Title: Multifactorial assessment in environmental metabolomics using multivariate ANOVA-based methods

Authors:

Laia Lopez (llfqam@idaea.csic.es), Carmen Bedia (cbgqam@idaea.csic.es), Joaquim Jaumot (jjsqam@idaea.csic.es)

Spanish National Research Council (CSIC), Environmental Chemistry, Jordi Girona 18-26, Barcelona, Spain

Presenter author: Joaquim Jaumot

Summary:

Untargeted metabolomics focuses on the sample exploratory analysis to detect differences between control and stressed samples and the detection of potential biomarkers causing the variation among them. The main advantage of these untargeted studies is that they allow generating new knowledge since are hypothesis-free approaches. However, this potential benefit presents the drawback of a more complex data analysis. Recently, some studies have introduced experimental design strategies to facilitate these data processing steps and to address underlying biological questions with more reliability. In the chemometrics field, these complex datasets have fostered the proposal of new methods able to give information regarding the significance of the biological factors tested, the differentiation among sample groups and the identification of potential biomarkers.

In this work, an assessment of the performance of these methods is performed by comparing the results obtained in the analysis of designed untargeted metabolomics studies using a liquid chromatography high-resolution mass spectrometry (LC-MS) platform. In these experiments, the effect of glyphosate exposure in the metabolome of carrots (Daucus carota) was evaluated considering experimental factors such as the dose applied, the exposure time or the analyzed tissue. Raw LC-MS data was processed using the ROI approach to obtain a feature matrix containing the compressed information of the different samples in the rows and the relevant m/z values in the columns.

These feature matrices were evaluated using different ANOVA-based multivariate modelling methods such as ANOVA Simultaneous Component Analysis (ASCA), ANOVA-PCA or regularized multivariate ANOVA (rMANOVA). First, the effect of different preprocessing strategies was evaluated, and models using a variety of preprocessing approaches were determined. All these methods showed similar results regarding the statistical significance of the assessed factors and gave the same main biological interpretation from the sample distribution in scores plots. However, significant differences between the methods were detected when the ability for identifying potential biomarkers associated with the exposure factors was evaluated.

From these results, the best combination of preprocessing approach and an ANOVA-based multivariate modelling method for the analysis of complex untargeted metabolomics datasets will be proposed.