MRS signal quantitation: A review of time- and frequency-domain methods

Jean-Baptiste Poullet *, Diana M. Sima, Sabine Van Huffel

Department of Electrical Engineering, SCD-SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

ABSTRACT

In this paper an overview of time-domain and frequency-domain quantitation methods is given. Advantages and drawbacks of these two families of quantitation methods are discussed. An overview of preprocessing methods, such as lineshape correction methods or unwanted component removal methods, is also given. The choice of the quantitation method depends on the data under investigation and the pursued objectives.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

These last two decades, Magnetic Resonance Spectroscopy (MRS) has shown increasing success in the MR community. One of the major goals of MRS is to quantify metabolite concentrations. However, despite tremendous efforts and numerous publications on the subject, it remains difficult to obtain accurate estimates of these concentrations, due to, inter alia, field inhomogeneities, relatively low signal-to-noise ratios (SNR), physiologic motion.

The goal of this paper is to give an overview of the existing MRS quantitation methods. Preprocessing methods, as part of the quantitation strategy, are also addressed. This includes macromolecule and solvent (or water) suppression and lineshape correction. MRS quantitation methods are usually divided into two principal categories: methods in the time domain [32,33] and methods in the frequency domain [34]. In theory, there are no differences between the two domains [35]. However, we will see that this is not totally true in practice due to some practical limitations. An introduction to the common processing methods in in vivo MR spectra is given in [36]. For sake of space, the scope of the paper is limited to post-acquisition methods, i.e., methods that are applied after signal acquisition.

The paper is organized as follows: time-domain and frequency-domain quantitation techniques are discussed in Sections 2 and 3, respectively. Section 4 gives an overview of the preprocessing methods and Section 5 describes the main quantitation features. A brief conclusion is given in Section 6.
Time-domain fitting methods are usually divided into two main classes: black-box or non-interactive methods (see, e.g., [21,39,20,10,15,12]) and methods based on iterative model function fitting or interactive methods (see, e.g., [2,37,38,31,1]), referring to the degree of interaction required by the method from the user.

2.1. Interactive methods

2.1.1. Global or local optimization

The objective of the interactive methods is usually to minimize the difference between the data and the model function, resulting in a typical non-linear least squares (NLLS) problem. This problem can be solved using local or global optimization theory. The main disadvantage of optimization procedures finding global optima, such as simulated annealing or genetic algorithms (used in MRS in [40–42]), is their poor computational efficiency. However, these methods decrease the risk of converging to a local minimum, which often occurs when the search space is of high dimension and when the starting values for the parameters are far from the global optimum. Most of the quantitation methods in MRS are based on local optimization techniques (see, e.g., [31,1,2]).

2.1.2. Use of a basis set of metabolite profiles in the model function or not

Another important feature of the interactive methods is whether they use a basis set of metabolite profiles or not. VARPRO [31], the local optimization procedure based on Osborne’s Levenberg–Marquardt algorithm [43], was the first widely used method for quantifying MRS data. It has been replaced later by AMARES, which proved to be better than VARPRO in terms of robustness and flexibility [1]. AMARES allows more prior knowledge and can also fit echo signals. These methods do not use a metabolite basis set even if the prior knowledge in AMARES can be derived from phantom data as suggested in [44]. In the presence of water components, the frequency-selective versions of VARPRO [45] and AMARES (AMARES wi [46]) are preferred and are expected to give good results for relatively well-separated peaks. However, these methods break down if nuisance peaks (i.e., peaks that are in the same frequency region but are unwanted) have large amplitudes or are close, in frequency, to the peaks of interest [21,46]. Although methods such as AMARES have been applied quite successfully to short-echo time MR spectra [47], the nuisance peaks and the more intensive user interaction tend to encourage methods based on the use of metabolite profiles since more prior knowledge is implicitly included in the model, especially information related to experimental conditions of acquisition.

On the other hand, methods such as AQSES [2] or QUEST [37] make use of a metabolite basis set, which can be built up from simulated spectra (e.g., via programs based on quantum mechanics such as NMR-SCOPE [25] or Gamma [9]) or from in vitro spectra. In [48], a spectral simulation method using Gamma for generating a priori information to be used in parametric spectral analysis is described. The use of a metabolite basis set facilitates the disentangling of overlapping resonances when the corresponding metabolite profiles also contain at least one non-overlapping resonance. Incorporating prior knowledge has been shown to provide better accuracy [49]. When adding prior knowledge one should take into account the acquisition specifications such as the type of external field $B_0$, temperature, echo time, repetition time, pH, pulse sequence, etc. If the metabolite profiles in vitro signals, the protocol used to acquire the in vitro signals should be similar to the one used to acquire the in vivo data. The influence of measured and simulated basis sets on metabolite concentration estimates, using QUEST as quantitation method, has been studied in [50]. In [38], Elster et al. proposed a semi-parametric model with an uncertainty analysis based on a Bayesian framework. They showed that this analysis yields a more appropriate characterization of the errors on the parameter estimates than the commonly used Cramér–Rao error bounds, which tend to overestimate accuracy.

2.1.3. How to choose the lineshape and the number of components in the model?

Even though individual metabolite signals can theoretically be represented by one or several complex damped exponentials (i.e., Lorentzians), in real-world situations, a perfect homogeneous magnetic field cannot be obtained throughout the sample. Therefore, Gaussian and/or Voigt lineshapes are sometimes preferred when substantial deviations from the ideal Lorentzian lineshape occur. In [51], the continuous wavelet transform is proposed to extract iteratively each resonance from the raw signal starting with the water peak, and is able to accommodate to both the Lorentzian and the Gaussian models. The model giving the best fit is selected. The problem with this approach is that if an error occurs in the first step it will be propagated all along the extracting process. The choice of the lineshape, which also determines the number of parameters per component in the model is a non-trivial problem, which is hardly solvable by a simple glance at the spectra.

