Assessing the repeatability and statistical advantages of homonuclear 2D-NMR spectra : a clustering approach

Baptiste Feraud (UCL, ISBA - MLG), Pascal de Tullio (CIRM, Chimie Pharmaceutique, ULg),
Bernadette Govaerts (UCL, ISBA), Michel Verleysen (UCL, MLG)

Metabolomics (and metabonomics) studies are increasingly being used in a variety of health, pharmaceutical and food applications including, among others, pharmacology, pre-clinical drug trials, toxicology, newborn screening, clinical chemistry, food and medicinal plants quality control. NMR spectroscopy is widely used together with multivariate analysis approaches in order to characterize perturbations in metabolic pathways occurring during biological processes. These predictive tools hold great promises for applications in a variety of scientific fields.

A large amount of recent scientific and statistical works are available concerning 1D spectra (principally ^{1}H-NMR spectra). More recently, two-dimensional NMR spectroscopy tools have been investigated: homonuclear ones (COSY, TOCSY, NOESY, etc…) and heteronuclear ones (HSQC, HMBC, etc…). It is commonly accepted by users (biologists, pharmacologists, healthcare professionals…) that the recent introduction of 2D-NMR methods represents a huge qualitative gap for metabolomics investigations in terms of metabolites and biomarkers identifications. In other words, it seems obvious and natural that more information (i.e. additional dimension) means more predictive power. But, until now, no statistical study clearly proved this assumption. Therefore, a fundamental question is "*Is this supplementary information equivalent to relevant and crucial information ?*".

In order to extend the statistical properties and tools developed for 1D spectroscopy to the new challenges raised by 2D spectra, a rigorous study of the repeatability and the robustness of 2D-NMR spectra is needed as a prerequisite. In the context of first homonuclear COSY experiments (spectra of complex biological media), we will present a methodology based on accurate multivariate clustering tools. Numerical quality indexes and graphical clustering results will be shown, obtained via binary vectors of positions, via recoded intensity vectors and through different levels of spectral resolution.

A second objective is to compare these 2D results with corresponding 1D results (^{1}H-NMR) obtained in the same conditions. Our preliminary results seem already promising: COSY appears to be a statistically robust tool and, in addition, appears to be more repeatable and reliable than one-dimensional classic method. Thus, additional information appears to be relevant.