

# The Bumpy Road towards Automatic Global Spectral Deconvolution (GSD)



MESTRELAB RESEARCH  
NMR Solutions

Carlos Cobas<sup>1</sup>, Stanislav Sykora<sup>2</sup>

<sup>1</sup> Mestrelab Research, Xosé Pasín 6, Santiago de Compostela, 15706 Spain; carlos@mestrec.com

<sup>2</sup> Extra Byte, Via Raffaello Sanzio 22/C, Castano Primo (Mi), Italy I-20022; sykora@ebyte.it

Everybody knows the meaning of the term “spectral multiplet deconvolution” and most are aware that the algorithms dedicated to this task are by no means simple, since they involve quite extensive – and often somewhat fuzzy – prior knowledge (input other than the experimental data themselves) such as the number of lines in the multiplet, their shapes, and artifacts such as baseline distortions. Presently, all such algorithms require a rather tight user control over the input parameters in order to provide meaningful results even when dealing with relatively modest multiplets (2-20) lines.

We have attempted a much more ambitious goal: a full, automatic decomposition of a complete spectrum, or of a part thereof, into a number of “peaks” and a peak-less residue. For the first time, this task can be satisfactorily handled by a computer algorithm which, however, turns out to be necessarily substantially more complex than what one might expect at first glance. This presentation illustrates the principle problems we have encountered and the strategies employed to overcome them in order to be able to cover spectra of virtually any complexity.

One problem consists in the necessity to define the total number of spectral peaks present in the spectrum prior to any fitting. These peaks should account for all spectral features recognizable by a trained human eye, such as local maxima, barely resolved peak splittings, and peak ‘shoulders’, but not ‘invent’ any peaks which are not really required. The reason is that sets of lineshape functions of any kind are never linearly independent. Consequently, a decomposition in which the number of peaks were itself fittable would never be unique. For example, a single Lorentzian can be fitted extremely well by three different Lorentzians, and any slightly distorted Lorentzian will be fitted by three Lorentzians much better than by a single one, even when the fit has no physical meaning.

To account for all qualitatively discernible spectral features prior to any fitting, we analyze in detail the experimental spectrum as well as its numeric derivatives (1st and 2nd). The latter can be computed automatically and reliably using Savitsky-Golay convolution filters with automatically set filter parameters (the settings are based a novel and robust mean-linewidth estimator). A reliable second derivative, for example, has typically a S/N ration which is just about 3 times worse than that of the original spectrum.

But even pre-determining all the discernible spectral peaks is still not enough to carry out a successful GSD. Another problem regards the spectral lineshapes which are generally poorly defined and quite far from true Lorentzians. A failure to account for lineshape distortions leads to residual deviations (especially in the vicinity of strong lines) which look as false spectral peaks with surprisingly large intensities.

The factors which affect the shapes of spectral peaks are: (i) field inhomogeneity (*shimming*), (ii) FID weighting prior to DFT (*weighed profiles such as Voight's*), (iii) the discrepancy between DFT and true lineshapes (*DFT distortion*) and, to a surprisingly large extent, (iv) the convolution of many unresolved quantum transitions under the same spectral peak (*transitions banding*). To account for all these factors, we use a *master lineshape function* which is sufficiently more general than a pure Lorentzian, but nevertheless involves only a limited number of additional lineshape parameters.

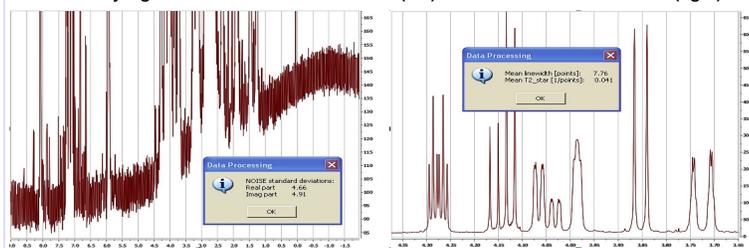
We have also found it extremely important to minimize (fit) not the square of the difference between the experimental spectrum and a theoretical one, but rather the variation of the said difference and of its first and second derivatives (*the smooth-residue principle*). Otherwise, one ends up with residuals full of large and unrealistic ‘bumps’.

