SUPPLEMENTARY MATERIAL

Section 1. Template Detection and FAST Procedures

The template matching procedure was as follows: We downsampled all day-long waveforms to 50 Hz and applied a zero-phase band-pass Butterworth filter between 5 and 10 Hz. From these, we built our templates library by cutting the waveforms 0.5s before the estimated P phase arrival and 4.5s after, effectively having templates with length of 5s. To find detections, we opted to scan the processed continuous data using the templates at each individual station. Potential detections are made whenever a template has an absolute cross correlation value equal or greater than 0.8 (See Example in Figure 2). Due to the lack of a large station coverage during some of the stages of the eruption, and since different stations had different noise levels or data quality and we are searching for small-magnitude earthquakes, it is advantageous to perform the match filter technique for each individual station separately. Furthermore, using the absolute cross correlation value as a metric for similarity is more intuitive than using other multi-channel approaches such as the median absolute deviation (MAD) threshold (e.g., Ross et al., 2019).

After we detected the potential new events at each station, we built an associator algorithm that allowed us to systematically find which detections are in fact the same earthquake. We separated all the detections into groups following the ID of the parent template. We then associated the detections of each group by looking at their differential time of arrival at each station and if this matched the travel time lag of the parent template at the same stations (allowing for up to 1 s of error) we grouped them together.

We separately count new events that were detected with a minimum of three stations and which ones were detected with two or even one single station. As it is shown in Section 3 of the Supplementary Materials, many earthquakes that are captured by our procedure by only one stations are in fact small earthquakes that happened close to that station, but due to the highly attenuating medium, the scattered station coverage, and the difference in noise levels at different stations, they either cannot be observed in other parts of the network or they can but do not exceed the absolute cross-correlation threshold of 0.8.

One way that one can estimate the fidelity of the association algorithm is to check for the templates themselves. As also shown in the Supplementary Material Section 3, we can verify that the templates, which were detecting themselves in the continuous waveform data with a correlation coefficient of 1 at each individual station, are indeed associated at all stations as reported in the original catalog. Additionally, since different templates could have found the same new event, we made sure that we are not double-counting events by comparing their detection times. If two or more events were within 10 s of each other, we kept the event that was detected at more stations. If, however, they had the same number of stations, we kept the event with the

highest mean of the cross-correlation values. Lastly, the final step of our process was to make a visual inspection of the waveforms of the detected events to manually discard as many false-positive detections as possible. At this step 611 events were discarded, which primarily stemmed from glitches or gaps in the data.

The FAST process is as follows: the continuous seismic time series is presented as a spectrogram, from which the data is divided into time windows. Then, a 2D Haar wavelet transform is performed on each one of these windowed spectral images to get their wavelet representation, from which the k most anomalous coefficients, i.e. the *k* coefficients that deviate the most from the median value of the *N* coefficients, are extracted. By doing so, one is essentially eliminating the noise of the data (most common values of the Haar wavelet coefficients), while extracting the most characteristic features of each window. For further data compression, all N-k coefficients that are not kept are represented with 0, and the remaining k Haar wavelet coefficients are represented by keeping only their sign, so that the negative values are all represented with -1 and the positive values with 1. Moreover, the new, vastly compressed images are turned into a binary fingerprint by transforming all the -1,0 and 1 into their binary representation. Finally, the similarity search is performed by using a min-wise independent permutation (Min-Hash) algorithm that applies a "hashing" function to map the sparse binary fingerprints into vectors of integers, called Min-Hash signatures. The normalized number of integers in common for two different fingerprints results in their similarity estimate. For more details on the FAST method please refer to Yoon et al. (2015) and Bergen and Beroza (2019).

