doi:10.1017/S1431927613006156
Microsc. Microanal. 19 (Suppl 2), 2013

R for X-ray Microanalysis

Jeffrey M. Davis

National Institute of Standards and Technology, Gaithersburg, MD USA

R is a high level statistical software package and programming language developed by the R Core Development Team [1]. The software is free to download, open source and requires minimal programming expertise. It also contains powerful plotting and data manipulation tools. This talk will focus on the use of R for processing and analyzing X-ray spectral data, and it will detail the development of a contributed extension package called microR, which improves the functionality of R for microanalytical tasks.

X-ray data, whether in the form of simple spectra, full hyperspectral data cubes or compositional data, present a number of challenges for analysis and data manipulation. The data are, for instance, heteroscedactic2, unit-sum constrained3, and frequently show non-normally distributed error terms (usually due to sample preparation or instrumentation error). The challenges of analyzing such data generally require techniques beyond what is offered in standard spreadsheet programs. The primary advantage of the R language is that it offers a diverse, well documented, open source library of functions for a variety of tasks such as time series analysis, generalized linear modeling, machine learning and chemometrics.

The microR package aims to provide a bridge between the specific data format needs of the microanalysis community and the data processing power provided by R. The package makes extensive use of other contributed packages, but re-writes the methods and applications to fit the specific needs of the microanalysis community. For instance, one of the major functions in the microR package is the analysis and processing of quality control data. A quality control measurement is simply a measurement made periodically on a well-characterized standard. The data are used the check the "health" of the instrument by applying a number of summarization and graphing tools. A suite of functions have has been implemented that perform time series analysis, archiving and anomaly detection for quality control data. The goal is to minimize the time required to sort and process the data using a spreadsheet program and improve the archival nature of the data. Figure 1 shows the results of nearly 3 years' worth of quality control data for a spectrometer measuring Mg (TAP crystal on a flow proportional counter). The figure demonstrates the use of the R "beanplot" function4, which provides a superior alternative to the common boxplot. Figure 2 demonstrates the use of the "robCompositions" package5 to perform a simplex based principal component analysis (PCA) of some compositional data. The centered log ratio transformation, performed prior to the PCA, results in a more informative biplot of the data than simply performing PCA on the data. Many other packages such as these have been imported into the microR package and adjusted so that they can be easily used by microanalysts to perform complex analyses.

References:
1. R Core Team. "R: A Language and Environment for Statistical Computing." http://www.R-project.org
2. Breusch, T.S. and Pagan, A.R. "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica* **47**, 5 (1979).
3. Aitchison, J. The Statistical Analysis of Compositional Data. (1986)
4. Kampstra, P. "Beanplot: A Boxplot Alternative for Visual Comparison of Distributions." *Journal of Statistical Software, Code Snippets* **28** (1), 1-9 (2008)
5. Filzmoser, P., Hron, K., Reimann, C. "Principal Component Analysis for Compositional Data with Outliers." *Environmetrics*, **20** (6), 621-632

