Processing harmonic EM noise with multiple or unstable frequency content in surface NMR surveys

Thomas Kremer, Jakob Juul Larsen, Frederic Nguyen

Geophysical Journal International

Geophys. J. Int. (2019) **0**, 1–0
doi: 10.1093/gji/ggz307
Advance Access publication 2019 July 10
GJI Marine geosciences and applied geophysics

Processing harmonic EM noise with multiple or unstable frequency content in surface NMR surveys

Thomas Kremer [1] Jakob Juul Larsen [2] and Frédéric Nguyen [1]

[1] *Department of Applied Geophysics, University of Liege, Liege, 4000, Belgium. E-mail: kremer.tho@gmail.com*
[2] *Department of Engineering, Aarhus University, 8000 Aarhus C, Denmark*

Accepted 2019 July 4. Received 2019 July 2; in original form 2018 October 18

SUMMARY

The harmonic electromagnetic noise produced by anthropic electrical structures is a critical component of the global noise affecting geophysical signals and increasing data uncertainty. It is composed of a series of harmonic signals whose frequencies are multiple integers of the fundamental frequency specific to the electrical noise source. To date, most model-based noise removal strategies assume that the fundamental frequency constraining the harmonic noise is single and constant over the duration of the geophysical record. In this paper, we demonstrate that classical harmonic processing methods lose efficacy when these assumptions are not valid. We present several surface nuclear magnetic resonance field data sets, which testify the increasing probability of recording the harmonic noise with such multiple or unstable frequency content. For each case (multiple frequencies or unstable frequency) we propose new processing strategies, namely, the *2-D grid-search* and the *segmentation* approach, respectively, which efficiently manage to remove the harmonic noise in these difficult conditions. In the process, we also apply a fast frequency estimator called the Nyman, Gaiser and Saucier estimation method, which shows equivalent performance as classical estimators while allowing a reduction of the computing time by a factor of 2.5.

Key words: Hydrogeophysics; Instrumental noise; Time-series analysis; Harmonic EM noise; Surface nuclear magnetic resonance.

1 INTRODUCTION

Surface nuclear magnetic resonance (SNMR) is a near-surface geophysical technique that has been gaining momentum as an efficient tool to non-invasively detect, locate and quantify groundwater resources, as well as estimate the hydraulic characteristics of the subsurface including porosity and permeability (Behroozmand *et al*. 2015).

As a result of its sensitivity to water content, interest in the method has been significantly widened during the last decade with major technologic and methodological advances such as multichannel acquisition devices (Walsh 2008, Dlugosch *et al*. 2011; Liu *et al*. 2019), QT-inversion scheme (Müller-Petke & Yaramanci 2010) and tomographic inversion algorithms (Hertrich *et al*. 2007; Dlugosch *et al*. 2014). Coupled with significant progress in signal processing techniques, those developments have been made available to the geophysical community in open-source SNMR data processing and inversion software (e.g. Müller-Petke *et al*. 2016; Irons *et al.* 2018a).

1.1 Improving the signal-to-noise ratio

The main bottleneck in SNMR surveying remains signal-to-noise ratio due to the relatively weak magnitude of the SNMR signal (tens to hundreds of nV) compared to the electromagnetic (EM) noise level, often several orders of magnitude stronger. The amplitude, phase and frequency characteristics of the EM noise are very site- and time-dependent due to the type of EM noise sources surrounding the study location (Dalgaard *et al*. 2014).

The EM noise signal is often described as the sum of three components (Larsen *et al*. 2013), which are, namely, the spiky noise, the random noise and the harmonic noise. We here focus on the latter component. We refer to Legchenko & Valla (2002), Costabel & Müller-Petke (2014), Larsen (2016) and Irons *et al*. (2018b) for studies dealing in the former two. The harmonic noise is composed of the harmonic EM signals associated with the fundamental frequency of the transmission powerlines or other electrical structures (railway, pipelines, etc.). This type of EM noise generally exhibits very high amplitudes and must therefore be removed from the acquisition signal, otherwise the SNMR data cannot be retrieved nor processed.

To do so, several techniques have been developed. Butler & Russell (1993) first presented and compared the methods of block subtraction and sinusoid subtraction applied to the removal of the harmonic noise in seismo-electric records, where the signal-to-noise ratio is also often very low. Later, Legchenko & Valla (2003) applied these techniques in the context of SNMR acquisition signals, and also compared them to the notch filtering approach. Both studies agree on the fact that the block subtraction method loses efficiency when the percentage of the harmonic noise in the signal decreases, which can be problematic in complex noise environments. The main drawback of the notch filtering approach is that the SNMR signal gets distorted if the Larmor frequency is close to one of the harmonic frequencies ($\Delta f < 8$ Hz), at which point a compensation procedure must be applied, which adds uncertainty to the filtering process and ultimately to the SNMR parameters' estimation. The sinusoid subtraction approach takes advantage of the periodic nature of the harmonic noise to predict a noise estimate that will be subtracted from the recorded signal. Butler & Russell (2003) provided a detailed description of the method in the seismo-electric context, which was then adapted to the processing of SNMR data by Larsen *et al*. (2013), who referred to it as the *model-based* harmonic processing method. They demonstrated the efficiency of this processing scheme (see also Larsen 2016), which has since been implemented to process the harmonic noise in the open-source code MRSMatlab (Müller-Petke *et al*. 2016).

Even though we will not address these aspects in this paper, it is important to note that Müller-Petke & Costabel (2014) proposed to use reference-based noise cancellation methods to remove the harmonic noise.Such a strategy may prove very efficient, but only at the condition that the correlation between the main coil and the reference coil is hig) Comparisonand optimal parameter settings of refernce based harmonic noise cancellation in time and frequency domains for surface-NMR. Neo deploy a reference coil. Also, it is worth mentioning that one potential difficulty with the harmonic noise removal arises when the Larmor frequency at the survey site gets close to one of the harmonic signals. In this case, the model-based subtraction might lead to a significant distortion of the SNMR signal. Fortunately, this phenomenon becomes critical only for frequency shifts of less than 1 or 2 Hz, and can be efficiently handled by extrapolation (Larsen *et al*. 2013), or with other techniques such as those proposed by Liu *et al*. (2018) or Wang *et al*. (2018), which can be easily adapted to the methods that we present within this paper.

1.2 Problem statement

Most of the methods presented previously and commonly used today to remove the harmonic EM noise are based on the classical assumptions that: 1) the harmonic noise is generated based on a single fundamental frequency value and 2) the fundamental frequency value is constant over the duration of the record (usually about 1 s in SNMR surveys). These assumptions are not necessarily true.

First, in some cases two (or more) harmonic signals with different fundamental frequencies can coexist due to the presence of several harmonic noise sources. Larsen & Behroozmand (2016) witnessed a very clear example of this possibility, when their data were contaminated by two harmonic series with two different fundamental frequencies of 50 Hz (origin from powerline) and 51 Hz (unknown origin), respectively. Fortunately, the two harmonic signals were clearly distinguishable and identifiable, and therefore easily removed one after the other using the classical model-based harmonic method described in Larsen *et al*. (2013). Yet, more complicated situations will arise when the two fundamental frequency values are closer to each other. For example, when both a railway electric line and a power network line are located close to the survey area. In the most European countries, the two harmonic signals produced by these two different noise sources will be both tuned to a nominal frequency of 50 Hz. In a few other European countries (Germany, Switzerland, Norway, Sweden and Austria), the two harmonic signals will be based on nominal fundamental frequencies of 16.66 and 50 Hz, respectively. Comparable situations will be encountered all around the world depending on the electrical grid specificities of the country of interest. The issues described below will also arise, for example, in the presence of two different powerlines close to the survey site (which would be both tuned to 50 Hz, in Europe), or any dual or multiple harmonic noise source group whose fundamental frequency values are identical or multiple of each other.

In all of these cases, the difficulty arises from the fact that the frequency content of the two harmonic groups will overlap, but not in a perfect way. That is, the two fundamental frequency values will independently deviate a few mHz from the nominal value (due to variations of supply and demand within the electrical grid), and the presence of two harmonic noise sources will therefore result in processing two harmonic signals with different (yet very close) fundamental frequencies. Recently, Wang *et al*. (2018) briefly addressed this issue in their study on the removal of co-frequency harmonics. However, it was only a minor aspect of their work about co-frequency harmonic removal, in which they presented a quite specific situation without addressing the general case. In this paper, we aim to investigate thoroughly the consequences of these situations on the harmonic processing accuracy and present in detail methods to efficiently handle these situations. In particular, we show with synthetic and field data that removing one harmonic signal after the other (sequential approach) using the classical model-based method with a 1-D line-search (Larsen *et al*. 2013) is likely to fail because of the mixing between the harmonic content of the two harmonic signals which deteriorates the frequency estimation accuracy. We demonstrate that the harmonic noise can be efficiently removed using the model-based approach and a 2-D optimized grid-search approach, where the optimal values of the two fundamental frequencies composing the two harmonic series are searched for simultaneously rather than sequentially. Such a method will increase substantially the processing time, which will still remain reasonable in the case of two fundamental frequencies (typically a few hours), but will become impracticable when three or more fundamental frequencies are present within the harmonic noise. Hence, for these more complicated cases we propose to adapt the frequency estimation strategy using other methods including nonlinear optimization methods.

The second objective of this study is to address the situation where the fundamental frequency cannot be considered stationary during the signal acquisition. Indeed, it is well known that the fundamental frequency of the harmonic noise slightly changes in time due to variations of supply and demand within the global electrical grid (Adams *et al*. 1982; Legchenko & Valla 2003). In SNMR surveys, the general assumption is that these variations are slowly changing within the duration of the signal acquisition, and that the harmonic noise can therefore be processed assuming a stationary fundamental frequency. However, the risk of this assumption being not valid is clearly acknowledged by several authors (Butler & Russell 2003; Jeng & Chen 2011). If the fundamental frequency changes too quickly, then the classical model-based approach will fail to remove the harmonic noise. Yet to our knowledge, no method has ever been proposed by the geophysical community to address this issue. In this paper, we propose to use what we will refer to as the *segmentation strategy*, which consists of dividing the acquisition signal in several equal parts of shorter duration and process each of these segments independently with the model-based approach. The idea behind this method being that the fundamental frequency can be considered stationary over these shorter periods of time. We will show that such an approach efficiently handles this kind of situation. However, applying the *segmentation strategy* using a 1-D line-search for the estimation of the fundamental frequency results in a significant increase of the computing time, since for each acquisition signal, the linear problem has to be solved as many times as the number of segments. To address this problem, we propose to use a faster frequency estimation method than the 1-D line-search, that we refer to as the Nyman, Gaiser and Saucier estimator (NGSE), and which we adapted to the SNMR context from the work by Saucier *et al*. (2006). All of these aspects will be investigated using synthetic and field data examples.

2 MATERIAL AND METHODS

2.1 Data acquisition and simulation

Data acquisition was performed using the SNMR instrument developed by Vista Clara, which is referred to in the following as the GMR device (Walsh *et al*. 2008). Different loop shapes (circular, square, 8-shape) and dimensions were used for each data set, and we will provide these characteristics progressively for each situation. The GMR allows recording 1 s duration signals, with a sampling frequency $f_s = 50$ kHz without any hardware-based bandpass filter centred at the Larmor frequency. This means that all the signals presented in this paper are raw signals, over which the only processing operations made are the ones presented here. Regarding the synthetic signals presented in this paper, they are all generated with a sampling frequency of 50 kHz and a 1 s duration, so as to match the field acquisition signals' characteristics.

2.2 Model-based harmonic processing

In classical noise conditions, that is, when the fundamental frequency of the harmonic noise is single and constant during the signal duration, the model-based approach gives very good results and is therefore largely used at this date within SNMR signal processing sequences. The method is based on subtracting from the acquisition signal an estimation of the harmonic noise components that is obtained through a nonlinear fitting process. The generally accepted model for harmonic interference is given by Larsen *et al*. (2013) as

$$h\ (n) = \sum_k A_k \cos\left(2\pi k \frac{f_0}{f_s} n + \phi_k\right), \tag{1}$$

where $h(n)$ denotes the harmonic model amplitude at time sample n, A_k and ϕ_k are the amplitude and phase of the kth harmonic component, f_s is the sampling frequency and f_0 is the fundamental frequency of the harmonic noise signal. The objective of the model-based approach is to determine the harmonic model that best reduces the energy of the acquisition signal. It is done by solving the following optimization problem:

$$\|S_{\mathrm{NMR}} - h_2\| \rightarrow \min, \tag{2}$$

where S_{NMR} denotes the acquisition signal and h the harmonic noise model. If the fundamental frequency f_0 of the harmonic noise is known, the problem can take a linear form and therefore be solved in a straightforward way. First, the cosine term of the harmonic model from eq. (1) has to be rewritten as

$$A_k \cos\left(2\pi k \frac{f_0}{f_s} n + \phi_k\right) = \alpha_k\ \cos\left(2\pi k \frac{f_0}{f_s} n\right) + \ \beta_k\ \sin\left(2\pi k \frac{f_0}{f_s} n\right), \tag{3}$$

where the variables α_k and β_k are related to the amplitude A_k and phase ϕ_k of each harmonic by the following relations:

$$A_k = \sqrt{{\alpha_k}^2 + {\beta_k}^2} \tag{4}$$

and

$$\tan\left(\phi_k\right) = -\frac{\beta_k}{\alpha_k}. \tag{5}$$

After this transformation, assuming that f_0 is known, the minimization problem from eq. (2) is linear and can be written in matrix notation as

$$\begin{bmatrix} \cos\left(k_0 2\pi \frac{f_0}{f_s} n_0\right) & \sin\left(k_0 2\pi \frac{f_0}{f_s} n_0\right) & \cdots & \cos\left(k_N 2\pi \frac{f_0}{f_s} n_0\right) & \sin\left(k_N 2\pi \frac{f_0}{f_s} n_0\right) \\ & \vdots & \ddots & \vdots & \\ \cos\left(k_0 2\pi \frac{f_0}{f_s} n_f\right) & \sin\left(k_0 2\pi \frac{f_0}{f_s} n_f\right) & \cdots & \cos\left(k_N 2\pi \frac{f_0}{f_s} n_f\right) & \sin\left(k_N 2\pi \frac{f_0}{f_s} n_f\right) \end{bmatrix} \begin{bmatrix} \alpha_{k_0} \\ \beta_{k_0} \\ \vdots \\ \alpha_{k_N} \\ \beta_{k_N} \end{bmatrix} = \begin{bmatrix} S_{\mathrm{NMR}}\ (n_0) \\ S_{\mathrm{NMR}}\ (n_1) \\ \vdots \\ S_{\mathrm{NMR}}\ (n_{f-1}) \\ S_{\mathrm{NMR}}\ (n_f) \end{bmatrix}, \tag{6}$$

where $[k_0,\ k_1 \ldots k_N]$ denotes the orders of the harmonics composing the model and $[n_0,\ n_1 \ldots n_f]$ are the time samples. This system can hence be solved by directly inverting the matrix from the left-hand side of the equation, after which the amplitude A_k and phase ϕ_k of each harmonic can be retrieved using eqs (4) and (5). In practice, however, the fundamental frequency f_0 is never precisely known, and solving for the best harmonic model h hence remains a nonlinear optimization problem, with f_0 being the only nonlinear parameter in eq. (3).

