Journal of Magnetic Resonance 196 (2009) 54–60

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Quantifying uncertainty in NMR T_2 spectra using Monte Carlo inversion

Michael Prange *, Yi-Qiao Song

Schlumberger-Doll Research, One Hampshire Street, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 31 July 2008
Revised 20 September 2008
Available online 12 October 2008

Keywords:
Monte Carlo
Laplace inversion
Uncertainty
Relaxation
Diffusion

ABSTRACT

Relaxation and diffusion data are often analyzed using a Laplace inversion algorithm that incorporates regularization. Regularization is used because Laplace inversion with finite and noisy data is an ill-conditioned problem for which many solutions exist for a given data set. This paper reports a different approach. Instead of finding a "best" solution by some *ad hoc* criterion, we developed an efficient Monte Carlo algorithm that generates thousands of probable solutions from which the statistical properties of the solution can be analyzed. We find that although all of the individual solutions are spiky, the mean solution spectrum is smooth and similar to the regularized solution. From the Monte Carlo solutions we obtain probability distributions for quantities derived from the spectrum, such as porosity and bound fluid. This ability to characterize the uncertainty of such quantities is novel.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

In the study of heterogeneous materials, both natural substances and manmade products, NMR spin relaxation spectra and diffusion constants are often used as a finger-print of the molecular species, structure and dynamics. For example, water and crude oils present in oil reservoirs can be distinguished by diffusion and relaxation experiments [1]. Typically, spin relaxation and diffusion are manifested as decaying signals. Data analysis often involves Laplace inversion to obtain a spectrum (or distribution) of relaxation times or diffusion constants. Such an inversion is ill-conditioned in the sense that for a given set of data with finite noise, many solutions will fit the data within the statistics of the noise. The well-established methods, for example Tikhonov regularization [2] and the maximum entropy method [3,4], find one solution that fits the data and satisfies some other simultaneous constraint. This type of approach effectively makes a choice of the class of solution based on independent criteria. In the case of the regularization solution, smoother spectra are preferred over more spiky spectra. Different algorithms essentially use different preferences and thus result in different "best" solutions. However, it is difficult to justify these choices. We will provide some specific examples to illustrate the multiplicity of the solutions.

NMR signals of T_2 decay in porous materials are well modeled as a sum of decaying exponentials:

$$m(t) = \int_0^\infty f(T)\exp(-t/T)\mathbf{d}\log(T), \quad f(T) \geqslant 0, \tag{1}$$

where $m(t)$ is the signal as a function of time and $f(T)$ is the spectrum as a function of relaxation time T. In the following we use the discrete form of this formula,

$$\mathbf{m} = \mathbf{G}\mathbf{f}, \quad \mathbf{f} \geqslant 0, \tag{2}$$

with $\mathbf{m} = \{m(t_i), i = 1, \ldots, M\}$, $\mathbf{f} = \{f(T_j), j = 1, \ldots, N\}$, and $\mathbf{G} = \{G_{ij} = \exp(-t_i/T_j)\Delta\log(T_j)\}$.

The principle commonly used in inversion is to use regression to find a solution $\mathbf{f}_0$ that fits the data "best" by some criteria. One approach is to find the minimum misfit solution by least-squares:

$$\mathbf{f}_0 = \arg\min_{\mathbf{f}\geqslant 0} \|\mathbf{m} - \mathbf{G}\mathbf{f}\|^2. \tag{3}$$

Contrary to popular opinion, the solution to this problem is unique. Non-uniqueness arises only when the min operator in Eq. 3 is replaced by a weaker condition that seeks all solutions satisfying the noise constraints. The solution to Eq. 3 can be efficiently found using the non-negative least-squares algorithm [5]. These solutions tend to have only a small number of non-zero elements [6]. The positions and amplitudes of these "spikes" are sensitive to the noise, creating the undesirable feature that the spectral solution that is not repeatable over different noise realizations.

The most commonly used approach regularizes the least-squares equations using

$$\mathbf{f}_0 = \arg\min_{\mathbf{f}\geqslant 0}(\|\mathbf{m} - \mathbf{G}\mathbf{f}\|^2 + \alpha\|\mathbf{f}\|^2), \tag{4}$$

where α is a scalar regularization parameter chosen to be just large enough to make the solution stable in the presence of noise. Choosing $\alpha = 0$ reverts to Eq. 3, while choosing α too large over-smoothes the solution, making the misfit unacceptably large with respect to the noise.

* Corresponding author.
E-mail address: prange@slb.com (M. Prange).

1090-7807/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2008.10.008

Examples are given in Fig. 1 to illustrate the behavior of these two approaches using a simulated spin echo decay. The spectrum and its associated simulated noisy data are shown in Fig. 1a. A least-squares T_2 spectrum resulting from these data is shown as the spiky black curve in Fig. 1b, and the corresponding regularized spectrum is overlain as the smooth black curve. New data were generated from the same spectrum but with a new noise realization generated with the same statistics; the least-squares and regularized spectra for these data are overlain as gray curves on this plot. A comparison of the least-squares solutions shows that both have three peaks whose amplitudes and positions vary with noise realization. The corresponding regularized spectra are less sensitive to noise realization. The smoothing due to regularization produces two peaks (one is an inflection), while the least-squares solutions have three peaks. The true spectrum is overlain as a dashed curve for comparison. It compares well in this case with the regularized solutions. However, in cases where the true spectrum is spiky, the least-squares solutions compares more favorably than the smoother regularized solutions.