Section 2. Data quality

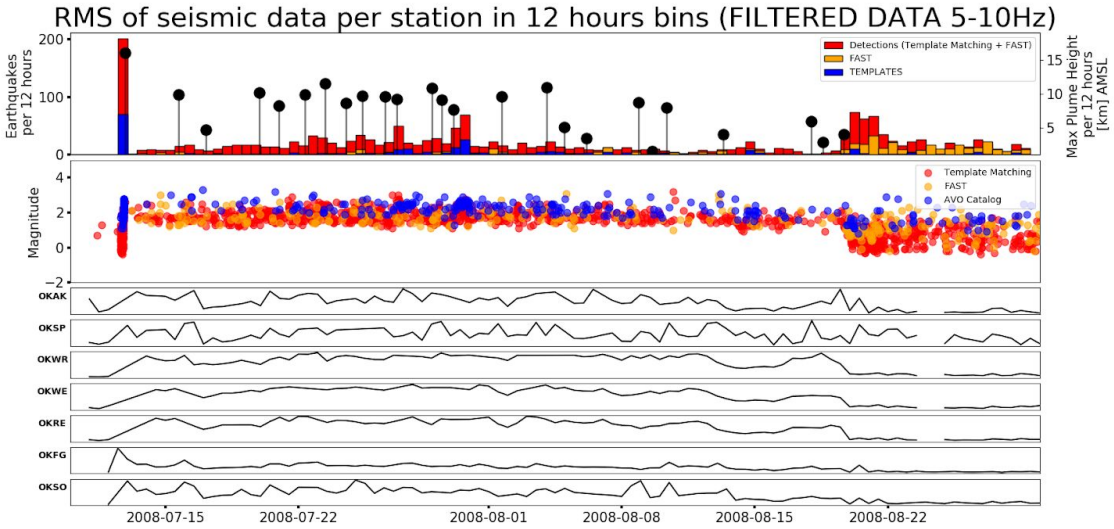

Figure S1. Earthquake rate and magnitude timelines compared to the root-mean-squared of the

seismic amplitude at each station, filtered between 5-10 Hz, which is the frequency band used for template matching. There is a clear increase in RMS throughout most stations at the beginning of the eruption and a clear drop at the end. This could explain the drastic difference in magnitude completeness between before, during and after the eruption.

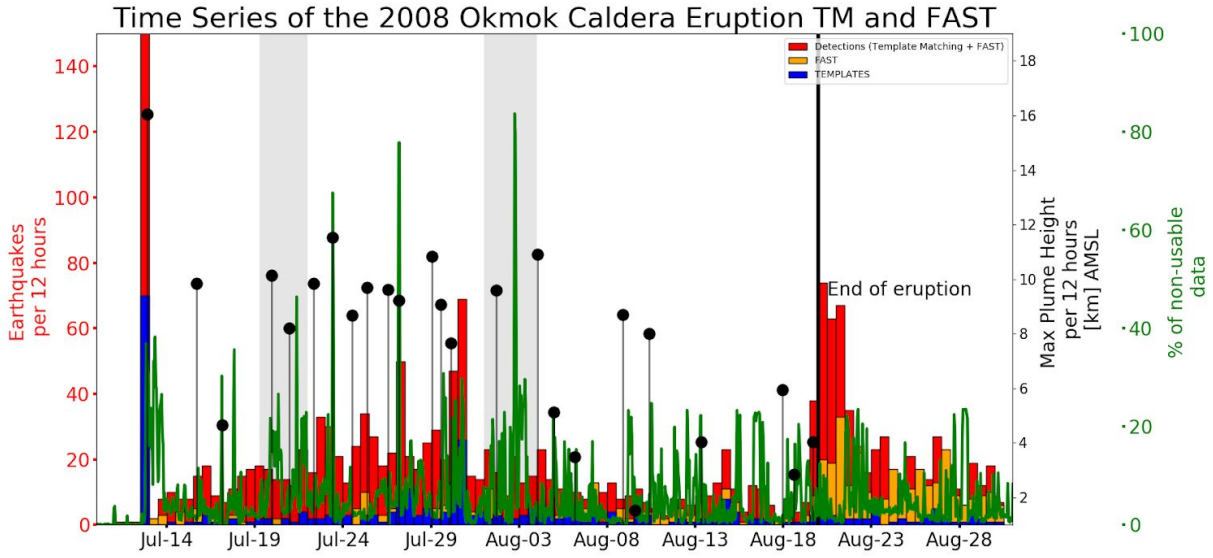

Figure S2. Earthquake rate (bars), plume heights (stem) and percentage of non-usable data (green line). The latter is composed of times when data is non-existent (i.e. gaps), when the traces are clipped, and the times of the false-positive detections that were manually removed by inspection. The lack of earthquakes following the exhalation of the first, and largest, plume can be explained by a high percentage of data outages. Data outages are prominent during the earthquakes bursts of July 27 and July 30, which could suggest that the number of earthquakes that we are finding with both template matching and FAST are a minimum of the actual number of occurrences for the same completeness level.

