

***Comparison of Work-flows for Spectral Data Pre-processing
and Multivariate Statistical Analysis by Means of Open Source
Solutions: Orange (Python) and R Studio (R language)***

Krzysztof Banas¹, Agnieszka Banas¹, and Mark B H Breese^{1,2}

¹Singapore Synchrotron Light Source, National University of Singapore, 5Research Link,
Singapore 117603

²Physics Department, National University of Singapore, 2 Science Drive 3,
Singapore 117542

Proper sample preparation, selection of correct sampling technique, tuning the parameters of the experiment and collecting the spectral data is very important phase in the analysis of biological samples. But it's only the first step. Next, and crucial for achieving the goal of any project, is spectral data evaluation stage. Comparison of the work-flows for pre-processing, univariate and multivariate statistical analysis of the spectral data based on open source solutions: R Studio [1] (using R language) and Orange platform [2] (using Python language) is presented in this contribution. There are certain advantages of each work-flow discussed here in terms of user-friendliness, speed, scalability and flexibility as well as the option to implement machine learning and BigData algorithms.

References

- [1] R: A Language and Environment for Statistical Computing, R DevelopmentCore Team, Vienna, Austria (2017) <http://www.R-project.org>
- [2] J. Demsar, T. Curk, A. Erjavec, C. Gorup, T. Hocevar, M. Milutinovic, M. Mozina, M. Polajnar, M. Toplak, A. Staric, M. Stajdohar, L. Umek, L. Zagar, J. Zbontar, M. Zitnik, B. Zupan, "Orange: Data Mining Toolbox in Python", Journal of Machine Learning Research 14(Aug), 2349-2353 (2013).