Machine learning analysis of seismograms reveals a continuous plumbing system evolution beneath the Klyuchevskoy volcano in Kamchatka, Russia

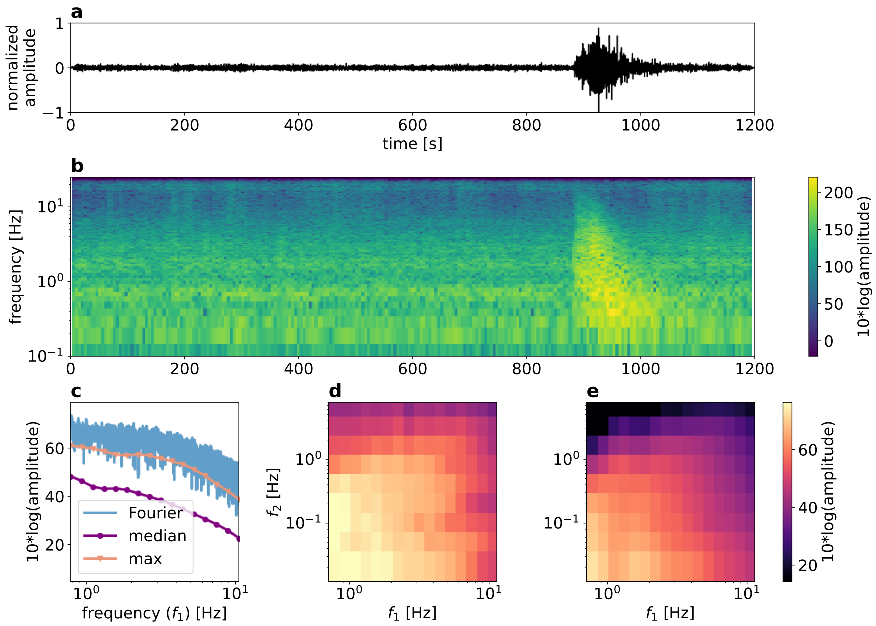
Rene Steinmann[1], L'eonard Seydoux[2], Cyril Journeau[3], Nikolai M. Shapiro[4], and Michel Campillo[5]

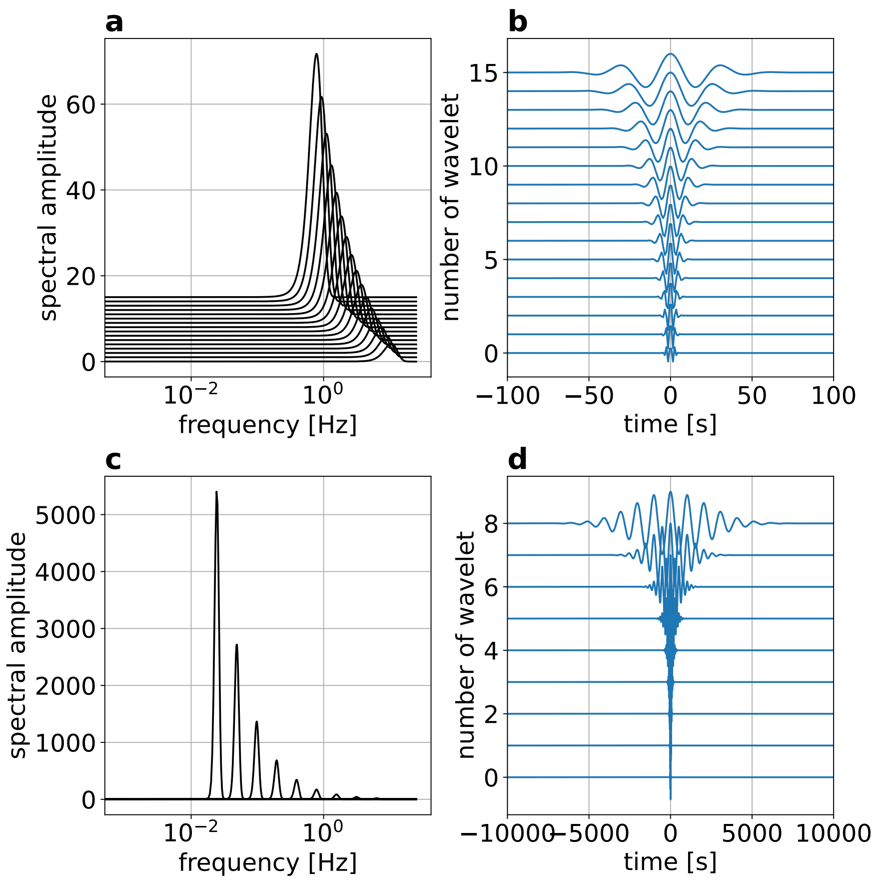
[1]Helmholtz center Potsdam
[2]Institut de Physique du Globe
[3]Institut des Sciences de la Terre (ISTerre)
[4]Institut de Sciences de la Terre, Universit e Grenoble Alpes, CNRS
[5]Université Grenoble Alpes

