A new approach to interpret non-negative least sqares (NNLS) relaxation results

Guilhem Pagès, Amidou Traoré, Jean-Marie Bonny

AgroResonance, UR 370 QuaPA–INRA, F-63122 Saint-Genès-Champanelle, France

The modelisation of an NMR signal decay in a sum of exponential terms is an ill-posed problem. Experimental data are not sufficient to find both relaxation times and amplitudes. Several solutions lead to similar least-square distance between the model and the experimental data. To reduce the number of solutions, an efficient strategy consists in adding a constraint of positivity on all the parameters. Non-negative least-squares (NNLS) algorithm (1) is the most popular algorithm incorporating this constraint. The relaxation time values are *a priori* set in the decomposition basis (DB), the algorithm returning an unique solution of positive amplitudes.

To obtain a smooth amplitude distribution, a Tikhonov regularization is most often performed after the NNLS analysis. The choice of the regularization parameters is most often operator-dependent and is based on both prior-knowledge and T_2 distribution hypothesis (2).

Considering that only amplitude positivity is an indisputable *a priori*, we propose here to scrutinize in details the solutions provided by NNLS without further regularization. We show both by simulations and experiments that interpreting NNLS results from the cumulative distribution function (cdf) leads to robust analyses.

Models / Simulations

1. Discrete models are simply multiple Dirac functions, continuous ones are uniform Heavyside distributions with constant amplitude while the pseudo-continuous function contains several Gaussian distributions.

2. Rician distributed noise was added to the simulated decay curves sampled with TE = 1ms. 1000 noisy dataset were generated at a signal-to-noise ratio (SNR) of 1000 and NNLS analysis was always performed till SNR ~ 3 for preventing biases due to asymptotic convergence towards a non-zero baseline. NNLS was performed with the algorithm proposed by Lawson and Hanson (2) under Matlab, with log-distributed DB ranging from 1 to 1000 ms at different densities (i.e. the DB contains different number of T_2 values). The 1000 inversions obtained without further constraints were finally averaged.

- Pdf analysis are highly subjected to bias, especially for the continuous distribution
- Cdf gives the true amplitudes when it reaches the plateaus (i.e. pdf~0)
- Amplitudes are independent of the DB used

