



**HAL**  
open science

## Machine Learning Reveals the Seismic Signature of Eruptive Behavior at Piton de la Fournaise Volcano

C. X Ren, Aline Peltier, V. Ferrazzini, B. Rouet-leduc, P. A Johnson, Florent  
Brenguier

► **To cite this version:**

C. X Ren, Aline Peltier, V. Ferrazzini, B. Rouet-leduc, P. A Johnson, et al.. Machine Learning Reveals the Seismic Signature of Eruptive Behavior at Piton de la Fournaise Volcano. *Geophysical Research Letters*, 2020, 47 (3), pp.e2019GL085523. 10.1029/2019GL085523 . hal-02930212

**HAL Id: hal-02930212**

<https://hal.univ-grenoble-alpes.fr/hal-02930212v1>

Submitted on 4 Sep 2020

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

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

# Geophysical Research Letters



## RESEARCH LETTER

10.1029/2019GL085523

### Key Points:

- Statistical features derived from time-windowed, filter-banked seismic data can be an effective way to characterize eruptive behavior of volcanoes
- Supervised learning allows us to determine the eruptive state of the volcano given a single time window of raw seismic data from a single station
- Spectral clustering can reveal different phases of eruptions and differences between various eruptions

### Supporting Information:

- Supporting Information S1

### Correspondence to:

C. X. Ren,  
cren@lanl.gov

### Citation:

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, e2019GL085523. <https://doi.org/10.1029/2019GL085523>

Received 24 SEP 2019

Accepted 15 JAN 2020

Accepted article online 21 JAN 2020

© 2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

## Machine Learning Reveals the Seismic Signature of Eruptive Behavior at Piton de la Fournaise Volcano

C. X. Ren<sup>1,2</sup> , A. Peltier<sup>3,4</sup> , V. Ferrazzini<sup>3,4</sup> , B. Rouet-Leduc<sup>2</sup> , P. A. Johnson<sup>4</sup> , and F. Brenguier<sup>5</sup> 

<sup>1</sup>Space Data Science and Systems Group, Los Alamos National Laboratory, Los Alamos, NM, USA, <sup>2</sup>Geophysics Group, Los Alamos National Laboratory, Los Alamos, NM, USA, <sup>3</sup>Université de Paris, Institut de physique du globe de Paris, CNRS, Paris, France, <sup>4</sup>Observatoire volcanologique du Piton de la Fournaise, Institut de physique du globe de Paris, La Plaine des Cafres, France, <sup>5</sup>ISterre, Université Grenoble Alpes, Gières, France

**Abstract** Volcanic tremor is key to our understanding of active magmatic systems, but due to its complexity, there is still a debate concerning its origins and how it can be used to characterize eruptive dynamics. In this study we leverage machine learning techniques using 6 years of continuous seismic data from the Piton de la Fournaise volcano (La Réunion island) to describe specific patterns of seismic signals recorded during eruptions. These results unveil what we interpret as signals associated with various eruptive dynamics of the volcano, including the effusion of a large volume of lava during the August–October 2015 eruption as well as the closing of the eruptive vent during the September–November 2018 eruption. The machine learning workflow we describe can easily be applied to other active volcanoes, potentially leading to an enhanced understanding of the temporal and spatial evolution of volcanic eruptions.

**Plain Language Summary** A good understanding of volcanic activity is key to managing volcanic hazards resulting from eruptive activity. Volcanic tremor is a continuous seismic signal often seen during eruptions associated with the flow of magma through the volcano and is thus an extremely useful tool in characterizing the progression and phases of eruptions. In this study we study this signal at the Piton de la Fournaise volcano, on La Réunion island. Using machine learning algorithms, we investigate characteristics of this signal emitted by the volcano during eruptions to reveal the fundamental frequency at which it occurs, as well as changes in eruptive state that occur during some eruptions in our data set. This workflow may be applied to other volcanos to further our understanding of eruptive dynamics.

## 1. Introduction

Long-lasting seismic signals known as volcanic tremors are almost ubiquitously present in eruptive episodes at volcanoes (Jellinek & Bercovici, 2011). These seismic signals are thought to be critical in characterizing the magma migration pathways in the internal plumbing system of volcanoes, as they are typically linked with magma propagation (B. A. Chouet, 1996) as well as degassing from eruptive vents (Battaglia et al., 2005) and are thus of interest in the context of both monitoring and forecasting volcanic activity. Attempts to forecast and monitor volcanic eruptions must take into account that the style and magnitudes of eruptions vary drastically, even throughout the duration of a single eruption (Battaglia, Aki, & Staudacher, 2005; Chardot et al., 2015; Kurokawa et al., 2016). Despite the inherent complexity of volcanic systems, there have been many demonstrations of the utility of volcanic tremor in monitoring eruptions. Battaglia and Aki (2003) demonstrated that the eruptive tremor sources at Piton de la Fournaise volcano were good indicators of the locations of eruptive fissures; in a subsequent study Battaglia et al. (2005) demonstrated that the cumulative amplitude of tremor recorded throughout an eruption could be used to estimate the volume of lava erupted. Several studies have also shown that analyzing the spectral content and location of volcanic tremor can help track the evolution of magma migration (Di Lieto et al., 2007; Jellinek & Bercovici, 2011; Kurokawa et al., 2016)

The application of machine learning (ML) techniques to the analysis of geophysical signals has become widespread in diverse settings such as the analysis of laboratory experiments (Hulbert et al., 2018; Rouet-Leduc et al., 2017; Rouet-Leduc, Hulbert, Bolton, et al., 2018), tracking slow-slip in real Earth (Rouet-Leduc, Hulbert, & Johnson, 2018), and phase association for the development of earthquake catalogs to

name a few (McBrearty et al., 2019; Ross et al., 2019). With regard to the applications of ML to study and characterization of volcanoes, the primary applications thus far have been in the classification of volcano-seismic signals (Hibert et al., 2017; Malfante et al., 2018; Titos et al., 2019). In this work, we describe how statistical features derived from the continuous seismic signal recorded at Piton de la Fournaise volcano can be utilized to build ML models that reveal the characteristic eruptive tremor and eruptive dynamics of volcanic eruptions.

## 2. Regional Setting

Piton de la Fournaise is an active volcano situated on La Réunion, a hot spot basaltic island in the western part of the Indian Ocean located approximately 800 km east of Madagascar. It is one of the most active volcanoes in the world, exhibiting 71 eruptions between 1985 and 2018 (Duputel et al., 2019; Roult et al., 2012).