Indeed, the nominal value of f_0 for electrical network powerlines is typically 50 Hz in Europe, 60 Hz in North America and alternates between these two values on other continents. For some railway powerlines in Europe, f_0 has a nominal value of 16.66 Hz. Yet the true fundamental frequency value of these harmonic signals is never exactly the nominal value because slight fluctuations in time happen due to variations of power demand and supply. Adams *et al*. (1982) reported variations that can reach 0.03 Hz in a matter of minutes. Legchenko & Valla (2003) also report site-dependent instability of the 50 Hz powerline fundamental frequency in Europe, estimating variations up to 0.5 Hz between several records.

Thus, to solve this nonlinear optimization problem, Larsen *et al*. (2013) propose to determine f_0 by solving several times the linear problem while running through an optimized grid-search and to choose the frequency value that yields the best signal reduction. This approach gives very satisfying results in the presence of the harmonic noise composed by a single, constant fundamental frequency, and was therefore implemented in the open-source code MRSMatlab (Müller-Petke *et al*. 2016).

However, the computing time resulting from the use of a 1-D optimized grid-search is not insignificant. It depends of course on the size and on the density of the grid-search, which is ultimately a choice from the user. With a relatively dense grid and optimal parameters, the processing time will typically be on the order of 1–2 hr depending on the number of signals to be processed and the computational power available, which is still quite reasonable in post-processing operations. However, as introduced earlier, in this paper we will propose from one hand the use of a 2-D optimized grid-search to process harmonic noise produced by two harmonic noise sources, in which case the increase in processing time will need to be discussed. And from the other hand, we will propose a specific strategy called the *segmentation strategy* to adapt to the case where the fundamental frequency is non-stationary. This strategy necessitates to solve the linear problem many more times than in classical noise conditions, potentially resulting in a problematic increase of the computing time. To face this issue, we will introduce a faster frequency estimator than the 1-D optimized grid-search, which we present in the next section.

2.3 Nyman, Gaiser and Saucier estimator

2.3.1 Description

A faster method than a 1-D line-search to determine the fundamental frequency of the harmonic noise was proposed by Saucier *et al*. (2006), who used the Nyman and Gaiser estimation technique (NGE, Nyman & Gaiser 1983) in order to construct a single, unbiased, lower-variance frequency estimator. This method has also been used successfully for the harmonic noise removal in the SNMR context by Ghanati & Hafizi (2017). In essence, the NGE method consists of solving a four-equation system with four linearly dependent unknowns to determine the amplitude, phase and frequency values of a single harmonic within the global harmonic noise. Once these values have been determined, that is, when satisfying convergence has been reached, the associated single harmonic model is subtracted from the signal processed and the operation is repeated for each of the harmonics composing the harmonic noise. The idea behind the work of Saucier *et al*. (2006) is to use the whole collection of frequency estimates provided separately for each harmonic with the NGE method, in order to build a more precise estimation of the fundamental frequency. Once this estimation is achieved, the amplitude and phase of all the harmonics are deduced at a later step, by solving the linear problem, as described in the previous section (eq. 6). Saucier *et al*. (2006) present the detailed derivation of the NGE method, which was not provided in the original work from Nyman & Gaiser (1983), and develop the mathematical framework necessary to construct their single lower variance frequency estimator. They demonstrate using Monte-Carlo simulations and magneto-telluric field data sets that the method exhibits an accuracy comparable to the least-squares estimation method (LSE), which is equivalent to the model-based approach with optimized grid-search described in the previous section. They also show that the computing time of the method is reduced by a factor of at least 2 compared to the LSE method. Later in this paper, we will provide more details on our own observations which confirm the reduction in computing time and the equivalent accuracy of this method compared to the use of a 1-D line-search.

2.3.2 Implementation

In the following, we describe the equations that need to be implemented in order to apply the NGSE to an SNMR signal. Furthermore, we provide in Appendix A a Matlab script which computes these equations and outputs the estimated fundamental frequency value given an input signal. For more details about the detailed derivation and construction of the estimator, we refer the reader to Saucier *et al.* (2006).

Let k_0 and $k_{\max}$ be integers representing the order of the first and of the last harmonic used to build the collection of frequency estimates, respectively. We will see later that the choice of k_{max} must be done carefully to ensure the efficiency of the method. The first step is to compute the four estimation variables U_k, V_k, X_k and Y_k, for each of the harmonics ($k = k_0,\ k_0+1,\ \ldots,\ k_{\max}$), using

$$\begin{cases} U_k = \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} \cos(k\omega n\Delta t)s(n) & \text{(6a)} \\ V_k = \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} \sin(k\omega n\Delta t)s(n) & \text{(6b)} \\ X_k = \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} \cos(k\omega n\Delta t)\sin\left(\dfrac{2\pi n\Delta t}{T}\right)s(n) & \text{(6c)} \\ Y_k = \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} \sin(k\omega n\Delta t)\sin\left(\dfrac{2\pi n\Delta t}{T}\right)s(n), & \text{(6d)} \end{cases}$$

where N is the number of samples composing the signal to be processed, $s(n)$ is the signal amplitude at sample n, Δt is the time step between each sample, T denotes the total duration of the signal processed, and $\omega = 2\pi f_{\rm nom}$ is the initial angular frequency estimate, usually computed based on the nominal fundamental frequency value $f_{\rm nom}$ expected to compose the harmonic noise (e.g. 50, 60 or 16.66 Hz). The variables U_k and V_k are similar to Fourier coefficients and lead to the estimation of the harmonics' phase and amplitude values. Variables X_k and Y_k are primarily sensitive to the frequency shift between the reference value ω and what would be the 'true' fundamental frequency of the noise ω_0. These variables are actually built in such a way that they approach zero when ω gets close to ω_0 (among other conditions, see Saucier *et al.* 2006).

Based on these four estimation variables, a frequency shift estimator $\hat{\epsilon}_k$ is computed for each harmonic using

$$\hat{\epsilon}_k = \frac{2\pi}{kT}\frac{U_k Y_k - V_k X_k}{{U_k}^2 + {V_k}^2}. \tag{7}$$

Simultaneously, the amplitude of each harmonic is estimated following

$$\hat{A}_k = 2\sqrt{{U_k}^2 + {V_k}^2}. \tag{8}$$

At this point, we have a collection of frequency shift estimations and a collection of amplitudes. Following Kay (1989), Saucier *et al.* (2006) show that the best linear unbiased frequency shift estimator based on this collection can be computed using

$$\hat{\epsilon} = \frac{\sum_{k=1}^{k_{\max}} k^2 \hat{A}_k^2 \hat{\epsilon}_k}{\sum_{k=1}^{k_{\max}} k^2 \hat{A}_k^2}. \tag{9}$$

This frequency shift estimation is then used to correct the initial angular frequency value $\omega = \omega + \hat{\epsilon}$, and the whole process (eqs 6–9) is repeated iteratively until a satisfying level of convergence is reached, that is, when $\hat{\epsilon}$ becomes sufficiently close to 0. Once the fundamental frequency value has been estimated, the last step simply consists of solving the linear problem from eq. (2), and hence determining the phase and amplitude values for all harmonics, and finally subtracting the resulting harmonic model from the acquisition signal.

Although easily implemented, the effectiveness of the NGSE method is also subject to compliance with certain conditions of use, without which the final frequency estimation is likely to be erroneous. In short, those conditions specify that the value of $k_{\max}$, which represents the number of harmonics used to build the frequency estimate collection, should not exceed a certain threshold above which the frequency estimate will almost certainly be inaccurate, and therefore the harmonic removal inefficient. We do not address this issue in this section, but we provide a more detailed investigation of these aspects in Appendix B, where we also provide the optimal number of harmonics that would ensure safe use of the NGSE algorithm in SNMR surveying.

3 RESULTS

In this section, we present the results of our investigations regarding the existence of the harmonic noise with more complex frequency features than usually assumed, which make the classical harmonic removal approaches inefficient, and we propose new strategies to adapt to these difficult conditions. We first address the overlapping of two harmonic signals based on different fundamental frequency values, and then we consider the issue of a fundamental frequency whose value is varying quickly with respect to the record duration.

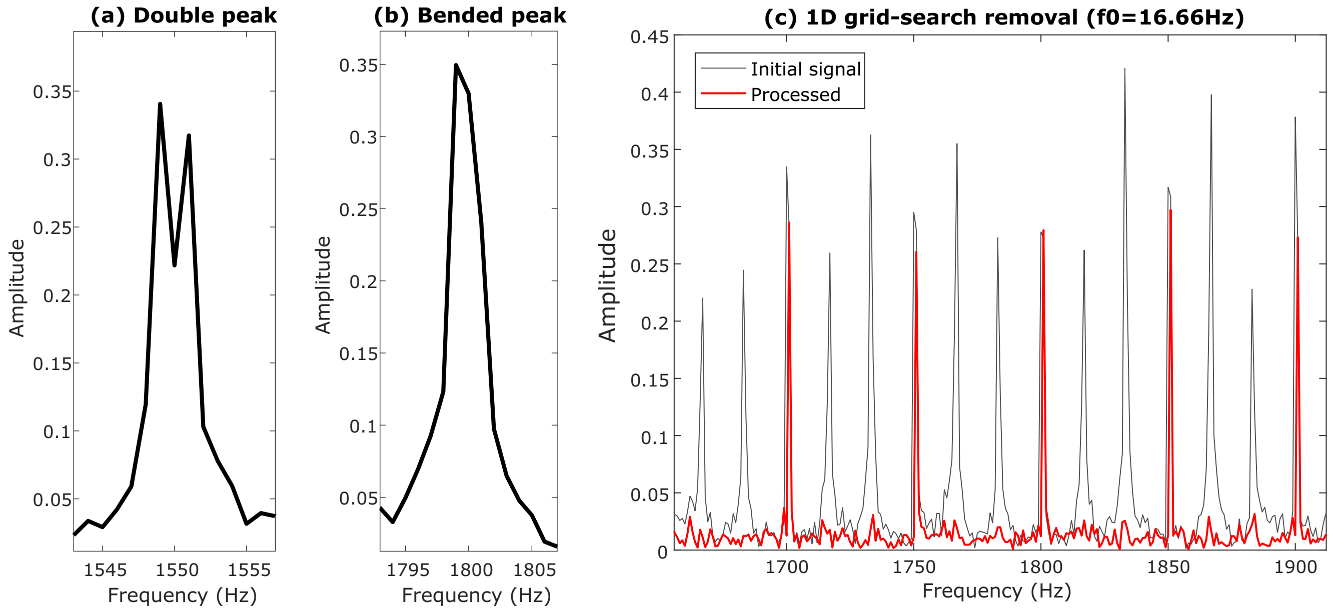

Figure 1. Synthetic harmonic noise data with random amplitudes and phases, and the Gaussian noise with zero mean and a standard deviation of 1 are shown in the Fourier domain to illustrate the issues raised by the presence of two different fundamental frequencies (16.66 and 50.02 Hz) within the harmonic noise. Panels (a) and (b) show 'double' or 'bended' frequency peaks that can appear at frequencies multiple of 50 Hz. Panel (c) shows the result of harmonic noise removal performed using a 1-D grid-search-model-based scheme, based on a nominal fundamental frequency of 16.66 Hz.

3.1 Harmonic noise with two different fundamental frequencies from different sources

Here, we investigate the situation where the harmonic noise is composed of two different fundamental frequencies, each originating from a different source. Such cases will be encountered increasingly frequently as the SNMR survey sites move closer to urban zones to address hydrogeological problematics in these areas, and due to the progressive densification of the electrical infrastructures network in the countryside. Many different situations can be encountered depending on the country where the survey takes place and on the type of electrical infrastructures producing the harmonic EM noise. Here we investigate the context of a survey located close to both a railway powerline and an electrical network powerline, which is a relatively common situation in Europe where the railway grid is dense.

