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

Rene Steinmann[1], Leonard 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

October 27, 2023

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

René Steinmann[1,2], Léonard Seydoux[3], Cyril Journeau[2,4], Nikolai M. Shapiro[2], Michel Campillo[2]

[1]GFZ German Research Centre for Geosciences, Potsdam, Germany
[2]ISTerre, Université Grenoble Alpes, CNRS, Université Savoie Mont Blanc, IRD, Université Gustave Eiffel, , Grenoble, France
[3]Université Paris-Cité, Institut de Physique du Globe de Paris, CNRS, Paris, France
[4]Department of Earth Sciences, University of Oregon, Eugene, OR, USA

Key Points:

- With machine learning, we analyze one year long seismic time series at individual stations at Klyuchevskoy volcano
- Continuous evolution of the signal characteristics over time reflects dynamic changes occurring in the volcano plumbing system
- Different episodes of volcanic activity are well distinguished on UMAP-based seismogram atlases

Corresponding author: René Steinmann, rene.steinmann@gfz-potsdam.de

Abstract

Seismic time series provide crucial information for monitoring the state of a volcano with discrete event catalogs describing impulsive seismic activity and manual 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 seismograms 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. Through additional data, we can relate areas in the atlas to various states of the volcano such as shallow and deep activity, deep reactivation, weak signals during quiet times, and eruptive activity. The time evolution represented by the atlas depicts continuous and sudden transitions between different states of the active Klyuchevskoy volcano. The atlases open new avenues to perceive large seismic time series visually and to connect the signal changes to physical processes.

Plain Language Summary

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.

1 Introduction

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). Volcano seismologists have classified seismic signals with volcanic origin into distinct classes based on the source mechanism and signal characteristics. These classes include volcano-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. In the recent years, supervised learning strategies have been used extensively for event detection and classification tasks (e.g., Curilem et al., 2009; Hibert et al., 2017; Maggi et al., 2017; Malfante et al., 2018; Titos et al., 2018; Lara et al., 2020; Falcin et al., 2021).

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; Beroza & Ide, 2011). 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, 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). 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.

Due to advancements in information processing and the introduction of continuous recordings, some studies have shown that continuous seismograms contain relevant information beyond classical event catalogs. The entropy of the seismic noise and the seismic background level seem to vary significantly prior to eruptions (Glynn & Konstantinou, 2016; Rey-Devesa et al., 2023; Ichihara et al., 2023). Machine learning strategies, in particular unsupervised learning, provide a promising approach for automatically analyzing large amounts of continuous seismograms and inferring such patterns without requiring predefined labels (e.g. Köhler et al., 2010; Holtzman et al., 2018; Ren et al., 2020; Seydoux 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)). Many volcano observatories use a single reference station situated in close proximity to the volcano, measuring in real-time the amplitude of raw continuous seismograms (RSAM) as an estimation of volcanic activity (e.g. Endo & Murray, 1991). Given the rich and various appearance of seismicity in a volcanic environment, we test the hypothesis that the data-driven analysis of continuous seismograms offers new and different insights into the inner workings of a volcano than what current discrete event catalogs, the continuous seismic amplitude, 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 the signal-altering effect of surface freezing and thawing onto a single independent component. 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 interesting information related to brain activity. 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). Besides the interpretation of independent components, 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.

2 Klyuchevskoy Volcano Group and its Seismic Activity

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 by fluids released from the thick, highly hydrated Hawaiian-Emperor crust subducted beneath this corner(Dorendorf et al., 2000). There is also evidence that the distinct volcanoes of KVG interact with each other on various time scales, affecting their steady state regimes and magma output (Coppola et al., 2021).

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).

3 From Continuous Seismograms to Data-Driven Features and Seismogram Atlases

3.1 An ideal data representation with a scattering network

In the following, we want to outline how we create data-driven features with ICA and seismogram atlases with UMAP from continuous seismograms. For two main reasons, waveform data is a poor data representation to perform those tasks. Firstly, waveform data is sensitive to translation, meaning it encodes information about the signal's position in time. Secondly, waveform data is sensitive to small signal deformations, meaning that small deformations cause large variations in the data representation. To address these challenges, it is essential to find a representation that is both translation-invariant and stable to small deformations. While the amplitude spectrum of the Fourier transform is translation-invariant, it lacks stability to signal deformations, particularly at higher frequencies Bruna and Mallat (2013). The wavelet transform replaces the non-localized sine waves of the Fourier transform with localized waveforms, offering stability to deformations. However, it is not inherently translation-invariant. By adding non-linear averaging operators to the wavelet transform, we achieve both translation invariance and stability and create an architecture similar to Convolutional Neural Networks (CNNs). However, it's worth noting that the non-linear averaging operator can potentially remove important signal information. To mitigate this information loss, we repeat the wavelet transform in combination with the averaging non-linear operators and extract information at each layer. This iterative approach allows us to recover most of the lost information, resulting in a representation that is both translation-invariant and stable to small deformations. This architecture has been mainly introduced in Bruna and Mallat (2013); Andén and Mallat (2014) and has been recently applied to continuous seismograms, capturing intriguing patterns (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). Moreover, scattering coefficients performed better for classification and data exploration tasks in comparison to spectral coefficients from the Fourier transform (Andén & Mallat, 2014; Steinmann, Seydoux, Beaucé, & Campillo, 2022). We apply a scattering network with a sliding window to the continuous three-component seismograms to retrieve the scattering coefficients (Figure 2). Each layer produces an output and the convolutional filters, classically learned in the case of CNNs, are restricted to a set of predefined wavelets. 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

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

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

$$S_1x(t, \lambda) = (U_\lambda \star \phi)(t) = (|x \star \psi_\lambda| \star \phi)(t), \tag{3}$$

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

$$S_2x(t, \lambda_1, \lambda_2) = (||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi)(t). \tag{4}$$

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 Gabor wavelets as initially proposed in Andén and Mallat (2014). The Gabor 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 Gabor wavelet, a defines the exponential drop-off of the waveform. We define a as a function of the wavelet's bandwidth d and its center frequency f, which in turn depends on the Nyquist frequency f_N of the signal $x(t)$ and the dilation factor λ

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

In this work, we design a two-layered scattering network with a sliding window of 20 min and an overlap of 10 min, resulting in a temporal resolution of 10 min of the scattering coefficient matrix. 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. The second layer wavelets start at much lower frequencies since they gather information about the modulation and shape of the signal. 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. With three-component seismograms, this results in a scattering coefficient matrix of 480 dimensions.

3.2 Pooling as a low pass filter in the scattering network

We use a pooling operation with a one-dimensional kernel as the low-pass filter $\phi(t)$ retrieving the scattering coefficients from the scalogram at each layer. 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) , ensuring a stable and translation-invariant representation for the scalograms. There are many different types of pooling operations, filtering different types of information and preserving different signal characteristics. In Seydoux et al. (2020) , the authors applied the scattering network with an average pooling and other possibilities are maximum pooling or median pooling, taking the maximum or median value of each scale in the scalogram. As an example case, we analyze the scattering coefficient retrieved with maximum and median pooling for a 20 min seismogram recorded at station SV13 (Figure 1). The dominant signal in this 20 min seismogram is a broadband transient event arriving after 800 s and lasting for 100 s. Moreover, there are also persistent harmonic signals , possibly of volcanic origin, around 0.8 and 2 Hz with a lower amplitude than the transient event. Besides the broadband transient and the harmonics, we can identify changing amplitudes at frequencies around 10 Hz. This example shows the variety of signals of a single seismogram and we must acknowledge that any representation without time information - such as the Fourier spectrum or scattering coefficients - will simplify the data. The information retrieved by the scattering coefficients depends largely on the settings of the scattering network: number of wavelets, frequency range of wavelets and pooling operation. Figure 1c, d and e show the median and maximum pooled scattering coefficients together with the Fourier spectrum. The first-order maximum pooled coefficients resemble a smoothed Fourier spectrum (Figure 1c). The first-order median pooled coefficients are lower in amplitude and contain different local maxima and minima. They show larger amplitude for the two harmonics and lower amplitudes in between. The transient event with large amplitudes between 0.2 and 10 Hz seems to have no influence on the median pooled coefficients. In contrast, the maximum pooled coefficients have an amplitude distribution that matches much better the transient event. The type of pooling operation, which transforms the scalogram into scattering coefficients, filters the data and stores different types of information. Median pooled coefficients contain the information of the background wavefield and ignore any short lived transients in the seismogram. Maximum pooled coefficients are sensitive to any type of short-lived transient in the seismogram which mask the background wavefield. Note also that maximum pooling would save the information of two transient events, if they appear in different frequency ranges. Thus, it could be a representation of a mixture of large amplitude events with different frequency content. Both pooling operations are valid, however, we need to acknowledge that both representations are biased and simplify the nature of the seismic data, an important fact to consider for exploratory data analysis. In this work, we will consider median pooling to mitigate the effects of impulsive short-term signals

3.3 Feature Extraction with Independent Component Analysis

The scattering coefficients are collected in a data matrix $\mathbf{X}$ in such a way that the rows contain the scattering coefficient time series and the columns contain all scattering coefficients for one sliding time window (Figure 2). We refer to the time series of the independent sources as features for the following text. The aim of ICA is the separation of multivariate signals into independent, non-Gaussian source signals, which can be formalized in the following way

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \tag{7}$$

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

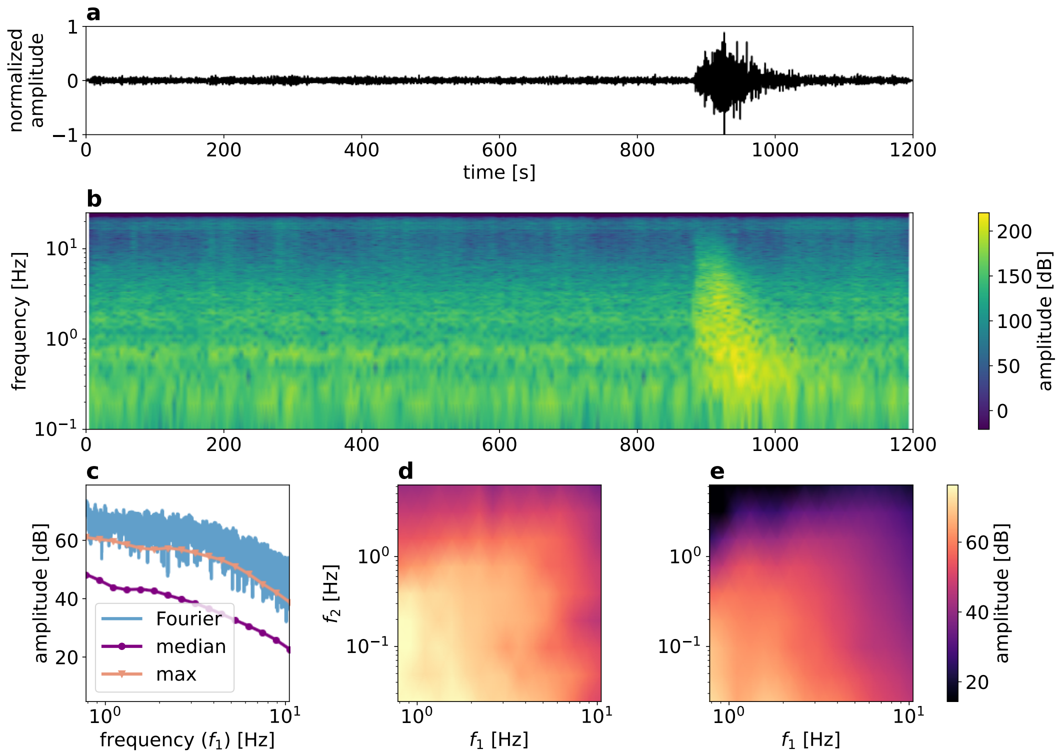

Figure 1. Comparison between Fourier spectrum and scattering coefficients of a seismic signal. **(a)** shows an example seismogram with normalized amplitude in time domain. **(b)** shows its corresponding Fourier spectrogram. **(c)** shows the Fourier amplitude spectrum and the first order median and 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.

