# Eruption Forecasting of Strokkur Geyser, Iceland, Using Permutation Entropy

Maria R.P. Sudibyo[1,1,1], Eva Patricia Silke Eibl[1,1,1], Sebastian Hainzl[2,2,2], and Gylfi Pall Hersir[3,3,3]

[1]University of Potsdam
[2]GFZ German Research Centre for Geosciences
[3]retired

November 30, 2022

**Abstract**

A volcanic eruption is usually preceded by seismic precursors, but their interpretation and use for forecasting the eruption onset time remain a challenge. Eruption processes in geysers are similar to volcanoes, but occur more frequently. Therefore, geysers are useful sites for testing new forecasting methods. We tested the application of Permutation Entropy (PE) as a robust method to assess the complexity in seismic recordings of the Strokkur geyser, Iceland. Strokkur features several minute-long eruptive cycles, enabling us to verify in 63 recorded cycles whether PE behaves consistently from one eruption to the next one. We performed synthetic tests to understand the effect of different parameter settings in the PE calculation. Our application to Strokkur shows a distinct, repeating PE pattern consistent with previously identified phases in the eruptive cycle. We find a systematic increase in PE within the last 15s before the eruption, indicating that an eruption will occur. We quantified the predictive power of PE, showing that PE performs better than seismic signal strength or quiescence when it comes to forecasting eruptions.

# Eruption Forecasting of Strokkur Geyser, Iceland, Using Permutation Entropy

**Maria R.P. Sudibyo[1], Eva P.S. Eibl[1], Sebastian Hainzl[2], Gylfi Páll Hersir[3,4]**

[1]University of Potsdam, Institute for Geosciences, Karl-Liebknecht-Str.24/25, Potsdam-Golm
[2]GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany
[3]ISOR, Iceland GeoSurvey, Urdarhvarf 8, Reykjavik, Iceland
[4]Presently gylfi.pall@outlook.com

**Key Points:**

- Permutation Entropy (PE) is a simple tool to assess the complexity of a time series.
- We analyzed the PE evolution for 63 eruptive cycles of Strokkur geyser and found characteristic changes in PE during recharge.
- PE is found to be an useful statistical predictor of the eruption times and highlights the precursor 15 s before eruptions.

Corresponding author: Maria R.P. Sudibyo, `pujiastutisudibyo@uni-potsdam.de`

## Abstract

A volcanic eruption is usually preceded by seismic precursors, but their interpretation and use for forecasting the eruption onset time remain a challenge. A part of the eruptive processes in open conduits of volcanoes may be similar to those encountered in geysers. Since geysers erupt more often, they are useful sites for testing new forecasting methods. We tested the application of Permutation Entropy (PE) as a robust method to assess the complexity in seismic recordings of the Strokkur geyser, Iceland. Strokkur features several minute-long eruptive cycles, enabling us to verify in 63 recorded cycles whether PE behaves consistently from one eruption to the next one. We performed synthetic tests to understand the effect of different parameter settings in the PE calculation. Our application to Strokkur shows a distinct, repeating PE pattern consistent with previously identified phases in the eruptive cycle. We find a systematic increase in PE within the last 15 s before the eruption, indicating that an eruption will occur. We quantified the predictive power of PE, showing that PE performs better than seismic signal strength or quiescence when it comes to forecasting eruptions.



## Plain Language Summary

When a volcano shows the first sign of activity, it is challenging to determine whether and when the actual eruption will occur. Usually, researchers create earthquake lists and locate these events to assess this. However, an alternative and simpler method can be directly applied to continuous seismic data. We tested a method that assesses the complexity of signals. We first created synthetic data to find reasonable parameter settings for this method. While volcanoes do not erupt very often, frequent eruptions at geysers allow us to systematically study and compare several eruptions. We analyzed the continuous record of 63 eruptions of the Strokkur geyser, Iceland. Our results show a distinct pattern that repeats from one eruption to the next one. We also find a clear pattern that indicates about 15 s before the next eruption that an eruption will occur. We show that this method performs better in eruption forecasting than assessing the seismic noise or silence caused by the geyser.


## 1 Introduction

When a volcano becomes restless, it is challenging to assess whether it will lead to an actual eruption and determine the timing of the eruption onset. A magmatic intrusion starting at depth can (i) remain at depth, (ii) stall just before reaching the surface, (iii) erupt in sluggish and viscous extrusion, or (iv) erupt rapidly or explosively (Moran et al., 2011). The process of magma migration involves interactions with the surrounding country rock, cooling magma bodies from previous eruptions, and (or) hydrothermal system (Moran et al., 2008). These interactions generate natural phenomena such as earthquakes, deformation, temperature changes, and gas emissions. These phenomena can be observed by geophysical and geochemical measurements (Moran et al., 2008) and integrated with the history of past eruptions in a framework of eruption forecasting (Whitehead & Bebbington, 2021).

From a seismic point of view, eruptions can show precursors such as accelerating or decelerating earthquake rates. To assess this, monitoring institutes conventionally use methods to tabulate daily event counts (McNutt, 1996) and calculate the average amplitude for a certain window length (Endo & Murray, 1991). The Failure Forecast Method estimates the onset time of eruption by using the rate and the acceleration of seismic precursors associated with the rock failure caused by magma propagation (Boué et al., 2015). However, this method cannot deal with complex precursory signals, e.g., that exhibits fluctuations or deceleration (Boué et al., 2015). Furthermore, due to the uncertainty of the eruption forecast and numerous false alarms (Bell et al., 2013), this method is not recommended to be stand-alone (Whitehead & Bebbington, 2021). Dempsey et al. (2020)

tested a real-time Machine Learning framework to detect eruption precursors of five major eruptions at Whakaari volcano, New Zealand, from 2011 to 2020. This framework derives the information from the seismic amplitude between different frequency bands to assess whether an eruption will occur. A challenge lies in the threshold determination: while increasing the threshold will eliminate false predictions, it leads to missing eruptions and vice versa.

Permutation Entropy, referred to as PE, quantifies the complexity of time series in a simple way, allowing us to characterize the evolution of a dynamic system (Bandt & Pompe, 2002). The calculation of PE relies on the number of permutations appearing in a data series, also known as ordinal pattern, which has been proved to be sensitive in detecting dynamical changes (Cánovas et al., 2011; Cao et al., 2004). PE is widely applied in biomedical science, such as epilepsy detection and prediction, discriminating depth level of anesthesia, and distinguishing heart rate to assess the possibility of heart failure, as has been summarized by Zanin et al. (2012). Although the original algorithm of PE is robust, efforts in PE modification have been made to improve its capability in handling structural changes in different data types. Examples are Tsallis Permutation Entropy to improve the characterization of different stochastic processes (Zunino et al., 2008) and Rényi Permutation Entropy, which uses Rényi Entropy (Rényi, 1960) in the calculation of PE, in order to distinguish rare from frequent events (Zhao et al., 2013). The PE modifications are not only limited to the amplitude information but also concern the signal's phase information, as recently proposed by Kang et al. (2021) as Phase Permutation Entropy.

A robust forecasting framework requires incorporating different forecasting attributes from multiple methods. Testing the application of new methods is important to improve the reliability of the forecasting framework. Glynn and Konstantinou (2016) successfully used the original PE algorithm to detect precursors prior to the 1996 Gjálp eruption. This motivates us to further assess PE's capability and limitation in detecting dynamical changes prior to eruptions.

Geysers are hot springs characterized by intermittent discharge of water that erupts turbulently and is accompanied by a vapor phase (White, 1967). The eruption process of geysers requires magmatism as a heat source, abundant water recharge, and a plumbing system (Hurwitz & Manga, 2017). While the type of liquid and gas phase in geysers differs from the liquid, gas, and solid phase in magma, the fluid is driven to eruption by the gases in both cases. Therefore, the knowledge gained from understanding geyser eruptions might provide useful insights for monitoring volcanic eruptions.

Here, we tested the application of PE for forecasting eruptions at Strokkur geyser, Iceland (Fig. 1a and b). The Strokkur geyser is an ideal site for three reasons: (1) Strokkur features a several-minute long eruptive cycle (Eibl et al., 2021) which allows us to check if PE behaves consistently from one cycle to the next one, (2) the features of the eruptive cycle were already described and interpreted multidisciplinaryly (Eibl et al., 2021) and provide a benchmark for our study, (3) the available instrument network (Fig. 1b) consists of seismometers located at a few meter distance from the geyser's conduit, providing signals with a high signal-to-noise ratio, and seismometers installed at 38.3 to 47.3 m distance, providing a good configuration to test the sensitivity of PE towards station distance.

In this publication, we first introduce the PE method (section 2.1) and perform several synthetic tests to choose the optimum parameters for PE calculations (section 2.2). We also introduce the Receiver Operating Characteristic (ROC) analysis (section 2.3) to assess the predictive power of PE. Then, the methods are applied to eruptions of the Strokkur geyser (section 3 and 4). We compare PE with seismic root-mean-square values (RMS) for one eruptive cycle (section 5.1) and stacked for all available single eruptive cycles (section 5.2). We assess PE for other eruption types (section 5.3) and the change of PE with distance (section 5.4). We discuss how PE relates to the seismic sources migration (section 6.1), the influence of source strength and path effects toward PE(section 6.2) and its predictive power for eruptions at the Strokkur geyser (section 6.3). We conclude that PE detects a clear precursory signal at stations at a few meter distance, making it a promising tool in eruption forecasting.

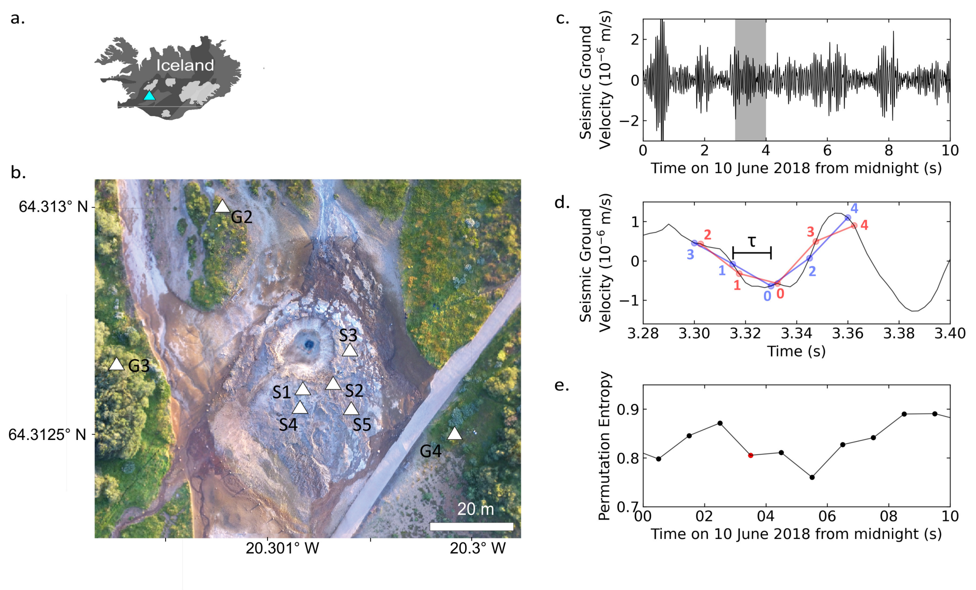


**Figure 1.** Overview of seismic network near Strokkur geyser, Iceland and the calculation of PE. (a) Location of the Strokkur geyser in Iceland (blue triangle) and (b) aerial map where white triangles indicate the location of the seismometers (7L network). (c) 10 s seismogram recorded by the vertical component of station S1. The seismogram is divided into 10 bins of 1 s. The shaded part is related to one of those bins. (d) A closer view of 0.12 s seismic data taken from the shaded window in subfigure (c). The blue and red dot-connecting-lines visualize two consecutive ordinal patterns, {3, 1, 0, 2, 4} and {2, 1, 0, 3, 4} respectively. Each pattern is constructed from five consecutive values selected using $m = 5$ and $\tau = 0.0015\,s$. The length of $\tau$ is visualized as a black horizontal scalebar. (e) The 10 PE values calculated for the consecutive 1 s time window in subfigure (c), where the red dot refers to the PE calculated for the shaded time window in subfigure (c).

