

Volcanic precursor revealed by machine learning offers new eruption forecasting capability

Kaiwen Wang¹, Felix Waldhauser², Maya Tolstoy³, David P. Schaff⁴, Theresa Sawi⁵, William S. D. Wilcock⁶, and Yen Joe Tan⁷

¹Lamont-Doherty Earth Observatory, Columbia University

²Lamont-Doherty Earth Observatory of Columbia University

³School of Oceanography, University of Washington

⁴Lamont-Doherty Earth Obs

⁵Lamont-Doherty Earth Observatory

⁶University of Washington

⁷Earth and Environmental Sciences Programme, Faculty of Science, The Chinese University of Hong Kong

March 07, 2024

Abstract

Seismicity at active volcanoes provides crucial constraints on the dynamics of magma systems and complex fault activation processes preceding and during an eruption. We characterize time-dependent spectral features of volcanic earthquakes at Axial Seamount with unsupervised machine learning methods, revealing mixed frequency signals that emerge from the continuous waveforms about 15 hours before eruption onset. The events migrate along pre-existing fissures, suggesting that they represent brittle crack opening driven by influx of magma or volatiles. These results demonstrate the power of novel machine learning algorithms to characterize subtle changes in magmatic processes associated with eruption preparation, offering new possibilities for forecasting Axial's anticipated next eruption. This novel method is generalizable and can be employed to identify similar precursory signals at other active volcanoes.

1 **Volcanic precursor revealed by machine learning offers**
2 **new eruption forecasting capability**

3 **Kaiwen Wang¹, Felix Waldhauser¹, Maya Tolstoy^{1,2}, David Schaff¹, Theresa**
4 **Sawi¹, William Wilcock², Yen Joe Tan³**

5 ¹Lamont-Doherty Earth Observatory, Columbia University

6 ²School of Oceanography, University of Washington

7 ³Earth and Environmental Sciences Programme, The Chinese University of Hong Kong

8 **Key Points:**

- 9 • Unsupervised learning separated regular earthquakes and precursory mixed fre-
10 quency earthquakes (MFEs) based on different spectral patterns
- 11 • The regular earthquakes have strong tidal modulation, corresponding to failures
12 on the caldera ring faults triggered by tidal stress changes
- 13 • The MFEs emerge 15 hours before eruption and migrate along pre-existing fissures,
14 likely associated with eruption preparation processes

Corresponding author: Kaiwen Wang, kw2988@columbia.edu

Abstract

Seismicity at active volcanoes provides crucial constraints on the dynamics of magma systems and complex fault activation processes preceding and during an eruption. We characterize time-dependent spectral features of volcanic earthquakes at Axial Seamount with unsupervised machine learning methods, revealing mixed frequency signals that emerge from the continuous waveforms about 15 hours before eruption onset. The events migrate along pre-existing fissures, suggesting that they represent brittle crack opening driven by influx of magma or volatiles. These results demonstrate the power of novel machine learning algorithms to characterize subtle changes in magmatic processes associated with eruption preparation, offering new possibilities for forecasting Axial's anticipated next eruption. This novel method is generalizable and can be employed to identify similar precursory signals at other active volcanoes.

Plain Language Summary

Our research used observations of small earthquakes to understand the dynamic behaviors of magma and fault systems before and during a volcano eruption. Specifically, we used machine learning techniques to search for signature waveform patterns that may inform us of their associated physical processes. At Axial Seamount, an active underwater volcano, we discovered distinct patterns in earthquake signals preceding and during the 2015 eruption. Based on event spectral patterns, we identified signals of mixed-frequency earthquakes that emerge about 15 hours before the eruption starts and migrate along pre-existing eruptive fissures. The spectral pattern involves a mixture of low frequency energy following the first arrivals, which we interpret to represent opening of cracks and being filled with magma or gases. Our study demonstrates that we can use machine learning algorithms to detect subtle changes in volcanic signals and help us better understand the processes leading up to an eruption. This may help us in forecasting Axial's upcoming eruption and can possibly be applied to other active volcanoes too.

1 Introduction

Seismic observations can provide important constraints on the structure of a volcano and its dynamic behavior in volcanic cycles (Wilcock et al., 2016; Wilding et al., 2023; Tan et al., 2019; Gudmundsson et al., 2016). Questions remain on how magma moves in the subsurface preceding an eruption and how soon before an eruption this process

66 begins. Recent advances in unsupervised machine learning methods (Holtzman et al.,
67 2018; Cotton & Ellis, 2011; Holtzman et al., 2021; Sawi et al., 2022; Yoon et al., 2015;
68 Seydoux et al., 2020; Jenkins et al., 2021) offer the opportunity to mine large waveform
69 archives to find subtle differences in the spectral content of seismic signals. These dif-
70 ferences can be interpreted with respect to changes in source characteristics and the volcano-
71 tectonic processes that drive brittle failure, providing a time-dependent image of phys-
72 ical processes that lead up to an eruption.

73 Axial Seamount is a well-instrumented, active submarine volcano on the Juan de
74 Fuca Ridge (Figure 1) with a long record of geophysical data that covers the last three
75 eruptions in 1998, 2011, and 2015 (Wilcock et al., 2018, 2016; Nooner & Chadwick, 2016),
76 including documentation of the eruptive fissures and lava flows of the recent 2015 erup-
77 tion (Chadwick et al., 2016), and 3-D images of its shallow magma chamber (Arnulf et
78 al., 2014). Five months before the most recent eruption in April 2015, seismicity at Ax-
79 ial Seamount has been recorded by a local, cabled, 7-station ocean bottom seismome-
80 ter (OBS) network operated in real-time by the Ocean Observatories Initiative (OOI)
81 (Kelley et al., 2014). The OBS array recorded signals from a variety of sources (Wilcock
82 et al., 2016). Here we apply unsupervised machine learning methods to the 4 months be-
83 fore and during the 2015 eruption to find precursory signals with distinct frequency con-
84 tent.

65 **2 Supervised and Unsupervised ML**

66 We combined supervised machine learning (ML) techniques (Zhu & Beroza, 2019;
67 Zhu et al., 2022) with cross-correlation-based, high-resolution earthquake relative loca-
68 tion methods (Waldhauser & Ellsworth, 2000; Waldhauser et al., 2020; Lomax et al., 2000,
69 2009) to develop a catalog of 240,000 earthquakes ($M = -1.74$ to 3.45) for Axial Seamount
70 from 2014 to 2021. The new earthquake catalog illuminates the caldera ring faults and
71 the fissures that were active during the previous eruptions (Figure 1) (Wilcock et al., 2016;
72 Waldhauser et al., 2020).

73 We then apply an unsupervised machine learning method (SpecUFEx, Holtzman
74 et al., 2018) to the 4 months of pre-eruption data to characterize spectral patterns in the
75 waveforms. SpecUFEx is an unsupervised spectral feature extraction algorithm origi-
76 nally developed using ML methods for audio pattern recognition (Cotton & Ellis, 2011)

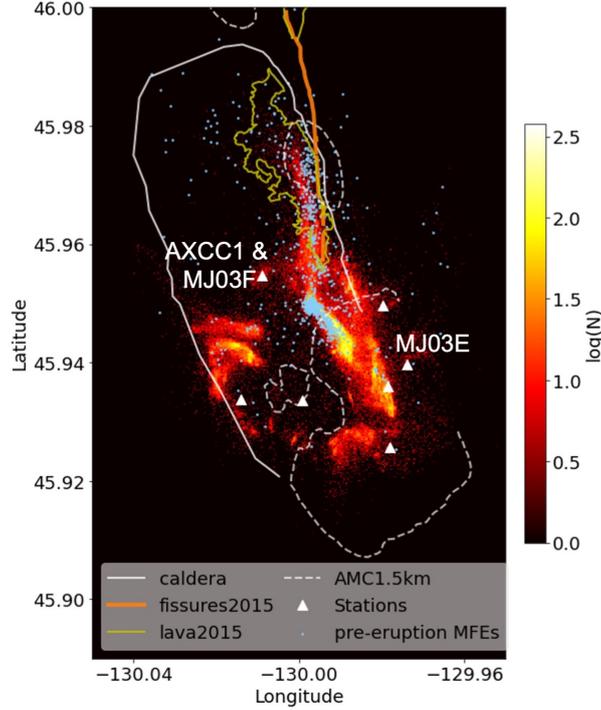


Figure 1. Heatmap of earthquake density at Axial Seamount from Nov 2014 to Dec 2021. Mixed-frequency earthquakes (MFEs) one day before the eruption are shown in light blue dots. Also shown are the caldera rim (white solid line), the 1.5 km depth contour of the Axial magma chamber (AMC) (dashed white line), eruptive fissures (orange lines), and lava flows (yellow lines) of the 2015 eruption and the OBS array (white triangles). The heatmap shows the number of earthquakes in each bin (bin size $25\text{m} \times 25\text{m}$).

77 and has been later adapted to characterize seismic waveforms of earthquakes (Holtzman
 78 et al., 2018), acoustic emissions (Holtzman et al., 2021), icequakes (Sawi et al., 2022),
 79 and repeating earthquakes (Sawi et al., 2023). It takes event spectrograms as input and
 80 applies nonnegative matrix factorization (NMF) and hidden Markov models (HMM) to
 81 reduce the dimensionality of the spectral features and remove features that are common
 82 to all signals. For each earthquake, we calculate a fingerprint matrix by counting the num-
 83 ber of state transitions in the state sequence matrix from the HMM output. The finger-
 84 prints are condensed representations of the original earthquake spectrograms while still
 85 keeping their time-dependent spectral information. We further compress the fingerprints
 86 by principal component analysis (PCA) and finally apply K-means clustering to iden-
 87 tify earthquake clusters that have common spectral features (Holtzman et al., 2018). We
 88 focus our analysis on the 4 months of seismicity leading up to the eruption on April 24,

89 2015. We use waveform data from broadband OBS station AXCC1 and learned the fea-
 90 ture dictionary on a representative subset of $\sim 9,000$ events in the week before the erup-
 91 tion. We then use the learned dictionary to calculate features of the ~ 4 months of seis-
 92 micity starting from the beginning of 2015 until the eruption onset.

93 **3 Spectral differences**

94 K-means clustering of the fingerprints separates the events into two main groups
 95 with small but distinct differences in spectral features in the waveforms between the groups
 96 (Figure 2A and 2D). To investigate which characteristic spectral features might contribute
 97 to the separation of the two earthquake clusters, we examine the representative patterns
 98 of the condensed fingerprints. By stacking the top 100 representative fingerprints in each
 99 cluster (Figure 2B and 2E), we identify the active states (bright spots in stacked finger-
 100 prints). These active states are the characteristic features that define the spectral fea-
 101 ture space. We project these characteristic features in the fingerprints back onto the HMM
 102 and NMF mappings (emissions matrix in Figure S3A and spectral dictionary in Figure
 103 S3B) to solve for their frequency-dependent sensitivity kernel (Figure S3C). Comparing
 104 the frequency dependency of the characteristic features in the two clusters, we find that
 105 one cluster has events with lower frequency content coming in shortly (~ 1 s) after the
 106 P-arrival. Thus we define the earthquakes in this cluster as mixed-frequency earthquakes
 107 (MFEs) and the events in the other cluster as regular earthquakes (EQs). The spectral
 108 differences can also be seen in the stacked spectrograms (Figure 2C and 2F) of the top
 109 100 representative events and their waveforms (Figure 2A and 2D).

110 **4 Spatio-temporal distribution**

111 The separation based on spectral characteristics reveals differences in the spatiotem-
 112 poral evolution of the earthquakes in the two groups (Figure 3, Movie S1). Approximately
 113 24 hours prior to the eruption, the MFEs start lighting up the eastern margin of the caldera
 114 along the southern segment of the eruptive fissures (Figure 1). These MFEs locate close
 115 to the roof of the Axial magma chamber (~ 1.5 km; 15). ~ 15 hours before the eruption,
 116 a distinct burst of MFEs migrates from the caldera center northward along the eastern
 117 margin of the caldera at a speed of 4.4 km/h (arrow in Figure 3B). The peak hourly mo-
 118 ment release of the MFEs during that burst is about two orders of magnitude above back-
 119 ground, and 40 times that released by all regular earthquakes in the same period. Af-

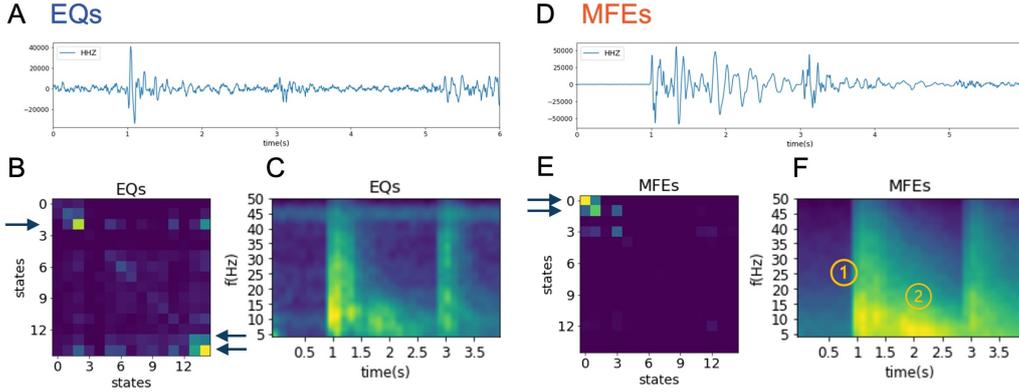


Figure 2. Spectral characteristics of events in the two main clusters. Waveform examples of one representative event in each cluster (A and D), stacked fingerprints (B and E), and stacked spectrograms (C and F) using 100 events in each cluster. The stacked fingerprints and stacked spectrograms are color scaled by their maximum value. ① and ② in F mark the impulsive P arrival and the low-frequency tail.

120 ter this initial burst, the MFE activity subsides for a couple of hours and then, about
 121 3.5 hours before the eruption, a second burst of MFES occurred, reversing the path of
 122 the previous burst and migrating southward (at a speed of 1.1 km/hr) and eventually
 123 upward towards the location where the lava first erupted on the seafloor (Wilcock et al.,
 124 2016) (Figure 3B). After that point and for the next hour, the MFES spread out across
 125 the entire fault system during the course of the eruption. The second MFE burst is char-
 126 acterized by a steep increase in seismic moment release starting about 4 hours and peak-
 127 ing 1 hour before the eruption onset. Peak hourly moment release is about 30 times that
 128 of the first burst, while the moment release from regular earthquakes leading up to the
 129 eruption is comparably insignificant. Once the eruption starts, MFE moment release con-
 130 tinuously decreases, while that from regular earthquakes increases.

