Bollettino di Geofisica Teorica ed Applicata Vol. 58, n. 3, pp. 181-204; September 2017
DOI 10.4430/bgta0207

Statistical de-spiking and harmonic interference cancellation from surface-NMR signals via a state-conditioned filter and modified Nyman-Gaiser method

R. Ghanati[1,2] and M.K. Hafizi[1]

[1] *Institute of Geophysics, University of Tehran, Iran*
[2] *Leibniz Institute for Applied Geophysics, Hannover, Germany*

(Received: April 17, 2017; accepted: August 9, 2017)

ABSTRACT The ability to recover subsurface information from surface-NMR measurements depends upon the signal quality which is adversely affected by ambient electromagnetic interference, i.e., power-line harmonics and impulse noises. We discuss two algorithms to isolate and then subtract these interferences. This study first tackles the use of the signal dependent rank-order mean filter for the detection and mitigation of noise spikes from highly corrupted surface-NMR signals. This algorithm estimates the likelihood the sample under inspection is corrupt relative to a threshold value derived from a statistical procedure and replaces a sample identified as impulse noise with an appropriate value. Then, the removal of power-line harmonics is implemented through a linear adaptive method, called a modified frequency-estimation approach stemming from the estimator proposed by Nyman-Gaiser. To verify the performance of the proposed algorithms to eliminate harmonics and spikes, the methods are tested on synthetic signals embedded in artificial noise and noise-only recordings derived from surface-NMR field measurements and a real data set. The results from the numerical simulations reveal an output signal-to-noise ratio increase with an accompanying enhancement in recovery of the surface-NMR signal parameters. Close agreement is also observed between the results of the field example and a borehole located at the sounding.

Key words: hydrogeophysics, time-series analysis, inverse theory, numerical approximations and analysis.

1. Introduction

The surface-NMR method, also called Magnetic Resonance Sounding (MRS), is a non-invasive and emerging geophysical technique providing key information regarding the distribution of water content in the shallow subsurface, pore geometry, and also hydraulic conductivity (Vouillamoz *et al.*, 2012; Behroozmand *et al.*, 2015). The main merit of the surface-NMR method compared with other geophysical tools is that the surface measurement of the surface-NMR signal responds directly to the presence of subsurface water. In other words, the surface-NMR procedure is a way to quantitatively determine water presence in the subsurface, which is impossible with other geophysical methods available today (Hertrich, 2005; Vouillamoz *et al.*, 2007). Surface-NMR is

© 2017 – OGS

the field scale implementation of the nuclear magnetic resonance method in which the nuclei of the hydrogen atoms of water molecules (i.e., protons) in the subsurface are energized by transmitting a resonance EM pulse, and then the energized protons generate a secondary magnetic resonance signal after the excitation pulse is switched off. The signal response of the hydrogen nuclei, which is an exponentially decaying function of time, resonates at the proton Larmor frequency (the resonance frequency of the water molecules in the geomagnetic field varying from 1 kHz up to 2.7 kHz) with signal phase ϕ_0. According to Legchenko and Valla (2003), this so-called detectable free induction decay (FID) signal is given by:

$$S(t) = S_0 \cos(2\pi f_0 t + \phi_0)\, exp\,(-t/T_2^*) \quad (1)$$

where ϕ_0 is the phase offset of the signal, S_0 is the initial amplitude, T_2^* denotes decay time of the FID signal and f_0 indicates the Larmor frequency. Both T_2^* and S_0 are key factors in discovering hydro-geological features. The initial amplitude S_0 is correlated with the amount of water and the relaxation time provides information on the pore structures that contain water. It is well-known that when recording surface-NMR signals, the two most important environmental electromagnetic noise sources are power-line harmonics and noise spikes. Power line harmonics are complicated by the time-varying nature of the fundamental frequency and harmonic content, as well as the varying characteristics across different power grids (Cohen *et al*., 2010). Noise spikes originate from natural and anthropogenic noise sources with the varying appearance and lessen the quality of data. It is well-known that when recording surface-NMR signals, the two most important environmental electromagnetic noise sources are power-line harmonics and noise spikes. Power-line harmonics are complicated by the time-varying nature of the fundamental frequency and harmonic content, as well as the varying characteristics across different power grids (Cohen *et al*., 2010). Noise spikes originate from natural and anthropogenic noise sources with the varying appearance and lessen the quality of data. Based on single-loop surface-NMR measurements, many practical methods have been developed for noise reduction (see Trushkin *et al*., 1994; Plata and Rubio 2002; Legchenko and Valla, 2003; Ghanati *et al*., 2014, 2016a, and references therein). Recently, multi-channel systems (the second generation of surface-NMR instruments consisting of a number of reference loops to record local noise conditions in addition to the primary loop for transmitting and receiving the surface-NMR signal) offer the possibility to measure the time series as broad-band data records at 50 kHz sampling rate instead of providing merely envelopes of the records. It makes to use more sophisticated noise cancelling techniques and it is possible to overcome the drawbacks from the single-channel surface-NMR filtering techniques (see Dalgaard *et al*., 2012; Walsh, 2008; Larsen *et al*., 2014; Müller-Petke and Costabel, 2014). In addition, a number of different spike detection approaches have been reported in the signal processing and geophysics literature, often taking inspiration from the closely related problem of identifying peaks in a time series. For instance, Legchenko (2007) suggested a spike exclusion method based on the simple obliterating of corrupted time sequences. A statistical approach called the Romanovsky criterion to identify and eliminate spiky noises was proposed by Jiang *et al*. (2011). Dalgaard *et al*. (2012) used a spike detection algorithm based on the non-linear energy operator so that the samples containing impulse noises are substituted with zeros. Costabel and Müller-Petke (2014) investigated three de-spiking methods where two of these schemes are applied in the time domain and the third de-spiking approach takes advantage of the wavelet-like nature of spike

events. Larsen (2016) addressed the removal of impulsive noise (spikes) from electric fences through a subtraction-based approach. It is noteworthy that the approach suffers from overfitting and subtraction of the NMR signal. In a recent paper, an efficient post-processing workflow for surface-NMR data using the singular spectrum analysis based de-noising algorithm for noise removal was proposed by Ghanati *et al*. (2016b).

1.1. Problem statement

Since the data acquisition relies on the principle of induction that generates a rather weak voltage in the range of nV to few μV in the surface loop. In addition, the weakness of the recorded signal also causes the measurements to be intensely noise corrupted: as a results, robust and effective noise attenuation approaches are required to preserve the signal of interest which leads eventually to an increase in the accuracy of the parameter estimation.

Let us assume that the surface-NMR data record, $S(t)$, contains spiky events and power-line interference of a quasi-periodic nature, i.e., consisting of a fundamental frequency (50/60 Hz) with harmonics at exactly integer multiples at any given time. The filtering is implemented by isolating and subtracting a reconstructed version of these interferences, i.e.,

$$S_{filtered}(t) = S_{raw}(t) - P(t) \tag{2}$$

where the second term of the above equation ($P(t)$) is related to the harmonic interference and spiky noise. Our task in mitigation the noise contaminators is to isolate $P(t)$ accurately and with computational efficiency. To that end, two techniques are presented. Initially, using a non-linear filter called the signal dependent rank-order mean (SD-ROM), noise spike is detected based on an appropriate thresholding criterion, and then replaced by the mean value of the signal amplitude of the measurement repetitions for the same sample on the time axis. Then, a modified Nyman-Gaiser estimation (MNGE) proposed by Saucier *et al*. (2006) is implemented to remove the harmonic noise. We also apply a non-linear inversion algorithm as the method of reference to verify the estimation accuracy of fundamental frequency, amplitudes, and phases through the MNGE procedure. Furthermore, a performance comparison between the proposed harmonic suppression algorithm and the model-based noise reduction (Larsen *et al*., 2014) is carried out. After de-spiking and power-line harmonic noise removal the surface-NMR signal is still corrupted by the background noise (i.e., random and Gaussian distributed white noise). The Gaussian noise can be reduced by standard averaging of multiple measurements (i.e., the stacking). Finally, the parameters of the surface-NMR signal can be retrieved by the envelope detection process. This includes signal extraction using the digital quadrature detection with additional phase correction (Neyer, 2010; Müller-Petke *et al*., 2011) and the estimate of the underlying surface-NMR signal parameters, such as initial amplitude S_0, decay time T_2^*, frequency, and phase ϕ_0 through a Marquardt-type damped Gauss-Newton method.

1.2. Summary of the paper

Section 2 gives a review of the methods and algorithms. An automatic and efficient thresholding function, that is proportional to the noise level of surface-NMR signals to detect spiky events, in particular those in which the period of their occurrence is too long, is presented.

Section 3 deals with the performance of the proposed strategies. It discusses simulations with

synthetic surface-NMR signals added to simulated noises and real noise-only measurements, as well as a field example.

Finally, a short conclusion summarizes the main ideas in section 4.

2. Methodology

2.1. Automatic detection of spikes

Spiky events of various origins in surface-NMR measurements degrade the performance of the power-line harmonic noise cancelling algorithm (that we present here). Therefore, the noise spikes must be detected and removed before further noise cancelling. In practice, the spikes usually corrupt only a part of the signal section, leaving most of it undisturbed, hence obliterating entire time series is not necessary. In this section, we introduce a fast, robust and automatic de-spiking procedure in the time domain, leading to a considerable reduction of impulse noises. As mentioned earlier, in a variety of de-spiking schemes, corrupted samples (the samples that are dissimilar with respect to their neighbors) are exchanged with zeros. This technique often causes notable distortions in the original signal. The signal dependent ranked-order mean (SD-ROM) filter (Abreu *et al*., 1996; Moor and Mitra, 2000) is a non-linear filter belonging to the class of decision-based filters, or state-conditioned filters (Ferahtia *et al*., 2009). The SD-ROM filter utilizes a spike detector to decide whether the sample under inspection is corrupted or not. In other words, the overall implementation can be viewed as special case of the spike noise-cleaning procedure in which the original sample of a spiky signal is replaced with an appropriate estimate if it is detected as corrupted. Otherwise, the corresponding sample (uncorrupted sample) is left unchanged. The spike detection process is implemented through the following 6 steps:

Step 1: choose a 1-D odd sliding window of size moving on the entire time series (for instance $k = 9$):

$$T = \{S_1(t), S_2(t), S_3(t), S_4(t), S_5(t), S_6(t), S_7(t), S_8(t), S_9(t)\} \tag{3}$$

Step 2: designate the central element of vector T $C(t)$ and then exclude it within the window:

$$\tilde{T} = \{S_1(t), S_2(t), S_3(t), S_4(t), S_6(t), S_7(t), S_8(t), S_9(t)\} \tag{4}$$