Another non-trivial choice is how many components should be used in the model, i.e., how many Lorentzians (or other lineshapes) in VARPRO or AMARES or which metabolite profiles in AQSES or QUEST. Knijn et al. [45] showed that the use of a variable projection method (used in VARPRO and AQSES and not in AMARES or QUEST) reduces the sensitivity to the absence of features in the model. A variable projection method does not encounter numerical problems either when some amplitudes are nearly zero [2]. It is therefore reasonable to prefer methods based on the variable projection algorithm when there is an uncertainty about the components present in the signal. Therefore, iterative time-domain quantitation methods such as AMARES, which are not based on the variable projection algorithm, are less appropriate for complex signals such as short-echo time in vivo MRS data. A method like peak picking to identify starting values for the parameters and the number of peaks can fail when several peaks are overlapping. In [52], more flexibility on the metabolite basis set is obtained by dividing each metabolite signal into groups of magnetically equivalent spins to form a new basis. This can be useful, for example, when temperature or pH variations are expected between the in vitro basis set and the signal undergoing analysis, resulting in different chemical shifts for the same group of spins. This method is particularly interesting in high resolution MR data such as magic angle spinning data, where the influence of pH and temperature on the chemical shifts is higher.

Intuitively, the number of components has an influence on the efficiency of the method. Some methods are particularly sensitive in terms of efficiency to the number of components. For example, in [53–55], the expectation–maximization (EM) algorithm is proposed to be applied to NMR. This algorithm divides the problem into K independent optimizations, $K$ being the number of components in the signal, and allows computations on parallel computers to reduce its characteristic high computation load. In [56], Bayesian probability theory is used to estimate the exponential parameters of a known model. Probability density estimation requires the computation of integrals for which no analytical solution exists and numerical estimation is needed. Due to its intrinsic high computation load, this method is only suitable for simple signals where only a few exponentials are present. A companion paper [57] extends [56] for determining the functional form of the model (i.e., the number of exponentials).
2.2. Black-box methods

The black-box methods, either based on the linear prediction (LP) principle or based on state-space theory like HSVD (both initially introduced in MRS applications by Barkhuijsen et al. [39,13]), allow less inclusion of prior knowledge than interactive methods, being thus less suitable for more complicated signals such as short-echo time MRS signals. Furthermore, these methods are limited to Lorentzian spectra. To overcome this limitation, Belkic et al. [58] proposed a method based on the Padé transform and capable to extract unequivocally the exact number of resonances directly from the time signal, but presenting the same limitations in terms of prior knowledge as the SVD-based methods. Indeed, if a single component identified by the Padé approximant has contributions from more than one biochemical source, there is no mechanism to separate these contributions. In addition, the Padé approximant is not able to extract components with amplitudes at the same level as the noise [59]. To improve the LP and total least squares (TLS) based methods [14], Zhu et al. [16] proposed the use of an iterative quadratic maximum likelihood (IQML) method and proved the superiority of IQML over LP or TLS based methods in terms of accuracy. One drawback of this method is that, similarly to LP, it needs to calculate the root of a polynomial which may generate numerical issues. By representing non-Lorentzian lineshapes as superpositions of Lorentzian lineshapes, these methods are not able to provide physical information. These limitations are inherent to this type of methods, constituting a serious drawback, since imposing prior knowledge related to specific physical parameters may be crucial for obtaining reliable and consistent results (see, e.g., [60]). Furthermore, these limitations make these techniques not appropriate for further classification problems since the extracted features will likely vary from one signal to another.

Although imposing prior knowledge is limited, some can however be incorporated into the model [15,61,17,62]. Chen et al. [15,61] derived an algorithm HTLS-PK able to include prior knowledge of known signal poles. This method has been outperformed by KNOB-TLS, a method proposed in [17], especially in terms of robustness. KNOB-TLS provides parameter estimates which are comparable to those obtained with AMARES, and which could be used as starting values in AMARES as suggested in [17]. In [21], Romano et al. proposed a frequency-selective method referred to as MeFreS (Metropolis Frequency-Selective), based on rank minimization of a Hankel matrix. The minimization procedure uses the down-hill simplex method implemented with simulated annealing. MeFreS does not use any preprocessing steps or filter to suppress nuisance peaks, but the signal model function is directly fitted. This method is compared to AMARES and VARPRO in [21]. Simulations show that MeFreS is able to correctly identify spectral parameters also in those cases where AMARES and VARPRO are expected to fail. The fitting process is also different since MeFreS fits only one spectral component/peak at a time by first selecting its single frequency, while AMARES and VARPRO need to fit all peaks that fall in the specified frequency range.

Another important limitation of SVD-based methods is their unsuitability for dealing with data that contains significant signal intensity from rapidly decaying resonances of macromolecules. SVD-based methods require manipulating the original data such that they follow a Lorentzian model. This is always inferior to a method that models the data as they were collected. Disentangling the signal of interest from the baseline requires prior knowledge often lacking (or not includable into the model) when using SVD-based methods. Moreover, these methods assume a Lorentzian-type model, which might be too limited for baseline signals, Gaussian lineshapes being often preferred to model the broad resonance signals from macromolecules (see, e.g., [63]).

A more detailed overview of the black-box methods is done in Section 4 since these methods are nowadays mainly used as solvent suppression methods.

3. Frequency-domain quantitation methods

The frequency domain is naturally suited for frequency-selective analysis with the advantage of decreasing the number of model parameters. Visual interpretation of the measured MRS signals and of the fitting results is best done in the frequency domain.