Recent activity has been focused in the Enclos Fouqué caldera, formed about 5,000–3,000 years ago (Ort et al., 2016), inside which a terminal cone was formed as a consequence of frequent effusive eruptions characterized by lava fountains and lava flow emissions. Following 41 months of rest between the end of 2010 and June 2014, eruptive activity at Piton de la Fournaise renewed on 20 June 2014. Between 2014 and March 2019, 14 eruptions occurred on the flank or at the base of the terminal cone. The locations of the fissures generated by these eruptions are shown in Figure 1. Eruptions are fed by a magma plumbing system consisting of a succession of reservoirs spreading from depth greater than 30 km below the external western flank of the volcano (Michon et al., 2015) to 2 km below the summit craters (Di Muro et al., 2014; Peltier et al., 2009), where most eruptions initiate.

### 2.1. Seismic Data

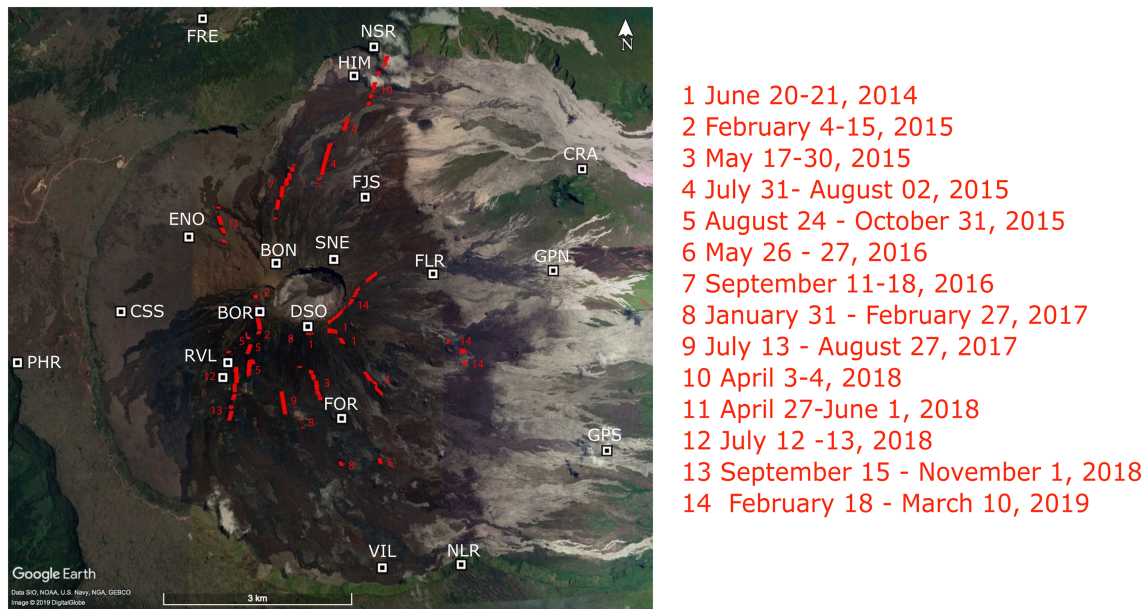
For this study, we leverage 6 years of continuous seismic data recorded across the Observatoire Volcanologique du Piton de la Fournaise (OVPF) network shown in Figure 1. The network consists of stations equipped with short-period seismometers and broadband seismometers, recording at a sampling rate of 100 Hz. The data shown in this manuscript are primarily from the seismic station of the Cratere Bory (BOR) site; although the analysis described in sections 2.2 and 2.3 has been performed for most of the seismic stations situated in the Enclos Fouqué belonging to the OVPF network caldera, we find that data recorded at the BOR site provide the most data for analysis in our period of interest and good performance for our eruptive state classifier while also demonstrating good variability among eruptive signals for the unsupervised learning. We note that in the period under study at BOR, seismic data for the 4 April eruption are missing, and thus, this eruption does not appear in our data set. We also use data from the Enclos Sery Sud, Chateau Fort, and the Faujas sites to characterize signals recorded throughout the eruptions, as these stations are approximately equidistant from the Dolomieu crater (De Plaen et al., 2016) and have a good amount of overlap in terms of available data during the periods of interest.

### 2.2. Feature Building

In order to reduce the continuous seismic signal recorded at various stations at Piton de la Fournaise to a set of tabular features, we first correct the seismic signal for each day and each station to remove the instrument response.

A filter bank is then applied in between the spectral range of 0.5 and 26 Hz with an initial spacing of 0.5–2 Hz and then a spacing of 1 Hz, in a similar approach to that described by Rouet-Leduc, Hulbert, and Johnson (2018). This results in 25 frequency bands. This frequency range is chosen, as it encompasses the typical frequencies of volcanic tremor and volcano-tectonic events reported at Piton de la Fournaise (Battaglia, Aki, & Ferrazzini, 2005; Battaglia, Aki, & Staudacher, 2005; Duputel et al., 2019), although eruptive signals have been reported below 1 Hz more recently on broadband instruments.

We then generate features by scanning a moving window of length 1 hr across the band-passed seismic signal, for each frequency band. The aim here is to capture distributions of the seismic data within the time windows and relationships across different frequency bands for a given window. This has been shown to be an effective technique to parametrize geophysical time series data in various settings such as simulations (Ren et al., 2019), laboratory experiments (Hulbert et al., 2018; Rouet-Leduc et al., 2017; Rouet-Leduc, Hulbert, Bolton, et al., 2018), and real Earth (Rouet-Leduc, Hulbert, & Johnson, 2018). In this case, these features consist of a range of percentiles and the range of the data, as well as normalized and nonnormalized



**Figure 1.** Map of the seismic stations located near the piton de la Fournaise volcano, with the locations of fissures resulting from eruptions highlighted and numbered in red.

