Data Evaluation Algorithms: Bayesian DOSY and ROSY Transforms (BDT and BRT)

MESTRELAB RESEARCH
NMR Solutions

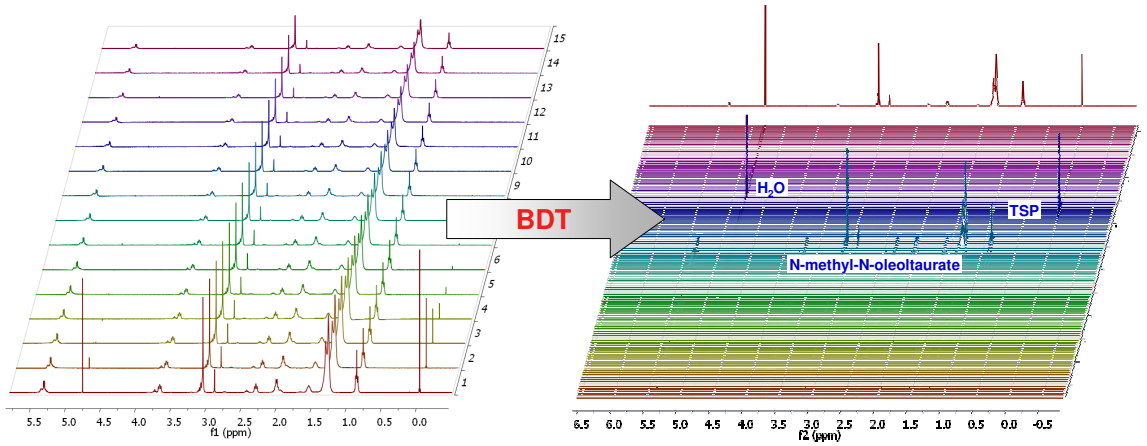
Carlos Cobas [1], María Sordo[1], Nikolay Larin[1], Stanislav Sýkora [2]

[1] Mestrelab Research, Xosé Pasín 6, Santiago de Compostela, 15706 Spain; carlos@mestrec.com
[2] Extra Byte, Via Raffaello Sanzio 22/C, Castano Primo (Mi), Italy I-20022; sykora@ebyte.it

BDT & BRT

A novel Bayesian-based approach to DOSY & ROSY Transformation

A well known problem with several types of **multi-array NMR techniques** regards the best visual presentation of the associations between physico-chemical parameters and individual spectral peaks. Typical examples are **DOSY**, in which the desired second parameter is the diffusion coefficient, and **ROSY**, for which it is the relaxation time. In such cases, the transformation of the original data set to a suitable final 2D graph (the DOSY or ROSY transform, respectively) is conceptually and mathematically difficult to manage. We propose a **Bayesian approach** which is computationally very efficient and physically eminently meaningful, and gives very satisfactory and artifact-free results. Applied specifically to the DOSY and ROSY data sets, it leads to what we call the **BDT** (Bayesian DOSY transform) and **BRT** (Bayesian **ROSY** transform) algorithms



BDT of an aqueous solution of potassium N-methyl-N-oleoltaurate (a surfactant) with TSP at 23 C
The original Varian FID file has been obtained from the *VARIAN NMR USER GROUP LIBRARY* (submitted by Brian Antalek as a sample for this DECRA algorithm)

The Bayesian aapproach in a nutshell

Let z be the gradient-dependent decay parameter associated with each DOSY spectrum and d the diffusion coefficient associated with a spectral line at some frequency f. Then the essence of a a Bayesian approach to the problem of transforming the experimental S[f,z] map to a probabilistic [f,d] map is embodied in the following question:

If one selects an arbitrary point in the [f,d] map, can one assign to it an amplitude and a probability which gives an idea of how compatible (congruent) that particular point in the [f,d] map is with the experimental [f,z] map?