3.1. Non-iterative methods

3.1.1. Peak integration

The oldest and still widely used quantitation method in the frequency domain is based on the integration of the area under the peaks of interest [64]. The advantage is that no assumptions have to be made concerning the lineshape of the signal. Unfortunately, this method is not able to disentangle overlapping peaks and therefore to extract information from individual peaks or metabolite contributions. Residual baseline signals and low SNRs will also hamper good quantitation. Furthermore, an appropriate phasing is necessary when dealing with the real part of the frequency-domain MRS signal, which is far from trivial. Peak integration depends widely on the defined bounds. The tail of the peaks is also neglected by peak integration and the area under the peaks will be therefore underestimated (possibly by up to 40% [64]).

3.1.2. SVD-based techniques

The frequency domain allows a straightforward selection of a frequency interval. SVD-based techniques are based on this observation and are therefore frequency-selective methods. Only the points in the frequency range of interest are considered for quantitation, resulting in faster algorithms. In [8], five methods are compared: the filter diagonalization method (FDM) [7,65], a modified version of MODE [22] to be usable in a selected frequency band (SELF-MODE), a data filtering and decimation approach FIDO (Filtering and Downsampling) [7], the ARMA-modeling based filtering and deminication technique called SB-HOYSVD [29], and the frequency-selective implementation of ESPRIT [6] (see, e.g., [66]) called SELF-SVD [30]. For moderately high SNRs, FDM seems to give better estimates than the four other methods. SELF-MODE and SELF-SVD have a stable parameter accuracy with relative root mean squared errors (RMSes) lying between FDM and the two filtering and decimation methods. SELF-SVD is the fastest method. SB-HOYSVD has the largest number of user parameters (i.e., the most interactive method), Djermoune et al. proposed an adapted version of SB-HOYSVD [67], which is intended to reduce the computational burden and to avoid the choice of the decimation factor (or the width of the spectral windows) which, in the case of a uniform decomposition, strongly conditions the estimation results. In [68], FDM has been shown to outperform LP-ZOOM [20]. The computational speed of these methods is generally superior to that of the time-domain SVD-based method HSVD, depending on the size of the frequency interval of interest, the number of components and the total number of data samples. As it is possible to decrease the computational load for time-domain SVD-based methods by using the fast Lanczos algorithm, it is equally possible to use the latter for these frequency-selective methods. The limitations regarding prior knowledge of time-domain SVD-based methods remain true for these frequency-domain methods.
3.2. Iterative methods

In parallel, methods based on model functions have been proposed (see, e.g., [35, 69, 70, 18, 71]). Although these methods are equivalent to time-domain fitting methods from a theoretical point of view, a simple exact analytical expression of the discrete Fourier transform (DFT) of the model function is often not available for the Voigt and/or Gaussian lineshapes, even if numerical approximations exist [72–74]. For example, in [73, 75], approximated Voigt lineshapes have been proposed, and the spectra were fit with the Levenberg–Marquardt algorithm. In any case, the model functions in the frequency domain are, in general, more complicated than in the time domain and necessitate thereby more computation time. Marshall et al. [76] show that the choice of the lineshape affects the metabolite peak areas and suggest the use of Gaussian lineshapes instead of Lorentzian lineshapes. The frequency-domain methods which only use the real part of the spectrum in their model, such as LCMODEL [18], require a very good phasing to get the spectrum in its absorption mode.

As for time-domain methods, many frequency-domain methods solve the NLLS problem by local optimization techniques, in particular using the Levenberg–Marquardt algorithm (see, e.g., [71, 18]).

3.3. Other techniques

A real-time automated way of quantifying metabolites in long-echo time in vivo NMR spectra using an artificial neural network (ANN) analysis is presented in [77, 78]. The performance of the ANN was compared with an established lineshape fitting (LF) analysis [19] using both simulated and experimental spectral data as inputs. The ANN quantified these spectra with an accuracy similar to LF analysis but was more easily automated.

Principal component analysis (PCA) has also been proposed as quantitation method in MRS [79]. PCA has the advantage of being model independent, making it well suited for the analysis of spectra with complicated or unknown lineshapes. It is not suitable if several overlapping peaks have to be quantified but might be useful when dealing with isolated peaks. PCA considers an entire data set at once, improving its precision in the presence of noise over methods that analyze one spectrum at a time. However, standard PCA will never give parameter information such as chemical shifts or linewidth and it will be accurate for low SNR only if the number of available spectra is large enough. A severe drawback of standard PCA was that all spectra in the data had to be in phase, which is often far from being trivial. To circumvent this issue, a modified PCA, which utilizes complex SVD to analyze spectral data sets with any amount of variation in spectral phase, has been developed [80]. More recent developments have extended this method to quantify all peak characteristics, including the linewidths [81]. In [82], a review of NMR spectra quantitation by PCA is given. Stoyanova et al. [83] proposed a superior method to the one in [81] in terms of stability, convergence and the range of variations it can determine. In [84], Ladroue et al. combined PCA and independent component analysis (ICA) and showed that signals with low occurrence and low SNR can be identified.

In [3], a quantitation algorithm for in vivo MR spectra based on the analysis of circles (CFIT) is described. The circular trajectories resulting from the projection of the peaks onto the complex plane, are fitted with active circle models. The use of active contour strategies allows incorporation of prior knowledge as constraint energy terms. The problem of phasing spectra is eliminated, and baseline artefacts are dealt with using active contours-snakes. A wide range of prior knowledge, including non-linear constraints, can be incorporated in CFIT. Slightly less good relative root mean squares errors (RRMSEs) have been reported for CFIT compared to AMARES. On the other hand, CFIT presents a better success rate for resolving the peaks of interest within specific intervals lying symmetrically around the true frequencies than AMARES, especially in the presence of baseline distortions.