131 Different from the MFES, the regular earthquakes locate primarily in the south-
 132 ern part of the caldera (Figure S5). They occur on both the eastern and western walls
 133 of the ring fault, which suggests that the spectral fingerprints are not sensitive to event
 134 location relative to the seismic station. The regular earthquake cluster also includes events
 135 during the pre-eruption inflation period as well as the rapid deflation period after the
 136 eruption started, that is, when the fault slip motion on the caldera ring faults reversed

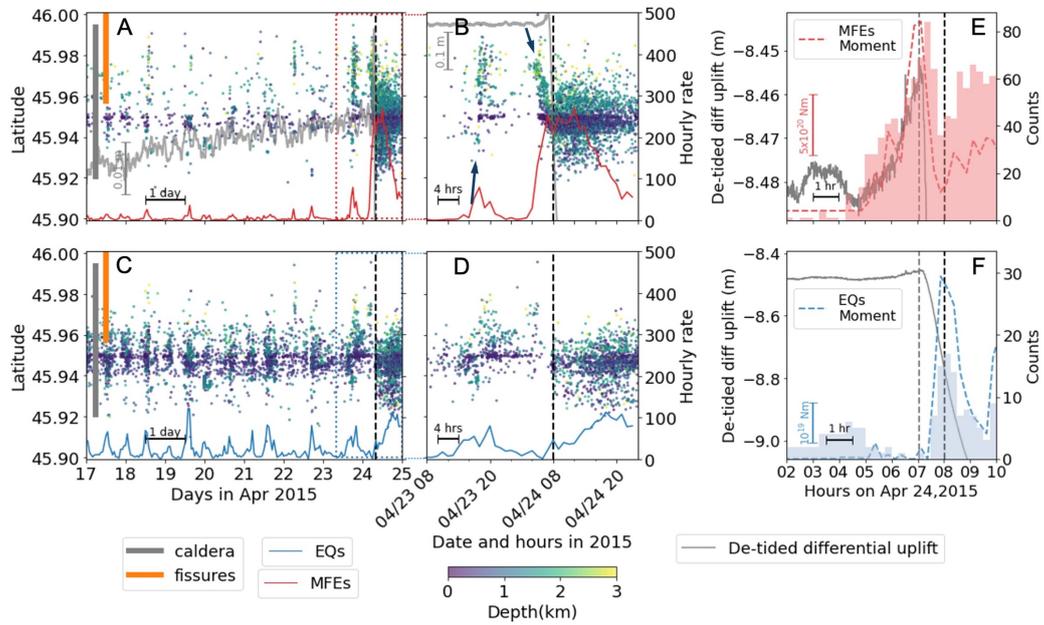


Figure 3. Pre-eruption temporal evolution of the two spectral clusters. The MFEs (top panels) and earthquakes (bottom panels) spatiotemporal distribution in ~ 1 week (A and C), ~ 2 days (B and D), and 8 hours (E and F) time scale. The top and bottom panels are plotted with the same time axis for MFEs (A, B, and E) and earthquakes (C, D, and F) activity. The red and blue curves (A - D) are hourly seismicity rates. Grey curves (A - F) show volcano inflation represented by detided differential uplift measurement between two bottom pressure recorders (BPRs) at sites MJ03E and MJ03F (Figure 1)(Chadwick et al., 2022). Dashed red and blue curves (E and F) show the binned seismic moment of the MFEs and earthquakes during the eruption. The dashed black vertical line (A - F) marks the time of eruption onset at 8:01 when the lava first reaches the seafloor(Wilcock et al., 2016). Arrows in (B) point to northward and southward migration prior to the eruption. The dashed gray vertical line (E and F) marks the time of maximum inflation at 7:03 (E and F).

137 from normal faulting to thrust faulting (Wilcock et al., 2016). This suggests that the fin-
138 gerprints are also not sensitive to the reversal of fault slip motion.

139 Tidal triggering of earthquakes is observed at Axial seamount, especially prior to
140 the eruption (Wilcock et al., 2016; Scholz et al., 2019; Tan et al., 2019; Tolstoy et al.,
141 2002). Here, we compare the temporal correlation between the two spectral clusters we
142 identified and the ocean tide to understand their driving mechanisms. We find that the
143 rate of regular earthquakes closely follows the tidal cycle over the observation period (Fig-
144 ure 3D, Figure S6). Given their locations and the temporal correlation with the tides,
145 we infer that these earthquakes generally occur on critically stressed ring faults and are
146 triggered by small stress changes. The MFE cluster, during the same period, shows rather
147 sparsely distributed bursts of events (Figure S4A) which mostly lie along the eastern edge
148 of the caldera to the north (Figure S7). Among these bursts, we do not observe a clear
149 migration pattern over long distances (Figure S7) as seen in the two very active bursts
150 that relate to the north-south migration ~ 15 hours and ~ 3.5 hours before the eruption
151 onset (Figure 3B). The timings of the MFE bursts correlate with the tides in many cases
152 (Figure S4), but they locate further to the north compared with the tidal driven earth-
153 quakes (Figure 3). We do not observe systematic offset between the timing of the MFE
154 bursts and the peak of regular earthquakes. This suggests that the underlying driver for
155 the two different types of earthquakes may be the same (e.g., magma pressure), or that
156 the drivers respond to the tidal forcing in a similar way.

157 **5 Possible mechanisms**

158 Two possible explanations of the spectral feature difference are path effect and source
159 effect. Spatially variant attenuation patterns, especially in local complex volcanic struc-
160 tures, may cause differences in frequency content if observed along different paths. How-
161 ever, we find that the same clustering analysis carried out at other stations (AXAS1, AXEC2)
162 in the OBS network gives similar groupings (Supplementary materials, Figure S2). If path
163 effects were causing the clustering, we would expect different event groupings at stations
164 that sample different source-receiver paths. It is also possible that attenuation or veloc-
165 ity changes occur in a region that is local to the source. However, we find that closely
166 located and timed events from the two groups still show different spectral behavior at
167 a common station. Therefore, we infer that the spectral difference between the two groups
168 is likely caused by differences in the source mechanisms.

169 There are many possibilities that can explain source differences, including differ-
170 ences in fault stress state, faulting mechanisms, or the effects of fluid. Wilcock et al. (2016)
171 detected southward migration of pre-eruption seismicity along the east wall in the hours
172 before the eruption and associated it with southward dike propagation and opening of
173 eruptive fissures. It is possible that the MFEs are tracking magma flows into the open-
174 ing cracks and thus include non-double-couple components from the crack opening mode
175 in contrast to simple shear failures of earthquakes on the ring faults (Foulger et al., 2004).
176 In this process, the low-frequency content in the MFE waveforms might be generated
177 by magma or volatiles filling the crack, as observed in studies at other regions (Chouet
178 & Matoza, 2013; Cui et al., 2021; Woods et al., 2018; Song et al., 2023).

179 When comparing the moment release of MFEs with available differential elevation
180 data (Nooner & Chadwick, 2016) we find that the peak moment coincides with the peak
181 in inflation about one hour before the eruption (Figure 3H). Moment release for the earth-
182 quakes, on the other hand, is highest during the time of rapid deflation after lava erupted.
183 This suggests the MFEs are associated with magmatic processes during the pre-eruption
184 inflation process, while the regular earthquakes are triggered by the stress change on the
185 ring faults as the magma chamber deflates. In the pre-eruption period, the MFEs in the
186 north also correlate with the region of maximum uplift observed in deformation measure-
187 ment (Nooner & Chadwick, 2016), illuminating the segment of the eruptive fissure where
188 the following eruption started.

189 Given that the MFEs locate along the eruptive fissures near the roof of the magma
190 chamber and the documented high CO₂ content at the Axial seamount (Dixon et al.,
191 1988), the MFEs are likely caused by brittle crack opening and subsequent movement
192 of magma and/or volatiles into the zones of weakness created by increasing magma pres-
193 sure. In fact, the observation that they distribute widely in space and time suggests they
194 are more likely related to CO₂ release as opposed to magma movement. Because MFEs
195 are detected for months prior to the eruption, it implies there is an extended period of
196 magma intrusion or volatile release possibly associated with inflating sills. However, the
197 behavior of early MFE bursts suggests that this magmatic process may occur at a small
198 scale at depth so that they do not show a clear migration pattern along the dike, con-
199 sistent with the presence of volatiles. As the magma pressure builds up, the dike finally
200 forms along the weakened zones and initiates the southward propagation, which is ob-

201 served as intense MFE activity starting ~ 3.5 hours before the eruption. Figure 4 shows
 202 a cartoon summarizing the physical processes and associated seismicity at Axial.

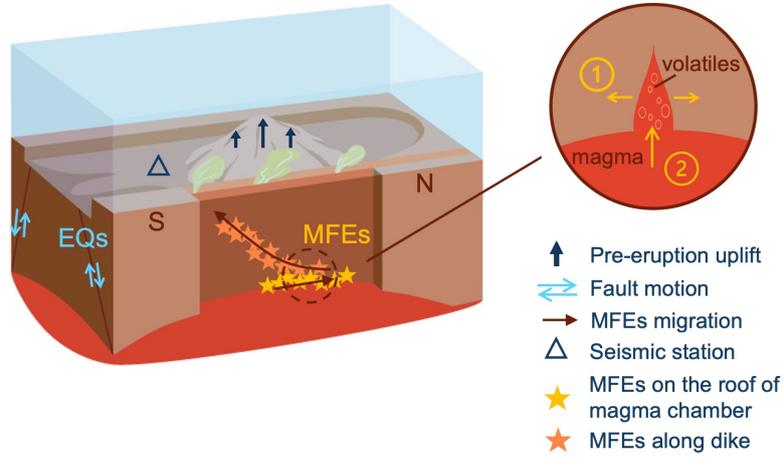


Figure 4. Cartoon summarizing observations. Tidal-driven earthquakes occur on caldera ring faults while the MFEs track movement of volatiles and magma prior to the eruption. Inset shows possible mechanisms of the MFEs. ① and ② correspond to the crack opening (brittle onset in Figure 2F) and volatile/magma influx (low-frequency tail in Figure 2F) processes.

203 Short-term volcano eruption forecasting has long been a challenging task due to
 204 the lack of clear and reliable precursory signals. Common prediction metrics include long-
 205 term deformation measurement, changes in tidal triggering, and short-term seismicity
 206 increase (Wilcock et al., 2018, 2016; Nooner & Chadwick, 2016). In this study, unsuper-
 207 vised ML revealed the emergence of a precursory signal defined as MFEs. These signals
 208 differ substantially from volcano-tectonic (VT; White and McCausland (2016)) or long-
 209 period (LP; Woods et al. (2018); Song et al. (2023)) earthquakes or tremors (Dempsey
 210 et al., 2020), as they contain both short and long period waves. Although they may re-
 211 semble some of the previously reported hybrid frequency earthquakes (HFE; Harrington
 212 and Brodsky (2007); Yu et al. (2021); Coté et al. (2010); Cui et al. (2021)), our obser-
 213 vations suggest that their mechanism might be different. Interpretations of previously
 214 observed hybrid earthquakes include path effects caused by strong attenuation or low-
 215 velocity layers, and source effects due to low stress drop, slow rupture speed, or fluid res-
 216 onance. Our analysis indicates that the characteristic spectral features of the MFEs likely
 217 originate from source effects rather than path effects, making them a potential precur-

218 sory signal to track magma movement or volatile release at depth. This precursory MFE
219 activity intensifies ~ 15 hours before the eruption and peaks ~ 1 hour before the magma
220 reaches the seafloor, which offers an opportunity to improve short-term eruption fore-
221 casting on time scales of hours to days. With the capability to identify such precursory
222 signal in real time, we can now monitor these signals as Axial is preparing for its next
223 eruption to occur within the 2025-2030 time period (Chadwick et al., 2022). More im-
224 portantly, the novel use of unsupervised machine learning opens up a new opportunity
225 to investigate whether such precursory seismic signals exist at other active volcanoes.

226 Data Availability Statement

227 Seismic waveforms used in this study were downloaded from the Incorporated Re-
228 search Institutions for Seismology (IRIS) Data Management Center (DMC) ([https://
229 ds.iris.edu/ds/nodes/dmc/data/](https://ds.iris.edu/ds/nodes/dmc/data/)). The earthquake catalog is available on Figshare
230 (<https://figshare.com/s/1cf6c6dadfa6cdefdbb1>).

231 Acknowledgments

232 This work received support from NSF awards OAC-2103741 and NSF OCE-1951448. We
233 would like to thank anonymous reviewers for providing constructive comments and sug-
234 gestions to improve the manuscript.

235 References

- 236 Arnulf, A., Harding, A., Kent, G., Carbotte, S., Canales, J., & Nedimović, M.
237 (2014). Anatomy of an active submarine volcano. *Geology*, *42*(8), 655–658.
- 238 Baillard, C., Wilcock, W. S., Arnulf, A. F., Tolstoy, M., & Waldhauser, F. (2019).
239 A joint inversion for three-dimensional p and s wave velocity structure and
240 earthquake locations beneath axial seamount. *Journal of Geophysical Research:*
241 *Solid Earth*, *124*(12), 12997–13020.
- 242 Chadwick, W., Paduan, J., Clague, D., Dreyer, B., Merle, S., Bobbitt, A., . . .
243 Nooner, S. (2016). Voluminous eruption from a zoned magma body after
244 an increase in supply rate at axial seamount. *Geophysical Research Letters*,
245 *43*(23), 12–063.
- 246 Chadwick, W., Wilcock, W. S., Nooner, S. L., Beeson, J. W., Sawyer, A. M., & Lau,
247 T.-K. (2022). Geodetic monitoring at axial seamount since its 2015 eruption

- 248 reveals a waning magma supply and tightly linked rates of deformation and
249 seismicity. *Geochemistry, Geophysics, Geosystems*, *23*(1), e2021GC010153.
- 250 Chouet, B. A., & Matoza, R. S. (2013). A multi-decadal view of seismic methods for
251 detecting precursors of magma movement and eruption. *Journal of Volcanology
252 and Geothermal Research*, *252*, 108–175.
- 253 Coté, D. M., Belachew, M., Quillen, A. C., Ebinger, C. J., Keir, D., Ayele, A., &
254 Wright, T. (2010). Low-frequency hybrid earthquakes near a magma cham-
255 ber in afar: quantifying path effects. *Bulletin of the Seismological Society of
256 America*, *100*(5A), 1892–1903.
- 257 Cotton, C. V., & Ellis, D. P. (2011). Spectral vs. spectro-temporal features for
258 acoustic event detection. In *2011 ieee workshop on applications of signal pro-
259 cessing to audio and acoustics (waspa)* (pp. 69–72).
- 260 Cui, X., Li, Z., & Huang, H. (2021). Subdivision of seismicity beneath the summit
261 region of kilauea volcano: Implications for the preparation process of the 2018
262 eruption. *Geophysical Research Letters*, *48*(20), e2021GL094698.
- 263 Dempsey, D., Cronin, S. J., Mei, S., & Kempa-Liehr, A. W. (2020). Automatic
264 precursor recognition and real-time forecasting of sudden explosive volcanic
265 eruptions at whakaari, new zealand. *Nature communications*, *11*(1), 3562.
- 266 Dixon, J. E., Stolper, E., & Delaney, J. R. (1988). Infrared spectroscopic measure-
267 ments of co₂ and h₂o in juan de fuca ridge basaltic glasses. *Earth and Plane-
268 tary Science Letters*, *90*(1), 87–104.
- 269 Foulger, G., Julian, B., Hill, D., Pitt, A., Malin, P., & Shalev, E. (2004). Non-
270 double-couple microearthquakes at long valley caldera, california, provide
271 evidence for hydraulic fracturing. *Journal of Volcanology and Geothermal
272 Research*, *132*(1), 45–71.
- 273 Gudmundsson, M. T., Jónsdóttir, K., Hooper, A., Holohan, E. P., Halldórsson,
274 S. A., Ófeigsson, B. G., . . . others (2016). Gradual caldera collapse at
275 bárdarbunga volcano, iceland, regulated by lateral magma outflow. *Science*,
276 *353*(6296), aaf8988.
- 277 Harrington, R. M., & Brodsky, E. E. (2007). Volcanic hybrid earthquakes that are
278 brittle-failure events. *Geophysical Research Letters*, *34*(6).
- 279 Holtzman, B., Groebner, N., Sawi, T., Xing, T., Pec, M., Ghaffari, H., . . . others
280 (2021). Unsupervised spectral feature extraction applied to acoustic emis-