3.1.1 Synthetic example

The main problem with double fundamental frequencies occurs when harmonics frequencies get close to one another but are not exactly the same, so that their frequency content overlaps but does not coincide exactly. In the case of a railway powerline and an electrical network powerline, the problem is plain: the railway fundamental frequency of 16.66 Hz was deliberately chosen to be one third of the electrical network frequency of 50 Hz. It follows that, from a Fourier space point of view, for each frequency that is a multiple of 50 Hz, the railway harmonic signal will overlap the electrical network harmonic signal. Obviously, if the nominal fundamental frequency of railway lines is 50 Hz instead of 16.66 Hz, the overlapping effect also occurs. In addition, this overlapping effect might not be 'perfect' since each harmonic signal will be slightly shifted from its nominal value due the varying nature of the fundamental frequency (caused by variations in power demand and supply), and because these variations are likely to be independent between each harmonic noise source due to their respective location within the electrical network grid and the varying load they might be subject to.

Visually, this will result in the appearance of 'double' or 'bended' frequency peaks in the frequency spectra, such as those illustrated by the synthetic data shown in Figs 1(a) and (b), respectively. These figures show the Fourier transform of a 1 s duration synthetic signal which was built as the sum of two harmonic noise models with respective fundamental frequencies of 16.66 and 50.02 Hz. For both models, the amplitude and phase of each individual harmonic are randomly selected within the range [0 0.5] and [0 π], respectively, and Gaussian noise with zero mean and standard deviation of 1 is added. Note that these 'bended' or 'double' frequency peaks are more likely to appear as the order of the individual harmonic increases, because the final frequency shift between two harmonics with different fundamental frequency increases for high order harmonics. In the present case, for example, the powerline fundamental frequency is shifted of 0.02 Hz from its nominal value (50.02 Hz) whereas the railway fundamental is exactly the nominal value (16.66 Hz). At 1500 Hz, which corresponds to the 30th powerline harmonic frequency, the frequency shift between the two signals is $\Delta f = 0.02$ Hz $\times$ 30 $= 0.6$ Hz whereas at 2500 Hz, the 50th harmonic, it becomes $\Delta f = 0.02$ Hz $\times$ 50 $= 1$ Hz.

However, the respective amplitudes of each single harmonic also influence the appearance of 'bended' or double frequency peaks. Similar amplitudes will more likely lead to double frequency shapes whereas significantly different amplitudes will rather result in bended

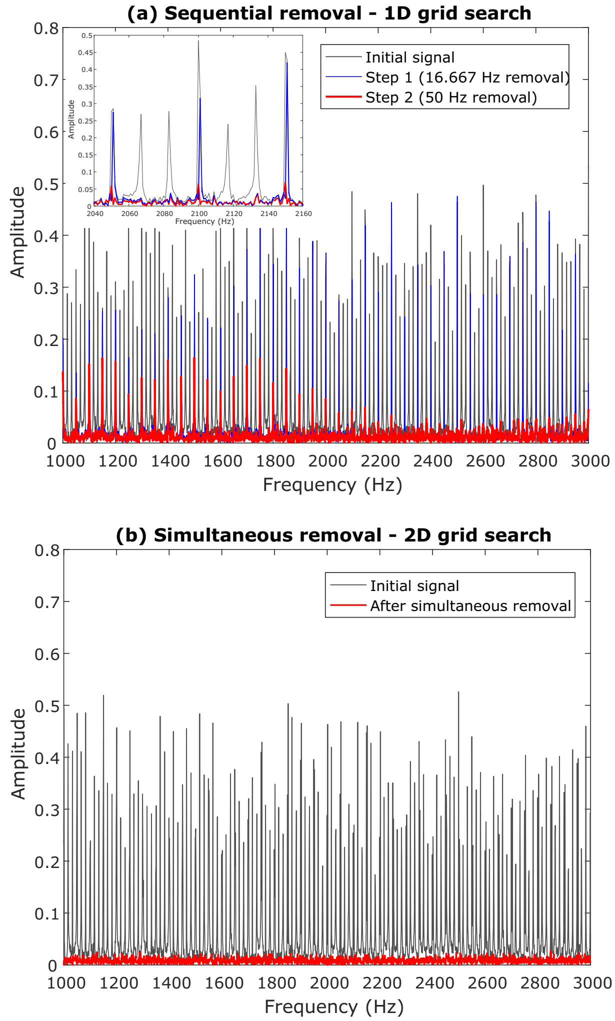

Figure 2. Processing of synthetic harmonic noise data composed of double fundamental frequencies (16.66 and 50.02 Hz). (a) Sequential processing in two subsequent steps: some harmonic content at frequencies multiple of 50 Hz remains in the low-frequency range of the graph. The graph on the upper left part is a zoom on the frequency range (2040–2160 Hz). (b) Simultaneous processing using a 2-D grid-search: all the harmonics are perfectly removed.

peaks or even single peaks undistinguishable from single frequency harmonic signals. This explains why the bended peak shown in Fig. 1(b) occurs at a higher frequency than the double peak from Fig. 1(a).

Fig. 1(c) shows the results of applying the model-based method with a 1-D grid-search on the synthetic noise signal, assuming a nominal fundamental frequency of 16.66 Hz. The process estimates precisely the true fundamental frequency (16.66 Hz), and it seems that the frequency peaks associated with this fundamental frequency are perfectly removed from the initial signal. However, some peaks are still visible at the frequencies multiple of 50 Hz, indicating that a removal process based on a single fundamental frequency estimation fails to suppress the harmonic noise composed of two different fundamental frequencies.

A logical attempt in the present situation could be to adopt a sequential two-step approach, by first removing the harmonic noise originating from the railway line (frequency search and model based on 16.66 Hz nominal value) and subsequently apply the same process in order to remove the harmonics originating from the powerline (based on 50 Hz nominal value). Unfortunately such an approach is very likely to fail, as illustrated by the example in Fig. 2(a), which shows the results of a sequential removal applied to a synthetic signal produced following the same criteria as previously (same frequency values, but different harmonic amplitudes and phase values since they are randomly chosen). As observed previously, even though the first step removes a significant part of the harmonic noise because it cancels optimally 66 per cent of the 16.66 Hz based signal, harmonics multiple of 50 Hz are still present. Subsequent removal based on 50 Hz fundamental frequency (step 2) globally improves the harmonic noise reduction, although significant harmonic peaks remain. Note that the final amplitude of the remaining harmonics as well as their position in the frequency spectra are strongly dependent on the value of the fundamental frequency shift between the two different harmonic signals (in this case 0.02 Hz), and are not easily predictable. We ran many different simulations (not shown here) among which the case presented in Fig. 2(a) appears as a quite favourable situation for sequential removal. In most cases the remaining harmonic noise after step 2 is of greater magnitude than this. Note also that the order in which the sequential removal is performed (16.66 Hz then 50 Hz or vice versa) has globally no effect on the method effectiveness. Again, we ran many simulations with different frequency shifts,

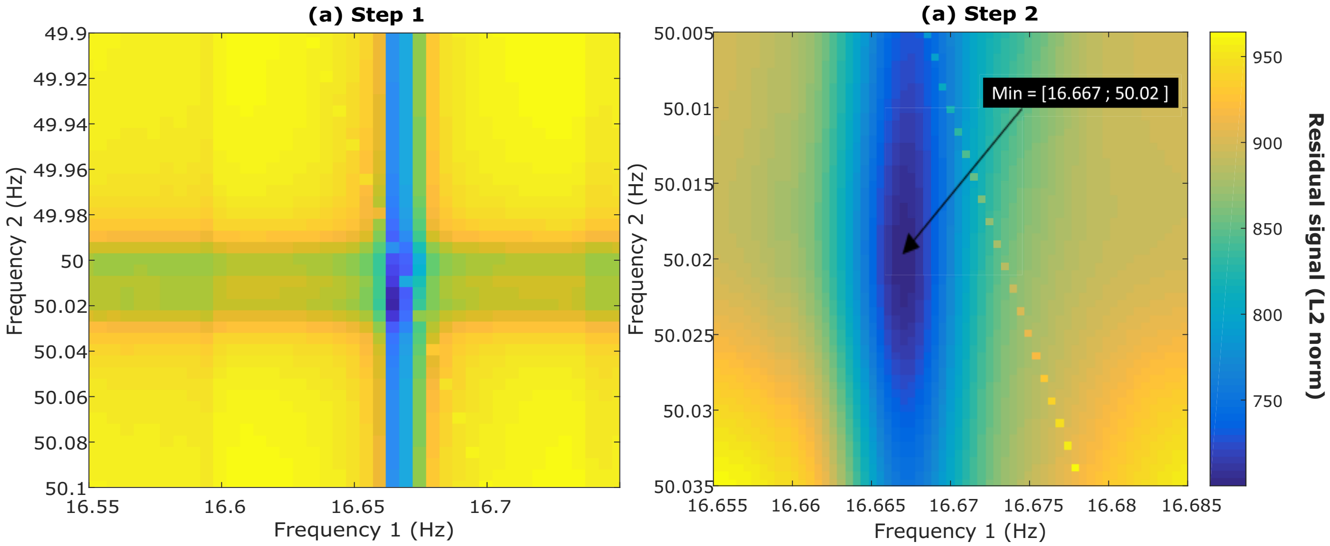

Figure 3. Residual signal matrix obtained after solving the model-based linear problem while travelling through the 2-D frequency space. (a) Residual signal distribution during the first step of the 2-D optimized grid-search (wide grid and coarse frequency step). (b) Residual signal distribution during the second step of the 2-D optimized grid-search (small grid and fine frequency step). The minimum residual signal yields the frequency pair used for final model subtraction. Note that for visualization purposes, these two grid searches have been deliberately sampled more densely than required for solving the problem. The 'dotted' diagonal in each matrix corresponds to the cases where the railway frequency is exactly one-third of the powerline frequency value.

amplitude and phase values (not shown here), and could not observe any clear superiority from one sequence order over the other. Both were alternatively more efficient than the other or of equivalent efficiency, and both where only rarely efficient. Those issues are discussed more in detail later in the paper.

To solve the issue of harmonic contamination composed of two fundamental frequencies, a straightforward and efficient method is to proceed to a brute force simultaneous search within the whole 2-D frequency space (Fig. 2b). This is simply an extension of the 1-D grid-search (or line-search) concept proposed by Larsen *et al.* (2013) for one fundamental frequency to the case where two frequencies need be estimated simultaneously. In this case, the harmonic model is written as

$$h\ (n) = \sum_k A_k \cos\left(2\pi k \frac{f_1}{f_s} n + \phi_k\right) + \sum_k B_k \cos\left(2\pi k \frac{f_2}{f_s} n + \varphi_k\right), \tag{10}$$

where f_1 and f_2 are the two different fundamental frequencies to be estimated, A_k and B_k are the amplitudes of the kth harmonic for each fundamental frequency, and ϕ_k and φ_k are the phases of the kth harmonic for each fundamental frequency. The cosine terms are transformed again following eq. (3), after what the problem becomes linear and can be written in matrix notation in a similar way as eq. (4).

Then, we can solve many times the linear problem while going through an optimized 2-D grid-search in which each dimension corresponds to one of the two nominal fundamental frequency values. As for the 1-D line-search, the search domain is explored in two steps. First a relatively wide and coarse grid is covered, computing each time the residual signal after model subtraction (Fig. 3a). Then, starting from the minimum obtained during this first step, a thinner and finer grid is explored, and the minimum residual signal obtained through this second step yields the best pair of frequencies estimations (Fig. 3b). These two frequency values are then used for the final harmonic subtraction. Fig. 2(b) shows how this approach efficiently removes all the harmonic components originating from each harmonic noise source. After harmonic removal, the standard deviation of the residual signal yields 1.078. Given the fact that the initial standard deviation value of the Gaussian noise was 1, the residual signal can be considered as composed primarily of the Gaussian noise, confirming that all the harmonic content has been removed. In the following we will refer to this approach as the 'simultaneous removal' or will mention the use of a '2-D grid-search'.

Note that the shape of residual signal matrix shown in Fig. 3(b) is globally convex, which should support the use of nonlinear optimization techniques such as conjugate-gradient or steepest-descent methods. However, the residual matrix from Fig. 3(a) does not seem so well adapted to these methods. Later in the paper we will discuss the possibility of using nonlinear optimization techniques to solve this kind of problems, and the prospective benefits and disadvantages of such approaches.