Another quantitation method which aims to circumvent the disadvantages of both time- and frequency-domain fitting has been proposed in [85], and referred to as time-domain frequency-domain (TDFD) fitting. The model is expressed in the time domain to keep flexibility for the lineshapes and for possible truncation or other typical time-domain processing. However, the fitting itself occurs in the frequency domain after Fourier transforming the discrete time-domain signals, which are the model and the signal under investigation. Due to the additional Fourier transform needed at each optimization iteration, TDFD fitting is approximately 20% slower than a pure time domain fitting method such as VARPRO. This difference is reduced when considering frequency-selective fitting for which time-domain methods require an additional method while frequency selection is straightforward in the frequency domain. TDFD fitting also allows non-analytical lineshapes.

4. Preprocessing techniques

Acquired MRS signals are rarely purely exponentially decaying due to experimental conditions (shimming imperfections, physiological motion, etc.) and need to be preprocessed to be suitable for analysis, i.e., such that the modified signals match the model. The influence of nuisance peaks in NLLS parameter estimation techniques such as VARPRO and AMARES has been studied in [45].

4.1. Correction for lineshape or model imperfections

Lineshape deviations from an exponentially decaying signal due to residual eddy currents and magnetic field inhomogeneities are often present in $^1$H spectroscopic data.

The eddy currents give rise to time-varying phase-shifts in the acquired data. One of the oldest and still widely used techniques was proposed by Klose et al. [4], inspired by [86], to correct pointwise the time-domain signal using, as reference, the water unsuppressed signal (no hardware suppression of the water signal). In [87], wavelets have been used to remove the phase distortion induced by eddy currents.

Other methods aim to correct for arbitrary lineshape imperfections (i.e., not satisfying a perfect exponentially decaying signal). In [88], a reference peak is chosen as one of the peaks in the experimental data. The time-domain reference signal is obtained by setting the spectral values outside the reference peak frequency region to zero and using the Fourier transform. A potential drawback is that the reference signal might be equal or close to zero in certain time points, resulting in spikes in the frequency domain. Moreover, setting points to zero boils down to multiplying the frequency signal by a rectangular window, generating the well-known ringing effect in the time domain. An algorithm based on the same principle as in [88] was proposed in [28]. The idea of this method, the so-called QUALITY method, is to pointwise divide the signal under investigation by an estimated lineshape deviation (from a pure decaying exponential) using either separated data or an isolated peak in the data to be quantitated. A further development for automating this method has been proposed in [89, 90]. The problem of the above methods including QUALITY is the potential risk of dividing by zero (spike effect described above). In [27], a method inspired by [26, 4], from which it takes its name QUECC (concatenation of “QU” for QUALITY and “ECC” for Klose’s eddy Current Correction method), is meant to benefit from the advantages of both methods, QUALITY for a complete correction of the lineshape such that it matches a decaying exponential and
ECC for avoiding the spike effect. The signal is separated in two parts defined by a crossover point in the time domain which depends on the slope and the SNR of points in the time domain reference data. The first part of the signal is corrected using QUALITY deconvolution, while the second part is corrected using ECC. To avoid discontinuity in the signal, an exponential damping constant is evaluated to equalize the magnitude of the last point that was QUALITY deconvolved with the magnitude of the first ECC point.

Instead of deconvolving the experimental signal, the lineshape can be incorporated into the fit by multiplying the model line-shapes or the metabolite profiles in the time domain with the reference lineshape (see, e.g., [85]). In the case where no information is available for the lineshape, the latter can be incorporated into the fit as an unknown vector which is convolved with the metabolite profiles in the frequency domain (see [18] for more details), modeled in the time domain (see, e.g., [85]), or estimated from the convolution of the raw data with a undamped spectrum (i.e., a simulated spectrum with zero linewidth) followed by measurement of the full width at half maximum (FWHM) value [71]. In [91], Maudsley proposed another method which does not require the use of a reference peak. The method is iterative and based on an initial estimate of the parameters of the spectral components.

4.2. Water peak removal

Biological or biochemical samples are generally recorded in aqueous solution. Due to the large proportion of water, the signal intensity of water is often several orders of magnitude larger than the signal intensities of the other metabolite components. Suppressing the water signal has been a key issue for designing spectrometers, acquisition sequences and post-acquisition methods (called preprocessing methods in this paper). An overview of these preprocessing methods for solvent suppression is given in this section. This section considers both cases: water-suppressed and water-unsuppressed signals. Note also that several pulse sequences achieve water suppression (see, e.g., [92–94]).

4.2.1. Water-suppressed signals

Using water-suppressed signals for quantitation is still the standard procedure, although recent publications (see, e.g., [95,96]) have shown that quantitation of water unsuppressed signals could also be carried out successfully. Most of the water suppression techniques have been developed based on water-suppressed signals and have been widely tested. With water-unsuppressed signals, gradients-induced artifacts, which originate from the switching of gradient pulses, cannot be totally removed, thereby reducing the accuracy of the parameter estimates. We can distinguish between methods based on the use of a finite impulse response filter and those based on a model function.