## 2 Methods and Synthetic Test

### 2.1 Calculation of Permutation Entropy (PE)

Permutation Entropy is a robust way to quantify the complexity of a time series (Bandt & Pompe, 2002; Zanin et al., 2012; Riedl et al., 2013). This PE method analyzes the probability distribution of ordinal patterns observed in the data (Bandt & Pompe, 2002). An ordinal pattern is a vector representing the relative order of amplitude of the successive samples in a sequence of time series (Bandt & Pompe, 2002; Zanin et al., 2012; Riedl et al., 2013). For example, a sequence of {0.5, 1.0, 3.5, 4.0, 5.7}, based on their amplitude order, is represented as an ordinal pattern of {0, 1, 2, 3, 4} and a sequence of {1.1, 0.8, 0.7, 1.3, 1.0} is represented as an ordinal pattern of {3, 1, 0, 4, 2}.

To construct an ordinal pattern, we basically downsample the time series using an embedding dimension and a delay time. The embedding dimension is the number of samples used to construct an ordinal pattern, i.e., the length of the ordinal pattern, while the delay time is the time gap between the successive samples constructing the ordinal pattern. The ordinal pattern is then defined by a vector of $x_s, x_{s+\tau}, ..., x_{s+(m-1)\tau}$, where $x_s$ is the first sample in the sequence, $m$ is the embedding dimension and $\tau$ is the delay time (Zanin et al., 2012; Riedl et al., 2013). If equal values of amplitude are selected, these values are ranked based on their temporal order (Zunino et al., 2017). To extract all ordinal patterns in a short time window, we continuously shift $x_s$ one sample forward until the last ordinal pattern reaches the end of the window. The PE for the time bin is then calculated as follows:

$$PE = \frac{-1}{\log m!} \sum_{k=1}^{m!} p_k \log p_k \tag{1}$$

where $p_k$ is the probability of the ordinal pattern $k$, and $m$ is the value of the embedding dimension. $p_k$ is estimated by the relative frequency $N_k/N$, where $N_k$ represents the number of recurrences of pattern $k$ and $N$ is the total number of ordinal patterns observed in the time window. The maximum number of different ordinal patterns in a time series signal is $m!$. Equation (1) is normalized with $\log(m!)$ to limit the value of PE to the range of 0 to 1. We then repeat the PE calculation for the next time bin that does not overlap with the previous one until the whole time period of interest is processed, and we can study the PE changes in time.

An example of PE calculated for seismic data of station S1 at Strokkur (see Fig. 1b) is illustrated in Fig. 1c-e. Here, we first divided the seismic time series into 1 s-windows (Fig. 1c), in which the ordinal patterns were extracted using $m = 5$ and $\tau = 0.015\,s$ (Fig. 1d). We define the delay time as the time gap in seconds as we deal with seismic time series that were recorded with different sampling rates. In each 1 s-window, we then estimated the probability distribution of the ordinal patterns and calculated the respective PE value (Fig. 1e).

### 2.2 Synthetic Test of Permutation Entropy

The calculation of PE requires the choice of the delay time, embedding dimension, and the length of time bins (e.g., the shaded window in Fig. 1c). We created several synthetic signals with and without noise to explore the role of these parameters and to define reasonable settings for the PE calculation. The synthetic signals were generated using the basic formula $x(t) = \sin(2\pi f t)$ and a sampling rate of 100 Hz. We set the length of the signals to 20000 s. For all tests, we used delay times $\tau$ ranging from $0.01T_0$ to $T_0$ with a step size of $0.01T_0$, where $T_0 = 1/f$ is the fundamental period of the signal, and embedding dimensions $m$ range from 3 to 9. Since one point cannot create any vectors, and two points can only construct a vector with two possible directions, up and down, $m = 3$ becomes the smallest embedding dimension to assemble ordinal patterns (Zanin et al., 2012). In this test, $m = 9$ was chosen as the upper limit due to the high computational cost. To find out whether the wavelength of the targeted signal should be considered when choosing the window length, we tested 8 different monochromatic signals with different wavelengths. All synthetic tests were performed using Python (Van Rossum & Drake, 2009).

We first tested a pure monochromatic signal with $f = 1$ Hz (Fig. 2a) to evaluate the effect of different delay times and embedding dimensions. We observed that the minimum PE is obtained when the shortest delay time, i.e. $\tau = 0.01$ s, and a delay time $\tau$ close to $T_0$ was used (Fig. 2c). We expected that the minimum PE is obtained when

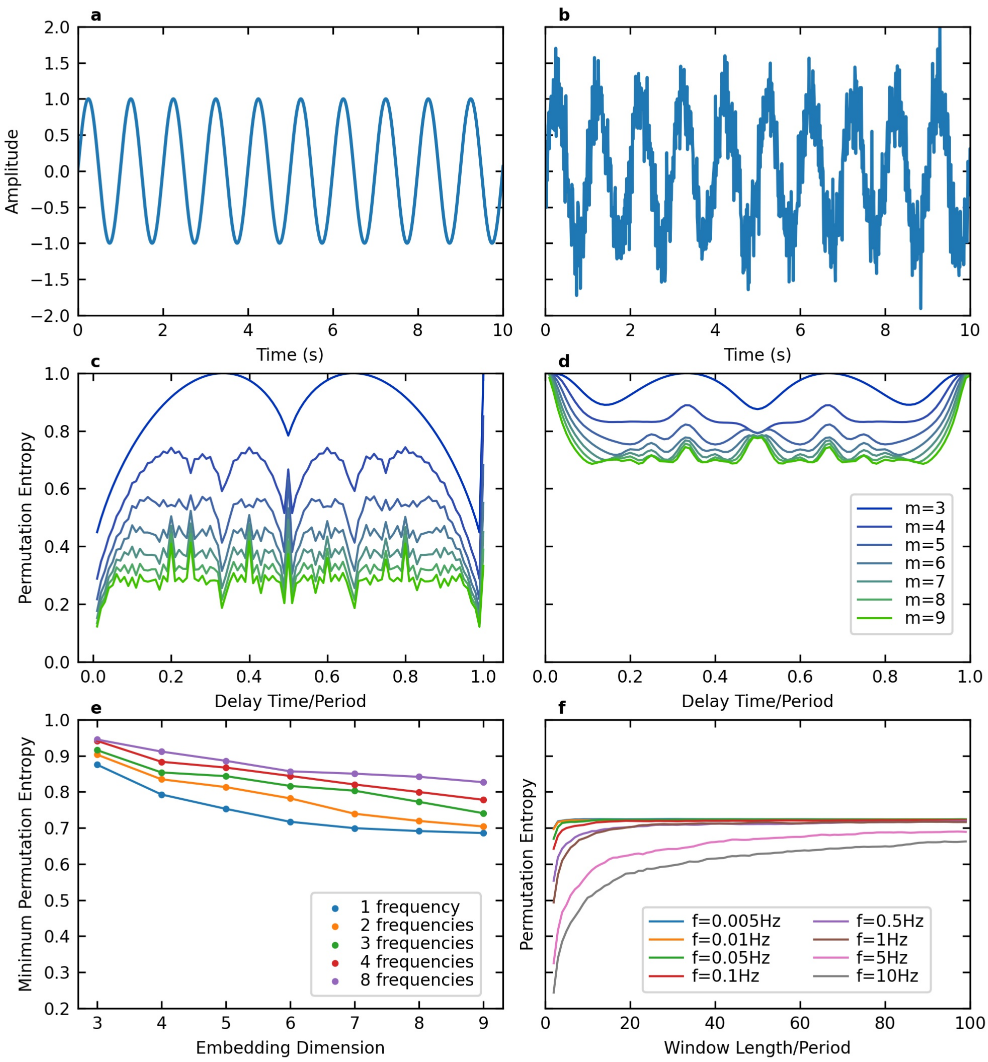


**Figure 2.** Synthetic test for PE calculation. 10 s zoom of the 2000 s synthetic signal with a frequency of $f$=1 Hz (a) without noise, (b) with SNR=5, (c) PE calculated from the signal in subfigure (a) using embedding dimensions $m$ from 3 to 9 and delay times $\tau$ from $0.01\,T_0$ to $T_0$ with step size $0.01\,T_0$. $T_0$=$1/f$ is the period of the signal. (d) Same as subfigure (c) for the signal in subfigure (b), (e) Minimum PE values for 5 synthetic signals, with different complexity and SNR=5, calculated using the same embedding dimensions and delay times as in subfigure (c), (f) PE calculated for 8 different monochromatic signals with frequencies $f$ between 0.005 and 10 Hz using $m$=7 and $\tau$=$0.2/f$. The synthetic signals used for subfigures (e) and (f) are shown in Fig. S1.

using $\tau = T_0$, since the delay time will select equal values of amplitude and construct a repeated ordinal pattern through the window. However, we obtained a very high PE, close to 1 (Fig. 2c) for $\tau = T_0$. After checking the synthetic sine wave constructed using the numpy library (Harris et al., 2020), we found that there are small differences in the order of $10^{(-16)}$ between the amplitudes of the same wave phase, due to the floatingpoint error. While the relative differences between values are negligible, the tiny differences disturb the ranking and create random ordinal patterns, resulting in PE close to 1.

To make the time series more complex, in the next step, we (i) added noise to the signal and (ii) added different frequencies to create different signal types. We quantified the noise level by the signal-to-noise ratio (SNR), defined as the ratio between the variance of signal and noise. The SNR hence can be calculated according to

$$SNR = \frac{\sigma_S^2}{\sigma_N^2} \tag{2}$$

where $\sigma_S$ is the standard deviation of the signal and $\sigma_S$ is the standard deviation of the noise. We used SNR=5 to create noise and added it to the monochromatic signal (Fig. 2b). The analysis of the synthetic signal shows that PE is equal to 1 when calculated using the shortest delay time and delay time equal to $T_0$ (Fig. 2d). We infer that the delay time should not be short nor equal to the fundamental period.

In the next step, we generated four different signals containing two, three, four, and eight frequencies, with and without noise (see Fig. S1 for the detailed information on the frequency content). The PE was calculated using the same delay time and embedding dimension as for the monochromatic signal. The result shows higher PE obtained for the signal containing more frequencies (Fig. 2e and Fig. S1). Similar to the monochromatic signal without noise, the minimum PE is obtained using $\tau = 0.001\,s$ and $\tau$ close to $T_0$, while the signals with noise reach PE close to 1 when using $\tau = 0.001\,s$ and $\tau$ close to $T_0$.

According to the PE result in Fig. 2c and d, and Fig. S1, using a higher embedding dimension will result in a lower PE. To see how the PE changes, we plotted the minimum PE for the monochromatic signal (Fig. 2b) and four different signals in Fig. S1 with SNR=5 in Fig. 2e. The minimum PE is obtained for each embedding dimension, calculated from different delay times ranging from $0.01T_0$ to $T_0$. PE generally converges for each signal, meaning that PE decreases less when using higher embedding dimensions.

Another requirement for PE calculation is that the window length has to accommodate the maximum number of possible ordinal patterns. Additionally, we need to consider the dominant period of the targeted signal. We tested eight different monochromatic signals, with the frequencies $f$ ranging from 0.005 Hz to 10 Hz (see Fig. 2d for the detailed list of frequencies) with SNR=5 and a sampling frequency of 100 Hz. PE was calculated using $m = 7$ and $\tau = 0.2T_0$ (see Fig. 2d). The delay time $\tau = 0.2T_0$ was chosen based on the result in Fig. 2f, where PE is minimum using $\tau = 0.2T_0$. The maximum possible number of different ordinal patterns related to the embedding dimension of 7 is 7! or 5040 ordinal patterns. The PE calculated for the signals with low frequencies, e.g. 0.005 Hz and 0.01 Hz, are stable when the window length is 3 $T_0$. In this case, the signal is much longer than required by $m = 7$. However, the number of points within 3 $T_0$ reduces with increasing signal frequencies given the fixed sampling frequency. Therefore, the signals with frequencies higher than 1 Hz require more than 3 $T_0$ to contain enough samples required by the embedding dimension. In conclusion, the window length should provide enough points for the embedding dimension and be longer than the targeted signal period.

