

# Machine learning models for aftershock forecasting: Application to the 2016 Kumamoto earthquake sequence

\*Hanyuan Huang<sup>1</sup>, Hiroe Miyake<sup>1</sup>

1. Earthquake Research Institute, University of Tokyo

Machine learning techniques are becoming increasingly prevailing in seismology contributed by the increasing amount of observable information. Considering that the research of seismology now is largely motivated by data from past earthquakes and machine learning mechanism enable itself to extract features and identify unseen signals which may help with earthquake prediction. This research is trying to use mathematically computed parameters known as seismicity indicator to forecast aftershocks on a certain day after the 2016 Kumamoto earthquake sequence. We here used the machine learning techniques as Multiple Linear Regression, Neural Network, k-Nearest Neighbors, and decision tree. To built the dataset based on the unified earthquake catalog extracted from Japan Meteorology Agency, we chose the data including time, magnitude, depth in the range from 1 January 2000 to 1 January 2020 whose magnitude is over 2.0 and depth is shallower than 30 km, in the region of the longitude of 130.0 to 131.3 and the latitude of 32.0 to 33.3, which is located around Kumamoto. To train the model, according to the original data, we refined the data in the unit of a day and count the maximum magnitude, sum number of earthquakes, mean magnitude in one day. Then we calculated the statistic information, the mean value, median value, and maximum value of the magnitudes during the time elapsed 30, 60, 90, 365, and etc., before the target date, and also the mean depth and earthquake energy. However, the linear correlation among energy, depth, and numbers of earthquakes are extremely low even leading the model to perform negatively. Finally, we only use the statistic value of magnitude in different time series as the training features. Few revised training datasets are also applied to different models. We divided the data set and the first 80%, from 1 January 2000 to 31 December 2015, is taken as the training set, the last 20%, from 1 January 2016 to 1 January 2020, is taken as the test set. The 2016 Kumamoto earthquake sequence occurred in April 2016, a series of aftershocks occurred in the Kumamoto region. However, earthquakes with magnitudes over 6 and depth shallower than 30 km have never shown up during the past 20 years in Kumamoto. So, we prepared to focus on the aftershock forecasting after 14 April 2016, the day when the earthquake with magnitudes greater than 6.0 firstly occurred. We tried to forecast the maximum magnitude sequence after the large earthquake in the Kumamoto region. About the results, when observing the training set and its prediction results in many models, because of the huge difference between the quantities of low magnitude earthquakes and high magnitude earthquakes, we found the predicted values are proportionally small compared to real value even when relatively large earthquakes occurred. Because we focus on the sequential series maximum magnitude of each day after the large earthquake, we multiple parameter alpha (=maximum magnitude in training set/maximum magnitude in prediction results) for each forecast results, which led the models confirmed real values better on high magnitudes. The evaluations for models are based on the loss between real maximum magnitude and predicted maximum magnitude in a day during different forecasting duration. We found the four models performed with relatively good accuracies during 30 days elapsed after 4 June 2016, the first M6.5 earthquake occurred, but deviation increased slightly with time elapsed which may be triggered by the increment of days without earthquakes with magnitude greater than 2.0 occur while these models performed less precisely on that. Focusing on the forecasting duration one month later, the four models also exhibited different characteristics. The Multiple Linear Regression and Neural Network models both predicted magnitudes with low loss, but the Multiple Linear Regression model seems strongly influenced by several certain features leading the prediction of model performed somehow similar to the maximum

magnitude on the last day, which showed low robustness. The decision tree predicted with relatively high loss, and it successfully fit the increasing trend of the M7.3 mainshock. In all the models, the Neural Network model which yielded good prediction accuracies whose mean the loss is around 0.5 performed good robustness and is considered can be applied for future magnitude forecast when earthquakes with magnitudes greater than 6.0 occur again in the Kumamoto region.