- 281 sions during brittle creep of basalt under dry and wet conditions. In *Agu fall*
282 *meeting abstracts* (Vol. 2021, pp. H12E–04).
- 283 Holtzman, B., Paté, A., Paisley, J., Waldhauser, F., & Repetto, D. (2018). Machine
284 learning reveals cyclic changes in seismic source spectra in geysers geothermal
285 field. *Science advances*, *4*(5), eaao2929.
- 286 Jenkins, W. F., Gerstoft, P., Bianco, M. J., & Bromirski, P. D. (2021). Unsupervised
287 deep clustering of seismic data: Monitoring the ross ice shelf, antarctica. *Journal of Geophysical Research: Solid Earth*, *126*(9), e2021JB021716.
- 288
- 289 Kelley, D. S., Delaney, J. R., & Juniper, S. K. (2014). Establishing a new era of
290 submarine volcanic observatories: Cabling axial seamount and the endeavour
291 segment of the juan de fuca ridge. *Marine Geology*, *352*, 426–450.
- 292 Lomax, A., Michelini, A., Curtis, A., & Meyers, R. (2009). Earthquake location,
293 direct, global-search methods. *Encyclopedia of complexity and systems science*,
294 *5*, 2449–2473.
- 295 Lomax, A., Virieux, J., Volant, P., & Berge-Thierry, C. (2000). Probabilistic earth-
296 quake location in 3d and layered models: Introduction of a metropolis-gibbs
297 method and comparison with linear locations. *Advances in seismic event*
298 *location*, 101–134.
- 299 Nooner, S. L., & Chadwick, W. (2016). Inflation-predictable behavior and co-
300 eruption deformation at axial seamount. *Science*, *354*(6318), 1399–1403.
- 301 Sawi, T., Holtzman, B., Walter, F., & Paisley, J. (2022). An unsupervised machine-
302 learning approach to understanding seismicity at an alpine glacier. *Journal of*
303 *Geophysical Research: Earth Surface*, e2022JF006909.
- 304 Sawi, T., Waldhauser, F., Holtzman, B. K., & Groebner, N. (2023). Detecting
305 repeating earthquakes on the san andreas fault with unsupervised machine
306 learning of spectrograms. *The Seismic Record*, *3*(4), 376–384.
- 307 Schaff, D. P., & Waldhauser, F. (2005). Waveform cross-correlation-based differen-
308 tial travel-time measurements at the northern california seismic network. *Bul-*
309 *letin of the Seismological Society of America*, *95*(6), 2446–2461.
- 310 Scholz, C. H., Tan, Y. J., & Albino, F. (2019). The mechanism of tidal triggering of
311 earthquakes at mid-ocean ridges. *Nature communications*, *10*(1), 2526.
- 312 Seydoux, L., Balestrieri, R., Poli, P., Hoop, M. d., Campillo, M., & Baraniuk, R.
313 (2020). Clustering earthquake signals and background noises in continuous

- 314 seismic data with unsupervised deep learning. *Nature communications*, *11*(1),
315 3972.
- 316 Song, Z., Tan, Y. J., & Roman, D. C. (2023). Deep long-period earthquakes at
317 akutan volcano from 2005 to 2017 better track magma influxes compared
318 to volcano-tectonic earthquakes. *Geophysical Research Letters*, *50*(10),
319 e2022GL101987.
- 320 Tan, Y. J., Waldhauser, F., Tolstoy, M., & Wilcock, W. S. (2019). Axial seamount:
321 Periodic tidal loading reveals stress dependence of the earthquake size distribu-
322 tion (b value). *Earth and Planetary Science Letters*, *512*, 39–45.
- 323 Tolstoy, M., Vernon, F. L., Orcutt, J. A., & Wyatt, F. K. (2002). Breathing of the
324 seafloor: Tidal correlations of seismicity at axial volcano. *Geology*, *30*(6), 503–
325 506.
- 326 Waldhauser, F., & Ellsworth, W. L. (2000). A double-difference earthquake location
327 algorithm: Method and application to the northern hayward fault, california.
328 *Bulletin of the seismological society of America*, *90*(6), 1353–1368.
- 329 Waldhauser, F., Wilcock, W., Tolstoy, M., Baillard, C., Tan, Y., & Schaff, D. (2020).
330 Precision seismic monitoring and analysis at axial seamount using a real-time
331 double-difference system. *Journal of Geophysical Research: Solid Earth*,
332 *125*(5), e2019JB018796.
- 333 White, R., & McCausland, W. (2016). Volcano-tectonic earthquakes: A new tool for
334 estimating intrusive volumes and forecasting eruptions. *Journal of Volcanology
335 and Geothermal Research*, *309*, 139–155.
- 336 Wilcock, W. S., Dziak, R. P., Tolstoy, M., Chadwick Jr, W. W., Nooner, S. L.,
337 Bohnenstiehl, D. R., . . . others (2018). The recent volcanic history of ax-
338 ial seamount: Geophysical insights into past eruption dynamics with an eye
339 toward enhanced observations of future eruptions. *Oceanography*, *31*(1), 114–
340 123.
- 341 Wilcock, W. S., Tolstoy, M., Waldhauser, F., Garcia, C., Tan, Y. J., Bohnenstiehl,
342 D. R., . . . Mann, M. E. (2016). Seismic constraints on caldera dynamics from
343 the 2015 axial seamount eruption. *Science*, *354*(6318), 1395–1399.
- 344 Wilding, J. D., Zhu, W., Ross, Z. E., & Jackson, J. M. (2023). The magmatic web
345 beneath hawai ‘i. *Science*, *379*(6631), 462–468.
- 346 Woods, J., Donaldson, C., White, R. S., Caudron, C., Brandsdóttir, B., Hudson,

- 347 T. S., & Ágústsdóttir, T. (2018). Long-period seismicity reveals magma path-
348 ways above a laterally propagating dyke during the 2014–15 b́ararbunga rifting
349 event, iceland. *Earth and Planetary Science Letters*, *490*, 216–229.
- 350 Yoon, C. E., O’Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake de-
351 tection through computationally efficient similarity search. *Science advances*,
352 *1*(11), e1501057.
- 353 Yu, H., Harrington, R. M., Kao, H., Liu, Y., & Wang, B. (2021). Fluid-injection-
354 induced earthquakes characterized by hybrid-frequency waveforms manifest the
355 transition from aseismic to seismic slip. *Nature communications*, *12*(1), 6862.
- 356 Zhu, W., & Beroza, G. C. (2019). Phasenet: a deep-neural-network-based seismic
357 arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–
358 273.
- 359 Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L., & Beroza, G. C.
360 (2022). Earthquake phase association using a bayesian gaussian mixture
361 model. *Journal of Geophysical Research: Solid Earth*, *127*(5), e2021JB023249.

Volcanic precursor revealed by machine learning offers new eruption forecasting capability

Kaiwen Wang¹, Felix Waldhauser¹, Maya Tolstoy^{1,2}, David Schaff¹, Theresa
Sawi¹, William Wilcock², Yen Joe Tan³

¹Lamont-Doherty Earth Observatory, Columbia University

²School of Oceanography, University of Washington

³Earth and Environmental Sciences Programme, The Chinese University of Hong Kong

Key Points:

- Unsupervised learning separated regular earthquakes and precursory mixed frequency earthquakes (MFEs) based on different spectral patterns
- The regular earthquakes have strong tidal modulation, corresponding to failures on the caldera ring faults triggered by tidal stress changes
- The MFEs emerge 15 hours before eruption and migrate along pre-existing fissures, likely associated with eruption preparation processes

Corresponding author: Kaiwen Wang, kw2988@columbia.edu

Abstract

Seismicity at active volcanoes provides crucial constraints on the dynamics of magma systems and complex fault activation processes preceding and during an eruption. We characterize time-dependent spectral features of volcanic earthquakes at Axial Seamount with unsupervised machine learning methods, revealing mixed frequency signals that emerge from the continuous waveforms about 15 hours before eruption onset. The events migrate along pre-existing fissures, suggesting that they represent brittle crack opening driven by influx of magma or volatiles. These results demonstrate the power of novel machine learning algorithms to characterize subtle changes in magmatic processes associated with eruption preparation, offering new possibilities for forecasting Axial's anticipated next eruption. This novel method is generalizable and can be employed to identify similar precursory signals at other active volcanoes.

Plain Language Summary

Our research used observations of small earthquakes to understand the dynamic behaviors of magma and fault systems before and during a volcano eruption. Specifically, we used machine learning techniques to search for signature waveform patterns that may inform us of their associated physical processes. At Axial Seamount, an active underwater volcano, we discovered distinct patterns in earthquake signals preceding and during the 2015 eruption. Based on event spectral patterns, we identified signals of mixed-frequency earthquakes that emerge about 15 hours before the eruption starts and migrate along pre-existing eruptive fissures. The spectral pattern involves a mixture of low frequency energy following the first arrivals, which we interpret to represent opening of cracks and being filled with magma or gases. Our study demonstrates that we can use machine learning algorithms to detect subtle changes in volcanic signals and help us better understand the processes leading up to an eruption. This may help us in forecasting Axial's upcoming eruption and can possibly be applied to other active volcanoes too.

1 Introduction

Seismic observations can provide important constraints on the structure of a volcano and its dynamic behavior in volcanic cycles (Wilcock et al., 2016; Wilding et al., 2023; Tan et al., 2019; Gudmundsson et al., 2016). Questions remain on how magma moves in the subsurface preceding an eruption and how soon before an eruption this process

66 begins. Recent advances in unsupervised machine learning methods (Holtzman et al.,
67 2018; Cotton & Ellis, 2011; Holtzman et al., 2021; Sawi et al., 2022; Yoon et al., 2015;
68 Seydoux et al., 2020; Jenkins et al., 2021) offer the opportunity to mine large waveform
69 archives to find subtle differences in the spectral content of seismic signals. These dif-
70 ferences can be interpreted with respect to changes in source characteristics and the volcano-
71 tectonic processes that drive brittle failure, providing a time-dependent image of phys-
72 ical processes that lead up to an eruption.

73 Axial Seamount is a well-instrumented, active submarine volcano on the Juan de
74 Fuca Ridge (Figure 1) with a long record of geophysical data that covers the last three
75 eruptions in 1998, 2011, and 2015 (Wilcock et al., 2018, 2016; Nooner & Chadwick, 2016),
76 including documentation of the eruptive fissures and lava flows of the recent 2015 erup-
77 tion (Chadwick et al., 2016), and 3-D images of its shallow magma chamber (Arnulf et
78 al., 2014). Five months before the most recent eruption in April 2015, seismicity at Ax-
79 ial Seamount has been recorded by a local, cabled, 7-station ocean bottom seismome-
80 ter (OBS) network operated in real-time by the Ocean Observatories Initiative (OOI)
81 (Kelley et al., 2014). The OBS array recorded signals from a variety of sources (Wilcock
82 et al., 2016). Here we apply unsupervised machine learning methods to the 4 months be-
83 fore and during the 2015 eruption to find precursory signals with distinct frequency con-
84 tent.

65 **2 Supervised and Unsupervised ML**

66 We combined supervised machine learning (ML) techniques (Zhu & Beroza, 2019;
67 Zhu et al., 2022) with cross-correlation-based, high-resolution earthquake relative loca-
68 tion methods (Waldhauser & Ellsworth, 2000; Waldhauser et al., 2020; Lomax et al., 2000,
69 2009) to develop a catalog of 240,000 earthquakes ($M = -1.74$ to 3.45) for Axial Seamount
70 from 2014 to 2021. The new earthquake catalog illuminates the caldera ring faults and
71 the fissures that were active during the previous eruptions (Figure 1) (Wilcock et al., 2016;
72 Waldhauser et al., 2020).

73 We then apply an unsupervised machine learning method (SpecUFEx, Holtzman
74 et al., 2018) to the 4 months of pre-eruption data to characterize spectral patterns in the
75 waveforms. SpecUFEx is an unsupervised spectral feature extraction algorithm origi-
76 nally developed using ML methods for audio pattern recognition (Cotton & Ellis, 2011)

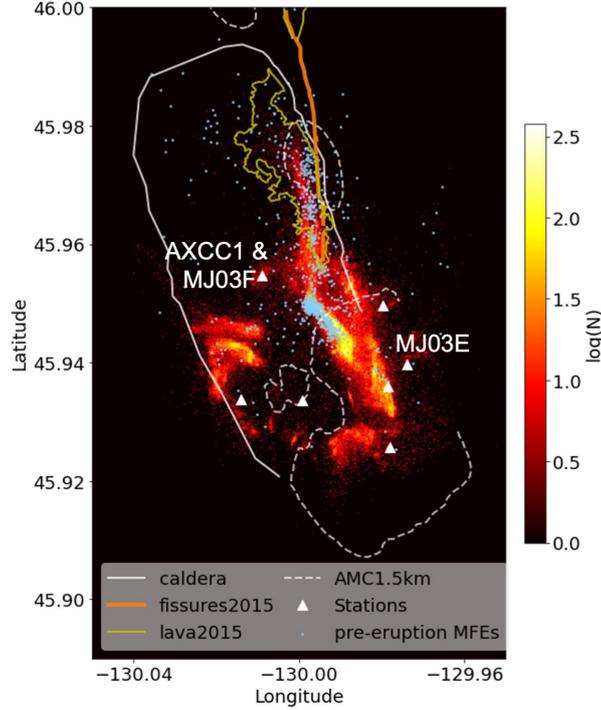


Figure 1. Heatmap of earthquake density at Axial Seamount from Nov 2014 to Dec 2021. Mixed-frequency earthquakes (MFEs) one day before the eruption are shown in light blue dots. Also shown are the caldera rim (white solid line), the 1.5 km depth contour of the Axial magma chamber (AMC) (dashed white line), eruptive fissures (orange lines), and lava flows (yellow lines) of the 2015 eruption and the OBS array (white triangles). The heatmap shows the number of earthquakes in each bin (bin size $25\text{m} \times 25\text{m}$).

77 and has been later adapted to characterize seismic waveforms of earthquakes (Holtzman
 78 et al., 2018), acoustic emissions (Holtzman et al., 2021), icequakes (Sawi et al., 2022),
 79 and repeating earthquakes (Sawi et al., 2023). It takes event spectrograms as input and
 80 applies nonnegative matrix factorization (NMF) and hidden Markov models (HMM) to
 81 reduce the dimensionality of the spectral features and remove features that are common
 82 to all signals. For each earthquake, we calculate a fingerprint matrix by counting the num-
 83 ber of state transitions in the state sequence matrix from the HMM output. The finger-
 84 prints are condensed representations of the original earthquake spectrograms while still
 85 keeping their time-dependent spectral information. We further compress the fingerprints
 86 by principal component analysis (PCA) and finally apply K-means clustering to iden-
 87 tify earthquake clusters that have common spectral features (Holtzman et al., 2018). We
 88 focus our analysis on the 4 months of seismicity leading up to the eruption on April 24,

89 2015. We use waveform data from broadband OBS station AXCC1 and learned the fea-
 90 ture dictionary on a representative subset of $\sim 9,000$ events in the week before the erup-
 91 tion. We then use the learned dictionary to calculate features of the ~ 4 months of seis-
 92 micity starting from the beginning of 2015 until the eruption onset.

