

Tsunami inundation prediction by regression model using tsunami database: A comparison between k-means and k-means++ methods

*Masato Kamiya¹, Toshitaka Baba²

1. Tokushima University Graduate School of Frontier Sciences, 2. Tokushima University Graduate School of Industrial Science and Engineering

Seafloor pressure gauges and GPS wavemeters can observe tsunamis propagating offshore before they reach the coast. Using Green's law, we can easily estimate the coastal tsunami height from the offshore tsunami height. As a more advanced method, regression models based on many tsunami simulations are known, such as Baba et al. (2014) and Yoshikawa et al. (2019). These are simple methods but can provide highly accurate predictions for a short processing time. However, the regression models predict the height at only one point on the coast and do not obtain the spatial distribution of the maximum inundation depth. For emergency response after a tsunami disaster, it is desirable to predict coastal tsunami height and the inundation depth distribution. To obtain the inundation depth distribution using the existing regression models, we can extend the method predicting all points in the target area. However, regressions of a large number of points require a long processing time. One solution to reduce the number of prediction points is pre-grouping areas where the inundation depths are always similar. Therefore, we clustered the inundation depths in a target region, Anan City, Tokushima Prefecture, using the k-means and k-means++ methods.

This study used a tsunami inundation database created using about 3500 source fault models in the Probabilistic Tsunami Hazard Assessment for Earthquakes along the Nankai Trough (National Research Institute for Earth Science and Disaster Prevention, 2020). We randomly selected 14 scenarios from the database. The k-means and k-means++ methods ask us to define the number of clusters. We defined the number of clusters being 18 because deviations of the inundation depth in all clusters were substantially small and obtained a cluster distribution. The average, standard deviation, and the maximum of the inundation depth for each clustered region were extracted from the 3500 scenarios. A regression model with power-law (Yoshikawa et al., 2019) applied to the data as response variables and the maximum tsunami height at DONET stations as explanatory variables. Finally, the regression models predicted the tsunami inundation calculated from the 11 cases of the Cabinet Office model.

The clusters obtained by the k-means and k-means++ methods were similar. Accordingly, the prediction results between the k-means and k-means++ methods were also similar. The prediction accuracy of the rias coast clusters is good with the best RMSE of 1.9 m, while that of the plain clusters is poor with the best RMS of 3.2 m. The classification of inundation areas may cause the difference in prediction accuracy. The plains clusters have a broader area compared to the rias coast clusters.

Keywords: cluster analysis, tsunami