Available online at www.sciencedirect.com
ScienceDirect
Journal of Magnetic Resonance 191 (2008) 226–230
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

A simulation based method to assess inversion algorithms for transverse relaxation data

Supriyo Ghosh [a,*], Kevin M. Keener [b], Yong Pan [c]

[a] Bruker Optics, Inc., The Minispec Division, 2700 North Crescent Ridge Drive, The Woodlands, Texas, TX 77381, United States
[b] Purdue University, West Lafayette, Indiana, IN 47907, United States
[c] Miami Valley Innovation Center, Procter & Gamble Co., Cincinnati, Ohio, OH 45253, United States

Received 7 May 2007; revised 17 December 2007
Available online 8 January 2008

Abstract

NMR relaxometry is a very useful tool for understanding various chemical and physical phenomena in complex multiphase systems. A Carr-Purcell-Meiboom-Gill (CPMG) [P.T. Callaghan, Principles of Nuclear Magnetic Resonance Microscopy, Clarendon Press, Oxford, 1991] experiment is an easy and quick way to obtain transverse relaxation constant (T_2) in low field. Most of the samples usually have a distribution of T_2 values. Extraction of this distribution of T_2s from the noisy decay data is essentially an ill-posed inverse problem. Various inversion approaches have been used to solve this problem, to date. A major issue in using an inversion algorithm is determining how accurate the computed distribution is. A systematic analysis of an inversion algorithm, UPEN [G.C. Borgia, R.J.S. Brown, P. Fantazzini, Uniform-penalty inversion of multiexponential decay data, Journal of Magnetic Resonance 132 (1998) 65–77; G.C. Borgia, R.J.S. Brown, P. Fantazzini, Uniform-penalty inversion of multiexponential decay data II. Data spacing, T_2 data, systematic data errors, and diagnostics, Journal of Magnetic Resonance 147 (2000) 273–285] was performed by means of simulated CPMG data generation. Through our simulation technique and statistical analyses, the effects of various experimental parameters on the computed distribution were evaluated. We converged to the true distribution by matching up the inversion results from a series of true decay data and a noisy simulated data. In addition to simulation studies, the same approach was also applied on real experimental data to support the simulation results.

Keywords: Relaxometry; Inverse problem; Simulation

1. Introduction

NMR relaxometry is a very useful tool for understanding various chemical and physical phenomena in complex multiphase systems. One of the key parameters that are highly exploited to indirectly measure several properties of the sample is the transverse relaxation time (T_2). This time constant is measured most commonly by a simple CPMG experiment [1]. In a CPMG experiment, echo heights are usually recorded at discrete linearly spaced time points and the decay of the echo heights are characterized by the time constant T_2, where time dependent echo signal can be represented as,

$$S(t) = S_0 \exp(-t/T_2) \tag{1}$$

Here, S_0 is the echo envelop height at initial time point, $S(t)$ is echo height as a function of time, and t is the time vector. When the sample under investigation has only one type of species in identical environments the echo heights would obey a mono-exponential decay pattern. But in most systems, there are several different species in multiple different environments. All these 'species–environment' combinations, in principle, will have characteristic T_2 values for their decaying echo signals. It is intuitive that, in natural systems, the transition from one T_2 value to the next would rather be smooth. This results in a T_2 distribution with multi-components (bi-modal, tri-modal, etc.) rather than mono-, bi-exponential, etc. Thus, the discrete-time decay data obtained through the CPMG experiment on a complex sample will have information regarding all the

* Corresponding author.
E-mail address: supriyoghosh@yahoo.com (S. Ghosh).

1090-7807/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2007.12.021

components of the entire distribution hidden in it. This raw data will be corrupt by some noise, as well. Hence, to have a clear understanding of a sample through NMR relaxometry, it is crucial to obtain a reliable deconvolution of all the T_2 components from the discrete, noisy, decaying echo signal. This is essentially an ill-conditioned inverse problem, which can be represented in the form of well known Fredholm integral equation of first kind (IFK) [4],

$$S_d = S_0 \int_a^b \exp(-t/T_2) f(T_2)\, dT_2 + \varepsilon \tag{2}$$

where S_d is signal raw data, S_0 is maximum signal, $f(T_2)$ is the frequency of a species with a specific characteristic T_2, and ε is random noise.