Section 3. Number of stations for detecting earthquakes

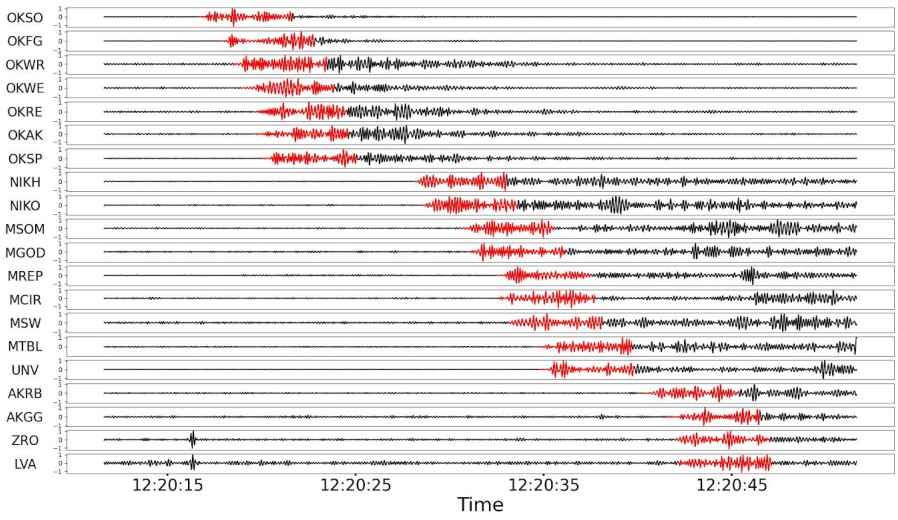

Figure S3. Example of a detection with cross-correlation=1, i.e. a template finding itself. The detection was made at different channels and grouped together by our association algorithm.

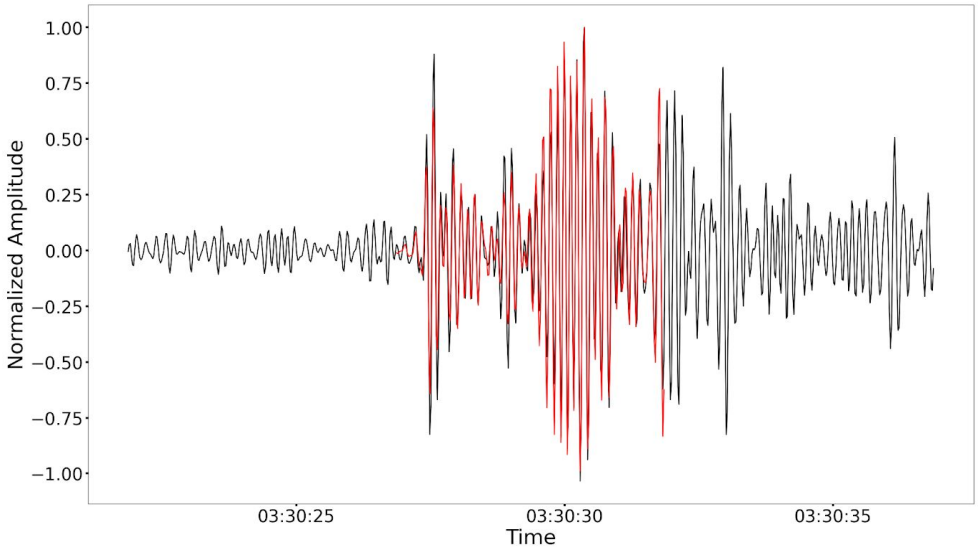

Figure S4. Example of a detection with cross-correlation=0.95 found at only one station (OKAK).