### 2.3 Receiver Operating Characteristic (ROC) Analysis

A well-known method to analyze the ability to predict an event, such as earthquakes or volcanic eruptions (DeVries et al., 2018; Spampinato et al., 2019), is the receiver operating characteristic (ROC) analysis (Fawcett, 2006). ROC analyzes the value of the predictor variable relative to a threshold. Four possible outcomes are possible: If the variable exceeds the threshold and an event (i.e., eruption in our case) follows within the alarm period (the subsequent $N_T$ time steps), it is a hit (true positive, TP); otherwise, it is a false alarm (false positive, FP). If no alarm is raised because the variable is below the threshold, either no event might occur (true negative, TN), or an event occurs (false negative, FN) within the next $N_T$ time steps. In this way, each value of the time series is associated with one of the values TP, FP, TN, or FN, and their counts are calculated for the whole time series. Based on these counts, the true positive rate TPR = TP / (TP + FN) and the false positive rate FPR = FP / (TN + FP) are determined. The ROC curve is finally created by plotting TPR against FPR for threshold values ranging from the minimum to the maximum value of the assessed variable (here, RMS or PE). Both TPR and FPR range between 0 and 1. For quantification, the area under the TPR curve (AUC) is calculated for FPR ranging from 0 to 1. An optimal predictor variable has AUC=1, while the ROC curve of a random variable scatters around the diagonal with AUC≈0.5. We applied this method to our PE and RMS time series, using a time window of 1 s to predict an eruption in the following 1 s window.

## 3 Overview of Instrument Network near Strokkur and Eruption Behaviour of Strokkur

Strokkur geyser is a part of the Geysir geothermal area in the Haukadalur valley in southwest Iceland (Fig. 1). On the surface, Strokkur hosts a water-filled pool of 12 m in diameter (Rinehart, 1986). In the middle of the pool, the uppermost part of the sinter conduit walls extends to the surface (Eibl et al., 2021). This conduit is 2.2 m wide and changes shape and width with depth (Walter et al., 2020). Strokkur features single to sextuple eruptions with one to six water fountains jetting into the air with an average interval of 16.1 s between fountains (Eibl, Hainzl, et al., 2020). Within this manuscript, we only assessed single to quadruple eruptions for which the waiting time after eruptions increases linearly from 3.7 ± 0.9 minutes to 11.3 ± 2.9 minutes (Eibl, Hainzl, et al., 2020).

We used seismic data recorded at 5 to 14 m distance south and east of the pool of Strokkur geyser, Iceland (Eibl, Walter, et al., 2020). The sensors are Nanometrics Trillium Compact Posthole 20 s seismometers at locations S2, S3, S5 and Nanometrics Trillium Compact 120 s at locations S1, S4 (see Fig. 1b) in the 7L seismic network (Eibl, Walter, et al., 2020). The seismometers were installed on 10 June 2018 for 4.5 to 5.25 hours and recorded at a sampling rate of 400 Hz. To assess the sensitivity of PE with respect to station distance from the source, we utilized the seismic data recorded at stations G2, G3, and G4 at a distance of 42.5 m, 47.3 m, and 38.3 m. For the latter stations, no data is available from 10 June, which does not hinder a comparison since the eruptive pattern does not change with time (Eibl, Müller, et al., 2020). The data used are recorded on 3 June 2018 using a sampling rate of 200 Hz.

Based on the same seismic dataset, Eibl et al. (2021) suggested that the conduit is linked to a horizontal crack and a bubble reservoir at 23.7 ±4.4 m depth, where the bubble reservoir extends from about 13 to 23 m west of the conduit and feeds eruptions of Strokkur. Strokkur passes through 4 phases during an eruptive cycle as laid out by Eibl et al. (2021) based on a multidisciplinary experiment (Eibl, Müller, et al., 2020). The illustration of the phases is shown in Fig. 3a).

The cycle starts with Phase 1 (P1), when an eruption is confirmed visually: a rising bubble slug reaches the surface, bursts, and pushes the water and steam upwards into a jetting water fountain. P1 ends when the eruption stops. Due to the water loss in the conduit, the water from the pool and water from a shallow aquifer flow back to refill the conduit. This process is identified as Phase 2 (P2). At the beginning of Phase 3 (P3), the water temperature in the bubble reservoir is low due to the heat loss during the eruption. Seismically, this phase features an eruption coda interpreted as steam entering the reservoir, which partly collapses (Eibl et al., 2021). The collapses release heat and therefore increase the temperature of the water in the bubble reservoir, eventually supporting the gas accumulation toward the end of P3. In Phase 4 (P4), bubbles regularly leave the bubble reservoir, migrate through the horizontal crack, and collapse at a temporal spacing of 21 to 26 s when reaching the water in the conduit that is not hot enough to preserve the steam bubble. With the water in the conduit heating up, the system eventually reaches conditions where steam bubbles burst on the surface, and the next eruption starts (P1).

# 4 Seismic Preprocessing and PE Setting at Strokkur

Previous volcano-seismic studies (Glynn & Konstantinou, 2016; Melchor et al., 2020) used only the vertical component of seismic data to calculate PE. We compared PE using the vertical and both horizontal components (Fig. S2) of the stations S1, S2, S3, S4, and S5. While the PE trends of the three components are generally the same, the vertical component exhibits larger variations in PE. We also checked and compared the seismogram and the spectrogram of the three components. The vertical components of these 5 stations display the largest amplitude. Therefore, we used the vertical components for the following analysis. Station G3 and G4 recorded larger amplitudes on the horizontal components while G2 on the vertical component. The seismic data were detrended, tapered, and instrument corrected to velocity. Afterward, a high pass Butterworth filter of order 4 with a corner frequency of 1 Hz was applied to remove the oceanic microseism.

Based on the eruption catalog compiled by Eibl et al. (2019), there were 63 eruptions recorded on 10 June 2018 from midnight to 04:17 in the morning. These eruptions consisted of 53 single eruptions, 8 double eruptions, one triple eruption, and one quadruple eruption. As the waiting times after eruptions are in the order of minutes, and changes within the cycle occur within less than a second (Eibl et al., 2021), we aim for PE with high temporal resolution. In that case, we need to find the shortest window length possible to calculate PE. We chose a window length of 1 s as it provides a good temporal resolution. The window length needs to contain more samples than the maximum possible $m!$ ordinal patterns constructed from the embedding dimension $m$. In this case, the highest embedding dimension that can be applied for a 1 s window length with a sampling frequency of 400 Hz is 5.

Since the stations are a few meters from the place where the bubbles burst (Fig. 1), the signal-to-noise ratio is high. According to our synthetic test of signals without noise in Fig. 2a, the minimum PE is obtained using the shortest delay time. To confirm this in the real seismic data, we compare five different estimations using small delay times, ranging from 0.0025 s to 0.0125 s (Fig. S3). The PE variations related to these five different delay times exhibit consistent patterns, with a difference in the absolute values. As we are only interested in relative PE changes during the eruptive cycle and not in its absolute values, it is safe to use one of them. In this paper, we present the result of PE using a delay time of 0.005 s.

In addition to PE, we calculated the Root-Mean-Square (RMS) of the ground motion in velocity using the same 1 s long time window. Both quantities will be further evaluated for their performance in eruption forecasting.

# 5 Results

## 5.1 PE and RMS Variation during an Eruptive Cycle

Repetitive patterns of the eruptive cycle for 63 eruptions recorded on 10 June 2018 are visible in seismogram, spectrogram, RMS, and PE. An exemplary single eruption starting at 00:24:39 recorded at station S1 is shown in Fig. 3b-e.

The RMS rises at the beginning of P1 and drops at the end of P1 (Fig. 3d). It stays low during P2 but increases again when P3 starts. In P3, RMS shows a so-called eruption coda composed of seismic peaks at a temporal spacing of 1.5 to 1.7 s featuring a fast increase and a slow decrease in amplitude. The RMS features regular peaks during P4 at an average temporal spacing of 22 to 27 s. Each of these peaks is followed by a weak eruption coda, while the seismic amplitude of the peaks tends to decrease towards the end of P4 (Eibl et al., 2021). The last peak is not followed by an eruption coda.

Fig. 3e exhibits a high PE of 0.89 at the beginning of P1, then increases to the maximum value of 0.94. PE slightly decreases at the start of P2 and suddenly drops towards P3. In P3, PE reaches a minimum value of 0.57, followed by a gradual increase towards P4. At the start of P4, PE reaches a value of 0.81 and sharply drops to 0.60. The following trend then repeats several times: The PE gradually increases to about 0.83 and sharply decreases to about 0.61. In the last 12 s of P4, PE reaches a value of 0.80 and remains high before it increases further and the next eruption (P1) starts. The double, triple, and quadruple eruptions also show similar patterns.

### 5.2 Stacked PE, RMS, and Hypocentral Distances of 53 Single Eruptions

To assess the repetitive pattern of PE and RMS, we stacked the PE and RMS of the 53 cycles, started with a single eruption, according to the start time of each phase. For better visualization, we calculated the mean and the 68% confidence interval (written as mean [lower bound, upper bound]) using a 1 s window. The 68% confidence interval is equivalent to plus/minus one standard deviation for a Gaussian distribution. If the pattern of PE and RMS in each phase is similar from one eruption to another eruption, stacking them will reduce the noise and enhance the pattern.

We aligned the RMS from 55 s before to 50 s after the onset of each phase (Fig. 4a-d). The stacked RMS on each phase shows a clear pattern. At 35 s and 15 s before the onset of P1, two seismic peaks reach the mean RMS of $8.2 \cdot 10^{-7}$ m/s and $9.4 \cdot 10^{-7}$ m/s, respectively. While both peaks are followed by a decrease in seismic amplitude, the second last peak is also followed by a weak eruption coda (Fig. 4a). At the onset of P1, the seismic amplitude increases toward the peak at the mean velocity of 7.9 [3.4, 11]$\cdot 10^{-6}$m/s (Fig. 4a). It drops rapidly to the onset of P2 (Fig. 4b). At the onset of P3, the seismic amplitude increases fast to the mean velocity of 1.2 [0.5, 1.9] $\cdot 10^{-6}$ m/s and slowly decreases towards the end of the phase (Fig. 4c). P4 starts with a sudden peak of mean velocity with a value of 6.7 [3.8, 9.9]$\cdot 10^{-6}$ m/s followed by a weak eruption coda (Fig. 4d).

The stacked PE shows a stable pattern during the different eruptive cycles with different behavior than RMS. Around 35 s before the eruption, we see the last peak reaching a value of 0.78 [0.72, 0.83] in P4. Then the PE value drops to 0.68 [0.59, 0.76] about 27 s before the eruption. Around 15 s before the eruption, the mean of PE reaches a similar value as the last peak of P4. However, instead of decreasing like after the previous peaks, PE remains high for about 6 s and then increases for 8 s to 0.90 [0.88, 0.93] at the start of P1 (Fig. 4e). The PE decreases slightly to P2 and drops to 0.70 [0.61, 0.78] at the beginning of P3 (Fig. 4f-g). PE continues declining for around 3 s to the minimum PE of 0.63 [0.57, 0.68]. After reaching the minimum, PE increases gradually for about 31 s to 0.80 [0.77, 0.82] at the onset of P4 (Fig. 4h). PE then rapidly decreases to 0.63 [0.59, 0.80] for about 8 s after the peak. This pattern repeats several times in P4 before the pattern changes about 14 s before P1.