A shortcoming of both least-squares and regularization approaches that is held in common with all "best fit" solution approaches is that they provide no indication of the uncertainty in the resulting T_2 spectrum, *i.e.*, they ignore the range of other spectral solutions that are also compatible with the measurements. The large range of such solutions is clearly indicated by the diversity of compatible spectra shown in Fig. 1b.

In this paper we use the Monte Carlo method to probabilistically sample the range of solutions that are consistent with the data. We then examine the statistical properties of these spectral solutions. After demonstrating that the T_2 spectral sampling problem is one of sampling from a truncated-multinormal distribution, where the truncation results from the non-negativity constraint, we show that the nature of the uncertainty distribution precludes the use of efficient truncated normal samplers already in the literature. The details of these truncated-multinormal samplers are discussed in Appendix A. We then present a new sampling algorithm that allows rapid sampling of T_2 spectral solutions. We use this algorithm to compute the uncertainty distribution of the inverted spectrum, and to compute the uncertainty distributions of other quantities derived from the spectrum. Previous work on computing the uncertainty of the T_2 spectrum [6] only provides estimates of the maximum and minimum values of quantities derived from the spectrum.

2. Theory

The inverse problem of determining $\mathbf{f}$ from a noisy measurement of $\mathbf{m}$ is one of finding the values of $\mathbf{f}$ for which the residuals $\epsilon = \mathbf{m} - \mathbf{Gf}$ are compatible with the measurement noise, *i.e.*, the covariance of ϵ should be consistent with the measured noise covariance. It is important to emphasize that many solutions exist which satisfy this criterion.

The noise in T_2 signals is often of the simple form of an uncorrelated normal distribution [6]. Thus the probability density function (*pdf*) of $\mathbf{f}$, $\pi(\mathbf{f})$, can be expressed succinctly as a truncated-multinormal distribution,

$$\pi(\mathbf{f}) \propto \exp\left[-\frac{1}{2}(\mathbf{m}-\mathbf{Gf})^T\mathbf{\Lambda}^{-1}(\mathbf{m}-\mathbf{Gf})\right], \quad \mathbf{f} \geqslant 0, \tag{5}$$

where Λ is a diagonal matrix containing the noise variance versus time. Truncation here refers to the $\mathbf{f} \geqslant 0$ condition. Since the noise variance is often well approximated as constant in time for a single measurement, in the following we simplify Eq. 5 with the approximation $\mathbf{\Lambda} = \sigma^2\mathbf{I}$, where σ^2 is the measured noise variance and $\mathbf{I}$ is the identity matrix.

In the parlance of Bayesian inference, Eq. 5 is called the likelihood function. It gives the probability of the measurement, $\mathbf{m}$, conditioned on a given model, $\mathbf{f}$. This is usually written as $\pi(\mathbf{m}|\mathbf{f})$. In Bayesian inference one would sample from the posterior, $\pi(\mathbf{f}|\mathbf{m})$, which is related to $\pi(\mathbf{m}|\mathbf{f})$ through Bayes' rule: $\pi(\mathbf{f}|\mathbf{m}) \propto \pi(\mathbf{m}|\mathbf{f})\pi(\mathbf{f})$, where $\pi(\mathbf{f})$ is the prior. Here we choose the prior to be constant, meaning that any value of $\mathbf{f}$ is as good as any other value in the absence of any data. If we had chosen the prior to be multinormal with a mean of zero and a diagonal covariance of σ^2/α, then the maximum value of the posterior would be that given by Eq. 4. This points out the commonality between the Bayesian approaches and the regularization approaches. We emphasize however that the goal of Bayesian methods is to quantify uncertainty while that of regularization methods is to find a single solution. An excellent tutorial on Bayesian methods is given in [7].

Since $\mathbf{m}$, $\mathbf{G}$ and σ are inputs to the inversion problem, Eq. 5 provides a complete description of the uncertainty in $\mathbf{f}$ and is thus the solution to the inverse problem. It allows the relative probabilities of any values of $\mathbf{f}$ to be computed and compared in describing the solution. If $\mathbf{f}$ were expressed in only two dimensions, a plot of these probabilities would provide a convenient display of the solution uncertainty. However, with a typical dimension of $\mathbf{f}$ being on the order of 100, other methods are needed to compute and describe