93 **3 Spectral differences**

94 K-means clustering of the fingerprints separates the events into two main groups
 95 with small but distinct differences in spectral features in the waveforms between the groups
 96 (Figure 2A and 2D). To investigate which characteristic spectral features might contribute
 97 to the separation of the two earthquake clusters, we examine the representative patterns
 98 of the condensed fingerprints. By stacking the top 100 representative fingerprints in each
 99 cluster (Figure 2B and 2E), we identify the active states (bright spots in stacked finger-
 100 prints). These active states are the characteristic features that define the spectral fea-
 101 ture space. We project these characteristic features in the fingerprints back onto the HMM
 102 and NMF mappings (emissions matrix in Figure S3A and spectral dictionary in Figure
 103 S3B) to solve for their frequency-dependent sensitivity kernel (Figure S3C). Comparing
 104 the frequency dependency of the characteristic features in the two clusters, we find that
 105 one cluster has events with lower frequency content coming in shortly (~ 1 s) after the
 106 P-arrival. Thus we define the earthquakes in this cluster as mixed-frequency earthquakes
 107 (MFEs) and the events in the other cluster as regular earthquakes (EQs). The spectral
 108 differences can also be seen in the stacked spectrograms (Figure 2C and 2F) of the top
 109 100 representative events and their waveforms (Figure 2A and 2D).

110 **4 Spatio-temporal distribution**

111 The separation based on spectral characteristics reveals differences in the spatiotem-
 112 poral evolution of the earthquakes in the two groups (Figure 3, Movie S1). Approximately
 113 24 hours prior to the eruption, the MFEs start lighting up the eastern margin of the caldera
 114 along the southern segment of the eruptive fissures (Figure 1). These MFEs locate close
 115 to the roof of the Axial magma chamber (~ 1.5 km; 15). ~ 15 hours before the eruption,
 116 a distinct burst of MFEs migrates from the caldera center northward along the eastern
 117 margin of the caldera at a speed of 4.4 km/h (arrow in Figure 3B). The peak hourly mo-
 118 ment release of the MFEs during that burst is about two orders of magnitude above back-
 119 ground, and 40 times that released by all regular earthquakes in the same period. Af-

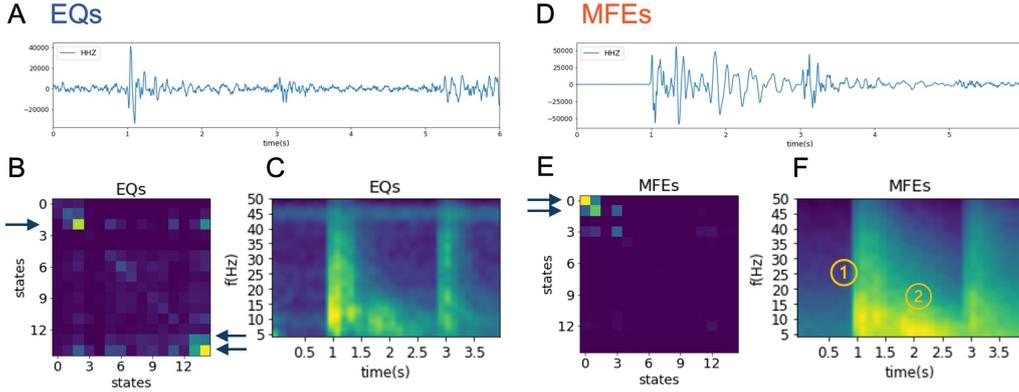


Figure 2. Spectral characteristics of events in the two main clusters. Waveform examples of one representative event in each cluster (A and D), stacked fingerprints (B and E), and stacked spectrograms (C and F) using 100 events in each cluster. The stacked fingerprints and stacked spectrograms are color scaled by their maximum value. ① and ② in F mark the impulsive P arrival and the low-frequency tail.

120 ter this initial burst, the MFE activity subsides for a couple of hours and then, about
 121 3.5 hours before the eruption, a second burst of MFES occurred, reversing the path of
 122 the previous burst and migrating southward (at a speed of 1.1 km/hr) and eventually
 123 upward towards the location where the lava first erupted on the seafloor (Wilcock et al.,
 124 2016) (Figure 3B). After that point and for the next hour, the MFES spread out across
 125 the entire fault system during the course of the eruption. The second MFE burst is char-
 126 acterized by a steep increase in seismic moment release starting about 4 hours and peak-
 127 ing 1 hour before the eruption onset. Peak hourly moment release is about 30 times that
 128 of the first burst, while the moment release from regular earthquakes leading up to the
 129 eruption is comparably insignificant. Once the eruption starts, MFE moment release con-
 130 tinuously decreases, while that from regular earthquakes increases.

131 Different from the MFES, the regular earthquakes locate primarily in the south-
 132 ern part of the caldera (Figure S5). They occur on both the eastern and western walls
 133 of the ring fault, which suggests that the spectral fingerprints are not sensitive to event
 134 location relative to the seismic station. The regular earthquake cluster also includes events
 135 during the pre-eruption inflation period as well as the rapid deflation period after the
 136 eruption started, that is, when the fault slip motion on the caldera ring faults reversed

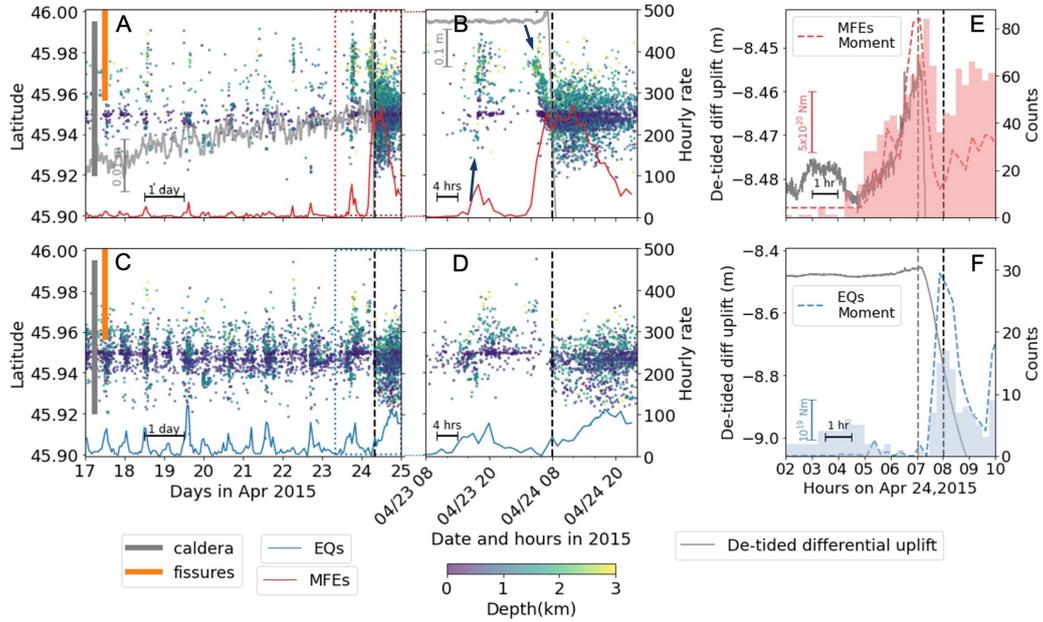


Figure 3. Pre-eruption temporal evolution of the two spectral clusters. The MFEs (top panels) and earthquakes (bottom panels) spatiotemporal distribution in ~ 1 week (A and C), ~ 2 days (B and D), and 8 hours (E and F) time scale. The top and bottom panels are plotted with the same time axis for MFEs (A, B, and E) and earthquakes (C, D, and F) activity. The red and blue curves (A - D) are hourly seismicity rates. Grey curves (A - F) show volcano inflation represented by detided differential uplift measurement between two bottom pressure recorders (BPRs) at sites MJ03E and MJ03F (Figure 1)(Chadwick et al., 2022). Dashed red and blue curves (E and F) show the binned seismic moment of the MFEs and earthquakes during the eruption. The dashed black vertical line (A - F) marks the time of eruption onset at 8:01 when the lava first reaches the seafloor(Wilcock et al., 2016). Arrows in (B) point to northward and southward migration prior to the eruption. The dashed gray vertical line (E and F) marks the time of maximum inflation at 7:03 (E and F).

137 from normal faulting to thrust faulting (Wilcock et al., 2016). This suggests that the fin-
138 gerprints are also not sensitive to the reversal of fault slip motion.

139 Tidal triggering of earthquakes is observed at Axial seamount, especially prior to
140 the eruption (Wilcock et al., 2016; Scholz et al., 2019; Tan et al., 2019; Tolstoy et al.,
141 2002). Here, we compare the temporal correlation between the two spectral clusters we
142 identified and the ocean tide to understand their driving mechanisms. We find that the
143 rate of regular earthquakes closely follows the tidal cycle over the observation period (Fig-
144 ure 3D, Figure S6). Given their locations and the temporal correlation with the tides,
145 we infer that these earthquakes generally occur on critically stressed ring faults and are
146 triggered by small stress changes. The MFE cluster, during the same period, shows rather
147 sparsely distributed bursts of events (Figure S4A) which mostly lie along the eastern edge
148 of the caldera to the north (Figure S7). Among these bursts, we do not observe a clear
149 migration pattern over long distances (Figure S7) as seen in the two very active bursts
150 that relate to the north-south migration ~ 15 hours and ~ 3.5 hours before the eruption
151 onset (Figure 3B). The timings of the MFE bursts correlate with the tides in many cases
152 (Figure S4), but they locate further to the north compared with the tidal driven earth-
153 quakes (Figure 3). We do not observe systematic offset between the timing of the MFE
154 bursts and the peak of regular earthquakes. This suggests that the underlying driver for
155 the two different types of earthquakes may be the same (e.g., magma pressure), or that
156 the drivers respond to the tidal forcing in a similar way.

157 **5 Possible mechanisms**

158 Two possible explanations of the spectral feature difference are path effect and source
159 effect. Spatially variant attenuation patterns, especially in local complex volcanic struc-
160 tures, may cause differences in frequency content if observed along different paths. How-
161 ever, we find that the same clustering analysis carried out at other stations (AXAS1, AXEC2)
162 in the OBS network gives similar groupings (Supplementary materials, Figure S2). If path
163 effects were causing the clustering, we would expect different event groupings at stations
164 that sample different source-receiver paths. It is also possible that attenuation or veloc-
165 ity changes occur in a region that is local to the source. However, we find that closely
166 located and timed events from the two groups still show different spectral behavior at
167 a common station. Therefore, we infer that the spectral difference between the two groups
168 is likely caused by differences in the source mechanisms.

169 There are many possibilities that can explain source differences, including differ-
170 ences in fault stress state, faulting mechanisms, or the effects of fluid. Wilcock et al. (2016)
171 detected southward migration of pre-eruption seismicity along the east wall in the hours
172 before the eruption and associated it with southward dike propagation and opening of
173 eruptive fissures. It is possible that the MFEs are tracking magma flows into the open-
174 ing cracks and thus include non-double-couple components from the crack opening mode
175 in contrast to simple shear failures of earthquakes on the ring faults (Foulger et al., 2004).
176 In this process, the low-frequency content in the MFE waveforms might be generated
177 by magma or volatiles filling the crack, as observed in studies at other regions (Chouet
178 & Matoza, 2013; Cui et al., 2021; Woods et al., 2018; Song et al., 2023).

179 When comparing the moment release of MFEs with available differential elevation
180 data (Nooner & Chadwick, 2016) we find that the peak moment coincides with the peak
181 in inflation about one hour before the eruption (Figure 3H). Moment release for the earth-
182 quakes, on the other hand, is highest during the time of rapid deflation after lava erupted.
183 This suggests the MFEs are associated with magmatic processes during the pre-eruption
184 inflation process, while the regular earthquakes are triggered by the stress change on the
185 ring faults as the magma chamber deflates. In the pre-eruption period, the MFEs in the
186 north also correlate with the region of maximum uplift observed in deformation measure-
187 ment (Nooner & Chadwick, 2016), illuminating the segment of the eruptive fissure where
188 the following eruption started.

189 Given that the MFEs locate along the eruptive fissures near the roof of the magma
190 chamber and the documented high CO₂ content at the Axial seamount (Dixon et al.,
191 1988), the MFEs are likely caused by brittle crack opening and subsequent movement
192 of magma and/or volatiles into the zones of weakness created by increasing magma pres-
193 sure. In fact, the observation that they distribute widely in space and time suggests they
194 are more likely related to CO₂ release as opposed to magma movement. Because MFEs
195 are detected for months prior to the eruption, it implies there is an extended period of
196 magma intrusion or volatile release possibly associated with inflating sills. However, the
197 behavior of early MFE bursts suggests that this magmatic process may occur at a small
198 scale at depth so that they do not show a clear migration pattern along the dike, con-
199 sistent with the presence of volatiles. As the magma pressure builds up, the dike finally
200 forms along the weakened zones and initiates the southward propagation, which is ob-

201 served as intense MFE activity starting ~ 3.5 hours before the eruption. Figure 4 shows
 202 a cartoon summarizing the physical processes and associated seismicity at Axial.

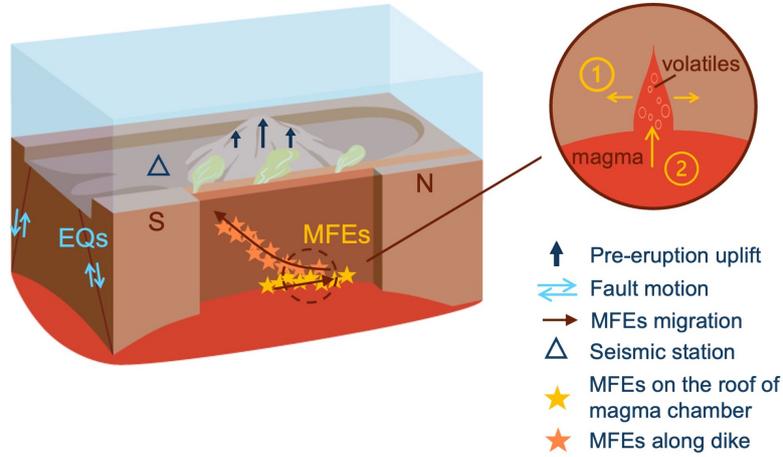


Figure 4. Cartoon summarizing observations. Tidal-driven earthquakes occur on caldera ring faults while the MFEs track movement of volatiles and magma prior to the eruption. Inset shows possible mechanisms of the MFEs. ① and ② correspond to the crack opening (brittle onset in Figure 2F) and volatile/magma influx (low-frequency tail in Figure 2F) processes.

