Contents lists available at ScienceDirect

Journal of Applied Geophysics

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



Uncertainty estimates for surface nuclear magnetic resonance water content and relaxation time profiles from bootstrap statistics



A.D. Parsekian^{a,*}, D. Grombacher^b

^a Department of Geology & Geophysics, University of Wyoming, Laramie, WY, United States
^b Department of Geophysics, Stanford University, Stanford, CA, United States

ARTICLE INFO

Article history: Received 23 October 2014 Received in revised form 28 March 2015 Accepted 1 May 2015 Available online 9 May 2015

Keywords: Surface NMR Uncertainty Hydrogeophysics

ABSTRACT

A method for estimating uncertainty in surface nuclear magnetic resonance (NMR) water content and relaxation times utilizing bootstrapping statistics is presented. Bootstrapping is particularly well suited for assigning uncertainty to the surface NMR data set due to the primary factor that degrades surface NMR data quality: ambient electromagnetic noise. We use synthetic forward modeled data with various noise levels applied (the "known uncertainty"), and then demonstrate that a bootstrap resampling of the observed synthetic data can produce an uncertainty estimate that closely represents the "known uncertainty". Finally, we present two field data sets collected under different magnitude ambient noise levels as examples illustrating the result of this approach for estimating uncertainty is computationally intensive, but straightforward to implement and produces useful uncertainty estimates on both water content and relaxation time results for smooth surface NMR sounding models.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Hydrogeophysical measurements have emerged as valuable tools for imaging the subsurface in hydrogeological investigations, estimating hydrologic parameter values and observing hydrologic processes. Evaluation of uncertainty in hydrogeophysical results has been recognized as important when producing data that may be used in a broader hydrogeologic context (Ferré et al., 2009), particularly when these data are used for parameterizing models, and when comparing the hydrogeophysical data to direct "traditional" hydrogeologic measurements. Surface nuclear magnetic resonance (NMR) is a valuable geophysical measurement due to the direct, unambiguous sensitivity to subsurface water content. Recent studies have leveraged this technology for a range of groundwater applications including, for example, aquifer characterization (e.g. Davis et al., 2013), glacial hydrogeology (Lehmann-Horn et al., 2011a), and parameterizing hydrological models (Baroncini-Turricchia et al., 2014).

One of the most common limitations to surface NMR data acquisition is ambient electromagnetic noise that may make signal analysis more challenging and contribute to inaccuracies (Trushkin et al., 1994). Many common sources of ambient electromagnetic noise are related to anthropogenic infrastructure near a measurements site including power lines, cars, trains or radio transmitters. Natural electromagnetic noise sources, such as lightning, are also frequently encountered. Although noise cancelation techniques involving reference loops

* Corresponding author. *E-mail address*: aparseki@uwyo.edu (A.D. Parsekian). (i.e. loops deployed specifically for the purpose of monitoring ambient electromagnetic noise for later digital subtraction) are often able to significantly reduce noise levels in the measured surface NMR data (Walsh, 2008), low signal to noise ratios (SNR) remain a common challenge. The uncertainty of a surface NMR measurement is dependent on the measurement quality, i.e. the SNR (Müller-Petke et al., 2011) and the result of low SNR is increased uncertainty in the estimated aquifer properties. Other factors may contribute to uncertainty in surface NMR measurements such as geometrically imperfect loop shapes that are modeled using simple loop geometries (Lehmann-Horn et al., 2011b), poorly known background magnetic field (B_0) at a measurement site (Walbrecker et al., 2011) or instrument bias, however here we focus only on uncertainty attributed to signal quality (i.e. ambient electromagnetic noise) because we assume it is the most dominant and frequently encountered factor. We aim to demonstrate an approach for estimating uncertainty that addresses the question: Under a given noise condition, how precisely can I estimate water content and relaxation time?

Currently, several common surface NMR inversion routines available to the geophysics community are deterministic (Muller-Petke and Yaramanci, 2010; Walsh, 2008; Behroozmand et al., 2012), meaning that the same result will be obtained each time the computation is executed. Often inversion routines that have a large, fixed number of layers, referred to as smooth inversions, are preferred for determining water content and T_2^* depth profiles because of the assumption that geologic properties change gradually through space. Uncertainty in smooth inversions has been evaluated previously using statistical parameters of the ensemble of stacked measurements and the diagonal elements of the covariance matrix (e.g. Müller-Petke et al., 2011), or by testing how parameters may be varied within the magnitude of the noise (Günther and Müller-Petke, 2012), however these approaches are not exclusively driven by the observed data. Alternatively, stochastic inversion schemes may be used to assess uncertainty, however they require a priori information about the subsurface (e.g. how many layers are present in the subsurface) (Guillen and Legchenko, 2002; Mohnke and Yaramanci, 2002) and typically produce blocky models with few layers. Smooth inversions may be preferable in situations where a priori knowledge of the number of layers in the subsurface is unavailable and also may result in gradual transitions in water content that are consistent with conceptual expectations of hydrostratigraphy. As surface NMR results become more readily utilized in the hydrogeology community there is a demand for methods to seamlessly assign uncertainty to the resulting data sets. Although existing approaches for assigning uncertainty are effective, given the importance of this topic to the utility of surface NMR measurements we believe that there is value in a stochastic assessment of uncertainty that results in statistical distributions of NMR parameters. Furthermore, to our knowledge, the existing approaches for assigning uncertainty have not been validated against synthetic data with noise at a known amplitude.

For this study we test a non-parametric bootstrap resampling to assess uncertainty in surface NMR parameters. This approach was chosen due to the ease of adding the bootstrap algorithm to existing open source inversion routines (i.e. MRSMatlab, Müller-Petke and Yaramanci, 2010), because the resulting uncertainty is driven directly by the data, and because no prior information is needed beyond the signal itself. The objective of this study is to demonstrate the effectiveness of bootstrap resampling for surface NMR using comparisons between synthetic data with known noise and synthetic data with bootstrap analysis applied. We also aim to illustrate the result of bootstrap uncertainty assignment on field data sets.