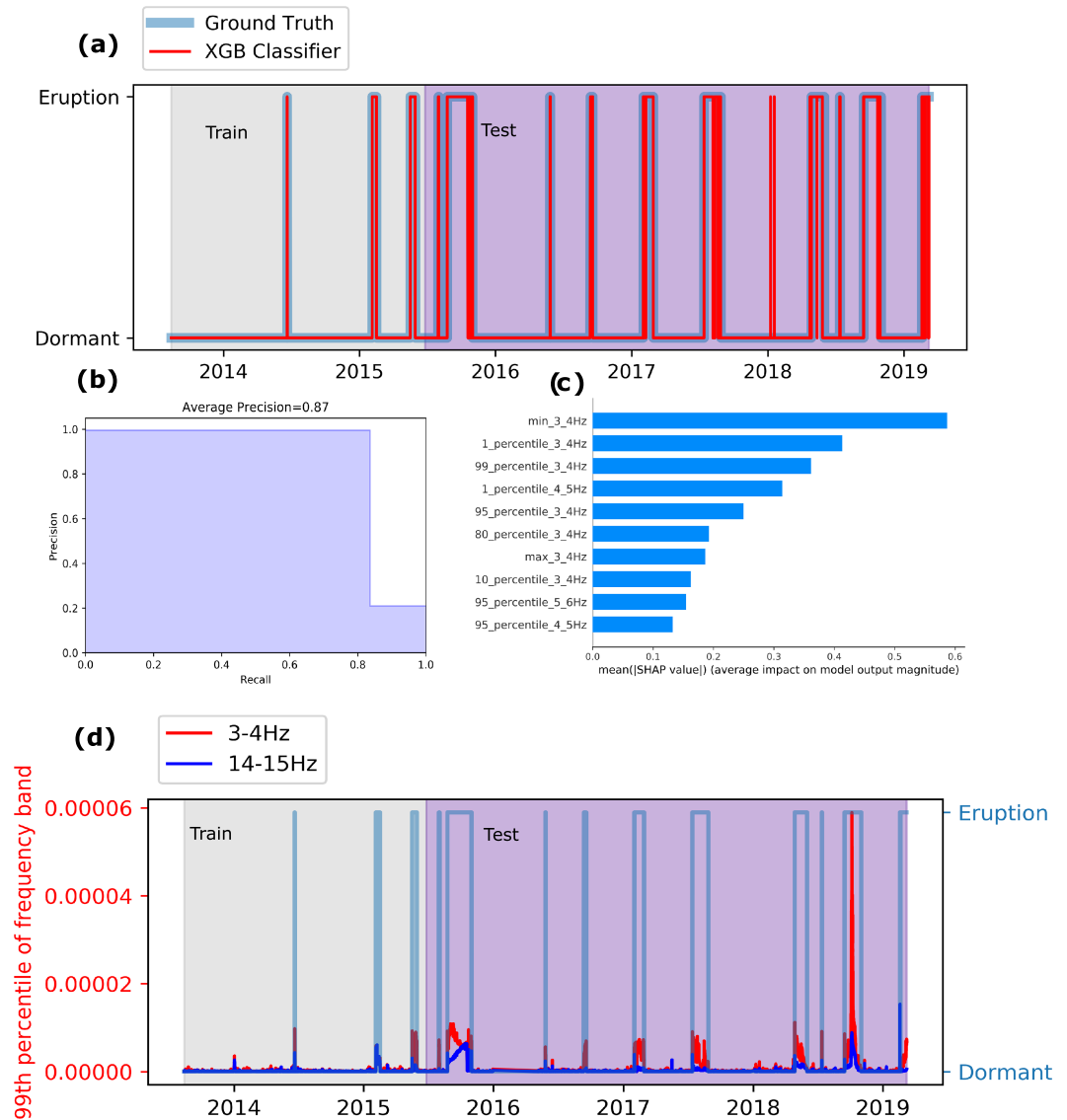
higher-order moments (see the supporting information for table describing features). For the Bory Crater seismic station (BOR), this results in 5,907 nonoverlapping 1 hr time windows of eruption data with 990 features generated for each window point, or 39 features per spectral band, and 38,021 windows in total over the 2013–2019 period.

### 2.3. Supervised Learning: Learning the Eruptive Signature of the Piton de la Fournaise

We first use a supervised learning approach to determine the characteristics of eruptive tremor detected at the Piton de la Fournaise. In this case, the target label is the eruptive state of the volcano (erupting or dormant). In applying this approach, we are essentially asking the following question: “For all eruptions in our data set, are there consistent underlying features/frequency bands for single stations which enable us to classify against non-eruptive signals?” For this binary classification task, we utilize a ML algorithm known as gradient-boosted decision trees (Friedman, 2002) with a cross-entropy loss function. We used the XGBoost implementation of gradient-boosted decision trees to perform the modeling described in this paper, which enables parallel and distributed computation of the model (Chen & Guestrin, 2016).

The XGBoost model takes the features derived from a sliding time window over the seismic signal recorded at a station as inputs and outputs a prediction on the eruptive state of the volcano. The training data in this case constitute approximately 30% of the total data available for the BOR station or 12,000 time windows. The trained model is blind tested on the remaining 70% of the data, outputting an estimate for the eruptive state of the volcano during this period, without ever having actually seen these data. This allows us to estimate the consistency of the characteristics of the eruptive tremor throughout this period: If the spectral characteristics of the tremor are different between the training and the testing set, the model will be unable to effectively predict the eruptive state in the testing set.

Figure 2(a) shows the ability of our classifier to predict whether Piton de la Fournaise is undergoing an eruption or dormant based on features generated from a single time window of the continuous seismic signal recorded at the BOR station. The classifier was trained on the gray portion of the data set (2013–2015) and tested on the violet section (2015–2019). The model performs well on the test set, with an average precision of 0.87, accuracy of 0.97, and a recall score of 0.83. Figure 2(b) shows the precision-recall curve for the XGBoost model. Figure 2(c) shows the top 10 most important features learned by the model during the training process, based on their Shapely Additive Explanations values (Lundberg et al., 2018; Lundberg & Lee, 2017). We can see from these features that the eruptive tremor detected at the BOR station occurs



**Figure 2.** (a) performance of the XGBoost classifier trained to predict the eruptive state of piton de la Fournaise volcano based on features derived from the continuous seismic signal recorded at the BOR station. The eruptive state (target) is shown in blue and the model predictions in red. The model is trained only on the gray portion of data. (b) precision-recall curve for the XGBoost model, with an average precision of 0.87. (c) top 10 informative features derived from the continuous seismic signal, based on shapely additive explanations importance value, demonstrating that the volcanic tremor signal resides mostly in the 3–5 Hz band. (d) the 99th percentile of the 3–4 Hz (red) and 14–15 Hz (blue) plotted with eruptive state. The 3–4 Hz band value increases beyond a consistent threshold for eruptive states across the data set, allowing the machine learning algorithm to recognize the occurrence of eruptive tremor.

consistently in the 3–5 Hz frequency range. This is confirmed in Figure 2(d), where we see that the 99th percentile of the 3–4 Hz frequency band (red trace) of the seismic signal spikes for all eruptions, as opposed to the 14–15 Hz band (blue trace).

