacked correlation is then returned as the source location for one particular tremor.

Fig. 3 shows a stacked correlation function over our region of interest. The maximum of the stacked correlation corresponds to the tremor's location. Because we rely on envelope cross correlation, depth is not well constrained. Note that if the reconstructed waveforms were not consistent with the speed of S-wave in Earth, the stacked correlation would not coalesce to a maximum, as it does for example in Fig. 3.

Besides enabling the location of signals of interest, the cross correlation of the extracted waveforms is also very useful to filter out false positives. In cases where noise signals are wrongly classified as tremor and return positive attribution values, they will only be selected by the algorithm if their propagation throughout the network matches the speed of S-waves. Signals that fool the model but do not travel at the speed of local velocity estimates will be discarded.

We use several selection criteria to improve confidence in a tremor's location as follows.

- 1) We only locate the source of waveforms that the neural network model classifies as tremor above a fixed "confidence" threshold (softmax ≥ 0.7 , on a minimum of two stations). For comparison, decision thresholds used for tremor or LFE detection tend to be included within the [0.5, 0.75] interval [29], [45].
- 2) To focus the analysis on informative stations, a station pair is only considered if its maximum correlation value reaches a given threshold. In line with the literature, we set a minimum correlation threshold of 0.7 [25], [43]. At least five station pairs must reach this minimum threshold. The stacked correlation values are weighted

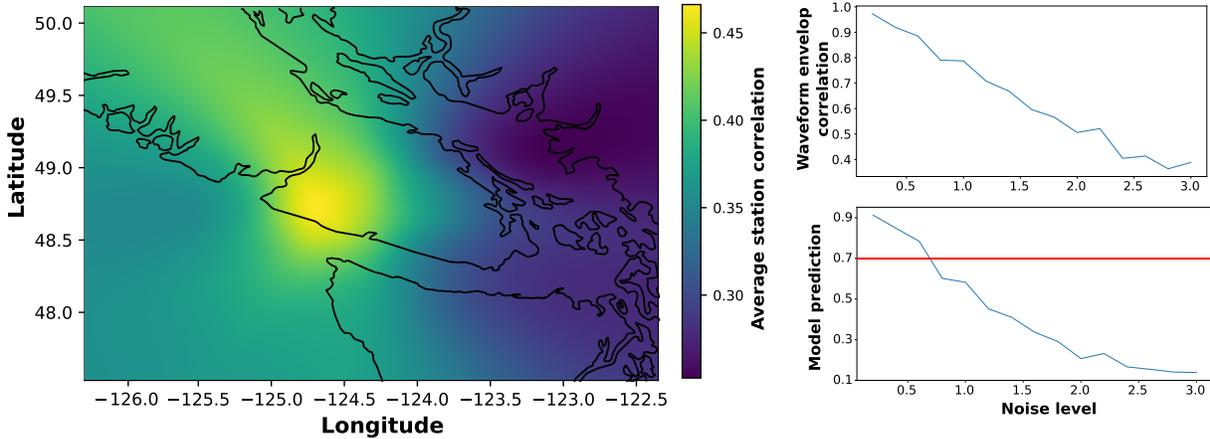


Fig. 3. **Tremor location and robustness.** (Left) Example tremor location. We compute the envelope of the cleaned signals; these envelopes are then sliced around each identified STA/LTA pick, and cross-correlated for all station pairs. This figure shows the stacked correlation plotted over the lat/lon grid for one tremor. If a number of criteria are met (see below), the maximum of the stacked correlation function is returned as the tremor’s location. (Right) Impact of noise on located signals. We overlay input tremor waveforms with seismic noise at different coefficients. We then perform the attribution analysis of the original and corrupted signals, and compare both extracted waveforms by computing their (top right) envelope correlation. We also compute the associated (bottom right) model prediction. In both plots, the blue line corresponds to the average value computed from approximately 5000 examples. When strong levels of noise are added, both the model’s prediction and the envelope correlation decay. Because we only try to locate signals above a “confidence” threshold of 0.7 (red line in bottom), our algorithm would not attempt to locate signals for which the added noise has strongly impacted the attribution analysis.

according to the maximum correlation measured for each station pair - as station pairs with higher waveform correlations are likely to be closer to the event source, or to exhibit a higher signal-to-noise ratio.

