Applying cluster analysis to seismic tomography models:
Uncovering similarities and differences in the spatial and spectral domains

Moloud Rahimzadeh Bajgiran\textsuperscript{1} and Lorenzo Colli\textsuperscript{2}

\textsuperscript{1}University of Houston, Earth and Atmospheric Science, Houston, United States of America (mrahimzadehbajgiran@uh.edu)
\textsuperscript{2}University of Houston, Earth and Atmospheric Science, Houston, United States of America (lcolli@central.uh.edu)

In recent years, several types of Machine Learning (ML) methods have been employed by Earth scientists to extract patterns and structures from multi-dimensional feature spaces. In this regard, images of the mantle obtained by different seismic tomography (ST) models are diverse datasets with varying structures due to their different theoretical approximations and input data. In this work, we apply an unsupervised ML method, K-means clustering, on ST models to explore their similarities and differences to improve our physical understanding of the Earth's interior. The K-means clustering method requires ST models to be standardized in a three-dimensional domain. For this purpose, we implement a weighted average technique to resample ST models to radial structural zones with uniform horizontal grid resolutions. However, the homogenized ST models still have $10^3$–$10^4$ parameters, which need to be distilled into a small number of summary features. Feature selection is thus a key part of this study: features should be independent from unphysical effects of inversion choices (e.g., the damping factor) and should instead capture the essence of the geological structure. Preliminary results obtained using the center of mass as the attribute to represent the longest wavelength part of the mantle structure show that P-wave and S-wave models do not cluster separately. Therefore, compositional anomalies do not play an essential role at these spatial scales. We plan to expand our analysis by including more summary attributes from both the spatial as well as the frequency domain.