Since the CPMG data collected in an experiment is discrete in time the IFK here is better represented in the discrete form of,

$$S_i = S_0 \times \left\{ \sum f_k(T_2) \exp(-t_i/T_{2k}) \right\} + \varepsilon \tag{3}$$

It is characteristic of this type of IFK equations that it is difficult to obtain useful solutions to them by straightforward methods like Gaussian elimination, Cholesky, or QR-Factorization. This inverse problem of T_2 deconvolution has been solved by linearization, followed by the use of different kinds of regularization parameters [5,6]. Several physically convincing constraints like non-negativity are also introduced to dampen the unnecessary oscillations in the solutions. The uniform-penalty inversion (UPEN) algorithm is one of the most successful regularization processes applied to NMR relaxation data [2,3]. This algorithm is fundamentally a least square minimization routine. To avoid excessive details in the computed distribution a penalty factor is also added to the squared error of fit, in this algorithm. A curvature penalty factor is adopted in UPEN within the data range. A coefficient (C_k) is also multiplied to the penalty factor. The value of this coefficient, C_k is iteratively adjusted in order to ensure a strict uniformity of the penalty factor. Thus, the main aim of this algorithm is to minimize,

$$\sum \left(S_0 \times \sum f_k(T_2) \exp(-t_i/T_{2k}) - S_i \right)^2 + \sum C_k (f_{k-1}(T_2) - 2f_k(T_2) + f_{k+1}(T_2))^2 \tag{4}$$

where the first term is the residual term and the second term is the penalty term. Similar to any other inverse problem cases, this deconvolution process has the drawback of non-uniqueness. Even if there exists an exact solution (i.e., a specific distribution of T_2) that has been used to generate the raw decay data, it may not be unique. The same raw data can satisfy several different model solutions very nicely (with very low magnitude of error). Rank-deficiency in the data is usually a major contributor to this problem [4]. Thus, the estimated model may get significantly smoothed or otherwise biased relative to the true model. Characterizing such bias through model resolution analysis is essential to properly applying models to real systems. The ill-conditioned nature of the inverse problem imparts instability into the computed solution. A small change in a single measurement can lead to an enormous change in the estimated model. Thus, as the noise level rises the computed model loses its precision in resolution. Because of these reasons, it is essential to perform systematic statistical analyses on the responses of a regularization algorithm like UPEN to a range of CPMG data similar to real experiments. If we synthesize data from a known model of T_2 distribution and then try to calculate back the true model using a regularization algorithm, it will provide an estimate of accuracy of the algorithm through comparison with the true model. This study defines limits on the UPEN algorithm based on a known T_2 distribution and a specific set of experimental conditions.

2. Methods

An algorithm is developed and the code is written in Matlab® (can be obtained by requesting the first author) to generate simulated CPMG T_2 exponential data. This algorithm takes the user inputs in terms of: number of components, relative contribution of each component, shape of each component, mean T_2 values of each component, variance of each component, and signal-to-noise (SNR) of the measured data. All raw discrete-time data are normalized. SNR effect is studied by creating simulated CPMG data with SNR level from 80 through 1000. The computed model resolution is studied by creating simulated CPMG data (with commonly encountered SNR level of 400) with component mean separations ranging from 450 through 900 ms. UPEN algorithm is run on these simulated data to obtain the computed distributions. The deviation of the computed solution from the true model is quantitatively analyzed through two-sample Kolmogorov–Smirnov test, which is an established statistical technique for comparison of two arbitrary distributions. The Kolmogorov–Smirnov test statistic is the maximum of the absolute differences between the two cumulative distribution functions of the distributions being compared [7]. If $F^*(x)$ is the true distribution and $F(x)$ is the computed distribution the hypotheses for our two-sided test would be:

$$H_0 : F(x) = F^*(x) \quad \text{for all } x \text{ from } -\infty \text{ to } +\infty \tag{5}$$

$$H_1 : F(x) = F^*(x) \quad \text{for at least one value of } x \tag{6}$$

The test statistic, T_1 is expressed as,

$$T_1 = \frac{\sup}{x} \mid F^*(x) - F(x) \mid \tag{7}$$

The decision rule is that we reject H_0 at the level of significance α if the test statistic, T_1 exceeds $1 - \alpha$ quantile $w_{1-\alpha}$ as given by the corresponding lookup table.