We note that the slightly lower recall score indicates that there are a number of false negatives produced by our model in the test set. Examining Figures 2(a) and 2(d), we see that during some eruptions, the level of the 3–4 Hz band can drop below a given threshold, resulting in false negative prediction. Despite this, the performance of the model in detecting the state of the volcano is surprising given the small training set consisting of only three eruptions and the fact that the features are derived from the signal detected at a single station. Given the feature importances, we can see the classifier is classifying based on the amplitude of

the 3–5 Hz frequency range, and one might be tempted to think that the same result could simply be achieved by inspecting the data. While this is true, analyzing 5,907 individual hours of seismic data to establish consistent patterns in the data can be a time-consuming task. Taking the supervised learning approach here allows the user to automate this process through the creation of features and application of XGBoost and interpret the model through the feature importances as shown in Figure 2(c).

#### 2.4. Unsupervised Learning: Spectral Clustering

Unsupervised learning is a term that describes a set of ML techniques used to learn relationships in data sets where no training labels are available (Ghahramani, 2004). Cluster analysis, which is a subset of unsupervised learning, can be defined as the task of grouping data using similarity measures that quantify the proximity between data vectors in a given feature space (Aminzadeh & Chatterjee, 1984). In this particular case, we have no explicit labeled data concerning different types or phases of eruptive behavior that Piton de la Fournaise may exhibit or the specific seismic fingerprints of these behaviors in our feature space, so this approach is particularly well suited for our problem. This can be considered an opposing problem to the one we address using supervised learning, here we pose the following question: “Given that there is some underlying common structure in our feature space concerning eruptive signals, are there any meaningful differences between our features during the eruptions?” In order to tackle the investigation into the seismic signature of the eruptive dynamics of Piton de la Fournaise, we turn to a clustering algorithm known as spectral clustering. Spectral clustering outperforms “standard” clustering algorithms such as  $k$ -means in cases where the user wishes to cluster based on connectivity rather than compactness (see Supporting Information S1 and Figure S2).

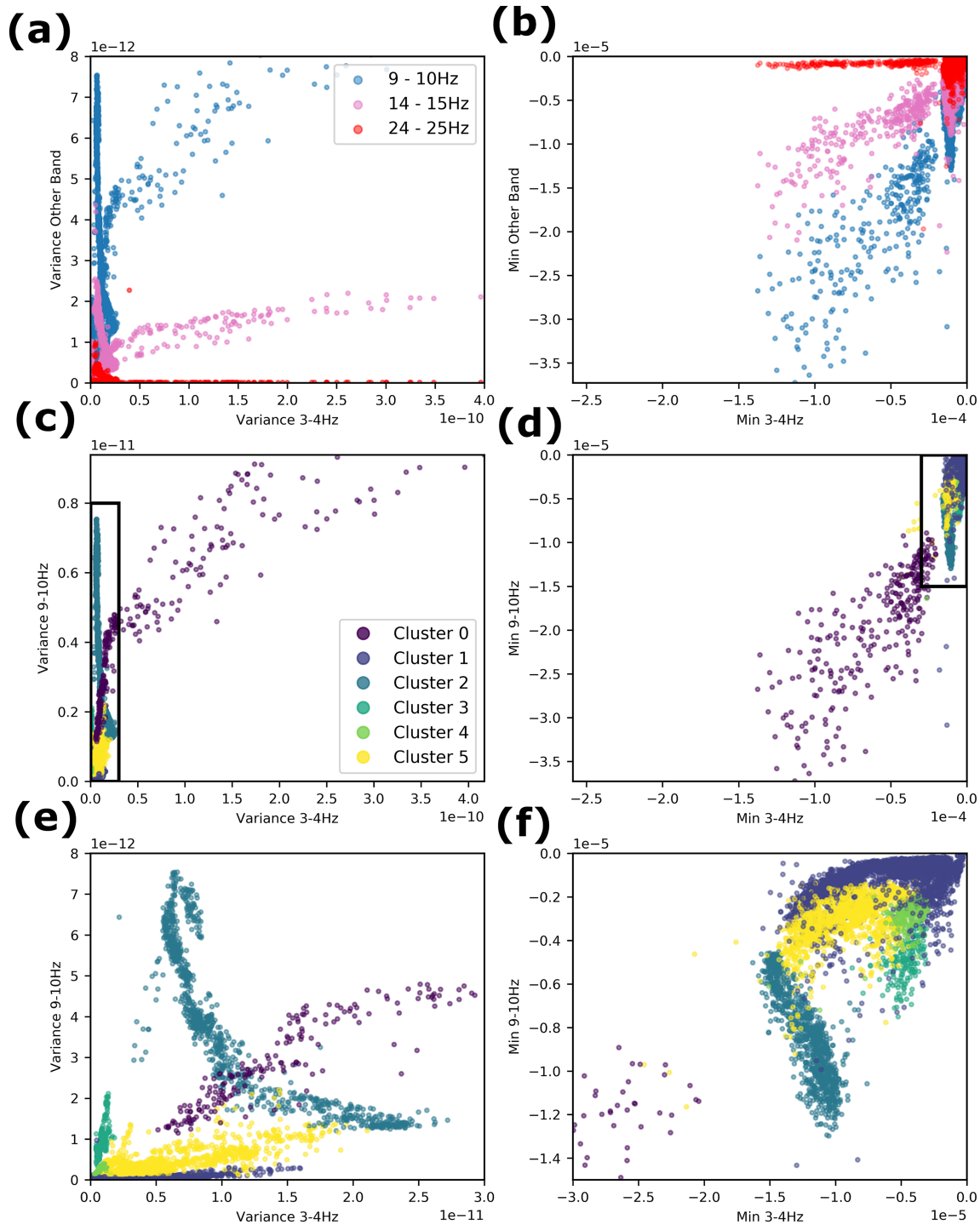
We utilize a connectivity-based clustering algorithm as we hypothesize that the spectral characteristics of the seismic signal are likely to change continuously during an eruption as the system transitions between phases; thus, the clusters of behavior are likely to be connected in our spectral feature space. This type of behavior in geophysical systems has previously been studied by Holtzman et al. (2018) who used unsupervised learning to characterize the spectral properties of seismic signals collected near geysers and identify different phases of behavior.