203 Short-term volcano eruption forecasting has long been a challenging task due to
 204 the lack of clear and reliable precursory signals. Common prediction metrics include long-
 205 term deformation measurement, changes in tidal triggering, and short-term seismicity
 206 increase (Wilcock et al., 2018, 2016; Nooner & Chadwick, 2016). In this study, unsuper-
 207 vised ML revealed the emergence of a precursory signal defined as MFEs. These signals
 208 differ substantially from volcano-tectonic (VT; White and McCausland (2016)) or long-
 209 period (LP; Woods et al. (2018); Song et al. (2023)) earthquakes or tremors (Dempsey
 210 et al., 2020), as they contain both short and long period waves. Although they may re-
 211 semble some of the previously reported hybrid frequency earthquakes (HFE; Harrington
 212 and Brodsky (2007); Yu et al. (2021); Coté et al. (2010); Cui et al. (2021)), our obser-
 213 vations suggest that their mechanism might be different. Interpretations of previously
 214 observed hybrid earthquakes include path effects caused by strong attenuation or low-
 215 velocity layers, and source effects due to low stress drop, slow rupture speed, or fluid res-
 216 onance. Our analysis indicates that the characteristic spectral features of the MFEs likely
 217 originate from source effects rather than path effects, making them a potential precur-

218 sory signal to track magma movement or volatile release at depth. This precursory MFE
219 activity intensifies ~ 15 hours before the eruption and peaks ~ 1 hour before the magma
220 reaches the seafloor, which offers an opportunity to improve short-term eruption fore-
221 casting on time scales of hours to days. With the capability to identify such precursory
222 signal in real time, we can now monitor these signals as Axial is preparing for its next
223 eruption to occur within the 2025-2030 time period (Chadwick et al., 2022). More im-
224 portantly, the novel use of unsupervised machine learning opens up a new opportunity
225 to investigate whether such precursory seismic signals exist at other active volcanoes.

226 Data Availability Statement

227 Seismic waveforms used in this study were downloaded from the Incorporated Re-
228 search Institutions for Seismology (IRIS) Data Management Center (DMC) ([https://
229 ds.iris.edu/ds/nodes/dmc/data/](https://ds.iris.edu/ds/nodes/dmc/data/)). The earthquake catalog is available on Figshare
230 (<https://figshare.com/s/1cf6c6dadfa6cdefdbb1>).

231 Acknowledgments

232 This work received support from NSF awards OAC-2103741 and NSF OCE-1951448. We
233 would like to thank anonymous reviewers for providing constructive comments and sug-
234 gestions to improve the manuscript.

235 References

- 236 Arnulf, A., Harding, A., Kent, G., Carbotte, S., Canales, J., & Nedimović, M.
237 (2014). Anatomy of an active submarine volcano. *Geology*, *42*(8), 655–658.
- 238 Baillard, C., Wilcock, W. S., Arnulf, A. F., Tolstoy, M., & Waldhauser, F. (2019).
239 A joint inversion for three-dimensional p and s wave velocity structure and
240 earthquake locations beneath axial seamount. *Journal of Geophysical Research:*
241 *Solid Earth*, *124*(12), 12997–13020.
- 242 Chadwick, W., Paduan, J., Clague, D., Dreyer, B., Merle, S., Bobbitt, A., . . .
243 Nooner, S. (2016). Voluminous eruption from a zoned magma body after
244 an increase in supply rate at axial seamount. *Geophysical Research Letters*,
245 *43*(23), 12–063.
- 246 Chadwick, W., Wilcock, W. S., Nooner, S. L., Beeson, J. W., Sawyer, A. M., & Lau,
247 T.-K. (2022). Geodetic monitoring at axial seamount since its 2015 eruption

- 248 reveals a waning magma supply and tightly linked rates of deformation and
249 seismicity. *Geochemistry, Geophysics, Geosystems*, *23*(1), e2021GC010153.
- 250 Chouet, B. A., & Matoza, R. S. (2013). A multi-decadal view of seismic methods for
251 detecting precursors of magma movement and eruption. *Journal of Volcanology
252 and Geothermal Research*, *252*, 108–175.
- 253 Coté, D. M., Belachew, M., Quillen, A. C., Ebinger, C. J., Keir, D., Ayele, A., &
254 Wright, T. (2010). Low-frequency hybrid earthquakes near a magma cham-
255 ber in afar: quantifying path effects. *Bulletin of the Seismological Society of
256 America*, *100*(5A), 1892–1903.
- 257 Cotton, C. V., & Ellis, D. P. (2011). Spectral vs. spectro-temporal features for
258 acoustic event detection. In *2011 ieee workshop on applications of signal pro-
259 cessing to audio and acoustics (wasppaa)* (pp. 69–72).
- 260 Cui, X., Li, Z., & Huang, H. (2021). Subdivision of seismicity beneath the summit
261 region of kilauea volcano: Implications for the preparation process of the 2018
262 eruption. *Geophysical Research Letters*, *48*(20), e2021GL094698.
- 263 Dempsey, D., Cronin, S. J., Mei, S., & Kempa-Liehr, A. W. (2020). Automatic
264 precursor recognition and real-time forecasting of sudden explosive volcanic
265 eruptions at whakaari, new zealand. *Nature communications*, *11*(1), 3562.
- 266 Dixon, J. E., Stolper, E., & Delaney, J. R. (1988). Infrared spectroscopic measure-
267 ments of co₂ and h₂o in juan de fuca ridge basaltic glasses. *Earth and Plane-
268 tary Science Letters*, *90*(1), 87–104.
- 269 Foulger, G., Julian, B., Hill, D., Pitt, A., Malin, P., & Shalev, E. (2004). Non-
270 double-couple microearthquakes at long valley caldera, california, provide
271 evidence for hydraulic fracturing. *Journal of Volcanology and Geothermal
272 Research*, *132*(1), 45–71.
- 273 Gudmundsson, M. T., Jónsdóttir, K., Hooper, A., Holohan, E. P., Halldórsson,
274 S. A., Ófeigsson, B. G., . . . others (2016). Gradual caldera collapse at
275 bárdarbunga volcano, iceland, regulated by lateral magma outflow. *Science*,
276 *353*(6296), aaf8988.
- 277 Harrington, R. M., & Brodsky, E. E. (2007). Volcanic hybrid earthquakes that are
278 brittle-failure events. *Geophysical Research Letters*, *34*(6).
- 279 Holtzman, B., Groebner, N., Sawi, T., Xing, T., Pec, M., Ghaffari, H., . . . others
280 (2021). Unsupervised spectral feature extraction applied to acoustic emis-

- 281 sions during brittle creep of basalt under dry and wet conditions. In *Agu fall*
282 *meeting abstracts* (Vol. 2021, pp. H12E–04).
- 283 Holtzman, B., Paté, A., Paisley, J., Waldhauser, F., & Repetto, D. (2018). Machine
284 learning reveals cyclic changes in seismic source spectra in geysers geothermal
285 field. *Science advances*, *4*(5), eaao2929.
- 286 Jenkins, W. F., Gerstoft, P., Bianco, M. J., & Bromirski, P. D. (2021). Unsupervised
287 deep clustering of seismic data: Monitoring the ross ice shelf, antarctica. *Journal of Geophysical Research: Solid Earth*, *126*(9), e2021JB021716.
- 288
- 289 Kelley, D. S., Delaney, J. R., & Juniper, S. K. (2014). Establishing a new era of
290 submarine volcanic observatories: Cabling axial seamount and the endeavour
291 segment of the juan de fuca ridge. *Marine Geology*, *352*, 426–450.
- 292 Lomax, A., Michelini, A., Curtis, A., & Meyers, R. (2009). Earthquake location,
293 direct, global-search methods. *Encyclopedia of complexity and systems science*,
294 *5*, 2449–2473.
- 295 Lomax, A., Virieux, J., Volant, P., & Berge-Thierry, C. (2000). Probabilistic earth-
296 quake location in 3d and layered models: Introduction of a metropolis-gibbs
297 method and comparison with linear locations. *Advances in seismic event*
298 *location*, 101–134.
- 299 Nooner, S. L., & Chadwick, W. (2016). Inflation-predictable behavior and co-
300 eruption deformation at axial seamount. *Science*, *354*(6318), 1399–1403.
- 301 Sawi, T., Holtzman, B., Walter, F., & Paisley, J. (2022). An unsupervised machine-
302 learning approach to understanding seismicity at an alpine glacier. *Journal of*
303 *Geophysical Research: Earth Surface*, e2022JF006909.
- 304 Sawi, T., Waldhauser, F., Holtzman, B. K., & Groebner, N. (2023). Detecting
305 repeating earthquakes on the san andreas fault with unsupervised machine
306 learning of spectrograms. *The Seismic Record*, *3*(4), 376–384.
- 307 Schaff, D. P., & Waldhauser, F. (2005). Waveform cross-correlation-based differen-
308 tial travel-time measurements at the northern california seismic network. *Bul-*
309 *letin of the Seismological Society of America*, *95*(6), 2446–2461.
- 310 Scholz, C. H., Tan, Y. J., & Albino, F. (2019). The mechanism of tidal triggering of
311 earthquakes at mid-ocean ridges. *Nature communications*, *10*(1), 2526.
- 312 Seydoux, L., Balestrieri, R., Poli, P., Hoop, M. d., Campillo, M., & Baraniuk, R.
313 (2020). Clustering earthquake signals and background noises in continuous

- 314 seismic data with unsupervised deep learning. *Nature communications*, *11*(1),
315 3972.
- 316 Song, Z., Tan, Y. J., & Roman, D. C. (2023). Deep long-period earthquakes at
317 akutan volcano from 2005 to 2017 better track magma influxes compared
318 to volcano-tectonic earthquakes. *Geophysical Research Letters*, *50*(10),
319 e2022GL101987.
- 320 Tan, Y. J., Waldhauser, F., Tolstoy, M., & Wilcock, W. S. (2019). Axial seamount:
321 Periodic tidal loading reveals stress dependence of the earthquake size distribu-
322 tion (b value). *Earth and Planetary Science Letters*, *512*, 39–45.
- 323 Tolstoy, M., Vernon, F. L., Orcutt, J. A., & Wyatt, F. K. (2002). Breathing of the
324 seafloor: Tidal correlations of seismicity at axial volcano. *Geology*, *30*(6), 503–
325 506.
- 326 Waldhauser, F., & Ellsworth, W. L. (2000). A double-difference earthquake location
327 algorithm: Method and application to the northern hayward fault, california.
328 *Bulletin of the seismological society of America*, *90*(6), 1353–1368.
- 329 Waldhauser, F., Wilcock, W., Tolstoy, M., Baillard, C., Tan, Y., & Schaff, D. (2020).
330 Precision seismic monitoring and analysis at axial seamount using a real-time
331 double-difference system. *Journal of Geophysical Research: Solid Earth*,
332 *125*(5), e2019JB018796.
- 333 White, R., & McCausland, W. (2016). Volcano-tectonic earthquakes: A new tool for
334 estimating intrusive volumes and forecasting eruptions. *Journal of Volcanology
335 and Geothermal Research*, *309*, 139–155.
- 336 Wilcock, W. S., Dziak, R. P., Tolstoy, M., Chadwick Jr, W. W., Nooner, S. L.,
337 Bohnenstiehl, D. R., . . . others (2018). The recent volcanic history of ax-
338 ial seamount: Geophysical insights into past eruption dynamics with an eye
339 toward enhanced observations of future eruptions. *Oceanography*, *31*(1), 114–
340 123.
- 341 Wilcock, W. S., Tolstoy, M., Waldhauser, F., Garcia, C., Tan, Y. J., Bohnenstiehl,
342 D. R., . . . Mann, M. E. (2016). Seismic constraints on caldera dynamics from
343 the 2015 axial seamount eruption. *Science*, *354*(6318), 1395–1399.
- 344 Wilding, J. D., Zhu, W., Ross, Z. E., & Jackson, J. M. (2023). The magmatic web
345 beneath hawai ‘i. *Science*, *379*(6631), 462–468.
- 346 Woods, J., Donaldson, C., White, R. S., Caudron, C., Brandsdóttir, B., Hudson,

- 347 T. S., & Ágústsdóttir, T. (2018). Long-period seismicity reveals magma path-
348 ways above a laterally propagating dyke during the 2014–15 b́ararbunga rifting
349 event, iceland. *Earth and Planetary Science Letters*, *490*, 216–229.
- 350 Yoon, C. E., O’Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake de-
351 tection through computationally efficient similarity search. *Science advances*,
352 *1*(11), e1501057.
- 353 Yu, H., Harrington, R. M., Kao, H., Liu, Y., & Wang, B. (2021). Fluid-injection-
354 induced earthquakes characterized by hybrid-frequency waveforms manifest the
355 transition from aseismic to seismic slip. *Nature communications*, *12*(1), 6862.
- 356 Zhu, W., & Beroza, G. C. (2019). Phasenet: a deep-neural-network-based seismic
357 arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–
358 273.
- 359 Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L., & Beroza, G. C.
360 (2022). Earthquake phase association using a bayesian gaussian mixture
361 model. *Journal of Geophysical Research: Solid Earth*, *127*(5), e2021JB023249.

Geophysical Research Letters

Supporting Information for

Volcanic precursor revealed by machine learning offers new eruption forecasting capability

Kaiwen Wang¹, Felix Waldhauser¹, Maya Tolstoy^{1,2}, David Schaff¹, Theresa Sawi¹, William Wilcock², Yen Joe Tan³

1 Lamont-Doherty Earth Observatory, Columbia University

2 School of Oceanography, University of Washington

3 Earth and Environmental Sciences Programme, The Chinese University of Hong Kong

Contents of this file

Materials and Methods
Supplementary Text
Figs. S1 to S7

Additional Supporting Information (Files uploaded separately)

Movies S1

Materials and Methods

Earthquake catalog development

We used continuous waveforms from November 16, 2014, to December 31, 2021, to build an ML-based earthquake catalog. This study focuses on analyzing the 4 months of pre-eruption seismicity rather than the entire cataloged period. The data were recorded by the OOI 7-station OBS network, which has two broadband stations and five short-period stations. We used PhaseNet (Zhu & Beroza, 2019), a deep-learning phase picker, to detect and pick the P and S arrivals. The input data are continuous waveforms sampled at 200 Hz. We use a 15-second-long sliding window with a 3-second stepping length. During periods of high seismicity rate (e.g., the day of eruption), we used a smaller window size of 6 seconds to improve picker performance for smaller events that are in the same detection windows with larger events. We use an ML associator (GaMMA, Zhu et al., 2022) to associate the picks with seismic events. We require at least 5 picks for an event to be associated. The ML workflow detects seismic signals not only from earthquakes but also from fin whale calls, seafloor impulsive events, and air-gun shots from active source experiments. We applied SpecUFEEx (Holtzman et al., 2018) to the raw catalog to discriminate earthquakes from other seismic sources.

The final ML-catalog includes 244,321 earthquakes with a total of 1,016,761 P- and 1,258,927 S-phase picks. We estimated moment magnitudes for the earthquakes following the same method used in Wilcock et al. (2016) and obtained magnitudes that range from -1.74 to 3.45. We computed initial hypocenter locations with a grid-search location algorithm (NonLinLoc, Lomax et al., 2000, 2009) together with a 3D tomographic velocity model (Baillard et al., 2019). The grid-search catalog of the earthquakes is then relocated using cross-correlation and double-difference methods following Waldhauser et al. (2020).

Cross-correlation-based double-difference earthquake location

In addition to the ML-based phase arrival times, we measure precise phase delay times using waveform cross-correlation following Waldhauser et al. (2020). We apply time-domain cross-correlation (Schaff & Waldhauser, 2005) to filtered (4–50 Hz), vertical and horizontal component seismograms of pairs of events recorded at the same station and separated by no more than 1 km. We chose 0.5 s long correlation windows for P waves and 0.75 s windows for S waves and search over lags that are ± 0.5 s. We compute delay times for a second pair of windows (0.75 s and 1 s) and retain only the measurements that agree within 0.01 s, thus reducing erroneous correlations due to cycle skipping, for example. From the 14.5 billion measurements, we keep only the correlation delay times for earthquake pairs with at least two measurements with cross-correlation coefficients $C_f \geq 0.8$. When S-delay times are available from both horizontal components, we use them both but set their weights to half of their initial weights (i.e., squared correlation coefficient). The resulting correlation time database includes a total of 1.4 billion delay times.

