Statistical Modeling of AIRS Level 3 Quantization Data

Tao Shi
Department of Statistics
Ohio State University
taoshi@stat.osu.edu

Abstract

Atmospheric Infrared Sounder (AIRS) has been collecting temperatures, water vapor mass-mixing ratios, and cloud fraction at various atmosphere pressure levels. It generates 35 dimensional vectors at each 45km ground footprint in each satellite path in its level-2 data. The level 3 quantization data (L3Q) summarize valid level-2 data in each 5 degree by 5 degree latitude-longitude grid box during a time period by a set of representative vectors and their associated weights. The specialty of the data set is that the observations are empirical distributions. Most statistical methods are mainly developed for handling datasets whose observations are in R^d. Statistical inference for this type of data is an open problem. We start with the commonly used Mallows distance as a measure of distance between two distributions and investigate the performance of on various type of multivariate analysis methods, such as MDS, Isomap, and Clustering.We then build a mixture model on empirical distributions with each component being a Gaussian type distribution. Finally, we will address some statistical questions such as classification and prediction on AIRS L3Q data. This is joint work with Dunke Zhou (OSU) and Amy Braverman (JPL).