To investigate the relation between PE and the distance to the source, we calculated the distances from the estimated median source locations (Eibl et al., 2021) to the station S1. S1 is located about 10 m to the south of the conduit on the surface. Eibl et al. (2021) estimated the source location by using the particle motion of the recorded seismic waves. The epicenters of the sources were estimated from the intersection of the azimuth angles derived from all 5 stations. Eibl et al. (2021) project the epicenter location vertically down and extract the source depth from the intersection point with the

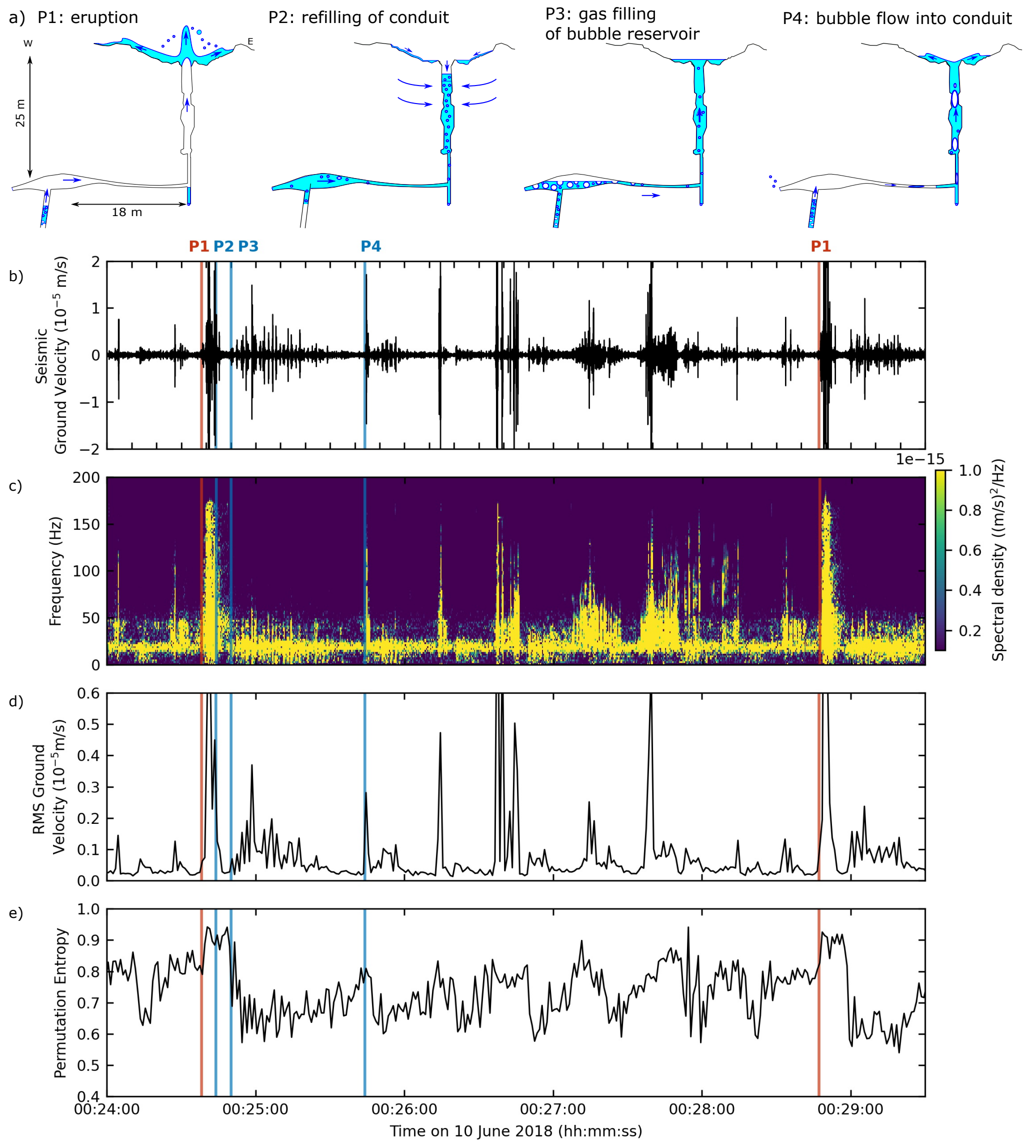


**Figure 3.** A typical eruptive cycle of a single eruption recorded on 10 June 2018. (a) Illustration of the phases of the eruptive cycle at Strokkur modified from Eibl et al. (2021), (b) Seismogram of the vertical component after high pass filtering with a corner frequency of 1 Hz. The two vertical red lines refer to the start of P1, while the blue lines refer to the start of P2, P3, and P4 as illustrated in subfigure (a), (c) Amplitude Spectrogram of subfigure (b) using a time window of 256 samples and overlap of 50 samples, (d) RMS and (e) PE calculated in non-overlapping 1 s long time windows for the seismic data shown in subfigure (b).

derived incidence angles for all stations. Note that the shallow source depths during P1 and peaks in P4 are poorly constrained since the particle motion shows an elliptical particle motion characteristic for Rayleigh waves when the seismic sources reach or approach the surface. We stacked the hypocentral distances from the sources to S1 and calculated their mean and the confidence interval (Fig. 4i-l).

We notice that from 15 s before the eruption, the seismic sources remain at about 10 m depth from the surface or about 20 m away from S1 until the eruption occurs (Fig. 4i).

The source gradually deepens in P2 and reaches a distance of 34 m from S1 (Fig. 4j-k). The sources in P3 are mostly located 13 to 23 m west of the conduit (Eibl et al., 2021), then hypocentral distances decrease toward P4. We checked the source depth and observed that the seismic sources migrate upwards towards the start of P4. P4 starts with seismic sources at a depth of about 10 m with a distance of 21 m to S1. It is likely that the seismic source reached less than 10 m depths during the peaks in P4 (Fig. 4l) and even more during P1, when the eruption occurs on the surface (Fig. 4i).

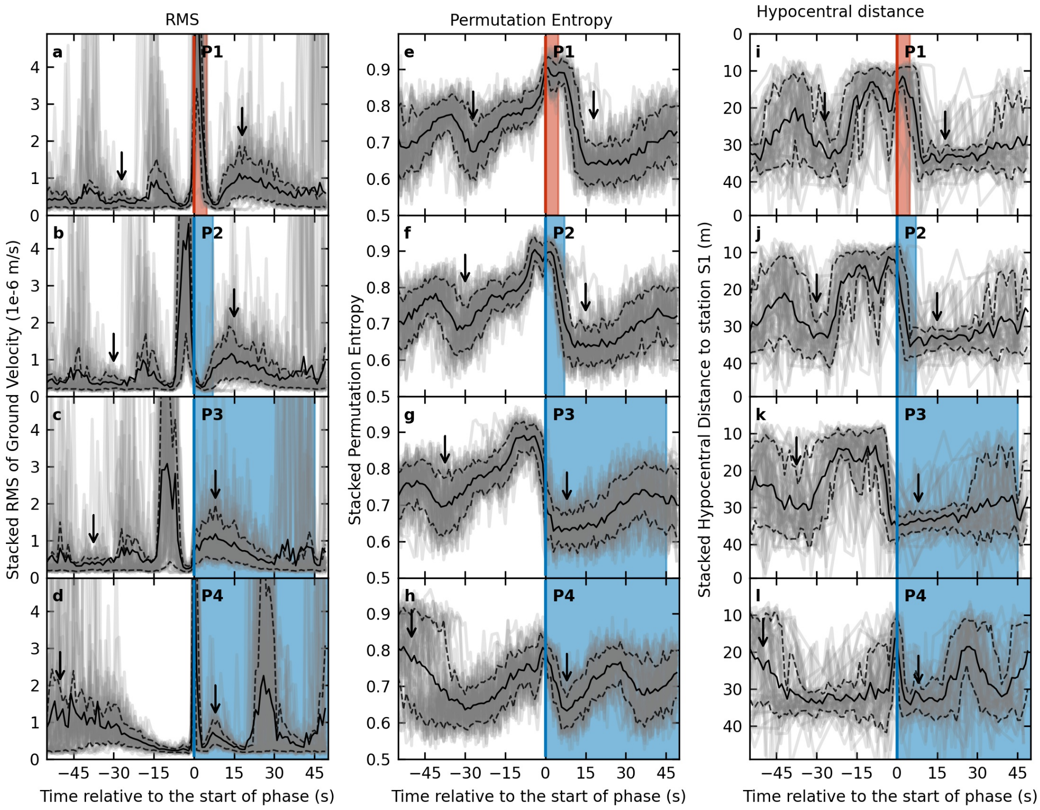


**Figure 4.** Stacked RMS, PE, and hypocentral distance values for the 53 cycles of single eruptions recorded at station S1. Grey lines mark the RMS values for each eruption aligned at (a) the start of the eruption (P1), (b) the end of the eruption (start of P2), (c) the start of the eruption coda (P3), and (d) the start of P4 with regular bubble collapses in the conduit at depth. The time is measured relative to the alignment time (i.e., the start of the red or blue area highlighting the mean duration of the phase). The black lines define the mean values in a 1 s window, while the dashed lines represent the 68% confidence interval. The black arrows point to the seismic eruption coda visible in P3 and P4. (e-l) Same as subfigures (a-d) for (e-h) PE and (i-l) the distance between the seismic source location and station S1 (Eibl et al., 2021).

### 5.3 PE Pattern with Respect to Double to Quadruple Eruptions

We also assessed the PE pattern of 8 double eruptions recorded on 10 June 2018. These eruptions consist of two water fountains at an average temporal spacing of 15.6 s, and the duration of phases P3 and P4 increase linearly with respect to single eruptions (Eibl et al., 2021). The PE pattern of double eruptions throughout the cycle is similar to single eruptions. Its variation is not systematically higher or lower than for single eruptions. While in single eruptions, the PE drastically drops, on average, after 8 s from the beginning of the eruptions, the PE of double eruptions remains high until the second water fountain. PE only drops when entering P3 on average 28 s after the beginning of the first water fountain (Fig. S4).

There was only one triple and one quadruple eruption during the whole recording period. In general, the PE patterns for both triple and quadruple are similar to the single and double eruptions, with PE remaining high in P1 until the last water fountain occurred.

### 5.4 Reliability of PE Results with Respect to Distance from the Source

To evaluate the performance of PE with respect to the station location, we compared the stacked PE variations obtained for the records at stations S1, S2, S3, S4, and S5. We also calculated the variations of the stacked source-station distance for the same stations in the same way. Supplementary Fig. S5 shows that PE is sensitive with respect to the stations' locations. The differences in source distance to each station are small, but the absolute values of PE for different stations are quite distinct. S1, which is located closest to the seismic sources, exhibits the lowest absolute values of PE compared to the other stations. S2, S3, and S4 display a similar temporal variation as S1 but with higher absolute values throughout the cycles. An exception is station S5. While the distance from S5 to the seismic sources is similar to the other stations, the temporal variation of PE does not reflect clearly the changing phases in the eruptive cycle. Overall, the PE at station S5 is dominated by high values except for the first half of P3. The PE in P4 is as high as in P1, making it difficult to see the transition to the eruption in the PE value.

To investigate further the performance of PE at stations with a larger distance, we calculated PE of seismic data recorded at stations G2, G3, and G4 (Fig. 1b) on 3 June 2018. These three stations are located at 42.5 m, 47.3 m, and 38.3 m north-west, west, and south-east of the conduit, respectively. PE values at G2 and G4 are mostly confined between 0.8 and 0.9 and exhibit more random patterns which do not correlate with the eruption phase (see Fig. S6.a and c). However at G3, PE behaves similarly to PE at S1 to S5, even though it is in a lower range and there is no clear transition toward eruptions (see Fig. S6.b).