3.1.2 Field example

The sequential and simultaneous approach were both applied to real noise signals recorded in areas where both railway and network powerlines are present and likely to contaminate the records. Noise data from Fig. 4(a) was acquired at a well characterized hydro-geophysical test site located in the municipality of Schillerslage, close to Hannover in Germany (see Dlugosch *et al.* 2014 for more details about the local hydrogeology), using a 50 m diameter circular loop. This location is particularly suited to our problem because the global EM noise level

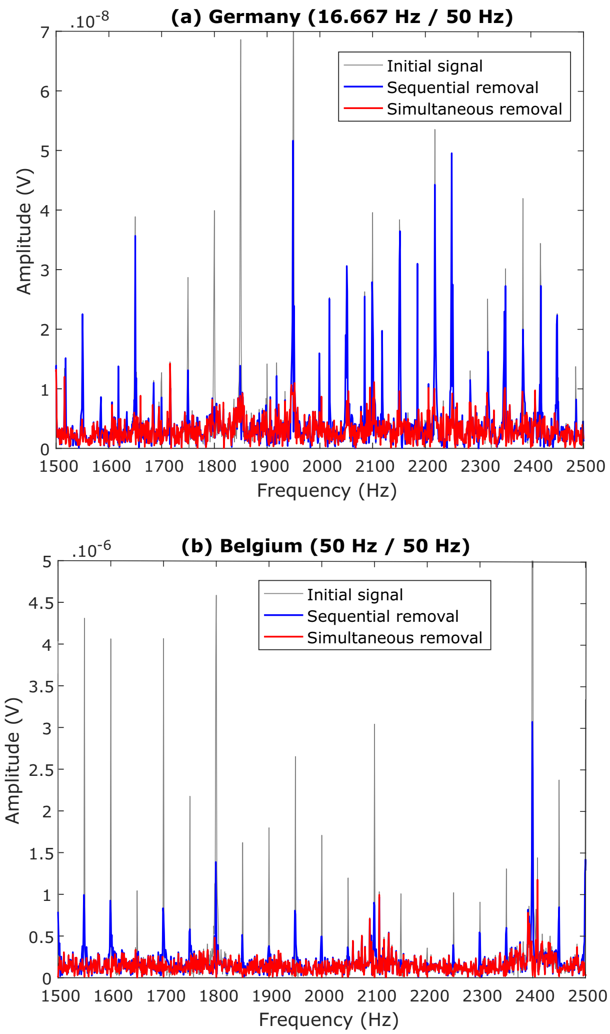

Figure 4. Processing of field data contaminated by harmonic noise with two different fundamental frequencies due to the presence of both a railway and an electrical network powerline close to the survey site. a) Data from the Schillerslage test site in Germany, contaminated by both 16.66 and 50 Hz harmonics. b) Data set from the Colonster Castle in Belgium, contaminated by two different 50 Hz harmonic signals. In each case, a sequential removal fails to suppress the harmonic noise whereas a simultaneous removal using a 2-D grid-search is much more efficient.

is remarkably low there, which allows to capture clearly harmonic noise emanating from a railway line located about 3 km away (16.66 Hz based harmonics) as well as some 50 Hz based harmonic signal likely to be originating from the low voltage powerline linking the two closest villages (about 1 km). The harmonics from each contamination source are clearly visible in Fig. 4(a), and some double frequency peaks can be observed with a closer look at the graph. The processing of this noise data shows that a sequential removal of first the 16.66 Hz harmonics followed by the 50 Hz harmonics fails to suppress the harmonic noise. Note that processing following the other sequence order (16.66 Hz then 50 Hz) gives similar results (not shown here). However, the simultaneous approach with a 2-D grid-search is much more efficient and eliminates almost completely all the frequency peaks. The fundamental frequency values estimated using the 2-D grid-search yield 49.99 and 16.679 Hz for the network powerline and the railway line, respectively.

The second data set presented in Fig. 4(b) was recorded using a 20 m edge square loop displayed at the Colonster Castle, located in a peri-urban area close to the city of Liege in Belgium. A high voltage powerline is located about 1 km from the area and a railway line about 500 m. In Belgium, both of these EM noise sources produce harmonic noise tuned at a fundamental frequency of 50 Hz. However, as explained previously, each of these electrical lines has a specific location within the global electrical grid, which makes them likely to have their true fundamental frequency vary independently from each other and hence compose a global harmonic signal with two different fundamental frequencies. The noise processing operations tend to confirm this hypothesis. A single frequency estimation (classical method, not shown here) failed to successfully remove the harmonic noise, and so did a sequential approach (Fig. 4b). However, searching simultaneously for the two fundamental frequency values again produced much better results. In this case, the two estimated frequencies were 49.986 and 49.997 Hz. These two values may seem very close to each other (11 mHz difference), yet the difference is much greater than the grid-search resolution of 1 mHz. Then, note that this data set was also subjected to the processing method presented in the following section to remove noise with a time-varying single fundamental frequency. This approach also failed to remove fully the harmonic content, hence confirming the hypothesis

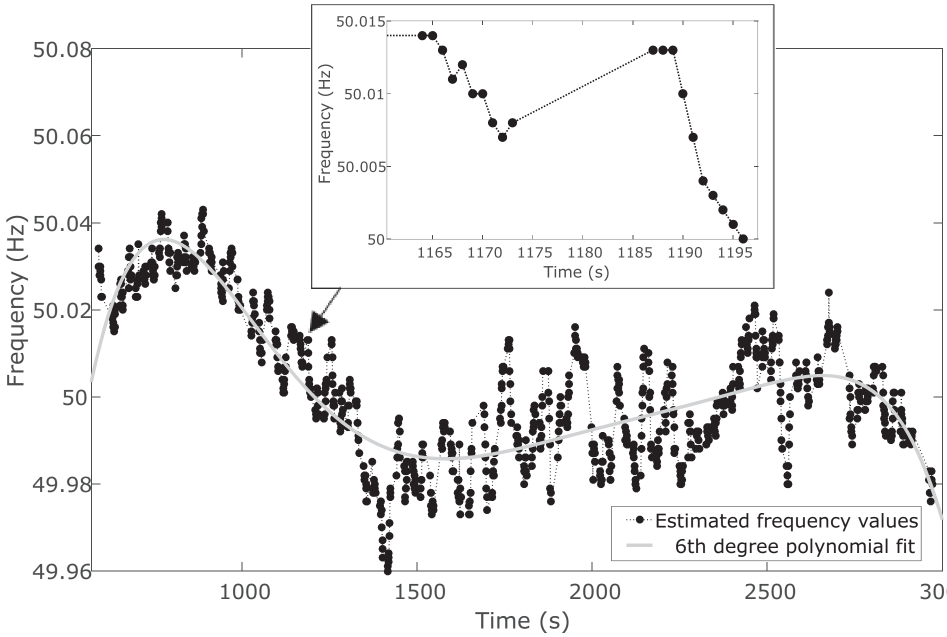

Figure 5. Temporal evolution of the fundamental frequency estimation performed with the NGSE method on 960 noise-only signals recorded quasi-continuously. The sixth degree polynomial fit underlines a 'large-period' variation pattern and the black rectangle offers a zoom on short-period variation patterns. The precision of the estimation for each record was confirmed through visual observation of the harmonic removal efficiency.

that the case presented in Fig. 4(b) corresponds indeed to two 50 Hz based harmonic signals with different fundamental frequencies, most likely produced independently by the railway and the network powerlines in the vicinity of the area.

These results highlight the fact that multiple fundamental frequencies can coexist in the harmonic EM noise recorded during geophysical data acquisition, and that specific methods such as the 2-D grid-search simultaneous removal may be needed to successfully remove the harmonic components originating from different noise sources.

3.2 Variations of the fundamental frequency

3.2.1 Long-period and short-period variations

As discussed and observed throughout the paper, the fundamental frequency value is varying around its nominal value, up to deviations of 0.03 Hz according to Adams *et al*. (1982), and possibly to even more important deviations according to our own observations (Fig. 5) and those of Legchenko & Valla (2003). Fig. 5 shows the results of estimating the fundamental frequency of powerline harmonic noise on 960 noise-only signals, which were recorded successively at a location close to the University of Liege, using a 20 m edge square-loop. The harmonic noise source in this case is a high voltage powerline located about 800 m from the acquisition area, and each signal was processed independently using the fast NGSE method presented earlier. The accuracy of the frequency estimation was confirmed visually for each record by analysing the results of harmonic subtraction based on these frequency estimates. The evolution curve shows that deviations from the nominal value as high as 0.04 Hz and above are possible, and should be taken in account in harmonic processing strategies. The observed pattern also suggest that the harmonic noise is a function of at least two temporally changing sources. One source with a slow variation pattern that could be symbolized by the grey sixth degree polynomial curve in Fig. 5, while much faster variations are occurring from one record to another, as illustrated on the black box from Fig. 5, where we observe variations of 0.014 Hz in less than 10 s.

Fig. 6 presents the evolution of the fundamental frequency estimations associated with noise only records, obtained this time with the grid-search approach. These signals were recorded at the Prakla site, close to the Leibniz Institute of Applied Geophysics (Hannover, Germany), using a 20 m edge eight-shape square loop. Although composed of less noise signals, Fig. 6 also illustrates how the fundamental frequency deviates significantly from its nominal value in very short time periods, showing deviations up to 0.04 Hz away from the nominal value, and a total variation of 0.07 Hz in less than 90 s. Overall, Figs 5 and 6 demonstrate how quickly the fundamental frequency of harmonic noise changes with time, in an unpredictable way.

Hence, based on these observations we would like to emphasize that the estimation of a maximum deviation of 0.03 Hz, provided by Adams *et al*. (1982), and very generally cited as a reference limit (e.g. Butler & Russel 1993, 2003; Saucier *et al*. 2006), may not be valid in many situations. This might be either because we are working in Europe instead of North America were Adams *et al*. (1982) initially assessed this limit, but more likely because the electrical network as a whole has become increasingly dense and complex since the 80 s. Either way, safe harmonic processing should, in our opinion, consider a possible deviation of at least 0.05 Hz.

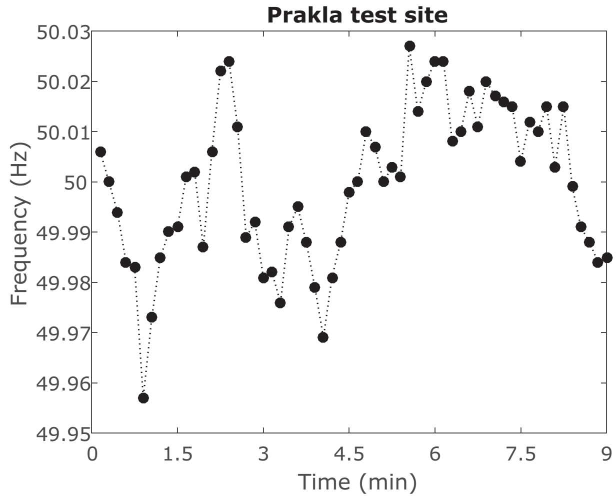

Figure 6. Temporal evolution of the fundamental frequency estimation performed with the grid-search method at the Prakla test site (60 records). The precision of the estimation for each record was confirmed through visual observation of the harmonic removal efficiency.

3.2.2 Segmentation strategy

Even though the varying nature of the fundamental frequency is well known and accepted, the harmonic subtraction strategies are nevertheless always based on the assumption that the fundamental frequency can be considered constant during the duration of the geophysical record (generally about 1 s in SNMR surveys). If this condition is not fulfilled however, the classic harmonic cancelling techniques based on single frequency estimations are likely to fail. To solve this issue we propose to use a so-called *segmentation strategy* that consists of dividing the signal in N equal parts, and apply classical harmonic cancellation techniques independently to each sample, so as to reconstruct the signal afterwards. The idea behind this strategy being that the constant frequency assumption is valid for these shorter duration signal segments. This method is illustrated in Fig. 7, where a harmonic noise signal recorded at the Larzac plateau, in France, is represented in the time domain before and after processing using the *segmentation strategy*. In this case the signal of interest was divided in 9 equal parts, and each of them was processed independently using the NSGE method for frequency detection. The processed signal reconstruction is done by simple concatenation. Even in the time-domain one can already observe that harmonic structures are removed from the initial signal. The lower part of the figure indicates the fundamental frequency values obtained for each segment, which do not seem to follow any specific trend.

We also want to emphasize that, even though this paper and this section are focused on the effect of fast frequency variations, the stability of the phase and amplitude values of the harmonic signals are also of importance for the effective processing of harmonic noise. If they change quickly with time (which they can), the effectiveness of the model-based approach is also likely to fail. In this paper we always consider them constant during the acquisition signal duration, and the synthetic signals presented in the following assume constant phase and amplitudes. However, note that since the *segmentation strategy* consists of processing each signal portion separately, it provides an independent estimation of the frequency, but also of phase and amplitudes values, so that variations of these parameters are also by default taken in account within this processing strategy.

3.2.3 Synthetic example

To investigate this approach we generate a synthetic harmonic noise signal as done in Section 3.1 (random phases within $[0\ \pi]$, amplitudes within [0 0.5] and same Gaussian noise), except that in this case the fundamental frequency is varying with time. The variation pattern is represented in Fig. 8(a) along with the exact governing equation, and consists of a quadratic evolution obtained with the *chirp* function from Matlab. Note that such a frequency variation pattern is not meant to be fully realistic, neither in terms of amplitude nor shape, since we have no information regarding what could be a realistic variation pattern (quick jumps, linear, quadratic, sinusoidal evolution. . . ?). Our primary goal here is in fact to put to the test the *segmentation strategy* on a synthetic signal with controlled frequency variation. A synthetic NMR signal was also added so as to represent a realistic geophysical record. It was generated at a Larmor frequency of 2015 Hz, with an initial amplitude set to 1, a relaxation time constant of 200 ms and zero phase.