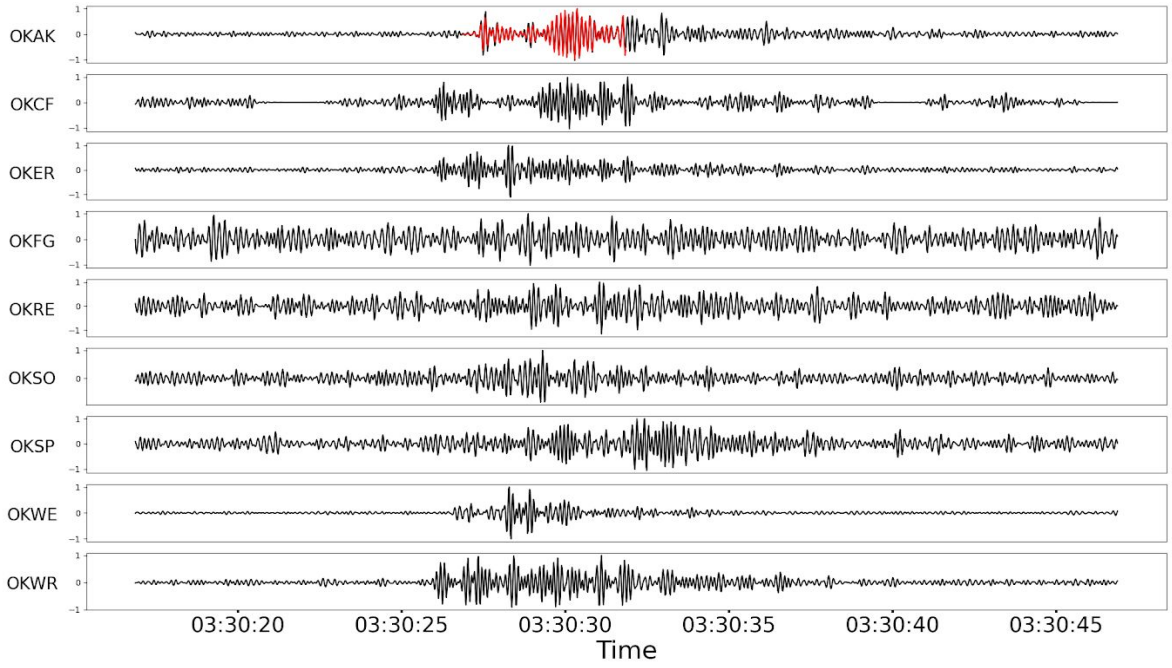

Figure S5. Same earthquake as shown in figure S4, but showing all stations on the Okmok volcano network. The only case where the cross-correlation exceeded our threshold is at station OKAK (template shown in red). We believe that the reason many of the earthquakes are one-station-detections is that small magnitude earthquakes lose most of their power through the attenuating medium, and by the time they get to the other stations they either get lost in the high-frequency noise or they have a low signal-to-noise ratio and so they do not exceed our cross-correlation threshold of 0.8.

Section 4. Earthquake relocation and depths

In this section we show the spatial distribution of all the Earthquakes found in Unmak Island. No magnitude of completeness or spatial constraint is imposed, contrary to Figure 3 in the main text which only shows the events within the caldera.

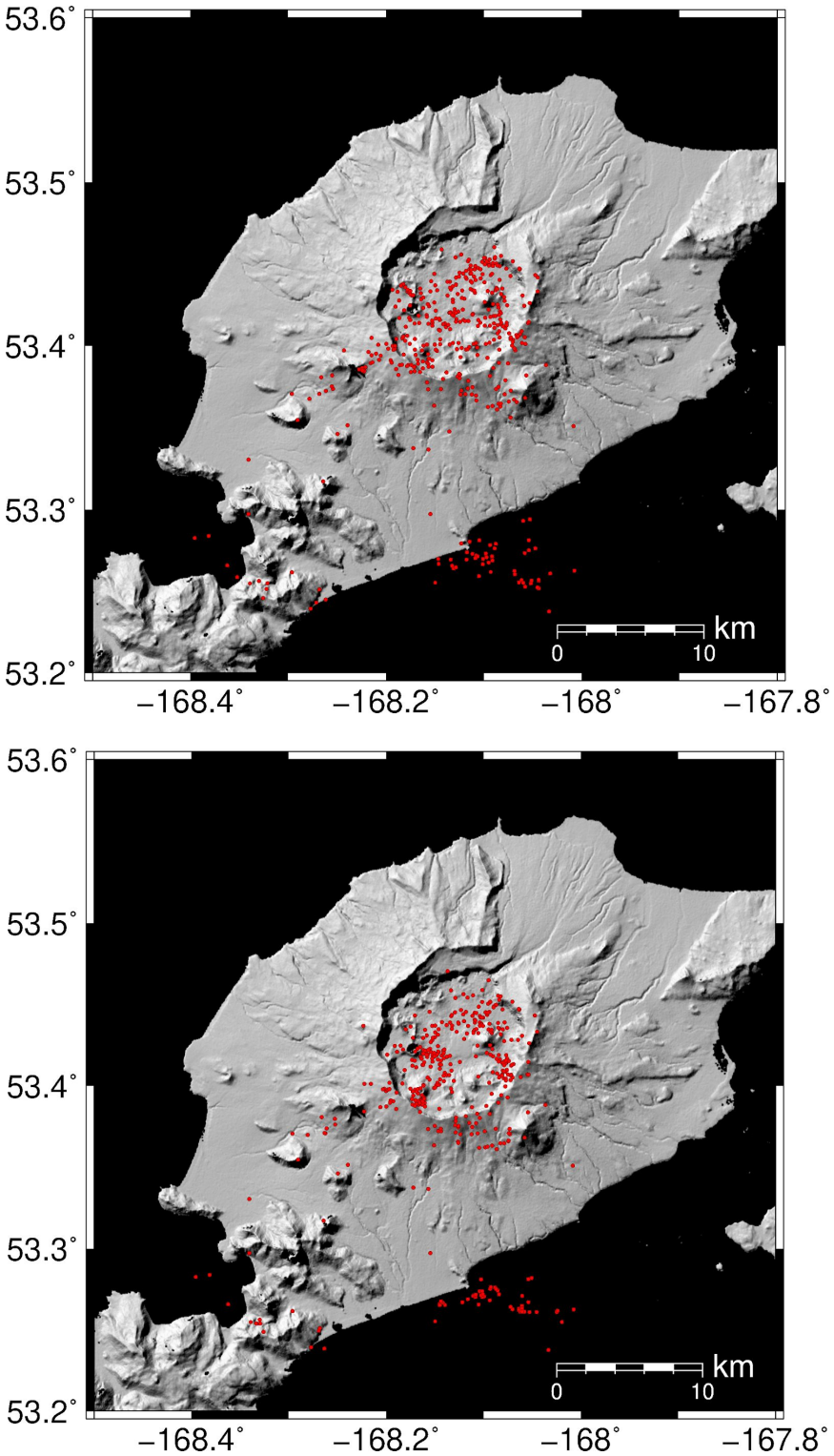

Figure S6. Top: map of the Earthquakes of the original catalog provided by AVO. Bottom: Events are relocated using the GrowClust algorithm.