Step 3: sort the samples of vector $\tilde{T}$ in ascending order:

$$\tilde{T}_{rearrenged} = \{\tilde{S}_1(t), \tilde{S}_2(t), \tilde{S}_3(t), \tilde{S}_4(t), \tilde{S}_5(t), \tilde{S}_6(t), \tilde{S}_7(t), \tilde{S}_8(t)\} \tag{5}$$

such that $\tilde{S}_1(t) \le \tilde{S}_2(t) \le ... \le \tilde{S}_8(t)$;

Step 4: compute the rank-ordered differences $W(t)$ between the elements of $\tilde{T}_{rearrenged}$ and $C(t)$:

$$W_i(t) = \begin{cases} \tilde{S}_1(t) - C(t) & if\ C(t) \le \mu \\ C(t) - \tilde{S}_{9-1}(t) & if\ C(t) > \mu \end{cases} \tag{6}$$

where $\mu = [\tilde{S}_4(t) + \tilde{S}_5(t)]/2$ and is called the rank-order mean. For a window of size 9, $i = 1,2,...4$;

Step 5: the algorithm decides $C(t)$ is a spike if any of the following conditions hold

$$W_i(t) \geq \sum, \quad i = 1,2,...4$$

where $\sum$ is the thresholding criteria calculated statistically using T-distribution checking, which is principally processed based on the distribution range of actual error in T-distribution. The characteristic of T-distribution is to first exclude a doubtful sample, and then to check based on T-distribution if the excluded sample has gross error. Suppose that one of the measured samples S_q is a doubtful datum, delete this sample and then recalculate the standard deviation of the other measurements as

$$\sigma = \sqrt{\frac{1}{N-2}\sum_{\substack{i=1 \\ i\neq q}}^{N}(S_i - (\frac{1}{N-1}\sum_{\substack{j=1 \\ j\neq q}}^{N} S_j))^2} \tag{7}$$

Then, based on the chosen conspicuousness (significant level) β and number of measurements N, the test coefficient $K(N, \beta)$ of T-distribution function can be acquired by using the T-distribution coefficient table (Ross 2009).

Hence, the threshold criterion is defined as:

$$\sum = \sigma K(N, \beta). \tag{8}$$

Note that generally the choice of threshold is a very delicate and important statistical problem, so that a small threshold leads to attenuating the surface-NMR signal of interest inside the noise spikes, while a large threshold provides inappropriate de-spiking. Hence, choice of the threshold based on the noise level and statistical characteristics of the recordings (i.e., data-driven threshold) might bring about a more efficient elimination of the spiky noises.

Steps 1 to 5 are implemented on all measurement repetitions of one pulse moment (i.e., the ensemble consisting of all single records to be stacked), and the time sequences identified as noise spikes are marked in all considered records;

Step 6: substitute the spiky sample (marked sample) detected in the previous steps with the mean values of the measurement repetitions at the corresponding position on the time axis excluding the marked samples, i.e., excluding the spiky features.

Our numerical experimentations show that a sliding window of size 9 gives better results. Furthermore, it should be noted that in the original SD-ROM algorithm, the detected spiky samples are replaced by the rank-order mean μ, leading to major distortions during the de-spiking process of surface-NMR data. On the other hand, the use of the proposed thresholding criteria enables more accurately identifying spiky events in surface-NMR recordings rather than the median absolute deviation and the standard deviation of the data. In Fig. 1, we present the design of the SD-ROM based de-spiking algorithm. In the following, we deal with the performance of our de-spiking strategy using the modelling of an exponentially decaying synthetic signal with initial amplitude $S_0 = 250$ nV , decay time $T_2^* = 100$ ms, the Larmor frequency $f_0 = 2237$ Hz, phase $\phi_0 = 1.03$ rad, and the time length of 500 ms and the sampling frequency of 20 kHz corrupted by spiky

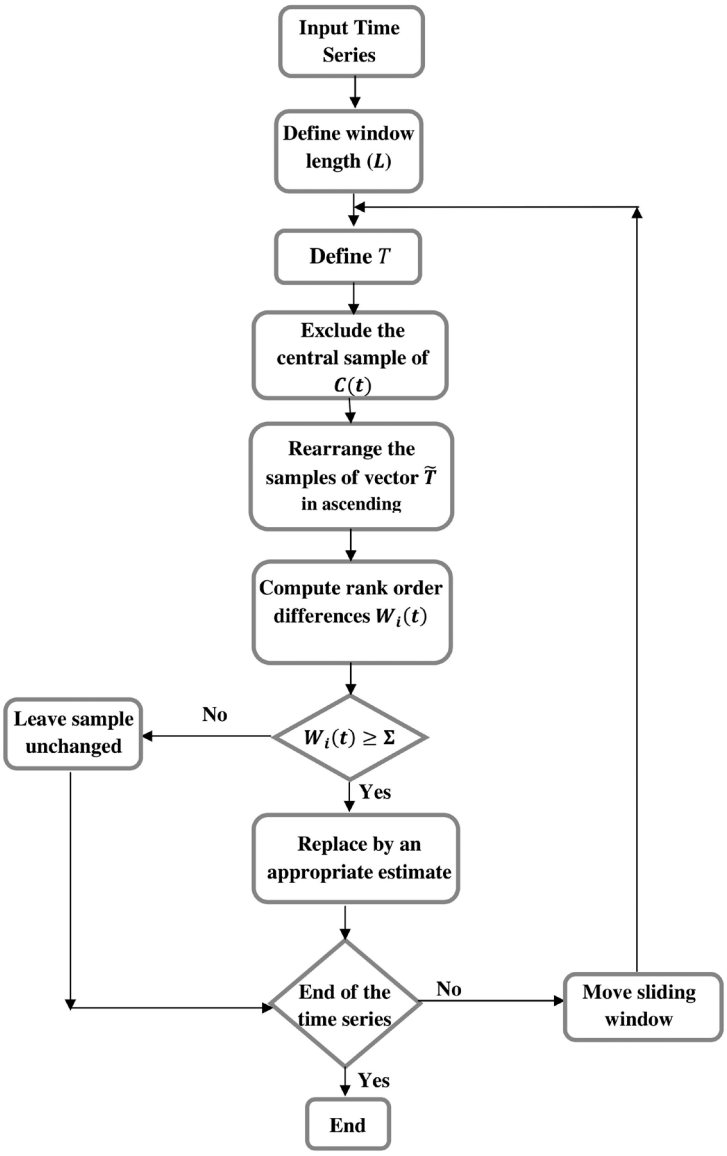

Fig. 1 - Flowchart of the signal dependent rank-order mean filter.

events and Gaussian-distributed noise with standard deviation σ = 150 nV and mean value m = 0. To simulate a spiky event as realistic as possible, we isolate a spiky signal from a real surface-NMR signal recording (see Fig. 2a). Then we simulate realistic measurements by repeating the numerical procedure to generate the synthetic signal 30 times with additional spiky events and Gaussian noise. Each time, the spiky signal occurs randomly at another position on the time axis. It should be noted that the presence of single spikes in the surface-NMR measurements is not a challenging issue, even if many such events occur in the data. In this case, the application of a simple de-spiking procedure provides satisfactory results. But the removal of such noise events from very long periods be might problematic (Costabel and Müller-Petke, 2014). Hence, for the following analysis, we consider a series of spikes like the feature at t = 241 ms to 262 ms shown in Fig. 2a. Fig. 2b illustrates the result of applying the proposed de-spiking method to the single stack shown in Figs. 2a. In addition, the resulting signal after de-spiking and staking 30 single records

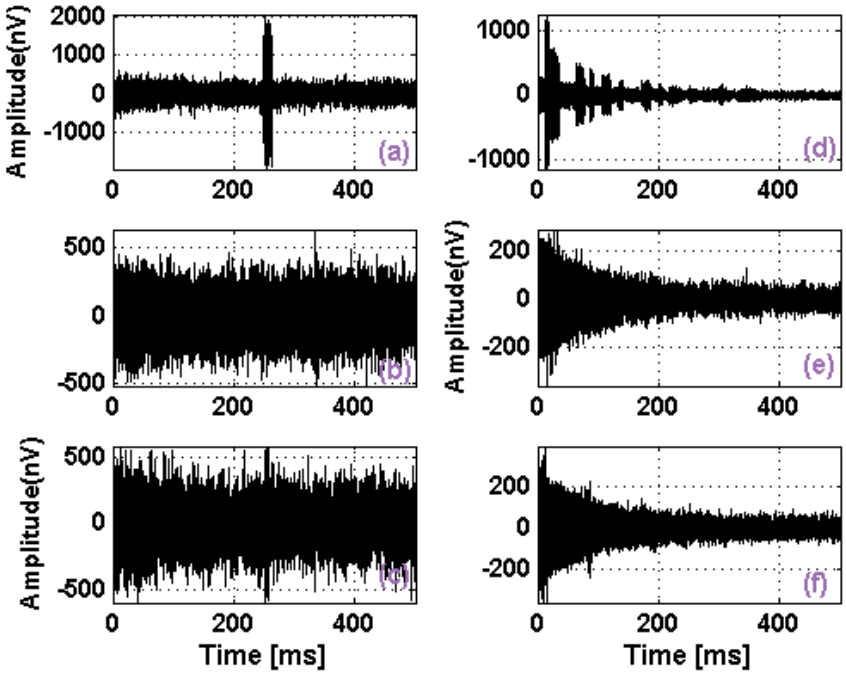

Fig. 2 - Results of applying different de-spiking strategies to a synthetic surface-NMR signal with initial amplitude $S_0 = 250$ nV and a decay time $T_2^* = 100$ ms contaminated with spiky events and Gaussian-distributed noise with standard deviation $\sigma = 150$ nV. Left column, single record; unprocessed (a), processed via SD-ROM (b), processed via TDmean (c). Right column, stacked signal; unprocessed (d), processed via SD-ROM (e), processed via TDmean (f).

and the unprocessed and stacked signal are shown in Figs. 2c and 2d, respectively. Referring to Fig 2c, it can be seen that the spiky noises have been substantially removed through the SD-ROM algorithm and the signal-to-noise ratio increases from -8.21 dB (related to the stacked noisy signal) to 4.67 dB. For comparison, the de-spiking process can also be implemented through the strategies proposed by Costabel and Müller-Petke (2014) on surface-NMR data based on two time-domain thresholding criteria called TDzero and TDmean. The main reason to use TD-based techniques is that in a variety of de-spiking schemes, corrupted samples are exchanged with zeros that may lead to notable distortions in the original signal, while in the TDzero and TDmean methods a more reasonable strategy is applied to substitute spiky samples. Both TDmean and TDzero contain three steps as follows: 1) the threshold value can be determined by relating it to the standard deviation of the data. Then, a user-specified factor is multiplied with the standard deviation of the time series and determines up to which maximum amplitudes the data are accepted and what amplitudes are regarded as spikes. Note that the above process is applied to all single records and the detected spiky samples are marked in all considered records. 2) For every single time sample, the mean value of the whole ensemble is calculated excluding the spiky features. 3) Subsequently, the spiky samples are substituted with the mean values at the corresponding position on the time axis. It should be noted that the only difference between TDmean and TDzero is in the calculation of the transfer function. This means that in TDzero, a time window with the length equal to 2 ms, depending on the data we have, around the corresponding sample on time axis is replaced with zeros, while in TDmean the spiky sample is exchanged with the mean values of uncorrupted samples. Whereas in our proposed processing workflow it is not required

to calculate a transfer function, for comparison we only consider the results derived from the TDmean algorithm. The processing result of the synthetic signal for TD-based strategy is shown in Figs 2e and 2f. Representative results from the proposed de-spiking strategies (i.e., TDmean and SD-ROM) to retrieve the surface-NMR signal parameters (the signal parameters are estimated by envelope detection and fitting the envelope to the mono-exponential decay that are discussed in a later section) are shown in Table 1. One can see that the values of mean absolute percentage error (MAPE) obtained by using the proposed algorithm are lower than those of TDmean, and that the signal parameters (i.e., T_2^* and S_0) are better estimated through SD-ROM. This validates the use of the SD-ROM based de-spiking method over TDmean.