2. Background: surface NMR and bootstrapping

A comprehensive presentation of the underlying physics related to surface NMR has been covered in several excellent reviews (Weichman et al., 2000; Legchenko and Valla, 2002; Hertrich, 2008), and therefore we briefly present only the most important points here. In the presence of a background magnetic field (B_0) , the magnetic moments of hydrogen atoms in water molecules tend to preferentially align along the direction of B₀ (Earth's magnetic field is used for surface NMR), resulting in the formation of a net magnetization. The surface NMR experiment involves the perturbation and measurement of this magnetization in order to gain insight into subsurface properties such as water content, pore-sizes, and permeability. To perturb the magnetization, an electromagnetic field is generated by pulsing an oscillatory current in a wire loop at the Earth's surface. If the oscillation frequency is selected to be equal to the Larmor frequency ω_0 ($|\omega_0| = |\gamma B_0|$, where γ is the gryomagnetic ratio of the hydrogen atom), the magnetization is perturbed out of alignment with B₀. After the excitation pulse is switched off, the component of the magnetization transverse to B₀ precesses at ω_0 , while the magnetization relaxes back to equilibrium, resulting in a measureable NMR signal.

To investigate subsurface properties at different depths a parameter called the pulse moment, q, determined by the product of the amplitude of the oscillatory current and the pulse duration, is varied; large q values are used to probe the greatest depths while small q values are sensitive to shallow depths. This allows surface NMR to produce depth profiles of the volumetric water content (VWC) and T_2^* relaxation time, a parameter that controls the time-dependence of the signal and that may be used in some situations to provide insight into pore-scale properties (e.g. Grunewald and Knight, 2011). In this paper we aim to quantify uncertainty in the VWC and T_2^* profiles. To avoid confusion, throughout the text we report all VWC values in terms of volumetric units [m³ m⁻³] while uncertainty is uniformly presented as a percentage [%].

The general surface NMR forward model described by Weichman et al. (2000):

$$V(q,t) = -\int_{\text{vol}} \omega_0 M_0 e^{2i\xi(r,\omega)} \sin(\gamma q B_{\perp}^+(r)) B_{\perp}^-(r) e^{-t/T_2 * (r)} w(r) d^3r$$
(1)

where V(q,t) is the measured voltage of the NMR signal in the surface loop at a time t following a pulse moment q. The M₀ term represents the amplitude of the equilibrium magnetization. The exponential term containing ξ describes the signal phase related to the subsurface conductivity. The sine term describes the component of the magnetization that has been rotated into a plane transverse to the B₀ direction. The $B_{\perp}^{+}(r)$ and $B_{\perp}^{-}(r)$ represent the co- and counter-rotating components of the applied magnetic field, and contribute to the perturbation of the magnetization and determine the receive sensitivity, respectively. The exponential containing T_2^* describes the decaying envelope of the NMR signal, while $w(\mathbf{r})$ represents the spatial distribution of the water content in the subsurface. In practice, a simplified version of Eq. (1) is most commonly used where the forward model is reduced from 3D to 1D by laterally integrating Eq. (1). The only spatial variable remaining in the forward model is the depth, z. This simplification contains the implicit assumption that the subsurface is laterally homogeneous (referred to as a layercake Earth). In this case, the forward model takes the form of M\char252ller-Petke and Yaramanci (2010)

$$V(q,t) = \int_{\text{depth}} K(q,z)w(z)e^{-t/T_{2^*}(z)}dz$$
(2)

where the K(qz) term is the kernel function that describes all the terms in Eq. (1) except the exponential containing T_2^* and w(z). Eq. (2) represents the forward model used in this study. The pulse moments, which are generally set by the hardware to logarithmically span from a low pulse moment (~0.1 A s) to the highest pulse moment (~10 A s) to allow an efficient time-saving use of the finite bus voltage, control each measurement spatial sampling of the subsurface. As such, the pulse moments used in a study influence the ability to resolve the w(z), and $T_2^*(z)$ profiles. The subsurface model is described as a series of depth layers, initially thin layers close to the surface and thicker layers at greater depths where the surface NMR measurement is less sensitive. In each depth layer, a single water content and a single T_2^* value is present. This simplification describes a mono-exponential decay within a single depth layer. However, multi-exponential signals are still well-described by this model given that multiple depth layers, and thus multiple independent T_2^* values contribute to the total signal. The goal of the standard surface NMR experiment is to estimate w(z)and $T_2^*(z)$. To estimate the depth profiles in this paper we utilize the QT inversion described by M\char252ller-Petke and Yaramanci (2010) that optimizes for both the water content and T_2^* profiles at once. The inversion involves iterating the w(z) and $T_2^*(z)$ profiles until the data misfit is below the predetermined threshold (typically until $\chi^2 \approx 1$). This is an deterministic process resulting in a best-fitting pair of w(z) and $T_2^*(z)$ profiles that describe the data. This inversion scheme improves the stability of the inversion and resolution of the result by accounting for the information shared between neighboring points in the data space. The challenge that we aim to address is to characterize the uncertainty in the estimated water content and relaxation time depth profiles.

Bootstrapping has been used for uncertainty assessment of geophysical measurements such as seismic (e.g. Sacchi, 1998), magnetics (Constable and Tauxe, 1990) and logging NMR (Parsekian et al., 2015). Our non-parametric bootstrap (e.g. Efron, 1979) strategy randomly resamples the entire surface NMR data set at a predetermined fraction of the original data and inverts this subsample. This statistical approach is well understood to be effective at estimating parameter variance when the distribution of the statistic of interest is unknown (e.g.