The means and standard deviation of the computed distributions were calculated using the log-normal probability distribution function:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (8)$$

where μ and σ are the mean and standard deviation of the variable. The computed distribution means and standard deviations were compared with gradual increase of SNR and the distance between the component means.

For experimental validation of the proposed new method, and to show the usefulness, we generated (some) a set of experimental (real) data, as well. Three solutions of $MnSO_4$ of varying strengths were made so that their T_2 values are 1 order of magnitude apart. Three different NMR tubes of diameter of 5 mm were filled with these $MnSO_4$ solutions, and placed together in a 15 mm diameter tube for generation of CPMG data using the minispec mq series bench-top NMR equipment tuned at 20 MHz (Bruker Optics Inc., TX, USA). We did the same experiment with the same sample with increasing number of scans, to observe the effect of SNR on computed distribution. We measured the SNR in the CPMG data with 128 scans. We generated simulated CPMG data from a model distribution with three components whose mean values correspond to that of the three different $MnSO_4$ solutions obtained from the data with very high SNR (2048 scans). We added noise to it of the same level as measured from the 128 scan experiment (SNR = 256). Similarly, we generated more simulated data with gradual increase in the component width. We did UPEN inversion on all these simulated data. We calculated the Kolmogorov–Smirnov statistic between the distribution computed by UPEN from the real data (128 scans), and the same from the simulated noisy data with varying component widths.

3. Results

Simulated data were created for one component T_2 distribution with mean value of 100 ms and standard deviation of 5 ms. SNR values of 40, 60, 80, 100, 200, 300, 400, 500, and 1000 were used to generate the simulated CPMG data. As the SNR reduced, the broadening of the computed peak is observed. Two extreme cases are shown by comparing with the true model in Fig. 1. The two-sample Kolmogorov–Smirnov tests are done between the true distribution and computed distributions obtained from simulated CPMG data with increasing SNR levels. The Kolmogorov–Smirnov statistic values are plotted against SNR levels in Fig. 2, to show how SNR contributes towards the deviation of computed model from the true model. For UPEN, we observed that if we achieve SNR level of about 400 that gave us best estimation of the T_2 distribution; we do not need to increase the SNR beyond that value. The relationship between SNR and the departure of the standard deviation (width of the peak) values of computed distribution from that of the true distribution was calculated and is shown in Fig. 3. It is observed to show a negative power law type relationship.

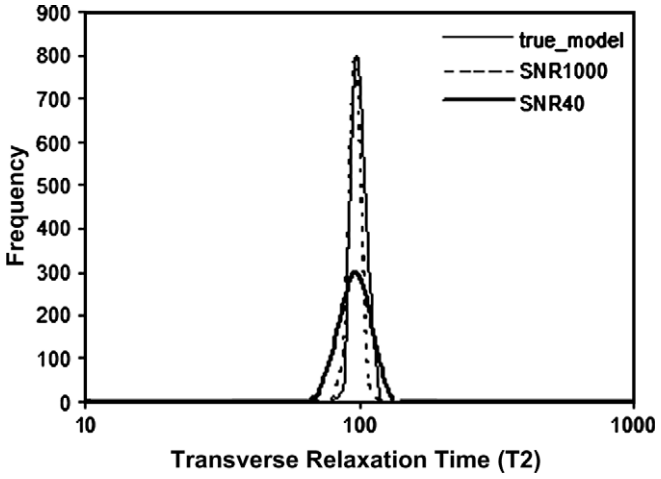

Fig. 1. Comparison of true T_2 distribution and UPEN computed T_2 distributions from data with varying noise levels.