4.2.2. FIR filter techniques

In [97] Kuroda et al. used first and second order differentiation to suppress the water peak. In order to improve this filter, Marion et al. [98] proposed a low pass FIR filter. The drawback of these filters is that they are linear phase filters which generate signal distortion due to the fact that the signals are composed of exponentially damped sinusoids and not pure sinusoids as shown in [24]. In order to reduce this distortion, Sundin et al. [24] proposed a maximum-phase FIR filter (MP-FIR). Although, these distortions are strongly reduced, they cannot be neglected when the stopband region is large or when the damping factor is high, as noticed by Poulet et al. [99]. Entire tails of frequency domain water signals can be removed by this method. A generalization of the method and advice to use it are given in [46]. Wavelets have also been used for water removal (see, e.g., [100–103]) and, in [101], the Gabor transform is proposed as a good alternative to the wavelets. In a review of filtering approaches to solvent suppression in MRS [102], 5 filtering methods are compared: a Gabor transform based method [101], the method of Marion et al. [98], the method of Sodano and Delepiere [104], the Cross method [105], the maximum-phase Finite Impulse Response (MP-FIR) filter method [24]. MP-FIR filter by Sundin et al. [24] has been shown to be the most accurate and efficient technique among them for quantifying long-echo time MRS spectra. In addition, MP-FIR allows the inclusion of prior knowledge that may be taken into account during quantitation (see [102] for more details). MP-FIR has also been successful in quantifying short-echo time in vivo MRS [2].

In [5], the ER-filter method is proposed. The idea is to select the frequency region of interest by filtering with a rectangular window in the frequency domain, and to get back to the time domain, reducing substantially the number of points in the signal. Although this technique inherently distorts the signal (ringing effect of the reduced FID due to rectangular filtering), it can be used when the wanted spectral region is smaller than the width of the full spectrum and the number of data points is large [46]. Its use might also be interesting for speeding up the quantitation process [106]. The estimation results are largely influenced by the choice of the filter type and filter order, for which only limited guidelines have been provided.

4.2.3. Based on a model function

Another approach is to model the water signal and subtract it from the original signal. The water signal is rarely a pure exponentially decaying signal due to field inhomogeneities and/or partial water suppression and is thereby not easily parameterized. However, the so-called black-box methods have been successful in reconstructing the water signal usually modeled as a sum of Lorentzians. The most common method is HLSVD developed by Pijnappel et al. [10] which reduces the computational load of the original HSVD method by computing only part of the SVD by using the Lanczos algorithm. An improved variant of HSVD is HTLS [14] which computes the TLS solution instead of the LS solution. In [107], HTLS is improved to deal with spectra which contain closely spaced sinusoids. From HLSVD, several variants have been developed (see, e.g., [11,12]). The main advantage of these methods compared to linear prediction methods [108] is that polynomial rooting and root selection are avoided. This is also the case for Matrix Pencil (MP) methods (see, e.g., [109]), since these methods also find (like space-state methods) the estimates of the signal poles as eigenvalues of a matrix. In [23], Rao reported no difference between estimates obtained by MP and state-space methods. The Cadzow method or minimum variance technique can also be used to preprocess the data to improve the basic HSVD and HTLS algorithm [110]. In [12], Laudadio et al. compare HLSVD with two other proposed variants: the method based on the Lanczos algorithm with Partial ReOrthogonalization (HLSVD-PRO) and the method based on the Implicitly Restarted Lanczos Algorithm (HLSVD-IRL) [111]. HLSVD-PRO and HLSVD-IRL outperform HLSVD in terms of computational efficiency and numerical reliability. Moreover, HLSVD-PRO is faster than HLSVD-IRL [111]. The user has to specify the model order and the frequency region of the water peak. These choices may influence the accuracy of the estimated parameters as shown in [46,112]. A drawback of these methods is their large computational complexity. Even fast methods such as HLSVD or HLSVD-PRO are much less efficient than FIR filter based methods (see, e.g., [46,99]). After subtracting the water signal from the original signal, most of the water tails are removed and the influence on the peaks of interest is small. However, as shown in [99], this influence exists and can reduce the quality of the parameter estimates.
HSVD and MP-FIR used with AMARES [1] are compared in [46] (so-called AMARES$_p$ and AMARES$_s$ for HSVD and MP-FIR, respectively). Combined with AMARES, HSVD is proved to be less accurate and efficient than MP-FIR for long-echo time MR data. Similar observations have been done in [99] for short-echo time MR data with HLSVD-PRO and MP-FIR combined with AQSES [2], where MP-FIR outperformed HLSVD-PRO in terms of accuracy and efficiency.

In frequency-domain quantitation methods, the residual water peak tails are often dealt with by considering them as an additional baseline (e.g., [69,85,113]).

4.3.1. Baseline distortions

Water-suppressed signals have become competitive thanks to high-resolution analog-to-digital converters (ADCs), which avoid digitizer overload (due to the high dynamic range, i.e., the high amplitude of the water compared to the metabolite amplitude) that results in severe digital noise. Dong et al. [95] report the following disadvantages of water-suppressed signals: (1) signals with small chemical shift differences to water are also partially suppressed, (2) magnetization transfer effects to metabolites [114–116] may cause systematic quantitation errors, and (3) RF pulses used for water suppression increase the total RF power deposition and may require additional adjustments. Furthermore, water-suppressed signals also present the following advantages: the water signal can be used as a reference for line-shape transformation and as an internal reference for absolute metabolite quantitation, both without additional measurements [95], but also for phase correction accounting for motion induced phase fluctuation between individual scans [117]. Note that additional preprocessing steps may be needed when using water-suppressed signals, for example to avoid nuisance peaks due to sideband artefacts (see, e.g., [96,95]). Although most of the methods used for water-suppressed signals should be applicable to water-unsuppressed signals, one should be careful when using FIR filtering techniques since these techniques may have limited performance in terms of attenuation of the water peak. Indeed, a water peak amplitude of 3 to 5 orders higher than the metabolite peak amplitudes requires an attenuation of ~60 to ~100 dB, which may be difficult to achieve due to the constraints imposed on the FIR filter (e.g., length of the filter or transition band width). SVD-based methods or MP method (used in [95]) are not affected by this problem.