Table 1 - Estimated parameters and MAPE performance of different de-spiking strategies implemented on a synthetic surface-NMR signal (T_2^*=100 ms and S_0=250 n_V) corrupted by Gaussian noise and spiky events.

Parameters Methods	TDmean	SD-ROM
T_2^* (ms)	106.2 ±7.71	104.1 ±6.51
S_0 (n_V)	262.16±18.94	259.13±14.61
MAPE[a] [%]	11.41±6.99	7.9 ±12.32

[a] MAPE: Mean Absolute Percentage Error

2.2. Harmonic interfering noise cancelling

One of the biggest challenges to estimate surface-NMR signal parameters accurately is the presence of harmonic interference that intensively affects the signal of interest. Based on the idea of using multi-channel systems, the data from reference loops are used to reduce the noise level from the surface-NMR measurements by harmonic noise cancellation. Although the reference-loop based noise cancellation methods have proved promising in enhancing the signal to noise ratio in surface-NMR measurements, the presence of spikes in any of the channels results in unreliable adaptive noise cancelling. In other words, for adaptive noise cancelling (ANC) to work efficiently, the primary signal and the reference signal must be highly linearly correlated, whereas the spikes from different sources considerably lessen the correlation between the primary channel and reference channel in the surface-NMR recordings. Furthermore, the adaptive noise canceller may become ineffective when the reference signal contains higher order harmonics (Keshtkaran and Yang, 2014). In this section, we present a relatively robust and computationally efficient processing technique based on the modified Nyman-Gaiser estimation (MNGE) method for suppressing power-line noise in the surface-NMR recordings. This technique involves subtracting an estimation of the harmonic component without distorting or enervating the signal of interest. Nyman and Gaiser (1983) demonstrated that it is possible to eliminate power-line noise in seismic reflection records by modelling the noise as a sum of stationary sinusoids and then subtracting them from each trace. Their technique includes finding a refined fundamental frequency and then sequentially finding phase and amplitude for each harmonic component through Fourier time series analysis.

A recorded surface-NMR signal $X(t)$ can be represented by the following model:

$$X(t) = S(t) + h(t) + \varepsilon(t) \tag{9}$$

where $S(t)$ is the ideal surface-NMR signal from the subsurface protons, $h(t)$ is the power-line interference, and $\varepsilon(t)$ is the non-harmonic noise (i.e., spiky events and Gaussian noise), all sampled at f_S Hz. $h(t)$ consists of a set of harmonic sinusoidal components with unknown frequencies, phases and amplitudes as:

$$h(t) = \sum_{k=1}^{M} C_k \cos\left(k\omega_f t + \Theta_k\right) = \sum_{k=1}^{M} h_k(t) = \sum_{k=1}^{M} \left[a_k \cos(k\omega_f t) + b_k \sin(k\omega_f t)\right] \tag{10}$$

Here, ω_f is the fundamental frequency in rad/s, the amplitude and phase of the k^{th} harmonic are denoted by C_k and Θ_k, respectively, and M indicates the number of harmonics present in the interference. An appropriate harmonic interference elimination algorithm should remove the interference $h(t)$, while preserving the free induction decay signal $S(t)$. Note that the amplitude, phase, and frequency of all harmonics are assumed to remain constant over the length of the record. This assumption is reasonable for recorded lengths of a few seconds or less (Butler and Russell, 1993; Larsen *et al*., 2014). Whereas the maximum signal length recorded by the surface-NMR instruments is less than 1 s, the above assumption is accepted as being valid for the surface-NMR measurements.

The harmonic noise $h(t)$ given by Eq. 10 may also be written in the form

$$h(t) = \sum_{k=1}^{M} \left[a_k \cos\left(k\omega_f t\right) + b_k \sin\left(k\omega_f t\right)\right], \tag{11}$$

where $a_k = C_k \cos(\Theta_k)$ and $b_k = -C_k \sin(\Theta_k)$. Putting Eq. 11 into Eq. 9, we arrive at the following equation. Note that the term $\varepsilon(t)$ is neglected for simplicity.

$$X(t) = S(t) + \sum_{k=1}^{M} \left[a_k \cos(k\omega_f t) + b_k \sin(k\omega_f t)\right]. \tag{12}$$

The recent equation can be written as a system of linear equation: $\boldsymbol{X} = H\boldsymbol{d} + \boldsymbol{S}$, where $\boldsymbol{S} = (s_1(t), s_2(t),..., s_N(t))^T$, $\boldsymbol{d} = (a_1, b_1,..., a_M, b_M)^T$, $\boldsymbol{X} = (x_1(t), x_2(t),..., x_N(t))^T$ and H is an $N\times 2M$ matrix with $H_{n,2i-1} = \cos(i\omega_0 t_n)$, $H_{n,2i} = \sin(i\omega_0 t_n)$, for $n = 1, 2,... N$, $i = 1, 2, 3,..., 2M$. If $S(t)$ is assumed to have Gaussian distribution, the maximum likely estimator of X is $H\tilde{\boldsymbol{d}}$ where $\tilde{\boldsymbol{d}}$ is the standard least-squares solution and hence

$$\tilde{\boldsymbol{d}} = \arg\min_{\mathbf{d}\in\mathbb{R}^{2\mathbf{M}}} \|\mathbf{X} - \mathrm{H}\mathbf{d}\|_2^2 = (H^T H)^{-1} H^T s. \tag{13}$$

For a certain value of the fundamental frequency f_0 an estimator $\tilde{\boldsymbol{d}}$ of the amplitude $\boldsymbol{d}$ can be computed by minimizing the above cost function.

Based on the Nyman-Gaiser estimation (NGE) methodology, frequency f_0 is known to be approximately equal to f_n = 50/60 Hz to estimate the amplitude and phase. Considering an initial guess of f_n of f_0, NGE generates an estimate $\tilde{\delta}$ of the fundamental frequency shift δ as:

$$f_0 = f_n + \delta. \tag{14}$$

Once $\tilde{\delta}$ is achieved, the frequency estimate is revised with $f_n^{r+1} = f_n^r + \tilde{\delta}$, where r is the number of iterations, and this value is used as a starting value for the next iteration. The optimization process is continued until convergence of $\tilde{f}$ is achieved. The NGE algorithm estimates each harmonic separately, leading to a collection ($\tilde{f}_1, \tilde{f}_2, ..., \tilde{f}_M$) of harmonic frequency estimates. Whereas these frequencies f_n are related to each other by $f_n = nf_0$ (n = 1, 2,..., M), it is possible to use a linear combination of the estimates $\tilde{f}_1, \tilde{f}_2, ..., \tilde{f}_M$ to construct a single lower variance estimator of the fundamental frequency f_0.

The proposed method is based on the Nyman and Gaiser's estimation variables defined as:

$$O_k = \frac{1}{T}\int_0^T \cos(k\omega t)\, X(t)dt \tag{15a}$$

$$P_k = \frac{1}{T}\int_0^T \sin(k\omega t)\, X(t)dt \tag{15b}$$

$$Q_k = \frac{1}{T}\int_0^T \cos(k\omega t)\sin\left(\frac{2\pi t}{T}\right) X(t)dt \tag{15c}$$

$$R_k = \frac{1}{T}\int_0^T \sin(k\omega t)\sin\left(\frac{2\pi t}{T}\right) X(t)dt \tag{15d}$$

where T is signal duration and k = 1, 2,..., M. The above estimation variables are used to construct an estimator of the k^{th} harmonic. Note that the amplitude C_k and phase Θ_k of the k^{th} harmonic are estimated using the variables O_k and P_k, respectively. The variables Q_k and R_k allow the estimation of the frequency shift $\delta = \omega_0 - \omega$. Using Eqs. 15 a collection of M estimators of $\tilde{\delta}_k$ δ of is expressed as follows:

$$\tilde{\delta}_k = F_k(O_k, P_k, Q_k, R_k) = \frac{2\pi(O_k R_k - P Q_k)}{kT(O_k^2 + P_k^2)} \tag{16}$$

Eq. 16 defines one estimator of δ for each harmonic of the fundamental frequency. It can be proved that $\tilde{\delta}_k$ are unbiased estimators of δ.

The estimators of $\tilde{C}_k$ and $\tilde{\Theta}_k$ of C_k and Θ_k are also represented by:

$$\tilde{C}_k = 2\sqrt{(O_k^2 + P_k^2)} \tag{17a}$$

$$\cos(\tilde{\theta}_k) = \frac{2(O_k - \pi Q_k)}{\tilde{C}_k} \tag{17b}$$

$$\sin(\tilde{\theta}_k) = -\frac{2(P_k - \pi R_k)}{\tilde{C}_k} \tag{17c}$$

In the statistical context, an optimum estimator needs to be unbiased, meaning that on the average the estimator will yield the true value of the unknown parameter, and minimizes the variance criterion (Kay, 1993). Such an estimator is called the minimum variance unbiased

(MVU) estimator. The determination of the MVU estimator is in general a difficult task because in practice the MVU estimator, if it exists, often cannot be found. For instance, we may not know the probability density function (PDF) of the data or even be willing to assume a model for it. In this case, even some methods which take advantage of the Cramer-Rao lower bound (Kay, 1993), which is a lower bound on the variance of any unbiased estimator, and the theory of sufficient statistics (Fisher, 1922) cannot be applied. Faced with our inability to determine the optimal MVU estimator, it is reasonable to resort to a suboptimal estimator. However, if the variance of the suboptimal estimator can be ascertained and if it meets our system properties, then its use may be justified as being adequate for the problem at hand. If its variance is too large, then we will need to consider other suboptimal estimators to detect one that meets our specifications. A common practice is to limit the estimator to be linear in the data and find the linear estimator that is unbiased and has minimum variance. This estimator is termed the best linear unbiased estimator (BLUE) and can be determined without complete knowledge of the PDF of the data (Kay, 1993).