Fig. 8 shows the results of applying the *segmentation strategy* to this 1 s duration synthetic record, with $N = 10$ segments, using both the Grid-search approach and the NGSE method for frequency estimation. A number of $N = 10$ segments consists of processing 10 independent short signals, which are re-assembled together afterwards. Fig. 8(b) illustrates how the segmentation strategy improves the removal of harmonic noise compared to a classical approach where the fundamental frequency is considered constant. With the classical approach, harmonics of about the same amplitude as the NMR signal are still present, whereas with the segmentation strategy the harmonics are almost completely removed, except for very small peaks still present at higher frequencies. Fig. 8(b) also demonstrates that the use of the segmentation strategy does not affect the SNMR signal. In the frequency domain, the associated frequency peak at the Larmor frequency

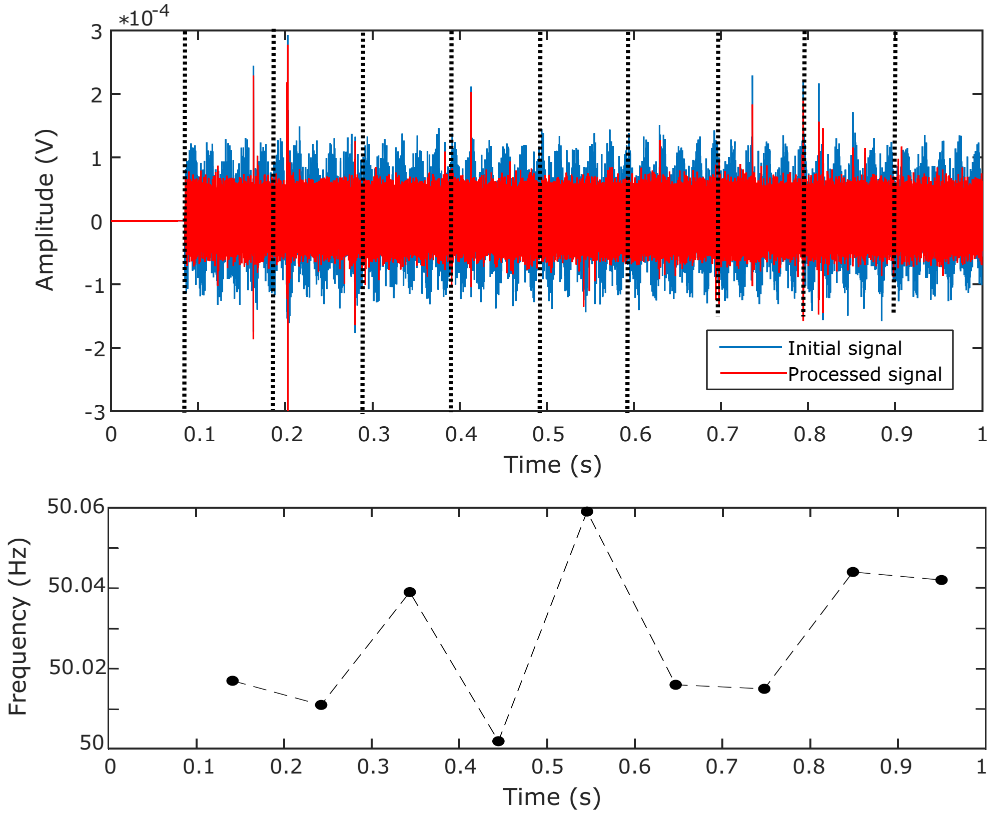

Figure 7. Illustration of the signal segmentation method. Top: the initial signal (blue) is divided in nine equal parts, which are all processed independently, and then concatenated together to compose the final processed signal (red). Bottom: the fundamental frequency estimated for each segment is plotted against the position in time of each segment.

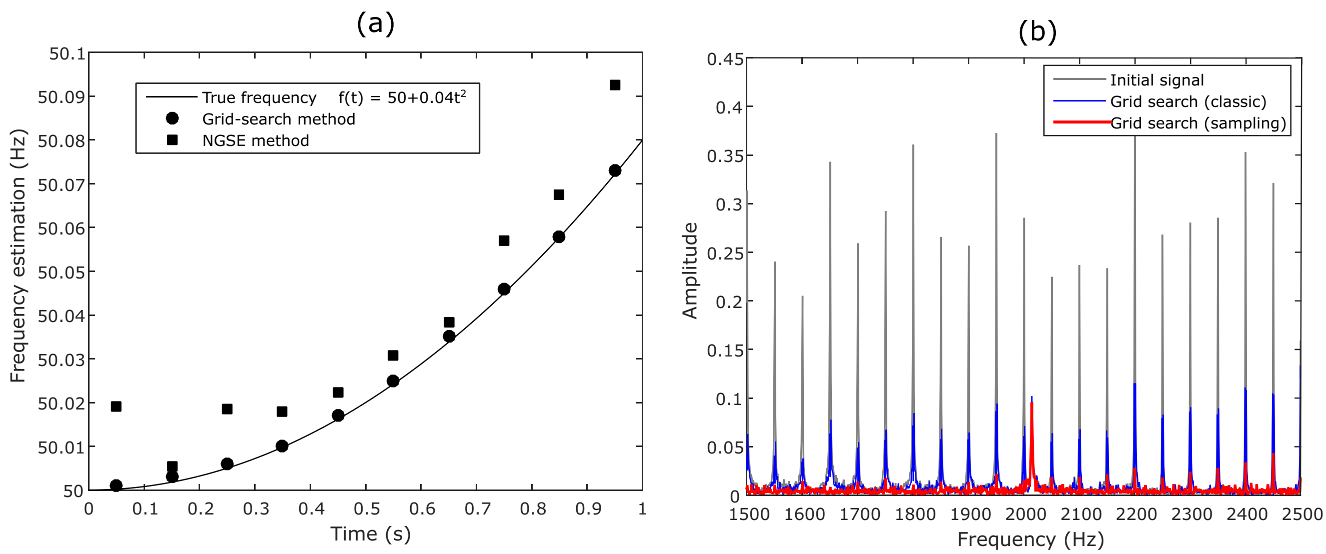

Figure 8. Application of the segmentation strategy with 10 segments to a synthetic NMR signal contaminated with synthetic harmonic EM noise. (a) Simulated frequency evolution following a quadratic shape, compared to the frequency estimation obtained for each sample, with the grid-search and the NGSE methods. (b) Frequency spectra of the initial synthetic signal, after processing using classical methods (i.e. assuming a constant frequency) and after processing using the segmentation strategy.

is still present and is well preserved. In the time-domain, a mono-exponential fit on the remaining signal yields an amplitude of 1 and a relaxation time constant of 204 ms, well in accordance with the initial parameters of the synthetic SNMR signal.

We see in Fig. 8(a) that the frequency values estimated for each sample using the grid-search approach follow closely the true frequency variation curve. With the NGSE method the estimated frequency values are globally higher but still follow the quadratic shape of the true evolution. Note that despite these differences, using the NGSE method yields in the end equivalent results and efficacy than using the

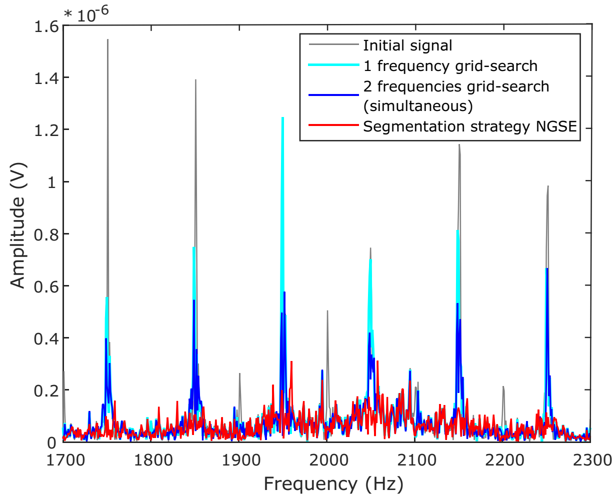

Figure 9. Results of different harmonic processing strategies applied to a noise-only signal recorded at Prieuré-du-Luc, France. The segmentation strategy is the only one able to successfully remove the harmonic noise, indicating that the poor efficiency of other methods are likely due to a quickly varying fundamental frequency composing the harmonic EM noise.

grid-search method, which is why only one of them was represented in Fig. 8(b) for visualization comfort. This indicates that even though the grid–search estimation is by nature the one that best minimizes the residual signal, the NGSE method still provides a frequency value that yields an equivalent minimization level. Thus, the computing time reduction provided by the NGSE method takes much value and interest in the frame of the *segmentation strategy*, because since a single signal requires as many frequency estimations as segments to be processed, the use of the Grid-search approach can lead to unreasonable processing duration when dealing with large data sets.

3.2.4 Field example—Larzac (France)

A probative data set regarding the issue of harmonic noise with varying fundamental frequency was obtained in the area of the Larzac plateau, a very remote area in France, at Prieuré du Luc. Several geophysical studies have taken place at this site to study the recharge dynamics of the unsaturated zone in karstic areas (e.g. Fores *et al*. 2016). A surface NMR survey whose results are still being processed was performed there to attempt quantifying the water content of the vadose zone at this site. An interesting characteristic of this site regarding the harmonic noise is that there are likely two possible sources of harmonic EM noise, and only two, due to the remote nature of the area. They consist of two electric powerlines, not connected locally, with different voltages (most likely medium and high), passing about 500 m and 1 km away from the survey site, respectively. Beyond this, there is no other electrical infrastructure in the area, and the closest village, located about 5 km away, is Campestre-et-Luc (107 inhabitants). Hence, the harmonic noise at this site is very likely to be composed of maximum two harmonic signals originating from the two powerlines, if not only one. However, Fig. 9 shows the results of different harmonic processing strategies applied to some noise-only signals recorded with a 28 m diameter circular figure-eight shape loop. Harmonics of the 50 Hz fundamental frequency are clearly visible on the raw initial signal. Applying the grid-search approach searching for one frequency fails to remove the harmonic peaks, and surprisingly, applying a 2-D simultaneous grid-search approach also fails. However, when applying the segmentation strategy, all harmonic peaks are correctly suppressed, indicating that the difficulty with this harmonic noise is not that it is contaminated by two different harmonic signals generated by the two sources, but more likely that the fundamental frequency is varying quickly within the duration of the recorded signal. These results also support the preliminary conclusion from the synthetic study, showing that the *segmentation strategy* appears to be suited to the processing of such harmonic noise with non-stationary fundamental frequency.

4 DISCUSSION

In this section we will discuss different aspects relative to the multiple and the non-stationary frequencies situations. Before doing so we would like to point out that, although we will discuss them separately here, these two situations are not necessarily exclusive and may very well occur at the same time, or obviously not occur at all. There is no direct way to identify which situation is dominant without proceeding directly to harmonic removal and assessing the efficiency of the processing. Typically we propose in Appendix D a very simple decision tree that should help the reader making decision about which method should be used and in which situation.

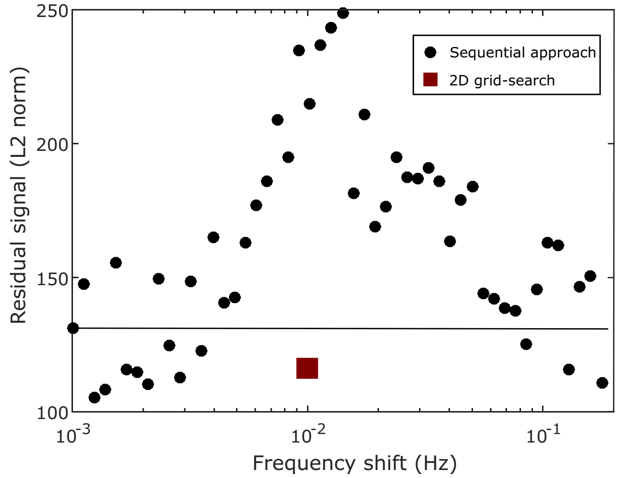

Figure 10. Impact of frequency shift upon the efficiency of the harmonic noise sequential processing in the presence of two fundamental frequencies. The red square indicates the residual signal level obtained using a 2-D grid-search when the frequency shift is 0.1 Hz, as an indicative result (the 2-D grid-search approach is always efficient, whatever the frequency shift is). Note that below a level of 130, indicated by the black line, the harmonic processing can be considered as successful (subjective threshold based on visual inspection of the frequency spectra after processing).

4.1 Multiple fundamental frequencies

4.1.1 Sequential removal versus 2-D grid-search

In Section 3.1, we addressed the issue of harmonic noise composed by two signals with different fundamental frequencies, through synthetic and field studies. We showed that such situations may lead to the appearance of bended or double peaks in the frequency spectra, and that classical processing methods will fail to remove successfully all the harmonic content. In particular, we demonstrated that even a sequential removal with a classical 1-D grid-search is likely to be unsuccessful whereas a 2-D grid-search is an optimal way of processing such kind of harmonic noise. Overall, we will always recommend using such processing schemes searching simultaneously for the two fundamental frequencies values. Yet it is worth pointing out that the sequential removal of two different frequencies can prove efficient, but only if the frequency shift Δf_0 between the two values is sufficiently high or sufficiently low. For example Larsen & Behroozmand (2016) processed successfully harmonic noise with double fundamental frequencies of 50 and 51 Hz, respectively. This is a relatively extreme frequency shift example though, and the authors were not able to identify the origin of the 51 Hz harmonic noise. Fig. 10 illustrates well how the frequency shift value affects the sequential removal efficiency. It shows the evolution of the residual signal level after harmonic removal performed on a synthetic harmonic signal composed of two fundamental frequencies with increasing difference. We see that the lowest residual values are obtained either for low-frequency shifts or high-frequency shifts, and that a progressive increase and decrease can be observed for intermediate values. Visual observations of each of these processing suggest that the harmonic noise can be considered as successfully removed when the residual signal gets below 130 (in this example). Above that level, some frequency peaks can still be distinguished within the frequency spectra. Hence, the results from Fig. 10 show that even though sequential removal is globally more efficient for low- and high-frequency shifts, the process remains highly nonlinear and unpredictable and might be ineffective even for low or high values, which is why we recommend to use a 2-D grid-search approach instead of a sequential one. If a sequential removal approach is nevertheless chosen, maybe for computing time considerations, we advise to check visually the processing results of each signal.

