Cluster analysis of tsunami inundation distribution for improvement of tsunami early prediction

*Masato Kamiya[1], Toshitaka Baba[1]

1. Tokushima University

Offshore tsunamis can be observed before arrival at the coast using seafloor pressure gauges or GPS wave meters in Japan. The use of these data with well-known Green's law can predict the height of the tsunami at the coast. Similar methods based on the regression of a large number of simulated tsunami results were proposed by Baba et al. (2014) and Yoshikawa et al. (2019). These models are simple but very practical in use because of good accuracy and short processing time. However, the regression models only predict the height at one point on the shore, while disaster prevention offices in charge of emergency operations after tsunami disaster need the information on the distribution of tsunami inundation. To extend the regression models to estimate the tsunami inundation distribution, we may be able to repeat regression analysis for enormous forecast points. But the processing time becomes longer. As a solution, we attempted to reduce the number of forecast points by pre-grouping areas where the tsunami inundation depths are always similar. Therefore, we applied clustering methods to make the pre-grouping area using tsunami inundation datasets.

In this study, we used the k-means method, which is a typical model of the non-hierarchical method of clustering analysis. The analysis area was near Anan City, Tokushima Prefecture. The tsunami inundation database used in the cluster analysis was created by Takeda (2019), which stores the tsunami calculation results in Tokushima Prefecture from 3967 fault models (M7-9) of the Nankai Trough earthquakes (Hirata et al., 2017). We selected the 18 earthquakes yielding particularly large inundation from the tsunami database. Also, we analyzed 11 cases of tsunami models proposed by the Cabinet Office, which are widely known as the Nankai Trough earthquake scenarios. It is noted that the k-means method requires the analyst to determine the number of clusters in advance. This would cause arbitrariness. Therefore, the x-means method (Ishioka, 2000) which automatically determines the number of clusters was also used. As a result, the studied area was divided into 10 clusters by the k-means method and 37 clusters by the x-means method. In the Cabinet Office scenarios, the cluster was divided into 6 clusters by the k-means method and 10669 clusters by the x-means method. Prediction accuracy using the regression model was evaluated for tsunami data in a cluster. The RMSE of tsunami height prediction using a regression model for 18 scenarios in the tsunami inundation database (Takeda, 2019) was improved by 4% from 1.164 m of k-means to 1.119 m of x-means. The RMSE for 11 scenarios of the Cabinet Office scenario was calculated to be 1.083 m with the k-means method.

Keywords: cluster analysis, tsunami