Geophysical Research Letters

Supporting Information for

Volcanic precursor revealed by machine learning offers new eruption forecasting capability

Kaiwen Wang[1], Felix Waldhauser[1], Maya Tolstoy[1,2], David Schaff[1], Theresa Sawi[1], William Wilcock[2], Yen Joe Tan[3]

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 SpecUFEx (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 $Cf \geq 0.8$. 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 Cf ≥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.

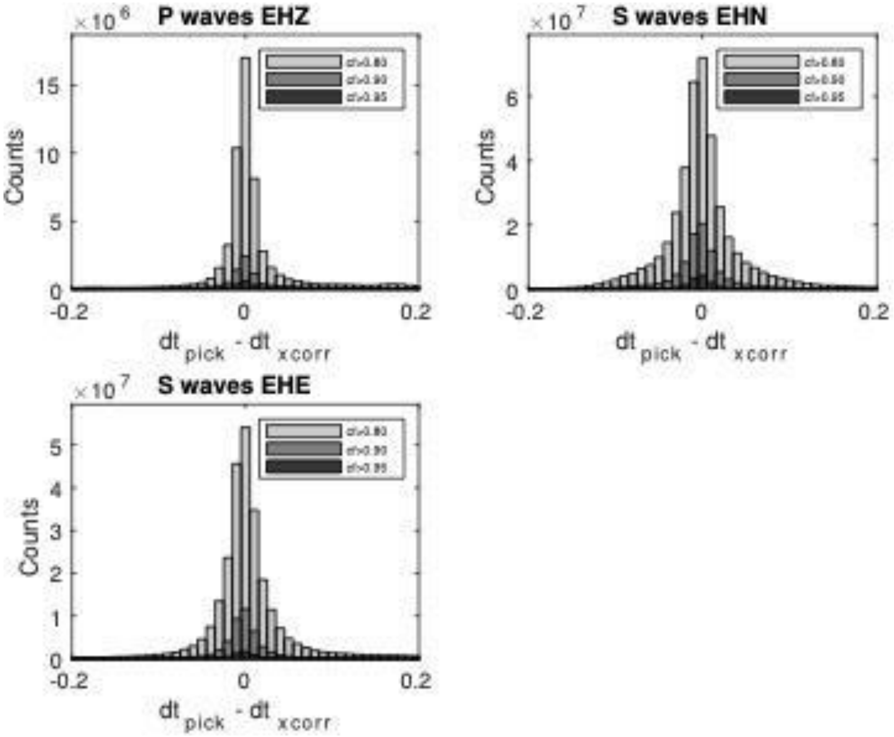


Figure S1. Difference distribution of the picks as compared with the cross-correlation delay times. The bars of light to dark gray colors show cross-correlation delay time measurements of different correlation coefficient thresholds. The three panels show the distribution of the difference between the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

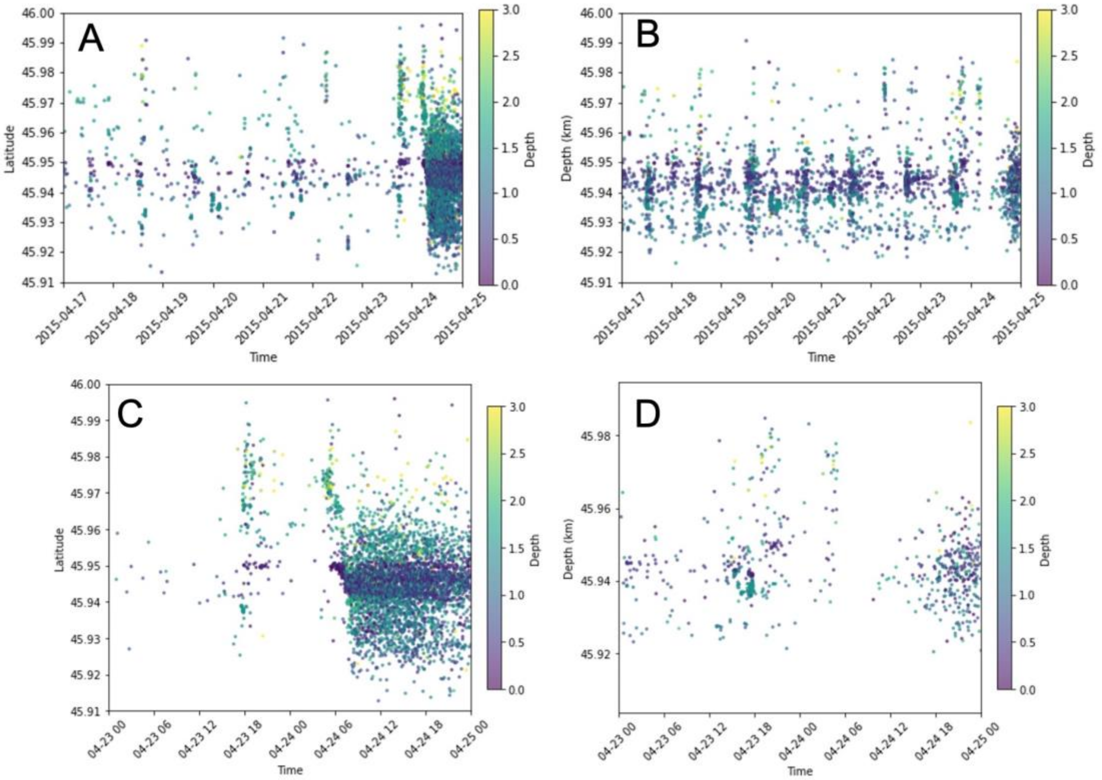


Figure S2. Clustering result using waveforms at AXAS1 station. Spatiotemporal plots of MFEs and earthquakes cluster at ~one week (A and B) and two days (C and D) time scale. (A) and (C) shows MFE activities. (B) and (D) shows regular earthquake activities.