Finally, the fitting itself cannot involve simultaneously all the parameters of all the peaks present in the spectrum – in many spectra, that would imply minimization of a function of up to several thousands of variables, a feat which is beyond the capability of any computer system. There are only two viable approaches: (i) *peak-by-peak removal* and (ii) a *sliding window* approach where only 1-3 peaks at a time are fitted. Even so, execution times may be a problem and special precautions need to be taken to keep them under control.

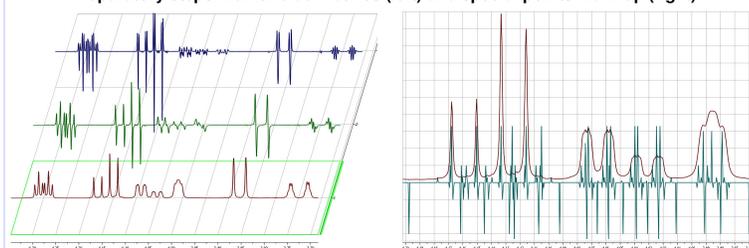
Once a spectrum has been decomposed into a set of generalized-shape spectral peaks and a reasonably smooth residue (baseline), any further data evaluation (integration, molecular structure verification/elucidation, factor analysis) is best carried out digitally on the peaks list rather than graphically on the spectrum itself. All meaningful information is in fact comprised in the peaks list, even though the exact shapes of the peaks may be difficult to interpret.

The GSD algorithm is so far limited to 1D spectra, but it can be certainly extended to any number of dimensions.

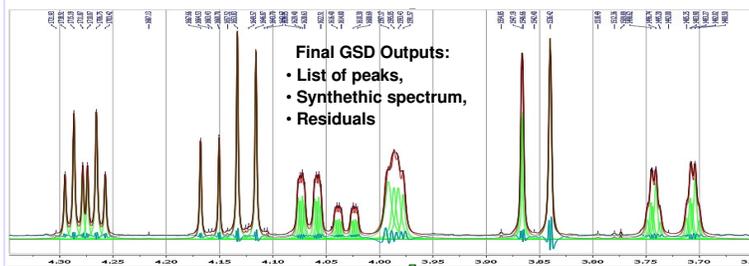
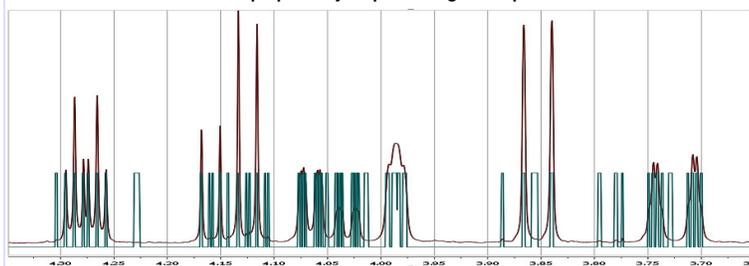
Auxiliary algorithms: robust noise estimate (left) and mean linewidth estimate (right)



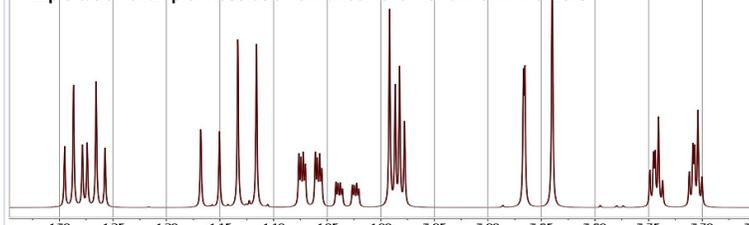
Preparatory steps: numeric derivatives (left) and special points mark-up (right)



More preparatory steps: boxing-in the peaks



Exploitation example: Resolution enhancement with uniform linewidths



Exploitation example: Stick spectrum and its integral