## 6 Interpretation and Discussion

### 6.1 PE extracting the dynamical information from seismic wave

PE does not depend on the absolute amplitudes, and multiplying a signal by a factor leads to the same PE value. In contrast, PE depends on the frequency bandwidth of the signal. Our synthetic test shows that a synthetic signal containing more frequencies, i.e., by superposing more harmonic signals, produces a higher PE than a signal containing fewer frequencies. We suggest that a signal with a broader frequency content has a higher PE compared to a signal with a narrower frequency band. Dávalos et al. (2021) investigated the effect of bandpass filters such as Butterworth and Chebyshev applied before the PE calculation and observed that lower PE corresponded to narrower bandwidths while higher PE corresponded to broader bandwidths. Our synthetic tests confirm their result.

Our observation at Strokkur shows that PE reaches the highest value during the eruption phase (P1) when the water jets into the air. In this phase, the amplitude peaks and the frequency content is broad. Once the last fountain stops (P2), the amplitude quickly drops and declines to narrower bandwidth. PE is still high at the end of the last fountain but then quickly drops to the next phase (P3). During P3, the eruption coda is composed of seismic peaks at a temporal spacing of 1.5 to 1.7 s. Whilst their frequency content is broad, it is not as broad as during seismic peaks in P1 and P4. Between these peaks in P3, the frequency content of the seismic signal is narrow banded, and the PE fluctuates and reaches minimum values. In P4, during the regular peaks and broad spectrum of the energy produced by the bubble collapses at depth, PE reaches the local maximum. Conversely, PE is smallest directly after the peaks in P4 despite a starting eruption coda that increases in amplitude and widens in frequency content. Shortly before the next peak in P4, it seems seismically quiet and with a narrow-banded frequency content, while the PE value keeps increasing. The PE hence does not solely depend on the broadness of the frequency spectrum.

During P4, the two last bubble collapses at depth in the conduit happen about 35 and 15 s before the start of the next eruption, respectively. Both collapses are recorded as a peak in seismic amplitude and are followed by a drop in seismic amplitude, as seen in the stacked RMS. During these collapses, the PE values reach a local maximum. Following the second last collapse, the PE value drops, while it remains high after the last bubble collapse. We further investigated the waveforms and spectrograms in the last 50 s before the eruption. The second last collapse is followed by a weak eruption coda. This coda is similar to the eruption coda in P3 in terms of the peaks' temporal spacing and frequency content. However, it is smaller in amplitude, and the duration is shorter than in P3. In contrast, the last collapse before the eruption is not followed by an eruption coda. Hence, the RMS value drops to a lower amplitude while the PE value remains high. With respect to the state of the geyser, this implies that the second last bubble collapse triggers recharge in the reservoir, while after the last bubble collapse at depth, the system has reached a state that is ready for eruption. At that stage, the water in the reservoir and conduit is most likely heated sufficiently - without further need to recharge - and contains small bubbles in the whole pipe system. The next large bubble that rises in the conduit can then reach the surface and burst into a jetting water fountain.

Eibl et al. (2021) observed a decrease in seismic peak amplitude during collapses in the conduit with time. They speculate that this is due to damping when more bubbles accumulate in the conduit and decouple the noise from the bubbles and the conduit walls. Here, an increasing amount of bubbles might then suggest that the PE values throughout P4 should increase. While in some eruptions, such a linear increase trend can be observed throughout P4, it is not always the case.

Glynn and Konstantinou (2016) observed an increase of PE for two days between a 5.6 Mw earthquake in Bárðarbunga on 29 September 1996 and the onset of a subglacial eruption in Gjálp on 1 October 1996. This PE increase was preceded by 8 days of PE decrease, which they associated with the lack of frequencies higher than 1 Hz. After the 5.6 Mw earthquake, earthquake swarms migrated to the Gjálp fissures featuring broader frequencies in the range of 0.1 to 9 Hz at station HOT23, located at 8 km distance (Konstantinou et al., 2020). Glynn and Konstantinou (2016) suggested that these higher frequencies increase the complexity, hence causing the PE to increase.

We tested the performance of PE using acceleration derived from the ground velocity and also find an increase in PE. Differentiating velocity into acceleration enhances the energy at higher frequencies. However, we found that the PE values obtained from acceleration are not only larger than PE obtained from velocity but also more confined to a narrower range featuring less variation throughout the eruptive cycle. An example is given in Fig. S7 for PE calculated using $m = 5$ and $\tau = 0.005\,s$ at station S1.

There are two possible reasons why PE obtained from acceleration is less sensitive toward the process inside the geyser than from velocity. First, acceleration enhances the part of the high-frequency signal which is susceptible to the scattering effect from the lateral heterogeneity of the upper crust. This path effect could blur the information of the source mechanism carried by the signal. Second, resolving the complexity of broader spectra requires a higher embedding dimension. In the case of Strokkur, as we aim for 1 s resolution and given the sampling frequency of 400 Hz, the highest embedding dimension ($m$) which we can use is 5.

### 6.2 The influence of source strength and path effects toward the PE performance

We observed that the PE at stations S1, S2, S3, and S4 correlates strongly with the distance between seismic sources and the station. As the seismic sources migrate to the surface and the source-station distance decreases PE increases. First, it should be considered that each phase in the eruptive cycle, which occurs at different depth intervals, is associated with different physical processes (see 3a). Those physical processes might be associated with different PE values. Second, high frequencies are attenuated with distance. If the attenuation eliminates energy and causes the frequency band to become narrower, PE will decrease. However, PE at station S5 exhibits high PE values and less change throughout all phases. Possible reasons are discussed in the following.

The seismic sources, mostly located at average depths of $23.7 \pm 4.4$ m and $9.9 \pm 4.1$ m (Eibl et al., 2021), are subject to the strong attenuation due to the lateral and vertical heterogeneity in Iceland's upper crust (Foulger et al., 2003; Menke et al., 1995). Sato and Fehler (1998) suggested that the particle motion of the P-wave should be linearly polarized if it travels through a path with no or small scattering. When P-wave particle motion is elliptical or even spherical, it indicates strong scattering. Eibl et al. (2021) observed linear particle motions at stations S1 to S4, while station S5 exhibits low linearity. This could suggest that the seismic waves arriving at S5 are much more scattered compared to the other four stations. Scattering attenuation could increase the complexity of the seismic waves due to the superposition between waves in a heterogeneous medium and lead to a more uniform frequency distribution, hence increasing PE.

At larger distances of 38.3 to 47.3 m, the PE performance deteriorates. When the seismic source only releases a small amount of energy, and the distance of the source to the station is large, PE seems to reflect more the filtering of the seismic wave during its propagation to the station. This is also supported by the findings of Eibl et al. (2021), who could not use these stations for the seismic source location due to low-quality particle motions. By contrast, the drop of PE prior to the 5.6 Mw earthquake at Bárðarbunga two days before the 1996 Gjálp eruption, could be detected by stations at a 100 km distance (Glynn & Konstantinou, 2016). This drop is thought to be caused by intrinsic attenuation when hot magma ascended to the upper crust. If the depth of the magma chamber feeding the eruption is estimated to be between 8 and 12 km (Konstantinou et al., 2020), then the seismic sources are located at depths between mid to upper crust. The attenuation at this depth is much lower compared to the uppermost 4 km of crust (Menke et al., 1995). Moreover, the pressurization of magma triggered the 5.6 Mw earthquake. The differences in the source strength and the path effect could explain the performance differences between PE at Strokkur and Bárðarbunga.

### 6.3 Predictive power of PE in comparison to RMS

We used the ROC analysis to quantify the predictive power of PE in comparison to RMS. The resulting curves are shown in Fig. 5 for alarms raised for the next time step when the variables exceed a certain threshold. PE demonstrates good predictive skills with AUC=0.846, while RMS is even worse than random with AUC=0.433. The latter is not surprising, having in mind that RMS tends to decrease prior to eruptions (see Fig. 4a). Thus, we also calculated the inverse of RMS as a measure of quiescence. However, 1/RMS yields AUC=0.567 which is only slightly better than a random forecast.

To rank the predictive power of the PE using only 1 s bin information, we also applied the statistical recurrence model of Eibl, Hainzl, et al. (2020) which was inferred from 20390 waiting times after eruptions of Strokkur geyser in December 2017 and January 2018. The analysis of this long sequence revealed log-normal recurrences with mean and standard deviations dependent on the eruption type of the last event. In particular, we determined the probability $p_T$ of the next event within the alarm time, knowing the time to the last eruption and its eruption style. A detailed description of the calculation of these probabilities is provided in the Appendix. This probability value is found to outperform PE with AUC=0.971. Of course, the comparison is unfair because $p_T$ is based on combined information over a very long time. However, PE can even improve the $p_T$-result if the product of both variables is considered. This result can be understood by considering that $p_T$ is monotonously increasing with increasing time to the last eruption. At the same time, PE is similarly high at intermediate bubble collapses at depth as before the eruptions (see Fig. 3e). The multiplication (shown in the black dashed and continuous lines in Fig. 5) suppresses the high values related to bubble collapses, leading to an enhanced forecast power. This effect is amplified, if the mean ($\langle$PE$\rangle$) value is removed from the PE signal, $\text{PE}_n = (\text{PE} - \langle\text{PE}\rangle)\,\text{H}(\text{PE} - \langle\text{PE}\rangle)$, with H the Heaviside function (H(x)=1 if x>0 and zero else). In this case, the AUC is 0.99, very close to the optimal value of 1.0.

Note that to test the predictive power of PE and RMS, we have only used so far the information in separate 1 s bins of the seismogram. We ignored the information encoded in the time evolution of these parameters. Analyzing the possible improvements using the full PE and RMS patterns requires machine learning techniques and is left for future studies.

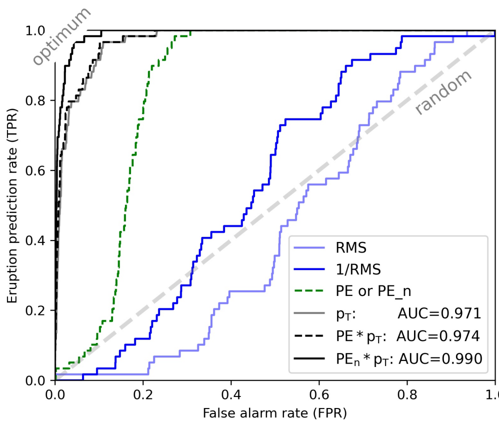


**Figure 5.** Assessing the predictive power of PE using ROC. ROC curves for PE and PE$_n$ (dashed green, note that the PE and PE$_n$ curves are identical), RMS (light blue), the inverse of RMS (blue), and the probability $p_T$ calculated for the recurrence model of Eibl, Hainzl, et al. (2020) (grey), as well as combinations of the latter with PE (solid black and dashed black). Here, PE$_n$ refers to the rescaled PE value, $\text{PE}_n = (\text{PE} - \langle\text{PE}\rangle)\,\text{H}(\text{PE} - \langle\text{PE}\rangle)$, with $\langle$PE$\rangle$ being the mean value of PE and H the Heaviside function. The alarm period is the next time step ($N_T$=1) with the corresponding AUC values given in the legend. The result of a random variable is indicated by the dashed diagonal with AUC=0.5, while the result of an optimal predictor is marked in the upper left corner.

## 7 Conclusions

In this research, we show a good capability of PE in characterizing different phases in the eruptive cycle of the Strokkur geyser. PE also performs better in predicting an eruption than RMS of the ground velocity. About 15 s before the eruption, PE indicates that the system is prone to erupt after the last collapse by increasing values. At the same time, the RMS indicates quiescence, and the seismic sources remain at a shallow depth. The PE reflects the seismic changes linked to a state with superheated water in the pipe system and small bubbles drifting in it. Hence, the PE might be indirectly sensitive to the number of small bubbles present in the water.

PE can characterize the different phases of the geyser's eruptive cycle for the near-field stations, but it seems that PE cannot resolve the dynamics for signals at larger distances. Depending on the signal strength at the source and the signal-to-noise ratio, our results indicate that this method requires seismic data recorded as close to the source as possible, in the case of Strokkur within 15 m. Defining suitable preprocessing steps for PE application on a volcano requires further research. While in a geyser, the interaction between the water and gas with the surrounding rock mostly generates tremors, the interaction between magma and the surrounding rock in a volcano generates more types of volcano-seismic signals with different complexities. For monitoring a volcano, the seismic stations are usually installed at larger distances, which will decrease the signal strength. These factors need to be taken into account. Nonetheless, PE has a strong potential to contribute to the framework of eruption forecasting. For this purpose, our study might help to define distinct precursory features in the temporal variation of PE prior to eruptions that are useful for eruption forecasting.