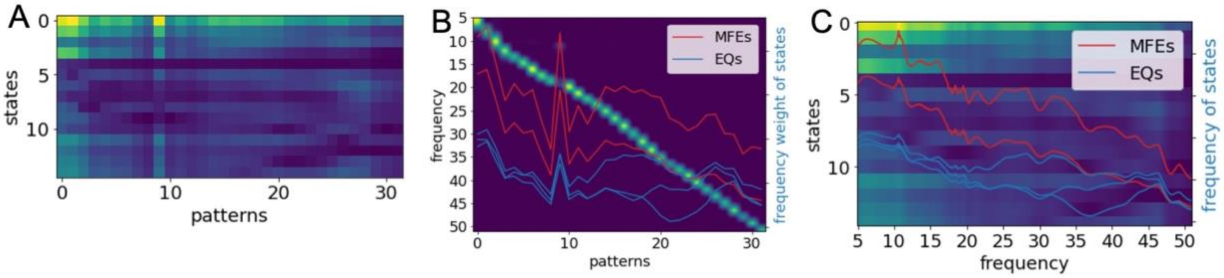


Figure S3. (A) HMM emissions matrix. (B) NMF spectral dictionary overlaid with curves showing frequency weights of the active states in the two stacked fingerprints. (C) Frequency dependent sensitivity kernel of the states in fingerprints. Red and blue curves show the characteristic states of the MFEs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

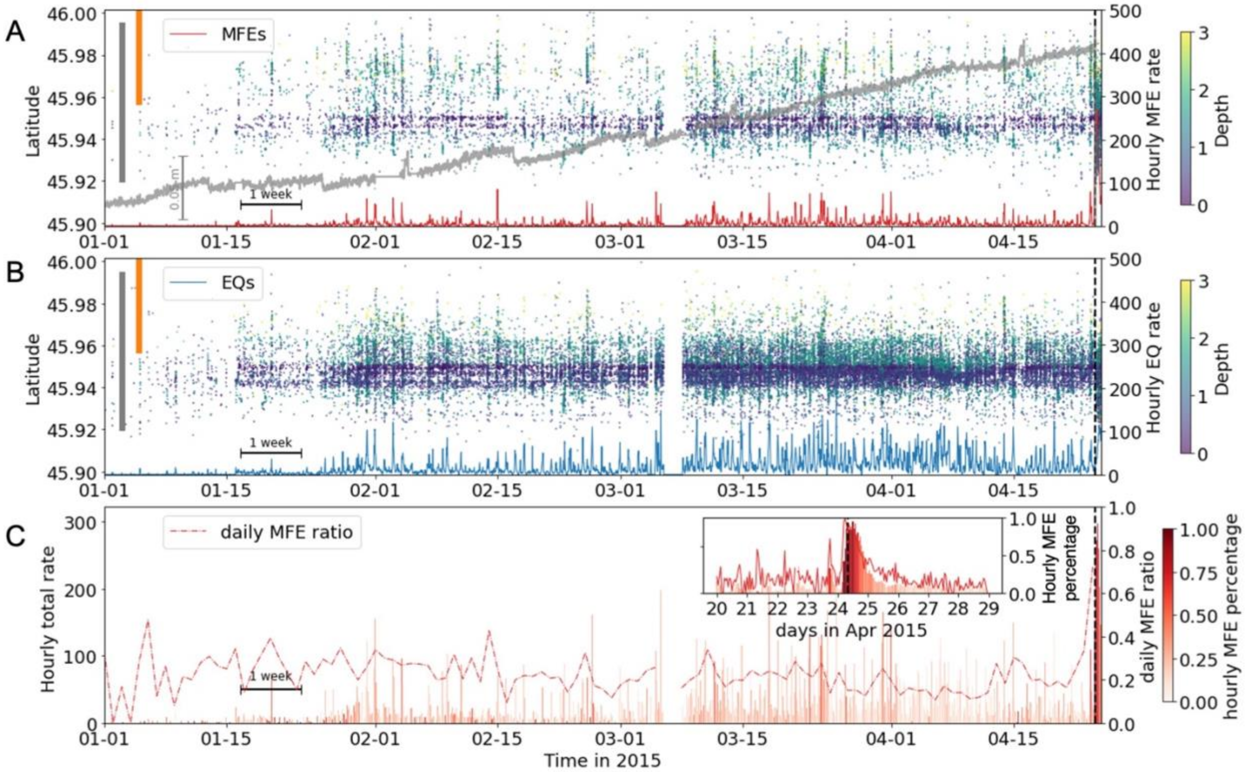


Figure S4. Pre-eruption spatiotemporal evolution of the two spectral clusters and their relative ratio. The MFEs (A) and earthquakes (B) spatiotemporal distribution in ~4 months prior to eruption. (C) Histogram shows hourly percentage of MFEs in all pre-eruption seismicity. Dashed red line shows the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