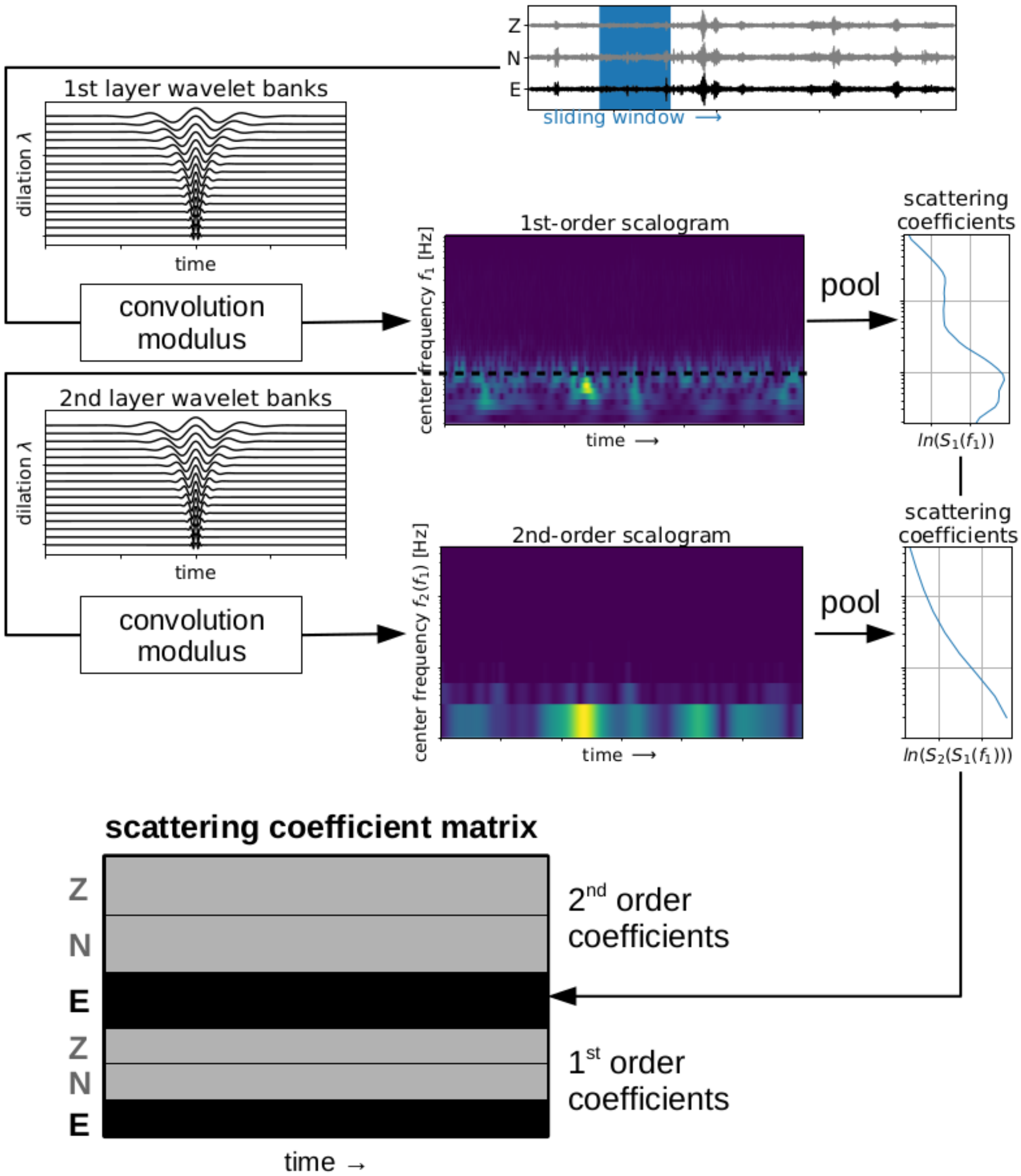

Figure 2. Detailed view of a two-layered scattering network applied to continuous three-component seismograms with a sliding window. The dashed line in the 1st-order scalogram indicates the data row which is convolved with the second-layer wavelet banks. The final scattering coefficient matrix contains the first- and second-order scattering coefficients from the three-component seismogram.

data in $\mathbf{X}$

$$\mathbf{S} = \mathbf{W}\mathbf{X}. \tag{8}$$

ICA solves this equation by maximizing the statistical independence of the sources in $\mathbf{S}$. 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 $\mathbf{S}$ and $\mathbf{A}$ 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 $\mathbf{X}$ since it constrains the unmixing matrix to be orthogonal and therefore the number of free parameters reduces (Hyvärinen & Oja, 2000).

3.4 Seismogram Atlases with UMAP

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 the t-distributed Stochastic Neighbor Embedding (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 (McInnes et al., 2018; 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), the mouse organogenesis cell atlas (Cao 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.

4 The Data: continuous seismograms, catalogs and lava discharge rate

In this work, we apply exploratory data analysis to the continuous seismograms of a joint Russian-German-French temporary seismic experiment named KISS (Klyuchevskoy Investigation – Seismic Structure of an Extraordinary Volcanic System; Shapiro, Sens-Schönfelder, et al., 2017), including short period and broadband sensors and covering the time period between August 2015 and July 2016. We focus on continuous three-component seismograms recorded by six individual broadband stations (Figure 3). The seismograms are demeaned, detrended, and down-sampled to a sampling rate of 25 Hz. Additional data such as a tremor catalog from Journeau et al. (2022) and the time averaged lava discharge rate (TADR) time series from Coppola et al. (2021) will support the exploratory data analysis, connecting signal patterns to known state changes of the volcanic system. Due to the remoteness of the KVG, seismic and satellite data are the only available data sources.

4.1 The scattering coefficients

In order to visualize the continuous seismograms and introduce the concept of scattering coefficients, we depict the first- and second-order coefficient time series of the east channel of station SV 13 in Figure 4a and b. We choose station SV13 since it is located directly above the cataloged seismic-volcanic activity (Figure 3). The first-order scattering coefficient time series resembles a spectrogram with spectral coefficients based on Fourier analysis (Figure 4a). The second-order scattering coefficient time series appear as multiples of the first-order scattering coefficients due to the application of the second-order wavelet transform to the first-order scalogram (Figure 4b).

4.2 The tremor catalog

The authors of Journeau et al. (2022) used the network's spectral covariance matrix (Seydoux et al., 2016) to detect and locate the coherent signals in a continuously moving time window and built a catalog of volcanic tremors including locations of their sources. We want to emphasize here that the catalog is a valuable source of information in validating the results of our work, however, it does not hold the ground truth, either. Figure 4c depicts the time-depth evolution of the catalog, showing periods of shallow and deep activity. Before December 2015, the tremors are mainly located in the deeper part of the Klyuchevskoy plumbing system with periods of increased tremor activity in August in September. The whole plumbing system becomes active after 4 December 2015 with shallow and deep tremor sources. After a short absence of tremors, they restart in March 2016 and last until the end of the cataloged time period. All scattering coefficient amplitudes are elevated during the presence of tremors (Figure 4a, b, and c).

4.3 Lava discharge rate time series and daily activity level

We also use the time-averaged lava discharge rate (TADR) time series of Coppola et al. (2021) , indicating the timing and strength of the eruption. The TADR data is estimated from infrared satellite data, assuming that the radiated energy of a lava body is linearly correlated to the bulk erupted volume. In the first half of April 2016, an eruption unfolded at the Klyuchevskoy volcano, indicated by TADR values above $0.1\,\mathrm{m^3\,s^{-1}}$ and lasted throughout the remaining recording time of the KISS experiment (Figure 4d). However, the exact starting time is not known. The scattering coefficients depict the largest amplitudes during the first half of April 2016 and show elevated amplitudes for the remaining recording time. Moreover, the Kamchatka Branch of the Russian Geophysical Survey (KBGS) determines the daily activity level for the Klyuchevskoy volcano based on detected seismic activity combined with visual and satellite observations, when available. The orange color in Figure 4d corresponds to an ongoing eruption.

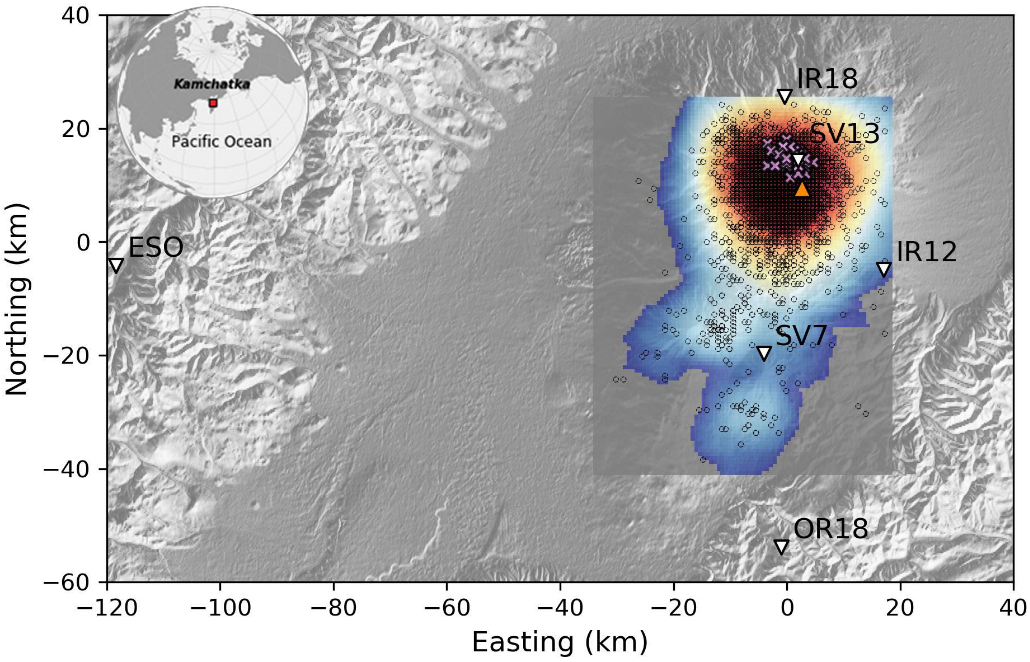

Figure 3. Map of Klyuchevskoy Volcano Group (KVG) with the seismic stations (SV13, IR18, IR12, SV7, OR18, and ESO) considered in this study, shown with white triangles. The orange triangle shows the location of the Klyuchevskoy volcano. Averaged spatial density of the tremor source location according 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.

