Factor analysis techniques in single-particle aerosol mass spectrometry

Maria A. Zawadowicz

Single-particle aerosol mass spectrometry

This figure has been removed due to copyright restrictions.

The figure is from
Thomson, D.S., M.E. Schein, and D.M. Murphy, Particle analysis by laser mass spectrometry WB-57 instrument overview, Aerosol Science and Technology, 33, 153-169, 2000. Please see Figure 1.

The challenge

This figure has been removed due to copyright restrictions.

The figure is from
Murphy, D. M., A. M. Middlebrook, and M. Warshawsky, Cluster analysis of data from the Particle Analysis by Laser Mass Spectrometry (PALMS) instrument, *Aerosol Sci. Technol.*, 37, 382-391, 2003.

Principal component analysis

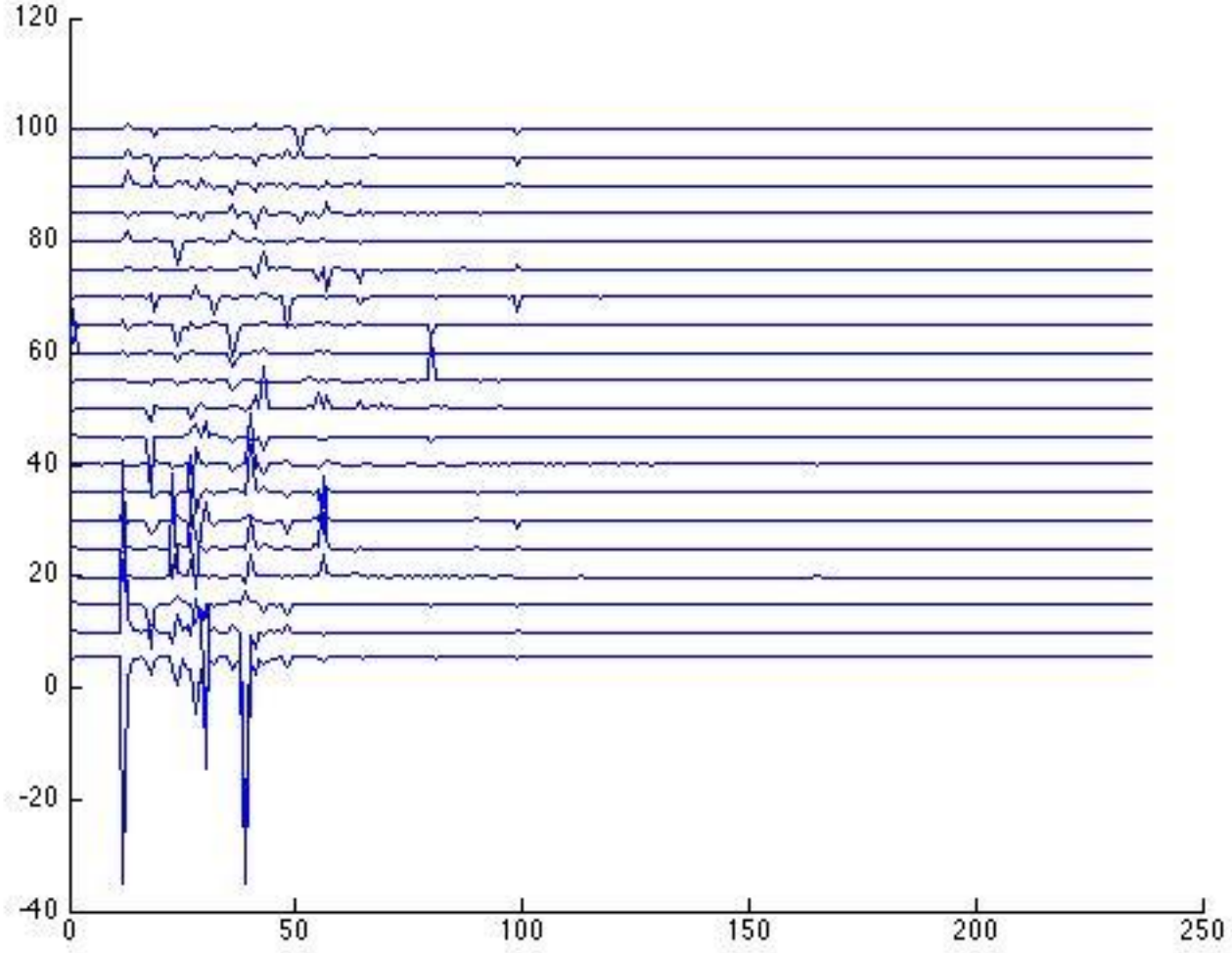
Probabilistic PCA: PCA together with EM to handle large data sets and missing values. Algorithm written by Jakob Verbeek, et al. 2002.