Efron, 1987; Bickel and Freeman, 1981). The VWC and T_2^* depth profiles produced using that particular subset of the data are stored and the resampling/inversion process is repeated many times to characterize the distribution of the estimated VWC and T_2^* profiles. We use a bootstrap without replacement approach that draws a number of samples from the original data but will not draw the same sample twice in each cohort. For example in the series of integers from 1 to 10, it would be possible to draw 3, 5 and 9 but it would not be allowed to draw 3, 3 and 3. The full surface NMR data set where V represents voltage observed in the receiving loop, V(q,t), is described by two parameters: the pulse moment q and time t. The pulse moment q controls the spatial sampling of the measurement, thus contributing to the uncertainty in the inversion's ability to determine the spatial origin of the measured signal (i.e. forming the depth profiles). The variable *t* controls the temporal sampling of the signal, which is noisy, and influences our ability to resolve the relaxation times. Therefore, the uncertainty in the predicted water content and relaxation time profiles depends on both parameters. The OT inversion scheme performs the optimization by fitting a model (described by w(z) and $T_2^*(z)$) to a 2D data space (described by q and *t*). We chose to resample both the *q* and *t* dimensions. Furthermore, resampling in both q and t dimensions allows for a greater range of models and thus a more conservative uncertainty estimate. The resampled subset of the original data used in each inversion consists of 90% of the time samples for 50% of the pulse moments. We chose these resampling rates based on an empirical comparison of the ability of several resampling rates to best capture the level of uncertainty. This was evaluated by testing permutations of the resampling rates [50% 70% 90%] for the time and pulse moment were tested; it was found that 90% and 50% sampling rates for time and pulse moments, respectively, balanced the representation of water content uncertainty and relaxation time uncertainty. In the following we aim to test the hypothesis that a bootstrapping approach may be used to characterize the uncertainty in surface NMR water content and T_2^* depth profiles through the comparison of bootstrap predicted uncertainties against known uncertainties. In each of the following cases, we illustrate the results for 1000 resamplings, a quantity that was observed to adequately capture the shape of the probability distribution. We note that in the context of this experiment, we present one set of resampling rates that represents our data well; it would be possible for future implementations of this method to incorporate an optimization to fully explore the effect of resampling in the time and pulse moment domains.

3. Methods

Our approach to validating the bootstrap method for approximating uncertainty is to demonstrate that the bootstrap uncertainty reflects the uncertainty related to a data set with "known" (simulated) noise as indicated in the flow chart shown as Fig. 1 and schematically in Fig. 2. Following this procedure, the first step is to conduct a synthetic experiment where surface NMR data is simulated. Using a simple subsurface structure consisting of three layers of contrasting water content and T_2^* and a survey employing a 25 m diameter loop and sixteen *q*'s logarithmically sampled on the interval [0.28 13.56] [A s], we generate a single synthetic surface NMR data set using the forward modeling framework within the open source surface NMR processing software MRSMatlab

(Müller-Petke and Yaramanci, 2010). Briefly, this process involves using a known set of w(z) and $T_2^*(z)$ in combination with the set of survey parameters (loop size, list of pulse moments) to simulate a data set V(q,t) using Eq. (2). This single data set is illustrated in Fig. 2a; the x- and y-axes indicate the pulse moment and time axes, respectively. The zaxis indicates the corresponding signal amplitude V(q,t). For synthetic data, we represent the subsurface as a 1D layercake, that only varies as a function of depth where one WC and one T_2^* are assigned to each depth layer. Properties were selected to provide contrast between layers. Specific details of the forward model and inversion can be found in Müller-Petke and Yaramanci (2010).

To create the synthetic data to be used to characterize the actual uncertainty we produce noise only data sets (no NMR signal) for four different root mean square noise levels: 1 nV, 10 nV, 20 nV, and 40 nV (corresponding to SNR of ~80, ~8, ~4, and ~2, respectively, estimated as the ratio of the signal's mean initial amplitude across all q's to the noise level at late times). An example of a noise only dataset for a 10 nV noise level is illustrated in Fig. 2b. Gaussian white noise values were drawn from a normal distribution assumed to be consistent with ambient noise. Not included in this type of noise are narrow band harmonics or spikey noise that would also be expected in experimental field data. To form a noisy synthetic data set for a given noise level, the noise free dataset in Fig. 2a is added to a noise only dataset (Fig. 2b) of the corresponding noise level. To illustrate the signal quality for each noise level consider Fig. 3; the left column illustrates the time domain signal and the right column illustrates the signal's spectrum; the effect of increased noise is clearly seen as random deviations from the simple exponential decay curve. To characterize the uncertainty in the predicted VWC and T_2^* depth profiles for each noise level we first generate 1000 independent noise only data sets (i.e. make 1000 independent versions of Fig. 2b) of the appropriate amplitude. Each of these independent noise only data sets are added to the simulated noise free data set (Fig. 2a). This results in 1000 noisy data sets, where the only variation between them is due to random noise fluctuations. This is done for each noise level. Note that only a single noise-free data set is used in this process. As such, any variation between noisy data sets is entirely due to the noise. Each resulting noisy data set is individually inverted to produce a single pair of VWC and T_2^* depth profiles. The cohort of all 1000 results for each noise level are combined to produce a probability distribution for the VWC and T_2^* profiles. We refer to these results as the "random noise" results. This serves as a synthetic analogue to the case where identical field experiments are performed 1000 times sequentially at the same site with unchanging ambient noise conditions. We treat this as representative of the "known" uncertainty in the estimated VWC and T_2 " depth profiles introduced during the inversion due to the noise level.

The next step is to characterize the uncertainty using a bootstrapping approach. We start by selecting a single noisy data set (e.g. Fig. 2c is equal to Fig. 2a plus Fig. 2b) from the cohort of 1000 noisy data sets produced to characterize the known uncertainties. A bootstrap without replacement inversion algorithm as detailed above is run on this single noisy data set to produce a single bootstrapped data set. This bootstrap without replacement procedure is performed 1000 times to produce a suite of 1000 bootstrapped data sets. Each bootstrapped data set represents a subset of the single noisy data set



Fig. 1. Flowchart showing the experimental method.