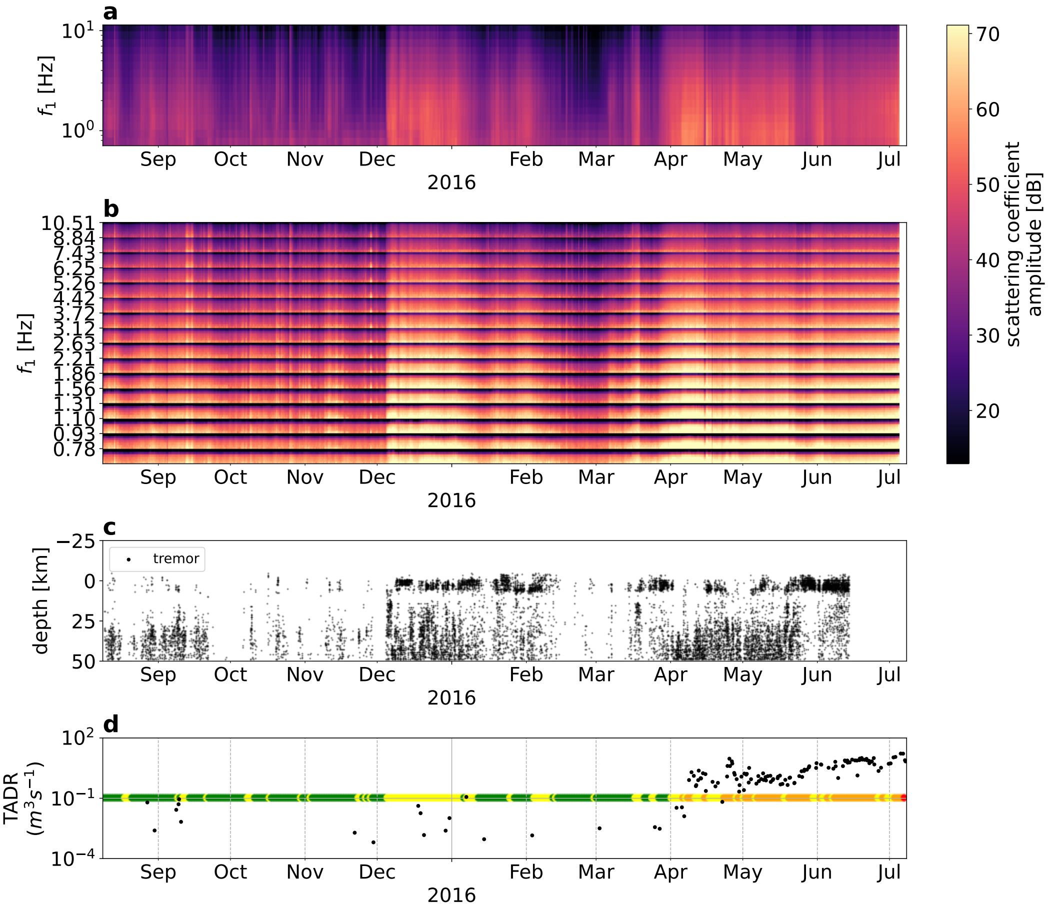

Figure 4. Time series of SV13 scattering coefficients and other data sources (a) Time series of first-order scattering coefficients of the east channel of SV13. **(b)** Time series of second-order scattering coefficients of the east channel of SV13. The y-axis represents the center frequency of the first-order wavelets f_1 and the center frequency of the second-order wavelets f_2 lies between two f_1 values (details in section 3.1). However, due to reasons of clarity, we do not label them on the y axis. **(c)** Localized tremors from Journeau et al. (2022) as a function of calendar time and depth. **(d)** TADR time series from Coppola et al. (2021) as black dots and the daily activity level from the KBGS as color-coded dots. Green corresponds to low volcanic activity, yellow corresponds to medium volcanic activity, orange corresponds to ongoing eruption, and red corresponds to an explosion on 7 July 2016.

5 Exploratory data analysis with Independent component analysis (ICA) and seismogram atlases

5.1 Hierarchical ICA with increasing number of components

The complete scattering coefficient matrix of station SV13, including the east, north and vertical channels, is the input for the ICA model. We apply ICA models with four (M_4), 12 (M_{12}), and 50 (M_{50}) independent sources to explore the impact of dimensionality. 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 99.6 %. Figure 5 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 scattering coefficient matrix. 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 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 5b and d), we can find similar features between M_4 and M_{12} such as feature 2 in both models (Figure 5b and c). M_{50} is very different from the other two models, since its features appear more sparse, i.e. they are mostly centered around zero except for a short duration. Moreover, it is striking that if one feature shows large amplitudes in a 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 a color-saturated diagonal line in the time-feature space (Figure 5d). It appears that each data point in this 50-dimensional feature space is located at the center of 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. The M_{50} model indicates that the seismic time series witness an ever-changing seismic wavefield with new signal characteristics throughout the recording time. The comparison with the tremor count per day and the lava discharge rate in Figure 5a reveals interesting relation between the features and known state changes. For instance, feature 2 of M_4 and M_{12} models shows positive values mainly during the tremor activity in August and September 2015, suggesting signal characteristics different from the later tremor periods. The reactivation of the plumbing system at 4 December 2015 is characterized by a rapid succession of features in the M_{50} model, indicating rapid pattern changes in the seismograms. In contrast, the eruption period starting in April 2016 is characterized by M_{50} features showing constant amplitudes for longer time periods (feature 40 to 50 in Figure 5d). This suggests more stable signal patterns lasting for longer time periods compared to the reactivation of the plumbing system in early December. Overall, we find correlations between feature changes – indicating signal pattern changes in the seismograms – and state changes of the Klyuchevskoy volcano. Moreover, we see that the number of components has an effect on the retrieved features, showing different time histories. 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.

5.2 Interpreting M_4 Features with The Mixing Matrix

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 $\mathbf{X}$ is estimated as the sum of rank-1 matrices, resulting from the outer product of a feature (rows in $\mathbf{S}$) with the corresponding columns in the mixing matrix $\mathbf{A}$. Hence, the columns of $\mathbf{A}$ reveal how each feature contributes to the estimation of $\mathbf{X}$. The visualization of the columns of $\mathbf{A}$ and its outer product with the corresponding feature can help to understand better the underlying signal characteristic of each feature. In (Steinmann, Seydoux, & Campillo, 2022), we used this method to reveal the changing signal patterns due to the freezing and

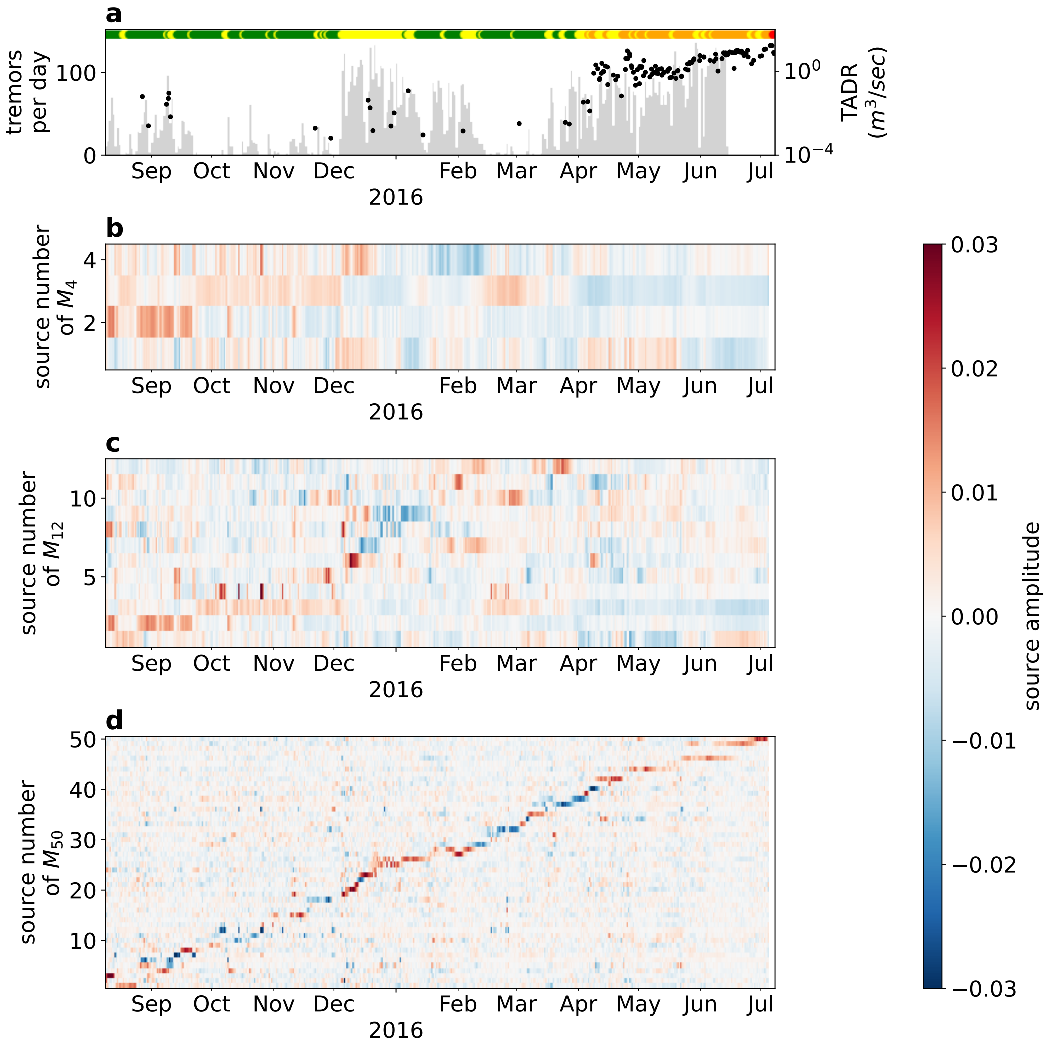

Figure 5. Hierarchical ICA for station SV13. (a) The grey histogram describes the daily number of localized tremors based on Journeau et al. (2022) and the colored circles indicate the daily activity level of the Klyuchevskoy volcano, where green represents low activity, yellow represents medium activity, orange represents an ongoing eruption, and red represents an explosion on 7 July 2016. The black dots correspond to the time-averaged lava discharge rate (TADR) from Coppola et al. (2021). By applying multiple ICA models to the SV13 scattering coefficient matrix, we retrieve time histories of the independent sources (features), describing signal patterns in the seismogram. **(b)** shows the features of the 4-component model M_4, **(c)** shows the features of the 12-component model M_{12}, **(d)** 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.

thawing of the near subsurface. In Figure 6, 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 5b. For example, feature 3 in Figure 5b shows a general correlation with the occurrence of tremors. 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 6). Figure 7b, c, and d show the reconstructed first-order scattering coefficients, resulting from the outer product of feature 3 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 (Figure 7b, c, and d). Note that both the mixing weights (source 3 in Figure 6) and the feature amplitudes during tremor-active periods (source 3 in Figure 5b) are negative, resulting in positive amplitudes of the reconstructed scattering coefficients due to the matrix multiplication (Figure 7b, c, and d). 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. The reconstruction shows that feature 3 of M_4, correlating with general tremor occurrence, relates to broadband amplitude changes. We do not need machine learning approaches to observe a correlation between a broadband amplitude increase and tremor occurrence, however, we want to show that the features represent meaningful patterns.

A more interesting example is feature 2, which correlates only with the tremor sequences in August and September (feature 2 in Figure 5b). The corresponding mixing weights show positive and negative amplitudes depending on the frequencies f_1 and f_2 and the channel (source 2 in Figure 6). 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 feature 2 (Figure 7e, f, and g). We see a clear anti-correlation for scattering coefficients below and above 1 Hz for the east channel (Figure 7e): 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 6). Weights with the same sign indicate 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 7e, f and g suggest that the deep tremor activity in August and September is different from the other tremor episodes mainly due to different patterns at the east channel around 1 Hz.