Saucier *et al*. (2006) demonstrated that the BLUE can be used to develop a single estimator for the error in the initial estimate of the fundamental frequency. This single estimator combines all the estimates for the frequency shifts $\tilde{\delta}_k$ and amplitudes for each of the harmonics. Whereas the variables $\tilde{\delta}_k$ are uncorrelated, which is a consequence of the orthogonality of the estimation variables (Eq. 14), according to Saucier *et al*. (2006) for mutually uncorrelated variables, the BLUE takes the form as

$$\tilde{\delta} = \frac{\sum_{k=1}^{M} \frac{\tilde{\delta}_k}{\sigma_k^2}}{\sum_{k=1}^{M} \frac{1}{\sigma_k^2}} . \tag{18}$$

Here, $\sigma_k^2 = Var\,(\tilde{\delta}_k)$ and $Var\,(\tilde{\delta}) = \frac{1}{\sum_{k=1}^{M} \sigma_k^{-2}}$. It can be proved that σ_k^2 is defined as [the readers are referred to Saucier *et al*. (2006), for more details about determination of variance of estimation variables]

$$\sigma_k^2 \approx \frac{4\pi^2\sigma^2}{A_k^2 k^2 T^3} \tag{19}$$

where σ^2 is the white process variance. Substituting Eq. 19 into Eq. 17, the BLUE takes the following form:

$$\tilde{\delta} = \frac{\sum_{k=1}^{M} \tilde{\delta}_k k^2 \tilde{C}_k^2}{\sum_{k=1}^{M} \tilde{C}_k^2 k^2} . \tag{20}$$

2.2.1. Algorithm implementation

First, the estimation variables for $k = 1, 2,..., M$ are calculated with the discrete form of Eq. 15 (see Appendix A) using the initial angular-frequency estimates $\omega_k = k2\pi f_n$. Then the frequency shifts $\tilde{\delta}_k$ are determined using Eq. 16. We next compute the amplitudes with the help of Eq. 17a and use the results in Eq. 20 to obtain the angular-frequency shift estimate $\tilde{\delta}$. We then use $\tilde{\delta}$ to revise the initial angular-frequency estimate with $\omega_{new} = \omega + \tilde{\delta}$. The optimization process is continued with this new angular-frequency estimate until ω converges. Once ω is calculated, the amplitude estimation problem is implemented using the cost function given by Eq. 12 so that the

final amplitudes obtained by using Eq. 12 are more accurate than those of Eq. 16a. Note that the convergence rate associated to the proposed method is relatively faster than the NGE algorithm so that the convergence of the proposed method is usually achieved with less than 6 iterations. In the following, a non-linear optimization problem is presented to estimate the harmonic-noise parameters.

2.2.2. Non-linear least squares estimation

In this section, we introduce a non-linear least squares problem based on least squares as the method of reference to assess the estimation accuracy of fundamental frequency, phase, and amplitude using the proposed method. Briefly, the determination of the model parameters f_0, Θ_k and C_k in Eq. 11 is a non-linear optimization problem. If we assume that the measurement errors are normally distributed, then the maximum likelihood principle leads us to minimizing the sum of squared errors normalized by their respective standard deviations. We seek to minimize:

$$\phi(\boldsymbol{m}) = \sum_{i=1}^{n} \left| \frac{X_i - F_i(\boldsymbol{m})}{\epsilon_i} \right|^2 = \|\boldsymbol{W}_X[F(\boldsymbol{m}) - \boldsymbol{X}]\|_2^2 \qquad \boldsymbol{m} \in \mathbb{R}^m, F \in \mathbb{R}^{n\times m} \,\&\, \boldsymbol{X} \in \mathbb{R}^{\boldsymbol{n}} \quad (21)$$

with $\boldsymbol{W}_X = diag\ (1/\varepsilon_i)$, where F is the forward modelling operator which is non-linear, $\boldsymbol{X}$ is the measured data vector of length n, W_X is an $n\times n$ data weighting matrix containing the reciprocal of standard deviation for each datum, $\boldsymbol{m}$ is the model parameters, consisting of the fundamental frequency, phase, and amplitude. The data-weighted misfit can also be expressed as chi-square function, $\chi^2 = \phi(\boldsymbol{m})/n$. A value of 1 means fitting the data within error bounds in a least-squares sense (Constable, 1987).

Eq. 20 may be written in the equivalent form:

$$\phi(f_0, \Theta_k, C_k) = \sum_{n=1}^{N} (X(t_n) - C_k cos(k2\pi f_0 t_n + \Theta_k))^2, \quad k = 1,2, \dots, M\,. \quad (22)$$

Expanding the cosine leads to

$$\phi(f_0, \Theta_k, C_k) = \sum_{n=1}^{N} (X(t_n) - C_k cos\Theta_k cos(k2\pi f_0 t_n) - C_k sin\Theta_k \sin(k2\pi f_0 t_n))^2, \quad k = 1,2, \dots, M \quad (23)$$

Note that the cost function ϕ is non-convex in C_k and Θ_k. Hence, we transform it to a convex function by letting $\alpha_k = C_k cos\Theta_k$ and $\beta_k = C_k sin\Theta_k$. Therefore, the new quadratic cost function is presented as:

$$\phi'(f_0, \alpha_k, \beta_k) = \sum_{n=1}^{N} (X(t_n) - \alpha_k cos(k2\pi f_0 t_n) - \beta_k \sin(k2\pi f_0 t_n))^2, \quad k = 1,2, \dots M\,. \quad (24)$$

Here, instead of directly estimating C_k and Θ_k in Eq. 22 we can equivalently estimate α_k and β_k in Eq. 23 to obtain $h(t)$. Accordingly, the unknown model parameters become f_0, α_k, and β_k. Since the problem is non-linear, an iterative minimization is applied.

3. Simulation results

In this section, we present experiments illustrating the performance of the proposed de-spiking and power-line interference removal algorithms in some numerical examples from the modelling of synthetic surface-NMR signals embedded in artificial and real noise recordings. We, then, compare the results with those of a non-linear least squares problem, which is used to determine the fundamental frequency and the amplitudes and phases of all harmonics.

In addition, to better verify the functionality of the MNGE method, particularly in predicting the fundamental frequency, a comparison is made between the proposed method and the model-based removal of harmonics provided by Larsen *et al.* (2014). It is noteworthy that, for both MNGE and the model-based method, the better the fundamental frequency estimation, the more accurate the harmonics retrieval. The processing steps presented in this paper for a surface-NMR recording include: 1) de-spiking, 2) harmonic noise cancelling, 3) stacking, 4) envelope detection, and 5) signal parameters estimation. Based on the proposed processing workflow, the synthetic surface–NMR signal is retrieved by first de-spiking the records corrupted by noise-only measurements. Then, power-line interference is eliminated with the MNGE method or non-linear least squares estimation. After applying the first two processing steps to each simulated single record, the data (the single records) are stacked. Subsequently, the parameters of the surface-NMR signal are determined through envelope identification and fitting the envelope to the mono-exponential decay.

Here, the digital quadrature detection with phase correction is used to extract the surface-NMR signal envelope. The quadrature detection multiplies the FID signal $S(t)$ by $e^{(-j2\pi f_R t)}$, where f_R denotes the frequency of the excitation signal. Then, the phase correction is done by multiplying the complex signal, obtained from the previous step, with $e^{(-j\alpha)}$, where α the phase offset of the signal is defined as:

$$\alpha = tan^{-1}\left(\sum_{n=1}^{N} imag(S(t)e^{(-j2\pi f_R t)}), \sum_{n=1}^{N} real(S(t)e^{(-j2\pi f_R t)})\right) \tag{25}$$

Finally, a low-pass filter (Butterworth filter of fifth order) of e.g., 500 Hz is implemented to improve the signal-to-noise ratio. The cutoff frequency of the low-pass filter is determined based on the idea proposed by Müller-Petke *et al.* (2011). Accordingly, two derived signals are obtained, one in phase (real part) and one in out-of-phase (imaginary part), where the real part of the signal is composed of noise and the FID, while the imaginary part includes only noise components (Müller-Petke *et al.*, 2011). Hence, the real part (amplitude) is used for inversion. The performance of the low-pass filter causes distortions in initial amplitude so that even a careful filter design does not counteract such an effect. To remedy this deficiency, the first few samples of the envelope signal are not included in the exponential fit based on the filter response to an impulse test (Müller-Petke *et al.*, 2016). To estimate the signal parameters, in synthetic experiments, the signal parameters are calculated using the complex mono-exponential fitting through a non-linear optimization problem based on the regularized Levenberg-Marquardt method (Chavent, 2009).

We use two metrics in terms of signal-to-noise ratio improvement and mean absolute percentage error (MAPE), which is a scale error metric and widely understood criterion for assessing estimation accuracy (Hyndman and Koehler, 2006). Mathematically, the MAPE is expressed as:

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|100\times\frac{(S(t_i)-\hat{S}(t_i))}{S(t_i)}\right| \tag{26}$$

where $\tilde{S}(t_i)$ is the reconstructed signal (the processed and stacked signal) and $S(t_i)$ is the ideal signal.

3.1. Application of the proposed schemes to synthetic signals with artificial noise

The applications consist of three synthetic surface-NMR signals generated through Eq. 1 with the parameters defined in Table 2. Then, the simulated signals are corrupted by Gaussian noise with a standard deviation of 110 nV and zero mean. Spiky events and harmonic signals with ten frequencies starting at 1700 Hz, equally spaced by 50 Hz, and the amplitude of each harmonic is randomly chosen between 100 nV and 150 nV. To simulate a spiky event as realistic as possible, we isolate a spiky signal from a real surface-NMR signal recording. Figs. 3a, 3b and 3c show the results of the simulation, which we will refer to as signals 1, 2 and 3, respectively. Then, we simulate realistic measurements (i.e., the stacking process) by repeating the numerical procedure to generate all three synthetic FID signals 15 times. It should be noted that for the simulated stacking the phases of the harmonics were randomly changed (uniformly distributed between $-\pi$ and π) in every single stack. Furthermore, each time the spiky signal (the isolated spike) occurs randomly at another position on the time axis. It is well-known that by increasing the stack size further suppression of the coherence noise is provided. This is because the phases of the power-line harmonics randomly change in every single stack, so that a part of the energy at the harmonic frequencies is diminished during the stacking process (Plata and Rubio, 2002; Strehl, 2006; Costabel and Müller-Petke, 2014). The post-processing sequence starts with the filtering based on the SD-ROM (in all applications, we set the window length of the SD-ROM algorithm equal to 9; in general, we have found that this choice leads to the best results) followed by the proposed harmonic cancellation algorithms that are applied to the single records before stacking. The results of applying the proposed combined procedure to single records associated with signals 1, 2 and 3 are shown in Figs. 3d, 3e and 3f, respectively. Referring to Figs. 3d and 3f, it can be observed that the time series now reveals the presence of a number of small spikes that were masked by the power-line interference. The remaining noise spikes can be removed by a second de-spiking process (Larsen *et al*., 2014). The stacked signals after 15 repetitions are depicted in Figs. 3g to 3i. The consequences of the filtering of signals 1, 2 and 3, i.e., the application of the de-spiking and subsequently of harmonic noise cancellation, in the frequency domain are shown in Fig. 4. From Fig. 4, we can see that the harmonic components are efficiently removed from the corresponding surface-NMR signals, while the Larmor frequencies remain unchanged. Note that the main energy of the desired surface-NMR signal is located at the Larmor frequency which should be preserved during the processing of the data.