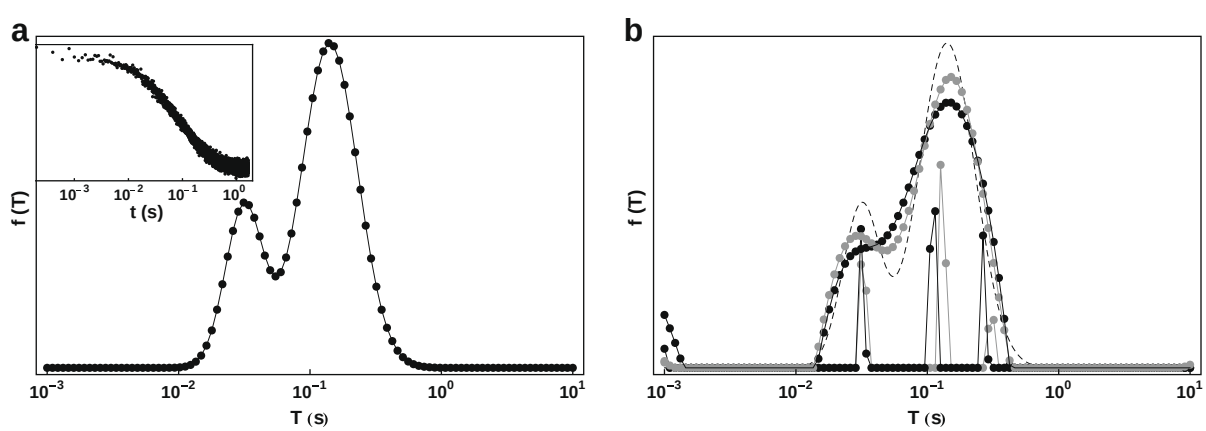

Fig. 1. Panel **a** shows the T_2 spectrum used to synthesize the T_2 echo data in the inset. The 100 T_2 spectral values are logarithmically spaced between 0.001 and 10 s. The 8192 echos are sampled at an echo spacing of 0.0002 s starting at 0.0002 s. The noise standard deviation is 0.025 of the maximum echo value. Panel **b** compares the spiky spectrum evaluated by the maximum likelihood method (results scaled to 10% to fit on plot) with the smooth regularized T_2 spectral solution. The black dots are solutions for the data shown in Panel **a**, and the gray dots are solutions for the same data but with a different noise realization. The original T_2 spectrum is overlain as a dashed curve for comparison.

the uncertainty. In the Monte Carlo method, random samples are drawn from the distribution. From these samples, statistics such as mean and variance are readily computed. This is the approach followed here.

The simplest method for finding random samples drawn from Eq. 5 is to use standard methods to first draw a large number of samples from the multinormal distribution (without the non-negativity constraint) [8], and then accept only samples which satisfy $\mathbf{f} \geqslant 0$. However, this method is astronomically inefficient for the T_2 spectral inversion problem since the fraction of multinormal samples that satisfy $\mathbf{f} \geqslant 0$ is miniscule for typical dimensions of $\mathbf{f}$ (*e.g.,* 100).

There are two Monte Carlo samplers in the literature for efficiently solving the more general problem of sampling from a truncated multinormal distribution in high dimensions [9,10]. These two methods (TN1 and TN2) are Gibbs samplers [11], *i.e.,* they sample an *N*-dimensional *pdf* as a sequence of one-dimensional sampling problems. This approach is particularly effective for a truncated-multinormal distribution because efficient algorithms are available for each one-dimensional sampling problem. Unfortunately, both methods perform poorly with the *pdf*s found in T_2 inversion. This issue is discussed in Appendix A.

3. Enhanced Gibbs sampler

In this section we overcome the limitations of the two Gibbs samplers mentioned above by making a modest extension of the Gibbs sampling approach. As discussed in more detail in Appendix A, TN1 is limited because it is unable to traverse in big steps along the major axes of the covariance ellipsoid, and TN2 is limited by the non-negativity constraint which causes it to become "stuck in the corner." We found that by generalizing the traditional Gibbs sampling approach to sequentially sample in two spectral dimensions at a time instead of just one, the resulting sampler overcomes all the limitations of TN1 and TN2. We have found it most effective to sample adjacent spectral dimensions, *i.e.,* we first sample values (f_1, f_2), then (f_3, f_4), and continue until the full spectrum has been sampled. Slice sampling is used for each of the two-dimensional sampling problems. An overview of two-dimensional slice sampling is presented in Appendix B along with suggestions for efficient implementation. Only one iteration of this Gibbs sampler is used for each sample update because we found that using more did not improve the solution. We call this algorithm TN3.

Algorithm TN3 was run on the T_2 data shown in Fig. 1. The run time to generate 10,000 samples on a 3.2 GHz Linux desktop machine was about 4 s. The speedup in TN3 over samplers TN1 and TN2 was examined by comparing the number of samples needed before statistical independence is reached (small sample correlation). A plot of sample correlation versus sample lag for spectral dimension 61, shown in Fig. 2, shows that the TN3 samples are only weakly correlated after about ten samples, whereas correlation plots for TN1 and TN2 for the same data (shown in Fig. A.3 in Appendix A) indicate substantial correlation remaining after thousands of samples. This demonstrates a speedup of a least three orders of magnitude.

Each of the 10,000 samples is a possible solution of the inverse problem, and the set of all samples represents the range of possible solutions. Four of these solutions are plotted in Fig. 3a–d. The spiky nature of these solutions is readily apparent. It is clear from the diversity of these solutions that interpretation algorithms that rely on well-determined positions and amplitudes of spectral peaks are unreliable unless they impose additional constraints on the solution such as smoothness.

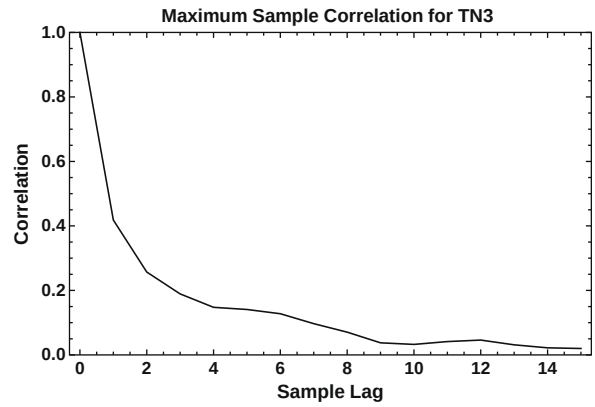

Fig. 2. Sampler TN3 was applied to the T_2 echo data shown in Fig. 1a, generating 10,000 samples. This is a plot of correlation between samples for spectral dimension 61 ($T = 0.266$ s), the dimension of maximum correlation.