## Appendix A Eruption probabilities based on the recurrence model of Eibl, Hainzl, et al. (2020)

We calculated the eruption probability for 1 s alarm times using the statistical model of Eibl, Hainzl, et al. (2020). The analysis of 20390 eruptions between December 2017 and January 2018 revealed a log-normal distribution $f_x(t)$ as the probability density function of the inter-eruption times $t$ at Strokkur, where the parameters depend on the type $x$ (single, double, triple, quadruple) of the last eruption. In particular, the mean ($\langle t \rangle$) and standard deviation ($\sigma_t$) of the inter-eruption times are $\langle t \rangle$=3.8, $\sigma_t$=0.8 ($x$=1), $\langle t \rangle$=6.6, $\sigma_t$=1.7 ($x$=2), $\langle t \rangle$=9.5, $\sigma_t$=2.5 ($x$=3), $\langle t \rangle$=12.4, $\sigma_t$=3.4 ($x$=4), $\langle t \rangle$=15.2, $\sigma_t$=4.1 ($x$=5), and $\langle t \rangle$=17.7, $\sigma_t$=4.5 ($x$=6).

Based on those log-normal distributions and knowing the actual waiting time ($t_w$) since the last eruption and its style ($x$), the probability ($p$) for an eruption in the period $[t_1, t_1 + T]$ (with $t_1 \geq t_w$) is calculated according to

$$p_x([t_1, t_1 + T]|t_w) = \frac{\int_{t_1}^{t_1+T} f_x(t)dt}{\int_{t_w}^{\infty} f_x(t)dt} \tag{A1}$$

Note that the denominator is the survival function of $f_x(t)$ for given $t_w$, which is necessary to normalize the distribution for $[t_w, \infty]$.

## Open Research

The seismic data used in this paper are available through GEOFON (Eibl, Walter, et al., 2020) via https://geofon.gfz-potsdam.de/doi/network/7L/2017. The scripts to calculate PE are available at https://gitup.uni-potsdam.de/pujiastutisudibyo/permutationentropy.

## Acknowledgments

We thank the Environment Agency of Iceland and the National Energy Authority for the research permit around Strokkur in the Geysir geothermal area. We thank the rangers and Daniel Vollmer for support and guidance in the field. This work was financially supported by DAAD under grant 57507871. We would like to thank Satoshi Ide, Kostas Konstantinou, and the anonymous associate editor and first reviewer for their constructive reviews that helped improving this manuscript. We declared that this work does not contain any conflict of interests.

## References

Bandt, C., & Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, *73*(17). doi: 10.1103/PhysRevLett.88.174102

Bell, A. F., Naylor, M., & Main, I. G. (2013, 06). The limits of predictability of volcanic eruptions from accelerating rates of earthquakes. *Geophysical Journal International*, *194*(3), 1541-1553. Retrieved from `https://doi.org/10.1093/gji/ggt191` doi: 10.1093/gji/ggt191

Boué, A., Cortés, G., Vallete, B., & G, R. (2015). Real-time eruption forecasting using the material failure forecast method with a bayesian approach. *Journal of Geophysical Research: Solid Earth*, *120*. doi: 10.1002/2014JB011637

Cao, Y., Tung, W.-w., Gao, J. B., Protopopescu, V. A., & Hively, L. M. (2004). Detecting dynamical changes in time series using permutation entropy. *Physical review. E, Statistical, nonlinear, and soft matter physics*, *70*. doi: 10.1103/PhysRevE.70.046217

Cánovas, J. S., Guillamón, A., & Ruiz, M. C. (2011). Using permutations to find the structural changes in time series. *Fluctuation and Noise Letters*, *10*(1). doi: 10.1142/S0219477511000375

Dempsey, D. E., Cronin, S. J., Mei, S., & A.W, K. (2020). Automatic precursor recognition and real-time forecasting of sudden explosive volcanic eruptions at whakaari, new zealand. *Nature Communication*, *11*. doi: 1038/s41467-020-17375-2

DeVries, P. M. R., Viégas, F., Wattenberg, M., & Meade, B. J. (2018). Deep learning of aftershock patterns following large earthquakes. *Nature*, *560*(7720), 632-634.

Dávalos, A., Jabloun, M., Ravier, P., & Buttelli, O. (2021). The impact of linear filter preprocessing in the interpretation of permutation entropy. *Entropy*, *23*. doi: 10.3390/e23070787

Eibl, E. P. S., Hainzl, S., Vesely, N. I. K., Walter, T. R., Jousset, P., Hersir, G. P., & Dahm, T. (2020). Eruption interval monitoring at Strokkur Geyser, Iceland. *Geophysical Research Letters*, *47*, e2019GL085266.

Eibl, E. P. S., Jousset, P., Dahm, T., Walter, T. R., Hersir, G. P., & Vesely, N. I. K. (2019). Seismic experiment at the strokkur geyser, iceland, allows to derive a catalogue of over 70,000 eruptions. *GFZ Data Services.* doi: https://doi.org/10.5880/GFZ.2.1.2019.005

Eibl, E. P. S., Müller, D., Allahbakhsi, M., Walter, T. R., Jousset, P., Hersir, G. P., & Dahm, T. (2020). *Multidisciplinary dataset at the strokkur geyser, iceland, allows to study internal processes and to image a bubble trap.* doi: 10.5880/GFZ.2.1.2020.007

Eibl, E. P. S., Müller, D., Walter, T. R., Allahbakhshi, M., Jousset, P., Hersir, G. P., & Dahm, T. (2021). Eruptive cycle and bubble trap of Strokkur Geyser, Iceland. *Journal of Geophysical Research: Solid Earth*, *126*. doi: 10.1029/2020JB020769

Eibl, E. P. S., Walter, T., Jousset, P., Dahm, T., Allahbakhshi, D., M.and Müller, & Hersir, G. (2020). 1 year seismological experiment at strokkur in 2017/18.

*GFZ Data Services. Other/Seismic Network*. doi: 10.14470/2Y7562610816

Endo, E. T., & Murray, T. (1991). Real-time seismic amplitude measurement (rsam): a volcano monitoring and prediction tool. *Bulletin of Volcanology*, *53*, 533-545. doi: 10.1007/BF00298154

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861-874.

Foulger, G. R., Du, Z., & Julian, B. R. (2003, 11). Icelandic-type crust. *Geophysical Journal International*, *155*(2), 567-590. Retrieved from `https://doi.org/10.1046/j.1365-246X.2003.02056.x` doi: 10.1046/j.1365-246X.2003.02056.x

Glynn, C. C., & Konstantinou, K. I. (2016). Reduction of randomness in seismic noise as a short-term precursor to a volcanic eruption. *Scientific Reports*, *6*(1), 37733. Retrieved from `http://www.nature.com/articles/srep37733` doi: 10.1038/srep37733

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020, September). Array programming with NumPy. *Nature*, *585*(7825), 357–362. Retrieved from `https://doi.org/10.1038/s41586-020-2649-2` doi: 10.1038/s41586-020-2649-2

Hurwitz, S., & Manga, M. (2017). The fascinating and complex dynamics of geyser eruptions. *Annual Review of Earth and Planetary Sciences*, *45*(1), 31-59. doi: 10.1146/annurev-earth-063016-015605

Kang, H., Zhang, X., & Zhang, G. (2021). Phase permutation entropy: A complexity measure for nonlinear time series incorporating phase information. *Physica A: Statistical Mechanics and its Applications*, *568*, 125686. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0378437120309845` doi: https://doi.org/10.1016/j.physa.2020.125686

Konstantinou, K., Utami, I., & Giannopoulos, D. (2020). A reappraisal of seismicity recorded during the 1996 gjálp eruption, iceland, in light of the 2014–2015 bárðarbunga–holuhraun lateral dike intrusion. *Pure and Applied Geophysics*, *177*, 2579–2595. doi: 10.1007/s00024-019-02387-x

McNutt, S. R. (1996). Seismic monitoring and eruption forecasting of volcanoes: A review of the state-of-the-art and case histories. In *Monitoring and mitigation of volcano hazards* (pp. 99–146). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-80087-0_3

Melchor, I., Almendros, J., Carniel, R., Konstantinou, K. I., Hantusch, M., & Caselli, A. (2020). On data reduction methods for volcanic tremor characterization: the 2012 eruption of copahue volcano, southern andes. *Earth, Planets and Space*, *72*, 134. doi: 10.1186/s40623-020-01270-7

Menke, W., Levin, V., & Sethi, R. (1995, 07). Seismic attenuation in the crust at the mid-Atlantic plate boundary in south-west Iceland. *Geophysical Journal International*, *122*(1), 175-182. Retrieved from `https://doi.org/10.1111/j.1365-246X.1995.tb03545.x` doi: 10.1111/j.1365-246X.1995.tb03545.x

Moran, S. C., Freymueller, J. T., LaHusen, R. G., McGee, K. A., Poland, M. P., Power, J. A., ... White, R. A. (2008). *Instrumentation recommendations for volcano monitoring at u.s. volcanoes under the national volcano early warning system* (Tech. Rep.). U.S. Geological Survey.

Moran, S. C., Newhall, C., & Roman, D. C. (2011). Failed magmatic eruptions: late-stage cessation of magma ascent. *Bulletin of Volcanology*, *73*(2). doi: 10.1007/s00445-010-0444-x

Rényi, A. (1960). *On measures of entropy and information* (Vol. 1). Berkeley: University of California Press.

Riedl, M., Müller, A., & Wessel, N. (2013). Practical considerations of permutation entropy: A tutorial review. *The European Physical Journal Special Topics*, *222*(2), 249–262. Retrieved from `http://link.springer.com/10.1140/epjst/e2013-01862-7` doi: 10.1140/epjst/e2013-01862-7

Rinehart, J. S. (1986). Seismic signatures of some icelandic geysers. *Journal of Geo-*

*physical Research*, *73*. doi: 10.1029/JB073i014p04609

Sato, H., & Fehler, M. C. (1998). *Seismic wave propagation and scattering in the heterogeneous earth / haruo sato, michael c. fehler.* New York: AIP Press.

Spampinato, S., Langer, H., Messina, A., & Falsaperla, S. (2019). Short-term detection of volcanic unrest at Mt. Etna by means of a multi-station warning system. *Scientific Reports*, *9*, 6506. doi: 10.1038/s41598-019-42930-3

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual.* Scotts Valley, CA: CreateSpace.