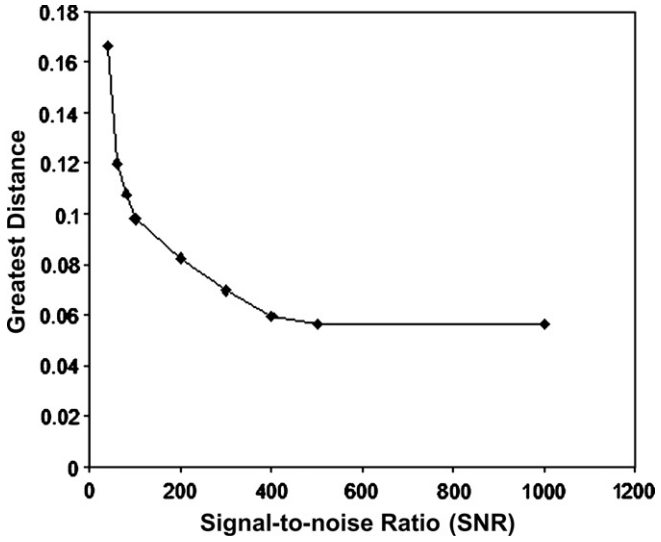

Fig. 2. Effect of signal-to-noise in raw data on deviation of computed mean T_2 value from the true mean T_2 value.

Simulated data were created for two component T_2 distributions. The distance between the component-means are set at 500, 550, 600, 700, and 900 ms. The standard deviations of each component were 5. The SNR level is kept constant at 400 for all simulations. This level of SNR is chosen because the Kolmogorov–Smirnov statistic value is observed to almost flatten out beyond 400 SNR. The evolution of the computed distribution as the two components approach each other is shown in Fig. 4. Kolmogorov–Smirnov test is done between the true and computed distributions.

The UPEN computed distributions from real experimental data of $MnSO_4$ solutions are shown in Fig. 5. The increased resolution of components as the number of scans, i.e., SNR increases is evident from the figure.

The calculated values of Kolmogorov–Smirnov statistic between the distribution computed by UPEN from the real

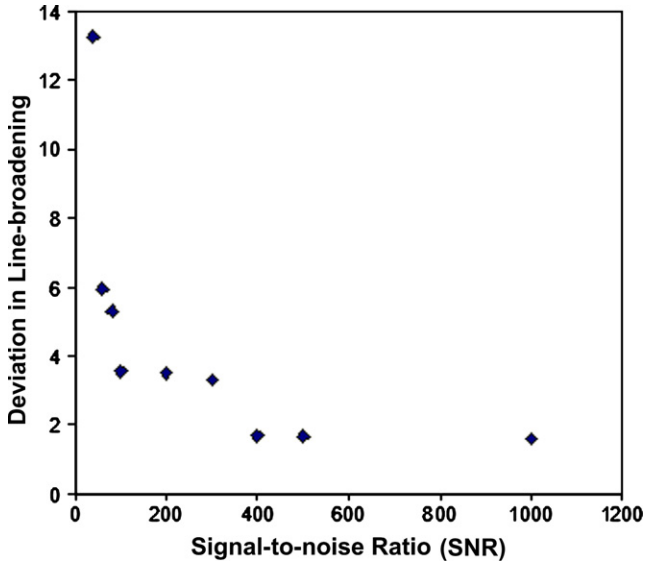

Fig. 3. Deviation in computed T_2 distribution width due to signal-to-noise ratio.

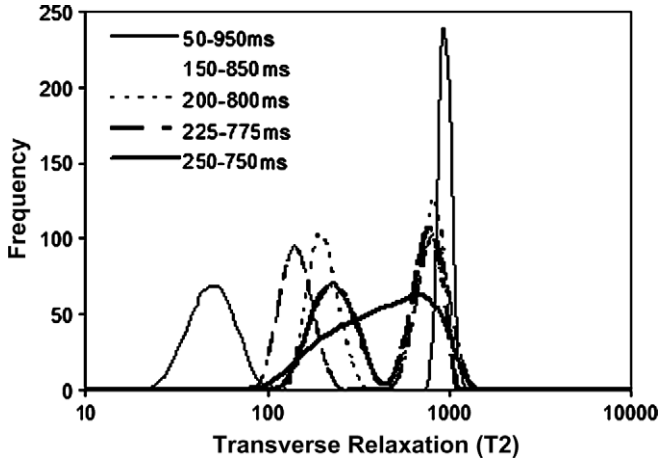

Fig. 4. Resolving efficiency of UPEN algorithm (while signal-to-noise is kept constant).

experimental data (128 scans), and the same from the simulated noisy data with varying component widths are plotted in Fig. 6.