Table 2 - Assumed parameters for synthetic signals 1, 2 and 3 contaminated with simulated noise signals.

Parameters Signals	Signal 1	Signal 2	Signal 3
S_0 (n_V)	170	170	170
T_2^* (ms)	100	100	100
f_0(Hz)	1916	1907	1902
ϕ_0 (rad)	1	1	1

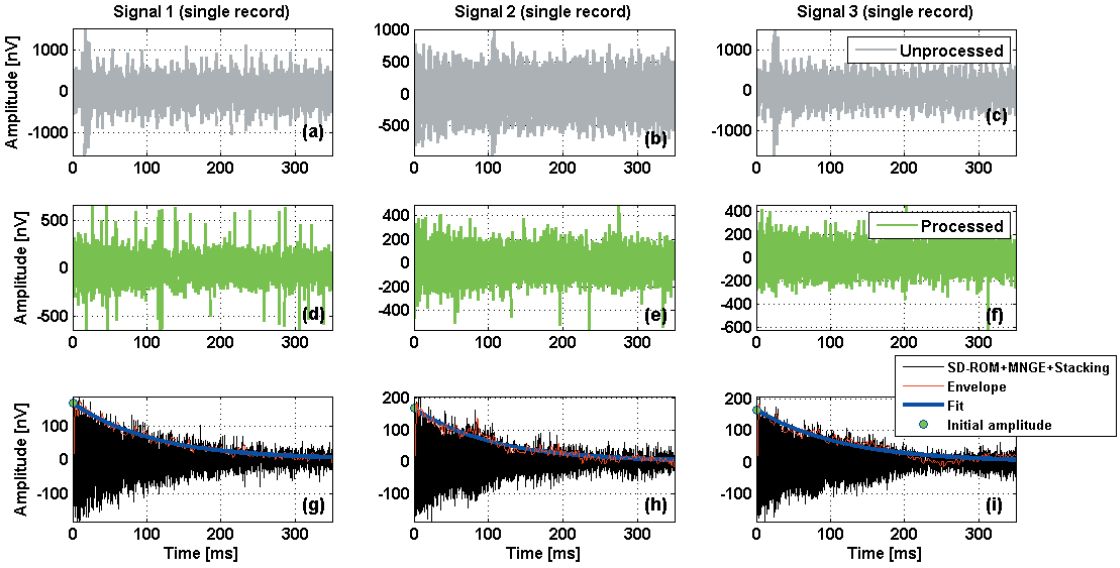

Fig. 3 - Results of applying the proposed filtering method to the synthetic signals contaminated with artificial noise: unprocessed single records (panels a to c), processed single records (panels d to f), and processed and stacked results with 15 repetitions (panels g to i).

For comparison, the harmonic noise subtraction can also be implemented using a non-linear least squares inversion (referred to as the reference method). Fig. 5 displays the resulting time series after implementing the de-spiking method and non-linear least squares estimation on the synthetic signals presented in Figs. 3a to 3c.

In addition, Fig. 6 shows the difference between the power spectrums of the filtered time series (i.e., signals 1, 2, and 3) obtained from the proposed method and the reference method. This difference is small almost everywhere, implying a close correlation between the spectrum obtained with the proposed method and that of non-linear least squares estimation. After applying the proposed filtering procedure, the next step is the envelope detection and fitting the envelope to the mono-exponential decay. The ability of the filtering strategy compared to the reference method to retrieve the surface-NMR signal parameters can be seen in Table 3. The surface -NMR signals have also been retrieved in the case where no filtering has been done. As indicated in Table 3, our

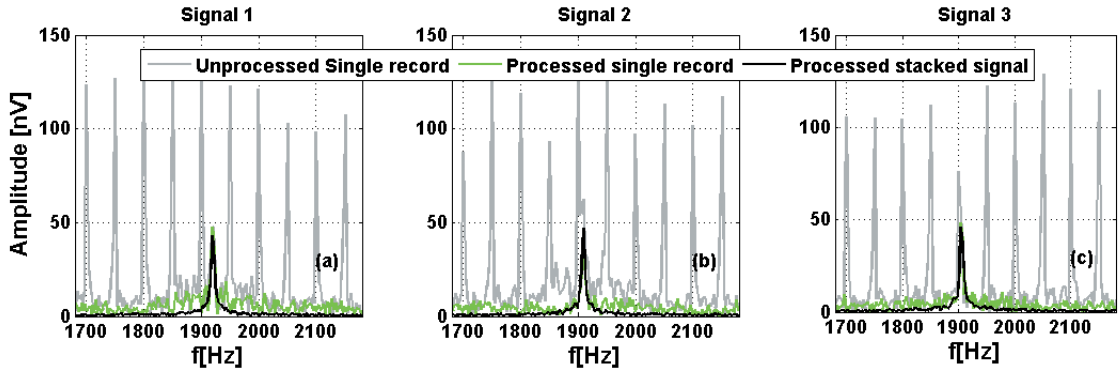

Fig. 4 - Results of applying the proposed filtering method to signals 1 (a), 2 (b), and 3 (c) shown in Fig. 3 in the frequency domain.

filtering scheme shows a close agreement between the assumed and estimated model parameters. We also note that the MNGE algorithm and the reference method give equal overall performance, indicating the relatively high accuracy of the proposed post-processing algorithm. Based on the above observations, it is worth noting that the proposed filtering procedure successfully works even if the Larmor frequency is close to one of the parasitic harmonics, particularly for signal 3 where the frequency offset is equal to 2 Hz. This is a promising result in surface-NMR signal processing, where the adjacency of the Larmor frequency to one of the power-line harmonics is a problem for many algorithms in the surface-NMR literature. Furthermore, Table 3 summarizes the results obtained using the model-based method, indicating a trivial difference between the parameters estimated after implementing the proposed method and those of the model-based method.

3.2. Application of the proposed schemes to synthetic signals with real noise

In this subsection, the proposed noise suppression algorithm is analyzed using synthetic surface-NMR signals superimposed on real surface-NMR noise records with special emphasis on the cases when one power-line harmonic frequency is close to the Larmor frequency. Numerical simulation is carried out using two synthetic FID signals generated through Eq. 1 and perturbed by two different real noise recordings from two different sites at the sampling frequency of 19,200 Hz, referred to as site 1 and site 2, with dissimilar noise conditions so that the noise levels at site 1 are relatively greater than at site 2. Fifteen real noise recordings, collected with a surface-NMR system from each of the sites 1 and 2, are used to corrupt the simulated signals with the parameters represented in Table 4. Figs 7a and 7b show the result of the simulation for one single record, which we will refer to as signal 1 and signal 2. From Fig. 7, it can be seen that the surface-NMR signals are completely masked by electromagnetic noises and the decaying exponential form is not observable; consequently, the retrieval of signal parameters may fail without noise removal. As explained, the filtering procedure of the surface-NMR signal includes two stages based on the proposed combined algorithm. Initially, by applying the SD-ROM method, the noise spikes are removed and in the second stage by applying MNGE or the reference method, then the harmonic components are removed. Figs. 7c and 7d show the processing result of the single records

Table 3 - Performance comparison results of a combination of SD-ROM and MNGE and reference method as well as model-based method implemented on signals 1, 2, and 3 with initial amplitude S_0=170 nV, relaxation time T_2^*=100 ms, phase ϕ_0=1 rad, and the Larmor frequencies equal to 1916, 1907, and 1902 Hz, respectively. The standard deviations of the derived signal parameters are calculated from 10 independent runs.

	Signal 1		Signal 2			Signal 3			
Parameters Methods	SD-ROM[b]+MNGE[c]	SD-ROM+ reference method	SD-ROM+model based method	SD-ROM+MNGE	SD-ROM+ reference method	SD-ROM+model based method	SD-ROM +MNGE	SD-ROM+ reference method	SD-ROM+model based method
$S_0(nV)$	167.64±3.16	167.74±4.51	168.57±5.1	164.7±2.74	164.91±3.09	166.43±4.12	163.46±2.07	162.37±4.12	161.82±6.15
$T_2^*(ms)$	108.52±3.1	104.25±3.01	103.87±6.12	109.9±3.86	108.21±3.14	107.06±3.46	112.28±2.76	109.78±3.19	113.71±5.3
f_0(Hz)	1915.95±0.04	1915.95±0.05	1915.95±0.05	1906.92±0.07	1906.89±0.061	1907.03±0.02	1902.06±0.031	1902.06±0.029	1902.94±0.045
ϕ_0 (rad)	1.02	0.96	1.021	0.99	0.985	1.01	0.97	0.974	0.981
SNR (dB)	25.34	25.01	24.67	23.43	25.01	26.53	19.55	17.01	16.01
MAPE[a] [%]	10.92±3.48	10.46±4.48	8.87±5.41	12.2±3.58	11.71±2.92	11.01±2.71	17.21±4.75	15.74±4.13	18.71±5.23

[a] MAPE: Mean Absolute Percentage Error, [b] Signal dependent rank-order mean, [c] Modified Nyman-Gaiser estimation.