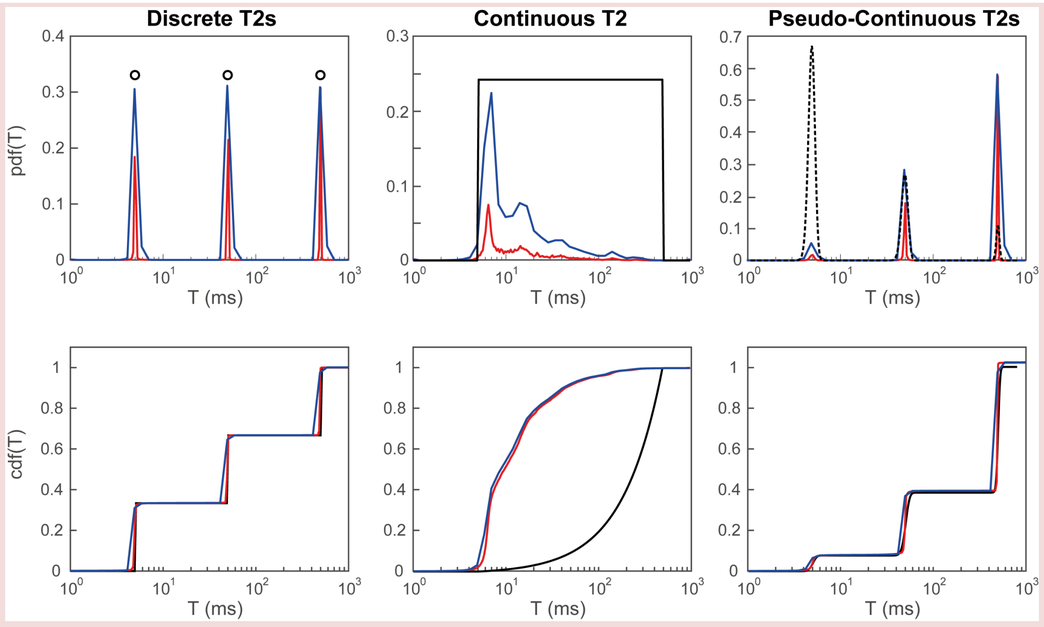


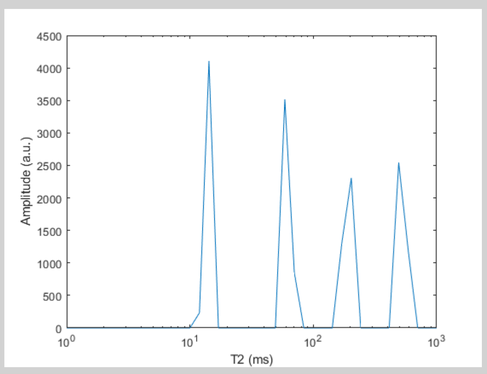
Figure 1: Averaged results for 1,000 simulated T_2 NMR signal decays analysed with NNLS algorithm and a DB containing either 40 (blue) or 200 (red) T_2 values. Black lines represent the theoretical values. Representation as a pdf (top) and cdf (bottom).

Experimental verification on a discrete sample model

1. 4 tubes containing water with different [$MnCl_2$] leading to T_2 values at 9.4 T of: 12, 57, 188 & 529 ms.

2. CPMG acquisition with 4196 echoes recorded with an echo time of 2 ms. Experiment was repeated 2000 times.

3. NNLS analysis performed with a DB containing either 40 or 200 T_2 values (logarithmically spaced from 1 to 1000 ms).



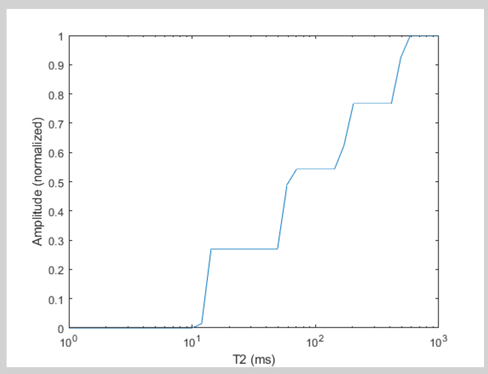


Figure 2: Illustration of both pdf (left) and cdf (right) results after NNLS analysis with a DB containing 40 T_2's of one of the experimental data.

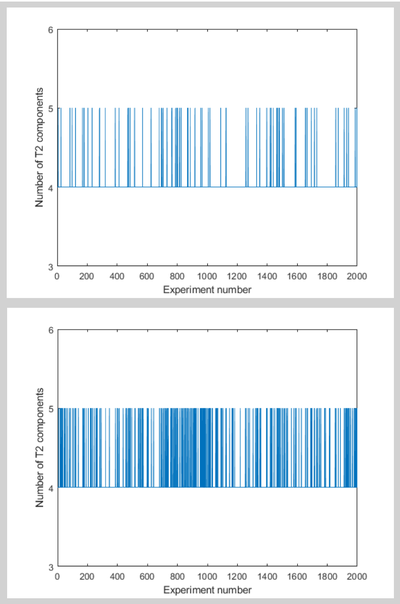


Figure 3: Number of plateaus found in the cdf after NNLS analysis for the 2000 repetitions with a DB of 40 (top) or 200 (bottom) T_2 values. Note that the initial amplitude of 0 is not scored as a plateau