4.3.2. Macromolecular signals

Macromolecular signals are often considered as nuisance components in MRS since they usually overlap with the metabolite contributions in the frequency domain. However, recent studies [125–127] show that strong correlation between the macromolecular concentration/composition and the location of the voxel in the brain have been found. Hoffmann et al. [125] also found significant correlation with age but not with gender, while no significant correlation with age could be detected by Mader et al. [126] (the correlation with gender was not studied in the latter reference). Similarly, several conditions such as stroke [128], brain tumors [63] and multiple sclerosis [129] show an altered macromolecule profile. Therefore, the macromolecular signal can provide relevant clinical information. The goal is thus to disentangle the macromolecule contributions from the metabolite signals in order to obtain accurate parameter estimates from quantitation while keeping the information provided by the extracted macromolecular signal.

In spite of our better knowledge of the macromolecular signal, it remains difficult to predict it in vivo MRS signals, and most of the classical methods just assume its smoothness in the frequency domain. The macromolecular signal can be removed in a preprocessing step (see, e.g., [37,71,130]) or can be modeled in the quantitation step (see, e.g., [2,38,18]).

4.3.2.1. In the preprocessing step. Different preprocessing approaches have been developed. The simplest one, based on the fact that the macromolecular components decay more rapidly than the metabolites, is to truncate some of the initial points in the time domain [131] (also called the “Trunc” method by Ratiney et al. [37]). This technique presents some drawbacks: the useful information is partially lost, selecting the number of points to be truncated is difficult and the spectrum may have an oscillating behavior due to discontinuities in the time domain after zero filling. More advanced techniques consist of subtracting a modeled macromolecular signal in the frequency domain from the original spectrum. Models may be generated with wavelets [71,132,133,95,75] or splines [134]. A comparison between wavelets and splines has been done in [130] but no significant differences have been found. The macromolecular baseline can also be measured in the time [135] or the frequency domain [136], then modeled as a sum of Gaussian peaks [137] or Voigt lines [125], and finally subtracted from the original signal. Ratiney et al. [37] proposed a three-step method (called “Subtract”) for subtracting the macromolecular baseline: (1) truncate the initial points and quantitate with QUEST [28] the metabolites, (2) estimate the baseline from the metabo-
lite-free signal by an SVD-based method or AMARES, and (3) subtract the parameterized baseline from the raw signal. The so-called time-domain frequency-domain (TDFD) methods follow the same principle [85,71,138] even if the wavelets or splines are usually preferred to the SVD-based methods for modeling the baseline. Other techniques such as SVD-based methods [139] have also been proposed. Although these methods have been shown to be rather successful for removing the baseline, they require an additional step prior to the quantitation, thereby increasing the risk of larger errors in the amplitude estimates.

4.3.2.2. In the quantitation step. On the other hand, the baseline can be modeled in the quantitation step. In parametric models, the profiles of the baseline components, obtained from measurements [126,63,140–144] or from theoretical considerations [145], are added to the database of metabolites. The authors of these papers conclude that including the baseline components in the basis set of metabolite profiles provides more accurate results. The baseline can be measured using specific sequences based on T1 relaxation such as the inversion-recovery [136,146] or the saturation-recovery sequences [125]. Baseline removal can also be based on T2 relaxation by increasing the echo time [147]. However, the in vivo determination of the exact relaxation times for both macromolecules and metabolites is complicated and time consuming. Furthermore, neither metabolites nor macromolecules necessarily present a narrow distribution of relaxation times. Williamson et al. used the Padé Transform to separate the baseline from the metabolite signals [59].

In semi-parametric models, the baseline signal is supposed to be smoother than the spectral components of interest. Different functions have been used to approximate the baseline: linear combination of splines (see, e.g., LCModel [18] or AQSES [2]), linear combination of reproducing kernels associated with a reproducing kernel Hilbert space (see, e.g., [38]). Incorporating the baseline into the fit via non-parametric modeling allows a one-step procedure which reduces the risk of accumulated error.

5. Discussion

A beginner in the field of MRS quantitation who needs to choose an appropriate quantitation method may face a big challenge. The choice is often made based on the availability (free or commercial, accessible via internet or not) of the method and its user-friendliness. In this paper and, in particular, in this section, we enlighten the general features of quantitation methods to help the reader to choose an appropriate method for his/her data. A better quantitation often results from better prior knowledge, and quantitation methods should be chosen in order to include as much prior knowledge as possible in the model. However, one should remember that incorporating prior knowledge is only beneficial when it is sufficiently close to the reality. There are indeed two reasons for ending up in an unwanted local minimum when using local optimization algorithms: bad initial estimates of the parameters, and wrongly implemented prior knowledge. Here is a list of key points for choosing a quantitation method. The features of the main quantitation methods are reported in Table 1, each column corresponding to one of the following features:

(i) Using an in vitro or simulated database of metabolite profiles

A first step is to identify the data to be analyzed, their complexity (i.e., high number of peaks? overlapping peaks?). As a rule of thumb, spectra with a large number of overlapping peaks are more easily modeled by a linear combination of metabolite profiles rather than a linear combination of Lorentzian, Gaussian or Voigt components. On the other hand, signals with a low number of resonances are easily handled by methods like AMARES or VARPRO. AMARES should be preferred to VARPRO particularly when constraints on the linear parameters (metabolite amplitudes, phases) have to be imposed. For example, complex signals such as short-echo time MRS data will be quantified by AQSES, LCModel, QUEST, TDFDFit or Elster’s while VARPRO or AMARES should be applied to long-echo time MRS data. When there is no prior knowledge, or at least, no reliable prior knowledge is available, non-parametric approaches [148] such as HLSVD can be used to quantify MRS data.