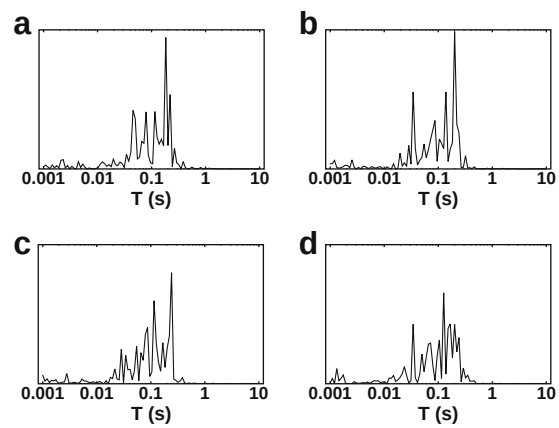

Fig. 3. Panels a–d present four of the 10,000 random samples from sampler TN3 for the T_2 echo data shown in Fig.1a. The spiky nature of these samples is characteristic of all of the samples.

One way to examine the diversity in these spectral solutions is by plotting the mean and standard deviation of each spectral value. These are plotted for the 10,000 samples in Fig. 4. The original spectrum is plotted for comparison as a dashed curve. The mean solution has two peaks, as does the original spectrum, but the peak

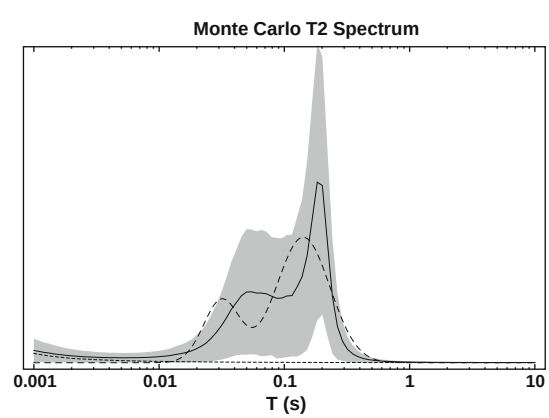

Fig. 4. The mean T_2 spectrum computed from 10,000 samples (using TN3) for the data in Fig. 1 is displayed as a solid black curve. The one-standard deviation region about the mean in indicated in gray. For comparison, the original T_2 spectrum is overlain as a dashed curve. The dotted curve shows the mean spectrum for data containing only noise (no signal) with the same noise statistics as the other curves.

positions are shifted upward from those of the original spectrum. Furthermore, the width of the second peak is substantially narrower than that in the original spectrum. However, the standard deviation is broad enough to encompass the original spectrum, while becoming very narrow outside the support of the original spectrum. This indicates that the original spectrum lies within our uncertainty-quantified solution to the inverse problem. It is important to understand that although the mean spectrum is *a* solution, it must not be taken as *the* solution.

A feature of the uncertainty-quantified solution is the increasingly upward bias in spectral values with decreasing T_2. This is a characteristic of the inverse problem, and is not an artifact of the TN3 sampler. It is produced by two factors. First, Eq. 1 indicates that only data with small t are used to determine $f(T)$ for small values of T. Having fewer data constraints at small T results in greater uncertainty. Second, the non-negativity constraint forces this uncertainty to be expressed as positive spectral values. In the limit, for values of T much smaller than the smallest t sample in the data, the spectral value is allowed to be any positive number. The dotted curve at the base of Fig. 4 highlights this bias by showing the mean spectrum corresponding to only the noise component of the spectrum (no signal), clearly showing the increasing bias at small values of T. This bias has long been recognized in the NMR community, where the smallest T_2 in the spectrum is typically taken to be several times larger than the smallest time present in the data. The best means for addressing this bias is to replace the constant prior in our formulation with one which pulls under-constrained spectral values toward zero. The standard regularization method uses this approach by using a zero-mean normal prior. We have chosen here not to use such a prior in order to demonstrate the full nature of the solution uncertainty.

Another feature of the solutions that is not indicated in Fig. 4 is the correlation relationships between spectral positions. This feature is difficult to capture in a static plot. An animated display over the 10,000 samples illustrates this correlation as adjacent spectral values bounce up and down in value. The presence of strong correlation is well known. For example, since the integral of the spectrum equals the data value at $t = 0$, if one spectral value increases, others must decrease in order to retain a constant integral value. One way to visualize the spectral uncertainty in way that includes these contributions is to plot the uncertainty of a functional of the spectrum, such as its total integral value. This is done in the next section.

4. Application to functionals of T_2 spectra

Using Monte Carlo sampling, we quantify the uncertainty in the T_2 spectral inverse problem in terms of a multiplicity of spectral solutions. However, uncertainty becomes most useful when it can be adequately summarized in terms of a few statistics. The plot of mean and standard deviation in Fig. 4 is one way to visualize the uncertainty in the entire spectrum, but more representative descriptions which also capture covariance are hampered by the large dimension of the spectral domain.

One way around this dimensionality problem is to use a functional to map each spectrum into a single number whose uncertainty can then be visualized as a histogram. Example functionals are total porosity and fractional porosity. Here we demonstrate this approach by estimating the uncertainty in some functionals of the T_2 spectrum using algorithm TN3.

The total porosity, ρ, is obtained as the integral over the T_2 spectrum. Computing this for each of the Monte Carlo samples for the data in Fig. 1a, the uncertainty in ρ is displayed as a histogram in Fig. 5, yielding the estimate $\rho \approx 0.1055 \pm 0.000996$. Estimates from the regularized T_2 spectra in Fig. 1 are 0.1017 and 0.1008, and from the least-squares spectra are 0.1025 and 0.1011. Since this is a synthetic example, we can compute the true value: $\rho = 0.100$. Our estimate of ρ is positively biased, being more than six standard deviations too high.