- 3) To ensure that the stacked correlation maximum is well defined, we compute the ratio between the maximum stacked correlation value and the median absolute deviation (MAD) of the correlation function. For tremor identification, in line with [12] we only keep events with an MAD ratio above 8 and discard the others.
- 4) To limit the locations’ spatial uncertainty, we use an approach similar to [25]. We randomly remove a subset of 10% of the stations pairs by bootstrapping, over a number of iterations. We discard signals for which the bootstrapped locations are not consistent spatially (outside a 5-km radius around the initial location).

Tremor signals are emergent, can overlap and/or last for a time potentially longer than our selected window duration. Therefore, there is always a risk that our algorithm could either identify a single tremor as two different ones, or miss an occurrence. The location algorithm considers 30-s long waveforms, and we rely on the two following criteria to minimize these potential issues: 1) we do not allow our algorithm to locate two tremors within a short time window (3 s, which corresponds to a minimum time separation) and 2) after a tremor is located, we wait for the envelope correlation function to drop below the afore-mentioned threshold before trying to locate another signal (criterion in practice for separating potential signals of interest). Once the correlation function has dropped and if it rises again, we consider that we are likely observing a new tremor. We tested the robustness of our approach by overlaying tremor waveforms with noise [see Fig. 3 (right) and the Supplementary Material]. We find that adding noise to tremor waveforms affects both the performance of the model and the attribution results. Strong noise amplitudes

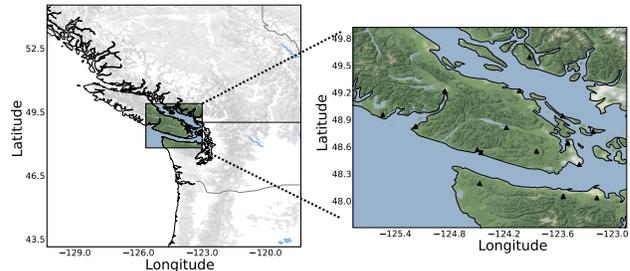


Fig. 4. **Area analyzed and seismic array used.** (Left) Map of the Southern Vancouver Island, Canada, analyzed in this study. (Right) Seismic array. The black triangles correspond to seismic stations used to locate tremors.

can generate artifacts in the extracted waveforms. However, because the associated signals are not classified as tremor by the model, our algorithm would not try to locate them, which suggests that the method is robust to strong noise amplitudes.

III. RESULTS

A. Area Analyzed

To test the performance of our location methodology, we apply it to the 2018 Cascadia slow slip. We focus on Southern Victoria Island, Canada, where clear tremor activity is typically recorded [5], [25]. In this region, the Cascadia subduction zone experiences an ETS approximately every 14 months, which typically lasts for one or two months. According to the PNSN tremor logs, the 2018 event started in early May, and paused for a while before resuming in mid-June and continuing until mid-July 2018. We, therefore, analyze three months of data (May, June and July 2018). Applying our method to a well-studied area allows us to validate it by comparing our results to existing tremor catalogs.

We rely on an array of 21 seismometers for locating the tremors, from the Canadian National Seismograph Network (CNSN) [46] and plate boundary observatory (PBO) [47]

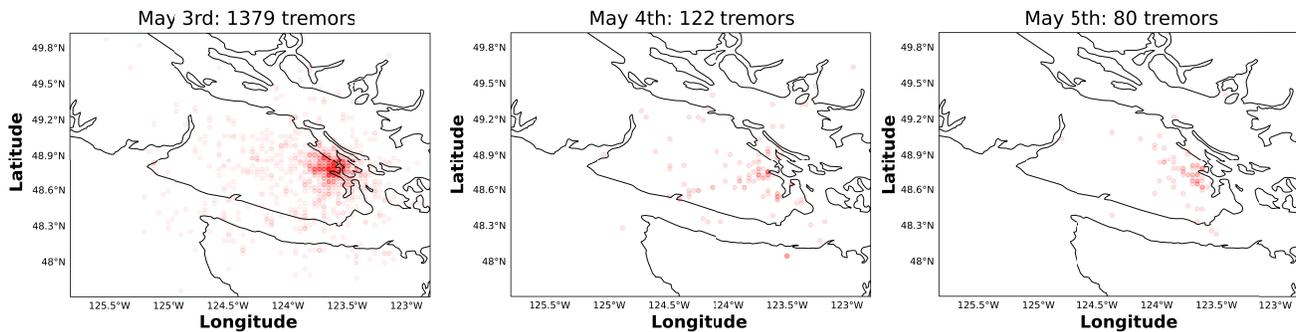


Fig. 5. **Example location.** Tremor locations (in red) for three consecutive days with consistent, local tremor patches, during the 2018 SSE.

seismic networks. About a third of the stations are borehole instruments. Fig. 4 shows a map of the area analyzed, as well as the seismic array used. We discretize the area with a regular ($100 \times 100 \times 25$) (lat, lon, dep) grid, with grid points of ($2.87 \times 2.96 \times 2.4$) km.

