Geophysical Research Abstracts
Vol. 21, EGU2019-1041, 2019
EGU General Assembly 2019

A novel approach for spatial clustering of 2-D earthquake-receiver raypaths using a machine learning algorithm

Himanshu Agrawal and Supriyo Mitra
Indian Institute of Science Education and Research Kolkata, Department of Earth Sciences, Mohanpur, India (himansh78@gmail.com)

Spatial data heterogeneity often introduces bias in the processing of observed data. Zonation of data sources or receivers in the concerned space is one of the major, yet unavoidable reason for this problem. For earthquake-receiver raypath based studies such biases can arise due to heterogeneous raypath density, which in turn can introduce artifacts in the results obtained from inversion of such dataset, eg. performing a linear tomography. In such problems, clustering of similar raypaths can be used to uniformly sample the data space. Our approach to deal with this problem down-weights the denser, more concentrated datasets by a density-based clustering model and then averaging out the nodes spatially among the clusters. We introduce the 2-D CLUST algorithm to minimize raypath heterogeneity spread over a broad area. Our dataset consists of 4432 earthquake-receiver raypaths spanning across India and Tibet. We use earthquake and receiver locations, slope, mid-point and length of the lines joining them, as our input data points to the clustering algorithm. Lines with the slope in some predefined range are grouped together according to input estimates or threshold values. After grouping by slope, the raypaths are clustered within each group using the mid-point values through a density-based clustering (eg. DBSCAN) algorithm. The input parameters for DBSCAN are 'minpts' and 'epsilon distance' used to cluster the raypaths. Following this, the raypaths are finally clustered using length grouping. For our dataset, 4432 earthquake-receiver raypaths are clustered and averaged to a final set of 2601 raypaths with relatively uniform spatial sampling density. The maximum variance in the mid-point of every cluster is 1.98, which is required to be below the input 'epsilon distance' value of 2 for obtaining reliable clustering of the dataset. Our results strongly suggest that 2D-CLUST is a potential tool for earthquake-receiver raypath cluster analysis. We tested our algorithm on surface wave group velocity dispersion measurement dataset for the above mentioned distribution of earthquake-receiver raypaths and observed a significant improvement in the spatial resolution in 2-D tomographic maps.