The reconstruction can indicate the underlying pattern changes in the seismogram and seems to be useful for ICA models with a low number of components. However, this becomes unfeasible for ICA models with a large number of components such as for M_{50}.

5.3 Comparing M_{50} models of multiple seismic stations

The M_{50} model of station SV13 (Figure 5d) pictures a seismic time series with many pattern changes , indicating an ever-changing seismic wavefield. 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 verify this assumption. The considered stations – named SV13, IR12, IR18, SV7, OR18 and ESO in Figure 3 – are located between 5 and 122 km away from the active volcano. Figure 8 shows the corresponding M_{50} models, revealing a diagonal line in the time feature space degrading with increasing distance to the volcano. This confirms

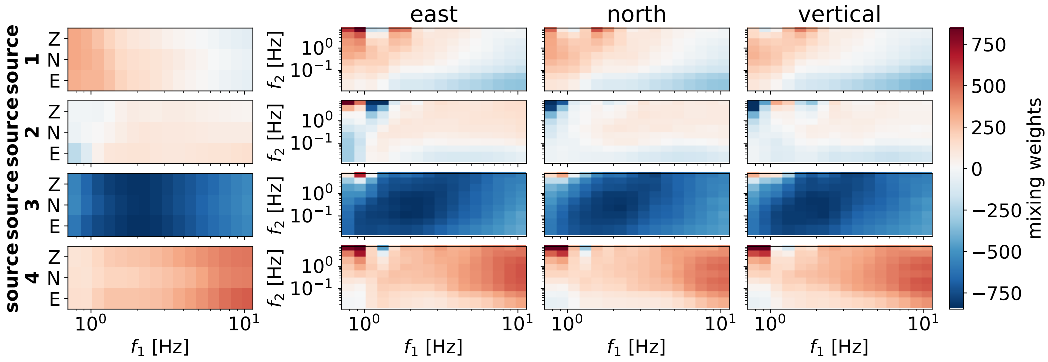

Figure 6. **Mixing weights** for the M_4 model for its four sources at station SV13. The matrix multiplication of the mixing weights with the M_4 features in Figure 5b estimates the scattering coefficient matrix (see equation 7). For visualization purposes, we reshaped the mixing matrix to display the 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.

our assumption that the large amount of pattern changes, represented by the diagonal line in the time-feature space, are only recorded in direct proximity to the active volcano. The activation of the whole plumbing system in December and the eruption period are characterized by dominant features even for stations further away such as SV7 and OR18. However, the large number of features makes it cumbersome to understand the pattern changes and relate it to physical processes in the volcano. Manifold learning techniques such as UMAP might help to overcome this issue by capturing more information on fewer dimensions.

5.4 Seismogram Atlases of individual stations

We obtain seismogram atlases for the 6 station with UMAP and color-code the data points with their corresponding calendar time, highlighting the temporal evolution of the seismic time series (Figure 9). 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 10). The atlases depict different cluster shapes and distances with regard to the hyperparameters, emphasizing the trade-off between local and global structures in the atlas. However, all the examples confirm the smooth time gradient and little to no overlap of different time periods. We opted for a minimum distance of 0.5 and 50 neighboring points, a decision primarily guided by visual assessment, much like the process commonly used in many other dimensionality reduction techniques. The chosen parameters seem to be a good choice for preserving local and global structures without resulting in too many disjoint clusters, as observed with 10 neighboring points. The seismogram atlases of SV13, IR18, and IR12 picture a variety of shapes with many linear and curved structures, where data points with different colors hardly overlap (Figure 9a-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 atlas. Therefore, neighboring data points in the atlas are likely neighboring data points 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: there are less linear or curved structures and different colors overlap more often (Figure 9e-f). The two-dimensional seismogram atlases confirm our interpretation of the 50-dimensional ICA models: close to

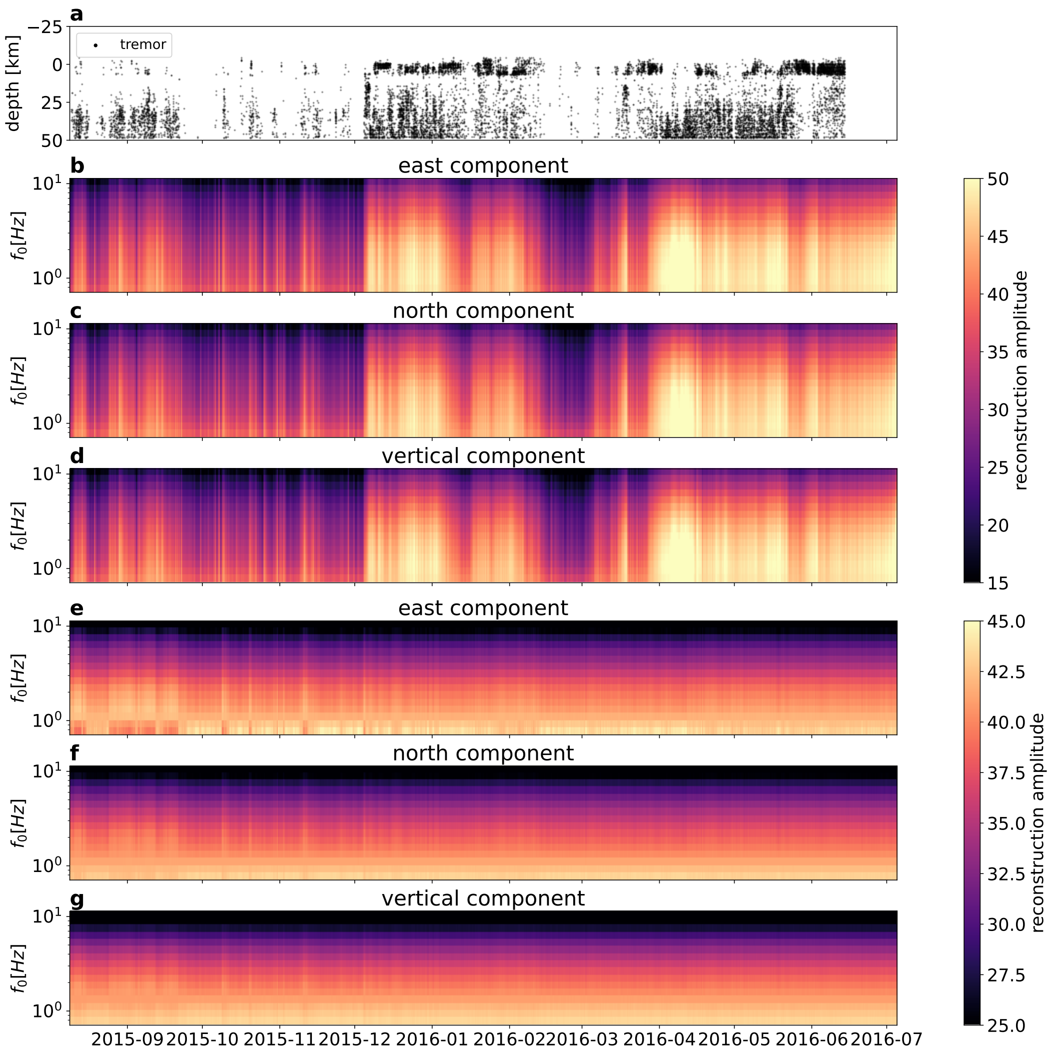

Figure 7. Reconstructed first-order scattering coefficients based on the outer product of a M_4 feature with its mixing weights for station SV13. Subfigure **a** shows the localized tremors, subfigures **b**, **c** and **d** show the reconstruction of the first-order coefficients of the east, north and vertical channel, based on feature 3 (Figure 5b). Subfigures **e**, **f** and **g** show the reconstruction of the first-order coefficients of the east, north and vertical channel, based on feature 2 (Figure 5b).

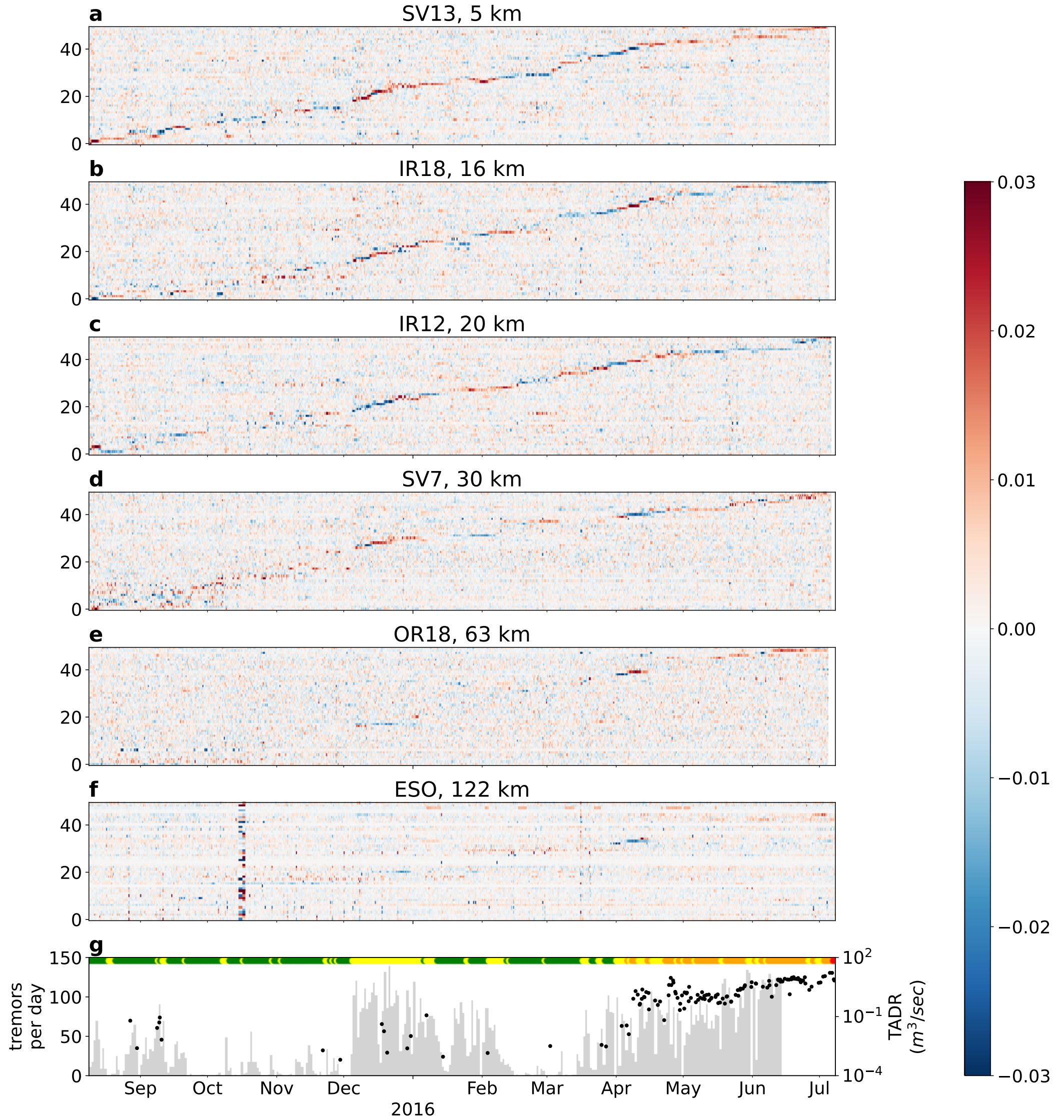

Figure 8. By applying a 50-component ICA model to the scattering coefficient matrices, we retrieve M_{50} features for the continuous seismograms recorded at station SV13 **(a)**, IR18 **(b)**, IR12 **(c)**, SV7 **(d)**, OR18 **(e)** and ESO **(f)**. Each feature represents a signal pattern, which dominates when the color is saturated. The results are ordered according to the distance to the active Klyuchevskoy volcano mentioned in the title of the subfigures. Subfigure **(g)** shows the number of localized tremors per day as a grey histogram, the time-averaged lava discharge rate (TADR) from Coppola et al. (2021) as black dots and the daily activity level from the KBGS as color-coded dots. Green corresponds to low volcanic activity, yellow corresponds to medium volcanic activity, orange corresponds to ongoing eruption, and red corresponds to an explosion on 7 July 2016.

