A machine-learning-based estimation of large amplitude regions for monitoring volcano seismicity

*Yuta Maeda[1]

1. Nagoya University

The seismic network of Mt. Ontake has improved after the 2014 eruption, producing one of the most dense observational networks in the world composed of 9 stations within 1 km of the summit, 17 within 4 km, and 29 within 10 km. This network has recorded long-period (LP) events and tremors which had been considered quite rare at Mt. Ontake (Maeda et al., 2018, JPGU). These signals are attributed to movements and volume changes of shallow volcanic fluids according to studies of similar signals at other volcanoes; thus they are important markers to understand the state and behavior of the fluids. However, despite finding several events, the entire activity of them are not known at Mt. Ontake; it is not easy to manually create an event catalog because of abundance of spurious events (i.e., local wind noise) and difficulty of distinguishing them from real volcano-seismic signals. This is partly related to a difficulty of comparing the records at many stations with a sufficient temporal resolution using a small number of figures. A new strategy to extract useful information from the large number of stations is thus needed.

In this study, we developed a method to investigate a spatiotemporal distribution of large amplitude regions to easily monitor the activities of earthquakes and tremors. The amplitudes could be a measure to monitor the seismicity including those without clear initial motions, but could be affected by local structures beneath the stations. To reduce this effect, we used information of whether the amplitudes are significantly larger than the normal level. We determined a threshold to distinguish significantly large and normal amplitudes from every 5-min-long window of each data trace, based on an idea that Gaussian and non-Gaussian portions of the frequency distribution of the amplitudes are normal and significantly large parts, respectively. This idea is implemented by creating a cumulative frequency distribution of absolute amplitudes in the 5-min-long window using only the smallest N data samples; calculating the fitting error between this distribution and an error function; and searching for an optimal N which minimizes the fitting error. The amplitude corresponding to this optimal N was regarded as the threshold. We then computed the ratio of large amplitude samples in every 1-sec-long window of each trace. We used the spatial distribution of this ratio as the input to a machine learning of a neural network model to investigate the large amplitude region for the 1-sec window. In this way, the spatial distribution of the large amplitude region is investigated for every 1-sec window, which could be used to distinguish true volcano-seismic signals from spurious ones.

We applied this method to the continuous records in November and December 2017 at Mt. Ontake. For most earthquakes occurred in the analysis domain, the large amplitude region distributed within a narrow area around the epicenter or spread with time from the epicentral area to a wider region. For distant earthquakes and local noise, the large amplitude region extended over the entire analysis domain with a small probability. These different patterns of the large amplitude regions were useful to distinguish the true volcano-seismic signals from spurious ones, although there were some missing detections of small earthquakes and tremors. Using the large amplitude region, only one figure is needed for each second, enabling a seismicity monitoring easier than detailed evaluations of continuous waveform records. We are now trying the 2nd stage of the machine learning to automatically detect earthquakes and tremors, which uses the spatiotemporal distribution of the large amplitude as the input, giving 93% success as a preliminary result.

Keywords: Volcano seismology, Machine learning, Mt. Ontake