B. Application to the 2018 Cascadia ETS

In Fig. 5, we illustrate our results by showing tremor locations for three consecutive days during the 2018 slow slip event (SSE). For comparison, the PNSN reported 223 tremors in this patch on May 3rd, 0 on May 4th, and 7 on May 5th. We chose to display these three days in particular because they contain consistent tremor patches, located around the same area. The fact that we identify consistent tremor activity in the same area over the three days suggests that our additional detections are likely to be accurate. Furthermore, local tremor patches as shown here are a better test to validate our methodology, as they are less forgiving regarding the location of false positives, or mislocations of real events. Most of the outliers to the main patches are relatively close to each other, which suggests that they may correspond to smaller, weaker tremor patches. Individual tremors that are far from each other may correspond to mislocated signals or false positives.

Overall, we find our locations to be consistent with tremor locations in existing catalogs [25], even for days characterized by small, local patches such as those displayed above. Figs. S1–S3 in the Supplementary Material show the distribution of tremor locations for all days analyzed, with comparisons to locations from the PNSN catalog. Note that running the exact same analysis with simple bandpassed waveforms (instead of waveforms extracted with integrated gradients) leads to only a small number of tremors to be cataloged (1500 versus 66 000). However, this striking difference is not simply the consequence of using an attribution scheme to clean the waveforms, and is likely to be mostly due to our reliance on STA/LTA analysis to identify a potential signal of interest. Running an STA/LTA analysis on bandpassed waveforms will only return a small number of picks, due to the fact that tremor signals are very emergent. It is, therefore, better to compare our results with other existing catalogs, that do not rely on picks to identify signals of interest.

Fig. 6 summarizes the comparison between our results and the tremor locations from the PNSN catalog for the 2018 SSE. In particular, Fig. 6(a) shows the number of tremors detected

by both approaches, for all days during the analysis. Variations in the number of detections are consistent, with a higher number of tremors detected between the 18th of June and the 11th of July. Overall our algorithm identifies six times more tremors than the PNSN catalog (67 000 versus 11 000 tremors over the period and the region analyzed). Part of this increase in detection numbers can be attributed to methodological differences between the approaches: the PNSN classifies 30 s of waveforms as tremor, whereas our algorithm looks at individual picks in the waveforms, and can therefore go below 30 s. However, a portion of the additional detections are likely to arise from the ability of our methodology to capture small signals below the noise level. This hypothesis is supported by the fact that for many days, such as the 4th of May shown above, our methodology finds a relatively high number of tremors (more than a hundred), while the PNSN reports no tremor activity. Note that both algorithms use the same correlation threshold as criterion for considering station pairs (0.7), and therefore the higher number of located tremors in our analysis is not related to the choice of a smaller threshold. As these particular tremors did not generate a coherent (standard) envelope cross correlation throughout the array, they are likely to be contaminated by noise, and therefore to correspond to low-amplitude signals below or close to the noise level.

Fig. 6(b) shows how close the identified tremor patches in both catalogs are, on a daily basis. We compute the density center of daily tremor distribution, for all days with tremor activity underneath Vancouver Island according to the PNSN catalog (when ten or more tremors are detected). We then compute how far our daily tremor density centers are compared to the tremor density centers reported in the PNSN catalog. For a few days characterized by several tremor patches, we rely on spectral clustering to separate the patches and compare individual patches (see the Supplementary Material).

For most days, distances between tremor patches identified in both catalogs are small, on the order of a few km. The median distance over the period analyzed is 4.42 km (less than one grid point offset), showing that indeed daily tremors are identified at very similar locations. However, for two particular days (9th of May and 21st of June), the distances are large, on the order of several dozens of km. These two days correspond to times when identified tremor patches in the PNSN catalog are small and located in the North-West and the South-East, at the edge or outside of our seismic array.