4.1.2 Speeding up the 2-D grid-search algorithm

We showed that the 2-D grid-search approach should be preferred for the processing of double fundamental frequency harmonic noise. It ensures the most precise frequency estimations and the most efficient reduction of the harmonic content. However, the main flaw of this relatively ‘brute-force’ approach is probably the significant increase of computing time that it requires compared to a sequential removal. Obviously, the magnitude of this increase will depend on the computational power available. As an example though, with our computational capabilities, an SNMR data set that would take about 1 hr to process with classical (1 frequency) model-based method, would require about 10 hr to be processed using a 2-D grid-search approach. Note that this could still be considered acceptable in the context of post-acquisition operations, where the time constraints are reduced compared to, for example, in the field. Yet we reckon that reducing this computing time would be a valuable improvement to this method.

To do so, it should first be noted that the 2-D grid-search problem is suitable to parallelization, if available. Two ways of handling it are possible. Several different stacks can be processed in parallel since they are independent, or, for a single stack, several areas of the search domain can be explored in parallel, since each iteration of the linear problem is independent of the others. Therefore, if a large number of processors is available, a significant reduction of the computing time can be achieved through parallelization.

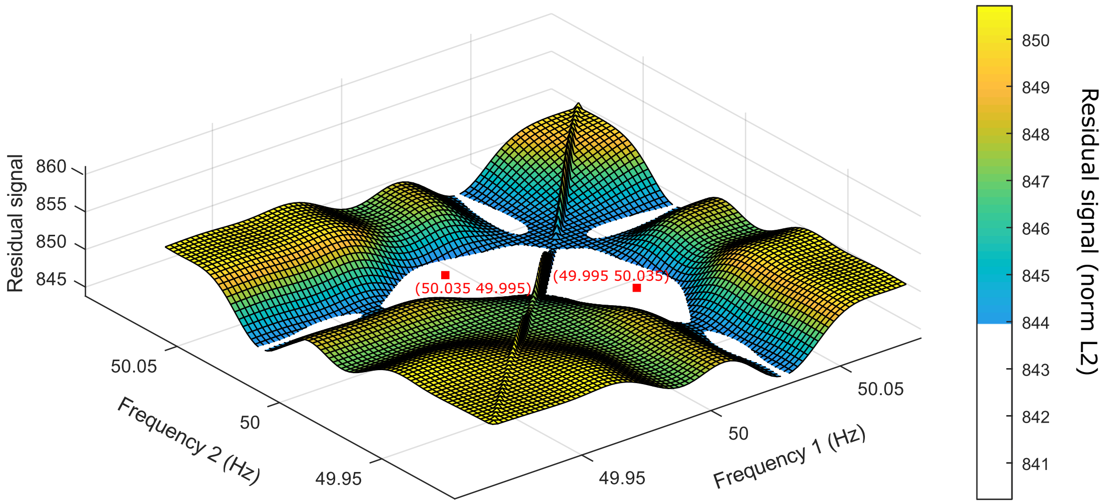

Figure 11. 3-D visualization of the residual signal matrix obtained through the first step of the optimized grid-search process. A threshold has been imposed to the colour bar to emphasize the presence of local minima on the external parts of the search domain. The red dots indicate the projection on the threshold plane of the location of the global minima corresponding to the exact frequency estimations.

Second, looking back at Fig. 3(b), we already mentioned that the residual signal matrix obtained during the second step of the optimized 2-D grid-search is largely convex, and seems therefore well suited to the use of nonlinear optimization techniques. Indeed a global minima looks easily and quickly reachable. However, it should be noted that this will usually not be true for the first step of the optimized 2-D grid-search. Either the convexity of the problem will not be obvious, as for example in the case of Fig. 3(a), where we see large zones of the search-domain where the gradient of the residual matrix is really weak, and not necessarily directed toward the global minimum. This may complicate the use and parametrization of descent methods, eventually leading to wrong frequency estimations. Similarly, Fig. 11 shows the residual signal matrix obtained after applying the first step of the 2-D grid-search resolution, on a synthetic noise signal composed of two fundamental frequencies (49.995 and 50.035 Hz), random amplitudes between 0.5 and 1 and null phases. We mark with red dots two minima that will eventually lead to correct frequency estimations, but we also show that reaching them could prove difficult depending on the starting point due to the presence of other local minima (emphasized by a colour threshold) and areas with very low gradient. This illustrates that if nonlinear optimization techniques are to be used, it should probably be done when the search domain has been sufficiently narrowed to ensure the existence of an easily reachable global minimum, for example, after going through the first step of the optimized grid-search, which would in the end not result in a great benefit in terms of computing time. If applied before the first step, on a wide search domain, the choice of the nonlinear optimization technique and its parametrization should be done very carefully. Some methods such as the neighbourhood algorithm (Sambridge 1999) or the CLM method (Badsar *et al*. 2007) may prove efficient, but this issue would require extended work to be fully investigated, and falls beyond the scope of this paper.

A probably simpler method to speed-up the 2-D grid-search could also be to trade accuracy for speed, for example by modelling only a very low amount of harmonics. Indeed, Larsen *et al*. (2013) rightly point out that in the model-based approach, a certain amount of harmonics should be used within the harmonic model to ensure capturing harmonics from the processed signal with significant energy. Similarly, we show in Appendix B that proceeding to the model-based approach using a very low number of harmonic (<10) can lead to a slight degradation of the results. In this case, the difference is small and would probably not impact the retrieval of an SNMR signal, however the magnitude of this deterioration can be more important if the harmonics used for modelling are absent from the processed signal. Nevertheless, if the presence of 1 or 2 harmonics can be assured, for example using a fast threshold detection algorithm (Jiang *et al*. 2011), then proceeding to the model-based approach using only few harmonics yields a very important reduction of the computing time. For example, with an Intel Core i3-7100U processor at 2.40 GHz, a synthetic data set composed of 50 stacks with 20 pulse moments, that is, 1000 signals, will be processed in about 10 hr when modelling 20 harmonics. However, using the same grid-search density but only 2 harmonics, the processing time goes down to 32 min, yielding equivalent results. Hence, if facing strong time constraints for the processing of a data set, this strategy can be considered if the magnitude of the chosen harmonics is significant enough.

4.1.3 Processing three or more fundamental frequencies

When even more complex situations arise, for example in urban areas, where the harmonic noise could be composed of more than two different fundamental frequencies (e.g. one railway line and two powerlines), the processing time will again be the most severe limitation for the use of accurate methods such as a 3-D or 4-D grid-search strategy. However, at this point in our investigations, those methods remain in our opinion the safest to use in terms of accuracy and the simplest to implement. Indeed, the residual matrix shown in Fig. 11 indicates that the nonlinear problem will not easily be solved using nonlinear optimization techniques in the case of two fundamental frequencies, and, according to our own observations on synthetic cases (not shown here), probably even less easily with three or four nonlinear parameters. Yet,

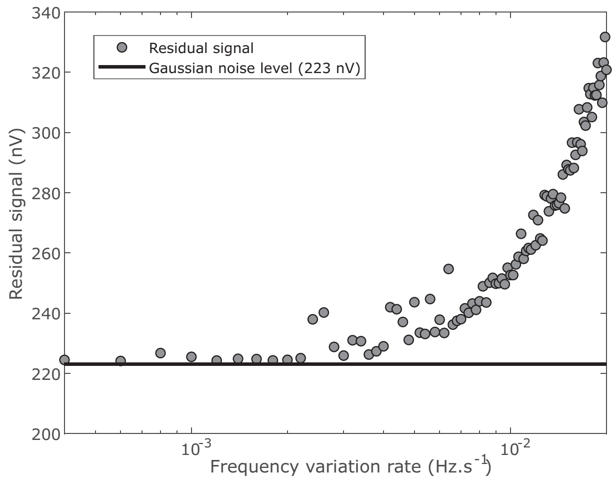

Figure 12. Evolution of the residual signal obtained after the processing of harmonic noise with a fundamental frequency varying linearly for different variation rates. The black line represents the residual signal corresponding to the Gaussian noise only.

as previously proposed, a very direct way to reduce the computing time could be to use only a low number of harmonics during the modelling phase. Synthetic tests performed on signals similar to those previously described, this time composed by three fundamental frequencies, yield the following results: processing 10 signals using 20 harmonics within the modelling took about 3.3 hr, which, by extrapolation, would result in a processing time of 330 hr to process 1000 signals (an average number of signals in SNMR surveying), which makes the method poorly efficient. However, using only 1 harmonic yields equivalent results and a processing time of only 410 s for 10 signals, which can be extrapolated to about 11.4 hr for 1000 signals, a processing time corresponding to one night of computation. Hence, provided that the harmonic used for the modelling phase is of sufficient magnitude, reducing the number of harmonics used within the Grid-search process may prove to be a suitable method to process harmonic noise composed of three fundamental frequencies.

4.2 Non-stationary fundamental frequencies

In Section 3.2, we investigated ways to process harmonic noise whose frequency content is not stationary over the duration of the record. While classical methods would fail to successfully remove such kind of harmonic contamination, we showed that these signals could be processed efficiently applying the *segmentation strategy*, where the acquisition signal is divided in several segments for which the stationary assumption is valid, and which are processed independently.

4.2.1 Number of segments in the segmentation strategy

A legitimate question regarding the implementation of the segmentation strategy concerns the optimal number of segments to be used for signal division. Ideally, the number of segments should be as low as possible to preserve the highest sampling resolution possible, while ensuring that the fundamental frequency can be considered stable over their duration. In the end, it goes down to the question of when can the fundamental frequency be considered stable? There is no definite answer to this, mainly because the way that the fundamental frequency variates over such short period of times is largely unknown (linearly, sudden changes, etc.). However we can bring some pieces of answer with a synthetic case, assuming for example a linear evolution. We generate 1 s duration synthetic harmonic noise signals, for which the fundamental frequency is increasing linearly with time, with different increasing rate, that is, with varying slopes. Gaussian noise with standard deviation of 1 and zero mean is added to these signals. Then each of these signals is processed independently and we compute the residual signal (norm L2) to assess the efficiency of the processing. Fig. 12 shows that, as the slope of the linear variation curve increases, the residual signal magnitude gradually moves away from the residual threshold corresponding to the Gaussian noise only, following what looks to be a linear trend. That is, the processing progressively loses efficiency when the fundamental frequency of the signal is changing more rapidly. In the present case, this seems to happen when the frequency variation rate exceeds about 2–3 mHz s^{-1}, and the residual signal is increased of about 10 nV. However, this value obviously depends on the level of the Gaussian noise composing the signal, and on the magnitude of the harmonic contamination. In presence of a high Gaussian noise level, the loss of efficiency will start being visible at higher variation rates, because for low variations rates the remaining harmonic components will be hidden within the Gaussian noise.

4.2.2 Computing time and accuracy of the NGSE

According to Saucier *et al*. (2006), the NGSE method offers equivalent performance as the grid-search method for frequency estimation, while providing a reduction of the computing time by a factor of at least 2. This has been the main argument to advocate the use of the

Table 1. Mono-exponential fitting results obtained after harmonic processing of the 60 noise records containing synthetic NMR signals. The uncertainty values are computed as the standard deviation of the results obtained on all the data sets processed with the same harmonic number.

	Synthetic signal	Grid-search estimation	NGSE estimation
E_0 (nV)	100	88.44 ± 0.53	89.11 ± 0.99
T_2^* (ms)	200	201.33 ± 1.28	198.55 ± 2.92

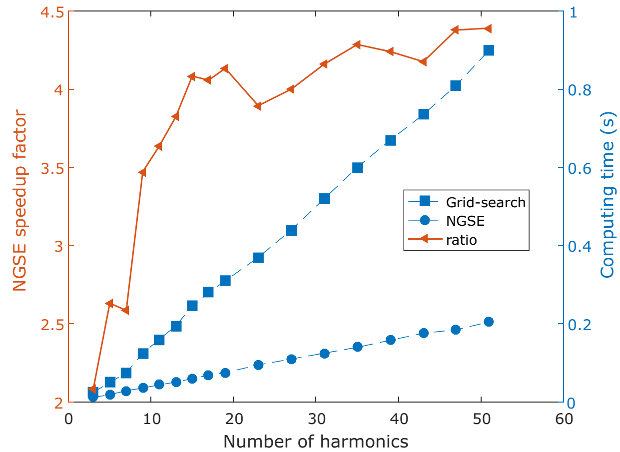

Figure 13. Comparison between the NGSE and the grid-search computing time evolution with different numbers of harmonics used within the method. Blue curves show the effective computing time evolution for the processing of a 1 s duration signal (averaged over the processing of 10 signals), which are fairly linear. The orange-triangle curve shows the ratio between the two methods, indicating that the NGSE method is always at least two times faster than the grid-search approach.

NGSE method within the segmentation strategy, so as to compensate the increase of computing time resulting from dividing the signals while ensuring equivalent processing accuracy than the grid-search method. In order to verify the claims of Saucier *et al*. (2006), we have conducted our own investigations about the efficiency and computing time relative to the NGSE method. In Appendix B, we provide a detailed explanation about some conditions of use, necessary to achieve optimal performance with the NGSE method. By doing so, we simultaneously demonstrated in Fig. B1 that the accuracy of the NGSE method was indeed equivalent to that of the grid-search method, provided that these conditions of use are respected, hence confirming the findings of Saucier *et al*. (2006).