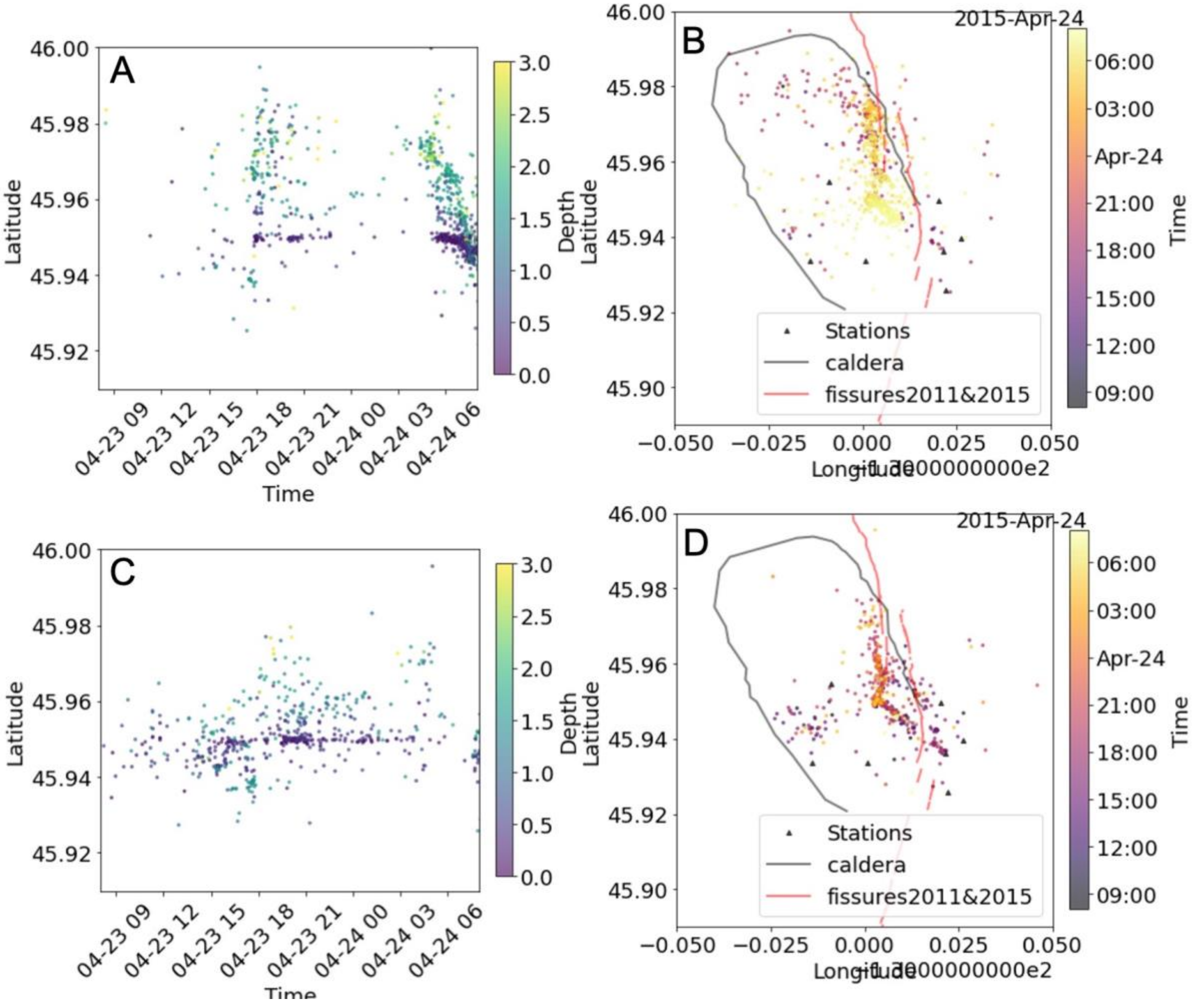


Figure S5. Comparison of earthquakes and MFEs locations. MFEs locations (B) and spatiotemporal plot (A) on the day before eruption. Locations of regular earthquakes (D) and their spatiotemporal distribution (C) in the same time period.

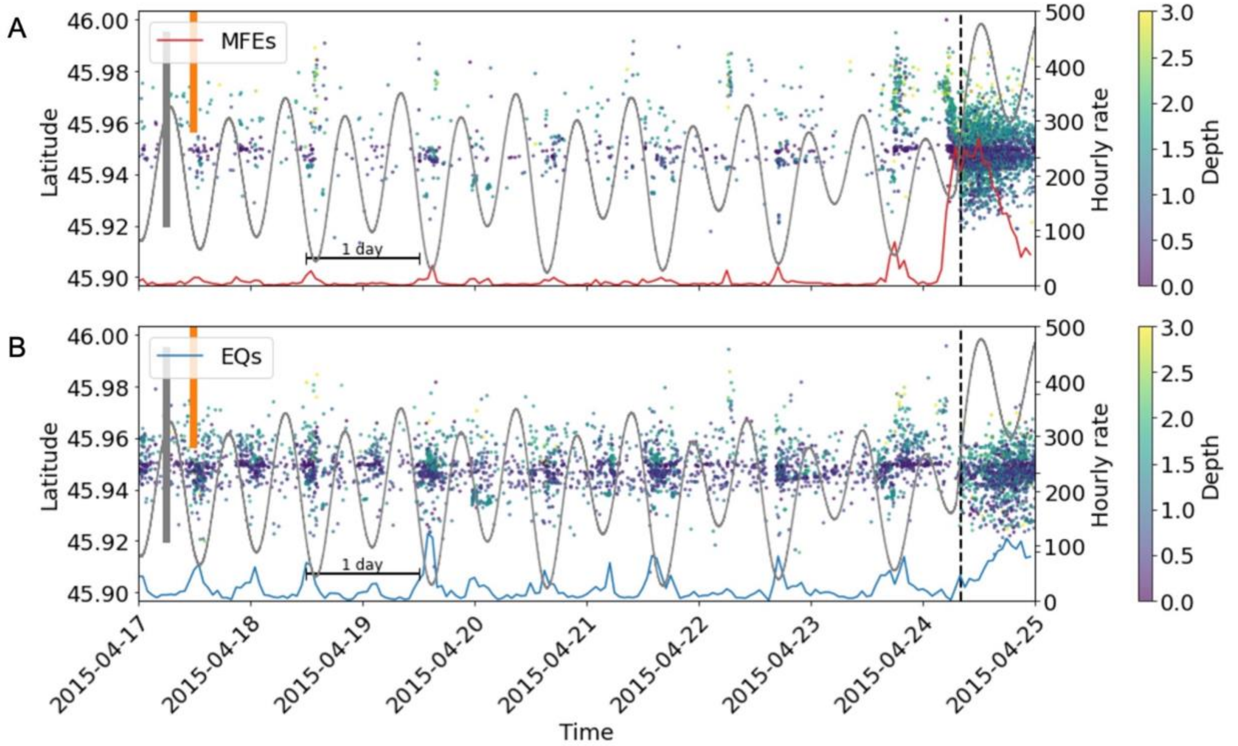
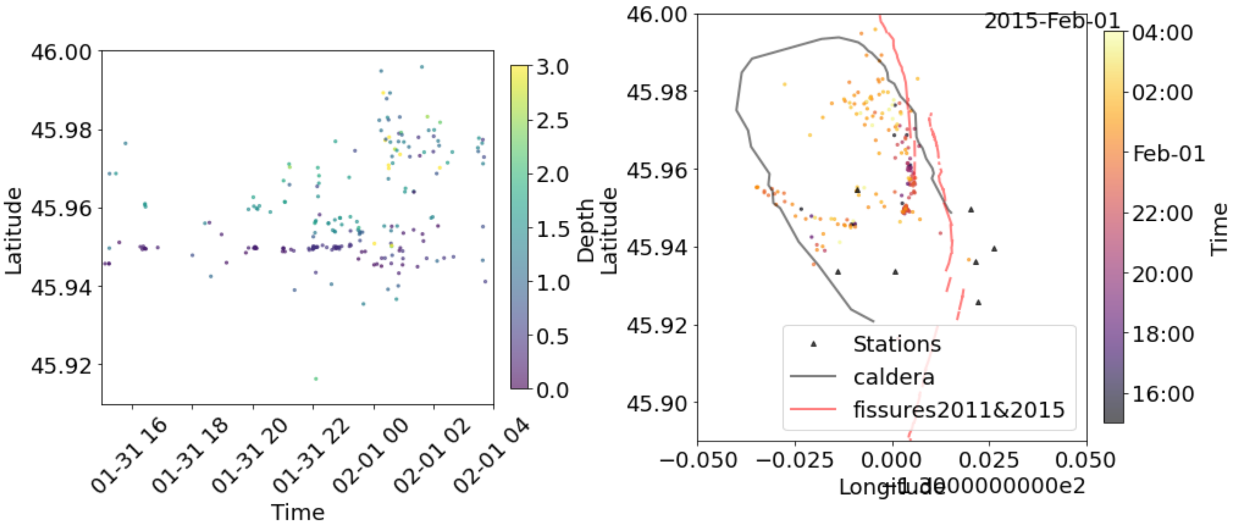
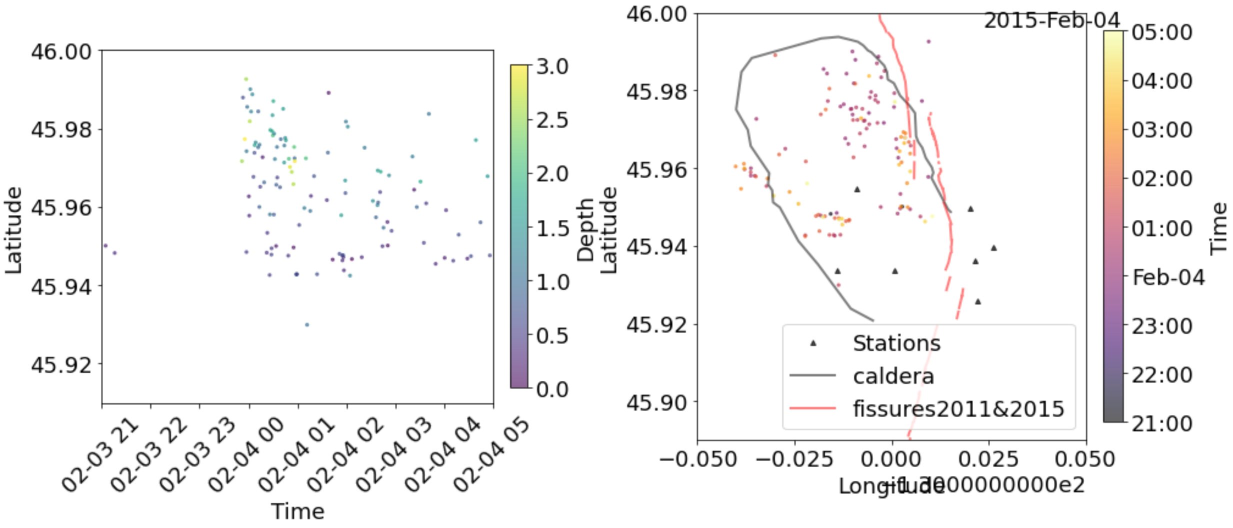