Figures 3(a) and 3(b) show features generated using the method described in section 2.2 on the vertical component of the continuous seismic signal recorded at the BOR station during eruptions occurring between 2013 and 2019. Here, we show two features across three different spectral bands, selected to illustrate the structure present in the 39-dimensional feature space generated for each spectral band: the variance and the minimum in a given time window for the band-passed signal. Each point in the scatterplots represents a single hour-long time window, and each color represents the relationship between the features generated for the dominant eruptive tremor band (3–4 Hz) and higher frequencies. The dominant eruptive tremor band was determined using the XGBoost model discussed in section 3.3.

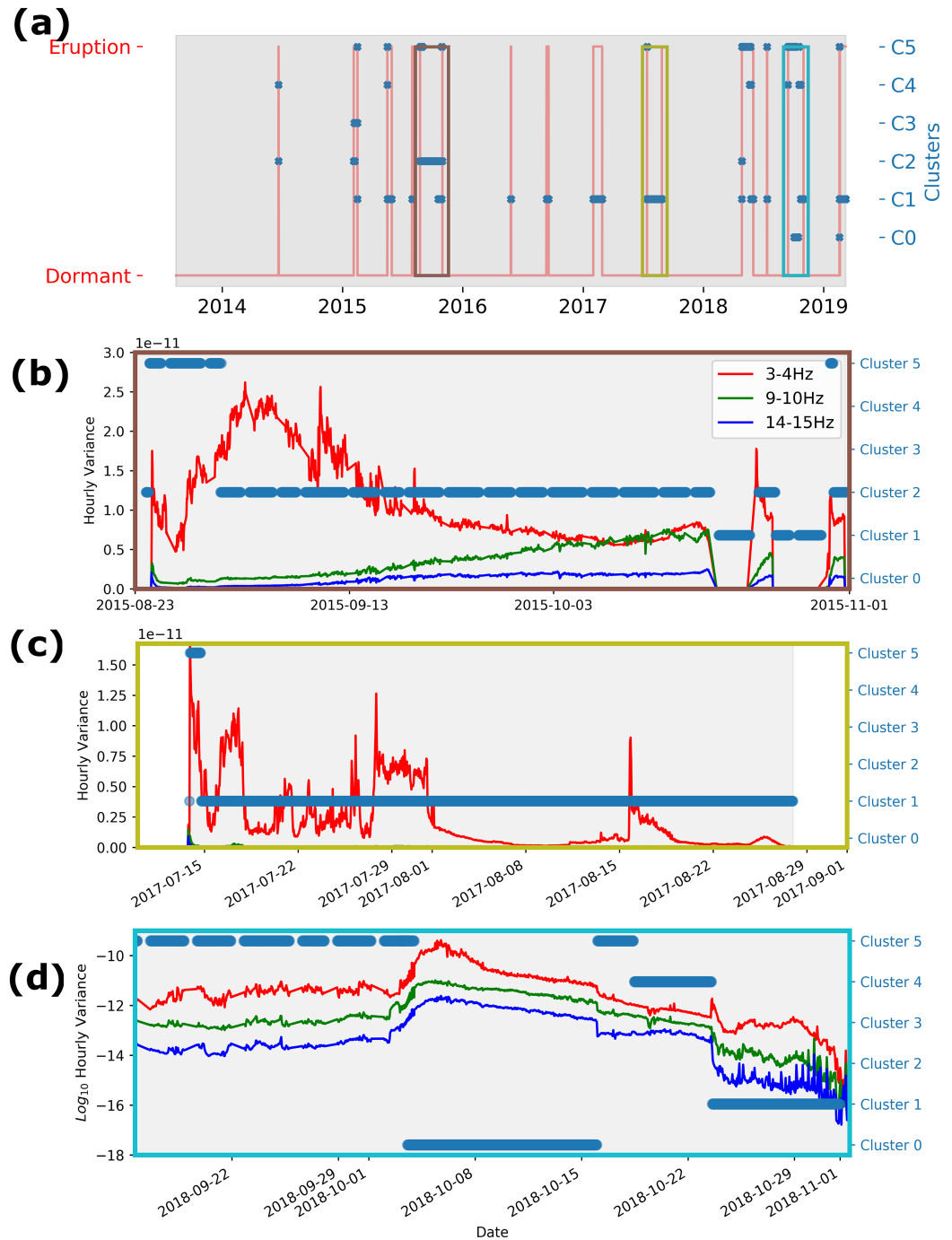
The results of the spectral clustering process are shown in Figures 3(c)–3(f). Here we choose 6 as the optimal number of clusters, determined by using the eigengap heuristic (von Luxburg, 2007). For the sake of illustration, we choose the 9–10 Hz spectral band (the blue scatter plot shown in Figures 3(a) and 3(b)), as this is the frequency band that shows the most structure when compared to the fundamental eruptive frequency band (3–4 Hz) and helps to illustrate the clustering results. We note that some clusters exhibit considerable overlap in the feature space visualization (C1, C5, and C4), there are also clusters that exhibit distinct separation in our feature space picked out by the algorithm (C0 and C2).

In order to evaluate the effectiveness of the clustering analysis on the eruptive dynamics of the Piton de la Fournaise volcano, we examine the temporal distribution of the clusters throughout the data set. Figure 4 (a) shows the Piton de la Fournaise eruptive state during the 2013–2019 period (red trace) and how the spectral clustering algorithm separated the features into six clusters (blue scatter plot). The colored bounding boxes are chosen to demonstrate the effectiveness of the combination of our feature generation method and spectral clustering in highlighting different eruptive behaviors of the Piton de la Fournaise volcano. We note that there are missing time windows in our analysis due to gaps in the available data; these appear as gaps in the clustering states in the following figures.