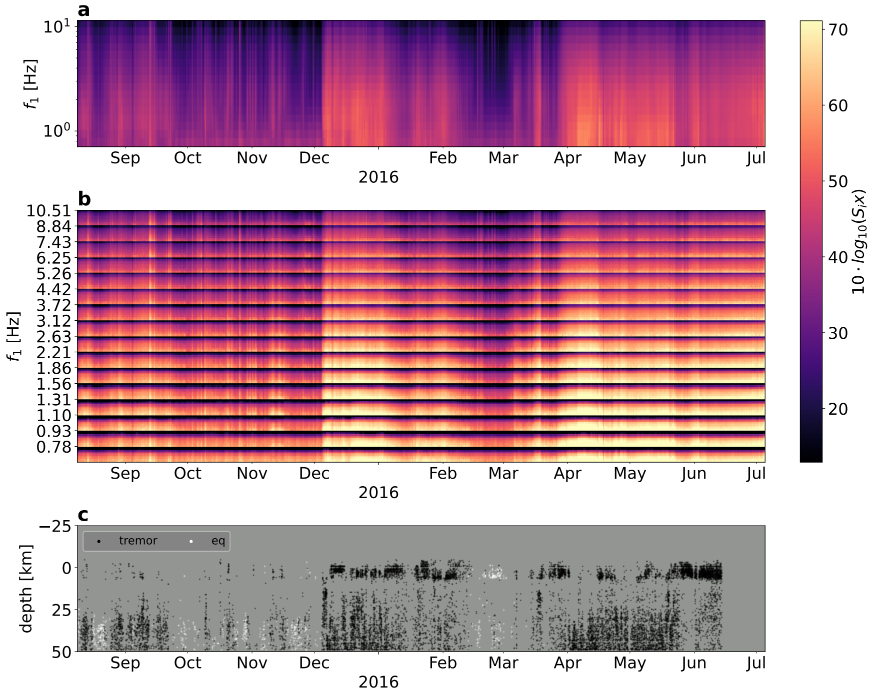
June 7, 2023

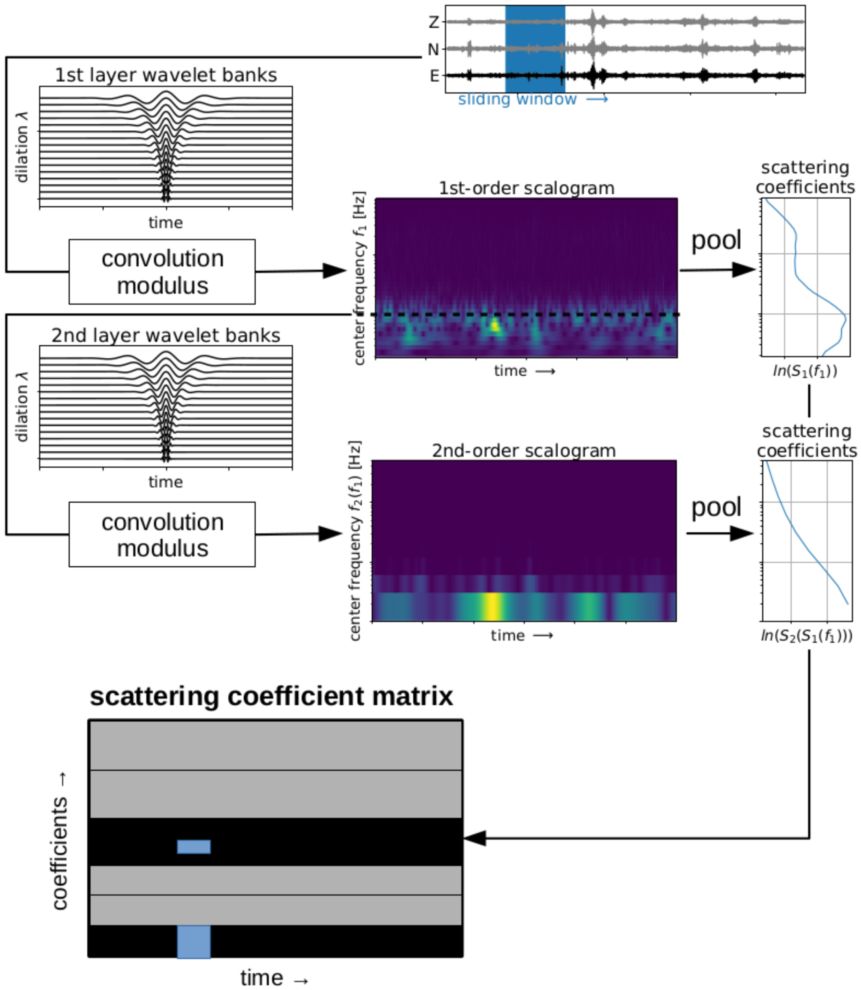
Abstract

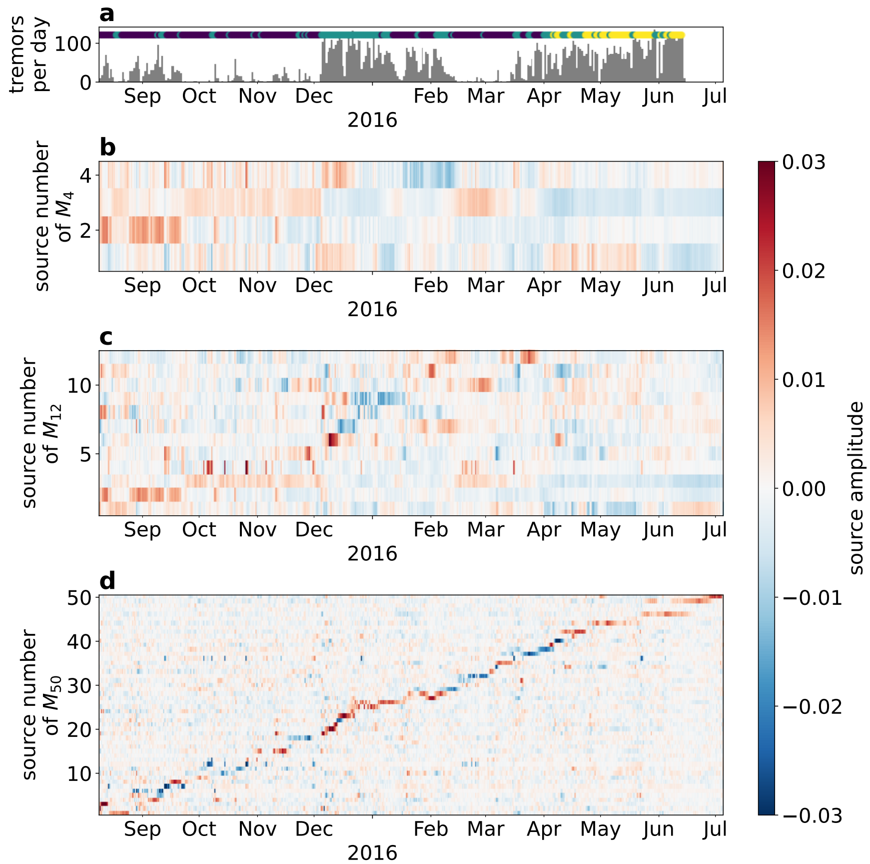
Seismic time series provide crucial information for monitoring the state of a volcano with discrete event catalogs describing impulsive seismic activity and hand-designed features describing more emergent signals (e.g. real-time seismic amplitude measurement for volcanic tremor signals). However, the emergent and long-term seismo-volcanic activity such as volcanic tremors are a complex and non-stationary phenomena that might contain more information than current methods can retrieve. In the present study, we consider the whole seismic time series as a valuable source of information by retrieving data-driven continuous features with an independent component analysis (ICA) and seismogram atlases with Uniform Manifold Approximation and Projection (UMAP). The data of interest are year-long seismic time series recorded at individual stations near the Klyuchevskoy Volcanic Group (Kamchatka, Russia). The features extracted from data recorded close to the active volcano depict a succession of short-lived patterns in the time series, indicating continuously changing signal characteristics. Additionally, the seismogram atlas reveals that, especially during periods of volcanic activation, the signal evolves continuously with some occasional sudden changes, resulting in new patterns throughout the recording time. The features and seismogram atlases reveal unique characteristics of the continuous seismograms recorded close to the volcano and related to its activity, suggesting that the complete seismic time series contains subtle but interesting information not captured by conventional methods. The seismogram atlases open new avenues to perceive large seismic time series visually and to connect the signal changes to physical processes.


a
normalized amplitude
time [s]
b
frequency [Hz]
10*log(amplitude)
c
10*log(amplitude)
Fourier
median
max
frequency (f_1) [Hz]
d
f_2 [Hz]
f_1 [Hz]
e
f_1 [Hz]
10*log(amplitude)

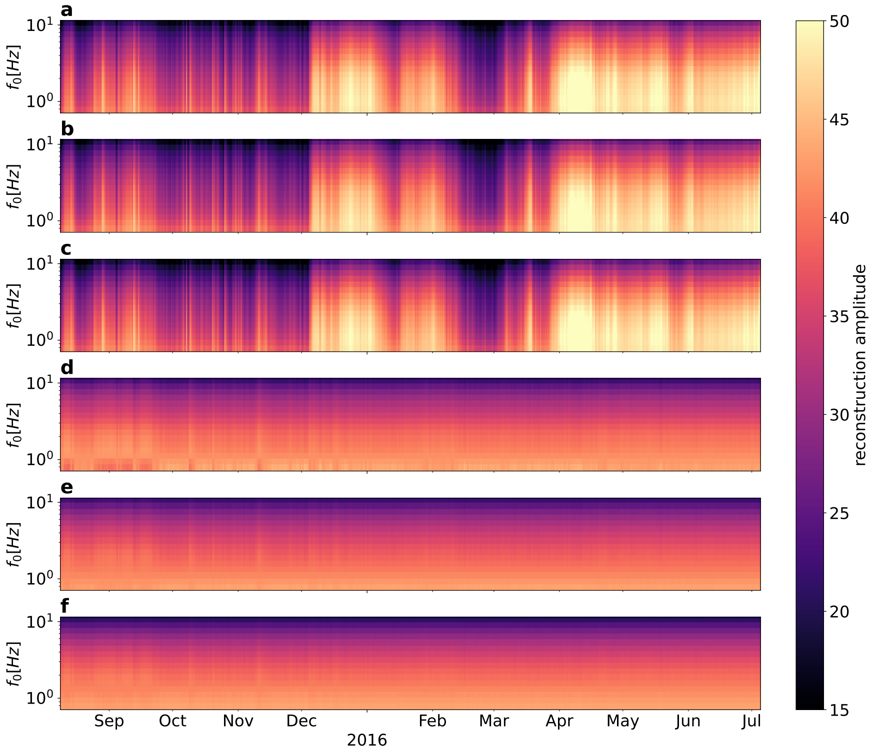

a
spectral amplitude
frequency [Hz]
b
number of wavelet
time [s]
c
spectral amplitude
frequency [Hz]
d
number of wavelet
time [s]

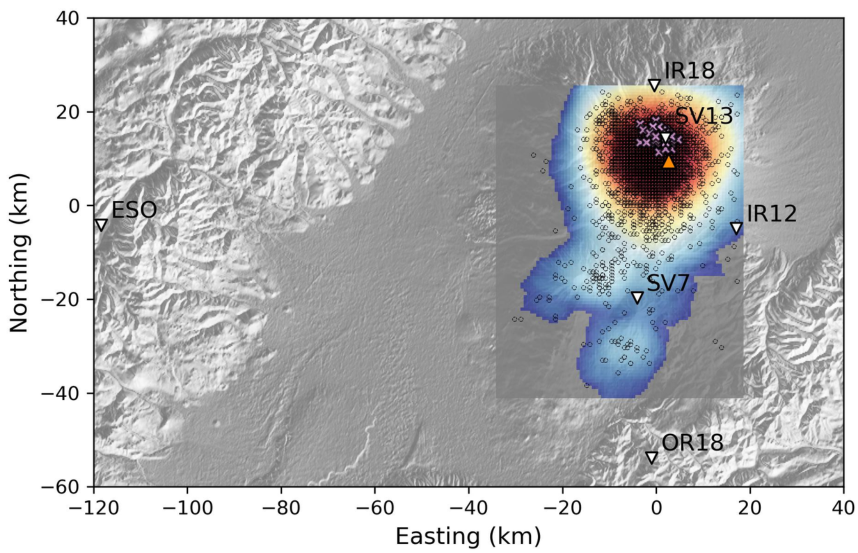

a
f_1 [Hz]
Sep Oct Nov Dec Feb Mar Apr May Jun Jul
2016
b
10.51 8.84 7.43 6.25 5.26 4.42 3.72 3.12 2.63 2.21 1.86 1.56 1.31 1.10 0.93 0.78
c
depth [km]
tremor
eq
−25 0 25 50
$10 \cdot log_{10}(S_i x)$
70 60 50 40 30 20

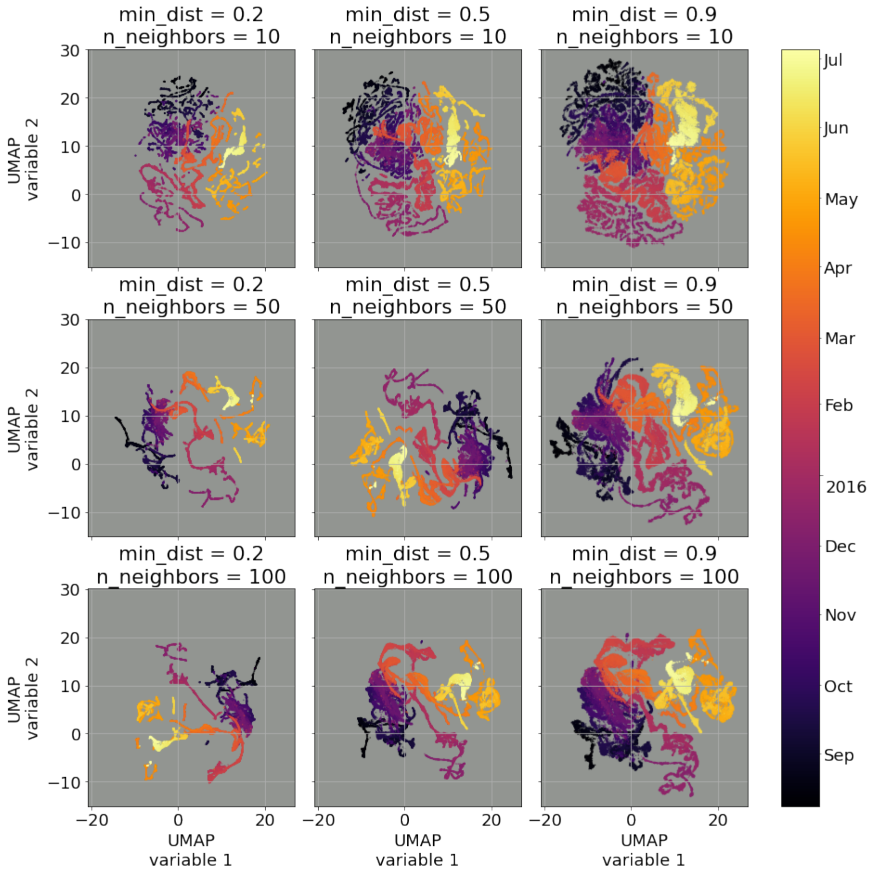

Z
N
E
sliding window ⟶
1st layer wavelet banks
dilation λ
time
convolution modulus
1st-order scalogram
center frequency f_1 [Hz]
time ⟶
pool
scattering coefficients
$ln(S_1(f_1))$
2nd layer wavelet banks
dilation λ
time
convolution modulus
2nd-order scalogram
center frequency $f_2(f_1)$ [Hz]
time ⟶
pool
scattering coefficients
$ln(S_2(S_1(f_1)))$
scattering coefficient matrix
coefficients →
time →



a
tremors per day
100
0
Sep
Oct
Nov
Dec
Feb
Mar
Apr
May
Jun
Jul
2016
b
source number of M_4
4
2
c
source number of M_{12}
10
5
d
source number of M_{50}
50
40
30
20
10
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
source amplitude


a
b
c
d
e
f
$f_0[Hz]$
10^1
10^0
Sep
Oct
Nov
Dec
2016
Feb
Mar
Apr
May
Jun
Jul
reconstruction amplitude
50
45
40
35
30
25
20
15