Fig. 2. Graphic representation of the QT data space and the steps taken to form the data sets used to characterize the uncertainty for the known uncertainty case (top row) and the bootstrap determined uncertainty (bottom row). In each subplot the x- and y-axes represent the pulse moment and time, respectively. Together these parameters span the data space. The z-axis describes the signal amplitude. To form a random noise data set (c), a data set consisting of only noise (b) is added to a noise free surface NMR data set (a). To form a bootstrapped data set (red dots in (d)), we randomly select 50% of the pulse moments, and 90% of the time samples for those pulse moments. These points form the bootstrapped data set. Note that to form a bootstrapped data set no new forward modeling is necessary; it is simply a subset of a full noisy data set.

formed by randomly sampling along the *q* and *t* axes. No new forward modeling occurs in this procedure, the single noisy data set is simply resampled in a random manner 1000 times (i.e. 1000 versions of Fig. 2d are produced from the single data set in Fig. 2c). This procedure involves selecting 50% of the pulse moments, and then selected 90% of the time samples in each of those pulse moments. To visualize this process consider Fig. 2d, where the remaining points indicate a bootstrapped data set formed by sampling the larger full data set. Each of the bootstrapped data sets is individually inverted to produce a single estimate of the water content and T_2^* depth profiles. The results of each of the 1000 bootstrap inversions are combined to produce a probability distribution of the water content and T_2^* depth profiles. We refer to these results as the bootstrap results. Since our objective is only to capture information about uncertainty, once we have a distribution of the VWC and T_2^* profiles for each depth layer described by the cohort of the 1000 bootstrapped VWC and T_2^* profiles, we take this distribution and shift it so that the median is placed at the VWC and T_2^* values determined by a normal deterministic smooth inversion of the full data set. This allows us to use the entire data set to represent the best-fit VWC and T_2^* information while using the bootstrap analysis only to assign uncertainty. The shift of the distribution is not a mathematical manipulation; rather it is simply a way to achieve the advantage of using the entire data set for the deterministic inversion while easily displaying the uncertainty results. We assume that the deterministic model using all of the data (rather than the mean of the bootstrap models) is most representative of the subsurface because this model is informed by the largest number of available data points. Finally, to test our hypothesis whether bootstrapping may be used to characterize the uncertainty of surface NMR, the random noise results are compared to the bootstrapping results under the assumption that if the bootstrap and random noise results have a similar character than the bootstrapping approach is an effective means for evaluating uncertainty for surface NMR parameters due to the influence of the noise level during the inversion.

We have also included a demonstration of the bootstrapping uncertainty evaluation using field data sets collected with different background noise levels to illustrate how this approach for estimating uncertainty works under true environmental noise conditions. For



Fig. 3. In the left column, simulated NMR relaxation time series are shown at 1 nV, 10 nV, 20 nV and 40 nV noise levels (from top to bottom). The spectra for each time series are shown in the right column. The time series shown are taken from data sets used in the suite of 1000 random realizations following the pulse moment of 2.87 A s.

these data sets, we chose a site with low noise (~10 nV noise floor after digital noise cancelation and filtering, SNR ~ 40) and a site with moderate noise (~20 nV noise floor after noise cancelation and filtering, SNR ~ 14). Both data sets were acquired using a Vista Clara GMR surface NMR instrument with multichannel digital noise cancelation functionality based on a coincident transmitting/receiving loop and separate dedicated noise reference loop. The low noise data set "Firehouse" was measured in northwest Wyoming, near the town of Dubois away from anthropogenic development that is typically associated with noise with the exception of a single power line about 500 m from the measurement location. A 30 m circle figure-8 loop was used, eight stacks were collected and digital noise compensation was utilized on one noise reference loop. The moderate noise data set "Red Buttes" was acquired in southeastern Wyoming, near the town of Laramie where the site has power lines on two sides at ~250 m, and a road, train tracks and buildings are nearby within 1 km. A 35 m circle figure-8 loop was



Fig. 4. Noise simulations. The top row shows true water content profiles (blue) and inverted water content frequency distributions from 1000 data sets with simulated random noise at various noise levels. The bottom row shows true T_2^* profiles (blue) and inverted T_2^* frequency distributions from 1000 data sets with simulated random noise. The gray scale indicates regions where many realizations produce similar results, dark gray indicates regions of many occurrences while white indicates regions of few occurrences. The red line corresponds to the VWC and T_2^* profiles for the data sets that were selected to be bootstrapped to produce the results shown in Fig. 5. Each column corresponds to the indicated noise level from 1 nV to 40 nV.

deployed, eight stacks where collected and digital noise compensation was utilized on one noise reference loop. The same bootstrapping method described for the synthetic data was applied to the field data.

4. Results and interpretation

4.1. Random noise uncertainty

First we consider the result for 1000 independent simulated data sets each with random noise applied during the simulation to form the random noise results: this is our "known" noise case. Fig. 4 illustrates the distribution of the estimated VWC and T_2^* profiles for each noise level, shown in grav scale. The true subsurface properties are illustrated by the blue line, while the uncertainty is illustrated in gray-scale. Darker and lighter shades indicate regions with a high and low frequency of occurrence, respectively. Low uncertainty is described by regions with a narrow dark cloud (left column of Fig. 4), while high uncertainty is described by broad regions of light gray (right column of Fig. 4). For the lowest noise level, we see that the smooth inversion results replicate the structure of the true model (blue lines) with low uncertainty, demonstrated by the narrow black uncertainty cloud at all depths in the VWC and T_2^* profile. At higher noise levels (10, 20 nV and 40 nV) the profiles become lighter gray and broader at all depths indicating increased uncertainty. As the noise level increases the VWC profile continues to be represented well, while the T_2^* profile begins to deviate from the true model, particularly at shallow depths. The first and third rows of Table 1 list the mean VWC and T_2^* standard deviations, respectively, for the each noise level averaged over all depths. The standard deviation is observed to grow with the noise level, as expected.