Figure S6. Tidal correlation of the regular earthquakes (B) and MFEs (A). The grey curve shows Bottom-pressure recorder (BPR) measurements at AXCC1. The red and blue curves show the hourly seismicity rate of MFEs (A) and regular earthquakes (B).

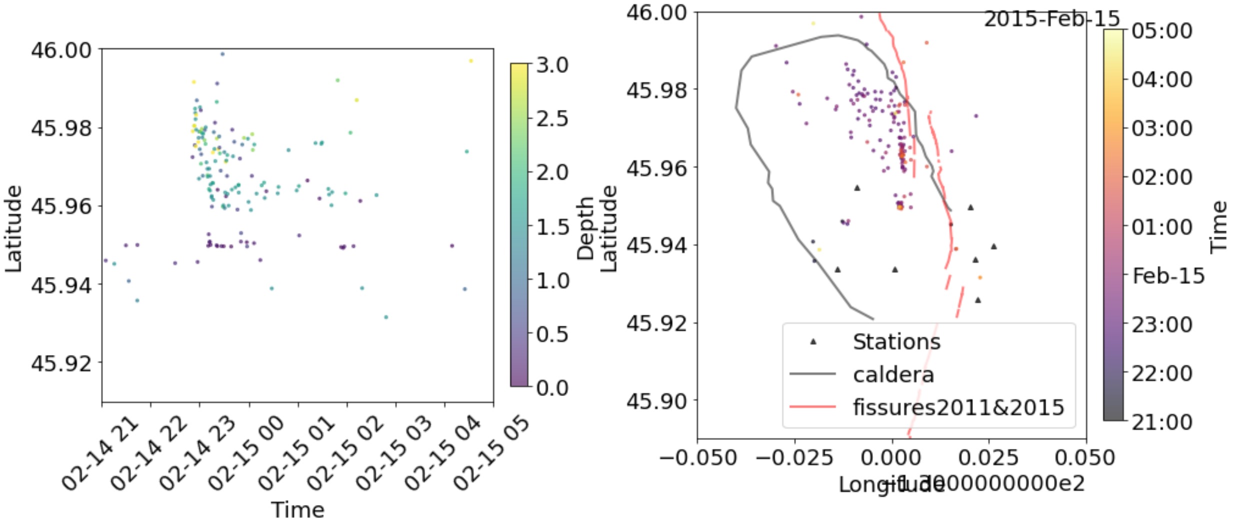
A


46.00
45.98
45.96
45.94
45.92
Latitude
01-31 16
01-31 18
01-31 20
01-31 22
02-01 00
02-01 02
02-01 04
Time
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Depth
2015-Feb-01
46.00
45.98
45.96
45.94
45.92
45.90
Latitude
−0.050
−0.025
0.000
0.025
0.050
Longitude
−1.3000000000e2
Stations
caldera
fissures2011&2015
04:00
02:00
Feb-01
22:00
20:00
18:00
16:00
Time

