Principal Component Analysis in EDS

D.L. West
ThermoFisher Scientific Inc., 81 Wyman Street, Waltham, MA 02451

Spectral imaging is a very powerful microanalysis tool for chemical phase identification in a variety of samples. However, a spectral imaging (SI) data set contains vary large amounts of information and traditional approaches to analyzing these data sets rely on element identification and X-ray mapping. These methods have the potential to misidentify or completely overlook minor phases. To efficiently analyze these large amounts of data, automated methods are needed. To be useful for routine analysis, these methods must make no assumptions about the sample chemistry, be fairly robust, work with low volume or sparse data sets and deal with noisy data. The microanalysis software program, COMPASS, automatically performs data reduction and analysis on SI data sets and conforms to these requirements [1, 2].

Compass is based on multivariate statistics and the application of multivariate statistics, multivariate analysis. Multivariate statistics is a form of statistics surrounding the simultaneous observation and analysis of more than one variable in a data set. Multivariate statistics is a subset of linear algebra which is the branch of mathematics concerning vector spaces and linear mappings between such spaces [3].

Within multivariate statistics there are a number of types of data analysis, one of which is principal component analysis (PCA). Principal component analysis creates a new set of orthogonal variables that contain the same information as the original data set but rotates the axes of variables to give a new set of orthogonal axes which are ordered such that they summarize decreasing proportions of the variation (in mathematics, orthogonality is the relation of two lines at right angles to one another, and the generalization of this relation into n dimensions; and to a variety of mathematical relations thought of as describing non-overlapping, uncorrelated, or independent objects of some kind).

Principal component analysis was developed by Karl Pearson in 1901[4] to show how variables act as the decisive factors in the way a data set collectively varies. PCA is a powerful statistical technique to determine the relationship between variables within a data set and reduce the data to a smaller set that describes those variables which contribute to the overall influence of these variables to the data set. That is, PCA uses sophisticated mathematical principles to transform a number of possibly correlated variables into a smaller number of variables called principal components.

References
1) This software is covered by U.S. patents 6,684,413 and 6,675,106.
2) P.G. Kotula, J.R. Michael and D.E. Newbury, (2007), Micros Microanal 13(Suppl 2), 1372-1373.
3) H. Anton (2005), *Elementary Linear Algebra (Applications Version)* (9th ed.), Wiley International
4) Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points is Space". *Philosophical Magazine Series 6* **2** (11): 559–572.