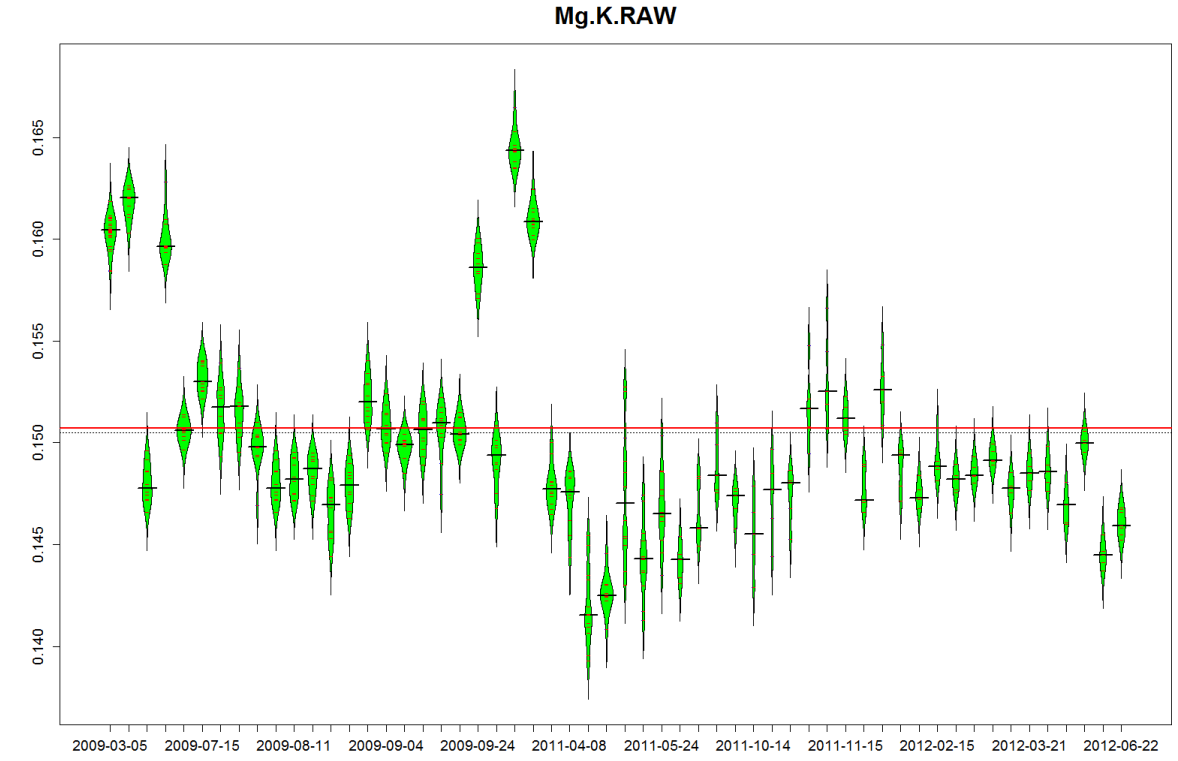

Figure 1: Beanplot of measured Mg k ratio data from K412 glass over time.

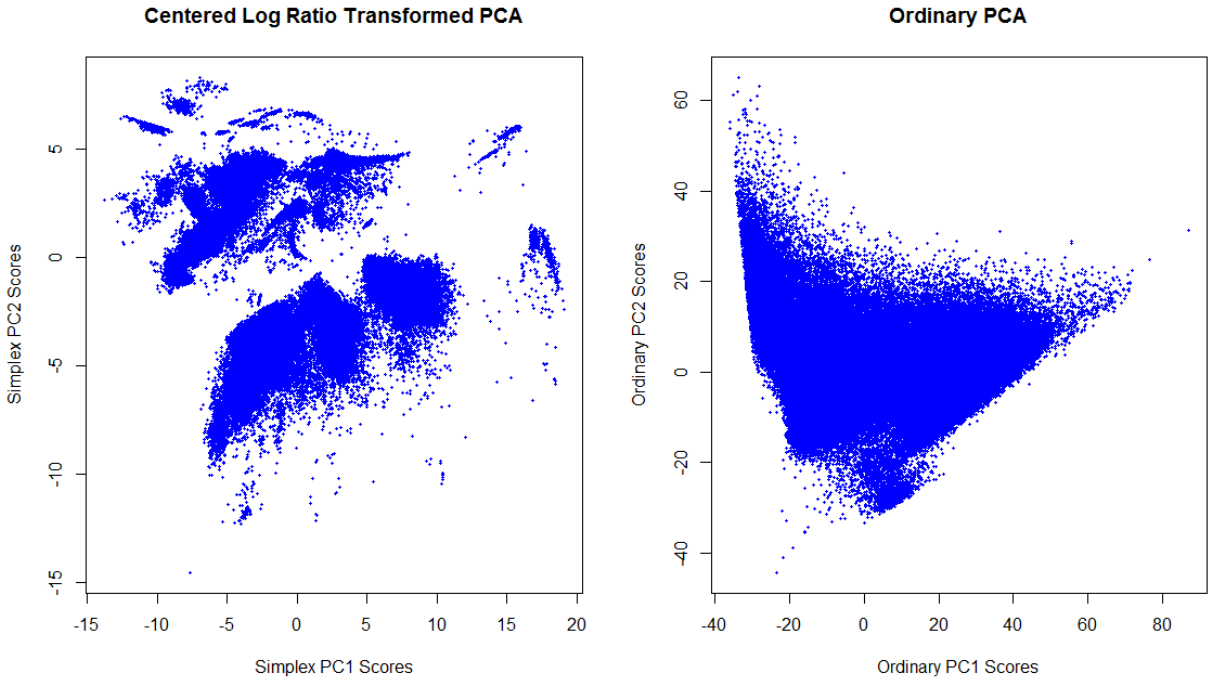

Figure 2: Comparison of centered log ratio transformed principal component analysis and ordinary principal component analysis for the same compositional data set.

https://doi.org/10.1017/S1431927613006156 Published online by Cambridge University Press