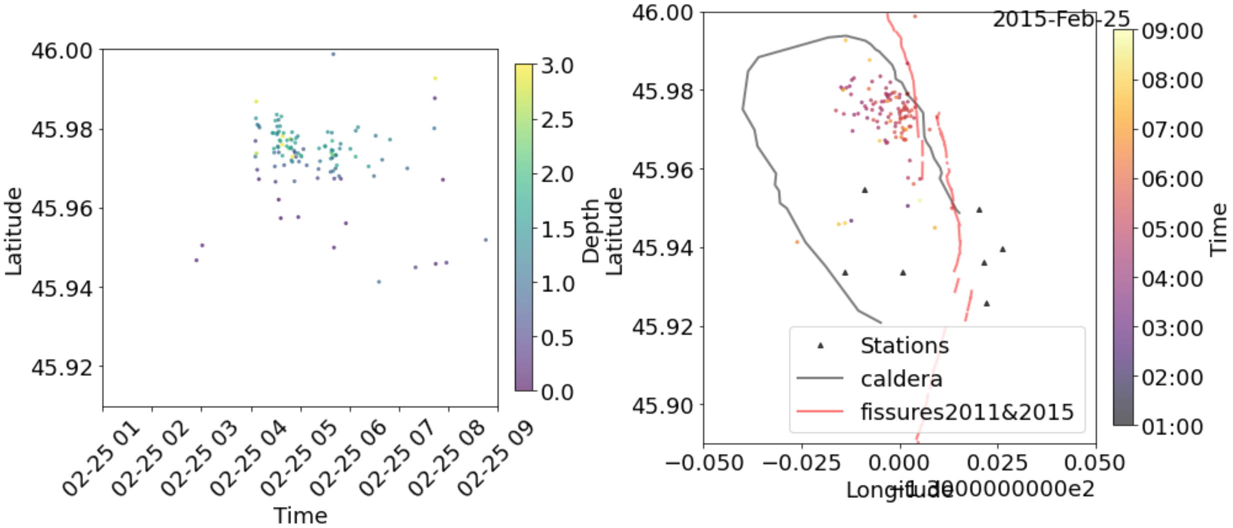
B


46.00
45.98
45.96
45.94
45.92
Latitude
02-03 21
02-03 22
02-03 23
02-04 00
02-04 01
02-04 02
02-04 03
02-04 04
02-04 05
Time
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Depth
2015-Feb-04
46.00
45.98
45.96
45.94
45.92
45.90
Latitude
−0.050
−0.025
0.000
0.025
0.050
Longitude
−1.3000000000e2
Stations
caldera
fissures2011&2015
05:00
04:00
03:00
02:00
01:00
Feb-04
23:00
22:00
21:00
Time

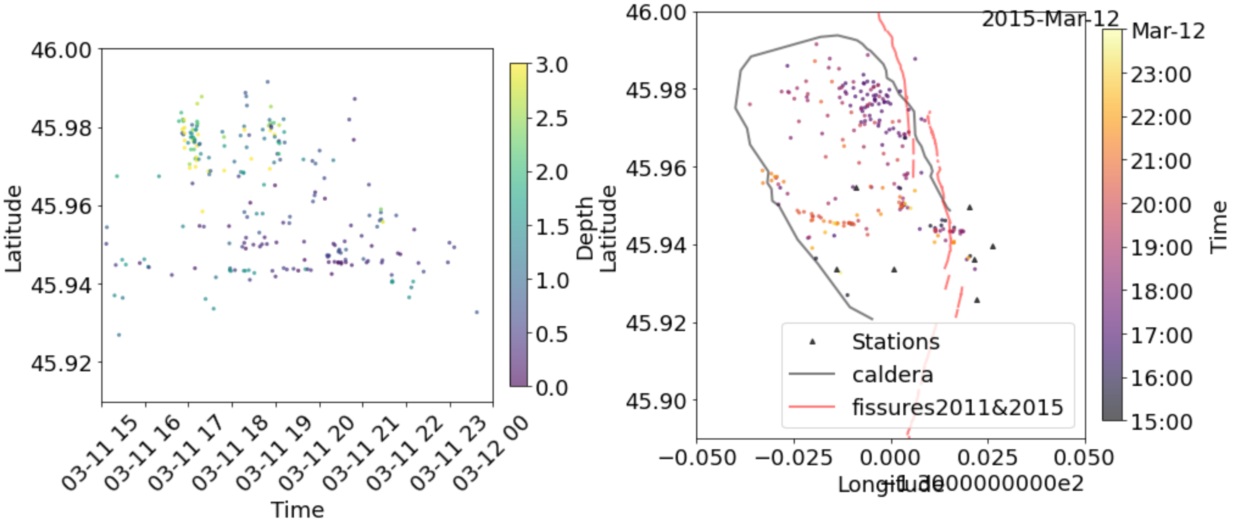
C


46.00
45.98
45.96
45.94
45.92
Latitude
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Depth
02-14 21
02-14 22
02-14 23
02-15 00
02-15 01
02-15 02
02-15 03
02-15 04
02-15 05
Time
2015-Feb-15
46.00
45.98
45.96
45.94
45.92
45.90
Latitude
Stations
caldera
fissures2011&2015
−0.050
−0.025
0.000
0.025
0.050
Longitude
−1.300000000e2
05:00
04:00
03:00
02:00
01:00
Feb-15
23:00
22:00
21:00
Time