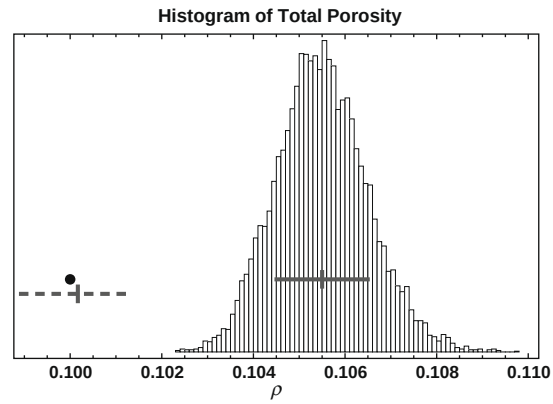

Fig. 5. Histogram of total porosity, ρ, computed from 10,000 samples (using TN3) for the data in Fig. 1. The mean and standard deviation for the histogram are indicated by the solid gray bars. The mean and standard deviation for the bias-corrected results are indicated by dashed bars. The true result for our synthetic model is overlain as a black dot.

As discussed in the previous section, this bias is due to the positive spectral bias for small values of T. The magnitude of this bias, found by computing spectral samples for data containing only the noise used in creating the synthetic data (as was done for the dotted curve in Fig. 4) and then computing total porosity, was found to be 0.00557 ± 0.000799. Subtracting this from the biased estimate yields the reasonable value $\rho \approx 0.0999 \pm 0.0013$. Since the true noise versus time is unknown, a more practical approach is to estimate it by subtracting the data predicted by the least-squares spectral estimate from the measured data, *i.e.,* $\mathbf{m} - \mathbf{G}\mathbf{f}_{LS}$. The total porosity corrected with this estimate of noise was found to be 0.100168 ± 0.00124. This bias-corrected result is plotted as dashed bars in Fig. 5, where it is compared with the true value and the uncorrected estimate. The mean of this bias-corrected result compares well with the true solution. Since the regularized solutions also provide good estimates, one might question the value of using our Monte Carlo approach; it lies in also quantifying the uncertainty of the estimate.

It is well-known that NMR T_2 spectra are sensitive to pore sizes in rocks and other porous media. As a result, T_2 spectra have been used to obtain capillary curves [1]. In particular, it was found that short T_2 corresponds to small pores where water will be held by capillary force. Only water in larger pores (*i.e.,* large T_2) will participate in flow. The fraction of water in large pores (unbound fluid) is important in determining the permeability of a rock from an NMR measurement. The unbound fluid volume is defined as the integrated volume with T_2 larger than T_c:

$$\rho_U = \int_{\log T_c}^{\infty} f(T) \mathrm{d} \log T, \tag{6}$$

where $T_c = 0.033$ s is a typical cutoff value for sandstones [1]. The histogram of ρ_U is shown in Fig. 6, yielding the estimate $\rho_U \approx 0.0870 \pm 0.0021$. Note that the true value, 0.0859, is comfortably within the standard error. No bias correction was needed because ρ_U does not depend on the biased portion of the spectrum. Estimates from the regularized T_2 spectra in Fig. 1 are 0.0847 and 0.0834, and from the least-squares spectra are 0.0725 and 0.0688. Note that both the regularized solutions and the least-squares solutions underestimate the value by more than one standard deviation.

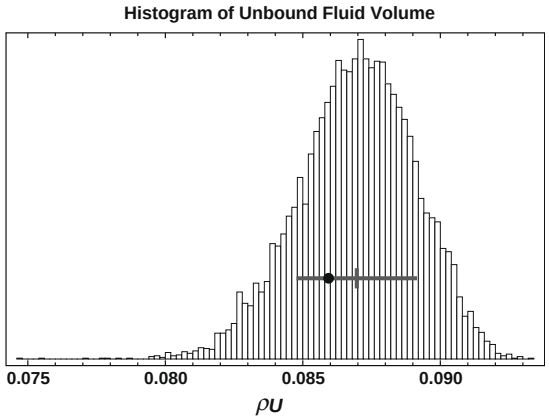

Fig. 6. Histogram of unbound fluid volume, ρ_U, computed from 10,000 samples (using TN3) for the data in Fig. 1. The mean and standard deviation are drawn as gray bars, and the true result for our synthetic model is overlain as a black dot.

5. Conclusions

Although the regularized approach to NMR T_2 spectral inversion yields a solution that is stable in the presence of noise, it fails to capture the considerable uncertainty present in the spectral inversion problem. We demonstrated that the spectral inversion problem can be expressed as a non-negative multinormal distribution when the Bayesian prior is a constant, and we use Monte Carlo sampling to characterize this uncertainty. This description contains no free parameters. Although this distribution is a special case of a truncated-multinormal distribution for which samplers exist in the literature, we show that these one-dimensional Gibbs samplers are ineffective for the T_2 spectral inversion problem. We propose a modest extension of these samplers in which the Gibbs sampling is done in two dimensions instead of one, and demonstrate that it is an efficient sampler for the T_2 spectral inversion problem.

We used this new sampler to generate 10,000 samples for a synthetic example T_2 inversion problem. From these samples we generated the mean and standard deviation of the spectral solution and compared this to the original spectrum used to generate the synthetic data. The original spectrum is shown to lie within the range of our inversion solution. While it is tempting to take the mean spectrum as the inversion result for comparison with, for example, the regularized solution, the inversion result is in fact represented by the *entire ensemble of the solutions* obtained by the Monte Carlo sampler (*e.g.,* 10,000 spectra). The mean spectrum is just one aspect of the solution. The solution ensemble describes the broad range of solution possibilities, *i.e.,* the uncertainty of the solution.