IR18
SV13
ESO
IR12
SV7
OR18
Northing (km)
Easting (km)
40
20
0
−20
−40
−60
−120
−100
−80
−60
−40
−20
0
20
40



min_dist = 0.2
n_neighbors = 10
min_dist = 0.5
n_neighbors = 10
min_dist = 0.9
n_neighbors = 10
min_dist = 0.2
n_neighbors = 50
min_dist = 0.5
n_neighbors = 50
min_dist = 0.9
n_neighbors = 50
min_dist = 0.2
n_neighbors = 100
min_dist = 0.5
n_neighbors = 100
min_dist = 0.9
n_neighbors = 100
UMAP
variable 2
UMAP
variable 1
30
20
10
0
−10
−20
Jul
Jun
May
Apr
Mar
Feb
2016
Dec
Nov
Oct
Sep

ac ine learning anal sis of seismograms re eals continuous lumbing s stem e olution beneat t e l uc e s o olcano in am c at a ussia

[1,2] [3] [2,4] [2]
[2]

[1]GFZ [illegible]
[2]ISTe, [illegible] CNRS, [illegible] IRD, [illegible]
[illegible]
[3][illegible] CNRS, [illegible]
[4][illegible] OR, USA

- With machine learning, we analyze one year long seismic time series at individual stations at Klyuchevskoy volcano
- It seems that the stations close to the volcano witness a constant flux of information coming from the volcano
- Throughout the recording time, the signal content close to the volcano always changes and rarely repeats itself

Corresponding author: René Steinmann, [illegible]

Seismic time series provide crucial information for monitoring the state of a volcano with discrete event catalogs describing impulsive seismic activity and hand-designed features describing more emergent signals (e.g. real-time seismic amplitude measurement for volcanic tremor signals). However, the emergent and long-term seismo-volcanic activity such as volcanic tremors are a complex and non-stationary phenomena that might contain more information than current methods can retrieve. In the present study, we consider the whole seismic time series as a valuable source of information by retrieving data-driven continuous features with an independent component analysis (ICA) and seismogram atlases with Uniform Manifold Approximation and Projection (UMAP). The data of interest are year-long seismic time series recorded at individual stations near the Klyuchevskoy Volcanic Group (Kamchatka, Russia). The features extracted from data recorded close to the active volcano depict a succession of short-lived patterns in the time series, indicating continuously changing signal characteristics. Additionally, the seismogram atlas reveals that, especially during periods of volcanic activation, the signal evolves continuously with some occasional sudden changes, resulting in new patterns throughout the recording time. The features and seismogram atlases reveal unique characteristics of the continuous seismograms recorded close to the volcano and related to its activity, suggesting that the complete seismic time series contains subtle but interesting information not captured by conventional methods. The seismogram atlases open new avenues to perceive large seismic time series visually and to connect the signal changes to physical processes.

Seismic time series are a valuable source for monitoring volcanic activity. Traditional methods rely on discrete event catalogs and hand-designed features to analyze seismic signals, but they may not capture all the valuable information, especially for long-term volcanic tremors. To overcome this limitation, we applied machine learning techniques on the continuous seismic time series, capturing patterns in a data-driven fashion. This approach reveals a continuously evolving seismogram close to the volcano, indicating ongoing changes in signal characteristics during and outside cataloged tremor periods. Additionally, a two-dimensional representation of the time series data –called a seismogram atlas –showed that, during periods of volcanic activity, the seismic signal evolved continuously with occasional sudden changes, resulting in new patterns throughout the recording period. These findings highlight the unique characteristics of continuous seismograms near the volcano, suggesting that there is valuable information in the complete seismic time series that conventional methods may miss. The seismogram atlases offer a new visual approach to analyze large seismic data and establish connections between signal changes and underlying physical processes.

Volcanoes produce a wide range of seismic signals providing valuable information about the underlying magmatic feeding systems and dynamics (e.g. Chouet & Matoza, 2013; Journeau et al., 2022; Wilding et al., 2022). Seismo-volcanologists have classified seismic signals with volcanic origin into distinct classes based on the source mechanism and signal characteristics. These classes include volcanic-tectonic earthquakes, long-period (LP) events, hybrid events, tornillos, rockfalls, and volcanic tremors (an overview of different signal classes is given by, e.g., McNutt, 2005; Chouet & Matoza, 2013). Tools adapted from earthquake seismology can detect the short-duration impulsive signals in continuous seismograms and most often locate their underlying sources, resulting in a discrete event catalog.

Long-duration signals such as volcanic tremors can last from minutes to months and have a varying appearance in frequency and amplitude (e.g. Julian, 1994; Konstantinou &

Schlindwein, 2003; Hotovec et al., 2013; Unglert & Jellinek, 2015). Some studies observed a continuous transition from discrete LP events to tremor episodes and back, making the boundary between these two signal classes blurry (e.g. Latter, 1979; Fehler, 1983). Often, an observed tremor signal in the data can not be directly linked to a single process, since many different source mechanisms may act simultaneously, with potential interactions, resulting in a non-stationary mixed signal (e.g. Konstantinou & Schlindwein, 2003; Chouet & Matoza, 2013). The complex nature of tremor signals makes it difficult to extract meaningful information from the data and link it to volcanic processes and challenge the notion of tremor signals as a single signal class. While volcano observatories often use simple single-station measurements based on the occurrence of volcanic tremors to monitor the state of the volcano, recent studies have developed more sophisticated methods to identify and locate tremor sources within a given time window (Seydoux et al., 2016; Soubestre et al., 2018, 2019; Journeau et al., 2020, 2022).

Machine learning provides a promising approach for automatically analyzing large amounts of continuous seismograms and inferring patterns in a data-driven fashion. Supervised models perform well for tasks such as signal detection and classification of cataloged signals (e.g., Malfante et al., 2018; Titos et al., 2018; Lara et al., 2020). However, for tremor signals, supervised models are problematic due to the a-priori information given by the label "volcanic tremor", referring to a complex signal with many possible source mechanisms. In contrast, unsupervised models can infer patterns from continuous seismograms without requiring predefined labels (e.g. Köhler et al., 2010; Holtzman et al., 2018; Seydoux et al., 2020; Ren et al., 2020; Jenkins et al., 2021; Steinmann, Seydoux, Beaucé, & Campillo, 2022; Steinmann, Seydoux, & Campillo, 2022; Zali et al., 2023).

In this study, we explore individual year-long continuous seismograms recorded in the vicinity of Klyuchevskoy volcano (Kamchatka, Russia) during an active tremor-dominated period using independent component analysis (ICA, Comon (1994)) and Uniform Manifold Approximation and Projection (UMAP, McInnes et al. (2018)). Given the complexity of seismic signals in a volcanic environment, we believe that continuous seismograms offer new and different insights into the inner workings of a volcano than current discrete event catalogs or supervised classification schemes can provide. ICA retrieves continuous features from the seismic time series, describing the temporal evolution of signal patterns. We are motivated by the results presented in Steinmann, Seydoux, and Campillo (2022) where the authors capture blindly the signal-altering effect of superficial surface freezing and thawing onto a single independent component. In addition, the seismogram atlas — a two-dimensional data representation of the seismic time series obtained using UMAP — offers a novel way to visualize the signal content of large seismic time series. By avoiding clustering and focusing on the analysis of the features and the seismogram atlas, we can observe the continuous evolution of the signal characteristics over time, providing a more complete picture of the mixing of different non-stationary seismic sources in seismo-volcanic signals.

The Klyuchevskoy volcano group (KVG) is one of the largest and most active clusters of subduction volcanoes in the World (e.g., Fedotov et al., 2010; Shapiro, Sens-Schönfelder, et al., 2017). Its origin is related to the unique tectonic setting at the corner between the Kuril-Kamchatka and Aleutian trenches. The enhanced supply of the melt from the mantle might be caused by the around-slab-edge asthenospheric flow (Yogodzinski et al., 2001; Levin et al., 2002) and related crustal extension (Green et al., 2020; Koulakov et al., 2020) or fluids released from the thick, highly hydrated Hawaiian-Emperor crust subducted beneath this corner(Dorendorf et al., 2000).

The sustained volcanic activity of the KVG results in nearly constantly occurring seismicity including long periods of seismo-volcanic tremors (Droznin et al., 2015; Soubestre et al., 2018, 2019; Journeau et al., 2022) and numerous earthquakes (Senyukov et al., 2009;

Thelen et al., 2010; Senyukov, 2013; Koulakov et al., 2021). In particular, the deep long-period earthquakes (DLP) have been observed at the crust-mantle boundary beneath the Klyuchevskoy volcano(Gorelchik et al., 2004; Levin et al., 2014; Shapiro, Droznin, et al., 2017; Frank et al., 2018; Galina et al., 2020; Melnik et al., 2020). The temporal correlation between the deep and shallow seismic activity has been attributed to the transfer of the fluid pressure from the deep-seated parts of the magmatic system towards shallow magmatic reservoirs beneath the active volcanoes (Shapiro, Droznin, et al., 2017; Journeau et al., 2022).

We use the data of a joint Russian-German-French temporary seismic experiment KISS (Klyuchevskoy Investigation –Seismic Structure of an Extraordinary Volcanic System) (Shapiro, Sens-Schönfelder, et al., 2017). We analyze continuous three-component seismograms, which were recorded by six individual stations (Figure 1) between July 2015 and July 2016. At the beginning of this period, all KVG and surrounding volcanoes were relatively quiescent. The Klyuchevskoy volcano showed signs of reactivation from December 2015 to January 2016 and its full eruption unfolded starting from April 2016. Journeau et al. (2022) used the network's spectral covariance matrix (Seydoux et al., 2016) to detect and locate the most prominent signal in a continuously moving time window. Detections with relatively well-constrained spatial location were labeled as earthquakes and those with poorly constrained locations as tremors. We use this catalog in the present study to relate the data-driven products of the continuous seismograms with known signal types. We want to emphasize here that the catalog is a valuable source of information in validating data-driven results, however, it holds the ground truth neither. The differentiation between earthquake- and tremor-like signals is based on a manually set threshold and the transition between these two event types is very likely continuous.

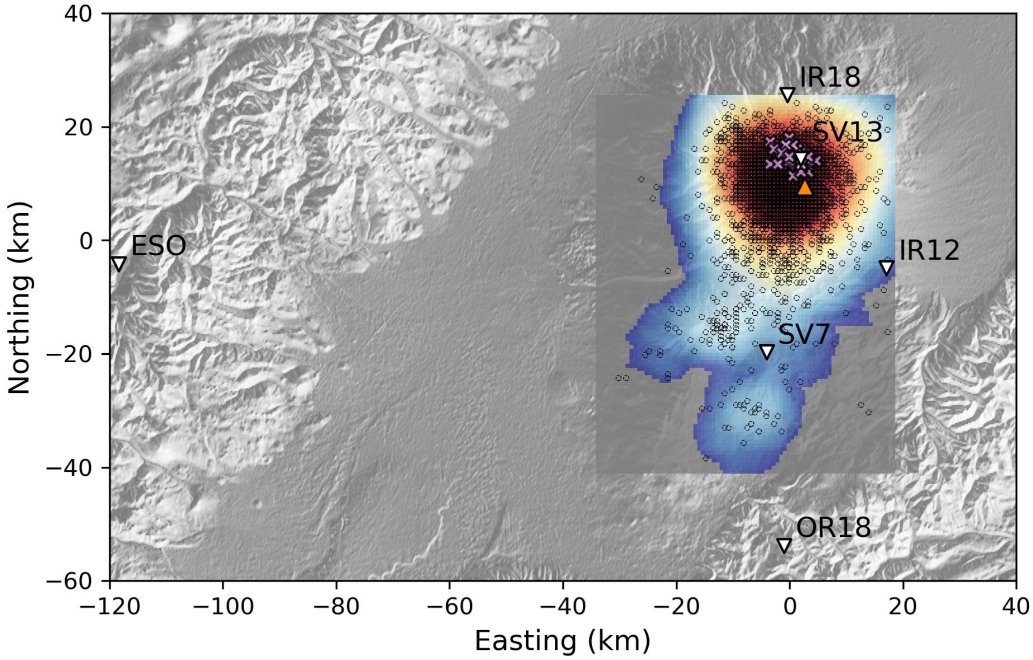
4

In the following, we want to outline how we create data-driven features and seismogram atlases from continuous seismograms. We first introduce the scattering network which retrieves a scattering coefficient matrix from the continuous seismogram. The scattering coefficient matrix serves as an input for ICA to create the data-driven features and for UMAP to create the seismogram atlases.