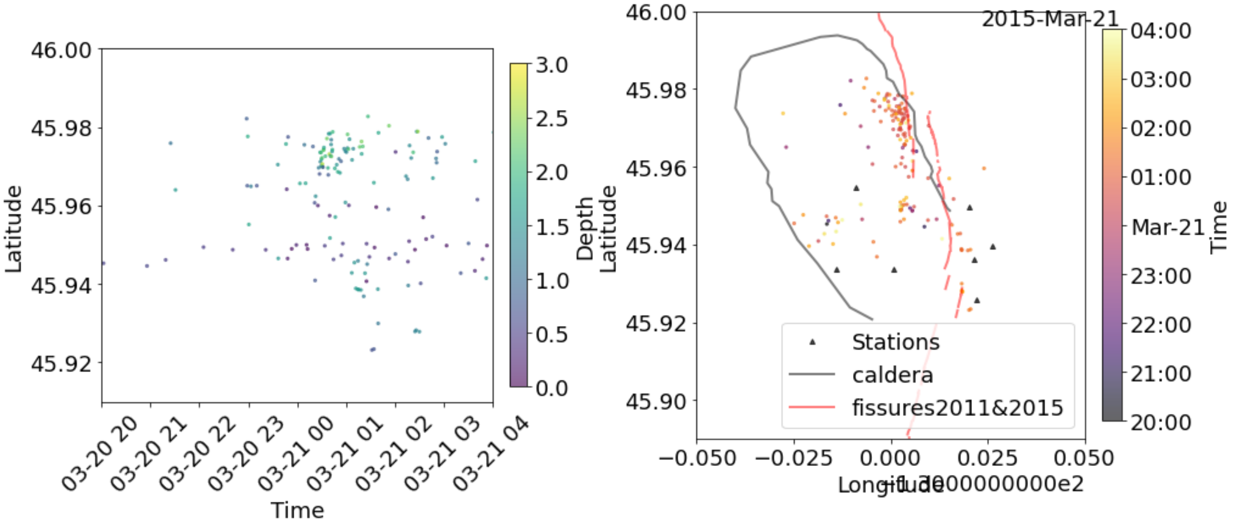
D


46.00
45.98
45.96
45.94
45.92
Latitude
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Depth
02-25 01
02-25 02
02-25 03
02-25 04
02-25 05
02-25 06
02-25 07
02-25 08
02-25 09
Time
2015-Feb-25
46.00
45.98
45.96
45.94
45.92
45.90
Latitude
Stations
caldera
fissures2011&2015
−0.050
−0.025
0.000
0.025
0.050
Longitude
−1.300000000e2
09:00
08:00
07:00
06:00
05:00
04:00
03:00
02:00
01:00
Time

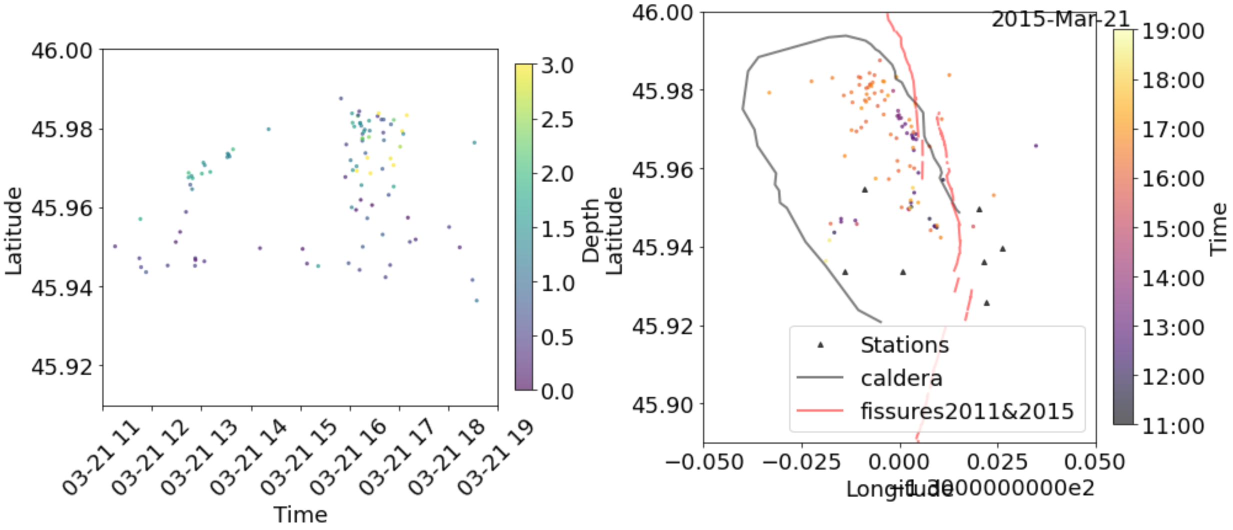
E


46.00
45.98
45.96
45.94
45.92
Latitude
3.0
2.5
2.0
1.5
1.0
0.5
0.0
03-11 15
03-11 16
03-11 17
03-11 18
03-11 19
03-11 20
03-11 21
03-11 22
03-11 23
03-12 00
Time
2015-Mar-12
Depth
Latitude
46.00
45.98
45.96
45.94
45.92
45.90
Stations
caldera
fissures2011&2015
−0.050
−0.025
0.000
0.025
0.050
Longitude
−1.300000000e2
Mar-12
23:00
22:00
21:00
20:00
19:00
18:00
17:00
16:00
15:00
Time