To support these conclusions, we also compared the ability of both methods to retrieve a synthetic SNMR signal embedded in real noise data. We generated an FID signal with initial amplitude $E_0 = 100$ nV, a Larmor frequency $f_l = 2015$ Hz, a relaxation time $T_2^* = 200$ ms, and a zero phase. This signal was embedded within 60 noise-only signals recorded at the Prakla site, close to Hannover in Germany, using a square shape figure-8 loop, with 20 m long edges. The noise level at this site is relatively high, being located about 200 m from a highway in a quite urbanized area (close to the Leibniz Institute of Applied Geophysics), and is primarily dominated by harmonic noise (very weak spike content). Those 60 signals were then processed using both the NGSE method and the grid-search approach for harmonic removal. After stacking, the MRSMatlab toolbox (Müller-Petke *et al*. 2016) was used to apply synchronous detection and finally carry out a mono-exponential fit on the envelope signal to retrieve the NMR parameters. The results of this fit are summarize in Table 1. We see that in both cases the estimations are quite good, again confirming that both methods can be considered to provide equivalent performances.

Regarding the computing time, note that a precise, definite comparison between the two methods is not easy because the computational cost of the grid-search approach depends on the density of the grid that is being used, which is ultimately a decision made by the user who will be weighting the risk of missing the best frequency estimation if the grid is too coarse with the need to achieve a fast data processing. In the present paragraph we used what we consider to be a relatively coarse step-length and narrow grid-search width, so as to provide a fair and rather optimistic comparison. The search is done using a two-step iterative solving with decreasing step length (0.015 Hz and 0.001 Hz) and narrowing frequency range, (0.1 and 0.01 Hz width). We compare it with the NGSE method using 10 iterations to ensure convergence, which is a relatively high value (Saucier *et al*. 2006 advise a minimum value of 4). Both methods are implemented in Matlab, running on an Intel Xeon CPU.E5-2630 v4 processor at 2.20 GHz. Fig. 13 shows the evolution of the computing time needed to process one signal of 1 s duration (average over the processing of 10 signals) depending on the number of harmonics used within each method. Both the NGSE and the Grid-search evolution curves are fairly linear. The NGSE method is always faster than the grid-search method, and the curve showing the ratio between each method demonstrates that the NGSE speedup factor is always superior to 2, and tends to increase with increasing number of harmonics. The computing time necessary to process a full data set (e.g. 50 stacks, with 20 pulse moments, that is 1000 signals) with the grid-search approach which is typically about 1 or 2 hr depending on the computational capabilities. This is generally acceptable in post-acquisition operations. Using the segmentation strategy with 10 segments however, the computing time is increased by about one order of magnitude, and becomes about 10–20 hr. In this case then, a reduction by a factor of 2–5 thanks to the use of the NGSE method is a significant advantage.

Thus, we conclude similarly to Saucier *et al*. (2006) that the NGSE method is at least 2 times faster than the classical descent methods such as the optimized grid-search strategy proposed by Larsen *et al*. (2013), while providing a frequency estimation of equivalent accuracy, and that it is therefore suited to the processing of large data sets or within specific strategies requiring a lot of computing time. That being said, it should be noted that other methods to estimate the fundamental frequency of harmonic noise also exist, such as the frequency correlation method proposed recently by Wang *et al*. (2018), which can also provide fast estimation of the fundamental frequency. A comparison between all the frequency estimation methods would be of great interest for the geophysical community dealing with harmonic removal, but it is beyond the scope of this paper.

5 CONCLUSION

Geophysical records can suffer contamination from many EM noise sources of natural or anthropic origin. Among them, the harmonic noise produced by electrical powerlines or other man-made electrical structures is particularly important in terms of amplitude and is being increasingly encountered because of the anthropic infrastructure network densification and the growing need to perform geophysical studies in urban or peri-urban areas. In this paper, we showed that the harmonic noise can often be more complex than usually assumed in most of the literature written on the subject of harmonic processing, and we proposed new strategies to handle appropriately these circumstances, mostly in the context of SNMR acquisitions.

First, it is possible to be dealing with harmonic noise that would be the summation of different harmonic signals, each based on a different fundamental frequency. We demonstrate using a synthetic example how such a situation will result in failure of the classical harmonic processing techniques. In the case of two different fundamental frequencies, we showed that a 2-D grid-search approach for frequency estimation should always be preferred over, for example, a sequential removal approach, and supported this conclusion with synthetic studies and two field examples. When more than two fundamental frequencies compose the harmonic noise, the grid-search approach faces limitations relative to the computing time necessary to cover the entire grid-search. We propose to overcome this limitation by modelling only one harmonic, appropriately chosen.

Second, we investigated the issue of a fundamental frequency that is quickly varying within the duration of the geophysical acquisition. Again, we used a synthetic example to illustrate the deterioration of the classical methods' efficiency and proposed a new approach to solve this problem to which we refer to as the *segmentation strategy*, which consists of processing short duration parts of the signal where the constant frequency assumption is still valid. This method proved efficient, and we presented a field example that demonstrates the efficiency of this approach to handle fast changing fundamental frequencies.

Finally, during this work we were led to evaluate different ways of estimating the fundamental frequency content of harmonic contamination. In particular, for computing speed, we adapted the work by Saucier *et al*. (2006) to the SNMR context and presented this alternative way of estimating the fundamental frequency of harmonic noise, which we call the NGSE method. We confirmed that it exhibits equivalent efficiency as classical methods for the processing of harmonic noise, while offering a reduction of the computing time by a factor of 2–5 (with identical computation strategies). Hence, the NGSE method is particularly suited for the harmonic processing of large data sets, and its timeliness can also prove to be a great advantage when implementing specific processing strategies that may require fast execution to be practical, like the *segmentation strategy* presented here.

ACKNOWLEDGEMENTS

We deeply thank Dr Raphael Dlugosch, Dr Mike Müller-Petke and the Leibniz Institute of Applied Geophysics for their help in acquiring some of the noise data presented here (Schillerslage and Prakla), as well as Dr Cédric Champollion, Dr Benjamin Fores and the University of Montpellier for their help in acquiring the data from the Larzac Plateau in France. We also thank the BEWARE program from the Walloon region, who funded this work, as well as Dr Trevor Irons and an anonymous reviewer for their very helpful and constructive comments.

REFERENCES

Adams, R.K., McIntyre, J.M. & Symonds, F.W., 1982. Characteristics of the eastern interconnection line frequency, *IEEE Trans. Power Appar. Syst.,* (12), 4542–4547.

Badsar, S.A., Degrande, G., Schevenels, M. & Lombaert, G., 2007. Application of the CLM method for the solution of the inverse problem in the SASW method, in *International Conference on Earthquake Geotechnical Engineering*, Thessaloniki, Greece.

Behroozmand, A.A., Keating, K. & Auken, E., 2015. A review of the principles and applications of the NMR technique for near-surface characterization, *Surv. Geophys.,* **36**(1), 27–85.

Butler, K.E. & Russell, R.D., 1993. Subtraction of powerline harmonics from geophysical records, *Geophysics,* **58**(6), 898–903.

Butler, K.E. & Russell, R.D., 2003. Cancellation of multiple harmonic noise series in geophysical records, *Geophysics,* **68**(3), 1083–1090.

Costabel, S. & Müller-Petke, M., 2014. Despiking of magnetic resonance signals in time and wavelet domains, *Near Surf. Geophys.,* **12**(2), 185–197.

Dalgaard, E., Christiansen, P., Larsen, J.J. & Auken, E., 2014. A temporal and spatial analysis of anthropogenic noise sources affecting SNMR, *J. Appl. Geophys.,* **110,** 34–42.

Dlugosch, R., Günther, T., Müller-Petke, M. & Yaramanci, U., 2014. Two-dimensional distribution of relaxation time and water content from surface nuclear magnetic resonance, *Near Surf. Geophys.,* **12**(2), 231–241.

Dlugosch, R., Mueller-Petke, M., Günther, T., Costabel, S. & Yaramanci, U., 2011. Assessment of the potential of a new generation of surface nuclear magnetic resonance instruments, *Near Surf. Geophys.,* **9**(2), 89–102.

Fores, B., Champollion, C., Moigne, N.L., Bayer, R. & Chery, J., 2016. Assessing the precision of the iGrav superconducting gravimeter for hydrological models and karstic hydrological process identification, *Geophys. J. Int.,* **208**(1), 269–280.

Ghanati, R. & Hafizi, M.K., 2017. Statistical de-spiking and harmonic interference cancellation from surface-NMR signals via a state-conditioned filter and modified Nyman-Gaiser method, *Boll. Geofis. Teor. Appl.,* **58**(3), 181–204.

Hertrich, M., Braun, M., Gunther, T., Green, A.G. & Yaramanci, U., 2007. Surface nuclear magnetic resonance tomography, *IEEE Trans. Geosci. Remote Sens.,* **45**(11), 3752–3759.

Irons, T., Kass, M.A. & Kremer, T., 2018a. Akvo: an open source surface NMR workbench, in *Symposium on the Application of Geophysics to Engineering and Environmental Problems*, Nashville, TN, USA, 1–1. https://www.xcdsystem.com/sageep/program/qA4ED9w/index.cfm.

Irons, T.P., McPherson, B.J. & Kass, M.A., 2018b. Reliable noise measure in time-gated NMR data, *Geophys. J. Int.,* **215,** 959–964.

Jeng, Y. & Chen, C.S., 2011. A nonlinear method of removing harmonic noise in geophysical data, *Nonlinear Process. Geophys.,* **18**(3), 367–379.

Jiang, C., Lin, J., Duan, Q., Sun, S. & Tian, B., 2011. Statistical stacking and adaptive notch filter to remove high-level electromagnetic noise from MRS measurements, *Near Surf. Geophys.,* **9**(5), 459–468.

Kay, S.M., 1989. A fast and accurate single frequency estimator: IEEE Transactions on Acoustics, *Speech, and Signal Processing,* **37,** 1987–1990.

Larsen, J.J., 2016. Model-based subtraction of spikes from surface nuclear magnetic resonance data, *Geophysics,* **81**(4), WB1–WB8.

Larsen, J.J. & Behroozmand, A.A., 2016. Processing of surface-NMR data from sites with high noise levels, *Geophysics,* **81**(4), WB75–WB83.

Larsen, J.J., Dalgaard, E. & Auken, E., 2013. Noise cancelling of MRS signals combining model-based removal of powerline harmonics and multichannel Wiener filtering, *Geophys. J. Int.,* **196**(2), 828–836.

Legchenko, A. & Valla, P., 2002. A review of the basic principles for proton magnetic resonance sounding measurements, *J. Appl. Geophys.,* **50**(1–2), 3–19.

Legchenko, A. & Valla, P., 2003. Removal of power-line harmonics from proton magnetic resonance measurements, *J. Appl. Geophys.,* **53**(2–3), 103–120.

Liu, L., Grombacher, D., Auken, E. & Larsen, J.J., 2018. Removal of co-frequency powerline harmonics from multichannel surface NMR data, *IEEE Geosci. Remote Sens. Lett.,* **15**(1), 53–57.

Liu, L., Grombacher, D., Auken, E. & Larsen, J.J., 2019. Apsu: a wireless multichannel receiver system for surface nuclear magnetic resonance groundwater investigations, *Geosci. Instrum., Methods Data Syst.,* **8**(1), 1–11.

Mueller-Petke, M. & Yaramanci, U., 2010. QT inversion—comprehensive use of the complete surface NMR data set, *Geophysics,* **75**(4), WA199–WA209.

Müller-Petke, M., Braun, M., Hertrich, M., Costabel, S. & Walbrecker, J., 2016. MRSmatlab—a software tool for processing, modeling, and inversion of magnetic resonance sounding data, *Geophysics,* **81**(4), WB9–WB21.

Müller-Petke, M. & Costabel, S., 2014. Comparison and optimal parameter settings of reference based harmonic noise cancellation in time and frequency domains for surface-NMR, *Near Surface Geophysics,* **12**(2), 199–210.

Nyman, D.C. & Gaiser, J.E., 1983. Adaptive rejection of high-line contamination, in *SEG Technical Program Expanded Abstracts 1983*, Society of Exploration Geophysicists, pp. 321–323.

Sambridge, M., 1999. Geophysical inversion with a neighborhood algorithm—I. Searching a parameter space, *Geophys. J. Int.,* **138**(2), 479–494.

Saucier, A., Marchant, M. & Chouteau, M., 2006. A fast and accurate frequency estimation method for cancelling harmonic noise in geophysical records, *Geophysics,* **71**(1), V7–V18.

Walsh, D.O., 2008. Multi-channel surface NMR instrumentation and software for 1D/2D groundwater investigations, *J. Appl. Geophys.,* **66**(3–4), 140–150.

Wang, Q., Jiang, C. & Müller-Petke, M.,, 2018. An alternative approach to handling co-frequency harmonics in surface nuclear magnetic resonance data, *Geophys. J. Int.,* **215**(3), 1962–1973.