4

First, we apply a scattering network with a sliding window to the continuous three-component seismograms to retrieve the scattering coefficients (Figure 2). Recently, this type of network has been used in a number of seismological studies, capturing intriguing patterns within continuous seismograms (Seydoux et al., 2020; Barkaoui et al., 2021; Rodríguez et al., 2021; Steinmann, Seydoux, Beaucé, & Campillo, 2022; Steinmann, Seydoux, & Campillo, 2022; Moreau et al., 2022). The architecture resembles a convolutional neural network with the difference that each layer produces an output and the convolutional filters, classically learned in the case of convolutional neural networks, are restricted to a set of predefined wavelets (Bruna & Mallat, 2013; Andén & Mallat, 2014). Considering a mother wavelet $\psi(t)$, we can define a set of filter bank $\psi_\lambda(t) = \lambda\psi(\lambda t)$ by dilating the mother wavelet $\psi(t)$ with a set of dilation factors $\lambda \in \mathbb{R}$. In the frequency domain, the set of wavelet banks would be $\hat{\psi}_\lambda(\omega) = \hat{\psi}(\omega/\lambda)$. The dilation factor λ can then be defined as

$$\lambda = 2^{\frac{k}{Q}}, \quad k = \{0, 1, \ldots, JQ - 1\}, \tag{1}$$



Map of Klyuchevskoy Volcano Group (KVG) with the seismic stations (SV13, IR18, IR12, SV7, OR18, and ESO) considered in this study, shown with white triangles. The orange triangle shows the location of the Klyuchevskoy volcano. Averaged spatial density of the tremor source location according to Journeau et al. (2022) is shown with a colormap. Black circles and purple crosses indicate hypocenters of individual detections of tremors and deep long-period earthquakes (DLP), respectively.

with $Q \in \mathbb{N}$ being the number of wavelets per octave and $J \in \mathbb{N}$ being the number of octaves. This definition of the dilation factor provides a logarithmic grid of the center frequencies for the set of wavelet filter banks.

By convolving a time series $x(t) \in \mathbb{R}$ with a set of wavelet filter banks $\psi_\lambda(t)$ and taking the modulus (which plays the role of an activation function), we obtain a real-valued time-frequency representation $U_\lambda(t)$ of the time series called a scalogram such as

$$U_\lambda(t) = |x \star \psi_\lambda|(t), \tag{2}$$

defining the first convolutional layer of the scattering network with $\star$ standing for convolution operation. In Andén and Mallat (2014) the authors introduce a low-pass filter $\phi(t)$ to retrieve the first-order scattering coefficients, as

$$S_1 x(t, \lambda) = (W_\lambda \star \phi)(t) = (|x \star \psi_\lambda| \star \phi)(t), \tag{3}$$

where the low pass filter $\phi(t)$ smooths the representation and makes it more stable for small deformation of the signal. However, it also removes other small-scale structures of the signal which might be important for pattern recognition tasks. This information is recovered by repeating the convolution and modulus operation, retrieving higher-order scattering coefficients. Note that the set of dilation factors λ differs with the layer of the scattering network. With two sets of wavelet filter banks $\psi_{\lambda_1}(t)$ at the first layer and $\psi_{\lambda_2}(t)$ at the second layer, we can calculate the second-order scattering coefficients

$$S_2 x(t, \lambda_1, \lambda_2) = (|\, x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi)(t). \tag{4}$$

By repeating this operation many times, we can retrieve higher-order scattering coefficients which add more and more information. However, Andén and Mallat (2014) already concluded that the information gain beyond second-order scattering coefficients is marginal compared to the increasing computational costs. Therefore, we build a two-layer scattering network recovering first- and second-order scattering coefficients. The wavelets of the scattering network are Morlet wavelets as initially proposed in Andén and Mallat (2014). The Morlet wavelet $\psi(t)$ with a center frequency f is a complex exponential multiplied with a Gaussian window, defined by

$$\psi(t) = \exp(-i2\pi\, ft) \exp(-t^2/a^2). \tag{5}$$

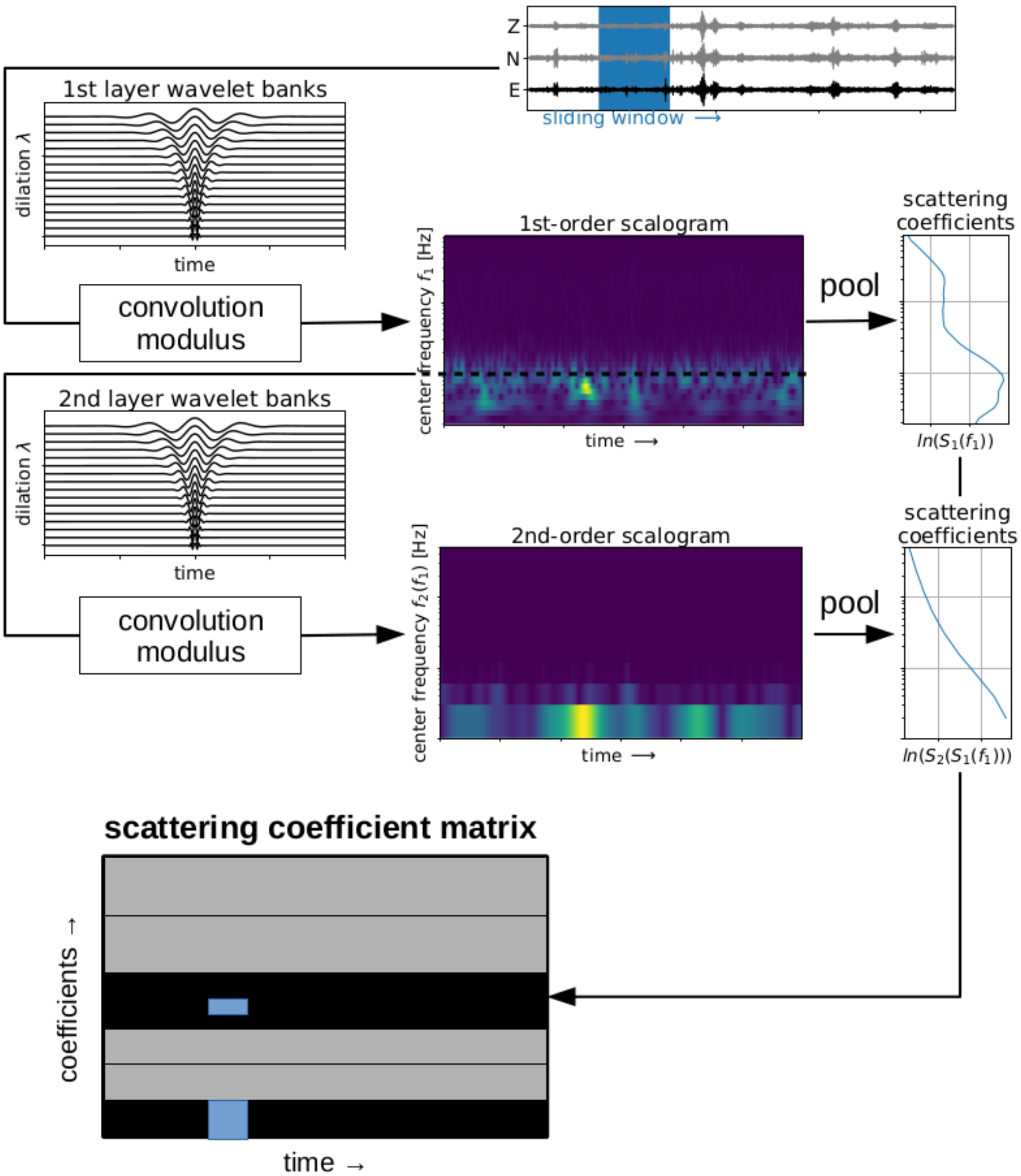
While f are the center frequencies defining the modulation of the Morlet wavelet, a defines the exponential drop-off of the waveform. We define a as a function of the bandwidth d and the center frequency f, which in turn depends on the Nyquist frequency f_N of the signal $x(t)$ and the dilation factor λ

$$a_j = \frac{d}{f} = \frac{d}{\lambda f_N}. \tag{6}$$

At last, we define our low-pass filter $\phi(t)$ retrieving the scattering coefficients from the scalogram at each layer. We use a pooling operation that ensures a stable and translation invariant representation for each window. The pooling operation retrieves a single value for each scale in the scalogram and, thus, acts as a low pass filter and downsampling operation (Dumoulin & Visin, 2016). There are many different types of pooling operations, filtering different types of information. In Seydoux et al. (2020), the authors applied the scattering network with an average pooling, which averages the scattering coefficients and collapses the time axis within the sliding window. Other possibilities are maximum pooling or median pooling where either the maximum or median value is taken for each scale in the scalogram. In this work, we will consider median pooling to mitigate the effects of impulsive short-term signals with tectonic or volcanic origin (see Appendix A for more details).

4

With ICA — a blind source separation method — we obtain the data-driven features from the scattering coefficients. The aim of ICA is the separation of multivariate signals



Detailed view of a two-layered scattering network applied to continuous three-component seismograms with a sliding window. The dashed line in the 1st-order scalogram indicates the data row which is convolved with the second-layer wavelet banks. The blue boxes in the scattering coefficient matrix show schematically where these specific scattering coefficients are stored.

into independent, non-Gaussian source signals, which can be formalized in the following way

$$ = \quad , \tag{7}$$

where $\in \mathbb{R}^{F\times N}$ are the N observations of dimension F, $\in \mathbb{R}^{F\times C}$ is the mixing matrix, and $\in \mathbb{R}^{C\times N}$ are the C independent sources. ICA estimates by applying the inverse or pseudo-inverse of the mixing matrix, called unmixing matrix, $\in \mathbb{R}^{C\times F}$ to the observed data in

$$ = \quad . \tag{8}$$

ICA solves this equation by maximizing the statistical independence of the sources in The independence is estimated by a measurement of non-Gaussianity such as the kurtosis or negentropy (Hyvärinen & Oja, 2000). The number of sources C is not known and is one of the most important parameters impacting the results of ICA. Often, this parameter is set according to a measurement estimating the information loss such as the explained variance score. Note that ICA is often described as a generalization of principal component analysis (PCA), since the independent components (sources) have no constraints of orthogonality (Comon, 1994). Also in contrast to PCA, the sign and amplitude of the independent sources can not be determined, because both and are unknown and a scaling factor can always be canceled out. Therefore, ICA does not provide any ranking to the retrieved sources. It is common practice to center and whiten the data in since it constrains the unmixing matrix to be orthogonal and therefore the number of free parameters reduces (Hyvärinen & Oja, 2000).