We can evaluate the consistency and accuracy of the two data sets by forming the difference between the correlation delay times and the delay times formed from the

picks for the corresponding event pair (Figure S1) (see Waldhauser et al., 2020). These differences have standard deviations of 96 ms (P waves) and 66 ms (S waves), indicating high consistency between the two data sets. Standard deviations of 81 ms (P waves) and 50 ms (S waves) for differences from data with $C_f \geq 0.95$ indicate the high accuracy of the PhaseNet picked arrival times, for both P and S arrivals.

Finally, we relocated the earthquakes using the double-difference location algorithm HypoDD (Waldhauser & Ellsworth, 2000; Waldhauser, 2001) to invert both phase pick and cross-correlation time delays for precise relative hypocenter locations (see Waldhauser et al., 2020 for details). The relocated 7-year-long earthquake catalog includes 162,111 with magnitudes between -1.74 and 3.45.

Spectral Clustering Analysis

We apply K-means clustering on the principal components of the fingerprints learned by SpecUFEx (see above). Here we focus on the characteristics of volcano-tectonic earthquakes, so we exclude other types of seismic signals (whale calls, seafloor impulsive events, tremors) in our analysis. We keep principal components that explained 80% of the variance. After inspecting the clustering results, we find that choosing the number of clusters as two would best cluster the earthquakes by their dominant spectral characteristics. While the first group includes signals that can be associated with typical earthquakes that represent shear failure, the second group includes signals that are similar to those of earthquakes, but have a lower frequency package arriving about 1 s after the P-wave onset (see Figure 2). We call these events MFEs (mixed frequency earthquakes). Increasing the number of clusters will subdivide the two main clusters into smaller clusters, still separating the signals from typical earthquakes from the MFE signals.

We tested other clustering algorithms such as Hierarchical clustering and Gaussian Mixture Model. We find different clustering algorithms in general give similar results that show the separation of MFE and earthquake signals, with the K-means results showing less leakage between the two groups.

To verify the spectral differences between MFEs and regular earthquakes identified by SpecUFEx, we run a test that takes the spectra of waveforms directly as input and clusters them by K-means. The clustering results still show the same general pattern of the two groups corresponding to MFEs and regular earthquakes. We compared the event cluster labels produced by clustering the event spectra and find ~90% of them have the same cluster label as defined by clustering SpecUFEx fingerprints. However, we find increased leakage between the two groups. This suggests that the MFEs and regular earthquakes can be separated by their differences in spectral content, but additional temporal information in the fingerprints extracted by SpecUFEx helps better define them in the feature space.

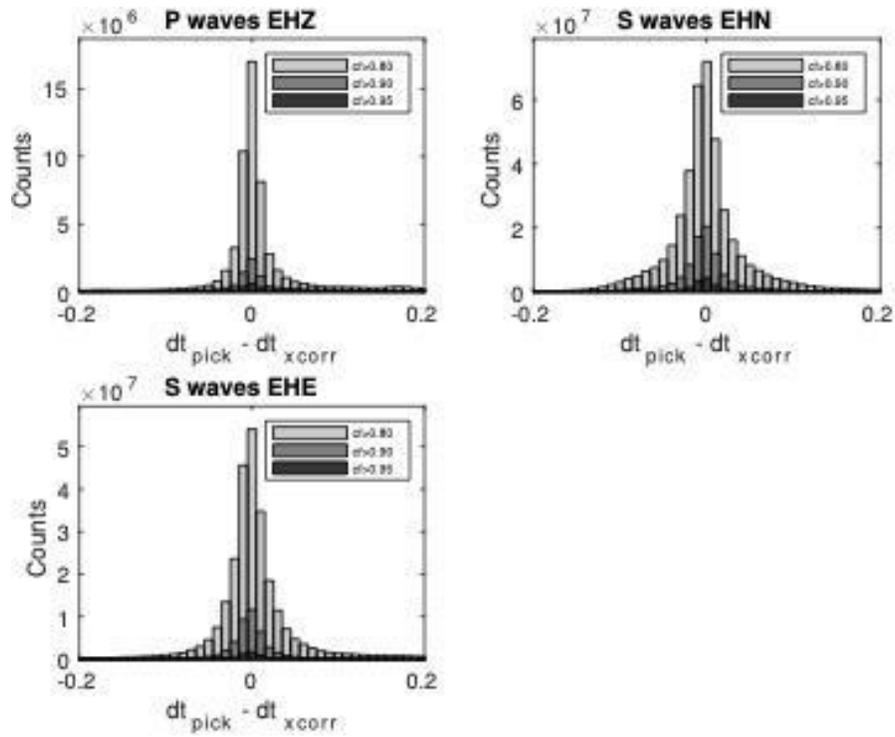


Figure S1. Difference distribution of the picks as compared with the cross-correlation delay times. The bars of light to dark gray colors show cross-correlation delay time measurements of different correlation coefficient thresholds. The three panels show the distribution of the difference between the P phase and correlation on the vertical component and the S phase and correlation on two horizontal components.

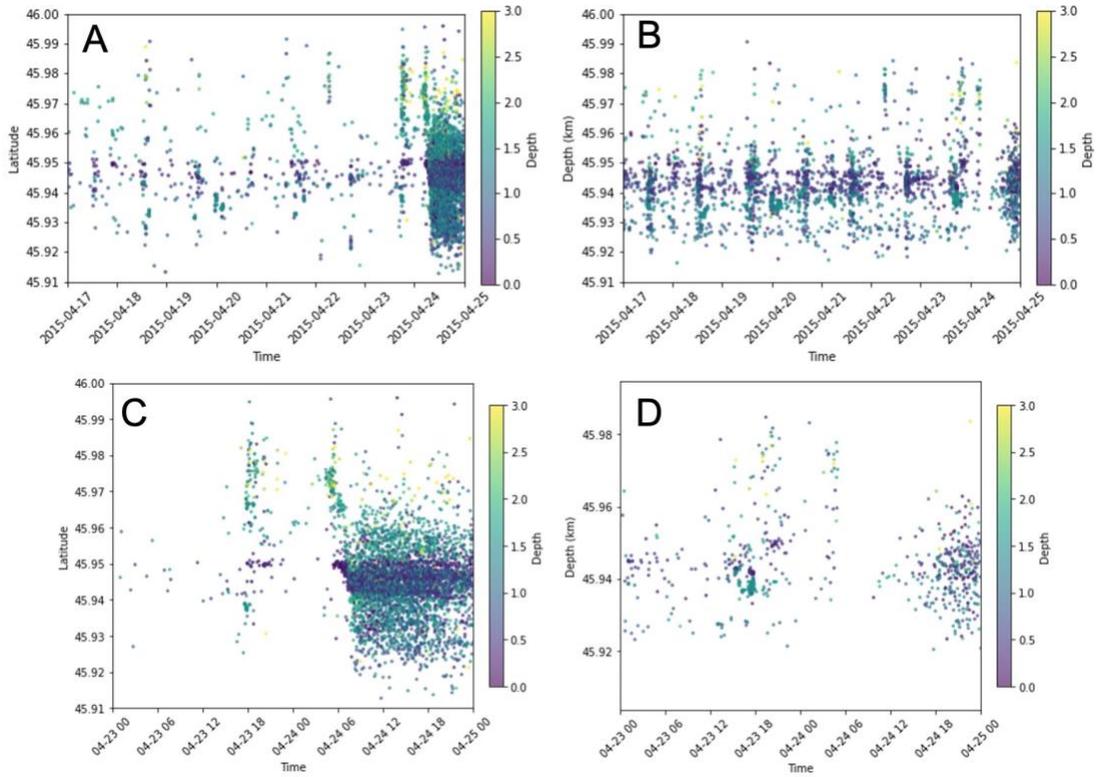


Figure S2. Clustering result using waveforms at AXAS1 station. Spatiotemporal plots of MFEs and earthquakes cluster at ~one week (A and B) and two days (C and D) time scale. (A) and (C) shows MFE activities. (B) and (D) shows regular earthquake activities.

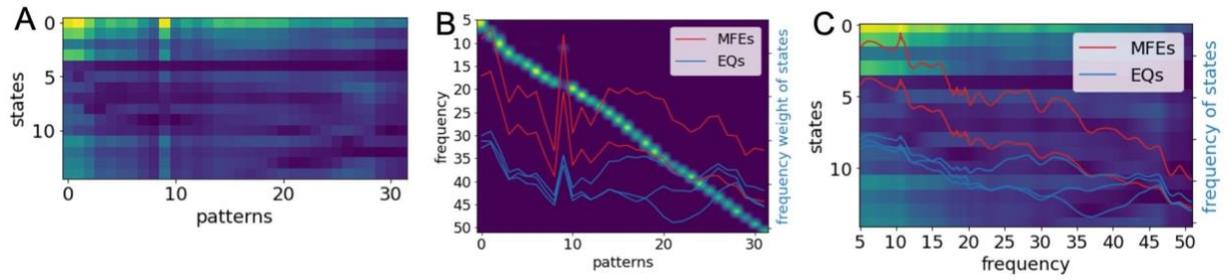


Figure S3. (A) HMM emissions matrix. (B) NMF spectral dictionary overlaid with curves showing frequency weights of the active states in the two stacked fingerprints. (C) Frequency dependent sensitivity kernel of the states in fingerprints. Red and blue curves show the characteristic states of the MFEs and earthquake group, respectively.

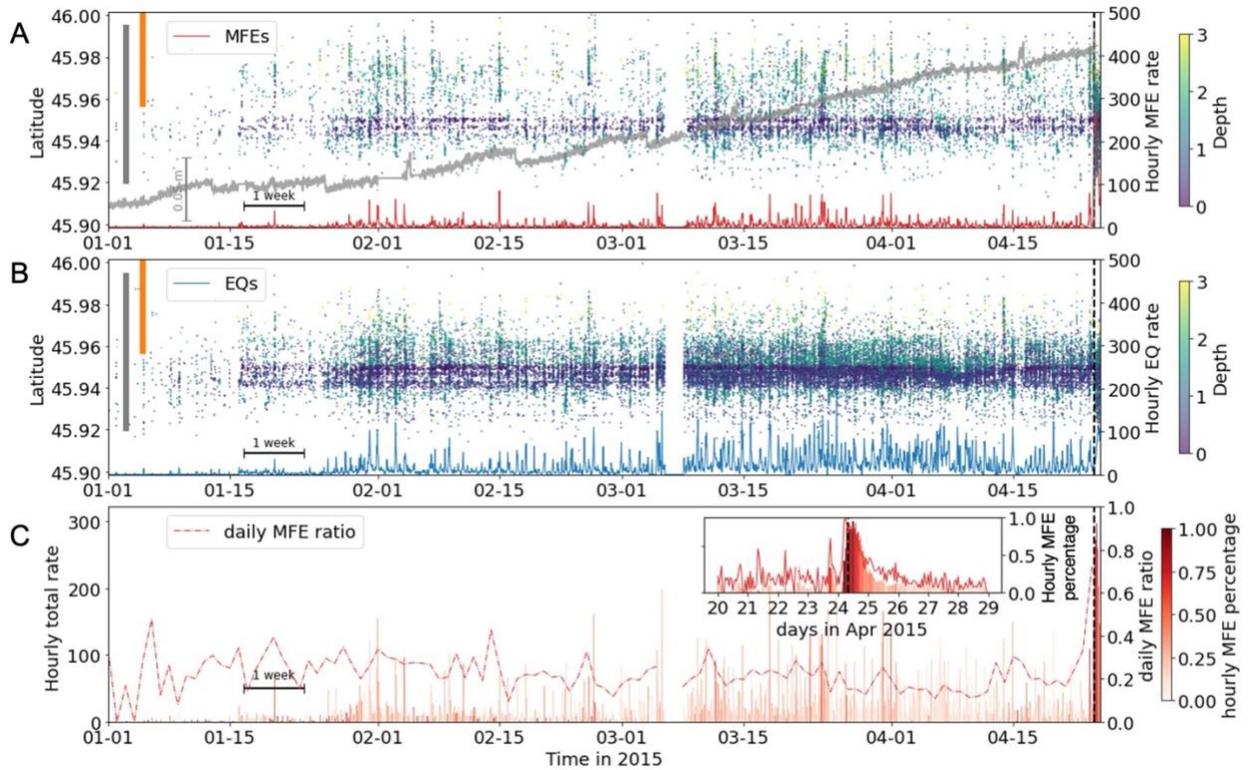


Figure S4. Pre-eruption spatiotemporal evolution of the two spectral clusters and their relative ratio. The MFEs (A) and earthquakes (B) spatiotemporal distribution in ~4 months prior to eruption. (C) Histogram shows hourly percentage of MFEs in all pre-eruption seismicity. Dashed red line shows the daily MFE ratio in the ~4 months prior to eruption. The inset shows hourly percentage of MFEs (red line) in a zoom-in window around eruption time.

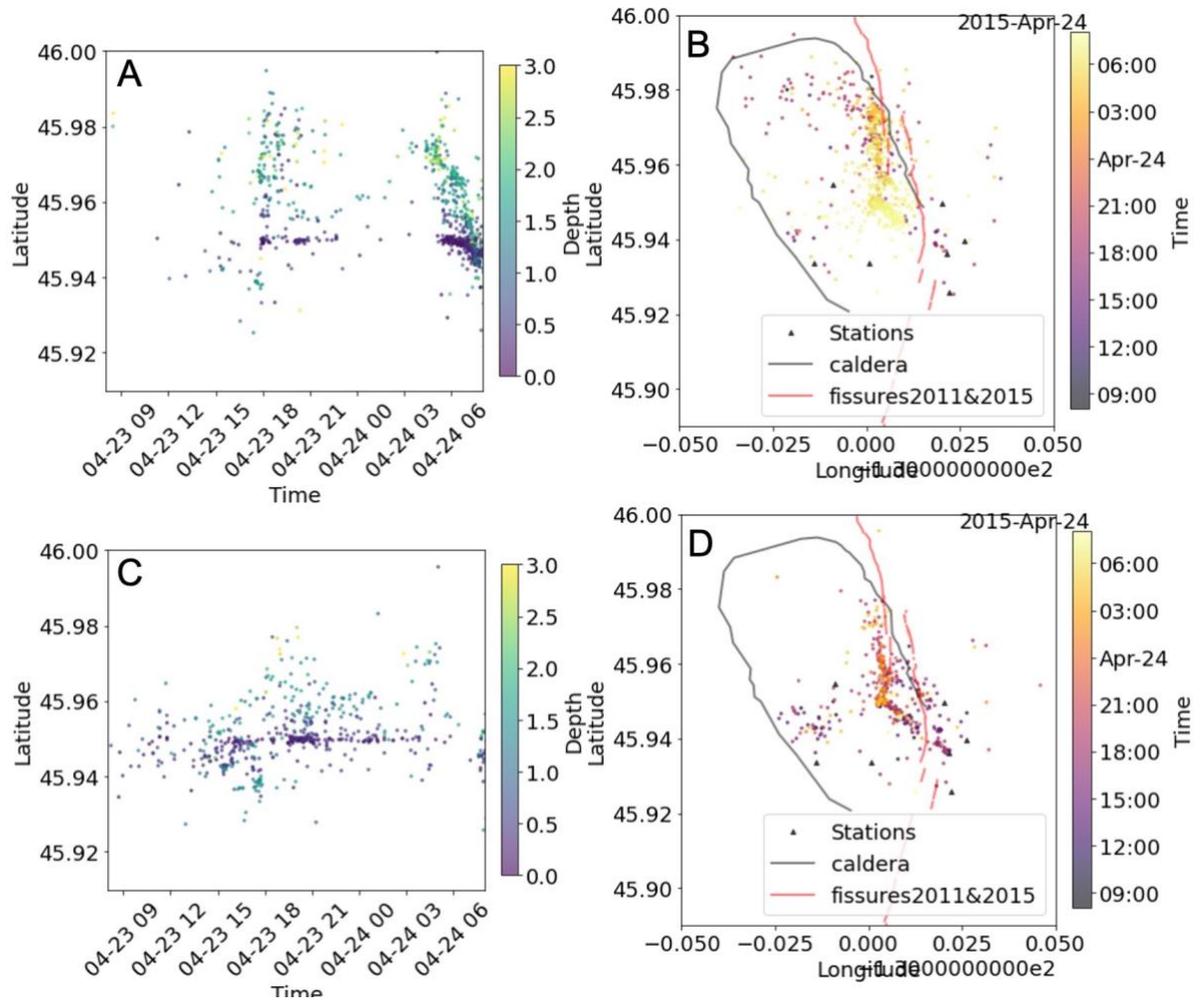


Figure S5. Comparison of earthquakes and MFEs locations. MFEs locations (B) and spatiotemporal plot (A) on the day before eruption. Locations of regular earthquakes (D) and their spatiotemporal distribution (C) in the same time period.