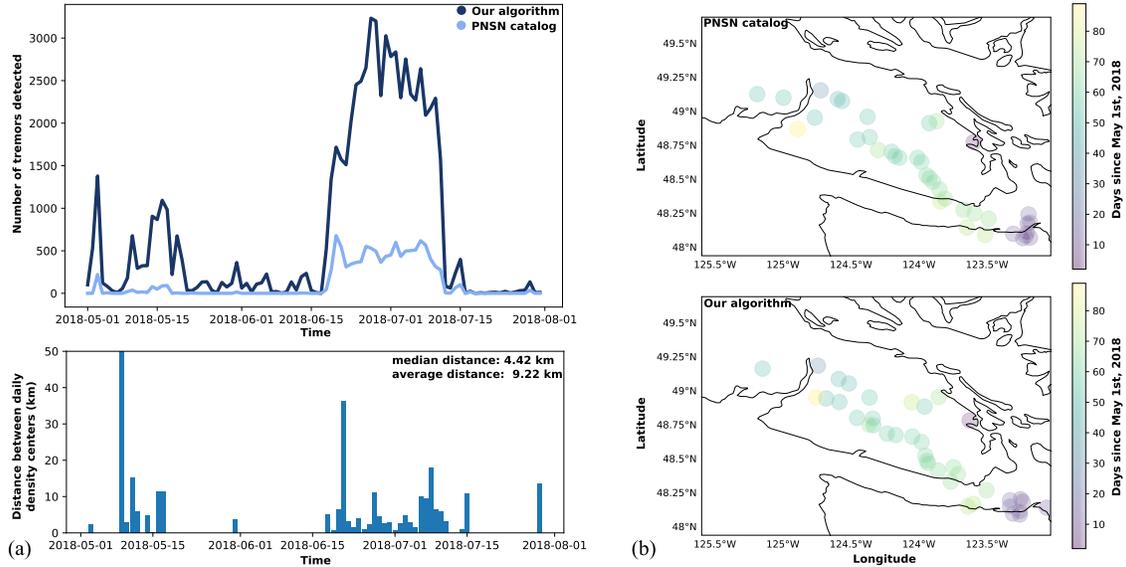


Fig. 6. **Comparison with the PNSN catalog.** (a) (Top left) Number of tremors detected by our algorithm (dark blue), compared to the number of detected tremors in the PNSN catalog (light blue). The timing of detections is consistent between both catalogs, with a high number of tremors detected between June 18th and July 11th. (Bottom left) Distance between daily tremor patches detected by our algorithm and by the PNSN catalog. For most days, the patches have similar locations (with a median distance of 4.42 km). The two days where the distance is high correspond to patches in the PNSN catalog that are on the border or outside of our seismic array. (b) Evolution of the center of daily tremor patches over the period analyzed, in (top right) PNSN catalog and (bottom right) in our analysis. Tremor migration patterns are similar with both algorithms.

This makes location difficult for our algorithm, as fewer stations are likely to see the tremors and reach a high correlation value. To overcome this issue in the future, we will work on locating events with overlapping subarrays, which should alleviate the issue associated with the borders of the network. On one day (July 17th), our algorithm also misses a tremor patch identified by the PNSN. As there are not enough datapoints in our catalog to compute a density center for July 17th, compared locations for this day can be found in the Supplementary Material.

IV. CONCLUSION

In contrast to earthquakes, tremor waveforms are emergent which makes arrivals extremely difficult to identify and pick. Therefore, traditional location techniques tend to break down, and coherent envelope signals observed across an array are often used instead for location. This makes tremor location particularly vulnerable to local sources of noise, that can impact drastically the correlation values observed between station pairs. Furthermore, for the same reason, sparse arrays are likely to be able to locate only tremors of high amplitude, that would remain correlated across the seismic network.

We developed a new tremor location approach, that is more robust to seismic noise. We show that waveform reconstruction through neural network attribution is a powerful tool for the detection and location of small tremors. We find that the reconstructed waveforms are the signatures of traveling waves, consistent with S-waves theoretical travel times. Our methodology is validated by the analysis of the 2018 slow slip events in Vancouver Island.

This methodology will be particularly helpful in areas with sparse seismic arrays, for which local noise sources are likely

to mask signals of interest and prevent the successful cross correlation of tremor envelopes. Finally, our approach is fully automatic and can be deployed as a systematic tool for tremor detection and location with minimal human intervention.