The scattering coefficients are collected in a data matrix in such a way that the rows contain the time series of one scattering coefficient and the columns contain all scattering coefficients for one sample (Figure 2). We retrieve independent sources from the scattering coefficients matrix and analyze their time series, which have been shown to reveal interesting patterns in seismic time series (Steinmann, Seydoux, Beaucé, & Campillo, 2022; Steinmann, Seydoux, & Campillo, 2022). We refer to the time series of the independent sources as features for the following text. In a similar mindset, Hyvärinen et al. (2010) applied ICA to the Short-Term Fourier Transform (STFT) of electroencephalography (EEG) and magnetoencephalography (MEG) time series data, revealing more interesting information related to brain activity than ICA applied to the actual time series data. Additionally, ICA has shown successful applications in analyzing various types of time series data, such as the examination of InSAR image time series (Ebmeier, 2016; Gaddes et al., 2018; Ghosh et al., 2021).

4

UMAP is a manifold learning technique, which has been introduced in the work of McInnes et al. (2018). Similar to ICA, UMAP is a tool to reduce the dimensions of a high-dimensional dataset for downstream tasks such as visualization. Since we are interested in a visualization of the high dimensional scattering coefficient matrix, we restrict the number of dimensions to two. Any dimensionality reduction technique comes with a loss of information and the loss depends on the objective of the dimensionality reduction technique. Because ICA performs a linear mapping, it preserves well the pair-wise distances but it loses information about local structures. UMAP learns the manifold of the given data and, thus, performs better in preserving local structures at the price of distorting the global structure. Hence, the distances between neighboring points are more reliable than distances between clusters of data points or the area of a cluster. Without going into further details, the inner workings of UMAP are based on topological data analysis and Riemannian Geometry, providing a complex but safe and sound mathematical background (see the original work of McInnes et al., 2018, for more details). It shares similarities with t-SNE, which has been used extensively for visualizations since its appearance in the 2000s (Van der Maaten & Hinton, 2008). However, compared to UMAP, t-SNE performs poorly in preserving global structures and its computation time is much slower (Becht et al., 2019). Despite its relatively recent introduction, UMAP has been already utilized in many scientific domains to create a two-dimensional representations, simplifying the visualization of large and high-dimensional datasets. The resulting two-dimensional UMAP spaces have been coined "atlases" such as the activation atlas of neural networks (Carter et al., 2019) or the metagenomic atlas (Lin et al., 2022).

UMAP comes with a set of hyperparameters to tune such as the number of neighbors and the minimum distance, drawing the focus either towards preserving local or global structures. The number of neighbors limits the number of neighboring points when UMAP learns the local manifold structure. A low number draws the focus to the local structure while losing the bigger picture. A large number draws the focus on the global structure while losing finer details. The minimum distance controls how closely UMAP is allowed to bring data points together. A low number results in a more dense and clumpier representation and preserves better the local structure of the data. A large number avoids putting points close to each other and draws a broader picture of the data.

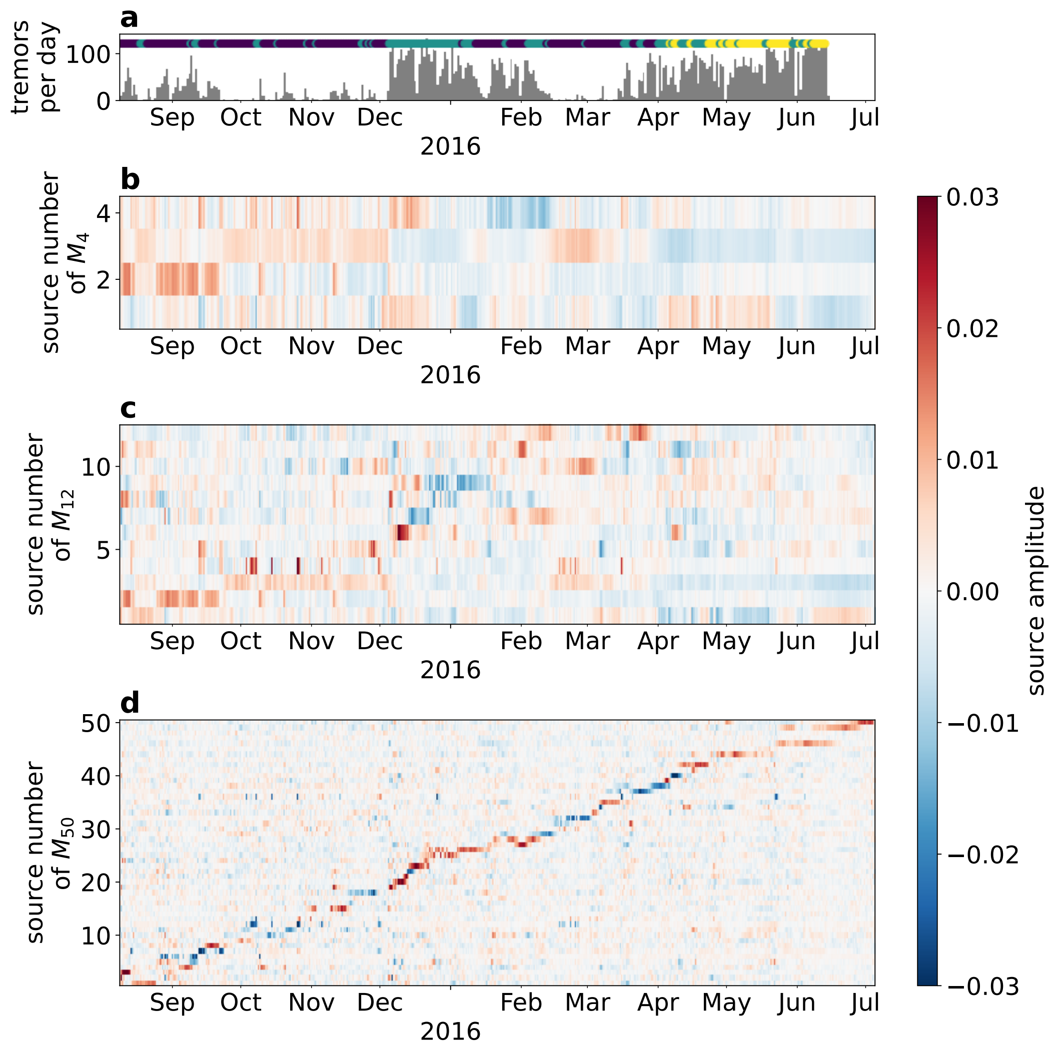
5

We apply the scattering network with a sliding window of 20 min and an overlap of 10 min to the three-component seismograms of station SV13, resulting in a temporal resolution of 10 min of the scattering coefficient matrix. The minimal amount of seismic preprocessing and the exact setup of the network is described in Appendix B and Figure B1 therein. We visualize the time series of the first- and second-order coefficients of the east channel together with the catalog of Journeau et al. (2022) in Figure B2. The scattering coefficient matrix is centered and whitened with a PCA before we apply ICA. We apply ICA models with four (M_4), 12 (M_{12}), and 50 (M_{50}) independent sources to explore the impact of the number of components. M_4 reaches an explained variance score of 94 %, M_{12} reaches an explained variance score of 98 % and M_{50} reaches an explained variance score of 99 %. Figure 3 shows the smoothed time history of the independent sources (features) of each model. The features show negative and positive values of arbitrary units centered around zero due to the centering and whitening of the input data. We sort the features according to their maximum absolute amplitude appearance in time, helping the visualization of any time-dependent processes.

The features of the three models show very different time series and in the following, we want to use the three models to understand better the underlying seismic data. First of all, we provide a qualitative comparison between the features of the three models. While there is no single feature matching between M_4 and M_{50} (Figure 3b and d), we can find similar features between M_4 and M_{12} such as the second feature in both models (Figure 3b and c). M_{50} is very different to the other two models, since its features appear more sparse, i.e. they are mainly centered around zero except for a short duration. Moreover, it is striking that if one feature shows large amplitudes in negative or positive direction (saturated blue and red colors), almost every other feature is centered around zero. These characteristics of M_{50} together with the sorting of the features result in this color-saturated diagonal line in the time-feature space. It appears that at each data point in this 50-dimensional feature space is located at the center for 49 dimensions. In contrast, the data points represented by the features of M_4 and M_{12} do have non-zero values for more than one dimension.

M

To understand better what the features represent, we recall the equation of ICA (see Equation 7 and 8). The whitened and centered scattering coefficient matrix is estimated



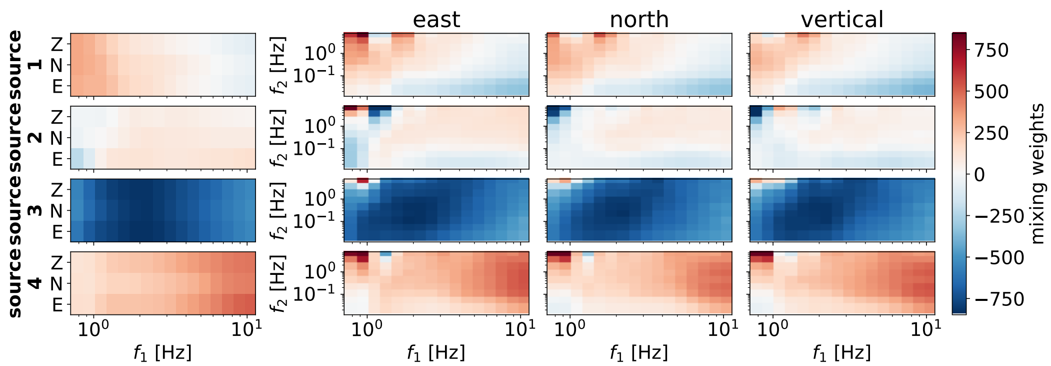
ICA results for station SV13. In , the histogram describes the daily number of tremor detections and the colored circles indicate the daily activity level of the Klyuchevskoy volcano, where yellow represents an ongoing eruption and dark blue represents low activity. shows the features of the 4-component model M_4, shows the features of the 12-component model M_{12}, shows the features of the 50-component model M_{50}. Note that the features were sorted with respect to their absolute maximum value in time for better visualization.

as the sum of rank-1 matrices, resulting from the outer product of a feature (rows in) with the corresponding columns in the mixing matrix. Hence, the columns of reveal how each feature contributes to the estimation of . The visualization of the columns of and its outer product with the corresponding feature can help to understand better the underlying signal characteristic of each feature. In Figure 4 we reorganize and visualize the mixing weights of the M_4 model according to the center frequencies of the first- and second-order wavelets. We can use the shown mixing weights to attribute signal characteristics to the features of M_4 in Figure 3b. For example, feature 3 shows mainly negative amplitudes during the occurrence of cataloged tremors and positive amplitudes during the absence of tremors (Figure 3a and b). The corresponding mixing weights show mainly negative amplitudes peaking at $f_1 = 2$ Hz in all components and for both first- and second-order scattering coefficients (Figure 4). Figure 5a, b, and c show the reconstructed first-order scattering coefficients, resulting from the outer product of the third feature with its mixing weights. We disregard the second-order scattering coefficients for visualization purposes, however, we want to emphasize that they contain important signal information. We also add the mean over the scattering coefficients, which we subtracted before the ICA during the whitening process. The reconstruction makes clear that the tremor periods are characterized by a broadband amplitude increase peaking around 2 Hz Note that both the mixing weights and the feature amplitudes during tremor-active periods are negative and the outer product reveals an amplitude increase. This example shows the ambiguity of the sign attached to the sources: any change of sign in feature 3 can be equalized with a change of sign of the corresponding column vector of the mixing matrix, resulting in the same rank-1 matrix.