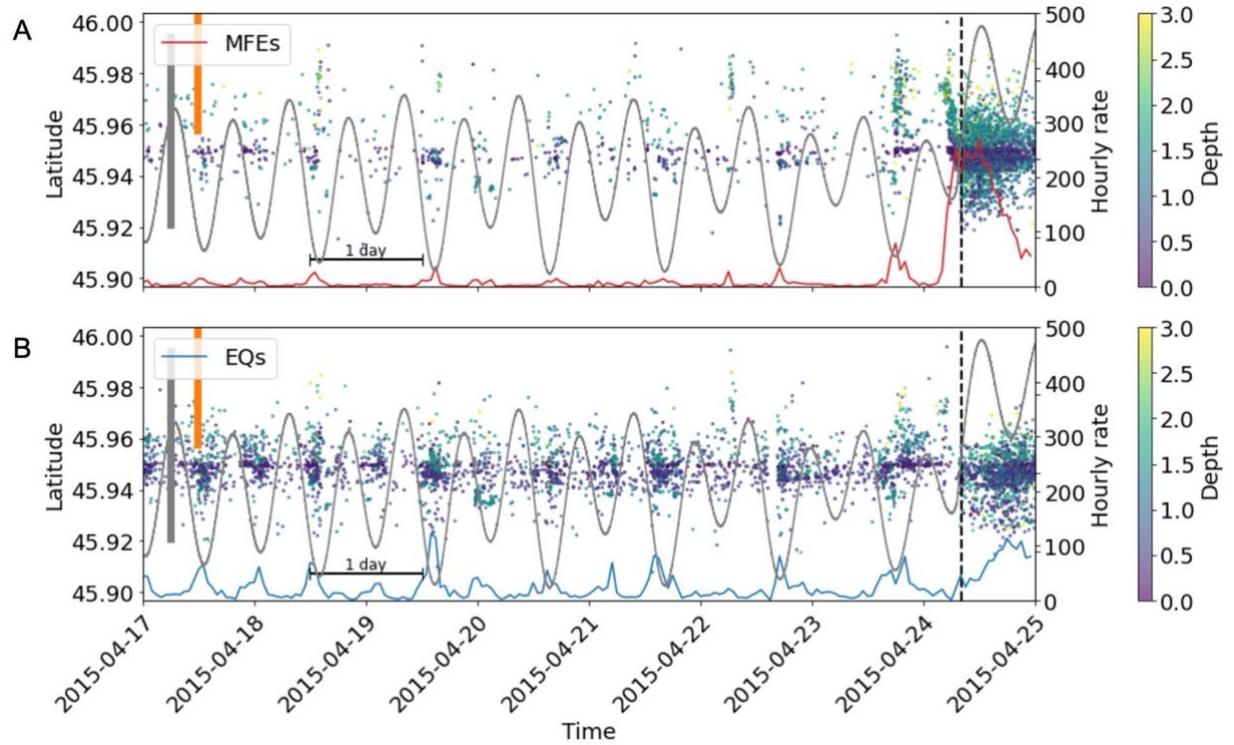
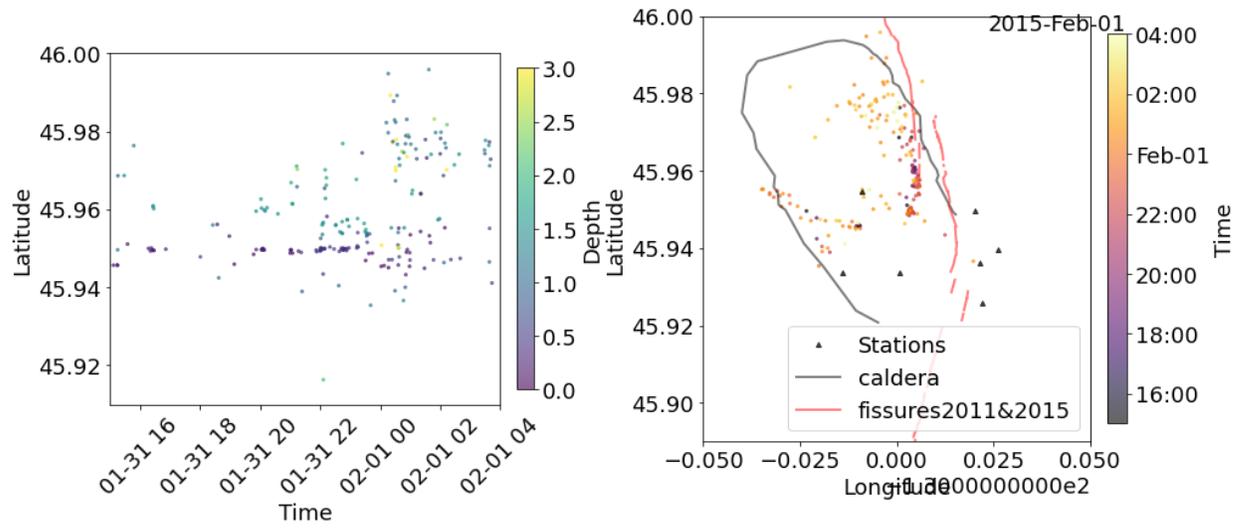
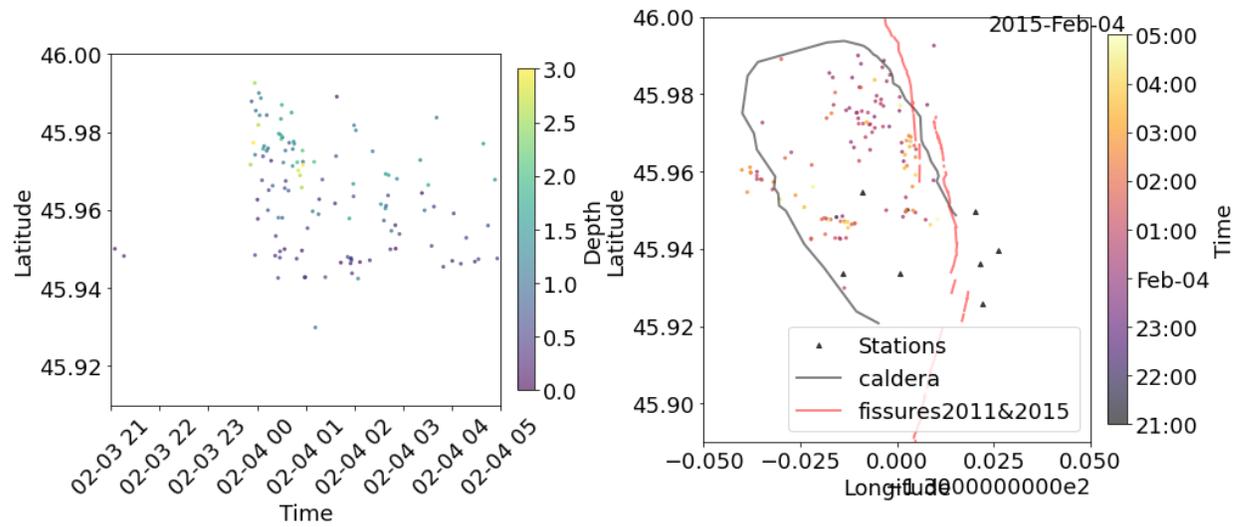


Figure S6. Tidal correlation of the regular earthquakes (B) and MFEs (A). The grey curve shows Bottom-pressure recorder (BPR) measurements at AXCC1. The red and blue curves show the hourly seismicity rate of MFEs (A) and regular earthquakes (B).