Figure 4(b) shows an irregular eruption, during which most of the features belonging to C2 occur. The eruption begins in C5, where we see the dominance of the 3–4 Hz frequency band in the eruptive tremor but also

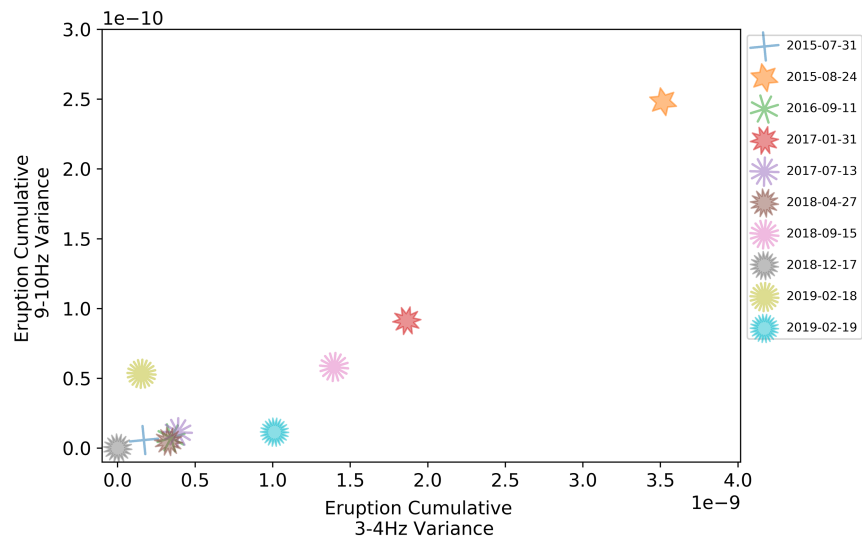


**Figure 3.** Spectral relationship between features generated at higher frequencies (9–10, 14–15, and 24–25 Hz) and the dominant volcanic tremor frequency (3–4 Hz). The features shown here are the (a) variance and (b) the minimum value for hour-long time windows of the band-passed continuous seismic signal for the station BOR in the aforementioned spectral bands. We show the results of the spectral clustering on these features in (c) and (d), respectively, with the dominant tremor band plotted against the 9–10 Hz band. The black bounding boxes are shown in greater detail in (e) and (f) for the spectral band variance and minimum value, demonstrating the structure present among the features as picked out by the clustering algorithm.



**Figure 4.** Temporal distribution of the clusters shown in Figure 3. (a) the temporal distribution for the entire data set (2013–2019). The three bounding boxes are chosen as representative examples of the clusters. (b) an “anomalous” eruption, exhibiting a transition between clusters 5 and 2: During this eruption, we see an initial dominant frequency range for the eruptive tremor between 3 and 4 Hz but as the eruption progresses higher frequencies start to dominate the eruptive tremor for the BOR station. We also note the abrupt decrease in amplitude of the eruptive signals in the waning stages of the eruption, where the eruptive signals detected at BOR transition between clusters 1 and 2 repeatedly. (c) a more “regular” eruption residing mostly in cluster 1 (see Figure 3(e)) where the variance of the 3–4 Hz frequency band of the eruptive tremor tends to be strong relative to higher frequencies for the BOR station. (d) an eruption time windows located in cluster 0. We note that both the dominant frequency band *and* the higher frequencies are strong in this cluster.





**Figure 5.** Cumulative variances for the 3–4 Hz band and the 9–10 Hz band, averaged for stations CSS, FOR, and FJS (see Figure 1) which are approximately equidistant from the Dolomieu crater, confirming the “anomalous” nature of the august 2015 eruption in terms of the seismic signals detected throughout the duration of the eruption. We have removed eruptions for which more than 50% of the time windows were missing across the three stations combined.

moderately elevated seismic energy in the higher-frequency bands. As the eruption progresses, the eruptive state transitions to C2 such that by the end of the eruption the dominant frequency in the eruptive tremor detected at BOR is no longer 3–4 Hz but closer to the 9–10 Hz range. The anomalous nature of the August 2015 eruption is further highlighted by the sudden decreases and increases in eruptive tremor amplitude toward the end of the eruption where we see consecutive transitions between C1 and C2.

A more “usual” eruption is shown in Figure 4(c), with the behavior of the volcano throughout this eruption occurring almost entirely in C1: During this type of eruptive behavior, we see the main eruptive tremor frequency band (3–4 Hz) dominates the seismic signal collected at BOR.

Figure 4(c) demonstrates the occurrence of another state transition revealed by the clustering algorithm: The eruption occurring in September 2018 initiates in C5, but transitions to C0 as the eruptive tremor amplitude increases in early October 2018. We show the features on a log scale here for clarity. Almost all time windows occurring in C0 for the whole data set occur during this eruption, indicating that the spectral clustering algorithm has successfully picked out a distinctive signature for this phase of the eruption. The clustering state then transitions through C4 as the eruptive tremor decays. Finally, the quiescent stages of the eruption occur in C1, where the features for the higher frequencies are extremely low ( $<10^{-14}$ ).

### 3. Physical Interpretation

The eruptive behavior of the Piton de la Fournaise volcano occurs in C5 during the initial phases of 7 out of 13 eruptions in our period of study (see Figure 4). Our interpretation of this result is that C5 contains time windows during which seismic signals associated with the dominant tremor band (3–5 Hz) are high in amplitude (see Figure 3) but also moderate amplitudes in the higher-frequency ranges, which may result from rockfall or volcano-tectonic events. It is also worth noting that broadband seismic signals can be related to impulsive events related to the migration of dykes and brittle failure within the edifice of the volcano (Duputel et al., 2019) but also subsidence of the caldera floor and lava tube collapse related to the lava load.

The anomalous signal recorded during the eruption of August–October 2015 can thus be interpreted in a similar manner: The sustained high frequencies belonging to C2, which occur almost entirely during the aforementioned eruption, are likely due to the subsidence of the caldera floor or lava tube collapses due to the load of the lava flow from the eruption. Indeed, this eruption resulted in a large total effusive volume of  $35.5 \times 10^6 \text{ m}^3$ , the largest amount recorded for the 2013–2018 period. Geodetic data indicate the volcanic edifice continually inflated until the end of September, indicating pressurization of the shallow magma

reservoir as a consequence of the impulsive ascent of deeper magma. Furthermore, Coppola et al. (2017) reported a magnesium-rich magmatic source becoming evident in correspondence with this new period of inflation, with an increase in SO<sub>2</sub> emissions and strong CO<sub>2</sub> enrichment in summit fumaroles during this phase of the eruption, providing further that the shallow reservoir was refilled by deep magma, leading to the repressurization of the magmatic plumbing system. This repressurization is likely responsible for the sustained flow of lava throughout the long duration of this eruption, inducing subsidence and lava tube collapse and the associated anomalous seismic signals observed during the second phase of this eruption. We note furthermore that the fissures and lava flow occurred in relatively flat areas of the volcano, reducing the flow of lava and inducing stacking and increasing the lava load in the area.

