Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Model order selection for quantification of a multi-exponential magnetic resonance spectrum

A. Devos, N. Bergans, T. Dresselaers, J. De Brabanter, D.M. Sima,
L. Vanhamme, F. Vanstapel, P. Van Hecke, and S. Van Huffel

Abstract— **Magnetic resonance spectroscopic signals analyzed by time-domain models in order to retrieve estimates of the model parameters usually require prior knowledge about the model order. For multi-exponential signals where a superposition of peaks occurs at the same resonance frequency, but with different damping values, model order selection criteria from information theory can be used. In this study, several generalized versions of information criteria are compared using Monte-Carlo simulation signals. The best criterion is further applied for selecting the model order of experimental ^{13}C glycogen signals.**

I. INTRODUCTION

Quantification of Magnetic Resonance Spectroscopic (MRS) signals can be done parametrically in the time domain by modeling the signal as a superposition of exponentially damped sinusoids. This supposes the model order to be known, which in practice is not the case. The determination of the model order is particularly difficult in the case that some of the exponentially damped sinusoids have the same frequency. This problem is encountered in the analysis of ^{13}C MRS signals where glycogen (GLY) is a superposition of an unknown number of exponentially damped sinusoids with the same frequency but with different damping factors: a multi-exponential signal [9]. Model order selection criteria would allow an objective evaluation of the model order rather than the operator-biased evaluation of the residue.

The glycogen signals were obtained during a ^{13}C-1 pulse-chase experiment, which followed the glycogen synthesis in a perfused rat liver and mainly consists of two pulse phases and a chase phase [1]. Processing these signals demonstrated that a sum of exponentials was necessary to accurately quantify the changing glycogen signals during the experiment.

Section II shortly describes some background about the considered NMR experiment, while section III mentions several concepts regarding the provided prior knowledge and model order selection. This analysis is then applied to simulation signals as well as experimental signals from the glycogen experiment. The simulation and experimental results are described and discussed in section IV. This includes the evaluation of the model order selection criteria, the optimal model order and the evolution of the signal parameters during the experiment.

A. Devos, J. De Brabanter, D.M. Sima, L. Vanhamme, S. Van Huffel are with Department of Electrical Engineering, SCD-SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Heverlee (Leuven), Belgium sabine.vanhuffel@esat.kuleuven.be

N. Bergans, T. Dresselaers, F. Vanstapel, P. Van Hecke are with the Biomedical NMR unit, Faculty of Medicine, Katholieke Universiteit Leuven, Herestraat 49, B-3000 Leuven, Belgium

II. MATERIAL

Glycogen synthesis from glucose in perfused livers of both fed and starved rats was studied by ^{13}C NMR. A so-called ^{13}C-1 pulse-chase experiment was carried out, followed by NMR. This experiment involves exposure to 100% ^{13}C enriched glucose (100% NMR visible), noted as ^{13}C-1 glucose, followed by a chase phase with normal glucose (unlabeled, *i.e.*, natural ^{13}C abundance, 1.1% NMR visible).

Spectra were acquired and preprocessed as described in [2], in a 4.7T, 30cm wide bore horizontal magnet equipped with a Biospec spectrometer (Bruker Spectrospin, Karlsruhe). Waltz broad-band ^{1}H-decoupled relaxed ^{13}C NMR spectra were acquired at 50.4 MHz in blocks of 512 scans using 90° pulses (pulse length 75 μs), a dwell time of 45 μs, TR=846 ms and a 270 μs dead time.

Experimental signals were available from fed and starved rats for the two pulse phases and the chase phase, at several noise levels. The SNR was increased by averaging the spectra from different livers (fed: 4 experiments, starved: 7 experiments) corresponding to the same time point in the experiment. To further increase the SNR, in certain cases a so-called sliding rule was applied in which three consecutive spectra within the sequence were averaged. The central time point was shifted (slided) over the consecutive spectra. The sliding rule was not applied at the transition between different phases of the experiment, to avoid mixing signals from different phases. (For details and examples, see [2].)