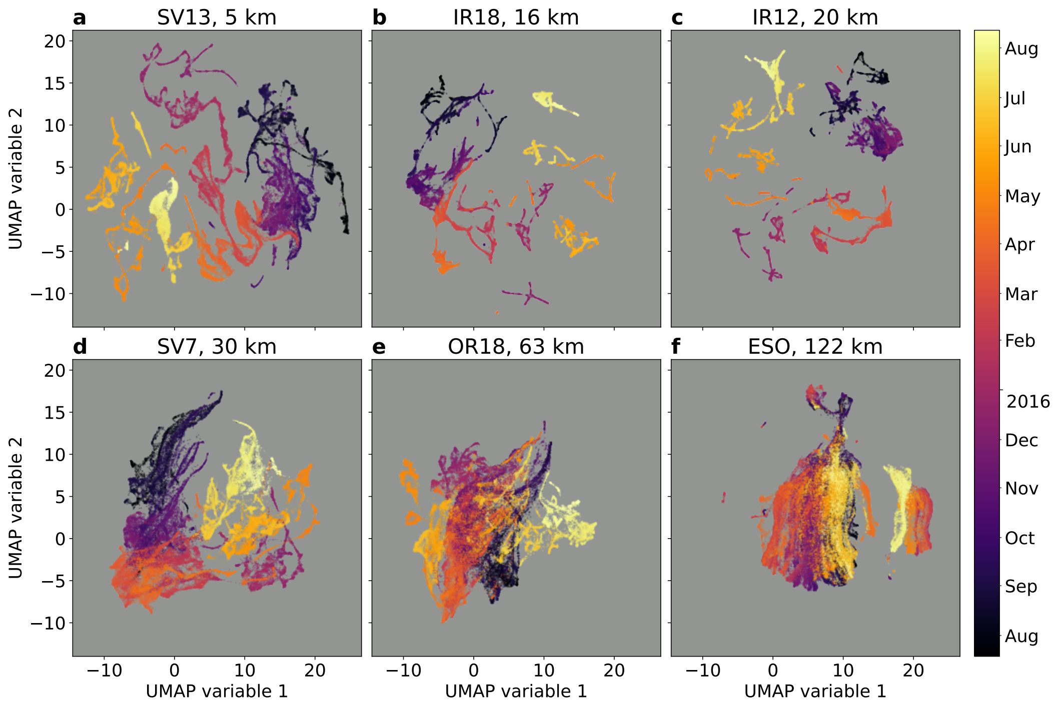

Figure 9. By applying the manifold learning technique UMAP to the scattering coefficient matrices, we obtain individual seismogram atlases for the continuous seismograms recorded at station SV13 **(a)**, IR18 **(b)**, IR12 **(c)**, SV7 **(d)**, OR18 **(e)** and ESO **(f)**. Each data point in the two-dimensional representation corresponds to 20 min three-component seismograms and the color-code reflects the calendar starting time of the 20 min window. The distance in the atlas reflects the similarity of the three-component seismograms. The subfigures are ordered according to the distance to the active Klyuchevskoy volcano mentioned in their titles.

the active volcano, the seismograms witness many pattern changes, which are not visible by stations further away.

5.5 Connecting the seismogram atlas to known physical processes

In order to provide a meaningful interpretation, we color-code the data points in the atlas with other physical parameters such as the TADR and the depth of the located tremors. Since the TADR time series are sampled irregularly, we interpolate the TADR data linearly, matching a TADR value with a data point in the atlas. We do that for the station closest to the volcano - SV13 - and for the station furthest from the volcano - ESO (see Figure 11a and b). For station SV13, we can clearly identify the eruption activity (TADR $> 0.1\,\mathrm{m^3\,s^{-1}}$) in the eastern area with multiple linear and cluster structures. For station ESO, we can only identify areas of strong eruption activity (TADR $> 1\,\mathrm{m^3\,s^{-1}}$) with no linear or curved structures, which would indicate continuous pattern changes in the seismograms. Figure 11c and d show the atlases color-coded with deep and shallow tremors. For station SV13, we can identify areas with either dominating deep or shallow tremor activity (areas with mainly blue or orange color in Figure 11c) and we can identify areas with a mix of deep and shallow activity. Moreover, the tremor signals cover a large part of the atlas and only a small part does not correspond to tremor signals (grey data points). For station ESO, deep and shallow tremor activity is nowhere separated, indicating that ESO is too far to sense fine tremor signal variations.

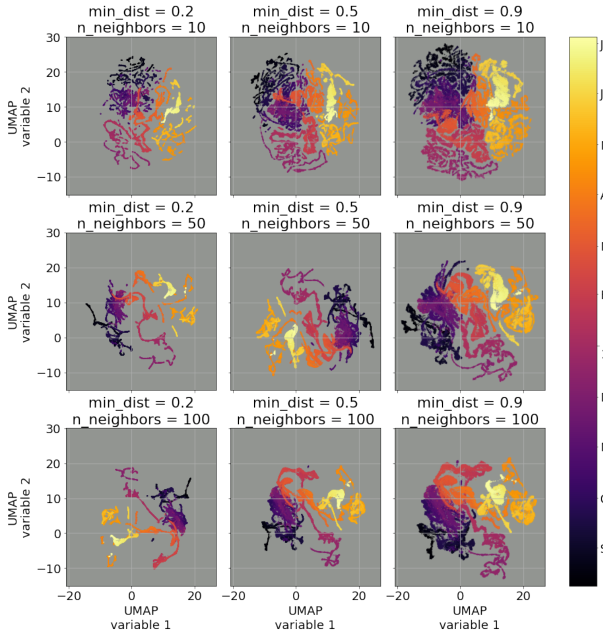

Figure 10. Seismogram atlases obtained with changing UMAP hyperparameters for the data recorded at SV13. Each data point in the two-dimensional representation corresponds to 20 min three-component seismograms and the color-code reflects the calendar starting time of the 20 min window. The results shown in Figure 9a correspond to min_dist = 0.5 and n_neighbors = 50.

6 Discussion

The seismogram atlases and the ICA features reveal a permanently evolving seismic wavefield with many pattern changes in the vicinity of the active Klyuchevskoy volcano. This suggests that the continuous seismograms from close stations contain relevant information about the dynamic processes occurring in the volcano plumbing system. We found particularly interesting the difference between the atlases obtained for stations located in the close vicinity of the active Klyuchevskoy volcano and those from distant stations. The latter look like "diffuse" clouds of points without distinctive structures. The former contain many well-defined "lineaments". We hypothesize that these "lineaments" in the seismogram atlases of nearby stations are associated with dynamic processes occurring within the volcano plumbing system. During such periods of "dynamic activity" the system evolution is characterized by certain "continuity" that is reflected in the "continuity of the seismogram atlas "lineaments".

In order to quantify this continuity, we use two parameters. We start with computing all vectors (steps) connecting two atlas points consecutive in time. Then, we compute the amplitudes of these steps. Small steps correspond to a more continuous atlas evolution. Therefore, we average the step amplitudes over N consecutive points and call this parameter as "seismogram atlas continuity". Then we evaluate if consecutive steps follow a preferential direction. For this, we compute an Euclidean distance between points separated by N steps and compare it with the sum of amplitudes of these N steps. If all steps are perfectly aligned in the same direction these two quantities are perfectly equal. If the directions of steps are random, the distance between the first and last points is much smaller than the sum of step amplitudes. We compute the ratio between these two quantities and call it "seismogram atlas gradient continuity". Its value maximizes at 1 for a perfectly straight atlas "lineament".

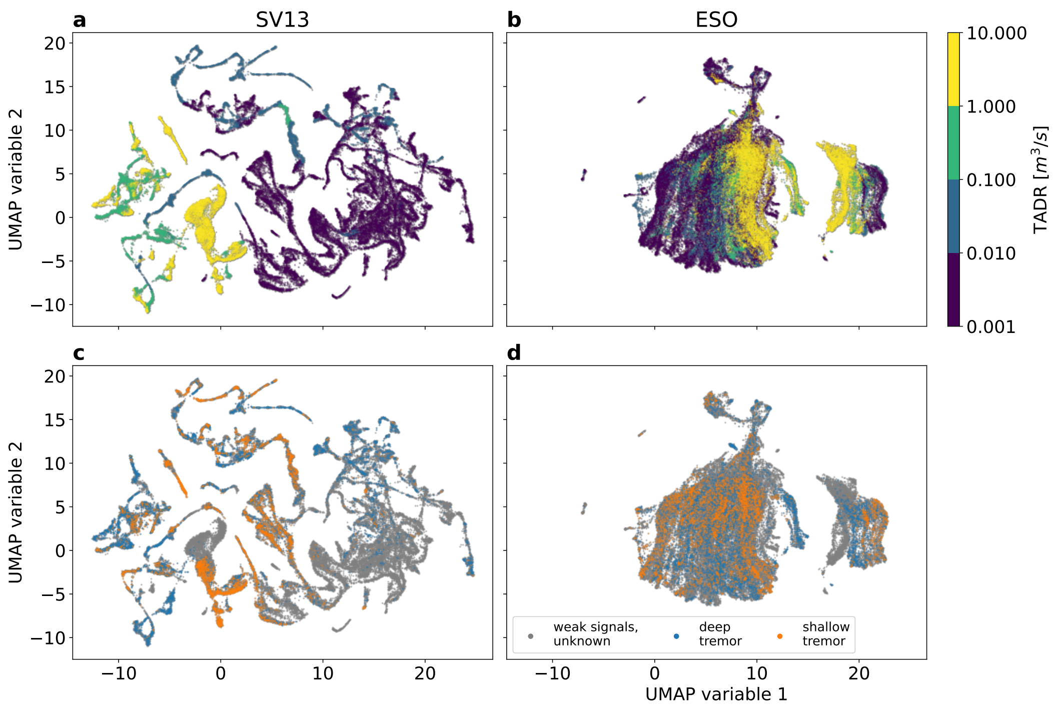

Figure 11. Seismogram atlases color-coded with interpolated time-averaged lava discharge rate (TADR) and localized tremors for two stations. Seismogram atlas with interpolated TADR for station SV13 in **(a)** and ESO in **(b)**, highlighting areas in the atlas with strong eruptive activity. Seismogram atlas with shallow and deep tremors for station SV13 in **c** and ESO in**d**, highlighting areas with rather shallow or deep volcanic activity. We color-coded a data point with tremor activity in case they matched in time. Deep and shallow tremors are separated at 10 km depth.