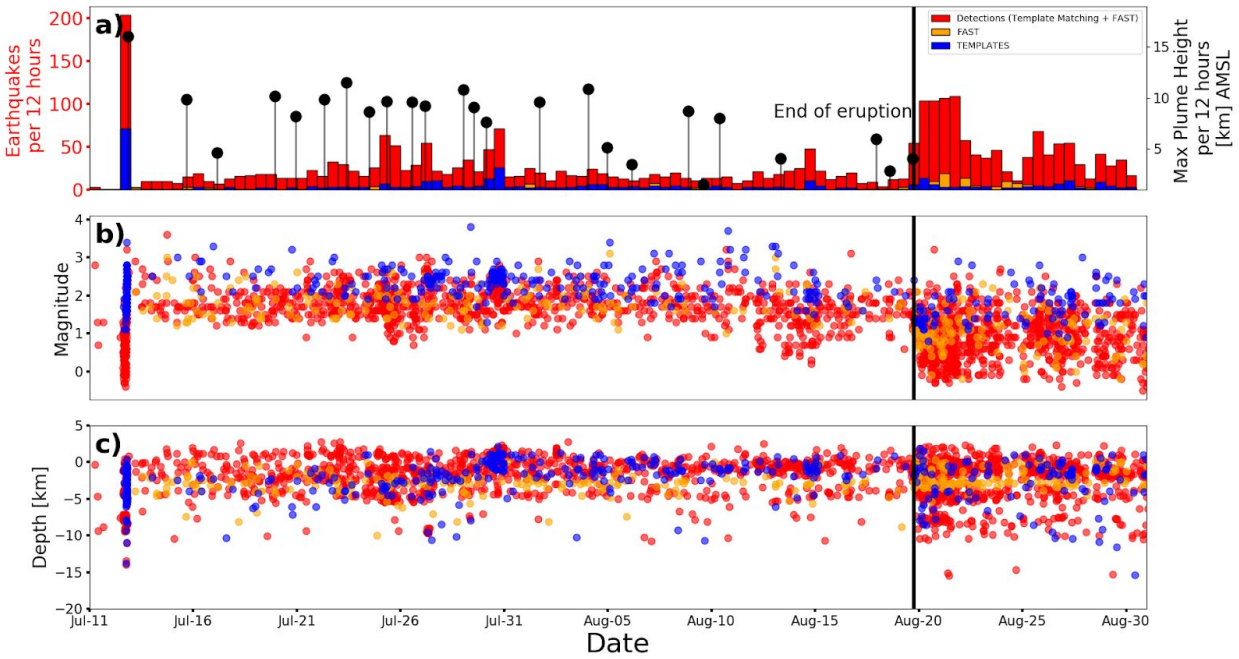

Figure S7. a) Earthquake rate and plume heights. b) Time series of magnitudes. c) Time series of depths.

