

Cluster Analysis of the GNSS Velocity Field in the Tohoku Area, northeastern Japan

*Yoshiaki Ito¹, Satoshi Miura¹, Yusaku Ohta¹

1. Graduate School of Science, Tohoku University

Recently, the cluster analysis technique has been applied to GNSS velocity data by Simpson et al. (2012) to reveal tectonic boundaries around the western USA. Assuming a number of clusters, the technique classifies GNSS velocities into some clusters, which shows similar characteristics. The optimal number of clusters can be determined by a statistical test (e.g. the gap statistic by Tibshirani et al., 2001). The advantage of this technique is to extract block-like behavior and to identify tectonic boundaries without considering geological and/or geographical informations. Loveless and Meade (2010) constructed a model composed of 20 blocks in Japan (JB1 model). This model in Tohoku area has two block boundaries, one along the Ou backbone range and the other along the eastern margin of the Japan Sea. The purpose of this research is to apply this clustering technique to the GNSS velocity field of the Tohoku Area, and compare with the JB1 model and the known fault system. GNSS data obtained from 298 continuous GNSS stations operated by the Geospatial Information Authority of Japan and Tohoku University are analyzed using the precise point positioning strategy of the GIPSY/OASIS-II software. We obtain the site velocities by fitting a linear function into coordinate time series from 1 January 2010 to 8 March 2011. We performed the cluster analysis for the horizontal components of the GNSS velocity field with the k-means clustering method. First, assuming the number of clusters we label every site with a cluster randomly. Then, we calculate the centroids of each cluster, and relabel each site with the closest centroid. This procedure is repeated until no more relabeling occurs. This method, however, has a disadvantage, namely the result sometimes depends on the initial random labeling. To avoid this problem, we carry out a thousand of clustering, and calculate the sum of L2 norm of every site pair in each cluster. Then the case of minimum sum is adopted as the optimal clustering. There is ambiguity in assuming the number of clusters. We decide the optimal cluster number by the gap statistic, which compares an ensemble mean of the logarithm of the sum of the L2 norm calculated from a random data, and the logarithm of the L2 norm calculated from the observed data. The gap statistic usually increases with the number of clusters and become stable around the optimal cluster number. The optimal number is 2 for our data set. The result demonstrates a cluster boundary, which runs along the Ou backbone range, and roughly coincides with the boundary of the JB1 model. There are two regions where our clustering result mismatches the boundary in the JB1 model. The reasons may be the ambiguity in the clustering method and/or the possible failure in estimating the site velocities.