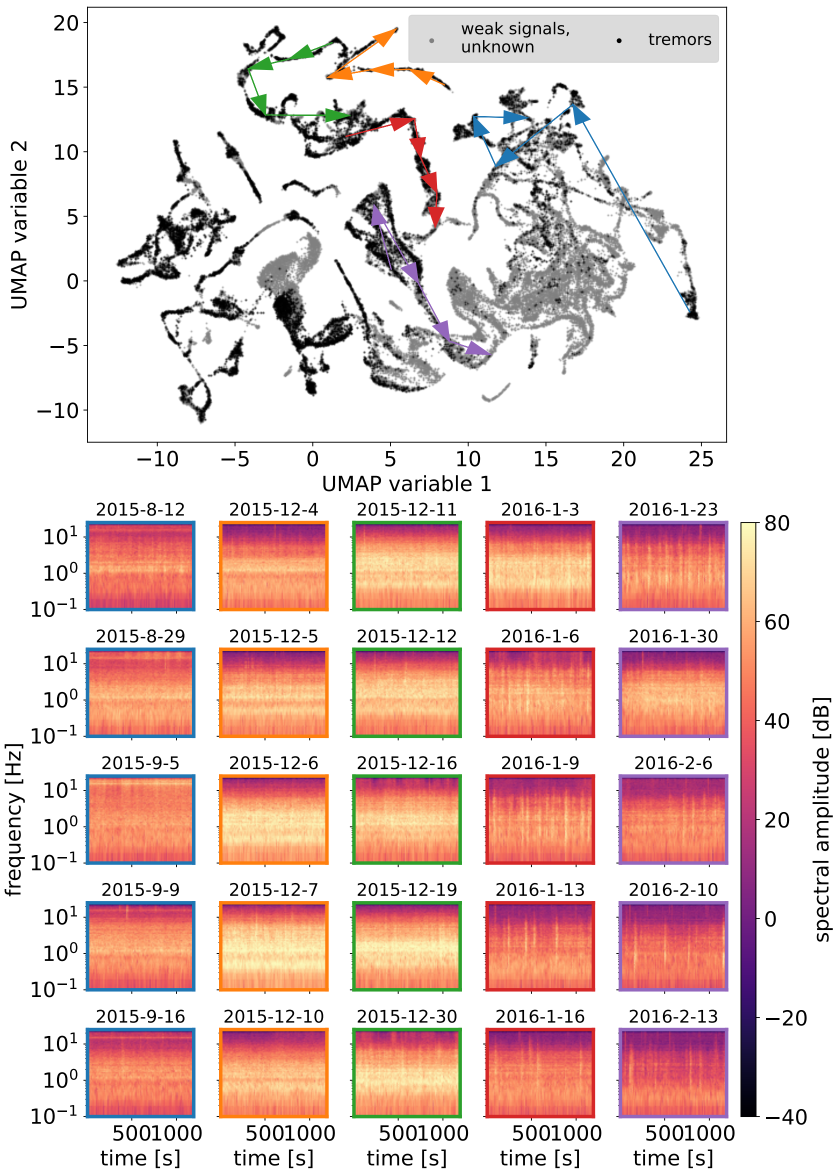

Figure 12. The seismogram atlas with example spectrograms of tremor signals from the period between August 2015 and February 2016. The upper image shows the SV13 seismogram atlas . Data points matching in time with localized tremors are black and data points with weak or unknown signals are grey. Each data point represents 20 min of three-component seismograms and for some data points (marked with the colored arrows) we visualized the east-component spectrograms with their calendar data on the top of the subfigure. 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.

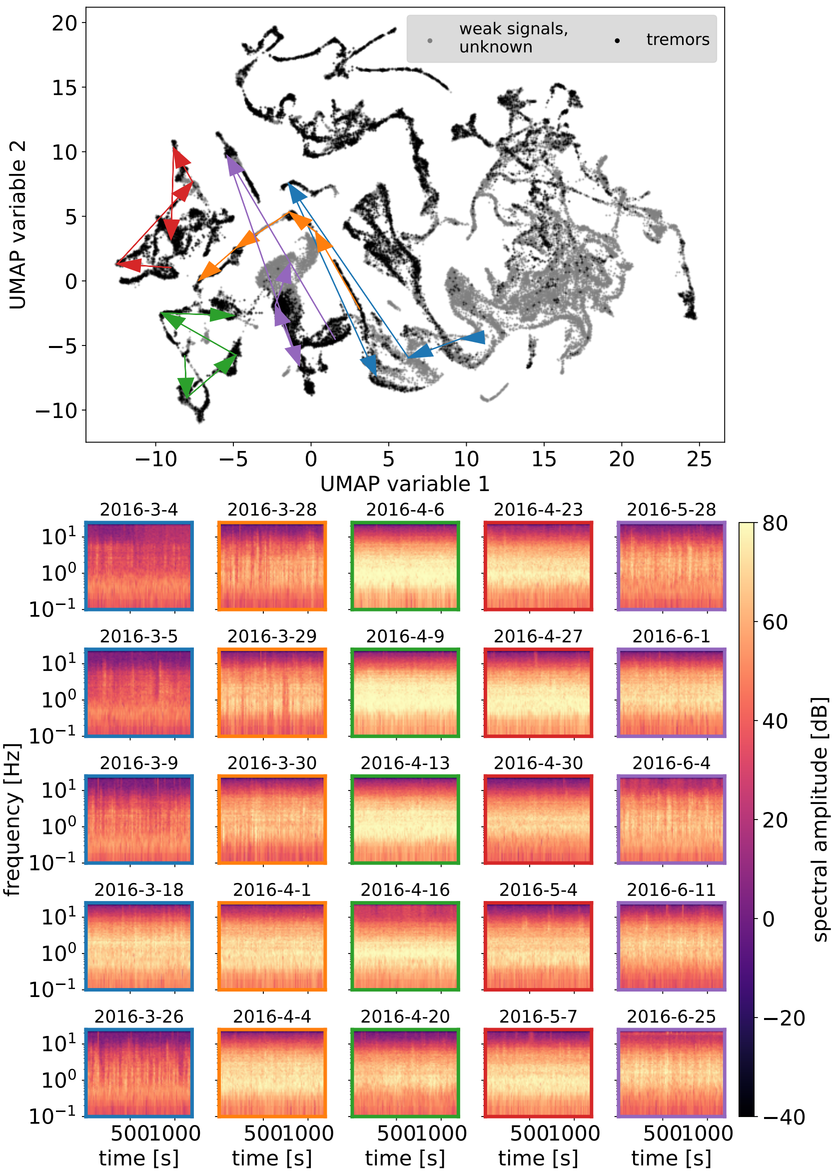

Figure 13. The seismogram atlas with example spectrograms of tremor signals from the period between February 2016 and July 2016. The upper image shows the SV13 seismogram atlas . Data points matching in time with localized tremors are black and data points with weak or unknown signals are grey. Each data point represents 20 min of three-component seismograms and for some data points (marked with the colored arrows) we visualized the east-component spectrograms with their calendar data on the top of the subfigure. 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.

6.1 Interpretation of the SV13 seismogram atlas

Our data analysis showed that station SV13, the closest station to the active Klyuchevskoy volcano, witnesses many pattern changes and we have seen correlations with other data sources such as the event catalogs and the TADR time series. In the following, we try to synthesize all data sources to identify meaningful areas within the SV13 seismogram atlas. Moreover, we visualize the spectrograms of some data points, showing characteristic examples of various atlas areas (Figure 12 and 13). In both Figures, we connect the data points in the atlas to known cataloged signals from Journeau et al. (2022). The example spectrograms focus on tremor active periods and we connect the different data points with arrows, highlighting the temporal evolution of the tremor signals.

The SV13 seismogram atlas shows an interesting data landscape of continuous and isolated linear and curved structures with temporal smooth gradients, correlating with the occurrence of tremors (Figure 11c). The time series of atlas continuity and atlas gradient continuity for the three stations in the vicinity of the active volcano are depicted in Figure 14a, b, and c. As a comparison, Figure 14d shows the daily count of shallow tremors in blue (above 10 km depth) and deep tremors in red (below 10 km depth), highlighting time periods of deep or shallow activity. Figure 15 provides an interpretation with the same labels for the SV13 seismogram atlas by connecting certain time periods to volcanic activity. The interpretations of both Figures are based on the temporal evolution depicted in Figure 9a, the catalog association with example spectrograms shown in Figure 12 and 13, the TADR and tremor depth association in Figure 11a and c, and the movie S1, highlighting the temporal evolution of the atlas. We also used information from the daily reports of the Kamchatka Branch of the Russian Geophysical Survey available at their web site (in Russian): http://www.emsd.ru/˜ssl/monitoring/main.htm and publications about the Klyuchevskoy activity in 2015-2016 (O. A. Girina et al., 2019; O. Girina et al., 2023).

From August to the end of September 2015, tremor activity occurs mainly at deeper depth (marked with a blue 1 in Figure 14d). The same time period corresponds to a connected point cloud with multiple linear and curved structures in the atlas (marked with a blue 1 in Figure 15). The spectrograms in that area picture narrow-banded continuous tremor signals centered around 1 Hz, which is different from the other tremor signals (blue framed spectrograms in Figure 12). This confirms our interpretation of the reconstructed scattering coefficients in Figure 7, indicating a signal difference around 1 Hz between the early deep tremor period and the tremors after the reactivation in December 2015. In the following we relate the pattern evolution of the continuous seismograms with distinct phases of the Klyuchevskoy volcano.

At the end of September 2015, the deep tremor activity decreases and only a few isolated days denote tremor activity. Since there is no strong seismic activity in this time period (marked with a grey 2 in Figure 14d), we can assume that it is mainly dominated by weak seismo-volcanic signals and ambient seismic noise. For the same time period, the atlas depicts a grey point cloud below the deep tremor activity (marked with a grey 2 in Figure 15). Interestingly, this time period is placed close to the deep tremor activity with overlapping structures. In fact, the grey linear structures reaching into the blue area are related to the deep tremor activity in October and November (see movie S1).

The reactivation of the whole plumbing system starts at 4 December 2015 and is characterized by a jump in the atlas from the seismic noise area to a linear structure in the northern area (marked by a blue 3 in Figure 15), suggesting a sudden pattern change in the seismograms. A high continuity value on all three stations reflects the temporal-spatial disconnection in the seismogram atlas (blue 3 in Figure 14a, b, and c). After that date the atlas depicts a complex trajectory where we can identify shallow and deep tremor phases (marked with 4, 5, and 6 in Figure 15 and Figure 14), indicating an ever-changing seismic wavefield with continuous pattern changes. In the same time period, the spectrograms in Figure 12 indicate a transition from pure continuous signals to continuous signals with impulses. The tremor catalog in Figure 14d indicates a deceasing tremor activity in February 2016, marked by neon-green 6 and grey 7.

From the deep reactivation in December 2015 until the quiet period in February 2016, the atlas depicts a continuous trajectory, ending close to the area of low seismo-volcanic activity in October/November 2015 (Figure 15). The purple-framed spectrograms of Figure 12 confirm the gradual decrease of continuous signals, which are characteristic of volcanic tremors. It is interesting to note that the linear trajectory in the atlas continues, even after the catalog does not show any tremor detections (see movie S1). This suggests continuing and not cataloged weak tremor activity. Since station SV13 is located close to the center of tremor activity, it is very likely that this station records weak tremor activity, which does not generate a coherent signal across an array of stations. It is also interesting that the data points of the quiet period in February/March 2016 (marked with a grey 7) do barely mix with the data points of the October/November 2015 period (marked with a grey 2), indicating a different type of ambient seismic noise or different weak seismo-volcanic signals. We can exclude the oceanic mircoseisms or large-scale meteorological phenomena for this behavior, since the center frequencies of the first-order wavelets do not cover frequencies below 0.78 Hz. Other studies have shown that the signal properties of the ambient seismic noise can change due to volcanic activity (Glynn & Konstantinou, 2016; Ichihara et al., 2023).