ACKNOWLEDGMENT

We thank C. Duverger and I. McBrearty for useful discussions. The seismic data used were obtained from the Canadian National Seismic Network [46] and the Plate Boundary Observatory [47]. All the data are publicly available.

CH developed the automatic location approach. RJ supervised the project and helped interpreting the results. CH and BRL developed the neural network. BG helped analyzing the waveforms. All authors contributed to writing the manuscript.

REFERENCES

- [1] Z. Peng and J. Gomberg, "An integrated perspective of the continuum between earthquakes and slow-slip phenomena," *Nature Geosci.*, vol. 3, no. 9, p. 599, 2010.
- [2] K. Obara and A. Kato, "Connecting slow earthquakes to huge earthquakes," *Science*, vol. 353, no. 6296, pp. 253–257, Jul. 2016. [Online]. Available: <http://science.sciencemag.org/content/353/6296/253>
- [3] R. Bürgmann, "The geophysics, geology and mechanics of slow fault slip," *Earth Planet. Sci. Lett.*, vol. 495, pp. 112–134, Aug. 2018.
- [4] R. Jolivet and W. B. Frank, "The transient and intermittent nature of slow slip," *AGU Adv.*, vol. 1, no. 1, Mar. 2020, Art. no. e2019AV000126.
- [5] G. Rogers and H. Dragert, "Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip," *Science*, vol. 300, no. 5627, pp. 1942–1943, Jun. 2003.
- [6] K. Obara, "Nonvolcanic deep tremor associated with subduction in southwest Japan," *Science*, vol. 296, no. 5573, pp. 1679–1681, May 2002.
- [7] W. B. Frank, "Slow slip hidden in the noise: The intermittence of tectonic release," *Geophys. Res. Lett.*, vol. 43, no. 19, p. 10, Oct. 2016.
- [8] R. M. Nadeau and D. Dolenc, "Nonvolcanic tremors deep beneath the San Andreas Fault," *Science*, vol. 307, no. 5708, p. 389, 2005.

