A novel approach for spatial clustering and averaging 2-D earthquake ray paths using a machine learning algorithm

Himanshu Agrawal1, Supriyo Mitra1,2

1Department of Earth Sciences, 2Centre for Climate and Environmental Studies, Indian Institute of Science Education and Research-Bhopal, Bhopal, Madhya Pradesh, 462026, West Bengal, India.

ABSTRACT

Spatial data heterogeneity often introduces bias in processing of observed data. Zonation of data sources or receivers in the concerned space is one of the major, yet unassailable 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, e.g. 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 depends on the density, more concentrated data sets in 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 an area. Our results strongly suggest that 2-D CLUST is a potential tool for 3-D earthquake-receiver raypath clustering analysis. We tested our algorithm on attenuation tomography computation by Xie and Mitchell (1999) for the dense distribution of earthquake-receiver raypaths and observed a significant improvement in the spatial resolution in 2-D point spreading function recovery maps. Significant improvement is also observed in the checkerboard tests of surface wave group velocity tomography.

PROBLEM IDENTIFICATION AND SOLVING APPROACH

Heterogeneity raypaths in the gridled area having propagation anomalies in the form shown by using the tomography technique. The yellow color is the central cell and white color is the surrounding grids representing the original parameter structure of the gridled area.

Identification raypaths representing the non-uniform propagation within the area shown by using the tomography technique. The yellow color is the central cell and white color is the surrounding grids representing the original parameter structure of the gridled area.

Figure 4: Illustration representing the observed results of the proposed tomography tomography methods also performing in the cluster raypaths (b) whereas being non-true in two of the clusters (c).

The general idea of solving the problem associated with the heterogeneity data is choosing the closely raypaths and then obtaining a spatially heterogeneous dataset out of which the cluster raypaths. It should be noted that the same data should also ensure that the resulting clusters should be highly homogeneous the region as compared to unevenly spread.

METHODOLOGY

Step 1: Grouping according the slopes of raypaths

Step 2: Clustering midpoints of raypaths using DBSCAN Algorithm

Step 3: Grouping according to the length

The clear lines are represented, which are made according to the chosen threshold value (t).

Representing the grouping of raypaths according to the slope.

Illustration representing the application of the algorithm on the example dataset.

TOMOGRAPHY COMPUTATION

Error Calculations

RESULT

1. Point spreading function results using the 4-raypaths tomography

2. Point spreading function results using the dense raypath coverage

3. Surface wave group velocity tomography checkerboard tests for India Eastern region

DISCUSSION

The figures illustrating improvements in the resolution test results, we can infer that the clustering algorithm (2-D CLUST) may serve a potential tool for improving tomography results.

References: Xie & Mitchell (1999), JQ, 100, 181-191

Elder et al. (1994), KQ, 304-330

Acknowledgments: We thank P. Sudhakar Reddy for the valuable inputs, Steven Madison and Srikanth for the keenness in our study of surface wave tomography. We acknowledge financial support from DST-INSPIRE and IISER Kolkata.

PACIFIC INSTITUTE FOR THE ENVIRONMENTAL STUDIES