Towards the end of the quiet period, the data points move away from the noise area towards an area where we have mainly pre-eruptive shallow tremors (yellow 9 in Figure 15 and movie S1), indicating slowly emerging and not cataloged tremor signals. This slow transition is interrupted by a deep reactivation starting at 17 March 2016, depicted by a jump in the atlas to the curved purple structure marked with an 8. In fact, the deep reactivation is characterized by a spatial-temporal disconnection in the atlas for the three stations close to the volcano, suggesting a sudden pattern change on the three stations similar to the 4 December 2015 reactivation (Figure 14). The deep reactivation is characterized by pure broadband continuous signals compared to the times before and after (blue-framed spectrograms in Figure 13). After the sudden pattern changes, the SV13 atlas indicates a more continuous pattern change until the start of the eruption (yellow 9 and magenta 10 in Figure 15). The spectrograms of the same time period show a continuous transition from repeating impulsive signals to more continuous tremor signals (orange-framed spectrograms in Figure 13). After the eruption started on April 4, 2016 (O. A. Girina et al., 2019; O. Girina et al., 2023), the spectrograms show mainly continuous tremor signals (green- and red-framed spectrograms in Figure 13) The April/May 2016 co-eruptive tremor period is located on the far left in the atlas and a neighboring linear structure characterizes a period of shallow tremor activity with an increase in the lava discharge rate occurring at the end of May 2016. After this event, the co-eruptive tremor signals build a point cloud with no linear structures (light-pink 14 in Figure 15), indicating no continuous pattern evolution for a longer time. This behavior is very different from the previous tremor-dominated periods, characterized by mainly linear or curved structures and continuous or sudden pattern changes. During that time in June 2016, most tremor sources are at a shallow depth (Figure 14d). The significant increase of the atlas continuity and gradient continuity seen at all three station at the beginning of July 2016 (Figure 14) marks reactivation prior to the strong explosion that took place on July 7, 2016 at the crater of the active Klyuchevskoy volcano. On this date, an ash column reached 10 km in altitude according to KBGS and the ash cloud traveled more than 400 km (O. A. Girina et al., 2019). The atlas denotes a jump back to the eruption area around 4 July 2016, indicating a reactivation prior to the explosion (red-yellow 15 in Figure 15).

The seismogram atlas of station SV13 shows that the seismograms follow a complex signal evolution with many pattern changes due to volcanic activity and a changing plumbing system. The atlas shows no overlap of the various tremor periods, indicating changing tremor signal patterns throughout the recording time. We see both smooth and sudden transitions between different activities, reflecting smooth and sudden regime transitions of the volcanic system. The interpretation shows that UMAP preserves global and local structures. Globally, we can identify seismograms with continuous signal characteristics in the outer circle of the atlas and seismograms with impulsive characteristics in an area inside the circle. Locally, we can identify certain periods of volcanic activity such as shallow or deep tremor activity, and see continuous or sudden changes between these different periods. In some areas we have continuous transitions between continuous and impulsive signals, indicating seismograms containing both types of signal characteristics. Journeau et al. (2022) observed a similar signal separation and transition within the variables space obtained from the network's covariance matrix. In our case, the second-order scattering coefficients contain information about impulsive and continuous signal characteristics and, therefore, the separation of seismograms according to these characteristics is reasonable. Our findings show that the majority of the recording time is dominated by tremor signals. This agrees with the findings of Makus et al. (2023) where their reference correlation function for the same dataset is dominated by tremor activity.

7 Conclusion

With data-driven features and seismogram atlases, we have analyzed the signal content of continuous seismograms recorded during the KISS experiment close to the active Klyuchevskoy volcano. A scattering network transformed the continuous seismograms into a stable data representation (scattering coefficient matrix) for exploratory data analysis. With an ICA, we extracted features describing data-driven signal characteristics of the seismogram. These features have shown a continuously evolving seismogram with many pattern changes, in particular for stations a few kilometers away from the active Klyuchevskoy volcano. A larger number of features was necessary to observe this behavior. Simultaneously, the large number made it cumbersome to understand this behavior and relate it to volcanic activity. Therefore, we utilized a non-linear dimensionality reduction technique called UMAP to create a two-dimensional representation of the scattering coefficient matrix, which we call a seismogram atlas. In the seismogram atlas, we find various structures with closeby data points representing similar seismograms and distant data points representing dissimilar seismograms. The atlases offer a visual tool to analyze long seismic time series and they confirmed our observation of an ever-changing wavefield. We were able to relate certain atlas areas, representing similar signal characteristics, to different volcanic activities and changes in the volcanic plumbing system. While the seismic wavefield seems to change throughout the recording time, the atlas helped us to identify sudden and continuous pattern changes. Deep reactivations are characterized by sudden pattern changes and tremor depth changes are mainly characterized by continuous pattern changes. We also want to emphasize that each tremor period seems to have its own signal characteristics, resulting in distinct structures in the seismogram atlas.

8 Open Research

The KISS dataset (Shapiro et al., 2015) is available from the GEOFON data center of GFZ-Potsdam (https://geofon.gfz-potsdam.de). The data was processed with Python and visualized with Matplotlib (Hunter, 2007), available under the Matplotlib license at https://matplotlib.org/. The python packages `ScatSeisnNet` (Steinmann, Seydoux, & Campillo, 2022), `ObsPy` (Beyreuther et al., 2010) and `Scikit-learn` (Pedregosa et al., 2011) were heavily used for processing the data.

Acknowledgments

This study was financially supported by the European Research Council (ERC) under the European Union Horizon 2020 Research and Innovation Program (grant agreements no 742335, F-IMAGE for R.S., L.S., and M.C. and 787399, SEISMAZE for C.J. and N.S.).

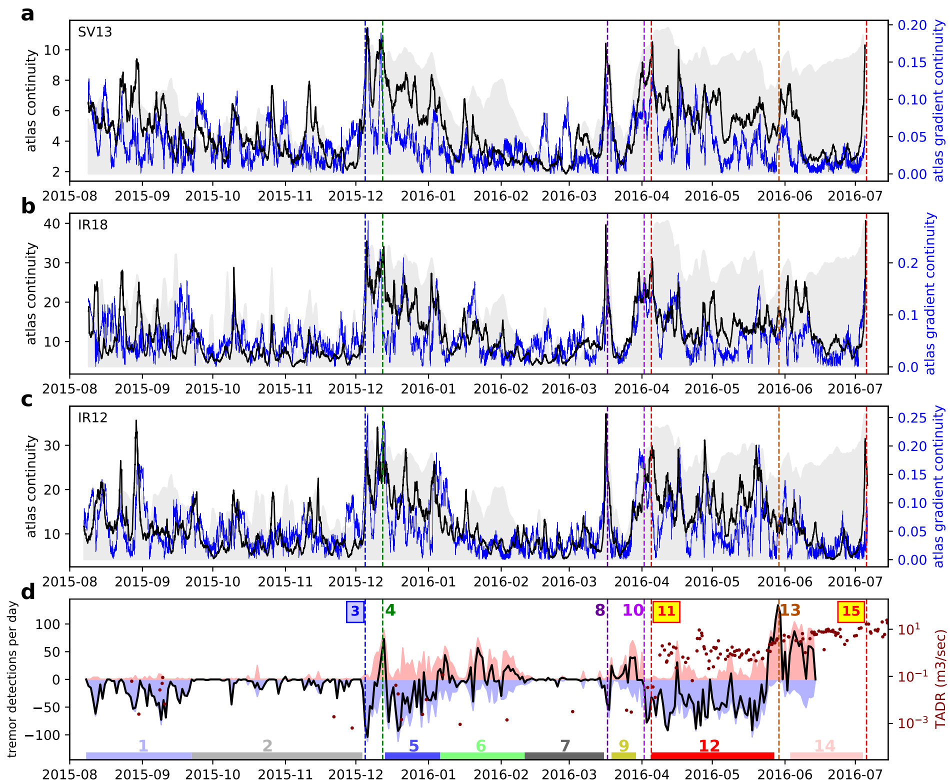

Figure 14. Time series of the atlas continuity and atlas gradient continuity of the three stations SV13 **(a)**, IR18 **(b)**, and IR12 **(c)**. The continuity has been averaged over consecutive steps (1.74 days). The gradient continuity has been averaged over 500 steps (3.47 days). Signal RMS amplitudes (averaged over 250 steps and normalized) are shown with light gray areas, for reference. Subfigure **d** shows the tremor count per day. Deep (below 10 km) and shallow (above 10 km) tremors are shown with blue and red shaded areas, respectively. The black continuous line shows the difference between the deep and shallow tremor counts. Colored numbers indicate different events and episodes of the activity of the Klychevskoy volcano-plumbing system. The vertical dashed lines correspond to specific events and the horizontal numbered bars correspond to specific time periods in the interpreted SV13 atlas in Figure 15.

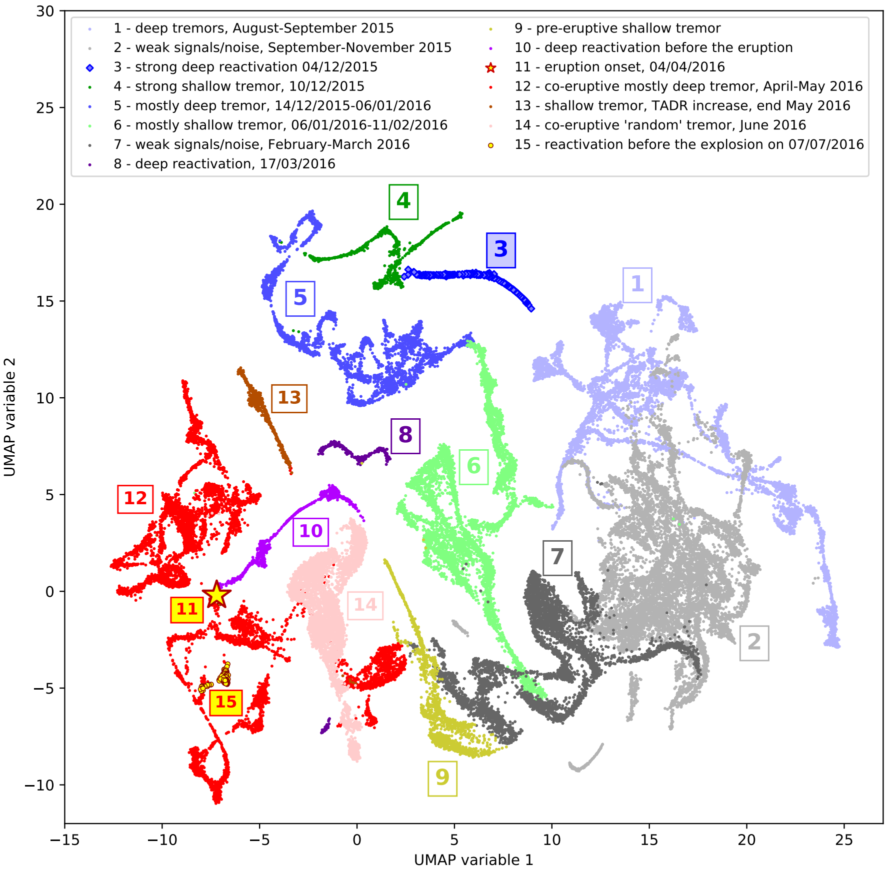

Figure 15. Synthesis of the results: interpreted seismogram atlas of SV13 with identification of volcanic activity and its relation to the plumbing system of the Klyuchevskoy volcano.