Using our new sampler, we quantify the well-known result that there is an increasing positive bias in the T_2 spectrum with decreasing T. The source of this bias is the non-negativity constraint combined with the diminishing number of data points which constrain the spectral solution with decreasing T. In one of our examples we showed that this bias results in the overestimation of total porosity. Although we could have removed this bias by imposing a prior that smoothed the spectral samples and hence underrepresented spectral uncertainty, we instead presented a method for bias-removal that preserves spectral uncertainty while generating estimates of total porosity that compare well with those produced by a regularization approach. Then what is the advantage of our approach over regularization approaches? We provide the uncertainty of our estimate, a desirable feature of any experimental result.

Appendix A. Review of truncated-multinormal samplers

In this appendix we examine use of existing Monte Carlo samplers for the non-negative normal distribution. Samplers in the literature solve the more general problem of sampling from a truncated normal distribution [10] defined by

$$\pi(\mathbf{f}) \propto \exp\left[-\frac{1}{2}(\mathbf{f}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{f}-\boldsymbol{\mu})\right], \quad \mathbf{Af} \leqslant \mathbf{b}, \tag{A.1}$$

where $\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{\Sigma}^{-1}$ is the inverse covariance matrix, and $\mathbf{A}$ and $\mathbf{b}$ provide the linear truncation constraints. $\mathbf{A}$ is a potentially non-square matrix. In our case, $\mathbf{b} = \mathbf{0}$ and $\mathbf{A} = -\mathbf{I}$. The approaches of [9] and [10] are compared in [10], which refers to the former approach as TN1 and the latter as TN2. A brief outline of these two approaches is given below for the special case of a non-negative normal distribution in order to understand why these approaches are inefficient for T_2 spectral inversion. A third approach [12], based on perfect sampling [11], is not suited to high-dimensional problems and is thus inappropriate for our application.

Both TN1 and TN2 are Gibbs samplers [11]. A Gibbs sampler samples an N-dimensional *pdf* as a sequence of one-dimensional sampling problems in which the sample for a particular dimension, say j, is drawn from the conditional *pdf* $\pi(f_j|\mathbf{f}_{-j})$, where $\mathbf{f}_{-j}$ represents the vector $\mathbf{f}$ with the j-th element removed. Each dimension is sampled, with each new sample replacing its old counterpart in $\mathbf{f}$. After a pass through all dimensions of $\mathbf{f}$, one sample is of $\mathbf{f}$ is generated, and this serves as the starting point for the next sample. This approach is particularly effective for a truncated multivariate normal distribution because each one-dimensional sample is drawn from a truncated univariate normal distribution for which efficient samplers are available [9].

In order to understand the difference between TN1 and TN2, consider a two-dimensional example problem with a high correlation between coordinates x and y. The TN1 sampler samples alternately in the x and y dimensions. The large correlation leads to slow convergence. This example is shown in Fig. A.1a, where the first 20 samples steps, starting from the point $(2, 6)$, are indicated by arrows. The samples remain clustered near the starting point because each one-dimensional update step has its variance limited to the narrow conditional variance in the x or y directions. This leads to a strong correlation between consecutive samples, a problem that is exacerbated with increasing correlation in $\boldsymbol{\Sigma}$.

The importance of having a sampler that is robust in the presence of strong correlation is illustrated in Fig. A.2. Here the eigenvalues of $\boldsymbol{\Sigma}$ for the example of Fig. 1 are seen to grow by a factor of 10^6 over just the first 10 eigenvalues. Since these are the lengths of the principal axes of the covariance ellipsoid, it is clear that the

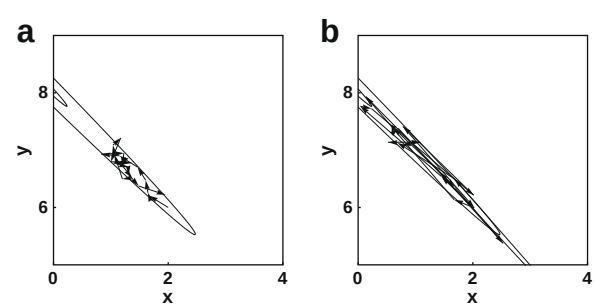

Fig. A.1. An illustration of the differences between Gibbs samplers TN1 and TN2 for our two-dimensional example problem. The contours, at Mahalanobis distances of one and two, indicate the covariance structure, and the mean is at $(-2, 10)$. Panel a shows the first 20 steps of the TN1 sampler starting at the point $(2, 6)$. Note the slow mixing over the long axis of the *pdf*. Panel b shows the first 20 steps of the TN2 sampler starting at the point $(2, 6)$, indicating much better mixing.

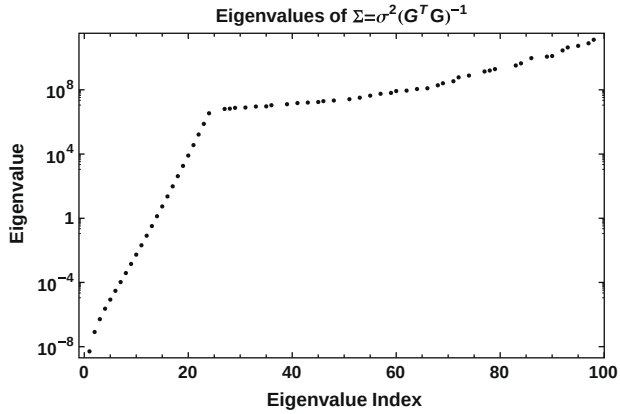

Fig. A.2. The eigenvalues of $\mathbf{\Sigma} = \sigma^2(\mathbf{G}^T\mathbf{G})^{-1}$ are plotted for the example of Fig. 1 to indicate that the uncertainty is effectively infinite along most of the principal axes of the covariance ellipsoid and tiny along most of the remaining axial directions. This indicates a high degree of correlation.