Having fixed f and d, one has only one parameter to fit, namely the decay-curve amplitude $a_k(d)$, k being the data index corresponding to the frequency f. This is a linear, one parameter fit which can be done explicitly. Assuming the decay curve

$$y_{k\alpha} = a_k(d)\exp(-dz_\alpha),$$ where α is the index within the z-set,

the value of a_k which minimizes the total square deviation $\Sigma_\alpha (y_{k\alpha} - S_{k\alpha})^2$ is

$$a_k(d) = [\Sigma_\alpha S_{k\alpha}\exp(-dz_\alpha)] / [\Sigma_\alpha \exp(-2dz_\alpha)].$$

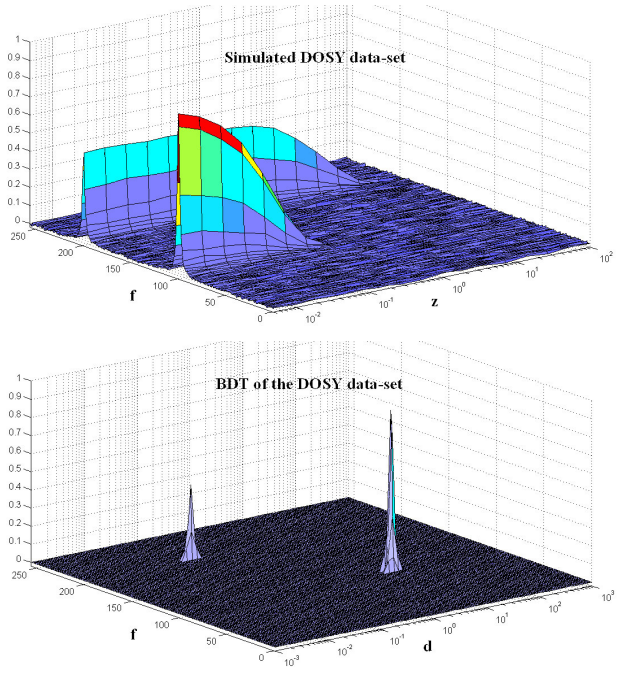
Even more important than the optimal amplitude is the Bayesian weight w_k associated with it. Denoting as σ the standard deviation of the experimental noise, the multiplicative contribution of the experimental data point $S_{k\alpha}$ to w_k is equal to $\exp(-(y_{k\alpha} - S_{k\alpha})^2/(2\sigma^2))$. Considering all z-points, we have

$$w_k = \Pi_\alpha \exp(-(y_{k\alpha} - S_{k\alpha})^2/\sigma^2) = \exp(-[\Sigma_\alpha (y_{k\alpha} - S_{k\alpha})^2]/\sigma^2),$$

Whose maximum $w_k(d)$ coincides with the $a_k(d)$ obtained by least-squares. Notice that, in principle, any d is legitimate, but the value of $w_k(d)$ is appreciable only when d is close to the 'correct' value. When the chosen d differs from the correct one, there are large systematic deviations (residues) and w(d) becomes very small. $w_k(d)$ is the proper 'vertical' value to be assigned to the (f,d) point in the DOSY 2D-plot, apart from the final normalization.

The final normalization, which is a characteristic step of any Bayesian evaluation, is very simple. Once the $y_k(d)$ values were calculated for all possible d-values, they are scaled so as to make their sum (vertical projection) equal to the intensity of the original spectrum at z = 0 (usually defined by S_{k0}).

Illustration of the algorithm using Matlab and simulated data



Conclusions

Compared to other approaches, the **Bayesian method** appears extremely promising. It automatically avoids having exact, unnatural zeros anywhere in the resulting map since *every point of the 2D map has has a well defined value of statistical congruence with the data*. Moreover, the **BDT maps** show 'normal' line widths in the f-direction, correctly positioned and resolved peaks in the d-direction and **quantitatively correct horizontal and vertical projections** – a combination difficult to achieved by any other means.

The approach can be easily extended to non-exponential cases arising from **overlapping lines** and, like all Bayesian methods, incorporate additional information available from other sources (*a-priori* knowledge). Likewise, it is possible to place a statistical premium on alignment of spectral peaks along horizontal lines in the [f,d] plot.

UNIÓN EUROPEA
Fondo Europeo de Desenvolvemento Rexional
XUNTA DE GALICIA
CONSELLERÍA DE INNOVACIÓN, INDUSTRIA E COMERCIO
Dirección Xeral de Investigación e Desenvolvemento

DOI: 10.3247/sl2nmr08.009

Developed in collaboration with www.ebyte.it