Walter, T. R., Jousset, P., Allahbakhshi, M., Witt, T., Gudmundsson, M. T., & Hersir, G. P. (2020). Underwater and drone based photogrammetry reveals structural control at geysir geothermal field in iceland. *Journal of Volcanology and Geothermal Research*, *391*. doi: 10.1016/j.jvolgeores.2018.01.010

White, D. E. (1967). Some principles of geyser activity, mainly from steamboat springs, nevada. *American Journal of Science*, *265*(8), 641–684. Retrieved from `https://www.ajsonline.org/content/265/8/641` doi: 10.2475/ajs.265.8.641

Whitehead, M. G., & Bebbington, M. S. (2021). Method selection in short-term eruption forecasting. *Journal of Volcanology and Geothermal Research*, *419*. doi: 10.1016/j.jvolgeores.2021.107386

Zanin, M., Zunino, L., Rosso, O. A., & David, P. (2012). Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy*, *14*, 1553-1577. doi: 10.3390/e14081553

Zhao, X., Shang, P., & Huang, J. (2013, may). Permutation complexity and dependence measures of time series. *EPL (Europhysics Letters)*, *102*(4), 40005. Retrieved from `https://doi.org/10.1209/0295-5075/102/40005` doi: 10.1209/0295-5075/102/40005

Zunino, L., Olivares, F., Scholkmann, F., & Rosso, O. A. (2017). Permutation entropy based time series analysis: Equalities in the input signal can lead to false conclusions. *Physics Letters A*, *381*(22), 1883-1892. doi: 10.1016/j.physleta.2017.03.052

Zunino, L., Pérez, D., Kowalski, A., Martín, M., Garavaglia, M., Plastino, A., & Rosso, O. (2008). Fractional brownian motion, fractional gaussian noise, and tsallis permutation entropy. *Physica A: Statistical Mechanics and its Applications*, *387*(24), 6057-6068. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0378437108006043` doi: https://doi.org/10.1016/j.physa.2008.07.004

# Supporting Information for "Eruption Forecasting of Strokkur Geyser, Iceland Using Permutation Entropy"

Maria R.P. Sudibyo[1], Eva P.S. Eibl[1], Sebastian Hainzl[2], Gylfi Páll Hersir[3,4]

[1]University of Potsdam, Institute for Geosciences, Karl-Liebknecht-Str.24/25, Potsdam-Golm

[2]GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany

[3]ISOR, Iceland GeoSurvey, Urdarhvarf 8, Reykjavik, Iceland

[4]Presently gylfi.pall@outlook.com

## Contents of this file

1. Figures S1 to S7

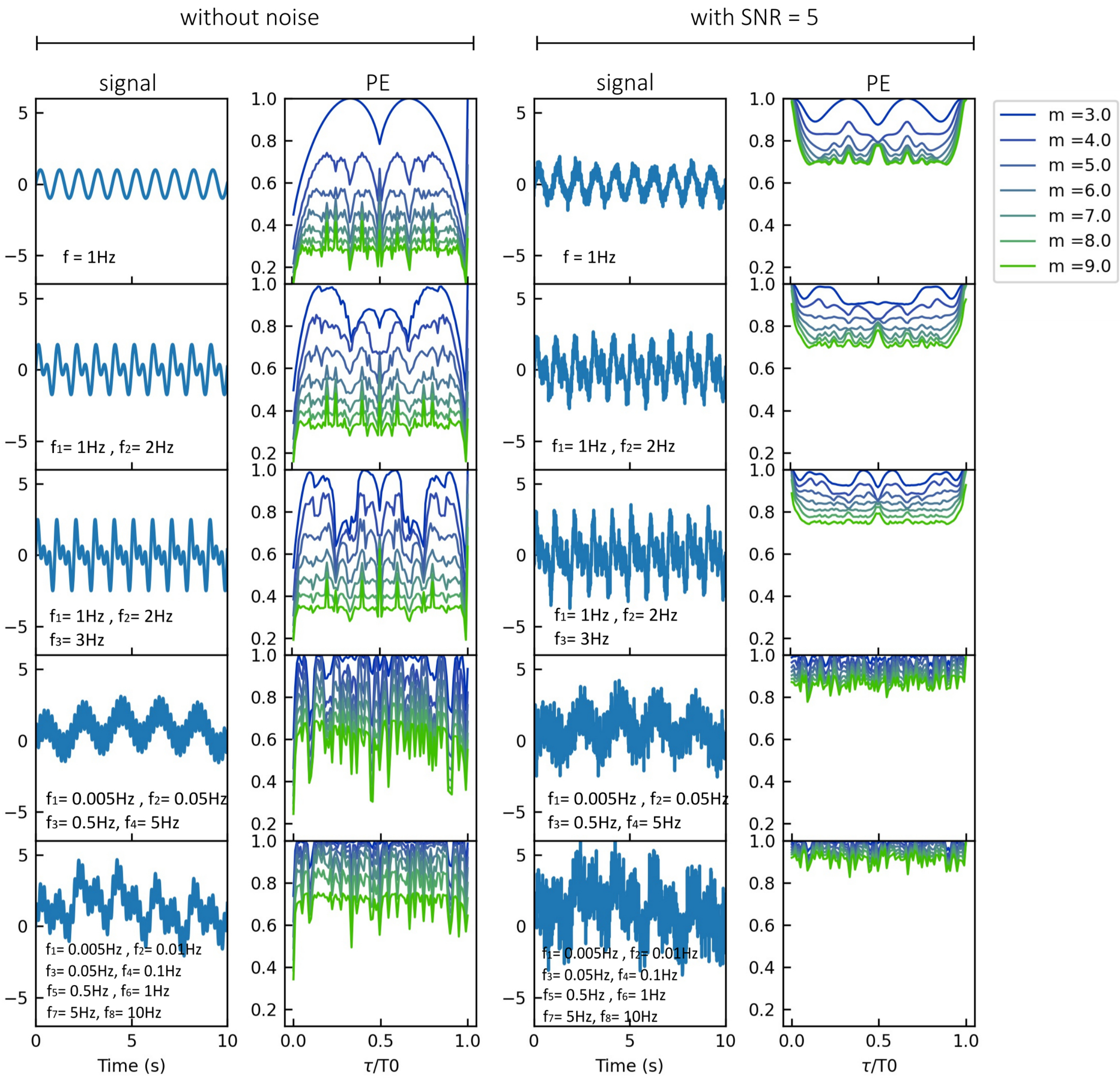


**Figure S1.** Synthetic test for the dependence of PE on the delay time using five different types of signals, with and without noise, and an embedding dimension ranging from 3 to 9. The frequencies used to create the signals are listed in the figure, the delay time ranges from $1/T_0$ to $T_0$, where $T_0$ is the fundamental period of the signal.

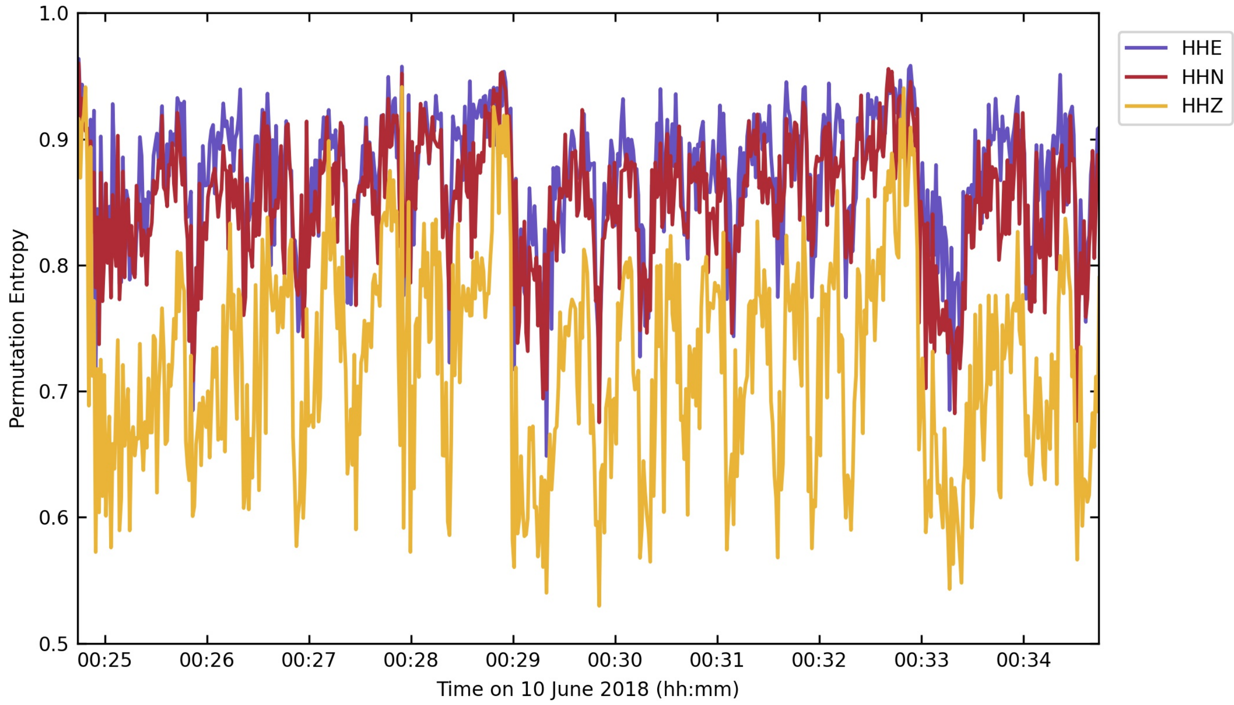


**Figure S2.** PE calculated for the horizontal and the vertical components at station S1, using embedding dimension 5 and a delay time of 0.005 s.

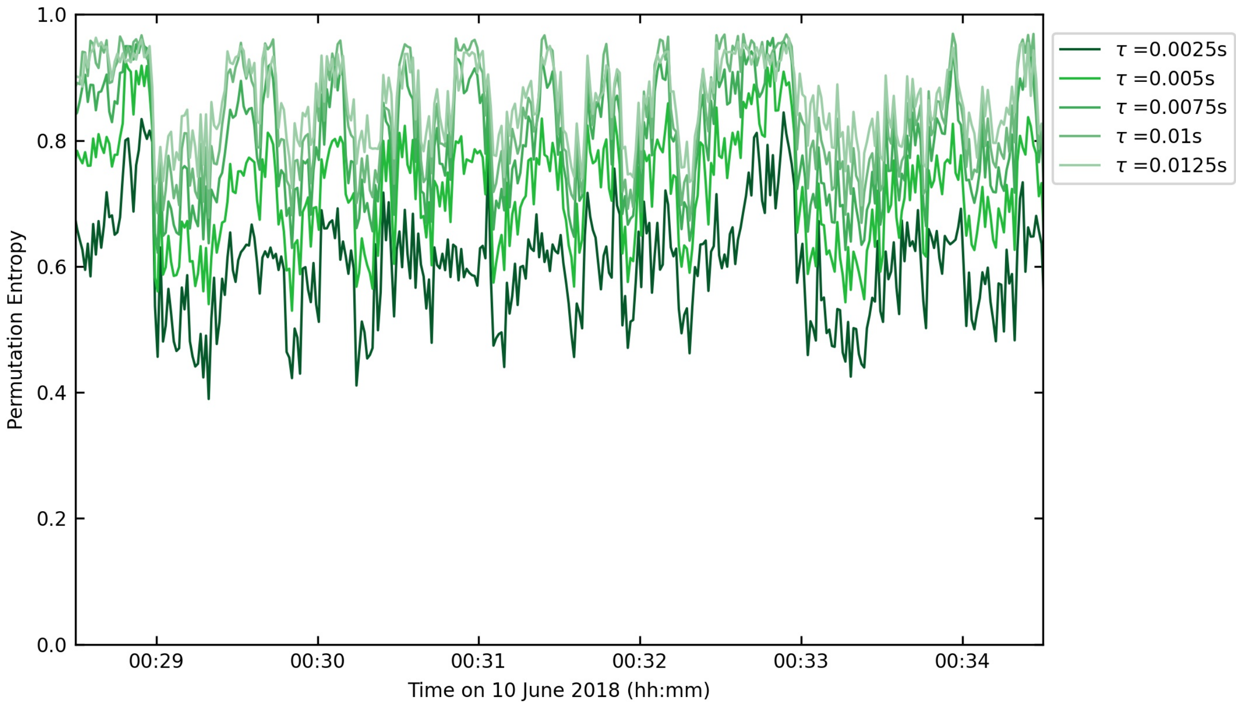


**Figure S3.** PE calculated for seismogram recorded at station S1 for the vertical component using an embedding dimension of 5 and delay times between 0.0025 s to 0.0125 s.