Figure 5 shows the unique position of this eruption in our cumulative feature space. To produce this figure, we calculate the features for the vertical components of three broadband stations (Faujas, Enclos Sery Sud, and Chateau Fort) that are approximately equidistant from the Dolomieu crater (De Plaen et al., 2016; see Figure 1). We average the feature values for time windows at which data are available across all three stations and sum these across the duration of the respective eruptions, to give a sense of the feature values coupled with the overall duration of the eruptions. For each eruption, we impose the condition that there needs to be over 50% of the duration of the eruption available in terms of the features. From Figure 5, we can see that features associated with the dominant tremor band as well as higher frequencies were consistently high throughout the duration of this eruption, even when compared to similar eruptions of longer duration (>1 month) during which repressurization was observed, such as the 27 April 2017 to 1 June 2018 eruption. We also note C0 represents another eruptive transition picked out by the clustering algorithm: According to OVPF reports concerning the September–November 2018 eruption, gas pistons were recorded from 3 October onward, marking a transition in the degassing of the eruption and accompanying the sharp increase in eruptive tremor recorded. The increase of tremor recorded, which corresponds to the C0 cluster, is likely related to the gradual closing of the eruptive cone (OVPF, 2018), leading to a higher amplitude of “resonance” (B. Chouet, 1988), thus increasing the tremor amplitude. We note that following the closing of the eruptive cone, the eruption continued mostly in lava tubes, which is likely the source of the higher frequencies observed for C0.

#### 4. Discussion

The use of both supervised and unsupervised learning methods in this work has helped to elucidate some of the structure present in seismic signals collected during eruptions at the Piton de la Fournaise volcano. Our XGBoost model isolates consistent frequencies present throughout the majority of the eruptions in our data set, with a surprisingly good performance for a single station. This type of analysis could potentially be utilized to search for previously undetected precursory pre-eruptive signals, in a similar approach to Rouet-Leduc, Hulbert, and Johnson (2018), and provide a path toward estimating the onset of eruptions, although it is important to note that there is little evidence for a continuously emitting source at the Piton de la Fournaise, unlike at the Juan de Fuca plate.

We believe the use of unsupervised learning to analyze seismic signals related to volcanic eruptions provides a complementary approach to this supervised learning approach, which allows users to establish physically meaningful differences in tremor characteristics, which can help to establish transitions between different eruptive states. The interpretation of unsupervised learning results must corroborate with the available data where possible: Here we show that the clusters are coherent temporally. We believe both approaches may provide an effective ML-based workflow for the analysis of eruptive tremor: Sources of consistency throughout the data set can be found by examining the feature importances of a supervised learning classifier, while sources of variation can be isolated by utilizing unsupervised learning.

#### 5. Conclusions

We have shown that both supervised and unsupervised ML techniques provide effective ways to probe volcanic behavior. Using XGBoost, we established the dominant eruptive tremor frequency at the Piton de la Fournaise volcano. By utilizing spectral clustering on features derived from the seismic signal recorded at the BOR seismic station, we show that we can uncover eruptive dynamics in the generated feature space

such as the August 2015 eruption during which a relatively large sustained lava flow was observed and the closing of the eruptive vent during the September 2018 eruption.

### Acknowledgments

C. R., B. R. L., and P. J. acknowledge institutional support (LDRD) at Los Alamos National Laboratory and the U.S. DOE Office of Science, Geoscience, for funding. This research was supported by the Los Alamos National Laboratory (LANL) through its Center for Space and Earth Science (CSES). CSES is funded by LANL's Laboratory Directed Research and Development (LDRD) program under project number 20180475DR. F. B. acknowledges support from the European Research Council under grant no. 817803, FAULTSCAN. We thank Claudia Hulbert, Ian McBrearty, James Theiler, Andrew Delorey, Daniel Trugman, João and Paulo for useful discussions. The data were obtained from the Observatoire Volcanologique du Piton de la Fournaise (OVPF) network. The data used are publicly available and can be found online. The seismic data used for this study were obtained from the Observatoire Volcanologique du Piton de la Fournaise (OVPF) network (<https://www.fdsn.org/networks/detail/PF/>) and can be downloaded from the IGP Data Center (<http://ws.ippg.fr/fdsnws/dataset/1/>), the IRIS Data Management Center (IRISDMC, <http://service.iris.edu/fdsnws/dataset/1/>), or RESIF (<http://ws.resif.fr/fdsnws/dataset/1/>).