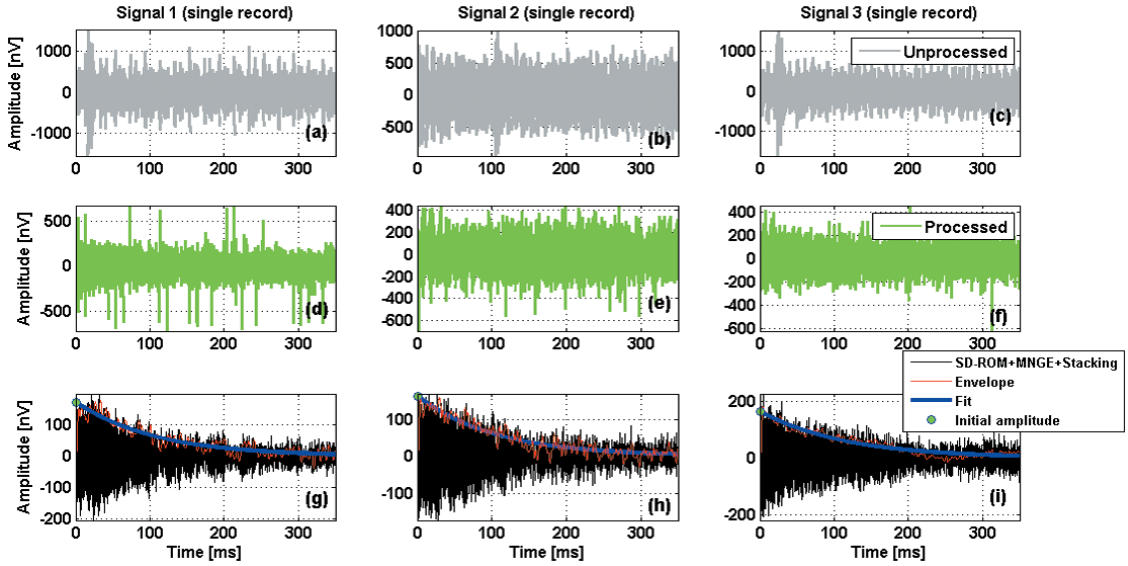

Fig. 5 - Results of applying the reference method to the synthetic signals contaminated with artificial noise: unprocessed single records (panels a to c), processed single records (panels d to f), and processed and stacked results with 15 repetitions (panels g to i).

associated to signals 1 and 2 for the proposed noise cancellation. Furthermore, the processed and stacked signals after 15 repetitions are shown in Figs. 7e and 7f. Inspection of Fig. 7 reveals that the proposed combined method adequately removes the electromagnetic interferences from the simulated FID signals. The consequences of the processing of signals 1 and 2, i.e., the application of the de-spiking and subsequently of harmonic noise cancellation, in the frequency domain are shown in Fig. 8. The figures show that the power-line harmonics are efficiently eliminated by the proposed algorithm and the peak at the Larmor frequency is left undisturbed. It should be recalled that whenever one power-line harmonic is close to the Larmor frequency, retrieval of the surface-NMR is a challenging task, so that the application of common de-noising strategies (e.g., narrow-band IIR notch filter) may lead to the distortion of the wanted signal. In addition, the resulting time series of signals 1 and 2 after applying the presented de-spiking method followed by the reference method are shown in Fig. 9. According to Fig. 10, showing the difference between the

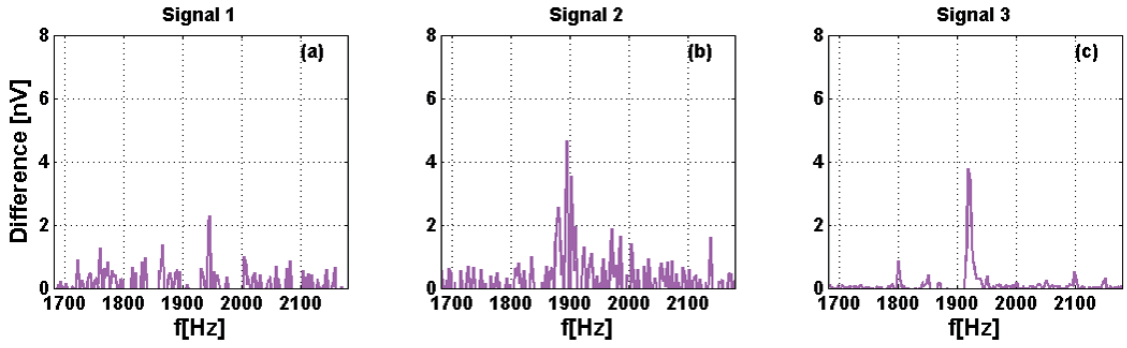

Fig. 6 - Difference between the spectrums obtained with the combination of SD-ROM and MNGE and the reference method associated with: a) processed and stacked signal 1; b) processed and stacked signal 2; and c) processed and stacked signal 1 shown in Figs. 3 and 5.

Table 4 - Assumed parameters for synthetic signals 1 and 2 corrupted by real-noise recordings form sites 1 and 2, respectively.

Parameters Signals	Signal 1	Signal 2
S_0 (n_V)	190	190
T_2^* (ms)	105	90
f_0 (Hz)	2108	2102
ϕ_0 (rad)	1	0.9

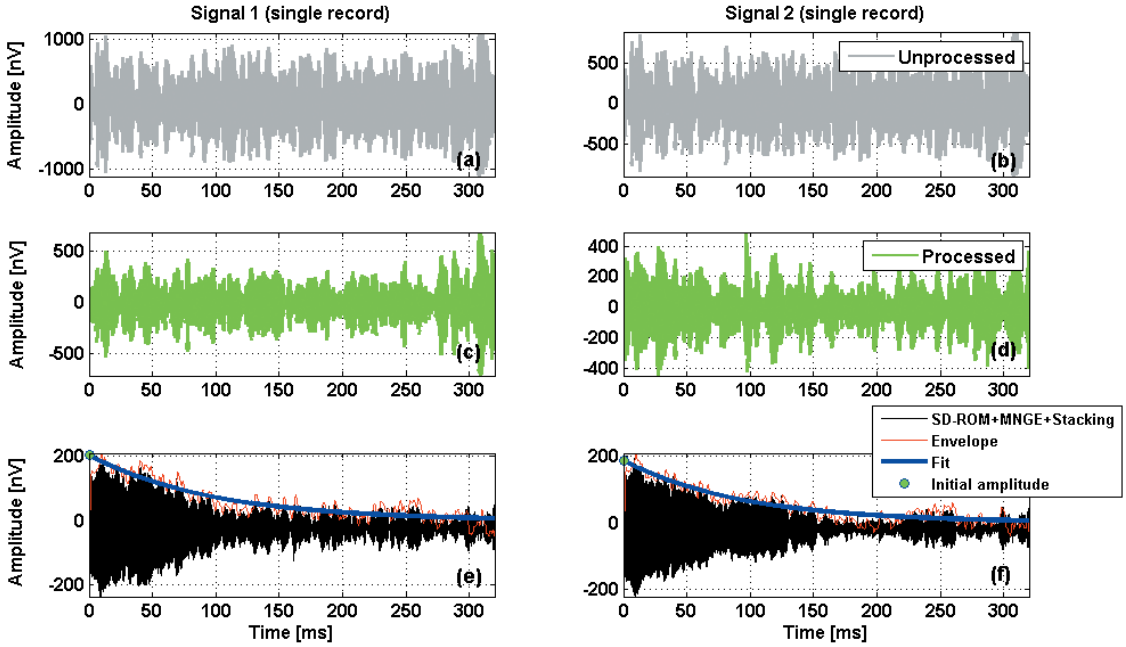

Fig. 7 - Results of applying the proposed filtering method to the synthetic signals contaminated with real-noise recordings: unprocessed single records (panels a and b), processed single records (panels c and d), and processed and stacked results with 15 repetitions (panels e and f).

power spectra associated with the processed and stacked time series (i.e., signals 1 and 2) from the proposed and reference method, one can see that the spectra obtained with our method and with non-linear inversion are comparable almost everywhere except around some of the higher-order harmonics of 50 Hz, where it can reach at most 3 nV. Representative results from the proposed filtering strategies (i.e., model-based method, MNGE and reference method) to retrieve the surface-NMR signal parameters are shown in Table 5. Moreover, Table 5 lists the standard deviation of the signal parameters obtained from 10 independent runs of signal creation. The results clearly show the signal improvements obtained using the presented approach, leading to improved parameter estimation. Referring to Table 5, a reasonable agreement is found between the values of estimated parameters obtained using the proposed scheme and those of the reference method. Moreover, after implementing the model-based method on signals 1 and 2, the resulting parameters are reported in Table 5. The fundamental frequencies predicted using the model-based method and MNGE algorithm associated with signal 1 with the time length of 1000 ms are 49.969 and 49.97, respectively. It is evident that the difference between the fundamental frequency obtained using

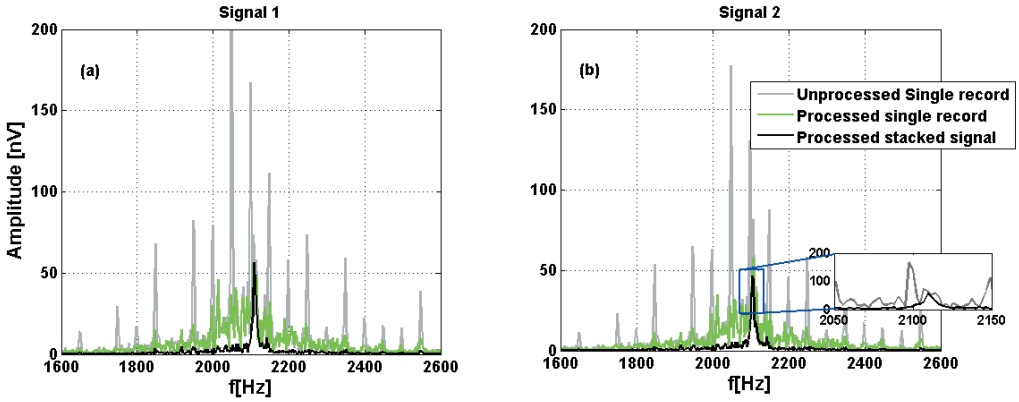

Fig. 8 - Results of applying the proposed filtering method to signals 1 (a) and 2 (b) shown in Fig. 7 in the frequency domain.

the model-based method and that of MNGE is trivial. It should be noted that the model-based method is implemented in a two-step process. First, all harmonics are removed, except for the harmonic close to the Larmor frequency, on the whole time-series. Then, and in the second step, amplitude and phase of the excluded harmonic are calculated based on fitting the model of the corresponding harmonic the last 500 ms of the time-series and extrapolating to the first 500 ms. Through the second step, the model parameters of the harmonic adjacent to the Larmor frequency are estimated using the last part of the FID signal where the MRS signal is low. Our numerical experiments show that often the application of this strategy to signals with duration less than 500 ms leads to unsatisfactory outcomes in accurately estimating the parameters (phase and amplitude) of the harmonic close to the Larmor frequency (this distortion increases for the smaller values of the difference between the Larmor frequency and the closest power-line harmonic).