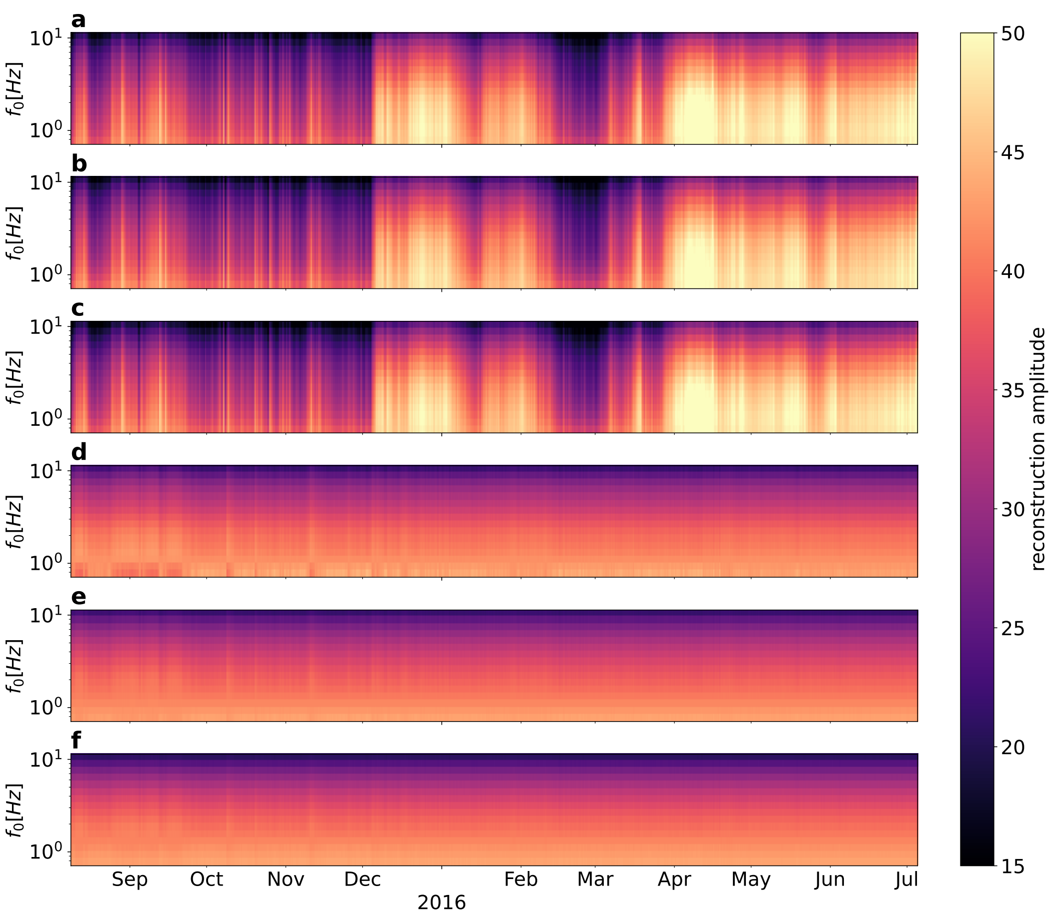
Another interesting example is the second feature, which shows strong positive amplitudes during the tremor sequences in August and September (Figure 3a and b). The corresponding mixing weights show positive and negative amplitudes depending on the frequencies f_1 and f_2 and the channel (Figure 4). Similar to before, we can visualize the first-order scattering coefficients of the obtained rank-1 matrix by the outer product of the mixing weights with the second feature (Figure 5d, e, and f). We see a clear anti-correlation for scattering coefficients below and above 1 Hz for the east channel (Figure 5d). An amplitude increase above 1 Hz occurs together with an amplitude decrease below 1 Hz (e.g. the tremor-dominated time periods in August and September). Similarly, an amplitude increase below 1 Hz occurs together with an amplitude decrease above 1 Hz (e.g. October to December 2015). This anti-correlation can be already observed by the negative and positive weights of the mixing matrix (source 2, Figure 4). Weights with the same sign show the scattering coefficients which correlate with the corresponding independent source. The observed anti-correlation is nothing physical and this rank-1 matrix reflects only a part of the data without taking into account the other independent sources. Nonetheless, Figure 5 suggests that the tremor period in August and September is different from the other tremor episodes mainly due to different patterns at the east component around 1 Hz

M

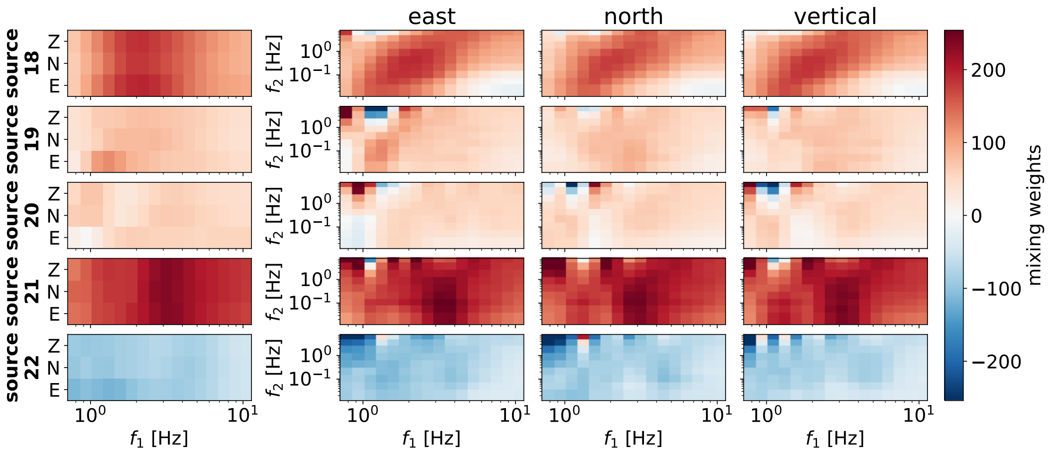
An outstanding characteristic of M_{50} is the successive domination of a single feature in time. The large number of features makes it unfeasible to visualize and interpret all mixing weights and, therefore, we only focus on a subset of features to show the difference to M_4. Figure 6 shows the mixing weights associated with the features 18 to 22, which describe the tremor onset at the beginning of December with negative and positive amplitudes. Compared to the mixing weights of the M_4 model, the mixing weights of the M_{50} show finer nuances and more complex patterns across neighboring scattering coefficients, revealing signal changes beyond pure amplitude increases or frequency shifts. At this point, we want to recall that M_4 reaches an explained variance score of 94 % and M_{50} reaches an explained variance score of 99.6 %. We suggested that the third feature of M_4 is related to the broadband energy peaking around 2 Hz and it correlates with the tremor detections. This indicates that all recorded tremors are characterized by this broadband amplitude variation and it explains a large part of the data's variance. However, this M_4 feature



Mixing weights for the M_4 model for its four sources at station SV13. For visualization purposes, we reshaped the mixing matrix to display the weights related to the first-order coefficients in the left column, and the weights related to the second-order coefficients are split into three different subplots according to the seismometer's component.



Reconstruction of the first-order scattering coefficients based on the outer product of an independent source of M_4 with its mixing weights for station SV13. Subfigures , and show the reconstruction of the first-order coefficients of the east, north and vertical component, based on the second source (Figure 3b). Subfigures , and show the reconstruction of the first-order coefficients of the east, north and vertical component, based on the third source (Figure 3b).



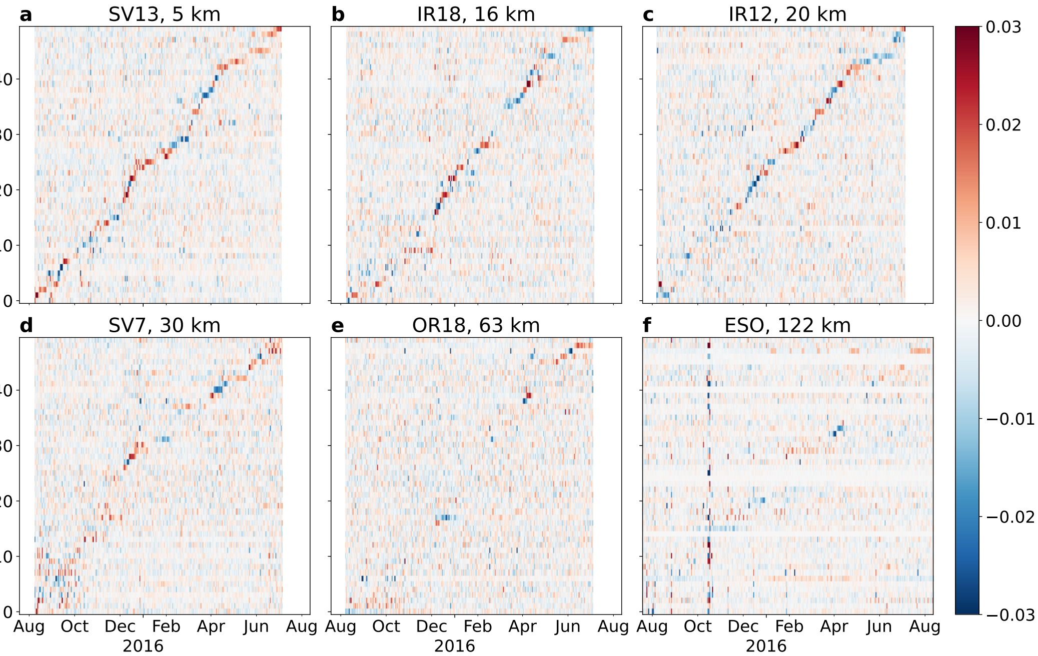
Mixing weights for five selected sources of the M_{50} model (source 18 - 22) for station SV13. For visualization purposes, we reshaped the mixing matrix to display the weights related to the first-order coefficients in the left column, and the weights related to the second-order coefficients are split into three different subplots according to the seismometer's component.

does not exist anymore with M_{50} and the tremor active periods are described by multiple features. M_{50} reveals that these tremor signals show complex variations beyond broadband amplitude variation. It seems that these more complex signal variations are a small part of the data's variance – we gain less than 6 % variance from M_4 to M_{50} – but they might contain crucial information about ongoing volcanic processes. Interestingly, these signal alterations captured by the features do not occur randomly in time but they occur for a certain amount of time before another pattern takes over.