4.2. Bootstrap uncertainty

Next we consider the bootstrapping results for a single data set taken from the cohort of 1000 random noise data sets for each noise level. The VWC and T_2^* profiles corresponding to the single random noise data set selected to be bootstrapped are shown as the red lines in Fig. 4. Fig. 5 illustrates the uncertainty determined by bootstrapping the single data set 1000 times. This process of selecting a single data set and then bootstrapping is performed for each noise level. We observed a similar uncertainty trends to that shown in Fig. 4; the low noise situation shows a narrow dark gray profile representing low uncertainty, while the profiles become lighter gray and fuzzier as the noise level increases. The rate of increasing uncertainty with increasing noise level appears qualitatively to be quite similar to that observed in Fig. 4. Here we see that bootstrapping at the lowest noise level (Fig. 5, left panels) results in a distribution of models that closely follow the known VWC and T_2^* profiles with low uncertainty, characterized by the narrow black cloud. Similar to the random noise results the uncertainty clouds are observed to become lighter gray and broader with increasing noise levels. Qualitatively we see that even under increased noise, the VWC curve generally replicates the subsurface structure, while in comparison, the distribution of T_2^* becomes increasingly less representative of the true model. This is attributed to the generally

Table 1

The depth averaged STD of the estimated VWC and T_2^* profiles. The bootstrapped results are formed by bootstrapping the first 100 noise realizations (from the suite of 1000 random realizations) 1000 times each.

	1 nV	10 nV	20 nV	40 nV
1000 random realizations of VWC STD	0.012	0.024	0.031	0.040
Bootstrapped VWC STD	0.020	0.023	0.031	0.032
1000 random realizations of T ₂ * STD	40 ms	56 ms	64 ms	88 ms
Bootstrapped T_2^* STD	43 ms	57 ms	110 ms	101 ms

short relaxation times assigned to the subsurface in this example meaning fewer points in the time dimension are available to fit the relaxation time in comparison with longer T_2^* values (given a constant sampling rate of the relaxation signal, a short relaxation time will naturally have fewer points to fit than a long relaxation before the signal is below the noise floor).

To quantitatively contrast the ability of the bootstrapping analysis to capture the uncertainty introduced by the noise level, we illustrate the results for a bootstrap analysis performed on the first 100 data sets within the cohort of 1000 random noise data sets used to produce Fig. 4. For each of the 100 data sets, we generate 1000 bootstrapped data sets and invert each to produce a distribution of VWC and T_2^* as a function of depth. This is done for each noise level. From the uncertainty produced using each single data set, we calculate the standard deviation of the VWC and T_2^* distributions within each depth layer to produce a VWC standard deviation and T_2^* standard deviation depth profile. We repeat this process for all of the 100 selected random noise data sets. This results in a set of 100 predictions of the VWC and T_2^* standard deviation depth profiles, illustrated in Fig. 6. The results are shown in gray scale similar to Figs. 3 and 4. The top and bottom rows of the figure show the VWC standard deviation and T_2^* standard deviation depth profiles, respectively. The red lines indicate the standard deviation associated with the known uncertainty results (the random noise results) illustrated in Fig. 4, and the metric against which we will compare the bootstrap results. When viewing Fig. 6, we consider the red lines based on the noise simulations to represent "true uncertainty" in the simulated data sets while the gray distributions represent the range of uncertainty predicted by bootstrapping. A perfect result of bootstrapping exactly replicating "true uncertainty" would be the red lines exactly in the middle of the gray distributions. Table 1 compares the depth averaged VWC standard deviation and T_2^* standard deviation for the 1000 random noise cases with the 100 bootstrapped cases for all noise levels. Contrasting the bootstrap results (gray scale) against the standard deviation resulting from the 1000 independent random noise data sets (red line) illustrates that bootstrapping uncertainty are similar, and increase as a function of noise level. Specifically, at low noise, bootstrapping slightly overestimates the uncertainty in water content and at higher noise (10 nV, 20 nV and 40 nV) the bootstrapping estimate falls approximately equally around the known uncertainty. For T_2^* relaxation times, bootstrapping slightly overestimates uncertainty at each noise level.

4.3. Results and interpretation of field data

The stacked NMR relaxation time series at Firehouse (Fig. 7a) shows the clear exponential decay until about 0.15 s at which point the noise is consistent and low for the duration of the recording. The water content inversion result for this site shows variable uncertainty throughout the profile (Fig. 7b). The high water content layer between 4 m-5 m shows slightly higher uncertainty, this is likely because it is a thin layer approaching the resolution limit in surface NMR. There is a low water content layer between 5 m–8 m. In this layer, we expect uncertainty to be higher because low water content results in a smaller signal, and therefore a locally lower SNR, thus the bootstrap finds a wider range of possible water content results. Below 12 m, the uncertainty for this zone of low water content is low, with the bootstrapping analysis consistently estimating the same VWC. In the T_2^* profile for this sounding (Fig. 7c) we see variable uncertainty with depth. The regions of greatest uncertainty appear to be collocated with regions of transition in the VWC profile. Below 12 m the bootstrap results consistently reproduce similar T_2^* estimates.

The stacked NMR decay for Red Buttes (Fig. 7d) shows an NMR relaxation curve to about 0.3 s after which point the noise is consistent, although distinctly higher amplitude than at Firehouse. The VWC curve (Fig. 7e) shows low water content until 14 m where there is a single high water content layer followed by the deepest layer with



Fig. 5. Bootstrapping results. The top row shows true water content profiles (blue) and inverted water content frequency distributions from 1000 bootstrap simulations (resampling rates of 0.9 and 0.5 were used for time and pulse moment, respectively) at various noise levels. The bottom row shows true T_2^* profiles (blue) and inverted T_2^* frequency distributions from 1000 bootstrap simulations at various noise levels. The gray scale indicates regions where many realizations produce similar results, dark gray indicates regions of many occurrences while white indicates regions of few occurrences. Each column corresponds to a different noise level. The same gray scale is used in Fig. 4. The red line indicates the VWC and T_2^* profiles produced using the full data set, and the location where the median of the bootstrapped probability distributions are centered.

intermediate water content. As would be expected since this site has higher noise, the estimated uncertainty is consistently larger (i.e. broader distribution) than at Firehouse (Fig. 7b). The T_2^* curve for Red

Buttes (Fig. 7f) has more structure than the VWC profile, with the T_2^* profile again showing the greatest uncertainty at regions of transition in the VWC profile.