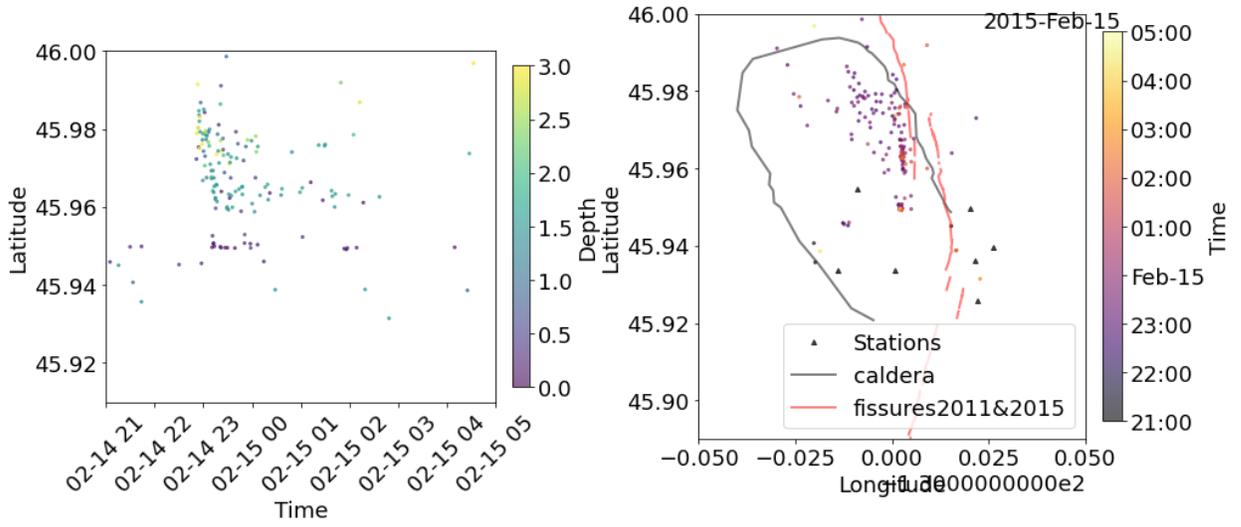
A



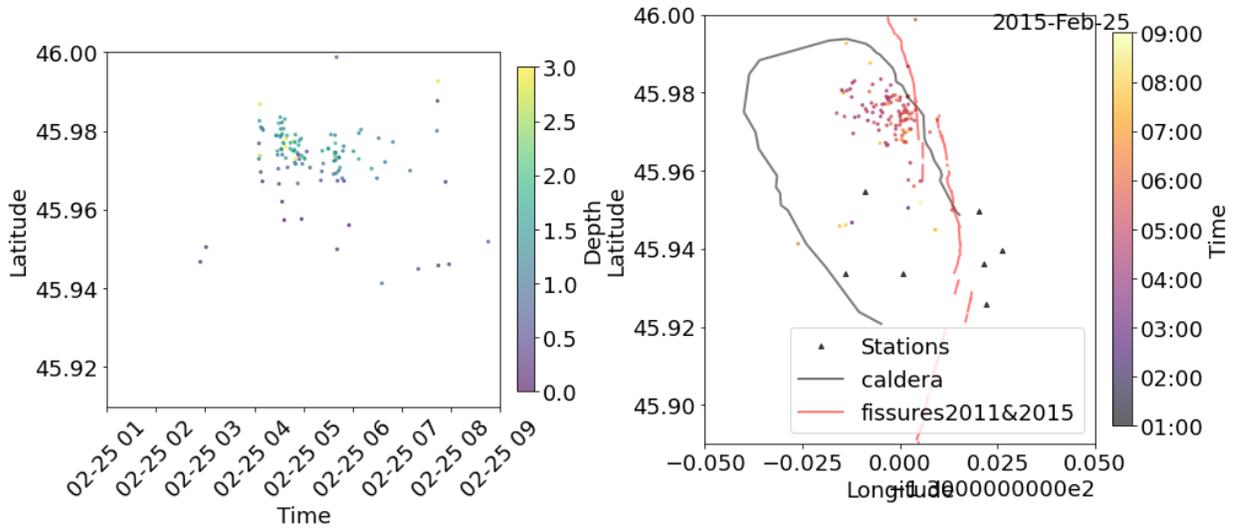
B



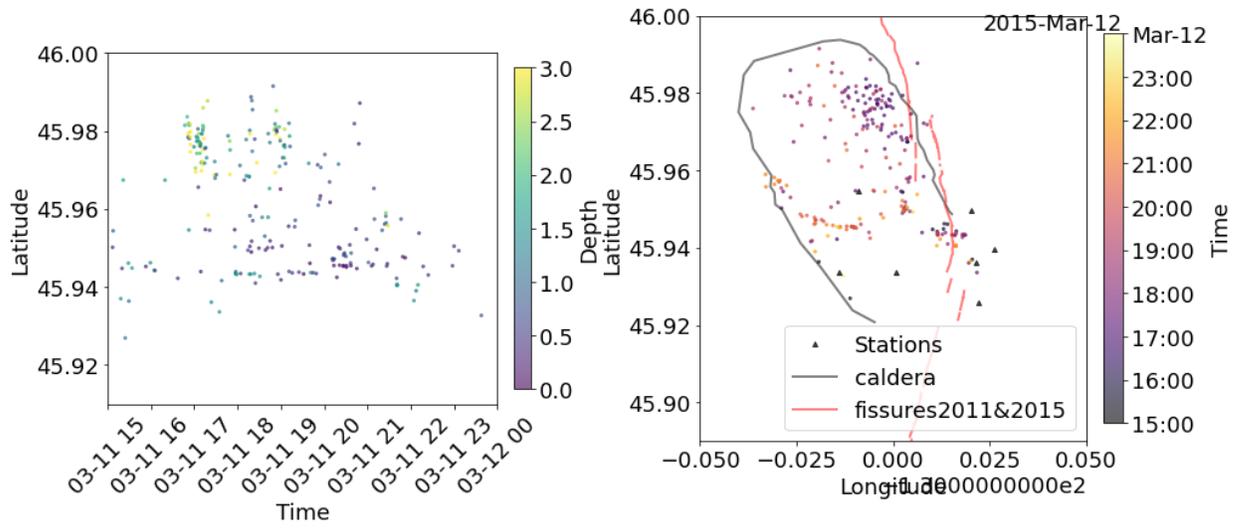
C



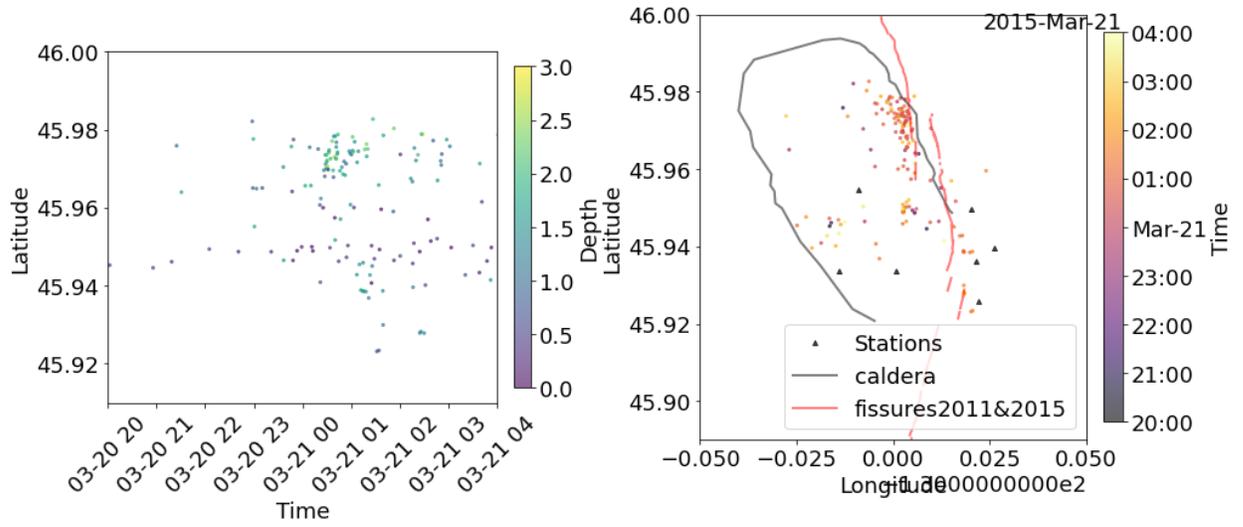
D



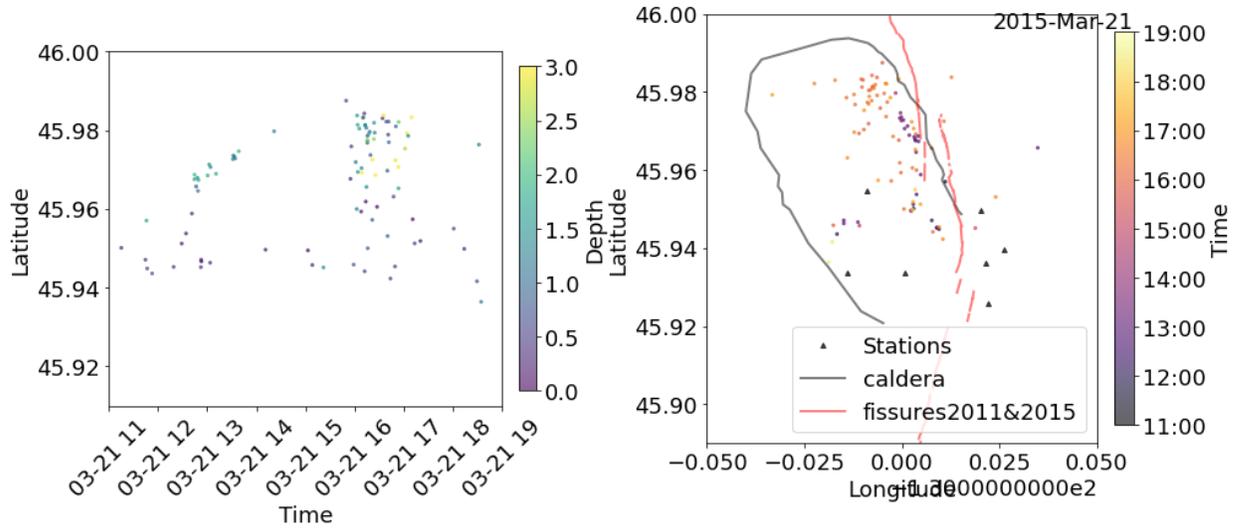
E



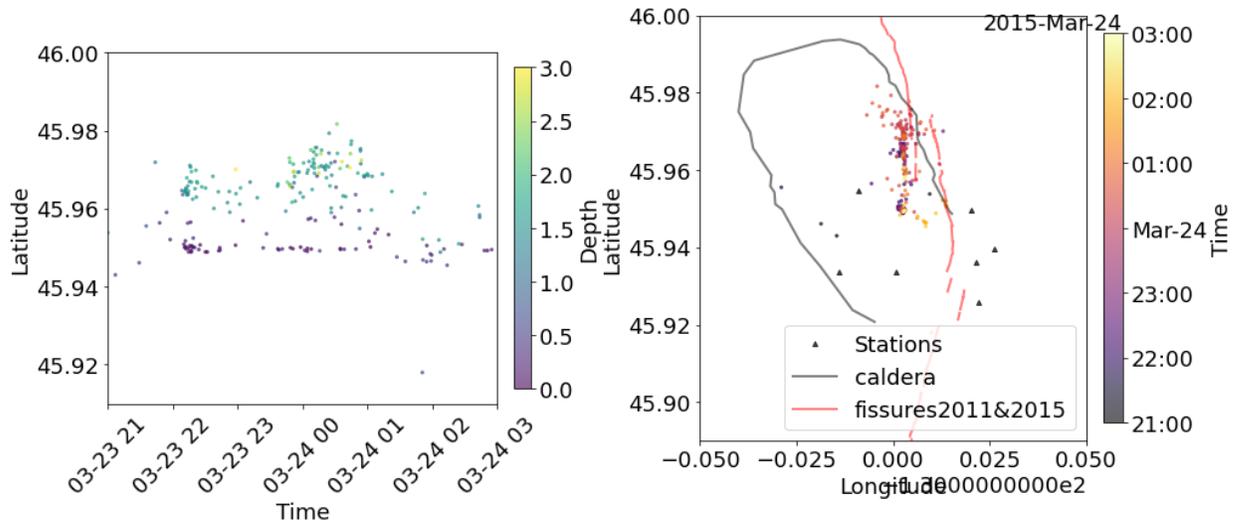
F



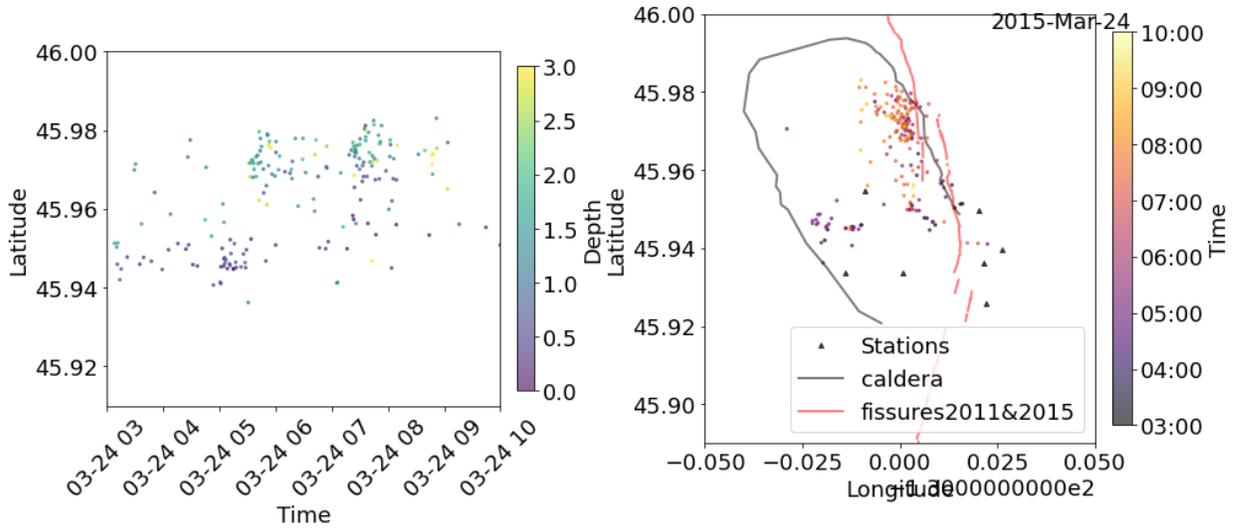
G



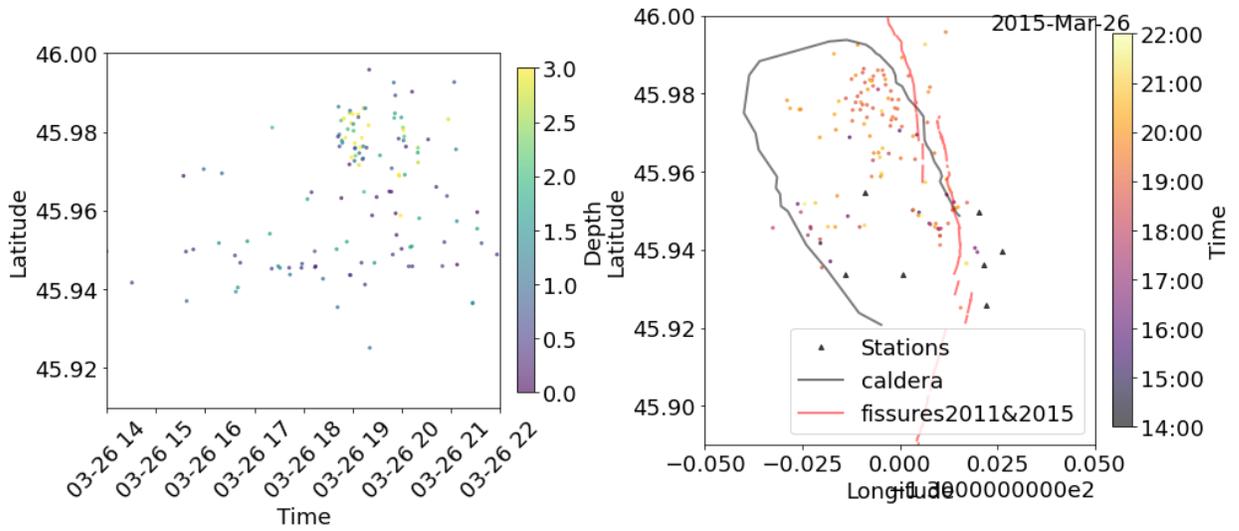
H



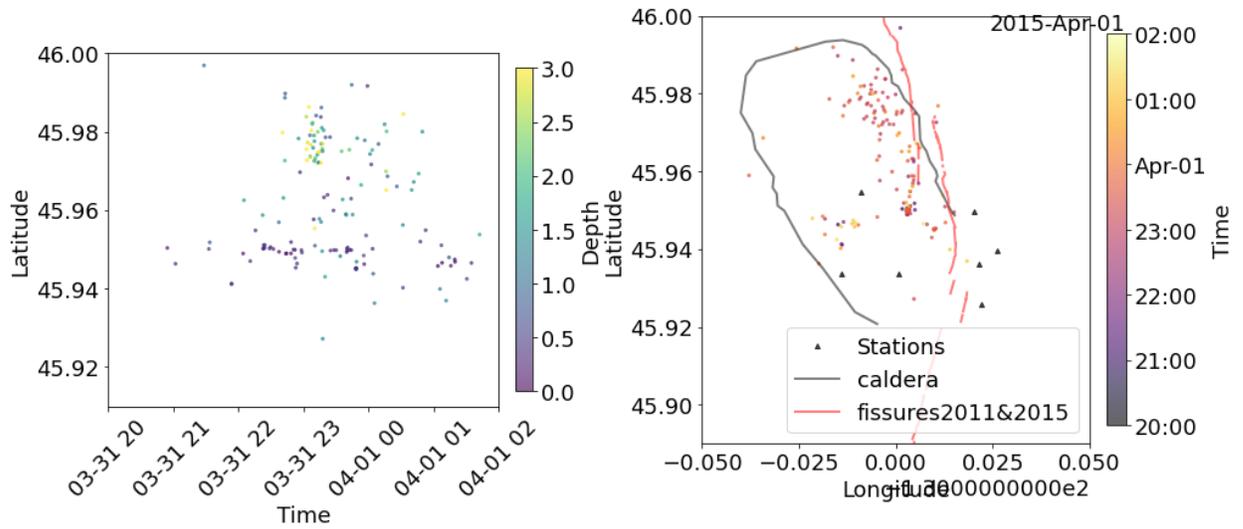
I



J



K



L

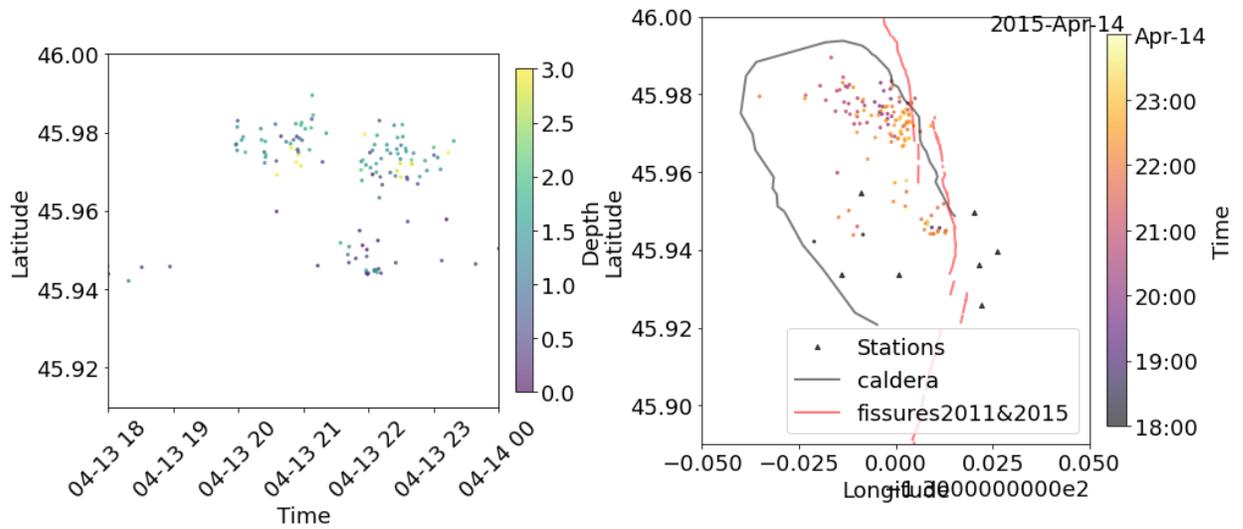


Figure S7. (A - L) Pre-eruption MFE bursts and their locations. Left panel: spatiotemporal evolution pattern of the MFE bursts. Right panel: the locations of MFEs in the same time period colored by time.

Movie S1. (separate file)

Animation of MFEs (left) and EQs (right) activity from Apr 23, 2015 to Apr 25, 2015. Outline of the caldera shown in black line, eruptive fissures of the 2011 and 2015 eruptions in red, and stations in black triangles.