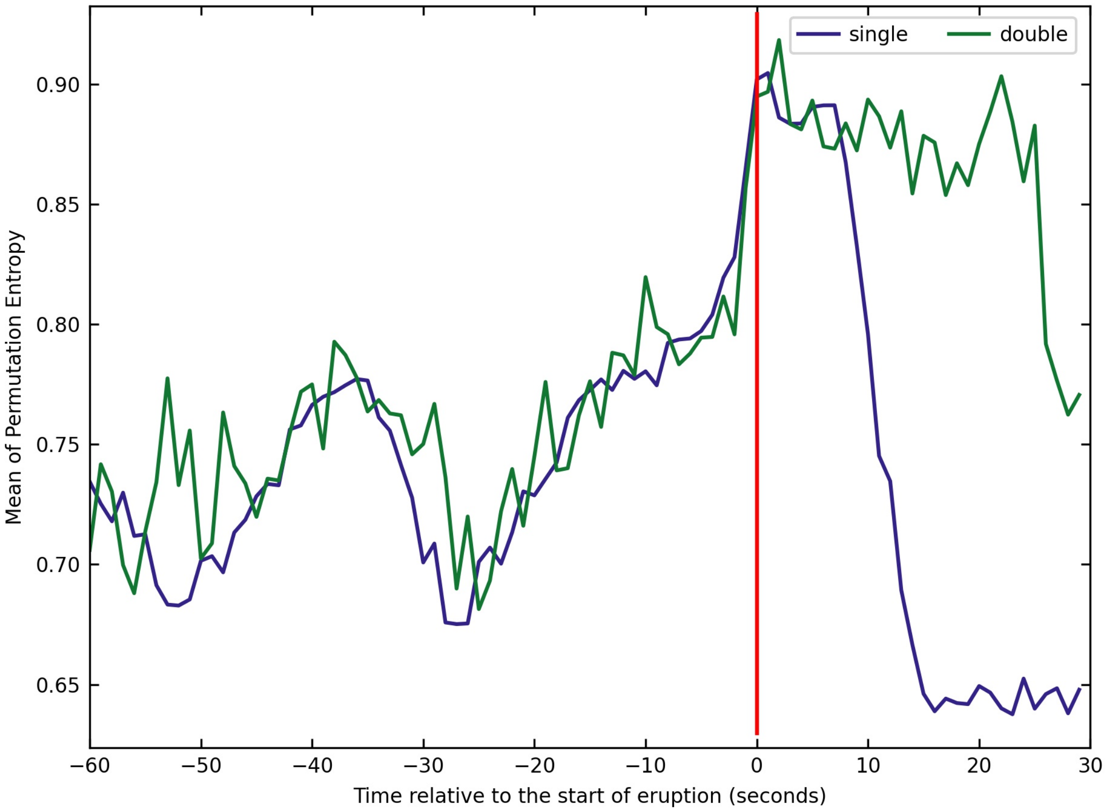


**Figure S4.** Mean of stacked PE aligned at the start of the eruption (red vertical line) related to the 53 single eruptions (blue) and 8 double eruptions (green).

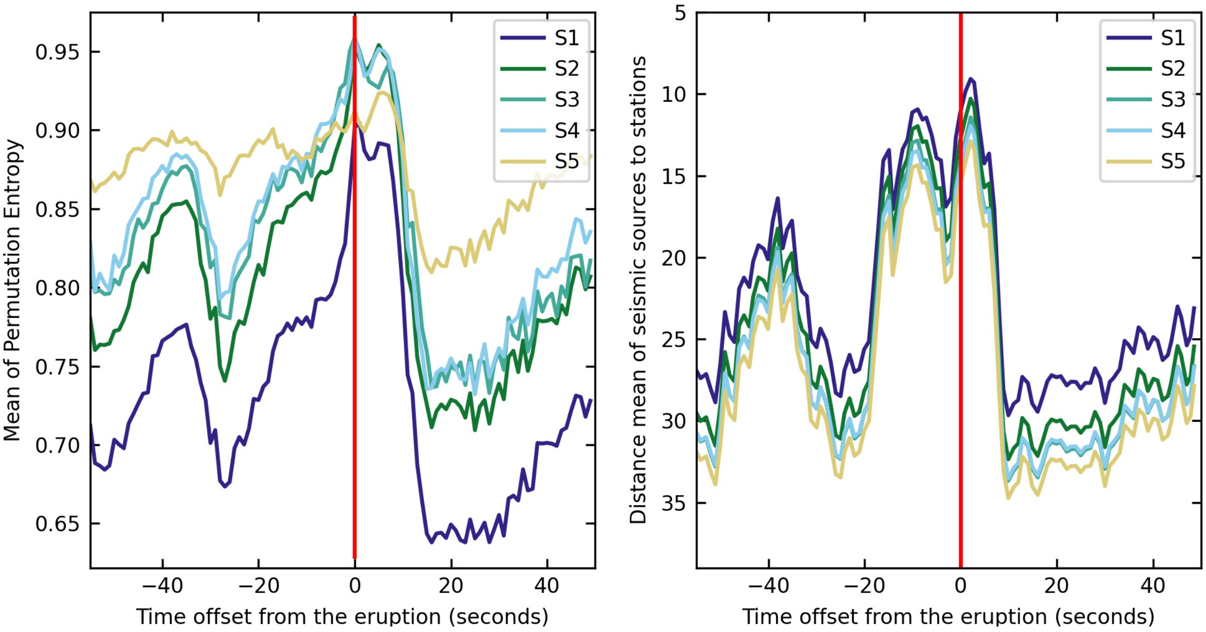


**Figure S5.** The mean of stacked PE (left) and source-station distance (right) aligned at the start of the 53 single eruptions calculated for five different stations: S1 (dark blue), S2 (green), S3 (light green), S4 (light blue), and S5 (yellow).

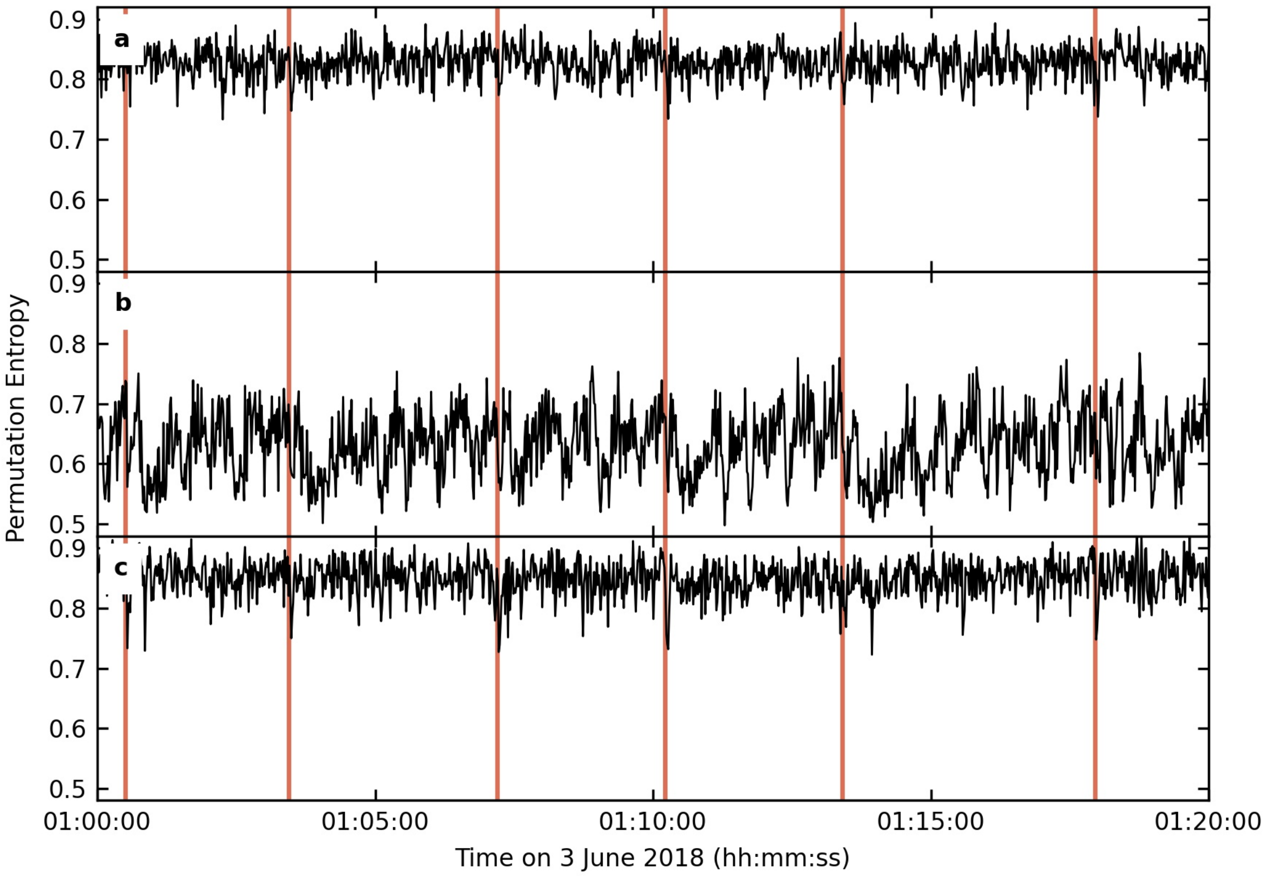


**Figure S6.** 20 minutes PE variation of station (a) G2, (b) G3, and (c) G4. The vertical lines represent the times of eruptions.

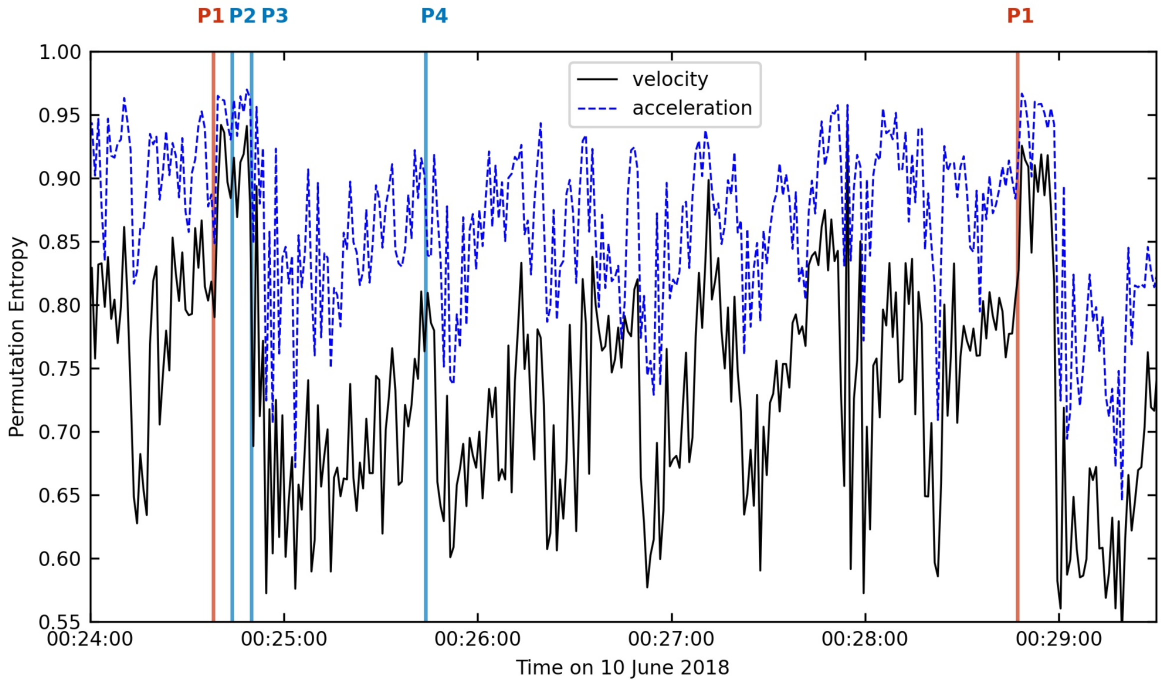


**Figure S7.** Comparison between PE estimated from seismic velocity (black line) and seismic acceleration (dashed blue) at S1. The vertical red lines represent the onset of eruptions (P1), blue lines represent the end of eruption (P2), steam entering the bubble trap (P3), and bubble collapses at depth (P4).