Fig. 6. The top and bottom panels show standard deviation depth profiles for the VWC and T_2^* estimates respectively. The red line indicates the STD for the case of 1000 random noise realizations (i.e. based on Fig. 4) and is the metric against which we compare the bootstrapped results. The first 100 data sets from the cohort of 1000 random noise realizations (i.e. based on Fig. 5) were each bootstrapped 1000 times to produce VWC and T_2^* profiles that are shown together in gray scale.



Fig. 7. Field data. a) Time series of stacked data for Firehouse site (52 pulse moments, 4 stacks). b) and c) show the uncertainty cloud for VWC and T_2^* profiles for Firehouse site, respectively. d) Time series of stacked data for Red Buttes (36 pulse moments, 8 stacks). e) and f) show the uncertainty cloud for VWC and T_2^* profiles for Red Buttes site, respectively. In each case the uncertainty clouds are composed of the results for 1000 bootstrapped VWC and T_2^* profiles where the median of the probability distributions is centered along the VWC and T_2^* profiles produced using the full data set (red lines). In each case, resampling rates of 0.9 and 0.5 were used for time and pulse moments, respectively.

5. Discussion

5.1. Bootstrapping estimate of uncertainty

As expected, increasing noise results in increased uncertainty for all cases. Table 1 illustrates the comparison of the depth averaged standard deviations at varying noise levels for the VWC and T_2^* profiles, indicating the trend that the T_2^* bootstrap uncertainty is slightly overestimated at higher noise levels, while the water content uncertainty is slightly underestimated at the highest noise and overestimated at the lowest noise. Overestimated uncertainty is preferable to underestimated given that it is likely better to err on the conservative side. However, the bootstrap results appear to capture the trends in uncertainty well and provide acceptable estimates of the uncertainty introduced into the VWC and T_2^* profiles by the noise level.

The primary objective of this study was to demonstrate that bootstrapping could reproduce surface NMR uncertainty similar to known simulated data under known random noise conditions. Comparing the results of the multiple noise simulations with the bootstrap resampling, the uncertainty estimate in VWC and relaxation time from bootstrapping appears visually similar to the multiple noise realizations. To be statistically complete, we use an F-Test – a common analysis designed for comparing standard deviations of two populations. With an $\alpha = 0.05$ (95% confidence interval), we find that the uncertainty

calculated by bootstrapping for VWC at 10 nV and 20 nV is not significantly different from the known uncertainty. Using the same test, we found that the uncertainty calculated by bootstrapping for T_2^* at 1 nV and 10 nV is not significantly different from the known uncertainty. This indicates that the bootstrapping was not able to match the known uncertainty at the 95% confidence level for all cases. We attempt to explain the discrepancy as follows: the 1 nV noise case is not realistic under even the most ideal field conditions and the uncertainty is small, so a perfect match in this case is of little consequence. For all of the T_2^* results (although they are not statistically equivalent) at least in all cases the bootstrapping is overestimating the uncertainty and, as was said before, a slight overestimate of uncertainty is acceptable. The only confounding result was the VWC uncertainty estimated for a noise level of 40 nV by bootstrapping predicted smaller uncertainty than the known value, although we note in this case that the data quality (see the bottom row of Fig. 3) is degraded considerably by the noise.

As noted by Muller-Petke and Yarmanci (2010), adjacent data in both the *t* and *q* directions have shared information because of how the kernel function integrates in space. Since information spans both the t and q space, it is justified to do the bootstrap resampling in both dimensions. This approach results in conservative estimate of uncertainty in the T_2^* profiles. We judge a slight overestimate of uncertainty to be satisfactory because being conservative in the confidence we have in our data is certainly preferred in comparison with overconfidence. We recognize that in certain circumstances, resampling in only the t dimension may be possible (similar to Monte Carlo analysis by Weichman et al., 2002), however given that information is shared in t and q dimensions, we argue that in absence of outside knowledge and as evidenced by our statistical comparisons, resampling in both of these dimensions is appropriate. The spatial sampling of the subsurface is determined by q, which is tied to the estimated depth profiles. We therefore conclude that the t and q resampling bootstrap method effectively assigns uncertainty in moderate noise levels to be expected at most field sites while giving a slightly optimistic uncertainty in very noisy sites.

It is possible to find examples (although quite different from a surface NMR data set) in the literature when multi-parameter data sets are bootstrapped and the uncertainty for one of the parameters is not well resolved (Efron and Tibshirani, 1986). We cannot entirely rule out that this is the case in our data, and perhaps this would partially explain why bootstrapping slightly overestimates uncertainty T_2^* relaxation times at each noise level. Nonetheless, this is certainly not enough evidence to rule out bootstrapping in two dimensions for all cases. Our claim that bootstrapping is effective at characterizing uncertainty in surface NMR is based on our observations that bootstrapping reproduces uncertainties based on "known" noise well.

Similar to the results of Guillen and Legchenko (2002), Müller-Petke et al. (2011) and Chevalier et al. (2014), neither our simulated data nor our field data necessarily result in increased uncertainty as a function of depth as has been shown for other geophysical soundings such as frequency-domain electromagnetic data (Minsley, 2011). At the greatest depths, regularization plays a strong role in determining the surface NMR inversion results, as well as a reduced spatial resolution inherently leading to a smoother inversion result (Müller-Petke et al., 2011). Both of these factors also influence the uncertainty at depth. In this paper we treat the uncertainty as the variation in inversion results driven by ambient electromagnetic noise, the most common contributor to reduced surface NMR signal quality.