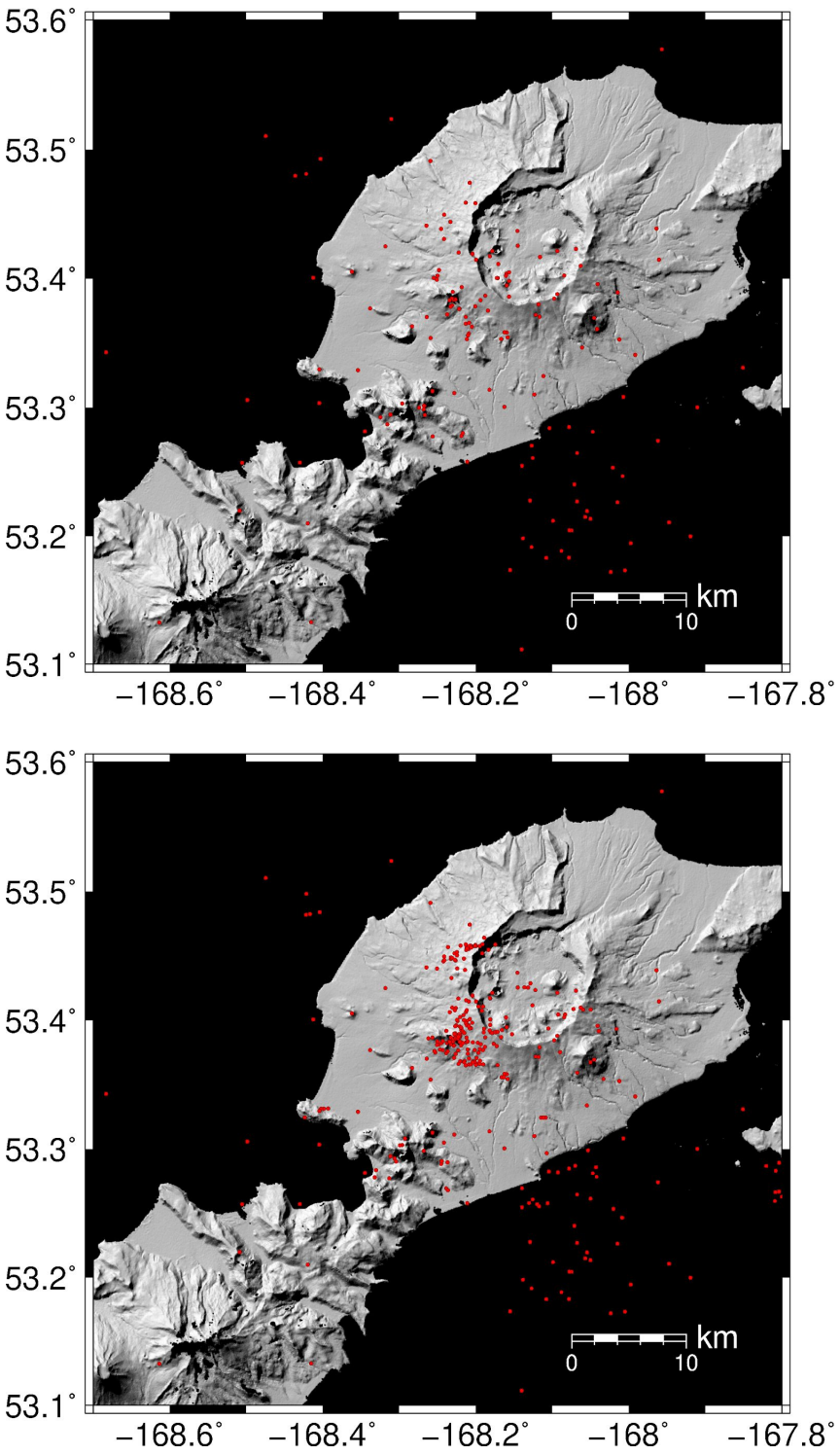

Figure S8. Topt: map of the earthquakes found by FAST and located with HYPOELLIPSE. Bottom: Events are relocated using the GrowClust algorithm.