### References

- Aminzadeh, F., & Chatterjee, S. (1984). Applications of clustering in exploration seismology. *Geoexploration*, 23(1), 147–159. [https://doi.org/10.1016/0016-7142\(84\)90028-0](https://doi.org/10.1016/0016-7142(84)90028-0)
- Battaglia, J., & Aki, K. (2003). Location of seismic events and eruptive fissures on the piton de la Fournaise volcano using seismic amplitudes. *Journal of Geophysical Research*, 108(B8), 2364. <https://doi.org/10.1029/2002JB002193>
- Battaglia, J., Aki, K., & Ferrazzini, V. (2005). Location of tremor sources and estimation of lava output using tremor source amplitude on the piton de la Fournaise volcano: 1. Location of tremor sources. *Journal of Volcanology and Geothermal Research*, 147(3–4), 268–290. <https://doi.org/10.1016/j.jvolgeores.2005.04.005>
- Battaglia, J., Aki, K., & Staudacher, T. (2005). Location of tremor sources and estimation of lava output using tremor source amplitude on the piton de la Fournaise volcano: 2. Estimation of lava output. *Journal of Volcanology and Geothermal Research*, 147(3–4), 291–308. <https://doi.org/10.1016/j.jvolgeores.2005.04.006>
- Chardot, L., Jolly, A. D., M. Kennedy, B., Fournier, N., & Sherburn, S. (2015). Using volcanic tremor for eruption forecasting at White Island volcano (Whakaari), New Zealand. *Journal of Volcanology and Geothermal Research*, 302, 11–23. <https://doi.org/10.1016/j.jvolgeores.2015.06.001>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 785–794). New York, NY, USA: ACM. <https://doi.org/10.1145/2939672.2939785>
- Chouet, B. (1988). Resonance of a fluid-driven crack: Radiation properties and implications for the source of long-period events and harmonic tremor. *Journal of Geophysical Research*, 93(B5), 4375–4400.
- Chouet, B. A. (1996). Long-period volcano seismicity: Its source and use in eruption forecasting. *Nature*, 380(6572), 309–316. <https://doi.org/10.1038/380309a0>
- Coppola, D., di Muro, A., Peltier, A., Villeneuve, N., Ferrazzini, V., Favalli, M., et al. (2017). Shallow system rejuvenation and magma discharge trends at piton de la Fournaise volcano (La Réunion Island). *Earth and Planetary Science Letters*, 463, 13–24. <https://doi.org/10.1016/j.epsl.2017.01.024>
- De Plaen, R. S. M., Lecocq, T., Caudron, C., Ferrazzini, V., & Francis, O. (2016). Single-station monitoring of volcanoes using seismic ambient noise. *Geophysical Research Letters*, 43, 8511–8518. <https://doi.org/10.1002/2016GL070078>
- Di Lieto, B., Saccorotti, G., Zuccarello, L., Rocca, M. L., & Scarpa, R. (2007). Continuous tracking of volcanic tremor at Mount Etna, Italy. *Geophysical Journal International*, 169(2), 699–705. <https://doi.org/10.1111/j.1365-246X.2007.03316.x>
- Di Muro, A., Métrich, N., Vergani, D., Rosi, M., Armienti, P., Fougereux, T., et al. (2014). The shallow plumbing system of piton de la Fournaise volcano (La Réunion Island, Indian Ocean) revealed by the major 2007 caldera-forming eruption. *Journal of Petrology*, 55(7), 1287–1315.
- Duputel, Z., Lengliné, O., & Ferrazzini, V. (2019). Constraining spatiotemporal characteristics of magma migration at piton de la Fournaise volcano from pre-eruptive seismicity. *Geophysical Research Letters*, 46, 119–127. <https://doi.org/10.1029/2018GL080895>
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Ghahramani, Z. (2004). Unsupervised learning. In O. Bousquet, U. von Luxburg, & G. Rätsch (Eds.), *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2–14, 2003, Tübingen, Germany, August 4–16, 2003, Revised Lectures*, (pp. 72–112). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-28650-9\\_5](https://doi.org/10.1007/978-3-540-28650-9_5)
- 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. <https://doi.org/10.1016/j.jvolgeores.2017.04.015>
- 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. <https://doi.org/10.1126/sciadv.aao2929>
- Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. (2018). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, 12(1), 69–74. <https://doi.org/10.1038/s41561-018-0272-8>
- Jellinek, A. M., & Bercovici, D. (2011). Seismic tremors and magma wagging during explosive volcanism. *Nature*, 470(7335), 522–525. <https://doi.org/10.1038/nature09828>
- Kurokawa, A., Takeo, M., & Kurita, K. (2016). Two types of volcanic tremor changed with eruption style during 1986 Izu-Oshima eruption: Tremor during 1986 Izu-Oshima eruption. *Journal of Geophysical Research: Solid Earth*, 121, 2727–2736. <https://doi.org/10.1002/2015JB012500>
- Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent Individualized Feature Attribution for Tree Ensembles.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*, (pp. 4765–4774). Red Hook, NY. Retrieved from: Curran Associates, Inc. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>
- Malfante, M., Mura, M. D., Metaxian, J., Mars, J. I., Macedo, O., & Inza, A. (2018). Machine learning for volcano-seismic signals: Challenges and perspectives. *IEEE Signal Processing Magazine*, 35(2), 20–30. <https://doi.org/10.1109/MSP.2017.2779166>
- McBrearty, I. W., Delorey, A. A., & Johnson, P. A. (2019). Pairwise association of seismic arrivals with convolutional neural networks. *Seismological Research Letters*, 90(2A), 503–509. <https://doi.org/10.1785/0220180326>
- Michon, L., Ferrazzini, V., Di Muro, A., Villeneuve, N., & Famin, V. (2015). Rift zones and magma plumbing system of piton de la Fournaise volcano: How do they differ from Hawaii and Etna? *Journal of Volcanology and Geothermal Research*, 303, 112–129.
- Ort, M. H., Di Muro, A., Michon, L., & Bachèlery, P. (2016). Explosive eruptions from the interaction of magmatic and hydrothermal systems during flank extension: The Bellecombe tephra of piton de la Fournaise (La Réunion Island). *Bulletin of Volcanology*, 78(1), 5.
- OVPF (2018, June 10). Bulletin d'activité du samedi 6 Octobre 2018 à 15h00 (Heure locale) IINSTITUT DE PHYSIQUE DU GLOBE DE PARIS. Retrieved December 3, 2019, from <http://www.ippg.fr/fr/ovpf/bulletin-dactivite-samedi-6-octobre-2018-a-15-h00-heure-locale>
- Peltier, A., Bachèlery, P., & Staudacher, T. (2009). Magma transport and storage at piton de La Fournaise (La Réunion) between 1972 and 2007: A review of geophysical and geochemical data. *Journal of Volcanology and Geothermal Research*, 184(1–2), 93–108.

- Ren, C. X., Dorostkar, O., Rouet-Leduc, B., Hulbert, C., Strebel, D., Guyer, R. A., et al. (2019). Machine learning reveals the state of intermittent frictional dynamics in a sheared granular fault. *Geophysical Research Letters*, *46*, 7395–7403. <https://doi.org/10.1029/2019GL082706>
- Ross, Z. E., Yue, Y., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2019). PhaseLink: A deep learning approach to seismic phase association. *Journal of Geophysical Research: Solid Earth*, *124*, 856–869. <https://doi.org/10.1029/2018JB016674>
- Rouet-Leduc, B., Hulbert, C., Bolton, D. C., Ren, C. X., Riviere, J., Marone, C., et al. (2018). Estimating fault friction from seismic signals in the laboratory. *Geophysical Research Letters*, *45*, 1321–1329. <https://doi.org/10.1002/2017GL076708>
- Rouet-Leduc, B., Hulbert, C., & Johnson, P. A. (2018). Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nature Geoscience*, *12*(1), 75–79. <https://doi.org/10.1038/s41561-018-0274-6>
- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, *44*, 9276–9282. <https://doi.org/10.1002/2017GL074677>
- Roult, G., Peltier, A., Taisne, B., Staudacher, T., Ferrazzini, V., & Muro, A. D. (2012). A new comprehensive classification of the piton de la Fournaise activity spanning the 1985–2010 period. Search and analysis of short-term precursors from a broad-band seismological station. *Journal of Volcanology and Geothermal Research*, *241–242*, 78–104. <https://doi.org/10.1016/j.jvolgeores.2012.06.012>
- Titos, M., Bueno, A., García, L., Benítez, M. C., & Ibañez, J. (2019). Detection and classification of continuous volcano-seismic signals with recurrent neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(4), 1936–1948. <https://doi.org/10.1109/TGRS.2018.2870202>
- von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416. <https://doi.org/10.1007/s11222-007-9033-z>