- [9] T. Seno and T. Yamasaki, "Low-frequency tremors, intraslab and interplate earthquakes in southwest Japan—From a viewpoint of slab dehydration," *Geophys. Res. Lett.*, vol. 30, no. 22, p. 2171, Nov. 2003.
- [10] S. Ide, G. C. Beroza, D. R. Shelly, and T. Uchide, "A scaling law for slow earthquakes," *Nature*, vol. 447, no. 7140, p. 76, 2007.
- [11] V. M. Cruz-Atienza, C. Villafuerte, and H. S. Bhat, "Rapid tremor migration and pore-pressure waves in subduction zones," *Nature Commun.*, vol. 9, no. 1, pp. 1–13, Dec. 2018.
- [12] D. R. Shelly, G. C. Beroza, and S. Ide, "Non-volcanic tremor and low-frequency earthquake swarms," *Nature*, vol. 446, no. 7133, p. 305, 2007.
- [13] S. Ide, "Variety and spatial heterogeneity of tectonic tremor worldwide," *J. Geophys. Res., Solid Earth*, vol. 117, no. B3, Mar. 2012.
- [14] C. L. Peterson and D. H. Christensen, "Possible relationship between nonvolcanic tremor and the 1998–2001 slow slip event, south central Alaska," *J. Geophys. Res.*, vol. 114, no. B6, 2009, Art. no. B06302.
- [15] J. S. Payero *et al.*, "Nonvolcanic tremor observed in the Mexican subduction zone," *Geophys. Res. Lett.*, vol. 35, no. 7, Apr. 2008, Art. no. L07305.
- [16] W. B. Frank, N. M. Shapiro, A. L. Husker, V. Kostoglodov, A. Romanenko, and M. Campillo, "Using systematically characterized low-frequency earthquakes as a fault probe in Guerrero, Mexico," *J. Geophys. Res., Solid Earth*, vol. 119, no. 10, pp. 7686–7700, Oct. 2014.
- [17] A. Gallego, R. M. Russo, D. Comte, V. Mocanu, R. E. Murdie, and J. C. VanDecar, "Tidal modulation of continuous nonvolcanic seismic tremor in the Chile triple junction region," *Geochem., Geophys., Geosyst.*, vol. 14, no. 4, pp. 851–863, Apr. 2013.
- [18] M. J. Kim, S. Y. Schwartz, and S. Bannister, "Non-volcanic tremor associated with the March 2010 Gisborne slow slip event at the Hikurangi subduction margin, New Zealand," *Geophys. Res. Lett.*, vol. 38, no. 14, pp. 1–4, Jul. 2011.
- [19] J. I. Walter, S. Y. Schwartz, J. M. Protti, and V. Gonzalez, "Persistent tremor within the northern Costa Rica seismogenic zone," *Geophys. Res. Lett.*, vol. 38, no. 1, pp. 1–5, Jan. 2011.
- [20] Z. Peng, H. Gonzalez-Huizar, K. Chao, C. Aiken, B. Moreno, and G. Armstrong, "Tectonic tremor beneath cuba triggered by the Mw 8.8 maule and Mw 9.0 Tohoku-Oki earthquakes," *Bull. Seismolog. Soc. Amer.*, vol. 103, no. 1, pp. 595–600, Feb. 2013.
- [21] H. Guo, H. Zhang, R. M. Nadeau, and Z. Peng, "High-resolution deep tectonic tremor locations beneath the san Andreas fault near Cholame, California, using the double-pair double-difference location method," *J. Geophys. Res., Solid Earth*, vol. 122, no. 4, pp. 3062–3075, Apr. 2017.
- [22] J. L. Rubinstein, M. La Rocca, J. E. Vidale, K. C. Creager, and A. G. Wech, "Tidal modulation of nonvolcanic tremor," *Science*, vol. 319, no. 5860, pp. 186–189, Jan. 2008.
- [23] N. J. van der Elst, A. A. Delorey, D. R. Shelly, and P. A. Johnson, "Fortnightly modulation of san Andreas tremor and low-frequency earthquakes," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 31, pp. 8601–8605, Aug. 2016.
- [24] J. Gomberg, J. L. Rubinstein, Z. Peng, K. C. Creager, J. E. Vidale, and P. Bodin, "Widespread triggering of nonvolcanic tremor in California," *Science*, vol. 319, no. 5860, p. 173, 2008.
- [25] A. G. Wech and K. C. Creager, "Automated detection and location of Cascadia tremor," *Geophys. Res. Lett.*, vol. 35, no. 20, pp. 1–5, 2008.
- [26] N. Poiata, C. Satriano, J.-P. Vilotte, P. Bernard, and K. Obara, "Multi-band array detection and location of seismic sources recorded by dense seismic networks," *Geophys. J. Int.*, vol. 205, no. 3, pp. 1548–1573, Jun. 2016.
- [27] H. Kao, S.-J. Shan, H. Dragert, G. Rogers, J. F. Cassidy, and K. Ramachandran, "A wide depth distribution of seismic tremors along the northern cascadia margin," *Nature*, vol. 436, no. 7052, p. 841, 2005.
- [28] M. La Rocca *et al.*, "Cascadia tremor located near plate interface constrained by S minus P wave times," *Science*, vol. 323, no. 5914, pp. 620–623, Jan. 2009.
- [29] B. Rouet-Leduc, C. Hulbert, I. W. McBrearty, and P. A. Johnson, "Probing slow earthquakes with deep learning," *Geophys. Res. Lett.*, vol. 47, no. 4, Feb. 2020, Art. no. e2019GL085870.
- [30] C. X. Ren, C. Hulbert, P. A. Johnson, and B. Rouet-Leduc, "Machine learning and fault rupture: A review," *Adv. Geophys.*, vol. 61, pp. 57–107, Jan. 2020.
- [31] B. Kim, J. Seo, and T. Jeon, "Bridging adversarial robustness and gradient interpretability," 2019, *arXiv:1903.11626*.
- [32] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.
- [33] D. Erhan, Y. Bengio, A. Courville, and P. Vincent, "Visualizing higher-layer features of a deep network," *Univ. Montreal*, vol. 1341, no. 3, p. 1, 2009.
- [34] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2014, pp. 818–833.
- [35] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, "Understanding neural networks through deep visualization," 2015, *arXiv:1506.06579*.
- [36] D. Baehrens, T. Schroeter, S. Harmeling, M. Kawanabe, K. Hansen, and K.-R. Müller, "How to explain individual classification decisions," *J. Mach. Learn. Res.*, vol. 11, no. 6, pp. 1803–1831, 2010.
- [37] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," 2013, *arXiv:1312.6034*.
- [38] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014, *arXiv:1412.6806*.
- [39] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," 2017, *arXiv:1704.02685*.
- [40] A. Binder, G. Montavon, S. Lapuschkin, K.-R. Müller, and W. Samek, "Layer-wise relevance propagation for neural networks with local renormalization layers," in *Proc. Int. Conf. Artif. Neural Netw.*, Cham, Switzerland: Springer, 2016, pp. 63–71.
- [41] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," 2017, *arXiv:1703.01365*.
- [42] G. Montavon, S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller, "Explaining nonlinear classification decisions with deep Taylor decomposition," *Pattern Recognit.*, vol. 65, pp. 211–222, May 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320316303582>
- [43] M. Nakamura, "Distribution of low-frequency earthquakes accompanying the very low frequency earthquakes along the Ryukyu trench," *Earth, Planets Space*, vol. 69, no. 1, p. 49, Dec. 2017.
- [44] R. S. Crosson and L. J. Nason, "Compilation of earthquake hypocenters in Western Washington-1975," Dept. Natural Resour., Internal Rep., 1978, vol. 64.
- [45] A. M. Thomas, A. Inbal, J. Searcy, D. R. Shelly, and R. Bürgmann, "Identification of low-frequency earthquakes on the san Andreas fault with deep learning," *Geophys. Res. Lett.*, vol. 48, no. 13, Jul. 2021, Art. no. e2021GL093157.
- [46] *Canadian National Seismograph Network*, Geological Survey of Canada, Ottawa, ON, Canada, 1989.
- [47] P. G. Silver *et al.*, "A plate boundary observatory," *Iris Newsl.*, vol. 16, no. 2, p. 3, 1999.