5.2. Comparison of field examples

The limited set of past studies available for comparison does not address uncertainty in T_2^* relaxation time, so here we only focus on VWC. Müller-Petke et al. (2011) investigated an ice-covered lake and found the NMR signal had 5% uncertainty at 1 m³ m⁻³ VWC using an ensemble of stacked measurements and the diagonal elements of the covariance matrix to calculate uncertainty. Günther and Müller-Petke

(2012) report that most VWC estimates in their aquifer (~0.35 $\text{m}^3 \text{ m}^{-3}$ VWC) measurements had between 10% and 20% uncertainty, and all uncertainties fell within half of an order of magnitude by testing how parameters may be varied within the magnitude of the noise. Chevalier et al. (2014) estimated uncertainty from a MCMC inversion of 3D surface NMR and found a maximum uncertainty of ~0.18 $m^3 m^{-3}$ in terms of VWC in an aquifer reported having up to 0.5 m³ m⁻³ VWC. Based on a visual inspection of their results, this resulted in a range of uncertainty values averaging about 40%. In this case we do not have a "true" uncertainty to compare our results with; instead we assume that the uncertainty replication shown in the synthetic data sets demonstrates that the uncertainty distributions calculated for the field data are reliable. Comparison with direct measurements is not relevant in this case because we are not trying to demonstrate how closely surface NMR can recreate the subsurface or how accurate the parameter estimates (i.e. VWC) are, as has been done many times before (e.g. Legchenko and Shushakov, 1998; Walsh, 2008). We only seek to reveal information already included in the geophysical measurement about the effect of noise on the data and how much variability could be introduced into the VWC and T_2^* profiles for a given noise level.

5.3. Additional considerations

One observation that may be made about the uncertainty estimates in our smooth models is that we do not explicitly represent uncertainty to subsurface interfaces as can be seen on blocky inversions such as Günther and Müller-Petke (2012). In a basic sense, we assume the assessment of uncertainty with depth is captured by the smoothness of the model. This can be clearly seen in Fig. 5 where elevated noise and associated uncertainty results in smoothing over larger depth intervals and therefore a broader depth range over which one might chose as the interface. However, this means that there are some cases where our bootstrapping approach as implemented may not be suitable: for example, if the depth to the interface is the primary measurement objective, than a method utilizing blocky inversion may be more appropriate than the smooth approach we demonstrate. However, we suggest that a revised version of our bootstrapping approach could be implemented into a blocky inversion framework.

As is common practice in surface NMR inversion, we choose to solve for a monoexponential signal with one T_2^* relaxation per depth layer (Müler-Petke and Yaramanci, 2010; Legchenko and Shushakov, 1998). This simplification results in a substantial reduction in the model space for an already ill-posed problem. Monoexponential signals in each depth layer still can fit a total signal that is multiexponential and if a multiexponential signal is measured, it is still possible to fit a single exponential relaxation decay. Although this is a simplification, it is a necessary one to make the model space smaller and more computationally manageable. A stretched exponential (e.g. Behroozmand et al., 2012) is a different simplification approach that may be preferred and could be attempted with our bootstrapping analysis in the future. We choose the monoexponential because our method is intended to be compatible with the most commonly used inversion approaches. Nonetheless, this monoexponential simplification detail is independent from our uncertainty framework and we suggest that this could be implemented with other inversion approaches in the future.

It is important to note that other considerations should also be considered when determining the uncertainty in the estimated VWC and T_2^* profiles. Bootstrapping allows us to characterize the uncertainty in the inversion results based on the noise level. The resolution of the survey at different depths must also be considered, at great depths resolution decreases resulting in an inherently smoother model at those depths due to the increased area contributing to the volume integration that determines the signal amplitude. In this paper we do not explicitly treat variation in measurement resolution with depth that may be accounted for by supplementing this technique with the use of the resolution matrices (Müller-Petke and Yaramanci, 2010). One last limitation of bootstrapping in general is that this statistical analysis may be computationally expensive given the need to run the many inversions to build the ensemble of resampled results. In comparison with deterministic inversions, the contrast in speed may be significant, however we note that the simulations run for the examples within this manuscript were conducted in several hours using a desktop computer. This is possible because MRSMatlab builds one kernel that can be used in each of the resampled inversions; other implementations of our approach that use a different strategy may not run as quickly.

6. Conclusions

Using forward models of surface NMR data based on a simulated subsurface, we have demonstrated that bootstrapping can reproduce the uncertainty in the inversion VWC and T_2^* results that would have been expected due to known noise conditions. Our approach is a resampling based on one data set - a single stacked sounding. We are not trying to generate new data sets or simulate noise; our uncertainty estimates are driven only by the variability in the data as it has been acquired. We have also applied this method to field data under variable noise. As expected, we find that uncertainty in the resolved NMR parameters increases with ambient electromagnetic noise. Although we did not find that the uncertainty estimates were statistically equivalent in all cases, we do find the uncertainty to be similar or overestimated using bootstrapping in almost all cases - an acceptable result. Furthermore, the visual comparison clearly shows that the bootstrap estimated uncertainty has a very similar character to the known uncertainty in all cases. We highlight this non-parametric bootstrapping approach as an easily implemented way to assess uncertainty due to ambient electromagnetic noise, the most frequently encountered detriment to surface NMR measurements in the field. Widespread implementation of uncertainty estimates to surface NMR data is important for the future integration of this geophysical data into hydrologic models and for transparency when reporting geophysical results for hydrologic science questions.

Acknowledgments

We thank D. Thayer, B. Gordon, S. Miller and G. Paige (University of Wyoming) for their assistance in the field while acquiring surface NMR data in Dubois, WY. We are grateful to J. Maurer (Stanford University) for discussion in bootstrapping statistics. We thank two anonymous reviewers for their comments and suggestions that have improved the quality of the manuscript. We also acknowledge the Wyoming Center for Environmental Hydrology and Geophysics (WyCEHG) EPSCoR Research Infrastructure Improvement (RII) Track-1 project, funded by NSF (Award #1208909) that provided surface NMR instrumentation.