(ii) Incorporating an unknown lineshape into the fitting model

Taking the lineshape into count is necessary in MRS quantitation. Marshall et al. gives a simple example in [76] where modeling a Gaussian peak with a Lorentzian peak results in a 26% error. This error decreases for 2 overlapping peaks with a minimum of 17% for a distance between the peaks of about twice their FWHM. Peaks or apparent signals in the fitting residuals are usually an indication of an inappropriate model. In general, the best way to correct for non-exponential decay is to use a reference signal (such as the water-suppressed signal) which has undergone the same distortions (see, e.g., [27,26,4]). Instead of deconvolving the original signal, one should, if possible, add the distortions to the profiles of the metabolite database to avoid any division by zero (see Section 4.1). If no reference signal is available, one can still include the lineshape estimation into the fitting process (see, e.g., [18,85]).

(iii) Incorporating water filtering into the fitting process

Frequency-domain methods usually consider the water tails overlapping with the metabolites of interest as part of the baseline and do not consider water filtering. On the contrary, time-domain methods need to remove the water components. As shown in [99], including water filtering inside the optimization process may improve the parameter estimates. When dealing with unsuppressed water signals, it is preferable to use SVD-based methods instead of FIR filtering techniques to avoid problems due to a too weak attenuation of the water signal (see Section 4.2.4).

(iv) Modeling the macromolecular signal and baseline distortions

The macromolecular signal should be included in the model (i.e., used in the quantitation step, see Section 4.3.2) when macromolecular contributions are present in the signals, either as a smooth function (see, e.g., [38,2,18]) or as additional “metabolite” profiles in the database. The latter is often preferred in recent publications (see, e.g., [126,63,140,141]). This can be explained by the fact that adding macromolecular profiles adds more prior knowledge.

<table>
<thead>
<tr>
<th>Table 1 Features of some quantitation methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
</tr>
<tr>
<td>HLSVD [10]</td>
</tr>
<tr>
<td>VARPRO [31]</td>
</tr>
<tr>
<td>AMARES [1,46]</td>
</tr>
<tr>
<td>QUEST [28]</td>
</tr>
<tr>
<td>Elster et al.’s [38]</td>
</tr>
<tr>
<td>TDFDFit [85]</td>
</tr>
<tr>
<td>CFIT [3]</td>
</tr>
<tr>
<td>LCModel [18]</td>
</tr>
<tr>
<td>Young/Soher et al.’s [132,48,71]</td>
</tr>
</tbody>
</table>

An ‘X’ indicates that the method *uses an in vitro or simulated database of metabolite profiles, *incorporates an unknown lineshape into the fitting model, *incoporates water filtering into the fitting process, *models the macromolecular signal and baseline distortions.
than the assumption of smoothness of the macromolecular signal. Disentangling the baseline from the rest of the signals before the quantitation process increases the risk of errors since any error due to disentangling will affect the parameter estimates and be superimposed to the fitting errors. The methods with an ‘X’ in the last column of Table 1 assume the smoothness of the baseline without distinguishing between baseline distortions and macromolecular signal. Moreover, frequency-domain methods such as LCModel [18] considers the tails of the water resonance as part of the baseline distortions.

5.1. Other important considerations

5.1.1. Time- or frequency-domain method?

Time- and frequency-domain methods are theoretically similar in performance even if the time-domain methods allow more flexibility in terms of model line shapes. Only a few took the risk of comparing time- and frequency-domain methods and no strong conclusions could be drawn. In [149], 4 methods have been compared on in vivo 31P MR data of tumors: VARPRO and HLSVD as time-domain quantitation methods, and peak integration and Lorentzian fitting as frequency-domain quantitation methods. The results suggest that VARPRO is the method of choice for quantitative analysis of tumor 31P MR spectra, giving the most reliable results at low SNR. Kanowski et al. [47], for instance, reported comparable results for AMARES and LCModel. It is also important to notice that the fast Fourier transform (FFT) is suboptimal if (1) the noise is not Gaussian, (2) the sampling time is not constant (different time steps), (3) samples are missing. In these cases, it might be preferable to avoid the FFT and to do the analysis in the measurement or time domain.

5.1.2. Lorentzian, Gaussian or Voigt model?

The choice of the model is a non-trivial question. One should first correct for lineshape imperfections as mentioned above. These corrections may not be sufficient to obtain pure Lorentzian signals and other lineshape models such as Gaussian or Voigt may be preferable. It is often complicated to judge whether the peaks in the signal are Lorentzian, Gaussian or Voigt. However, one can test different lineshape models and choose the one which gives the best residuals (small residuals with no peak or signal in it) and the best success rate in the sense of Gabr et al. [3] (see Section 3.3). As Marshall et al. showed numerically [76], choosing a wrong model is less important when modeling two Gaussians than one unique Gaussian. This is explained by the fact that the large Lorentzian tails compensate the natural overestimation of the amplitudes when modeling Gaussians by Lorentzians. One can also intuitively imagine that adding noise (smaller SNR) or baseline distortions will also reduce the effect of a wrong model (which does not mean that the error will be smaller). However, it would be very challenging to fix a threshold value for the SNR at which we can consider the choice of the lineshape as important since this value depends on the signal under investigation (number/shape of peaks, artefacts in the signal, macromolecular signal, etc.).