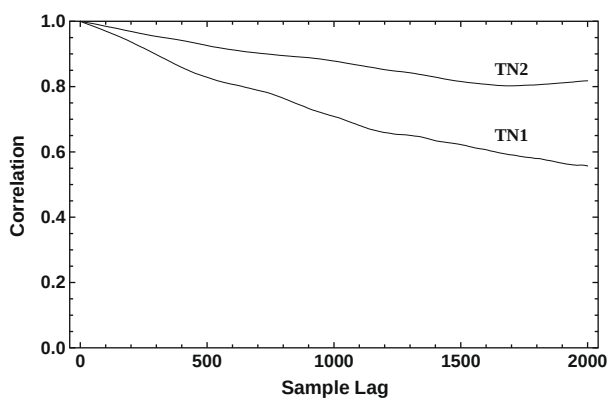

Fig. A.3. Samplers TN1 and TN2 were applied to the T_2 echo data shown in Fig. 1 generating 10,000 samples each. Plotted here is correlation between samples for TN1 and TN2, respectively, for spectral dimension 61 ($T = 0.266$ s). Both algorithms have slow mixing when applied to our NMR data.

uncertainty along most of these axes is effectively infinite, while the uncertainty along a few of these axes is tiny. This indicates a high degree of correlation.

TN2 improves on the efficiency of the TN1 sampler by sampling along the directions of the principle axes of the covariance ellipse (the eigenvector directions) instead of along the coordinate directions. This allows large steps to be taken along directions where the uncertainty is large. Fig. A.1b shows the first 20 samples from the TN2 sampler for our two-dimensional problem. Note that TN2 has much better mixing than TN1, indicating much faster convergence.

The degree of mixing is indicated by the correlation between samples for each dimension. We examine this for the noisy T_2 echo data shown in Fig. 1. Since each dimension has its own correlation plot, we have chosen to plot the correlation for the dimension of maximum observed correlation, dimension 61 ($T = 0.266$ s), which nearly coincides with the peak of the true T_2 spectrum. The correlation plots samples from TN1 and TN2, shown in Fig. A.3, indicate that both samplers require thousands of samples between each independent sample. This is a clear illustration of the inefficiency of these algorithms for T_2 spectral inversion.

The reason for this poor convergence can be understood in terms of the covariance structure in the T_2 spectral inversion problem. This structure is illustrated by two-dimensional conditional covariance plots in Fig. A.4. Starting with $\mathbf{f}$ defined by Eq. 3 for the data shown in Fig. 1a, we examined all consecutive pairs of conditional *pdf*s and found 86 of the form shown in Fig. A.4a and 13 of the form shown in Fig. A.4b, with the latter more likely near the peak of the spectrum. We demonstrated earlier (see Fig. A.3) that TN2 has much better mixing than TN1 for latter form of covariance because the sample steps can jump long distances along the major axes of the covariance ellipse. However, this strategy fails for the covariance structure in Fig. A.4a. For example, when the current state is at the origin in Fig. A.4a, it is clear that no jump will be allowed along the major axis of the ellipse (along the contour direction) because that direction is blocked by the non-negativity constraint, and jumps along the minor axis (perpendicular to the contour direction) will feel a strong pull toward the origin, perpetuating the problem. Hence neither TN1 nor TN2 will converge well when both types of covariance are present.

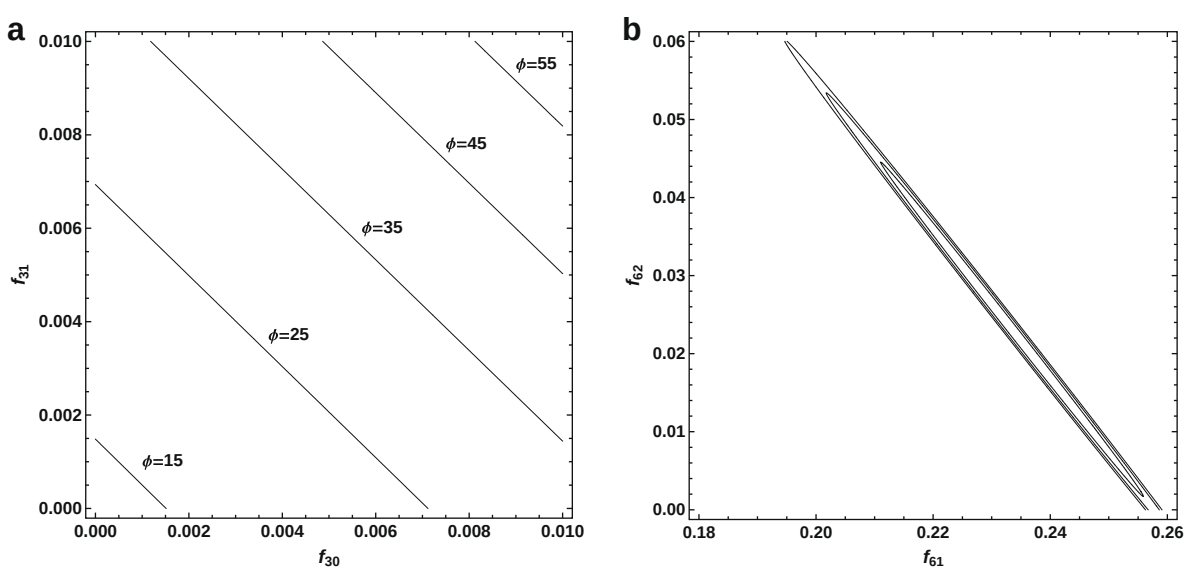

Fig. A.4. Panels a and b show the conditional *pdf*s for T_2 spectral dimensions $(30, 31)$ and $(61, 62)$, respectively, conditioned to the least-squares spectral solution for the T_2 echo data shown in Fig. 1. The contours in (a) are Mahalanobis distances indicated by ϕ, and in (b) $\phi = 1, 2, 3$.