The width of the model distribution was increased from 0 to 2 with linear intervals of 0.25. The width of 1.25 is observed to give the lowest value of Kolmogorov–Smirnov statistic. The value of Kolmogorov–Smirnov statistic in this case is 0.052 which is less than 0.06072 the table value of $w_{1-\alpha}$ quantile with the level of significance (α) as 0.01. as a result, we can accept the null hypothesis, i.e., the computed distribution can be considered equivalent to the true distribution. Hence, we can conclude that the true distribution of T_2 in the sample will have a width very close to this value.

4. Discussion

A major issue in using inversion algorithms for deconvolution of T_2 data in NMR relaxometry, is that when we get a computed distribution by applying the estimation algorithm we cannot determine how accurate the solution is or in other words how close the computed distribution is to the true distribution present in the sample under study. Variation in several parameters like SNR, number of T_2s we are solving for, data spacing, etc. have influence on the final solution. For many real systems we have no '*a priori*' knowledge of the T_2 distribution, as well. Thus we cannot determine solution accuracy right away. Many a time it is not practically possible to start developing a new inversion algorithm suited for the experimental data in hand; rather we need to make use of an existing algorithm. In this case, the knowledge of efficiency of the existing algorithm in inverting the noisy CPMG decay to the true T_2 distribution is critical. The knowledge of the data quality (e.g., SNR) required in order to get sufficient accuracy in inversion also becomes very useful. In these situations, an easy, quick and practically feasible technique to check the efficiency of the inversion algorithm, in hand, and to determine the necessary set of experimental parameters in order to obtain data those will give satisfactory inversion becomes a powerful tool.

Following our approach of using simulated CPMG data to evaluate the efficiency of the deconvolution achieved by an inversion algorithm, we were able to do detailed quantitative statistical comparison between the true distribution

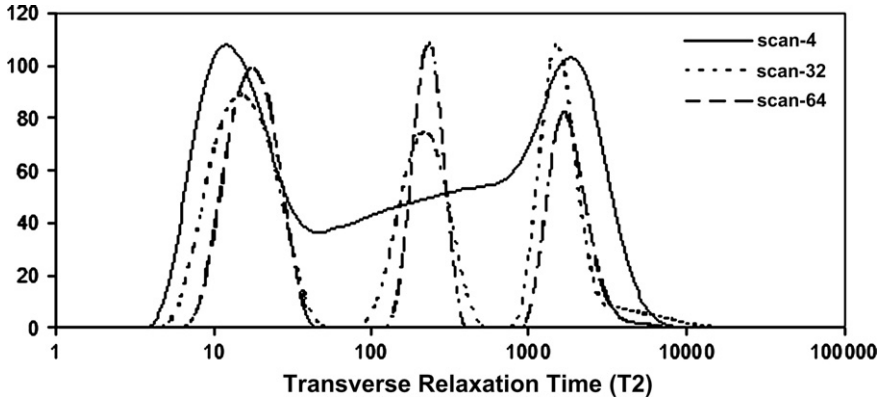

Fig. 5. Effect of signal-to-noise ratio on the peak resolving efficiency.

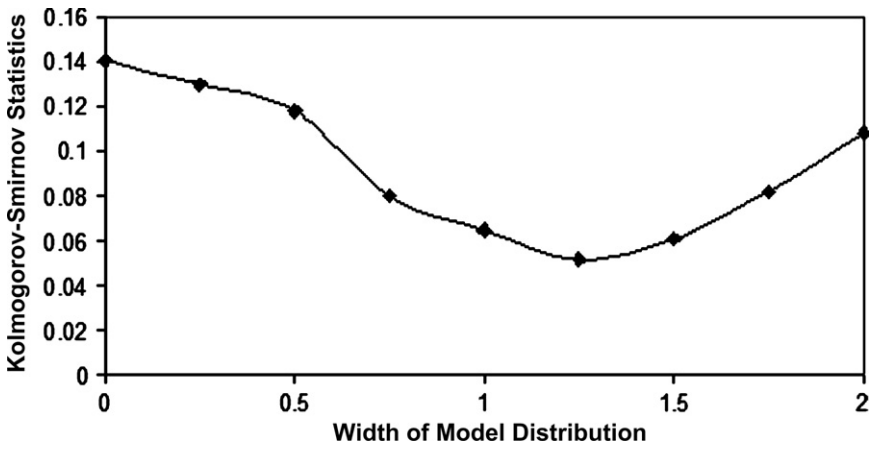

Fig. 6. Evolution of Kolmogorov–Smirnov statistics between true distribution and computed T_2 distribution as starting distribution width is gradually changed keeping the signal-to-noise ratio constant.