The different ICA realization can be seen as a hierarchical ICA, where a model with a larger number of components – such as M_{50} – can account for smaller differences in the signal characteristics. In fact, this approach shares similarities to the hierarchical exploration of continuous seismograms with hierarchical clustering (Steinmann, Seydoux, Beaucé, & Campillo, 2022). However, hierarchical clustering assigns a cluster to a data point which can contain multiple types of signals. The present study shows an approach where a single data point is described by multiple features, which correspond to specific signal characteristics captured with scattering coefficients.

4

The M_{50} model of station SV13 pictures a continuous seismogram where a pattern occurs mainly once or twice in a constrained time window during the whole recording time. This seems surprising and it might be a particular characteristic of the data recorded close to the active Klyuchevskoy volcano. By retrieving M_{50} models from different stations with an increasing distance to the volcano, we can make a qualitative comparison to station SV13 (see Figure 7). The considered stations are located between 5 and 122 km away from the active volcano. In general, the diagonal line in the time feature space degrades with increasing distance to the volcano. The periods with a high detection rate of tremors (December to February and from March onwards) show this characteristic diagonal line even for stations further away such as SV7 and OR18. This suggests that tremor signals are mainly responsible for the diagonal line in the time feature space. Interestingly, stations close to the volcano, show an almost continuous diagonal line, indicating a continuous flow of information coming from the volcano. The catalog relies on a stable signal seen by the majority of stations within the network (Journeau et al., 2022). Therefore, tremor signals



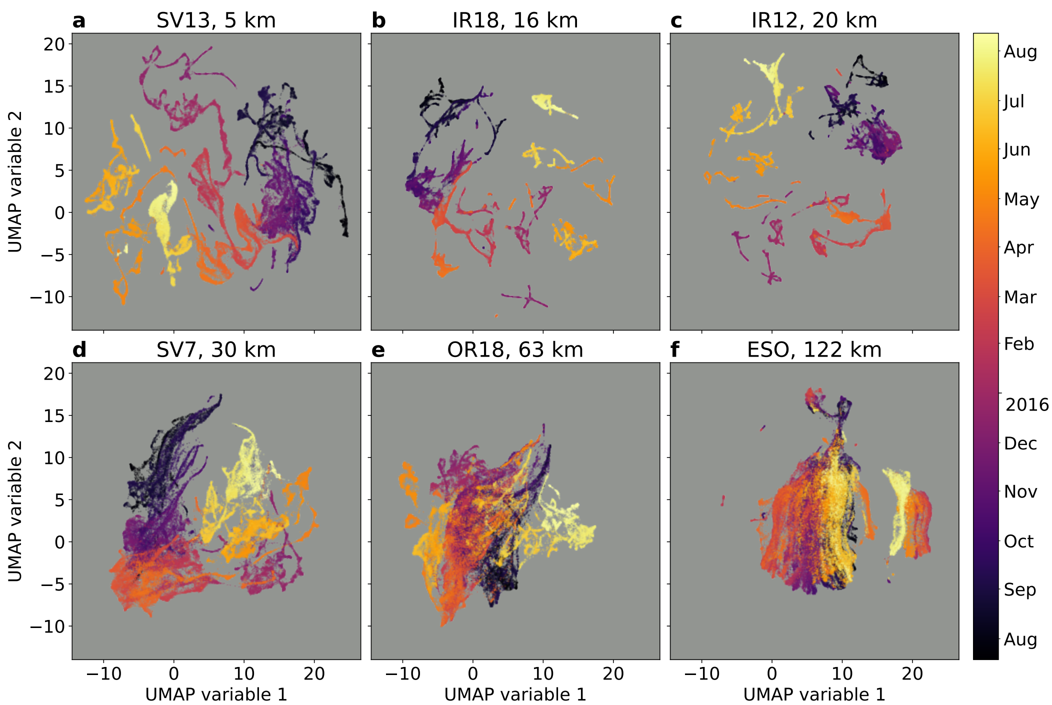
The 50 independent sources for data recorded at station SV13 , IR18 , IR12 , SV7 , OR18 and ESO . The results are ordered according to the distance to the active Klyuchevskoy volcano mentioned in the title of the subfigures.

with a weak amplitude witnessed by only close-by stations are not detected and located. The 50 features of the individual stations indicate an almost constant flux of information coming from the volcano, not captured by the catalogs.

We can complement the observations and interpretation of the features with seismogram atlases for each station obtained with UMAP (Figure 8). We color-code the data points with their corresponding calendar time, highlighting the time evolution of the seismic time series. The seismogram atlases of SV13, IR18, and IR12 picture complex structures where data points with different colors hardly overlap (Figure 8a-c). Many data points with a similar color seem to be located close to each other, which gives rise to a smooth color gradient across the complex data structure. Therefore, neighboring data points are likely close to each other in time, suggesting smooth and slow signal changes. However, there are also isolated or disconnected structures indicating more sudden signal changes from time to time, especially for stations IR12 and IR18. The atlases of the data recorded at SV7, OR18, and ESO look different: the data points are more concentrated at the center with partly overlapping colors (Figure 8e-f).

The seismogram atlases share the same hyperparameters: the minimum distance is set to 0.5 and the number of neighbors to 50. We tested different hyperparameters for the data of station SV13 (see Figure B3). The atlas depicts different cluster shapes and distances with regard to the hyperparameters, emphasizing the limited notion of global distances and structures in the atlas. However, all the examples confirm the smooth time gradient and little to no overlap of different time periods.

SV13 is the closest station to the volcano and shows a variety of connected and disconnected patterns, picturing a complex signal evolution due to the active volcano. The



The seismogram atlases obtained with UMAP for data recorded at station SV13 , IR18 , IR12 , SV7 , OR18 and ESO . The results are ordered according to the distance to the active Klyuchevskoy volcano mentioned in the title of the subfigures.

catalogs of Journeau et al. (2022) provide signal labels that can help to identify known areas in the atlas. In the upper image of Figure 9 we mark data points that match in time with a known detection of tremor signals, DLP swarms or earthquake-like signals. We want to note that Journeau et al. (2022) provides these labels based on thresholding variables of the network's covariance matrix. According to that, they label the event detections as either earthquake-like and tremor signals. In total, we have 47.360 data points and we can connect 13.027 data points to known tremor signals, 682 data points to earthquake-like signals and 38 data points to DLP swarms. Therefore, most of the 47.360 data points do not match with a cataloged event. Nevertheless, the matching data points cover almost the same area as the non-matching data points, indicating that the non-matching data points contain similar signals to the cataloged signal types. Thus, the data contains almost no quiet period (the classic seismic 'noise'), and tremor- or earthquake-like signals seem to be recorded all the time. This agrees with the findings of Makus et al. (2023) where their reference correlation function for the same dataset is dominated by tremor activity. The scattering coefficients contain information about impulsive and continuous signal characteristics and, therefore, the earthquake-like signal are separated from the tremor-like signals in the atlas. However, we see that the shared boundaries are continuous and blurry, indicating that we have continuous transitions between the two signal types. Journeau et al. (2022) observed the same characteristic within the variables space obtained from the network's covariance matrix. The tremor signals picture complex data point structures, which mainly show a continuous evolution in time (see Figure 8a and Figure 9). The earthquake-like signals form a data cloud structure in the south-east of the seismogram atlas with no strong time evolution. The DLP swarms are located close to each other at one of the blurry boundaries between tremor- and earthquake-like signals. This suggests that the different DLP swarm detections share signal similarities and they have signal characteristics of both earthquake- and tremor-like signals, confirming the common perception of DLP swarms.

To visualize the signals behind the data points and to understand better the UMAP space, we want to follow the active tremor period from the beginning of December to the end of February in the atlas and depict the spectrograms behind some data points as an example (Figure 9). Note that the atlas contains the information of the three component data but we only visualize here the spectrograms of the east component. The start of the tremor period begins with a separated linear structure in the north of the atlas, indicating a sudden onset of the tremor signals. During the first weeks until the end of December, the tremor signal is continuous and shifts in frequency and amplitude. The connected and curved structure indicates continuous changes in the signal characteristics. At the end of December repetitive impulsive events seem to superimpose the continuous signal and the connected and curved structures end in a relatively small data cloud, indicating a temporary stop of the continuous signal change. We can see something similar happening for the 50 features: a rapid succession of a few features occur at early December compared to late December (Figure 3d). From the small data cloud, the data points form another curved elongated structure during January before they connect to another data cloud structure at the end of January (Figure 9). The spectrograms show that during these times the repetitive impulsive signals change their interval time and frequency content, while the continuous broadband signals change their amplitude and frequency content (green and red framed spectrograms in Figure 9). After mid-February the continuous and impulsive signals decease and the data points enter the cluster of earthquake-like signals. The time evolution of the seimogram atlas is even better captured in the attached movie S1.

By combining ICA with a scattering network, we retrieved continuous and data-driven features from seismic time series recorded at individual stations in the vicinity of the KVG. An ICA model with 4 independent sources obtained an explained variance score of 94 % and two of its features correlated with the general occurrence of tremor signals. Thanks to the mixing weights we were able to tie one of the correlating features to a broadband amplitude increase centered around 2 Hz The low number of sources can not account for the smaller differences in the tremor signals and it finds features that describe most of the tremor signals. An ICA model with 50 independent sources obtained an explained variance score of 99.6 % and depicts a seismic time series with a succession of different short-lived patterns across the whole recording time. A comparison to other stations located further away from the active Klyuchevskoy volcano suggested that this behavior is unique to the data recorded close to the volcano. The seismogram atlases obtained with UMAP depict continuous and sudden signal characteristic changes, in particular during times of cataloged tremors. It revealed a sudden onset of tremor signals with the first occurrence of shallow tremors in early December, while the stop and re-start of tremors in February and March, respectively, are characterized by a more continuous emergence of these signals (see also movie S1). During the one-year recording time, the data points related to tremor signals continuously explore new terrain in the seismogram atlas, indicating that there is always a minimal amount of signal difference. This is in agreement with the M_{50} model where each feature dominates for a short period of time. The seismogram atlas adds information if the seismic signals change rather continuously or suddenly by placing the next data point far away or close by. One of the main limitations of this study is that we are not able to link the changing characteristics to physical processes. The mixing weights and the reconstructed scattering coefficients help linking the data-driven features to information encoded in the scattering coefficient space, however, the interpretation is limited as we have seen for the second feature in the M_4 model. Moreover, the M_{50} model provides too many features with complex mixing weight patterns, making an individual inspection not feasible. Nonetheless, the features revealed unique characteristics of the seismic time series recorded close to the volcano, indicating that the complete seismic time series contains interesting and subtle information which are not captured by conventional methods such as discrete event catalog or hand-engineered features. The seismogram atlases offer interesting and novel ways to

Figure 9. The upper image shows the seismogram atlas of station SV13. Data points matching in time with a cataloged earthquake- or tremor-like signals are colored, data points with no match are grey. Each data point represents 20 min of three component waveform data and for some data points (marked with the colored arrows) we visualized the spectrogram of the east component. The color-coding of the arrows matches the color-coding of the spectrogram's frame and the arrows point towards the next data point in time.