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, É., 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.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W., Ng, L. G., ... Newell, E. W. (2019). Dimensionality reduction for visualizing single-cell data using umap. *Nature biotechnology*, *37*(1), 38–44.

Beroza, G. C., & Ide, S. (2011). Slow earthquakes and nonvolcanic tremor. *Annual review of Earth and planetary sciences*, *39*, 271–296.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). Obspy: A python toolbox for seismology. *Seismological Research Letters*, *81*(3), 530–533. doi: https://doi.org/10.1785/gssrl.81.3.530

Bruna, J., & Mallat, S. (2013). Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1872–1886.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., ... others (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, *566*(7745), 496–502.

Carter, S., Armstrong, Z., Schubert, L., Johnson, I., & Olah, C. (2019). Activation atlas. *Distill*, *4*(3), e15.

Chouet, B. A., & Matoza, R. S. (2013). A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *Journal of Volcanology and Geothermal Research*, *252*, 108–175.

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, *36*(3), 287–314.

Coppola, D., Laiolo, M., Massimetti, F., Hainzl, S., Shevchenko, A. V., Mania, R., ... Walter, T. R. (2021). Thermal remote sensing reveals communication between volcanoes of the klyuchevskoy volcanic group. *Scientific reports*, *11*(1), 13090.

Curilem, G., Vergara, J., Fuentealba, G., Acuña, G., & Chacón, M. (2009). Classification of seismic signals at villarrica volcano (chile) using neural networks and genetic algorithms. *Journal of volcanology and geothermal research*, *180*(1), 1–8.

Dorendorf, F., Wiechert, U., & Wörner, G. (2000). Hydrated sub-arc mantle: a source for the kluchevskoy volcano, kamchatka/russia. *Earth and Planetary Science Letters*, *175*(1), 69-86. doi: https://doi.org/10.1016/S0012-821X(99)00288-5

Droznin, D., Shapiro, N., Droznina, S. Y., Senyukov, S., Chebrov, V., & Gordeev, E. (2015). Detecting and locating volcanic tremors on the klyuchevskoy group of volcanoes (kamchatka) based on correlations of continuous seismic records. *Geophysical Journal International*, *203*(2), 1001–1010.

Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*.

Ebmeier, S. (2016). Application of independent component analysis to multitemporal insar data with volcanic case studies. *Journal of Geophysical Research: Solid Earth*, *121*(12), 8970–8986.

Endo, E. T., & Murray, T. (1991). Real-time seismic amplitude measurement (rsam): a volcano monitoring and prediction tool. *Bulletin of Volcanology*, *53*(7), 533–545.

Falcin, A., Métaxian, J.-P., Mars, J., Stutzmann, É., Komorowski, J.-C., Moretti, R., ... others (2021). A machine-learning approach for automatic classification of volcanic seismicity at la soufrière volcano, guadeloupe. *Journal of Volcanology and Geothermal Research*, *411*, 107151.

Fedotov, S. A., Zharinov, N. A., & Gontovaya, L. I. (2010). The magmatic system of the klyuchevskaya group of volcanoes inferred from data on its eruptions, earthquakes, deformation, and deep structure. *J. Volcanol. Geotherm. Res.*, *4*(1), 1–33. doi: 10.1134/S074204631001001X

Fehler, M. (1983). Observations of volcanic tremor at mount st. helens volcano. *Journal of Geophysical Research: Solid Earth*, *88*(B4), 3476–3484.

Frank, W. B., Shapiro, N. M., & Gusev, A. A. (2018). Progressive reactivation of the volcanic plumbing system beneath tolbachik volcano (kamchatka, russia) revealed by long-period seismicity. *Earth and Planetary Science Letters*, *493*, 47-56. doi: https://doi.org/10.1016/j.epsl.2018.04.018

Gaddes, M., Hooper, A., Bagnardi, M., Inman, H., & Albino, F. (2018). Blind signal separation methods for insar: The potential to automatically detect and monitor signals of volcanic deformation. *Journal of Geophysical Research: Solid Earth*, *123*(11), 10–226.

Galina, N. A., Shapiro, N. M., Droznin, D. V., Droznina, S. Y., Senyukov, S. L., & Chebrov, D. V. (2020). Recurrence of deep long-period earthquakes beneath the klyuchevskoi volcano group, kamchatka. *Izvestiya, Physics of the Solid Earth*, *56*(6), 749–761. doi: 10.1134/S1069351320060026

Ghosh, B., Motagh, M., Haghighi, M. H., Vassileva, M. S., Walter, T. R., & Maghsudi, S. (2021). Automatic detection of volcanic unrest using blind source separation with a minimum spanning tree based stability analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 7771–7787.

Girina, O., Manevich, A., Loupian, E., Uvarov, I., Korolev, S., Sorokin, A., ... Burtsev, M. (2023). Monitoring the thermal activity of kamchatkan volcanoes during 2015 - 2022 using remote sensing. *Remote Sensing*, *15*(19). Retrieved from `https://www.mdpi.com/2072-4292/15/19/4775` doi: 10.3390/rs15194775

Girina, O. A., Manevich, A. G., Melnikov, D. V., Nuzhdaev, A. A., & Petrova, E. G. (2019). The 2016 eruptions in kamchatka and on the north kuril islands: The hazard to aviation. *Journal of Volcanology and Seismology*, *13*(3), 157–171. Retrieved from `https://doi.org/10.1134/S0742046319030047` doi: 10.1134/S0742046319030047

Glynn, C., & Konstantinou, K. (2016). Reduction of randomness in seismic noise as a short-term precursor to a volcanic eruption. *Scientific reports*, *6*(1), 37733.

Gorelchik, V. I., Garbuzova, V. T., & Storcheus, A. V. (2004). Deep-seated volcanic processes beneath klyuchevskoi volcano as inferred from seismological data. *Journal of Volcanology and Seismology*, *6*, 21-34.

Green, R. G., Sens-Schönfelder, C., Shapiro, N., Koulakov, I., Tilmann, F., Dreiling, J., ... Gordeev, E. (2020). Magmatic and sedimentary structure beneath the klyuchevskoy volcanic group, kamchatka, from ambient noise tomography. *Journal of Geophysical Research: Solid Earth*, *125*(3), e2019JB018900. doi: https://doi.org/10.1029/2019JB018900

Hibert, C., Provost, F., Malet, J.-P., Maggi, A., Stumpf, A., & Ferrazzini, V. (2017). Automatic identification of rockfalls and volcano-tectonic earthquakes at the piton de la fournaise volcano using a random forest algorithm. *Journal of Volcanology and Geothermal Research*, *340*, 130–142.

Holtzman, B. K., Paté, A., Paisley, J., Waldhauser, F., & Repetto, D. (2018). Machine learning reveals cyclic changes in seismic source spectra in geysers geothermal field. *Science advances*, *4*(5), eaao2929.

Hotovec, A. J., Prejean, S. G., Vidale, J. E., & Gomberg, J. (2013). Strongly gliding harmonic tremor during the 2009 eruption of redoubt volcano. *Journal of Volcanology and Geothermal Research*, *259*, 89–99.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. doi: 10.1109/MCSE.2007.55

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, *13*(4-5), 411–430.

Hyvärinen, A., Ramkumar, P., Parkkonen, L., & Hari, R. (2010). Independent component analysis of short-time fourier transforms for spontaneous eeg/meg analysis. *NeuroImage*, *49*(1), 257–271.

Ichihara, M., Ohminato, T., Konstantinou, K. I., Yamakawa, K., Watanabe, A., & Takeo, M. (2023). Seismic background level (sbl) growth can reveal slowly developing long-term eruption precursors. *Scientific reports*, *13*(1), 5954.

Jenkins, W. F., Gerstoft, P., Bianco, M. J., & Bromirski, P. D. (2021). Unsupervised deep clustering of seismic data: Monitoring the ross ice shelf, antarctica. *Journal of Geophysical Research: Solid Earth*, e2021JB021716.

Journeau, C., Shapiro, N. M., Seydoux, L., Soubestre, J., Ferrazzini, V., & Peltier, A. (2020). Detection, classification, and location of seismovolcanic signals with multicomponent seismic data: Example from the piton de la fournaise volcano (la réunion, france). *Journal of Geophysical Research: Solid Earth*, *125*(8), e2019JB019333. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JB019333` doi: https://doi.org/10.1029/2019JB019333

Journeau, C., Shapiro, N. M., Seydoux, L., Soubestre, J., Koulakov, I. Y., Jakovlev, A. V., ... others (2022). Seismic tremor reveals active trans-crustal magmatic system beneath kamchatka volcanoes. *Science advances*, *8*(5), eabj1571.

Julian, B. R. (1994). Volcanic tremor: Nonlinear excitation by fluid flow. *Journal of Geophysical Research: Solid Earth*, *99*(B6), 11859–11877.

Köhler, A., Ohrnberger, M., & Scherbaum, F. (2010). Unsupervised pattern recognition in continuous seismic wavefield records using self-organizing maps. *Geophysical Journal International*, *182*(3), 1619–1630.

Konstantinou, K. I., & Schlindwein, V. (2003). Nature, wavefield properties and source mechanism of volcanic tremor: a review. *Journal of Volcanology and Geothermal Research*, *119*(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. *Scientific Reports*, *11*(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. *Journal of Geophysical Research: Solid Earth*, *125*(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. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 1322–1331.

Latter, J. H. (1979). Volcanological observations at tongariro national park. ii: Types and classification of volcanic earthquakes, 1976-1978. *Report-Geophysics Division*, *150*.

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. *Geol.*, *42*(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. *Nature*, *418*(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. *bioRxiv*, 2022–07.

Maggi, A., Ferrazzini, V., Hibert, C., Beauducel, F., Boissier, P., & Amemoutou, A. (2017). Implementation of a multistation approach for automated event classification at piton de la fournaise volcano. *Seismological Research Letters*, *88*(3), 878–891.

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. *Journal of Geophysical Research: Solid Earth*, e2022JB025738.

Malfante, M., Dalla Mura, M., Mars, J. I., Métaxian, J.-P., Macedo, O., & Inza, A. (2018). Automatic classification of volcano seismic signatures. *Journal of Geophysical Research: Solid Earth*, *123*(12), 10–645.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

McNutt, S. R. (2005). Volcanic seismology. *Annual Review of Earth and Planetary Sciences*,

33(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. *Nature Communications*, *11*(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. *The Cryosphere Discussions*, 1–19.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, *12*, 2825–2830. doi: 10.5555/1953048.2078195
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. *Geophysical Research Letters*, *47*(3), e2019GL085523.
Rey-Devesa, P., Prudencio, J., Benítez, C., Bretón, M., Plasencia, I., León, Z., ... Ibáñez, J. M. (2023). Tracking volcanic explosions using shannon entropy at volcán de colima. *Scientific Reports*, *13*(1), 9807.
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. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 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 fluid-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
Shapiro, N. M., Sens-Schönfelder, C., Lühr, B. G., Weber, M., Abkadyrov, I., Gordeev, E. I., ... Saltykov, V. A. (2015). *Klyuchevskoy volcanic group experiment (kiss).* GFZ Data Services, [Dataset]. Retrieved from `http://geofon.gfz-potsdam.de/doi/network/X9/2015` doi: 10.14470/K47560642124
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 classification 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., García, L., Benítez, M. C., & Ibañez, J. (2018). Detection and classification 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