

# Multivariate evaluation, feature extraction, classification and clustering of images and spectral data sets by using R environment

Krzysztof Banas<sup>1</sup>, Agnieszka Banas<sup>1</sup> and Mark B. H. Breese<sup>1,2</sup>

<sup>1</sup>*Singapore Synchrotron Light Source, National University of Singapore, 5 Research Link, Singapore 117603*

<sup>2</sup>*Physics Department, National University of Singapore, 2 Science Drive 3, Singapore 117551*

Analysing the large number of variables in data sets (images and spectra) collected in contemporary laboratories is unquestionably a major challenge. In order to be able to extract essential information from a plethora of recorded bytes one must turn to statistical methods. Multivariate evaluation of the experimental results enable for significant dimension reduction (principal component analysis [1], partial least squares [2]), segmentation or clustering (hierarchical clustering analysis, k-means [3], density-based clustering, self-organizing maps [4]) and finally classification (linear discriminant analysis [5], decision trees and random forest [6], support vector machine [7]). These chemometric approaches adapt to spectral data and images can provide new insights on the spatial and spectroscopic information. In this contribution various examples of multivariate statistical analysis of spectral data and images by using R open source [7] environment will be presented.

- 
- [1] J. E. Jackson, A user guide to principal components. Wiley: New York (1991).  
[2] D. Haaland, E. Thomas, Analytical Chemistry **60**(11) 1193 (1988)  
[3] C. Fraley, A. Raftery, Journal of the American Statistical Association **97**(458), 611 (2002).  
[4] T. Kohonen, Biological Cybernetics **43**(1), 59 (1982).  
[5] A. Martinez, A. Kak IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(2) 228 (2001)  
[6] L. Breiman, Machine Learning **45**(1) 5 (2001)  
[7] C. Cortes, V. Vapnik, Machine Learning, **20**(3) 273 (1995).  
[8] R Core Team, R: A Language and Environment for Statistical Computing, <http://www.R-project.org/> (2013)