If y is a data vector and x is a vector of principal components,

- E-step: estimate x vectors and missing values of y

- M-step: maximize the expected joint log-likelihood of x and y

Principal component analysis



Independent component analysis

1. Starts with the principal components derived from the previous algorithm.

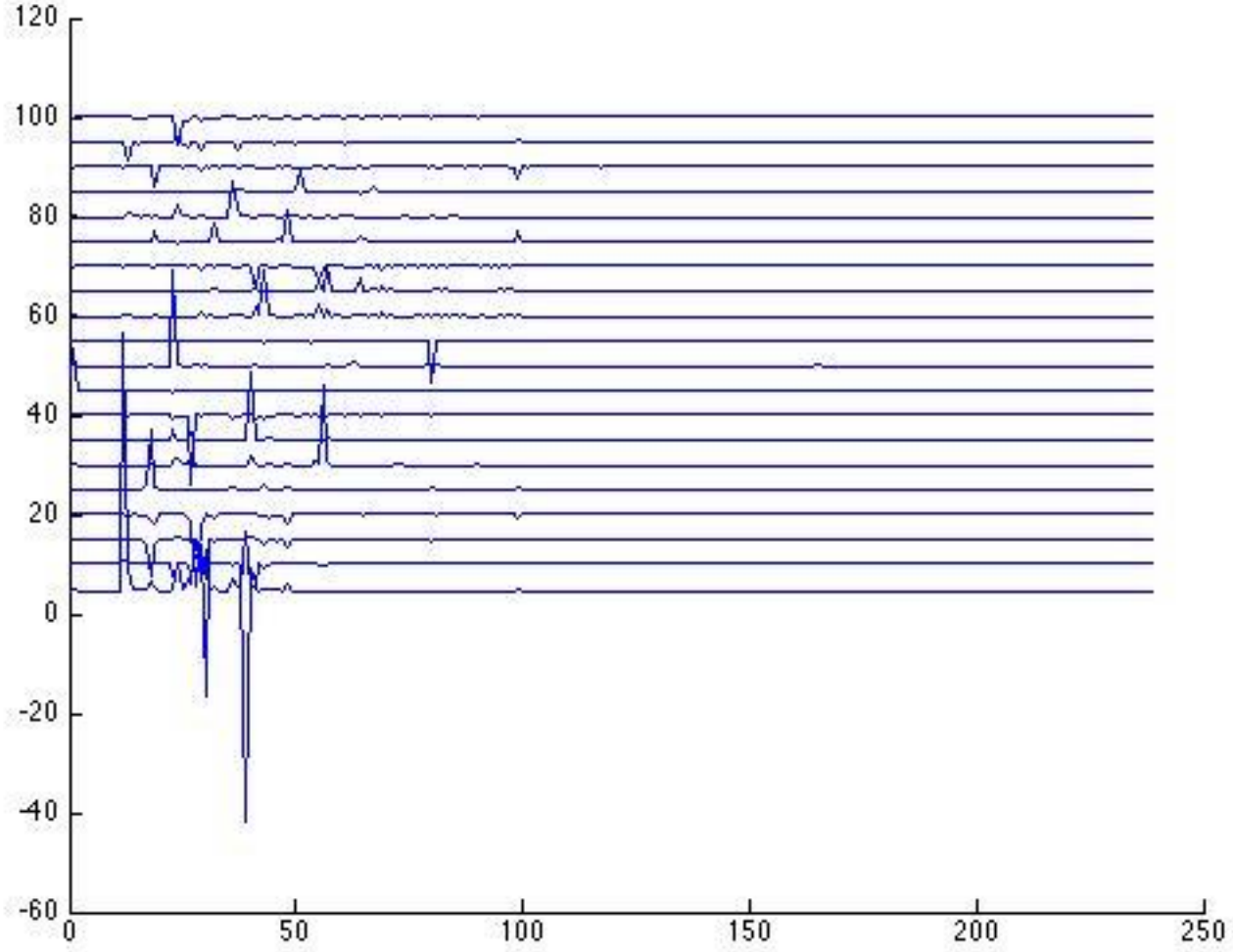
1. Assumes non-Gaussian distributions, maximizes the kurtosis for randomly-generated weighing factors.

