Machine Learning Geochemical Discrimination of Muddy Tsunami Deposits in NE Japan
by Using SVM, Random Forest, and UMAP.

tomohisa shimada[1], Matsuno Satoshi[1], Diana Mindaleva[1], *Noriyoshi Tsuchiya[1]

1. Graduate School of Environmental Studies, Tohoku University

One of the most important issues in recent earthquake disaster research is the reconstruction of hazard maps by estimating past tsunami inundation areas. For this purpose, it is essential to identify muddy sediments, which are said to account for 90% of the tsunami inundation area, and the purpose of this research is to clarify the identification of muddy sediments using data science, which has been considered difficult in the past.

In the previous study by Kuwatani et al.2014 using data science, SVM (a machine learning method) and CV (a method for evaluating results) were used as identification methods. They found that the identification of tsunami deposits is 100% possible if 11 out of 32 elements are known.
However, there are two problems with this research. The first is that it requires a huge computational cost to find every element that can be used to draw a boundary by brute force.
Secondly, it is not very versatile because the learning machine is built for each sample.

As a solution to this problem, nonlinearization of the boundary and unification of reference parameters are mentioned, and as the novelty of this research is the use of random forest, which is a machine learning method having advantages in terms of versatility and rigor of identification, compared to SVM. Qualitative identification was performed using UMAP, a dimension compression method. Results of SVM and random forest were compared to tsunami deposits described in the outcrop showing good correlation. UMAP clusters allowed to recognize geochemical characteristics of tsunami deposits, such as high Zr content compared to non-tsunami data. Local feature of tsunami deposits from different area clearly displayed by UMAP clustering.

As a result, we confirmed that SVM and random forest can be used as for more precise discrimination of tsunami sediments at the time of observation, and it was confirmed that the tsunami sediment clusters generated on the UMAP two-dimensional graphs were separated by grain size, tsunami generation time, location, and chemical composition.
Finally, we were able to increase the resolution of these three machine learning-based tsunami deposit certifications.