5.1.3. Is my method robust?

Most of the quantitation methods claim to be robust, but are not necessarily robust against the same type of disturbances (noise, baseline, water peak, etc.). Moreover, they usually base their conclusions on simulated spectra that do not reflect all the artefacts or distortions present in a measured signal. In order to analyze the robustness of a method on in vivo signals, Gabr et al. [3] proposed to study the success rate (or failure rate) to resolve the peaks of interest within specific intervals lying symmetrically around the true frequencies. They show that CFIT is less sensitive than AMARES to baseline distortions. When considering only non-failure cases, AMARES presents lower RRMS than CFIT. That is why it is important to identify the components in the signal, known and unknown: the rolling baseline is visible and is not part of the metabolite signal, therefore it should be removed before using AMARES. Gabr et al. confirm that much better success rates are obtained when using AMARES after filtering the rolling baseline. High failure rates may be an indication of a wrong model, or remaining artefacts (in this case the rolling baseline) that should be removed prior to quantitation. Signals with non-Gaussian noise can also lead to non-optimal parameter estimation. Indeed, the least squares problem yields the smallest estimation errors when the distribution is Gaussian and is suboptimal otherwise. MR scanner noise is supposed to be Gaussian, but perturbations or deviations from Gaussian distribution may occur due, for example, to body motion [150]. These perturbations may be considered as acquisition artefacts which are beyond the scope of the paper. In [150], Slotboom et al. proposed a method to detect and discard signals with non-Gaussian noise.

5.1.4. Variable projection or not in the optimization algorithm?

When no prior knowledge about the linear parameters (amplitudes, phases) of the model is available, an optimization method using variable projection (like in [2,31]) is preferable because all linear parameters are projected out thereby reducing the number of parameters to be optimized by one half or more. If equal phases are assumed, variable projection can still be used in a modified form [151]. In other cases, a more general optimization algorithm (like the non-linear least squares method NL2SOL as used in AMARES [11]), which optimizes all parameters (linear and non-linear) is recommended.

5.1.5. Weighting and normalization

If one wants to give more importance to particular frequency regions, a weighting matrix can be used in the minimization function, which multiplies the vector of squared errors (see, e.g., Eq. (11) in [85]). The largest weights are assigned to the frequency points of interest.

One may also want to give the same importance to the same weight to all the peaks. In that case the least squared error can be normalized in order to balance the peak contributions with respect to this error (see, e.g., Eq. (12) in [85]). This might however be dangerous since quantitation methods like LCModel or TDFD Fit tend to overestimate the amplitude of low concentration metabolites [152].

6. Future improvements and conclusion

Improving quantitation means increasing prior knowledge. Hardware improvements can also contribute to better prior knowledge. Here are some hints for possible improvements:

- One of the main weaknesses of the quantitation methods is their way of dealing with the baseline. Only little prior knowledge regarding the baseline is currently used in the model, resulting in a poor separation between the baseline and the metabolites of interest. Furthermore, macromolecular components have to be clearly distinguished from baseline distortions since the former may provide useful information for pathology diagnosis (see Section 4.3.2).
- Spatial information in MRSI data is also not sufficiently exploited. Quantitation is often done on individual voxels without taking into consideration the surrounding voxels.
– Finally, quantitation methods have to be continuously refined due to new hardware and new acquisition schemes. For example, quantitation of brain HRMAS signals using QUEST has been recently proposed [153].

In spite of numerous publications on the topic, quantitation of MRS data remains an important issue. No satisfactory systematic study of the accuracy of the methods has been performed. One of the obstacles is the lack of gold standard simulated signals which would mimic real-world signals and permit fair comparisons between the methods. In this paper, advantages and drawbacks of the different methods have been depicted and it appears clearly that none of the methods outperforms the others in all cases. However, the choice of the quantitation method should result from the objectives that the analyst pursues (e.g., which data he/she wants to analyze, etc.) and tips are given in that respect (see Section 5).

Acknowledgments

We thank Dirk van Ormondt of Delft University of Technology, Netherlands, for help and discussions, as well as the referees for their suggestions. Research supported by:

– Research Council KUL: GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering (OPTEC), IDO 05/010 EEG-IMRI, IOF-KP06/11, several PhD/postdoc and fellow grants;

– Flemish Government:
  • FWO: PhD/postdoc grants, projects, G.0407.02 (support vector machines), G.0360.05 (EEG, Epileptic), G.0519.06 (Noninvasive brain oxygenation), FWO-G.0212.06 (Tensors/Spectral Analysis), G.0302.07 (SVM), G.0341.07 (Data fusion), research communities (iCCoS, ANMM);
  • IWT: Ph.D Grants;
  • Belgian Federal Science Policy Office IUAP P6/04 (DYSOC, ‘Dynamical systems, control and optimisation’, 2007-2011); EU: BIOPATTERN (FP6-2002-IST 508803), ETUMOUR (FP6-2002-LIFESCIHEALTH 503094), Healthagents (IST-2004-KP06/11, several PhD/postdoc and fellow grants; research communities (Data fusion), research communities (ICCoS, ANMM);

– EU: BIOPATTERN (FP6-2002-IST 508803), ETUMOUR (FP6-2002-LIFESCIHEALTH 503094), Healthagents (IST-2004-KP06/11, several PhD/postdoc and fellow grants; research communities (ICCoS, ANMM);

– Flemish Government:
  • FWO: PhD/postdoc grants, projects, G.0407.02 (support vector machines), G.0360.05 (EEG, Epileptic), G.0519.06 (Noninvasive brain oxygenation), FWO-G.0212.06 (Tensors/Spectral Analysis), G.0302.07 (SVM), G.0341.07 (Data fusion), research communities (iCCoS, ANMM);