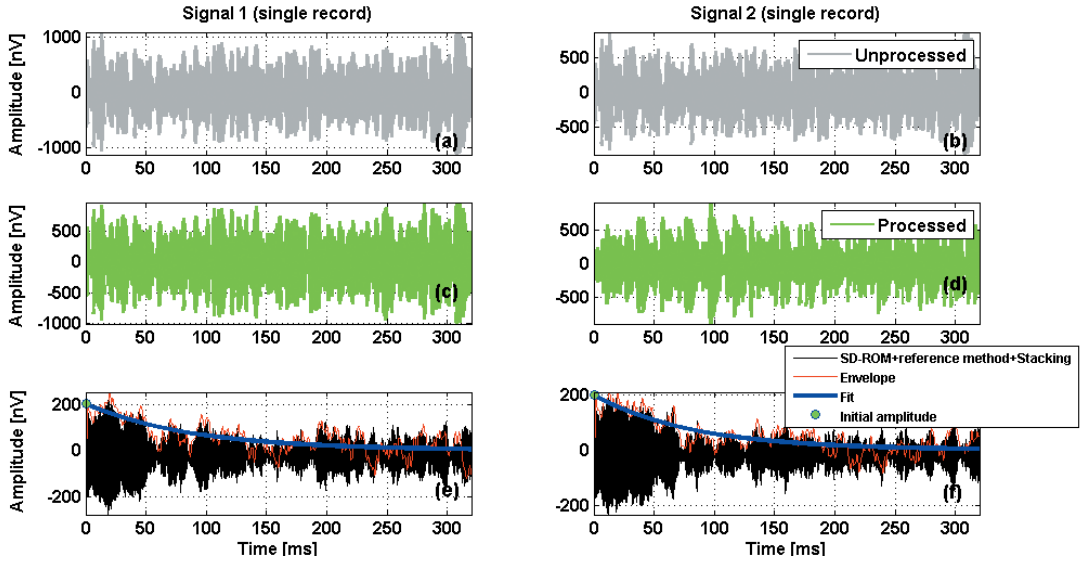

Fig. 9 - Results of applying the reference method to the synthetic signals contaminated with real-noise recordings: unprocessed single records (panels a and b), processed single records (panels c and d), and processed and stacked results with 15 repetitions (panels e and f).

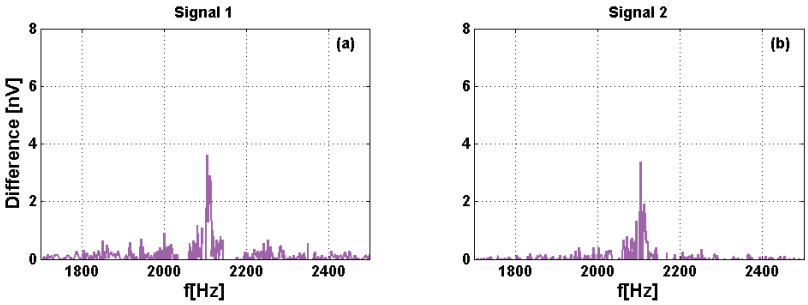

Fig. 10 - Difference between the spectra obtained with the combination of SD-ROM and MNGE and the reference method associated with: a) processed and stacked signal 1, and b) processed and stacked signal 2 shown in Figs. 7 and 9, respectively.

Considering the above observations, the following may be noted:

i) the processing of Surface Nuclear Magnetic Resonance signals is a highly challenging task because the main energy of the wanted signal is located at the Larmor frequency, and hence great care must be taken in preserving the Larmor frequency during the electromagnetic noise cancellation from the surface-NMR measurements;

ii) the main issues related to the proposed de-spiking strategy are: 1) the reliability of the thresholding criterion and 2) the substitution of the detected spiky samples with appropriate values. The de-spiking method with a data-driven threshold was presented aimed at responding to both the first and second issue;

iii) a surface-NMR survey is rather time-consuming, so an arbitrary increase of the stacking rate is not possible. Referring to the synthetic experimentations, it was shown that using the proposed filtering strategy a small number of stacks is required to achieve reliable results. Furthermore, in cases where the power-line harmonics from different sources consist of more than one fundamental frequency, the MNGE method is implemented only once with an initial guess equal to 50/60 Hz, which provides a computationally fast harmonic noise cancellation method;

iv) based on the numerical experiments, in cases where the signal length does not allow estimating the harmonic signal parameters (phase and amplitude) using the last part of the FID signal where the surface-NMR signal mostly includes noise, the proposed method outperforms the model-based method.

Table 5 - Performance comparison results of pure stacking, a combination of SD-ROM and MNGE and reference method as well as model-based method implemented on signal 1 with S_0=195 nV, T_2^*=105 ms, f_0=2108 Hz and ϕ_0=1 rad and signal 2 with S_0=190 nV, T_2^*=90 ms, f_0=2102 Hz and ϕ_0=0.9 rad. The standard deviations of the derived signal parameters are calculated from 10 independent runs.

	Signal 1				Signal 2			
Parameters Methods	Pure stacking	SD-ROM[b]+ MNGE[c]	SD-ROM+ reference method	SD-ROM+model-based method	Pure stacking	SD-ROM+ MNGE	SD-ROM+ reference method	SD-ROM+model based method
$S_0(nV)$	214.14±27.77	191.09±9.16	192.79±7.21	191.74±5.14	232.1±33.95	183.18±6.54	184.14±8.89	186.39±7.45
$T_2^*(ms)$	118±14.02	104.06±7.03	103.27±5.85	105.2±8.15	81.41±16.85	97.75±5.4	96.43±5.12	95.61±2.54
f_0(Hz)	2107.091±0.08	2108.02±0.032	2108.028±0.041	2107.97±0.029	2102.094±0.074	2102.05±0.018	2101.94±0.031	2102.04±0.043
ϕ_0 (rad)	1..089	0.978	1.012	0.97	0.96	0.91	0.991	0.984
SNR (dB)	-1.089	23.47	21.96	24.12	-0.321	27.05	25.74	26.29
MAPE[a] [%]	32.51±396	12.21±6.71	12.13±6.33	12.25±6.02	27.51±7.53	9.76±4.07	11.04±3.59	9.11±6.67

[a] MAPE: Mean Absolute Percentage Error, [b] Signal dependent rank-order mean, [c] Modified Nyman-Gaiser estimation.

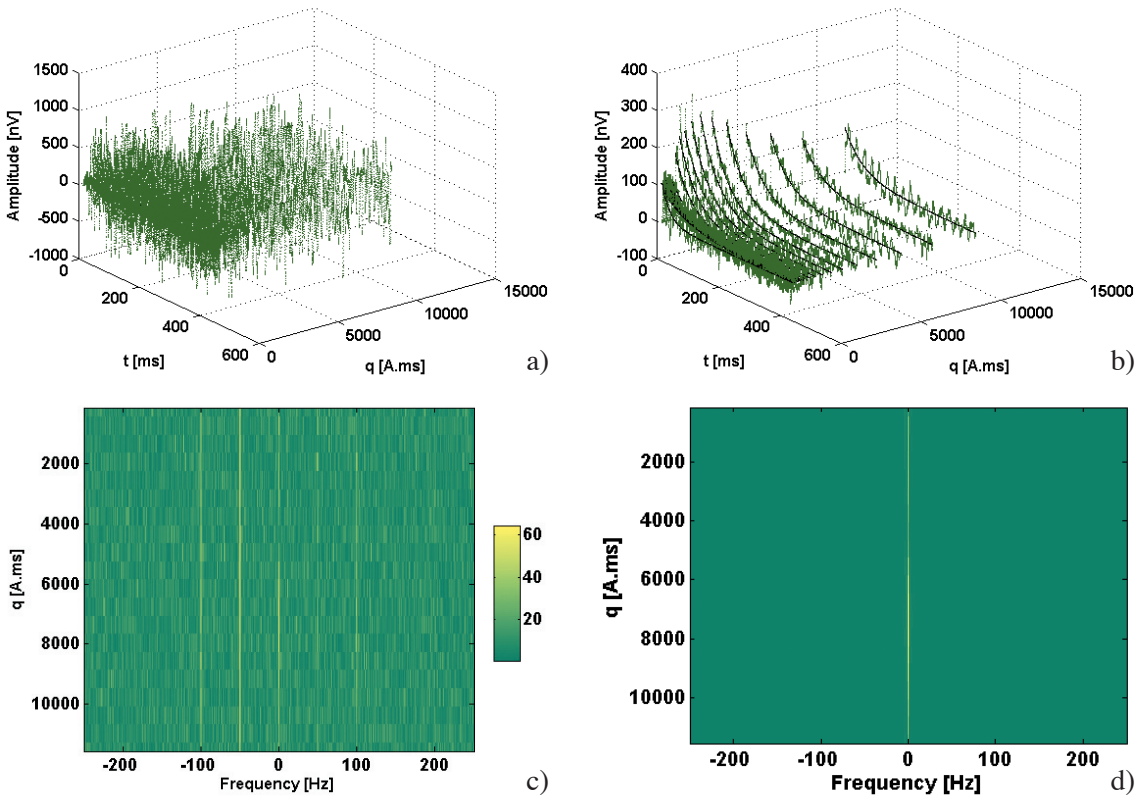

Fig. 11 - Representation of the envelopes detected (green lines) from all the unprocessed (a) and processed (b) FID signals S as functions of measurement time t and pulse moment q, processed (c) and unprocessed (d) data set in the frequency domain, as well the exponential fit (black lines) to the processed signals with data fit residual 11 nV.

3.3. Field example

In the previous subsections, the performance of the proposed algorithm was demonstrated by presenting the results of implementing the combination of the MNGE method and SD-ROM-based de-spiking on several synthetic MRS signals embedded in both artificial and real noises. In the following, we will show the results from the inverted parameters associated with the processed real MRS signals. As explained, the de-noising procedure of the MRS signal includes two stages based on the proposed method. Initially, by applying the SD-ROM algorithm, the MRS signal is recovered by de-spiking the records and in the second stage by applying the MNGE method, then the parasitic harmonics are cancelled. Finally, the data are stacked and the MRS signal parameters are retrieved by an envelope detection technique based on the digital quadrature detection with phase correction followed by a non-linear fitting algorithm. The field data used here was carried out in cooperation with the U.S. Geological Survey, in the Platte River alluvial valley, central Nebraska, U.S.A., within the framework of investigating hydrologic parameters in the near-surface aquifer system. The survey was conducted with GMR instruments (Vista Clara Inc.) using a 46-m-side-eight loop with one turn. The Earth's magnetic field had an intensity of 54013 nT at an inclination of 68.33° and a declination of +7.98°. The MRS field measurements consisted of 20 pulse moments ranging from 134.33 to 11566.94 A·ms. Initially, by applying the proposed filtering algorithm, data are processed. In the second stage, the parameters water content and relaxation time of the processed FID signals, are inverted. Figs. 11a, 11b, and 11c

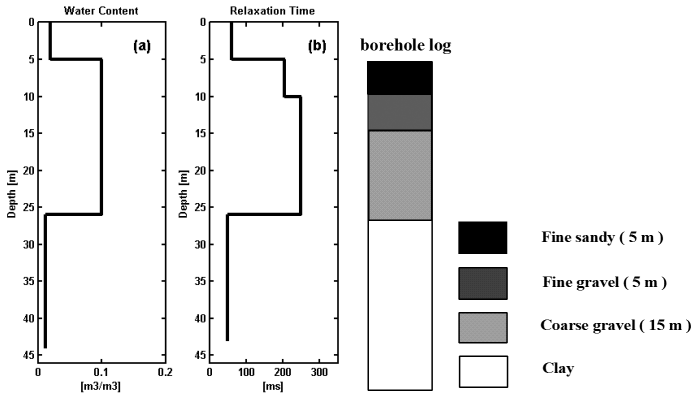

Fig. 12 - Left: inversion results of the field example: a) free water content and b) relaxation time. Right: subsurface according to a borehole lithology located at the sounding with the thickness of each layer. The water table measured in the borehole can be found at a depth of 4.5 m from the subsurface.

show the fit to all the processed signals as functions of measurement time and pulse moment, processed and unprocessed data set in the frequency domain, respectively. The data fit residual, as the L2-norm of the difference of observed data from calculated data divided by the square root of number of measurements, is calculated at 11 nV. The MRS inversion results, using the rotated amplitude data set corresponding to water content and relaxation time, are represented in Figs. 12a and 12b, respectively. Referring to the inversion result, an aquifer layer is detected containing 10% of water, while using the inverted parameter of relaxation time, two layers with decay time of 205 and 251 ms, indicating a significant difference between pore size distribution of the corresponding two layers, are observed. Note also that the water table measured in the borehole lithology approximately corresponds to the depth at which a 205 ms decay time is reached. The information obtained from a borehole located at the sounding corresponds reasonably closely with the resulting inverted parameters from the MRS measurements.

