

Tsunami height prediction using multiple linear regression and L1 regularization

*Shin Murata¹, Yasuhiko Igarashi¹, Toshitaka Baba², Ken-ichiro SATO³, Takane Hori³, Masato Okada¹

1.Graduate School of Frontier Sciences, The University of Tokyo, 2.Institute of Technology and Science, The University of Tokushima, 3.R&D Center for Earthquake and Tsunami, Japan Agency for Marine-Earth Science and Technology

The Dense Oceanfloor Network system for Earthquakes and Tsunamis (DONET) is constructed in the Nankai trough for real-time earthquakes and tsunamis monitoring (Kaneda et al., 2015). DONET1 has 20 monitoring points with seismometers and ocean-bottom pressure gauges on the ocean floor. It is important to make an accurate estimate of maximum tsunami height from monitoring data of the DONET1, because real-time tsunami forecasting could reduce the damage of tsunami disaster. In previous studies, Baba et al., used linear regression to model the relationship between the maximum tsunami height and the average of maximum absolute values of the hydrostatic pressure (Baba et al., 2013). Although the clear linear relationship has been shown in the previous work, we consider that the prediction is more accurate by using the data observed at all the 20 points rather than using only the average value of the sensors.

We use the data observed at all the 20 stations to make an accurate tsunami prediction by applying multiple linear regression. Multiple linear regression provides the relationship between two or more explanatory variables, e.g. the observed data, and a response variable, e.g. the maximum tsunami height. Moreover, we also adopt L1 regularization. L1 regularization is widely used to identify unnecessary input variables in regression and enables us to find useful sensory points for the tsunami prediction.

We found that the generalization error of the predictions in this study is smaller than that of the previous study. Furthermore, linear regression with L1 regularization provides more accurate predictions. Using L1 regularization, we also found that almost all the sensor constructing DONET1 are necessary for making the prediction and it indicates evidence to the value of all the 20 points DONET stations for tsunami prediction. Furthermore, we compare the linear regression with the nonlinear regression, Gaussian Process (GP) [Igarashi et al., submitted] for the tsunami prediction of expected tsunami scenarios. We found that the linear regression provides better prediction when the prediction data is beyond the observed data.

Keywords: Tsunami height prediction, Linear regression, L1 regularization, DONET