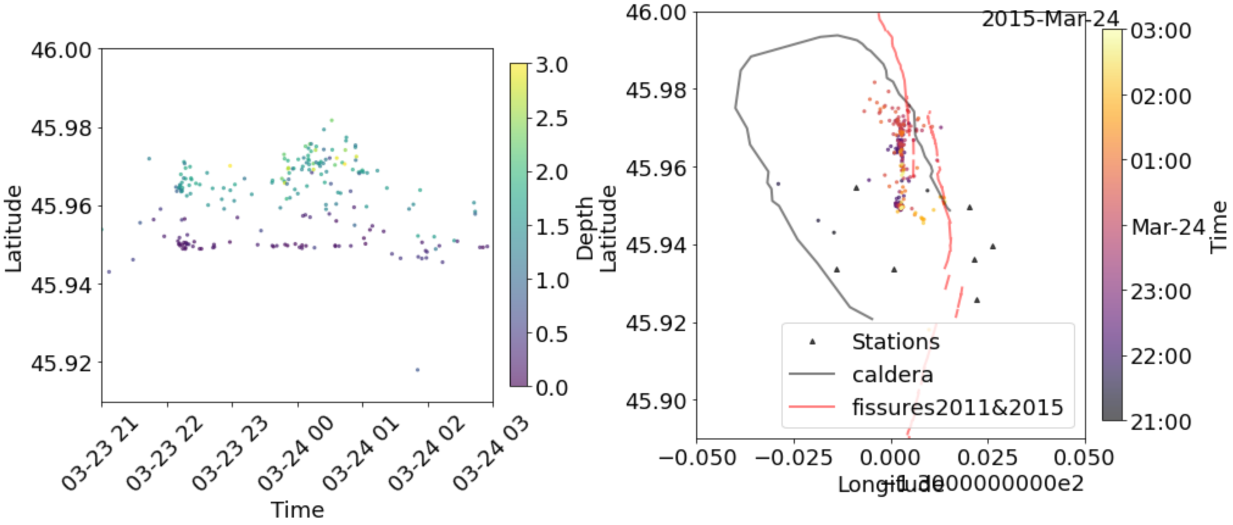
F


46.00
45.98
45.96
45.94
45.92
Latitude
3.0
2.5
2.0
1.5
1.0
0.5
0.0
03-20 20
03-20 21
03-20 22
03-20 23
03-21 00
03-21 01
03-21 02
03-21 03
03-21 04
Time
2015-Mar-21
Depth
Latitude
46.00
45.98
45.96
45.94
45.92
45.90
Stations
caldera
fissures2011&2015
−0.050
−0.025
0.000
0.025
0.050
Longitude
−1.300000000e2
04:00
03:00
02:00
01:00
Mar-21
23:00
22:00
21:00
20:00
Time

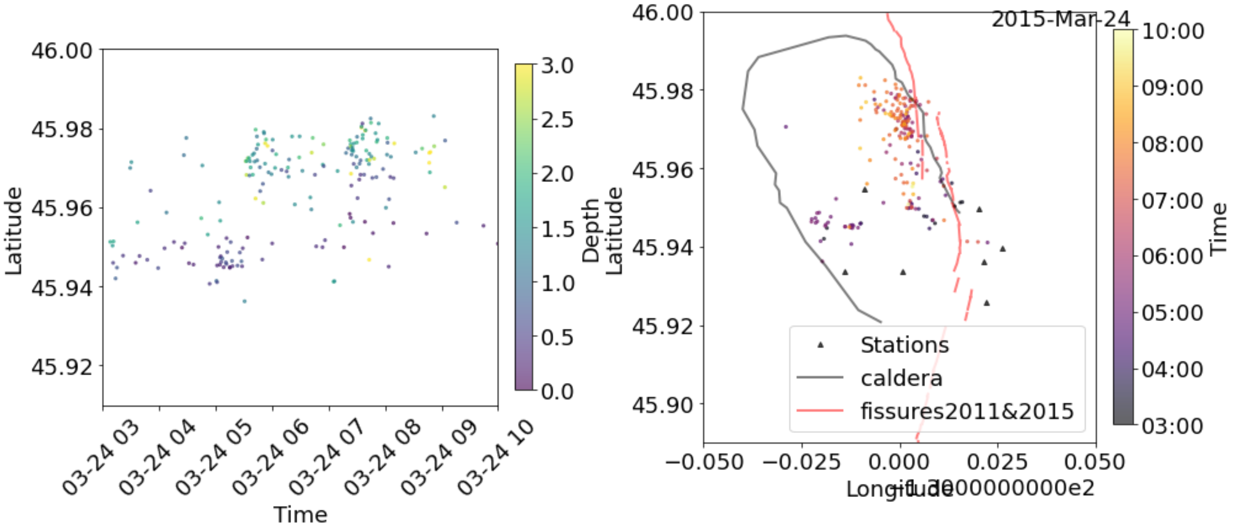
G


46.00
45.98
45.96
45.94
45.92
Latitude
03-21 11
03-21 12
03-21 13
03-21 14
03-21 15
03-21 16
03-21 17
03-21 18
03-21 19
Time
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Depth
2015-Mar-21
46.00
45.98
45.96
45.94
45.92
45.90
Latitude
Stations
caldera
fissures2011&2015
−0.050
−0.025
0.000
0.025
0.050
Longitude
−1.3000000000e2
19:00
18:00
17:00
16:00
15:00
14:00
13:00
12:00
11:00
Time

H

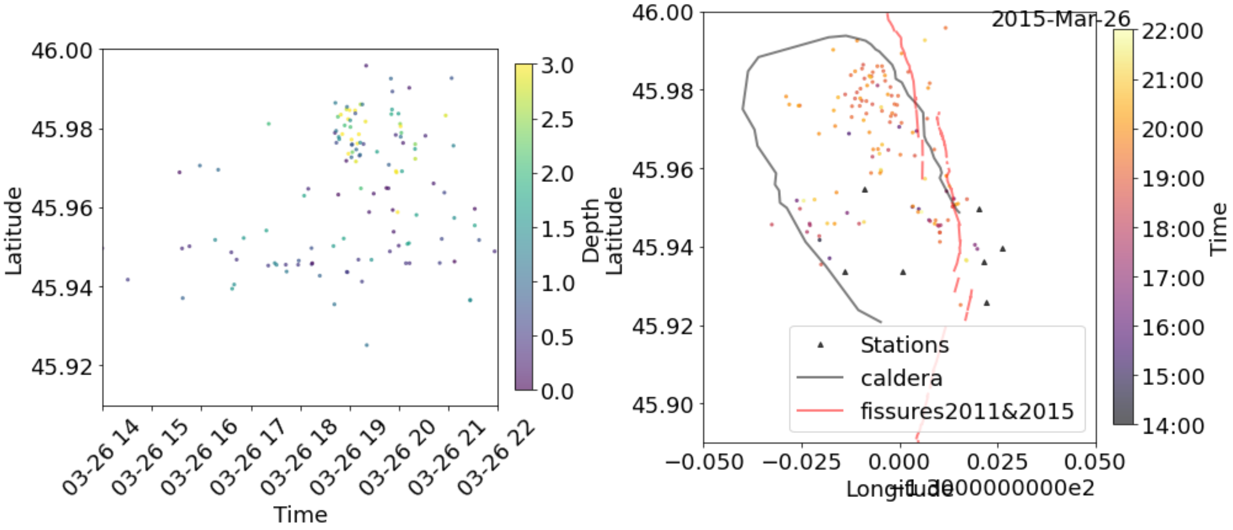

46.00
45.98
45.96
45.94
45.92
Latitude
03-23 21
03-23 22
03-23 23
03-24 00
03-24 01
03-24 02
03-24 03
Time
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Depth
2015-Mar-24
46.00
45.98
45.96
45.94
45.92
45.90
Latitude
Stations
caldera
fissures2011&2015
−0.050
−0.025
0.000
0.025
0.050
Longitude
−1.3000000000e2
03:00
02:00
01:00
Mar-24
23:00
22:00
21:00
Time