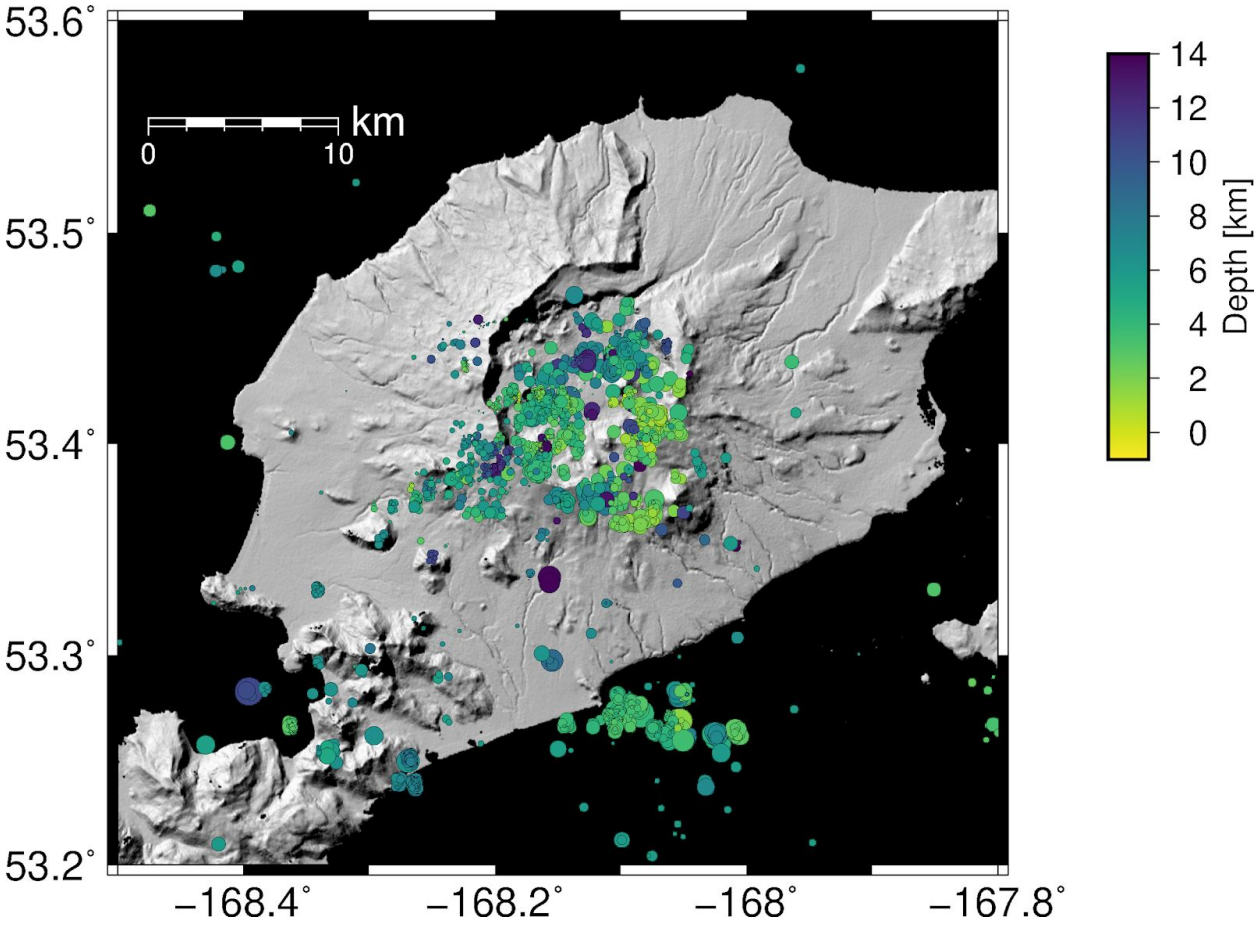

Figure S9. Map showing the spatial distribution of all earthquakes in the final catalog. The main features that are highlighted by the triggered earthquakes during the eruption are the ring-fault structure of the caldera, a long >15km lineation striking SW-NE starting at the edge of the caldera near Cone A and ending close to Inanudak Bay in the south-west sector of the volcano, a ~10km long NW-SE striking cluster off-shore to the south-east of the volcano, and a group of clusters in the geothermal area surrounding Steeple Point, also in the south-east of Okmok volcano.

Section 5. Template matching vs FAST

This section of the supplement shows the difference in efficiency of template matching and FAST at finding earthquakes during the eruptive sequence. Template matching

performs better during the times of the eruption when there were many more templates available. FAST is superior toward the end of the eruption, the location of these earthquakes can be found in Figure S8 of this supplement. We conclude that the FAST method should be complementary to template matching when trying to enhance seismic catalogs and neither should be prefered over the other. Perhaps a more exhaustive search could be performed by utilizing the new events that are discovered by FAST and run them through the template matching machinery.

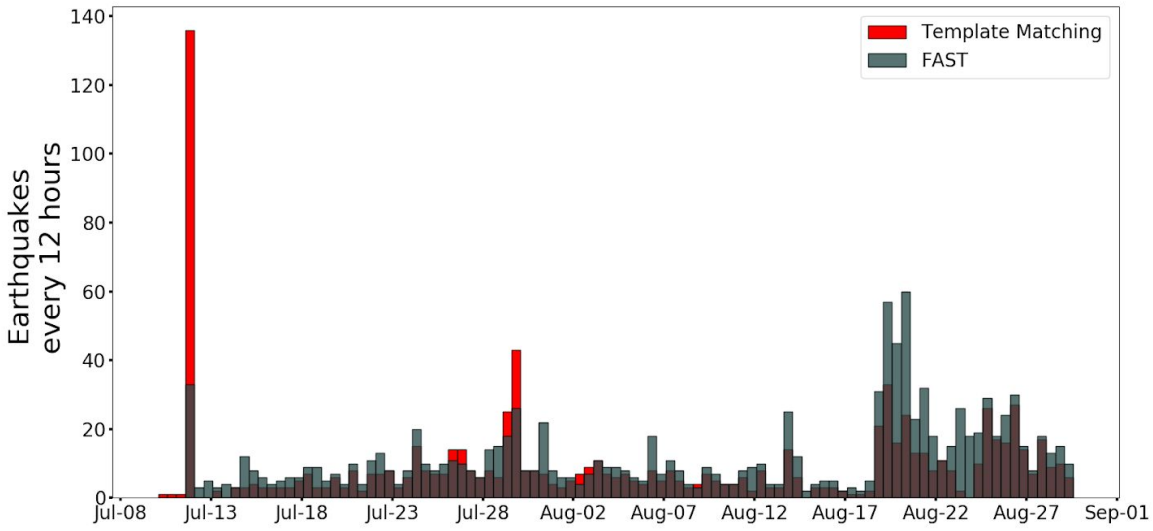

Figure S10. Earthquake rate for template matching (red) and FAST (grey). The most notable differences are that template matching is superior during the bursts of the eruption and FAST performs better after the end of the eruption on August 19.