APPENDIX A: NGSE ALGORITHM—MATLAB SCRIPT

```
%%% Applying Saucier et al. (2006) method for
%%% amplitude / phase / frequency estimation for each harmonics.
%%% The input are: signal, clock vector (t=[0 T])
%%% and nominal fundamental frequency (16.66 Hz, 50 Hz, 60 Hz...).
%%% The output is the estimated fundamental frequency.

function [f0_estimation]=NGSE(s,t,f_nom)

%% NGSE parameters
kmax=10; % number of harmonics (recommended value)
maxiter=15; % number of iterations (minimum should be 5)
%% Estimation procedure
% Initialization

Est=0; Eps_est=0;
U=zeros(kmax,1); V=zeros(kmax,1);
X=zeros(kmax,1); Y=zeros(kmax,1);
eps=zeros(kmax,1); A=zeros(kmax,1); phi=zeros(kmax,1);
N=length(t); T=max(t); wn=2*pi*f_nom;

for iter=1:maxiter
  wn=wn+Eps_est; % update frequency

  for k=1:kmax
```

```
      % 1. Estimation variables(equation 32)
      U(k)=1/N*dot(cos(k*wn*t)',s);
      V(k)=1/N*dot(sin(k*wn*t)',s);
      X(k)=1/N*dot(cos(k*wn*t)'.*sin(2*pi*t/T)',s);
      Y(k)=1/N*dot(sin(k*wn*t)'.*sin(2*pi*t/T)',s);
      % 2. Compute the frequency shift (equation 22)
      eps(k)=2*pi/(k*T)*(U(k)*Y(k)-V(k)*X(k))/(U(k)^{2}+V(k)^2);
      % 3. Compute the amplitude estimation (equation 23a)
      A(k)=2*sqrt(U(k)^{2}+V(k)^{2});
      phi(k)=atan(-(U(k)+pi*Y(k))/(V(k)+pi*X(k)));
    end

  % 4. Angular shift estimate (equation 25)
  for k=1:kmax
      Eps_est=k^{2}*A(k)^{2}*eps(k)/(k^{2}*A(k)^{2});
    end
  end
  disp(strcat('Estimated fundamental frequency = ',num2str(wn/(2*pi)),' Hz'));
  f0_estimation=wn/(2*pi);
```

APPENDIX B: CONDITIONS OF USE OF THE NGSE METHOD

To ensure optimal efficiency, the NGSE method is subject to the respect of some specific conditions of use, which concerns mainly the number of harmonics used within the process. Indeed, the derivation of the NGSE method involves some first-order Taylor approximations, which are not accurate and therefore not valid anymore if the following condition is not fulfilled (Saucier *et al.* 2006):

$$k < \frac{1}{2T\Delta f}, \tag{B1}$$

where k denotes the harmonic of the kth order, T is the total signal duration and Δf is the uncertainty inherent to the fundamental frequency. In essence, eq. (B1) provides an upper boundary constraining the number of harmonics that can be used to build the frequency shift estimator, that is, it defines the quantity $k_{\max}$. This condition implies that if the 'true' fundamental frequency of the harmonic noise is far from the nominal frequency value initially used to solve eq. (6), that is, if Δf is high, a fewer number of harmonics can be used. According to Saucier

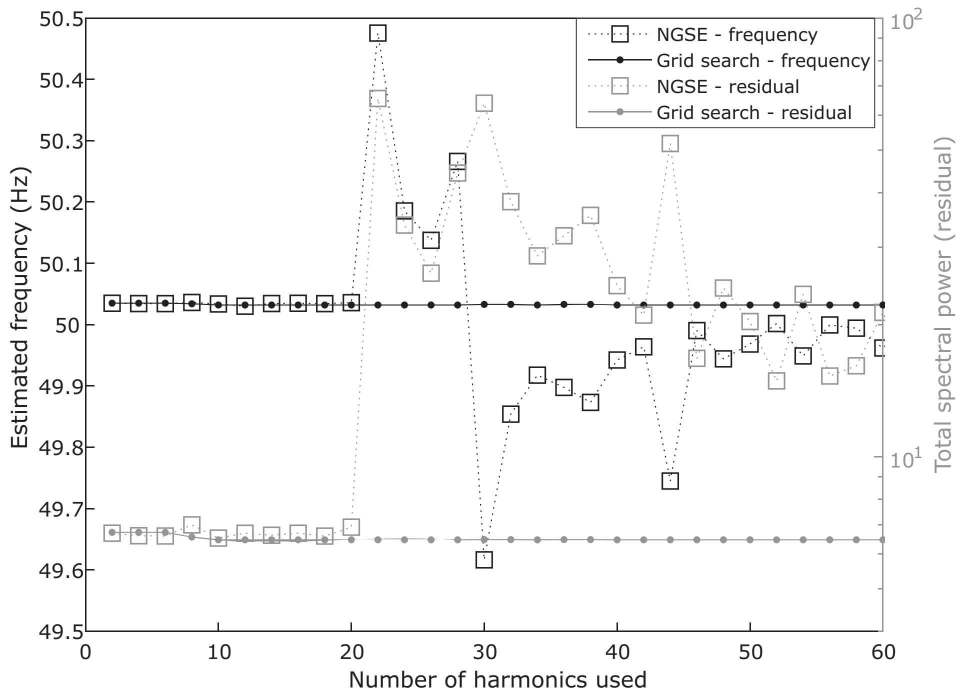

Figure B1. Comparison between the NGSE and the grid-search method for the fundamental frequency estimation evolution versus the number of harmonics used in each method. The signal used is a real noise-only record contaminated with harmonic noise. Black lines and markers represent the evolution of the estimated frequency value and grey lines and markers show the residual signal after harmonic cancellation based on the associated frequency estimate.

et al. (2006), being able to use a higher number of harmonics will improve the accuracy of the frequency estimator. However, we found that when condition (B1) does not hold, it causes drastic deterioration of the processing efficiency and should prevail as a selection criteria to determine an optimal value for $k_{\max}$. This aspect is well illustrated in Fig. B1, which shows the results of frequency estimation and harmonic removal performed on a 1 *s* duration noise-only signal recorded in a peri-urban area close to the University of Liege, where powerline harmonic noise is present. The same signal was submitted independently to the NGSE method and the grid-search approach, while varying the number of harmonics used within each method for frequency estimation. Each time, the harmonic removal is applied on the frequency range [0–4000 Hz] and the spectral power of the residual signal is computed to assess the quality of the frequency estimation.

The grid-search approach yields a frequency evolution much more stable than the NGSE method. Yet it is worth mentioning that a slight improvement is observed when the number of harmonics used exceeds 10, that is, the residual signal is slightly reduced. After that, it remains almost constant, indicating that, at least in this case, the number of harmonics used within the grid-search approach does not affect much the quality of the processing, although a number at least above 10 is advised. The mean value obtained with the grid-search method considering all harmonics is a fundamental frequency of 50.032 Hz with insignificant standard deviation. Considering this as the 'true' value of the harmonic noise, it would correspond to an uncertainty of $\Delta f = 0.032$ Hz compared to the nominal value, leading to a maximum number of harmonics $k_{\max} \approx 16$ for the application of the NGSE method. Fig. B1 shows the evolution of the estimated frequency value obtained with NGSE using an increasing number of harmonics, and of the energy of the resulting signal after harmonic subtraction. One can clearly observe that the estimated frequency is relatively stable until the number of harmonics used exceeds 20, and remains very close to the value obtained with the grid-search approach. The best energy reduction is achieved in this range, and is also logically very close to the energy reduction obtained with the grid-search estimation, confirming that the two frequency estimation methods lead to harmonic removal with similar efficiency. Then, above the threshold of 20, the frequency estimation curve for NGSE method becomes erratic and the processing efficiency is drastically reduced.

Fig. B1 illustrates the necessity of optimally selecting the maximum number of harmonics $k_{\max}$. Unfortunately, Δf is by nature unknown, therefore $k_{\max}$ cannot be determined in advance. If no *a priori* information regarding the expected value of Δf is available, which is most generally the case due to the varying nature of the fundamental frequency, we recommend to use a value of $k_{\max} = 10$, based on empirical evaluations obtained with noise-only signals recorded on different sites all presenting high levels of harmonic noise. Such a value ensures to respect the condition dictated by eq. (B1) up to a frequency uncertainty Δf of 0.05 Hz, which is, in our experience, a relatively safe uncertainty range given the expected deviations from the nominal value (Adams *et al*. 1982 propose a maximum of 0.03 Hz). A larger gap between the nominal and the true fundamental frequency value might require to use even lower values for $k_{\max}$. In difficult cases, another possibility would also be to reduce the duration T of the processed signal, hence increasing the possibilities of $k_{\max}$. However, it is not recommended to do so in the SNMR context since it would mean excluding the late time part of the NMR signal and losing the associated information.

APPENDIX C: 2-D GRID-SEARCH ALGORITHM—MATLAB SCRIPT

The 2-D grid-search algorithm presented below is largely based on the scripts and functions from the MRSMatlab software, developed by Müller-Petke *et al*. (2016). In the present form it is adapted to the removal of harmonic noise composed of two fundamental frequencies whose nominal values are both 50 Hz. If the nominal value of the harmonic noise signals is different, then some parameters may require to be changed (e.g. nominal value, number of harmonics, or the grid-search definition).

```
function [sp] = HNM_2freq(time, signal)
% This function organizes the optimized 2D grid-search process.
% The grid-search width and step are defined, and the search domain is
% covered in two steps, with varying step-size. At each iteration, the
% residual signal is computed, calling the function "fit_model" below.
% Input : (time vector (column) and signal to be processed (row vector))
% Output : [processed signal]

% Initialisation
f1=50; f2=50; % Nominal frequency values
fH1=20; fH2=20; % First harmonic to be modeled
nH1=40; nH2=40; % Number of harmonics to be modeled
bfs1 = [f1-0.05:.015:f1+0.05]; bfs2=bfs1; % Grid-search definition 1st step
e=zeros(length(bfs1),length(bfs1)); % Residual signal matrix initialisation

% Two steps optimized-grid search
for n=1:2
  for ibf=1:length(bfs1)
    iB = bfs1(ibf);
    for ibf2=1:length(bfs2)
```

```
      iB2 = bfs2(ibf2);
  e(ibf,ibf2) = norm(signal - fit_model(time,signal,fH1,nH1,fH2,nH2,iB,iB2));
      end
    end
    emin=min(e(:));
    [b,c] = find(e == emin);
    iB = bfs1(1,b(1));
    iB2 = bfs2(1,c(1));
    bfs1 = [iB-0.01:.001:iB+0.01]; % Grid-search definition 2nd step
    bfs2 = [iB2-0.01:.001:iB2+0.01]; % Grid-search definition 2nd step
  end
  disp(['base frequency n°1 estimated: ' num2str(iB)])
  disp(['base frequency n°2 estimated: ' num2str(iB2)])
  [ms,fit,f] = fit_model(time,signal,fH1,nH1,fH2,nH2,iB,iB2); % Final model
  sp=signal-ms; % Subtraction of the final model
  end
  function [ms,fit] = fit_model(time, signal, fH, nH, fH2, nH2, iB, iB2)
  % This function determines the harmonics amplitudes and phases for each
  % frequency, the problem being overdetermined and linear we do that
  % directly.
  % Inputs : (time vector as a column, signal to be processed, first harmonic
  % order,number of harmonics to be modeled for the first frequency, first
  % and number of harmonics for the 2nd frequency, fundamental frequency
  % value for the first, and second frequency).
  % Outputs : [modeled harmonic noise, fit]

  f = fH*iB + iB*[1:nH];
  f2 = fH2*iB2 + iB2*[1:nH2];
  F = repmat(f,length(time),1); F2 = repmat(f2,length(time),1);
  T = repmat(time,1,nH); T2 = repmat(time,1,nH2);
  A = [cos(T.*2.*pi.*F) sin(T.*2.*pi.*F) cos(T2.*2.*pi.*F2) sin(T2.*2.*pi.*F2)];
  fit = A\signal;
  ms = A*fit;
  end
```

APPENDIX D: DECISION TREE

In this appendix, we provide a decision tree (see Fig. D1) to help the reader proceeding to harmonic removal in difficult noise conditions. In essence, if the 1-D grid-search processing is not efficient, one should examine the possibility that several harmonic noise sources are contaminating the signal. If the environment suggests that this possibility exists, one should consider applying a 2-D grid-search approach. Otherwise, the inefficiency may be due to a non-stationary fundamental frequency, in which case the segmentation strategy should be attempted. If both methods fail, two possibilities remain: either the fundamental frequency of the harmonic noise is out of the search domain or the noise is composed by more than two fundamental frequencies. The first possibility can be checked by increasing the width of the search-domain and try again the basic processing and/or the segmentation strategy or the 2-D grid-search. The second possibility can be checked by implementing a 3-D or a 4-D grid-search.

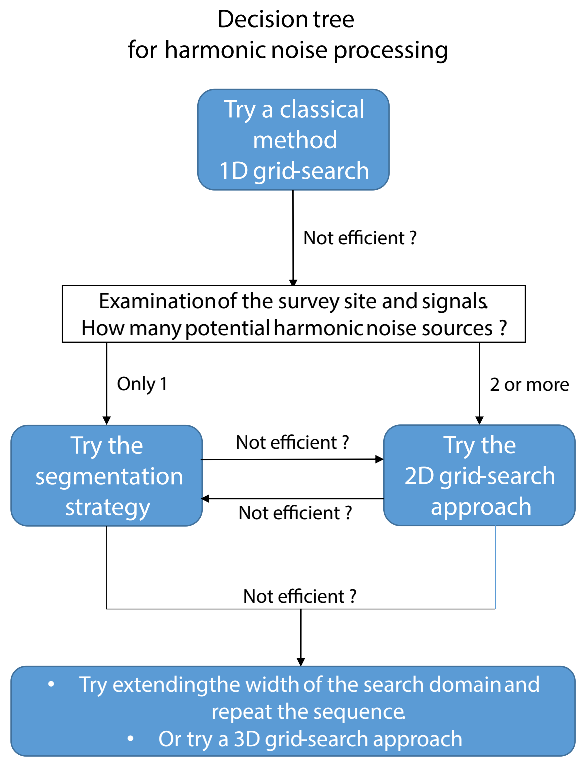

Figure D1. Decision tree for harmonic processing of SNMR acquisition signals.