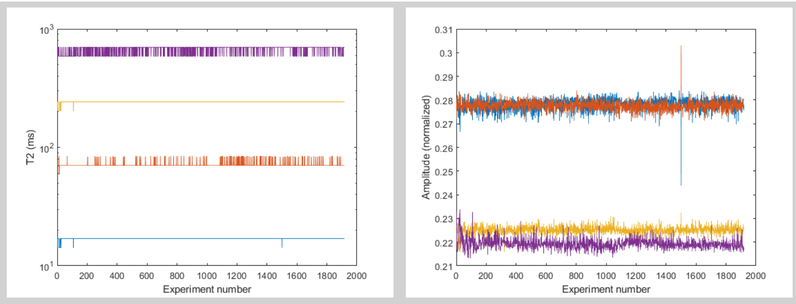


Figure 4: T_2 values (left) at the beginning of the plateaus and their amplitudes (right) for all experiments leading to 4 plateaus (1919 dataset, i.e. ~96% of the data) after NNLS analysis with a DB of 40.

- Highly consistent and reproducible results
- Data giving 5 plateaus split the last component but do not lead to significant changes for the amplitude values.
- Results independent of the DB.

Analysis based on cdf is robust

Application to food products

1. Large dataset (818 different samples) of T_2 measurements on a raw cereal product harvested at different locations and different level of maturation.

2. CPMG data acquired on a MQ20. The echo sampling rate is 7 ms and 200 echoes are recorded

3. NNLS performed with a 100 T_2 DB ranging from 10 to 1000 ms. Results analyzed based on cdf. A plateau is detected when 2 consecutive cdf values are similar.

- Reproducible number of T_2 components
- Variation in the T_2 amplitudes most likely due to biological differences

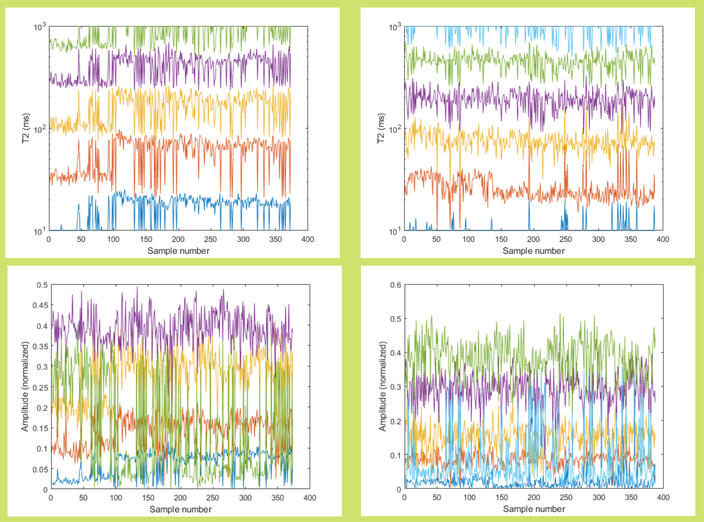


Figure 5: T_2 values at the beginning of the plateau for dataset giving 5 (left) and 6 (right) T_2 components. 373 and 387 dataset lead to 5 and 6 T_2 components, respectively. 93% of the data are represented on these 2 graphs.

Figure 6: Normalized amplitudes for each of the T_2 component in function of the number of plateaus found in the cdf. 5 were found for the left Figure while 6 were detected for the right one.

Conclusions

- To limit user-inputs into NNLS analysis, we push the idea that the cdf distributions are sufficient for obtaining useful information provided that the amplitudes are considered at the plateaus
- Result interpretation based on cdf led to non-biased results as shown by simulations
- Experimental results on a well-known sample showed that analysing results based on cdf was reproducible and rather insensitive to DB density
- Our interpretation approach highlighted similar behaviours without masking sample dependence information in food product

AgroResonance, UR370 QuaPA – INRA
F-63122 Saint-Genès-Champanelle, France
http://www6.inra.fr/agroresonance

(1) C.L. Lawson, R.J. Hanson, Solving least squares problems, Prentice-Hall, 1974, 23, 161.

(2) K.P. Whittall, A.L. MacKay, J. Magn. Reson., 1989, 84, 134.