Large Scale Analysis of Archival Data

Gerald T. Ruch (University of St. Thomas), Charles E. Woodward (University of Minnesota)

The existence of large archives of astronomical data combined with the increasing complexity and diversity of theoretical models implies a need for new sophisticated analysis techniques. We combine Principal Component Analysis (PCA) and Artificial Neural Networks (ANN) to create a powerful set of data visualization and analysis tools. We apply our tools to circumstellar disk spectra from the Spitzer Cores to Disks (c2d) program using the two layer disk model proposed by Chiang et al. (2001) and Dullemond et. al. (2001) (CGPLUS) as well as the model developed by Whitney et al. (2003).

Model Space
Performing PCA on ~40,000 spectra from the model (the sample grid) provides a new space to visualize observational data and compare it with the model. In the model space plot on the right, 71 c2d spectra (diamonds) are superimposed on a 2D histogram of the sample grid (the model locus). This reveals that the c2d sources are divided into two populations. Additionally, many Population I sources lie outside of the model locus indicating that they are not well represented by CGPLUS.

ANN - Mapping Model Space to Parameter Space
An ANN trained on the sample grid, since it can interpolate between grid points, acts as a fast proxy for the model. Above, it is used to show how a data point moves when a single model parameter is adjusted (the parameter trajectory). The large crosses are the original data points (one Pop I and one Pop II). The diamonds are parameter trajectory. The first panel shows that Pop I and Pop II sources are related through the Inner Wall Height. Pop I sources have little or no inner wall while Pop II sources have a moderate wall. The second and third panels show redundancy between the Grain Size Index and the Minimum Grain Size.

ANN-Constraint
Because the ANN is fast (2500 samples/sec) it can quickly characterize the error surface for a given observation. This allows us to determine the degree to which model parameters are constrained by the observational data.

Whitney
The 3D model of Whitney et al. (2003) includes sources with a circumstellar envelope. The green points are samples with no envelope while the red points have an envelope. Pop II sources have a moderate wall. The second and third panels show redundancy between the Grain Size Index and the Minimum Grain Size.

References