References

- Baroncini-Turricchia, G., Francés, A.P., Lubczynski, M.W., Martínez-Fernández, J., Roy, J., 2014. Integrating MRS data with hydrologic model-Carrizal Catchment (Spain). Near Surf. Geophys. 12 (2), 255–269.
- Behroozmand, A.A., Auken, E., Fiandaca, G., Christiansen, A.V., Christensen, N.B., 2012. Efficient full decay inversion of MRS data with a stretched-exponential approximation of the T₂* distribution. Geophys. J. Int. 190 (2), 900–912.
- Bickel, P.J., Freedman, D.A., 1981. Some asymptotic theory for the bootstrap. Ann. Stat. 1196–1217.
- Chevalier, A., Legchenko, A., Girard, J.F., Descloitres, M., 2014. Monte Carlo inversion of 3-D magnetic resonance measurements. Geophys. J. Int. ggu091.
- Constable, C., Tauxe, L., 1990. The bootstrap for magnetic susceptibility tensors. J. Geophys. Res. Solid Earth 95 (B6), 8383–8395 (1978–2012).
- Davis, A., Munday, T., Somaratne, N., 2013. Characterisation of a coastal aquifer system in the Eyre Peninsula, South Australia, using nuclear magnetic resonance methods. Groundwater in the Coastal Zones of Asia-Pacific. Springer, Netherlands, pp. 89–120.
- Efron, B., 1979. Bootstrap method: another look at the jackknife. Ann. Stat. 7, 1–26. Efron, B., 1987. Better bootstrap confidence intervals. J. Am. Stat. Assoc. 82 (397),
- 171-185. Detter bootstrap connected for the data static Associate (577)
- Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. 54–75.

- Ferré, T., Bentley, L., Binley, A., Linde, N., Kemna, A., Singha, K., Holliger, K., Huisman, J.A., Minsley, B., 2009. Critical steps for the continuing advancement of hydrogeophysics. EOS Trans. Am. Geophys. Union 90 (23), 200–201.
- Grunewald, E., Knight, R., 2011. The effect of pore size and magnetic susceptibility on the surface NMR relaxation parameter T-2*. Near Surf. Geophys. 9, 169–178.
- Guillen, A., Legchenko, A., 2002. Inversion of surface nuclear magnetic resonance data by an adapted Monte Carlo method applied to water resource characterization. Journal of Applied Geophysics 50 (1), 193–205.
- Günther, T., Müller-Petke, M., 2012. Hydraulic properties at the North Sea island of Borkum derived from joint inversion of magnetic resonance and electrical resistivity soundings. Hydrology and earth system sciences 16 (9), 3279–3291.
- Hertrich, M., 2008. Imaging of groundwater with nuclear magnetic resonance. Prog. Nucl. Magn. Reson. Spectrosc. 53 (4), 227–248.
- Legchenko, A.V., Shushakov, O.A., 1998. Inversion of surface NMR data. Geophysics 63, 75–84.
- Legchenko, A., Valla, P., 2002. A review of the basic principles for proton magnetic resonance sounding measurements. J. Appl. Geophys. 50 (1), 3–19.
- Lehmann-Horn, J.A., Hertrich, M., Greenhalgh, S.A., Green, A.G., 2011a. Three-dimensional magnetic field and NMR sensitivity computations incorporating conductivity anomalies and variable-surface topography. IEEE Trans. Geosci. Remote Sens. 49 (10), 3878–3891.
- Lehmann-Horn, A., Walbrecker, J.O., Hertrich, M., Langston, G., McClymont, A.F., Green, A.G., 2011b. Imaging groundwater beneath a rugged proglacial moraine. Geophysics 76 (5), B165–B172. http://dx.doi.org/10.1190/geo2011-0095.1.
- Minsley, B.J., 2011. A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data. Geophys. J. Int. 187 (1), 252–272.

- Mohnke, O., Yaramanci, U., 2002. Smooth and block inversion of surface NMR amplitudes and decay times using simulated annealing. J. Appl. Geophys. 50 (1–2), 163–177.
- Müller-Petke, M., Yaramanci, U., 2010. QT inversion comprehensive use of the complete SNMR dataset. Geophysics 75 (4), W199–W209 (doi:0016–8033).
 Müller-Petke, M., Dlugosch, R., Yaramanci, U., 2011. Evaluation of surface nuclear magnet-
- ic resonance-estimated subsurface water content. New J. Phys. 13 (9), 095002.
- Parsekian, A.D., Dlubac, K., Grunewald, E., Butler, J.J., Knight, R., Walsh, D.O., 2015. Bootstrap calibration and uncertainty estimation of downhole NMR hydraulic conductivity estimates in an unconsolidated aquifer. Groundwater http://dx.doi.org/10.1111/ gwat.12165.
- Sacchi, M.D., 1998. A bootstrap procedure for high-resolution velocity analysis. Geophysics 63 (5), 1716–1725.
- Trushkin, D.V., Shushakov, O.A., Legchenko, A.V., 1994. The potential of a noise-reducing antenna for surface NMR groundwater surveys in the earth's magnetic field. Geophys. Prospect. 42 (8), 855–862.
- Walbrecker, J.O., Hertrich, M., Green, A.G., 2011. Off-resonance effects in surface nuclear magnetic resonance. Geophysics 76 (2), G1–G12.
- Walsh, D.O., 2008. Multi-channel surface NMR instrumentation and software for 1D/2D groundwater investigations. J. Appl. Geophys. 66 (3), 140–150.
- Weichman, P.B., Lavely, E.M., Ritzwoller, M.H., 2000. Theory of surface nuclear magnetic resonance with applications to geophysical imaging problems. Phys. Rev. E 62 (1), 1290.
- Weichman, P.B., Lun, D.R., Ritzwoller, M.H., Lavely, E.M., 2002. Study of surface nuclear magnetic resonance inverse problems. J. Appl. Geophys. 50 (1), 129–147.