

# Applying Multivariate Statistical Tools to 2D-PAGE Data

Dave Bramwell, Anna Kapferer, Ian Reah, Dave Miller,  
Martin O’Gorman and Christian Andersson Ståhlberg  
Nonlinear Dynamics, UK.

In order to extract relevant information from proteomics experiments that reflect the complex interactions taking place within the cell, multivariate statistics has established itself as a powerful analysis tool. However, current workflows to statistically evaluate images from 2DGE are restricted due to the missing value problem, which limits the application of higher lever statistics. Missing values may arise from inadequately matched or undetected spots - a problem that is reinforced by the inability to sufficiently align spot patterns. Furthermore, the number of missing values increases in proportion to the number of images included in the experiment, thus encouraging the use of far fewer replicates than would be desirable for achieving statistical significance.

By employing an advanced image alignment step coupled to improved analysis workflows we are able to produce statistically robust 2DGE data with no missing values. Consequently, we are now in a position to extract the most meaningful information using a variety of multivariate data analysis tools, including Principal Component Analysis (PCA), Discriminant Analysis (DA), Multidimensional Scaling (MDS), and Correlation Analysis.