NMR spectroscopy is an important tool in many metabolomics applications. Its potential capability to handle complex mixtures of metabolites makes it a prime choice in both identification and quantification of the multitude of different species constituting unprocessed biological mixtures. In quantitative metabolomics (targeted metabolic profiling) all detectable compounds in a biosample are identified and quantified by comparing the biosample spectrum to a library of reference spectra of pure compounds [1]. The underlying assumption is that a spectrum is the weighted sum of the spectra of all the individual metabolites which constitute the mixture (spectral additivity). In practice, the inverse problem of decomposing an experimental spectrum into its component parts corresponding to individual metabolites is arduous due to their large number and the complexity of their spectra, massive overlap of spectral peaks, non-trivial deviations of peak shapes from the ideal Lorentzian profile [2], lack of a suitable orthonomal basis in the vector space spanned by sets of Lorentzian peaks, and presence of artifacts such as receiver noise, irregular baseline drifts, and magnetic field inhomogeneity effects. Due to all these factors, attempts to overcome peaks overlap problems by means of conventional deconvolution (fitting) of selected spectral areas have in general limited success. For this reasons, and because the actual nature of all the metabolites is rarely known in advance, metabolomics often uses alternative statistical evaluation methods, such as multivariate factor analysis [3], which sidestep the need for a complete interpretation of the spectra and a full solution of the inverse problem, while still permitting the correlation of the spectra with specific biological aspects. However, such approaches require integration over predefined intervals (bins) and a meaningful integration of such intricate and artifact-burdened spectra may often be just as arduous as peaks fitting. Recently, a new algorithm called GSD (Global Spectrum Deconvolution) has been developed [4] and made available in the Metyal software package of Mestrelab. GSD is capable of identifying even poorly resolved spectral peaks and of fitting all recognizable peaks in even a very complex 1D spectrum in a surprisingly short time (typically a dozen seconds for up to 1000 peaks). Moreover, it is fully automatic and objective (no human intervention is required) and produces a table of all detectable spectral peaks and their parameters. Such a table can be then used for various purposes such as generation of artifact-free synthetic spectra (with or without resolution enhancement), stick spectra, artifact-free integrals, as well as accurate binning void of any bin-crossover problems due to the overlapping wings of spectral peaks.

Because of these attractive features, GSD is likely to become a very important pre-processing tool for all metabolomic approaches to the evaluation of NMR spectra of whole biosamples.

In addition, we have taken advantage of GSD to implement under Mestrelab the quantitative referencing strategy known as PULCON [5]. It is well known that when dealing with biological samples containing lipids and proteins, one can not use an internal reference compounds (such as TSP or DSS) interact with carbohydrates, lipids and proteins, one can not use an internal reference compounds. We then compare the metabolite concentrations obtained from experimental spectra by means of the GSD - PULCON algorithm with the PCA plots obtained by binning the synthetic spectra generated by the GSD algorithm.

References