Claudia Hulbert received a B.S. degree in mathematics from University Pierre and Marie Curie, Paris, France, in 2015, a M.S. degree in applied mathematics, specialized in machine learning, from Ecole Polytechnique, Palaiseau, France, in 2017, and a Ph.D. degree in geophysics from Ecole Normale Supérieure, Paris, in 2022.

Her primary research interests include applications of machine learning in seismology to study earthquake nucleation and the connection between slip modes.



Romain Jolivet received the Ph.D. degree in geophysics from the Université de Grenoble Alpes, Grenoble, France, in 2011.

He was a Post-Doctoral Researcher at the California Institute of Technology, Pasadena, CA, USA, and Cambridge University, Cambridge, U.K. He is currently a *maître de conférences* (Associate Professor) with the Laboratoire de Géologie, Département de Géosciences, Ecole normale supérieure, Paris, France. He is also a Junior Member of the Institut Universitaire de France, Paris. His research interests include the analysis of surface deformation from satellite and seismic data, with a focus on active faults and associated earthquakes.



Blandine Gardonio received the Ph.D. degree in geophysics from the Université de Grenoble Alpes, Grenoble, France, in 2017.

She was a Post-Doctoral Researcher with the Laboratoire de Géologie, Ecole normale supérieure, Paris, France. She is currently a *chargée de recherche* (Researcher) at the CNRS, Université de Lyon, Villeurbanne, France. Her research interests include the analysis of all kinds of seismological signals for the study of active faults and earthquakes.



Paul A. Johnson is currently a Staff Member at the Los Alamos National Laboratory, Los Alamos, NM, USA. His interests include the physics of faulting and fracturing, and the elastic and nonlinear elastic and plastic properties of materials and application of machine learning methods to geophysical problems.

Dr. Johnson is a fellow of the Los Alamos National Laboratory, the American Geophysical Union, the Acoustical Society of America, and the American Physical Society.



Christopher X. Ren received the Ph.D. degree in materials science from Cambridge University, Cambridge, U.K., in 2016.

He was a Post-Doctoral Researcher and then a Scientist with the Los Alamos National Laboratory, Los Alamos, NM, USA. He is currently an Assistant Professor with the Disaster Prevention Research Institute, Kyoto University, Kyoto, Japan. His research interests include machine learning applications to geophysics, with the goal of identifying signatures of slow earthquakes and earthquake nucleation from

laboratory to field scale.



Bertrand Rouet-Leduc received the Ph.D. degree in materials science from Cambridge University, Cambridge, U.K., in 2018.

He is currently a Scientist at the Los Alamos National Laboratory, Los Alamos, NM, USA. His research interests include machine learning applications in geophysics, with the goal of identifying signatures of slow earthquakes and earthquake nucleation in laboratory, seismic, and interferometric synthetic aperture radar (InSAR) data.