I



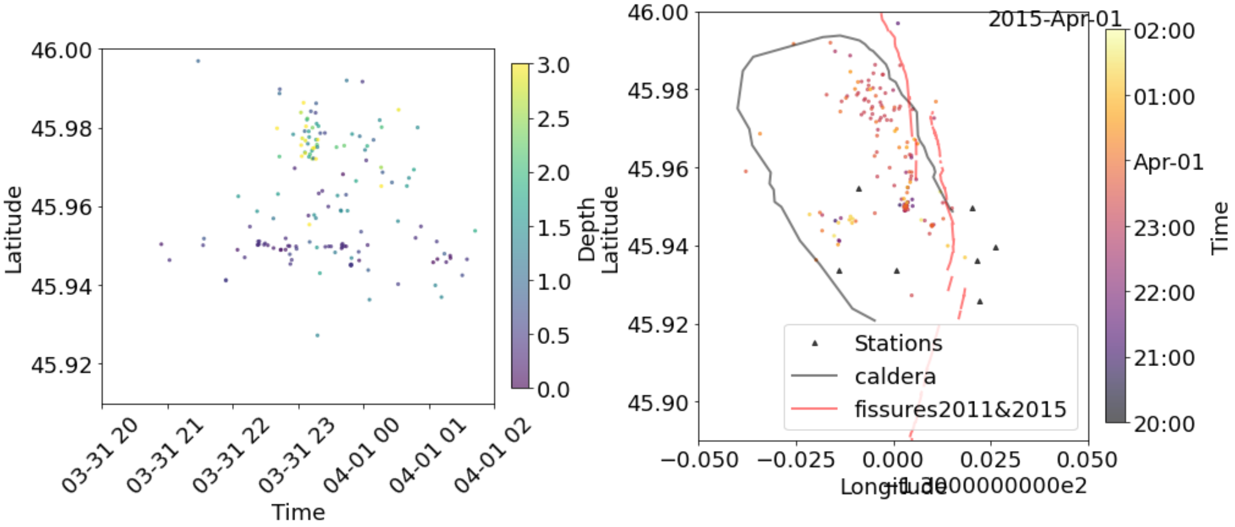
46.00
45.98
45.96
45.94
45.92
Latitude
03-24 03
03-24 04
03-24 05
03-24 06
03-24 07
03-24 08
03-24 09
03-24 10
Time
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Depth
2015-Mar-24
46.00
45.98
45.96
45.94
45.92
45.90
Latitude
−0.050
−0.025
0.000
0.025
0.050
Longitude
−1.300000000e2
Stations
caldera
fissures2011&2015
10:00
09:00
08:00
07:00
06:00
05:00
04:00
03:00
Time

J



46.00
45.98
45.96
45.94
45.92
Latitude
03-26 14
03-26 15
03-26 16
03-26 17
03-26 18
03-26 19
03-26 20
03-26 21
03-26 22
Time
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Depth
2015-Mar-26
46.00
45.98
45.96
45.94
45.92
45.90
Latitude
−0.050
−0.025
0.000
0.025
0.050
Longitude
−1.300000000e2
Stations
caldera
fissures2011&2015
22:00
21:00
20:00
19:00
18:00
17:00
16:00
15:00
14:00
Time

K



L

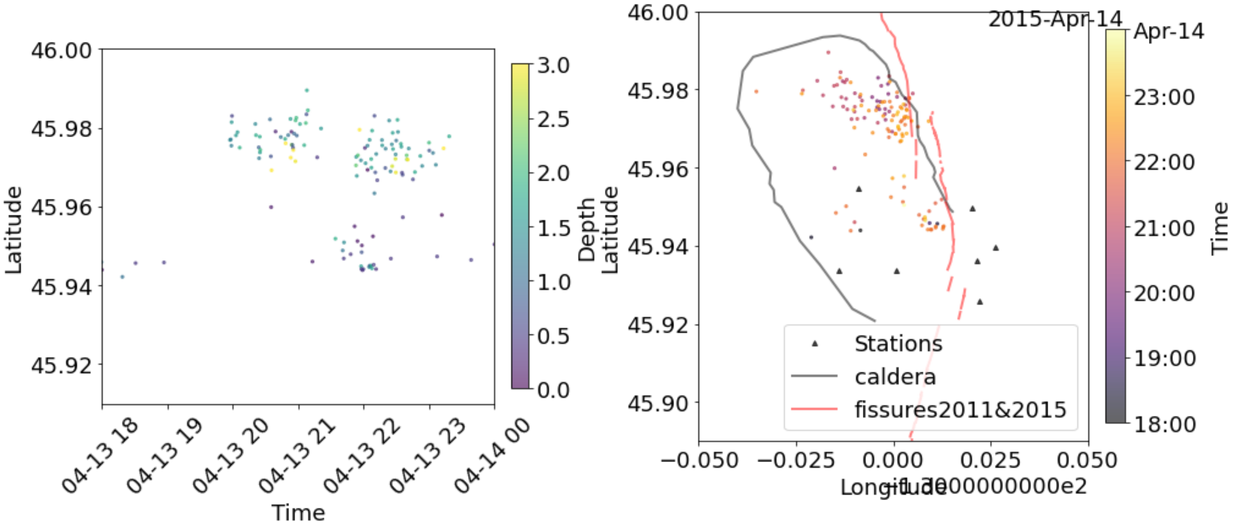


Figure S7. (A - L) Pre-eruption MFE bursts and their locations. Left panel: spatiotemporal evolution pattern of the MFE bursts. Right panel: the locations of MFEs in the same time period colored by time.

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.