Appendix B. The two-dimensional slice sampler

An excellent description of the slice sampler is given in [11]. A summary of the algorithm is given here. For a *pdf* given by $\pi(\mathbf{x})$ and a beginning state $\mathbf{x}_0$, a uniform random sample, u, is drawn from the interval $0 \leqslant u \leqslant \pi(\mathbf{x}_0)$. Then a uniform random sample of $\mathbf{x}$ is drawn from the domain defined by $\pi(\mathbf{x}) \geqslant u$. This new state, labeled $\mathbf{x}_1$, is an estimate of the random sample satisfying $\pi(\mathbf{x})$. Iterating this procedure leads to an independent sample from $\pi(\mathbf{x})$.

For the truncated normal distribution we can simplify the slice sampler algorithm by defining $\pi(\mathbf{x}_0) \propto e^{\phi(\mathbf{x}_0)}$ with $\phi(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{H}(\mathbf{x} - \boldsymbol{\mu})$ and $\mathbf{H} = \boldsymbol{\Sigma}^{-1} = \sigma^{-2}\mathbf{G}^T\mathbf{G}$. Then a uniform sample of u is drawn from the interval $0 < u < 1$, and $\phi(\mathbf{x}_1) = \phi(\mathbf{x}_0) - 2\log u$. A uniform sample of $\mathbf{x}$ is then drawn from the domain $\phi(\mathbf{x}) \leqslant \phi(\mathbf{x}_1)$ with $\mathbf{x} \geqslant 0$. We use rejection sampling to obtain a uniform sample from this truncated elliptical domain using the algorithm presented in Appendix A of [12].

The two-dimensional conditional values of $\boldsymbol{\mu}$ and $\mathbf{H}$ for dimensions i and j, denoted $\hat{\boldsymbol{\mu}}_m$ and $\hat{\mathbf{H}}_m$, are given by

$$\begin{aligned} \hat{\boldsymbol{\mu}}_m &= \boldsymbol{\mu}_m + \boldsymbol{\Sigma}_{-m}^T (\boldsymbol{\Sigma}_{-mm})^{-1} (\mathbf{f}_{-m} - \boldsymbol{\mu}_{-m}) \\ \hat{\mathbf{H}}_m &= \mathbf{H}_{mm}. \end{aligned} \tag{B.1}$$

The symbol m represents the pair of indices i and j. The subscript m on a vector or matrix indicates the selection of those elements. The subscript $-m$ on a vector indicates the selection of all elements except those of m, and on a matrix it indicates the selection of the columns m minus the rows m. The subscript $-mm$ indicates the selection of all rows and columns except those of m.

To find an expression for $\hat{\boldsymbol{\mu}}_m$ that avoids computing the inverse of the poorly conditioned $\mathbf{G}^T\mathbf{G}$ matrix, we use the inverse formula for a partitioned matrix to get

$$\begin{aligned} (\boldsymbol{\Sigma}_{-mm})^{-1} &= \mathbf{H}_{-mm} - \mathbf{H}_{-m}\mathbf{H}_{mm}^{-1}\mathbf{H}_{-m}^T \\ \boldsymbol{\Sigma}_{-m} &= -\boldsymbol{\Sigma}_{-mm}\mathbf{H}_{-m}(\mathbf{H}_{mm})^{-1}, \end{aligned} \tag{B.2}$$

yielding

$$\hat{\boldsymbol{\mu}}_m = \boldsymbol{\mu}_m - (\mathbf{H}_{mm})^{-1}\mathbf{H}_{-m}^T(\mathbf{x}_{-m} - \boldsymbol{\mu}_{-m}). \tag{B.3}$$

References

[1] R.L. Kleinberg, Well logging, Encyclopedia of Nuclear Magnetic Resonance, vol. 8, Wiley, New York, 1996, pp. 4960–4969.
[2] A.N. Tikhonov, V.A. Arsenin, Solution of Ill-posed Problems, Winston and Sons, 1977.
[3] J. Skilling, Classic maximum entropy, in: J. Skilling (Ed.), Maximum Entropy and Bayesian Methods, Cambridge, Kluwer, 1989, pp. 45–52.
[4] S.F. Gull, Developments in maximum entropy data analysis, in: J. Skilling (Ed.), Maximum Entropy and Bayesian Methods, Kluwer Academic, Dordrecht, 1989, pp. 53–71.
[5] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.
[6] R.L. Parker, Y.-Q. Song, Assigning uncertainties in the inversion of NMR relaxation data, Journal of Magnetic Resonance 174 (2005) 314–324.
[7] D.S. Sivia, J. Skilling, Data Analysis: A Bayesian Tutorial, 2nd ed., Oxford University Press, New York, 2006.
[8] L. Devroye, Non-Uniform Random Variate Generation, Springer-Verlag, 1986.
[9] E. Kermidas, S. Kaufman (Eds.), Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints, Computing Science and Statistics. in: Proceedings of the 23rd Symposium in the Interface, Interface Foundation of North America, Fairfax, VA, 1991.
[10] G. Rodriguez-Yam, R. Davis, L. Scharf, Efficient Gibbs sampling of truncated multivariate normal with application to constrained linear regression, Tech. Rep., Colorado State University, 2004.
[11] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, New York, 2004.
[12] A. Philippe, C.P. Robert, Perfect simulation of positive gaussian distributions, Statistics and Computing 13 (2) (2003) 179–186.