2. Decorrelates and re-normalizes factors after every iteration:

$$w_{p+1} = w_{p+1} - \sum_{j=1}^{p} w_{p+1}^{T} w_j w_j$$

$$w_{p+1} = w_{p+1} / \sqrt{w_{p+1}^{T} w_{p+1}}$$

Independent component analysis



Non-negative matrix factorization

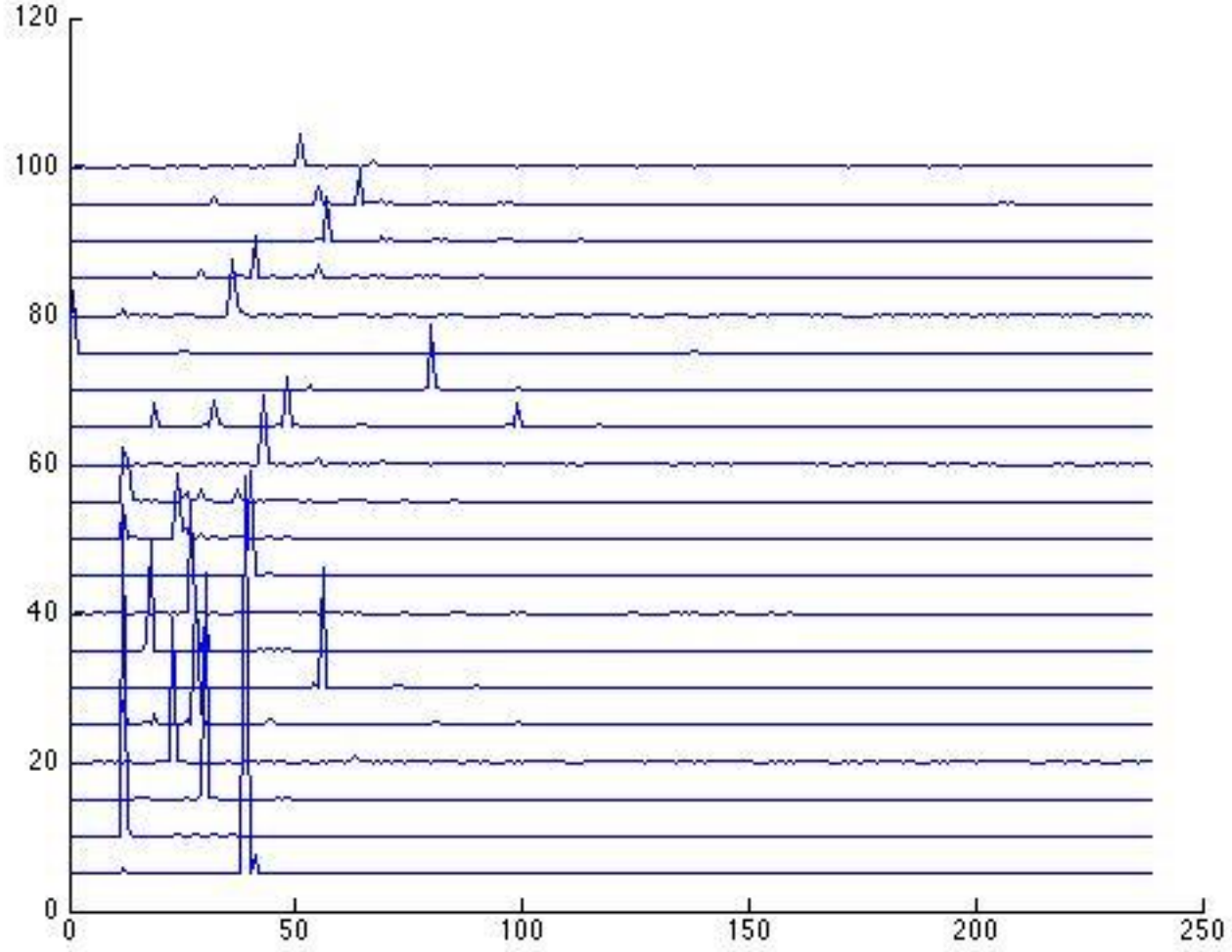
1. Factors an *n*-by-*m* matrix A into nonnegative factors W (*n*-by-k) and H (k-by-*m*).

2. Starting with random W and H, iteratively minimize the residual:

$$D = \sqrt{\frac{\|A - WH\|_F}{nm}}$$

3. Re-normalize W and H.

Non-negative matrix factorization



Future work

1. Try a different way of measuring non-gaussianity in the ICA algorithm (negentropy)

2. Combine ICA and NNMF?

3. Validate results with laboratory data and analyze multiple aircraft flights

4. Clustering algorithms?

12.S990 Quantifying Uncertainty
Fall 2012