Figure A1. Comparison between Fourier spectrum and scattering coefficients of a seismic signal. (a) shows an example seismogram with normalized amplitude in time domain. (b) shows its corresponding Fourier spectrogram. (c) shows the Fourier amplitude spectrum and the first order median and maximum pooled scattering coefficients of the signal shown in (a). (d) shows the second order maximum pooled scattering coefficients and (e) shows the second order median pooled scattering coefficients as a function of the center frequencies f_1 and f_2.

want to focus on tremor signal.

Appendix B Data Processing and Scattering Network Design

The seismic data is demeaned, detrended, and down-sampled to a sampling rate of 25 Hz. The first layer wavelets are adapted to the possible frequency content of the tremors; their center frequencies range from 0.78 to 10 Hz with a logarithmic grid (see Figure B1a and b). The second layer wavelets start at much lower frequencies since they gather information about the modulation and shape of the signal (Figure B1c and d). The first layer covers 4 octaves and is densely spaced with 4 wavelets per octave. The second layer covers 8 octaves and is sparsely sampled with 1 wavelet per octave.

References

Andén, J., & Mallat, S. (2014). Deep scattering spectrum. *IEEE Transactions on Signal Processing*, *62*(16), 4114–4128.

Barkaoui, S., Lognonné, P., Kawamura, T., Stutzmann, E., Seydoux, L., Maarten, V., ... others (2021). Anatomy of continuous mars seis and pressure data from unsupervised learning. *Bulletin of the Seismological Society of America*, *111*(6), 2964–2981.

continuous seismic wavefield records using self-organizing maps. [illegible], [illegible] (3), 1619-1630.

Konstantinou, K. I., & Schlindwein, V. (2003). Nature, wavefield properties and source mechanism of volcanic tremor: a review. [illegible], [illegible] (1-4), 161-187.

Koulakov, I., Plechov, P., Mania, R., Walter, T. R., Smirnov, S. Z., Abkadyrov, I., … Droznina, S. Y. (2021). Anatomy of the bezymianny volcano merely before an explosive eruption on 20.12.2017. [illegible], [illegible] (1), 1758. doi: 10.1038/s41598-021-81498-9

Koulakov, I., Shapiro, N. M., Sens-Schönfelder, C., Luehr, B. G., Gordeev, E. I., Jakovlev, A., … Stupina, T. (2020). Mantle and crustal sources of magmatic activity of klyuchevskoy and surrounding volcanoes in kamchatka inferred from earthquake tomography. [illegible], [illegible] (10), e2020JB020097. doi: 10.1029/2020JB020097

Lara, P. E. E., Fernandes, C. A. R., Inza, A., Mars, J. I., Métaxian, J.-P., Dalla Mura, M., & Malfante, M. (2020). Automatic multichannel volcano-seismic classification using machine learning and emd. [illegible], [illegible], 1322-1331.

Latter, J. H. (1979). Volcanological observations at tongariro national park. ii: Types and classification of volcanic earthquakes, 1976-1978. [illegible], [illegible].

Levin, V., Droznina, S. Y., Gavrilenko, M., Carr, M. J., & Senyukov, S. L. (2014). Seismically active subcrustal magma source of the klyuchevskoy volcano in kamchatka, russia. [illegible], [illegible] (11), 983-986. doi: 10.1130/G35972.1

Levin, V., Shapiro, N., Park, J., & Ritzwoller, M. (2002). Seismic evidence for catastrophic slab loss beneath kamchatka. [illegible], [illegible] (6899), 763-767. doi: 10.1038/nature00973

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., … others (2022). Evolutionary-scale prediction of atomic level protein structure with a language model. [illegible], 2022-07.

Makus, P., Sens-Schönfelder, C., Illien, L., Walter, T. R., Yates, A., & Tilmann, F. (2023). Deciphering the whisper of volcanoes: Monitoring velocity changes at kamchatka's klyuchevskoy group with fluctuating noise fields. [illegible], e2022JB025738.

Malfante, M., Dalla Mura, M., Mars, J. I., Métaxian, J.-P., Macedo, O., & Inza, A. (2018). Automatic classification of volcano seismic signatures. [illegible], [illegible] (12), 10-645.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. [illegible].

McNutt, S. R. (2005). Volcanic seismology. [illegible], [illegible] (1), 461-491.

Melnik, O., Lyakhovsky, V., Shapiro, N. M., Galina, N., & Bergal-Kuvikas, O. (2020). Deep long period volcanic earthquakes generated by degassing of volatile-rich basaltic magmas. [illegible], [illegible] (1), 3918. doi: 10.1038/s41467-020-17759-4

Moreau, L., Seydoux, L., Weiss, J., & Campillo, M. (2022). Analysis of micro-seismicity in sea ice with deep learning and bayesian inference: application to high-resolution thickness monitoring. [illegible], 1-19.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others (2011). Scikit-learn: Machine learning in python. [illegible], [illegible], 2825-2830.

Ren, C. X., Peltier, A., Ferrazzini, V., Rouet-Leduc, B., Johnson, P. A., & Brenguier, F. (2020). Machine learning reveals the seismic signature of eruptive behavior at piton de la fournaise volcano. [illegible], [illegible] (3), e2019GL085523.

Rodríguez, Á. B., Balestriero, R., De Angelis, S., Benítez, M. C., Zuccarello, L., Baraniuk, R., … Maarten, V. (2021). Recurrent scattering network detects metastable behavior in polyphonic seismo-volcanic signals for volcano eruption forecasting. [illegible], [illegible], 1-23.

Senyukov, S. L. (2013). Monitoring and prediction of volcanic activity in kamchatka from seismological data: 2000{2010. Journal of Volcanology and Seismology, 7(1), 86{97. doi: 10.1134/S0742046313010077

Senyukov, S. L., Droznina, S. Y., Nuzhdina, I. N., Garbuzova, V. T., & Kozhevnikova, T. Y. (2009). Studies in the activity of klyuchevskoi volcano by remote sensing techniques between january 1, 2001 and july 31, 2005. Journal of Volcanology and Seismology, 3(3), 191{199. doi: 10.1134/S0742046309030051

Seydoux, L., Balestriero, R., Poli, P., De Hoop, M., Campillo, M., & Baraniuk, R. (2020). Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. Nature communications, 11(1), 1{12.

Seydoux, L., Shapiro, N. M., de Rosny, J., Brenguier, F., & Landès, M. (2016). Detecting seismic activity with a covariance matrix analysis of data recorded on seismic arrays. Geophysical Journal International, 204(3), 1430{1442.

Shapiro, N. M., Droznin, D. V., Droznina, S. Y., Senyukov, S. L., Gusev, A. A., & Gordeev, E. I. (2017). Deep and shallow long-period volcanic seismicity linked by uid-pressure transfer. Nat. Geosci., 10(6), 442{445. doi: 10.1038/ngeo2952

Shapiro, N. M., Sens-Schönfelder, C., Lühr, B. G., Weber, M., Abkadyrov, I., Gordeev, E. I., . . . Saltykov, V. A. (2017). Understanding kamchatka's extraordinary: Volcano cluster. EOS: Transactions, American Geophysical Union, 98(7), 12{17. doi: 10.1029/2017EO071351

Soubestre, J., Seydoux, L., Shapiro, N., De Rosny, J., Droznin, D., Droznina, S. Y., . . . Gordeev, E. (2019). Depth migration of seismovolcanic tremor sources below the klyuchevskoy volcanic group (kamchatka) determined from a network-based analysis. Geophysical Research Letters, 46(14), 8018{8030.

Soubestre, J., Shapiro, N. M., Seydoux, L., de Rosny, J., Droznin, D. V., Droznina, S. Y., . . . Gordeev, E. I. (2018). Network-based detection and classi cation of seismovolcanic tremors: Example from the klyuchevskoy volcanic group in kamchatka. Journal of Geophysical Research: Solid Earth, 123(1), 564{582.

Steinmann, R., Seydoux, L., Beaucé, E., & Campillo, M. (2022). Hierarchical exploration of continuous seismograms with unsupervised learning. Journal of Geophysical Research: Solid Earth, 127(1), e2021JB022455.

Steinmann, R., Seydoux, L., & Campillo, M. (2022). Ai-based unmixing of medium and source signatures from seismograms: Ground freezing patterns. Geophysical Research Letters, 49(15), e2022GL098854.

Thelen, W., West, M., & Senyukov, S. (2010). Seismic characterization of the fall 2007 eruptive sequence at bezymianny volcano, russia. Journal of Volcanology and Geothermal Research, 194(4), 201-213. doi: https://doi.org/10.1016/j.jvolgeores.2010.05.010

Titos, M., Bueno, A., Garca, L., Bentez, M. C., & Ibañez, J. (2018). Detection and classi cation of continuous volcano-seismic signals with recurrent neural networks. IEEE Transactions on Geoscience and Remote Sensing, 57(4), 1936{1948.

Unglert, K., & Jellinek, A. M. (2015). Volcanic tremor and frequency gliding during dike intrusions at kilauea|a tale of three eruptions. Journal of Geophysical Research: Solid Earth, 120(2), 1142-1158. doi: 10.1002/2014JB011596

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. Journal of machine learning research, 9(11).

Wilding, J. D., Zhu, W., Ross, Z. E., & Jackson, J. M. (2022). The magmatic web beneath hawai`i. Science, eade5755.

Yogodzinski, G., Lees, J., Churikova, T., Dorendorf, F., Wöerner, G., & Volynets, O. (2001). Geochemical evidence for the melting of subducting oceanic lithosphere at plate edges. Nature, 409(6819), 500-504. doi: 10.1038/35054039

Zali, Z., Mousavi, S. M., Ohrnberger, M., Eibl, E., & Cotton, F. (2023). Tremor clustering reveals precursors and evolution of the 2021 geldingadalir eruption. Research Square. doi: 10.21203/rs.3.rs-2716246/v1