4. Concluding remarks

There has been progress in developing procedures for noise cancellation on multi-channel surface-NMR signals including field and analytical techniques, but little has been done for the noise cancelling of single-channel surface-NMR signals. In this paper, we have presented a new signal dependent rank-order mean filter and MNGE method, which potentially improve these techniques for removing spiky events and power-line harmonics from surface-NMR signals. The proposed de-spiking algorithm was compared with a time domain de-spiking method (i.e., TDmean) so that the application of the signal dependent rank order mean filter for removing noise spikes from the surface-NMR measurements revealed a better performance in terms of MAPE, in addition to improving the fitting parameters. We also compared the results using the proposed harmonic noise cancelling method with those of a non-linear optimization problem as the method of reference to verify the estimation accuracy of fundamental frequency, amplitudes, and phases through the MNGE procedure. The results of numerical experiments from applying the proposed filtering strategy to the removal of electromagnetic noise induced by power-line harmonics, as well as spiky events from synthetic signals corrupted by artificial and real noise recordings, confirmed the relatively good capability of the proposed scheme, leading to more accurate recovery of the signal parameters. It was also shown that the parameters estimated from signals restored with

the proposed harmonic noise cancelling method have comparable accuracy to those of signals restored with the algorithm based on the model-based method. Furthermore, the application of the presented method on a real data set, followed by the inversion of the processed data, resulted in reasonable match between the estimated parameters of the aquifer and those obtained from a borehole lithology located at the sounding.

Acknowledgements. The authors would like to thank Carla Braitenberg (Associate Editor) and the reviewers for their insightful and constructive comments and help in improving the manuscript.

REFERENCES

Abreu E., Lightstone M., Mitra S.K. and Arakawa K.; 1996: *A new efficient approach for the removal of impulse noise from highly corrupted images*. IEEE Transactions on Image Processing, **5**, 1012-1025.

Behroozmand A A, Keating K and Auken E.; 2015: *A review of the principles and applications of the NMR technique for near-surface characterization*. Surveys in Geophysics, **36**, 27-85.

Butler K.E. and Russell R.D.; 1993: *Subtraction of power-line harmonics from geophysics records*. Geophysics, **58**, 898-903.

Chavent G.; 2009: *Nonlinear least squares for inverse problems*. In: Scientific Computation: Theoretical foundations and step-by-step guide for applications. Springer, New York, pp. 209-211.

Cohen M.B., Said R.K. and Inan U.S.; 2010: *Mitigation of 50–60 Hz power line interference in geophysical data*. Radio Science, **45**, RS6002.

Constable S.C., Parker R.L. and Constable C.G.; 1987: *Occam's inversion: a practical algorithm for generating smooth modes from electromagnetic sounding data*. Geophysics, **52**, 289-300.

Costabel S. and Müller-Petke M.; 2014: *Despiking of magnetic resonance signals in time and wavelet domain*. Near Surface Geophysics, **12**, 185-197.

Dalgaard E., Auken E. and Larsen J.; 2012: *Adaptive noise cancelling of multichannel magnetic resonance sounding signals*. Geophys. J. Int., **191**, 88-100.

Ferahtia J., Djarfour N., Baddari K. and Guérin R.; 2009: *Application of signal dependent rank-order mean filter to the removal of noise spikes from 2D electrical resistivity imaging data*. Near Surface Geophysics, **7**, 159-169.

Fisher R.A.; 1922: *On the mathematical foundations of theoretical statistics*. Philosophical Transactions of the Royal Society, **A222**, 309-368.

Ghanati R., Fallahsafari M. and Hafizi M.K.; 2014: *Joint application of a statistical optimization process and empirical mode decomposition to magnetic resonance sounding noise cancellation*. Journal of Applied Geophysics, **111**, 110-120.

Ghanati R., Hafizi M.K. and Fallahsafari M.; 2016a: *Surface nuclear magnetic resonance signals recovery by integration of a non-linear decomposition method with statistical analysis*. Geophysical Prospecting, **64**, 489-504.

Ghanati R., Hafizi M.K., Mahmoudvand R. and Fallahsafari M.; 2016b: *Filtering and parameter estimation of surface-NMR data using singular spectrum analysis*. Journal of Applied Geophysics, **130**, 118-130.

Hyndman R.J. and Koehler A.B.; 2006: *Another look at measures of forecast accuracy*. International Journal of Forecasting, **22**, 679-688.

Hertrich M.; 2005: *Magnetic resonance sounding with separated transmitter and receiver loops for the investigation of 2D water content distributions*. Ph.D. thesis, School of Civil Engineering and Applied Geosciences, Technical University of Berlin, 125 pp.

Jiang C., Lin J., Duan Q., Sun S. and Tian B.; 2011: *Statistical stacking and adaptive notch filter to remove high-level electromagnetic noise from MRS measurements*. Near Surface Geophysics, **9**, 459-468.

Kay S.M.; 1993: *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, 593 pp.

Keshtkaran M.R. and Yang Z.; 2014: *A fast, robust algorithm for power line interference cancellation in neural recording*. J. Neural Eng, **11**, 026017.

Larsen J.J., Dalgaard E. and Auken E.; 2014: *Noise cancelling of MRS signals combining model-based removal of power-line harmonics and multi-channel Wiener filtering*. Geophys. J. Int., **196**, 828-836.

Larsen J.J.; 2016: *Model-based subtraction of spikes from surface nuclear magnetic resonance data*. Geophysics, **81**, WB1-WB8.

Legchenko A.; 2007: *MRS measurements and inversion in presence of EM noise*. Boletín Geológico y Minero, **118**, 489-508.

Legchenko A. and Valla P.; 2003: *Removal of power-line harmonics from proton magnetic resonance measurements*. Journal of Applied Geophysics, **53**, 103-120.

Moore M.S. and Mitra S.K.; 2000: *Statistical threshold design for the two-state signal dependent rank order mean filter*. In: Proceedings of the International Conference on Image Processing, vol. 1, pp. 904-907.

Müller-Petke M. and Costabel S.; 2014: *Comparison and optimal parameter setting of reference-based harmonic noise cancellation in time and frequency domain for surface-NMR*. Near Surface Geophysics, **12** ,199-210.

Müller-Petke M., Dlugosch R. and Yaramanci U.; 2011: *Evaluation of surface nuclear magnetic resonance-estimated subsurface water content*. New Journal of Physics, **13**, 165-185.

Müller-Petke M., Braun M., Hertrich M., Costabel S. and Walbrecker J.; 2016: *MRSmatlab - A software tool for processing, modeling, and inversion of magnetic resonance sounding data*. Geophysics, **81**, WB9-WB21.

Neyer F.M.; 2010: *Processing of full time series, multichannel surface-NMR signals*. MSc thesis, ETH Zurich.

Nyman D.C. and Gaiser J.E.; 1983: *Adaptive rejection of high-line contamination*. In: 53rd Annual International Meeting, SEG, Expanded Abstracts, pp. 321-323.

Plata J. and Rubio F.; 2002: *MRS experiments in a noisy area of a detrital aquifer in the south of Spain*. Journal of Applied Geophysics, **50**, 83-94.

Ross S.M.; 2009: *Introduction to probability and statistics for engineers and scientists*. Elsevier Academic Press, 664 pp.

Saucier A., Marchant M. and Chouteau M.; 2006: *A fast and accurate frequency estimation method for canceling harmonic noise in geophysical records*. Geophysics, **71**, V7-V18.

Strehl S.; 2006: *Development of strategies for improved filtering and fitting of SNMR- signals*. MSc thesis, Technical University of Berlin, 103 pp.

Trushkin D., Shushakov O. and Legchenko A.; 1994: *The potential of a noise-reducing antenna for surface NMR groundwater surveys in the Earth's magnetic field*. Geophysical Prospecting, **42**, 855-862.

Vouillamoz J., Chatenoux B., Mathieu F., Baltassat J. and Legchenko A.; 2007: *Efficiency of joint use of MRS and VES to characterize coastal aquifer in Myanmar*. Journal of Applied Geophysics, **61**, 142-154.

Vouillamoz J.M, Sokheng S, Bruyere O., Caron D. and Arnout L.; 2012: *Towards a better estimate of storage properties of aquifer with magnetic resonance sounding*. Journal of Hydrology, **458**, 51-58.

Walsh D.O.; 2008: *Multi-channel surface NMR instrumentation and soft-ware for 1D/2D groundwater investigations*. Journal of Applied Geophysics, **66**, 140-150.

APPENDIX A
DISCRETE FORM OF ESTIMATION VARIABLES

$$O_k = (1/N)\sum_{n=1}^{N-1} cos(k\omega n\Delta t)\, X(n) \tag{A-1}$$

$$P_k = (1/N)\sum_{n=1}^{N-1} \sin(k\omega n\Delta t) X(n) \tag{A-2}$$

$$Q_k = (1/N)\sum_{n=1}^{N-1} \cos(k\omega n\Delta t)\sin\left(\frac{2\pi n\Delta t}{T}\right) X(n) \tag{A-3}$$

$$R_k = (1/N)\sum_{n=1}^{N-1} \sin(k\omega n\Delta t)\sin\left(\frac{2\pi n\Delta t}{T}\right) X(n) \tag{A-4}$$

Corresponding author: Reza Ghanati
Institute of Geophysics, University of Tehran, Iran
Phone: +98 21 61118573; email: rghanati@ut.ac.ir