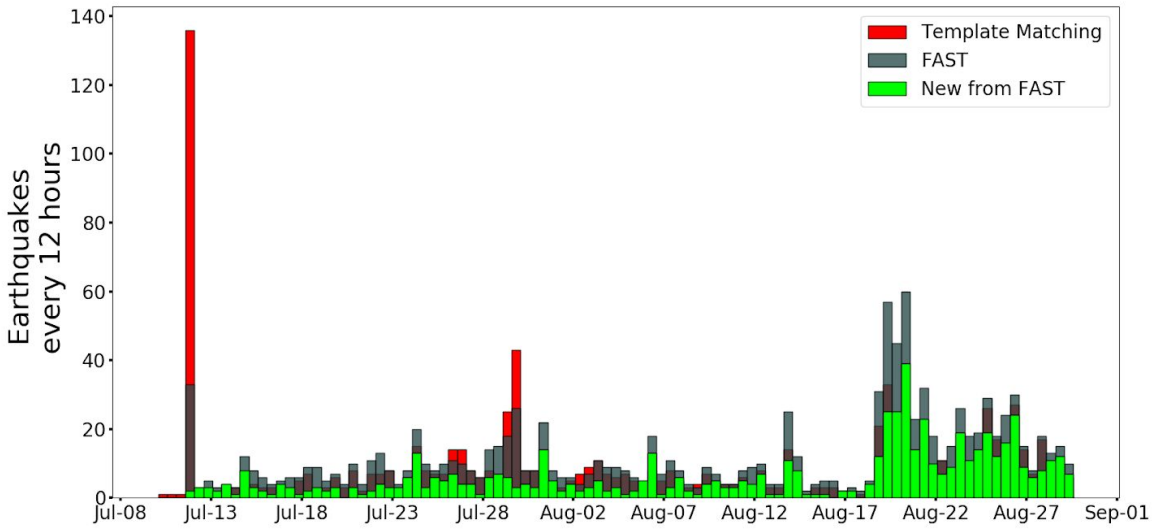

Figure S11. Earthquake rate for template matching (red) and FAST (grey) and the earthquakes that FAST was able to capture and template matching did not (green).

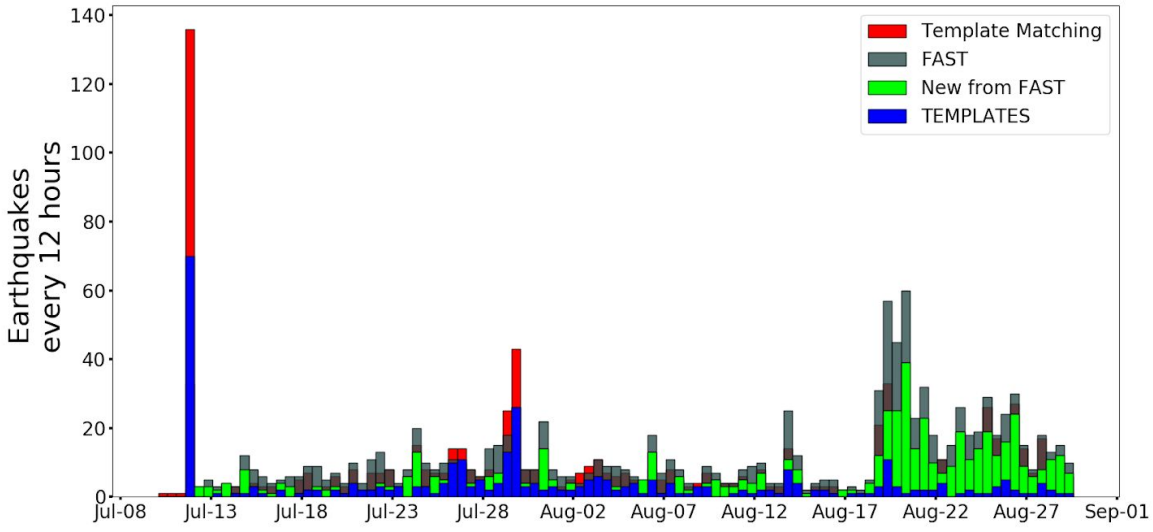

Figure S12. This figure includes the earthquake rate of the original AVO catalog (blue) to show that the times when template matching outperforms FAST are those times when there were much more templates available to search the continuous data.