

REFINEMENT OF THE SIGNIFICANCE LEVEL USED IN DETERMINING A FISHER RATIO THRESHOLD FOR PEAK TABLE GENERATION OF COMPLEX GC×GC-TOFMS DATA

Lena Dubois, Pierre-Hugues Stefanuto, Jean-François Focant

CART, Organic and Biological Analytical Chemistry Group, Department of Chemistry - University of Liège, Allée du Six Août B6C, Quartier Agora 11, B-4000 Liège, Belgium

The non-targeted analysis of volatile organic compounds emanating from complex biological samples using GC×GC (comprehensive two-dimensional gas chromatography) coupled to TOFMS (time-of-flight mass spectrometry) produces vast amounts of data. As a result, the interpretation of chromatograms requires sophisticated data treatment. The comparison of several chromatograms belonging to different classes is an especially challenging task, as it can be difficult to determine compounds of interest amongst the numerous candidate compounds. For this reason, a statistical approach was developed based on a calculation of the Fisher Ratio (FR) for each candidate compound to reduce the peak table complexity and to identify compounds showing a high variance between classes. The FR for each compound is compared against a threshold known as the critical value (F_{crit}) which is determined using a single-factor analysis of variance (ANOVA) incorporating the number of samples in each class, the number of classes, and the significance level (α). Following alignment of processed chromatograms, compounds exceeding F_{crit} are generally retained as compounds of interest and applied to multivariate statistical analysis like principle component analysis (PCA). In this study, various significance levels were investigated to determine the effect on the resulting multivariate analyses. The standard significance level ($\alpha = 0.05$) and three additional α values ($\alpha = 0.1, 0.01$ and 0.001) were tested. Lower significance levels resulted in fewer candidate compounds being retained as compounds of interest. The analysis confirmed that the suggested FR filtering approach provides a valuable tool for multivariate visualization by reducing the amount of data required for class discrimination. The three higher significance levels ($\alpha = 0.1, 0.05$ and 0.01) resulted in a similar PCA score dispersion over the plot, yet resulted in higher clarity of the correlation loadings plot with more stringent significance levels. Applying $\alpha = 0.001$, the number of retained compounds were not sufficient to obtain class discrimination using PCA. However, facilitation of data handling can be performed by adjusting the significance level to control the data reduction depending on the amount of detected compounds.