and the computed distribution, and quantify the limitations of the algorithm. The UPEN algorithm has been chosen here just to illustrate the technique. The same technique can be utilized if the inversion algorithm, in hand, to analyze the T_2 data is a different one. The discrete-time decay data is distorted by noise before the inversion algorithm is operated. Just like any other inverse problems, this is a mathematical challenge. Small variation in raw data introduce(d) appreciable oscillation in the computed model, resulting in broadening of the peaks. From Fig. 3 we can see a power law type trend in the departure from the true standard deviation as SNR changes. Using the Kolmogorov–Smirnov test, we could observe that as the SNR approaches 400 level and increases beyond that, the difference between the computed and true model flattens out and falls below the quantile table value for significance level of 0.10. From this observation we can conclude that achievement of an SNR level of 400 can give us enough accuracy in the computed distributions generated by UPEN algorithm.

The bi-modal simulation studies showed us that the UPEN algorithm can distinguish two different components (with 5 standard deviations) until their mean values are separated by at least ~550 ms. This conclusion is valid only for an SNR level of 400, which we kept fixed during this part of simulation study. 400 SNR level being commonplace in many NMR relaxometry studies it can be concluded that if two populations of T_2s are $\geqslant 550$ ms apart then the computed distributions obtained through UPEN algorithm are reliable.

By matching up the real data with varying width simulated data, we could also converge to a very good estimate of the true T_2 distribution in an unknown sample. From this analysis we conclude that, when we have data with a specific SNR level, the model distribution which had the lowest Kolmogorov–Smirnov statistic value would be the closest prediction for the true distribution of T_2 in the real sample. This way we were able to alleviate the uncertainty in the estimated distribution computed by an inversion algorithm, even in the absence of complete '*a priori*' knowledge of the sample.

This approach of systematic evaluation of regularization algorithms used for deconvolution of T_2 data, enabled us to quantify the accuracy of the computed distributions, and helped us reduce the uncertainty in the estimated distribution when very little is known about the sample. This study also provides guidance regarding the extent of SNR we need to achieve to get a reliable distribution. The approach described here can also be very useful in optimizing the regularization parameter for a specific kind of relaxation data.

Acknowledgments

We acknowledge the help of Dr. R.J.S. Brown and Dionne Swift for sharing the UPEN algorithm TrueBASIC code. We also acknowledge the help we received through the discussions regarding statistical analysis with Dr. Jason A. Osborne.

References

[1] P.T. Callaghan, Principles of Nuclear Magnetic Resonance Microscopy, Clarendon Press, Oxford, 1991.
[2] G.C. Borgia, R.J.S. Brown, P. Fantazzini, Uniform-penalty inversion of multiexponential decay data, Journal of Magnetic Resonance 132 (1998) 65–77.
[3] G.C. Borgia, R.J.S. Brown, P. Fantazzini, Uniform-penalty inversion of multiexponential decay data II. Data spacing, T_2 data, systematic data errors, and diagnostics, Journal of Magnetic Resonance 147 (2000) 273–285.
[4] R.C. Aster, B. Borchers, C.H. Thurber, Parameter Estimation and Inverse Problems, Elsevier Academic Press, New York, 2005.
[5] W. Provencher, A constrained regularization method for inverting data represented by linear algebraic or integral equations, Computer Physics Communications 27 (1982) 213–227.
[6] G.C. Borgia, V. Bortolotti, R.J.S. Brown, P. Castaldi, P. Fantazzini, U. Soverni, A comparison among different inversion methods for multi-exponential NMR relaxation data, Magnetic Resonance Imaging 12 (1994) 209–212.
[7] W.J. Conover, Practical Nonparametric Statistics, Wiley, New York, 1998.