III. METHODS

We consider only models with one, two and three peaks for the glycogen resonance frequency, noted as M_1, M_2 and M_3, respectively. All models are obtained by time-domain modeling using AMARES [11].

A. Prior knowledge

Quantification of the signals was carried out in order to approximate the resonances from glycogen as well as from the neighboring α- and β-glucose. The FID was modeled as a Lorentzian model, which is a sum of complex damped exponentials:

$$y_n = \bar{y}_n + e_n = \sum_{k=1}^{K} (a_k e^{j\phi_k})\, e^{(-d_k + j2\pi f_k)t_n} + e_n, \qquad (1)$$

where the model order K corresponds to the number of different resonances and $j = \sqrt{-1}$. The term $\bar{y}_n$ is the noiseless signal and e_n is the Gaussian noise term. The parameters of the different components were labeled as a_k

(amplitude), d_k (damping), f_k (frequency) and ϕ_k (phase), where $k = 1, \ldots, i$ ($i = 1$ for M_1, $i = 2$ for M_2, or $i = 3$ for M_3) are the peaks of glycogen, and $k = i+1, i+2$ are the peaks of β- and α-glucose. Prior knowledge was used:

- The first 5 points of the FID were truncated to filter out ringing (dwell time is 45 μs).
- All phases were constrained to be equal. The frequencies were constrained between 100 and 102 ppm.
- Amplitudes of the glycogen peak were unconstrained. Amplitudes of β- and α-glucose were also estimated.
- The dampings of β and α-glucose were set to be equal. For the chase phase, soft constraints can be imposed on the dampings (linewidth between 46 and 52 Hz). The linewidth lw_k is related to the damping as $lw_k = d_k/\pi$.
- The frequency of β-glucose was constrained between 95.5 ppm and 97.5 ppm, while that of α-glucose was imposed to be between 91.5 ppm and 93.5 ppm.

B. Model order selection for MRS quantification

Consider the selection of a model from a set, $M_i \in \mathcal{A}$, where M_i is a nonlinear model for experimental data:

$$y_n = g_i(t_n, \theta_i) + e_n,\ n = 0, \ldots, N-1. \quad (2)$$

The vector θ_i contains all model parameters and $g_i(t_n, \theta_i)$ is the model function depending on t_n and determined by θ_i, and e_k are independently and identically distributed errors with mean 0 and constant variance σ_e^2.

The model M_i that minimizes the following Generalized Prediction Error (GPE) function [8] is selected:

$$RSS_i + \lambda d_{\text{eff},i} \hat{\sigma_e}^2, \quad (3)$$

where RSS_i is the residual sum of squares, $d_{\text{eff},i}$ the effective number of parameters (for nonlinear and complex models) and $\hat{\sigma_e}^2$ is an estimate of the error variance σ_e^2. The first term in (3) indicates how well the model is fitting the signal, while the second term penalizes the complexity of the model. Under the above assumptions concerning the distribution of the error term e_n, and due to having equidistant samples in the time domain, an unbiased estimator of the error variance σ_e^2 can be calculated [4]:

$$\hat{\sigma_e}^2 = \frac{1}{6(N-2)} \sum_{k=2}^{N-1} \left(2y_k - y_{k-1} - y_{k+1}\right)^2, \quad (4)$$

The effective number of parameters d_{eff} generally differs from the true number of parameters d in the vector θ and depends on the nonlinearity of the model, the prior model constraints and the amount of model bias [8]. (For linear models $d_{\text{eff}} = d$.) Optimization-based methods like AMARES provide an estimate of the covariance matrix $\hat{\mathbf{\Gamma}}$ of the unknown variables. This can be used in information criteria for the computation of the "hat" matrix $\mathbf{S} = \hat{\mathbf{\Gamma}}\hat{\mathbf{\Gamma}}^{\dagger}$, which relates the model $\hat{\mathbf{y}}$ and the data $\mathbf{y}$ as $\hat{\mathbf{y}} \approx \mathbf{S}\mathbf{y}$. The effective number of parameters d_{eff} is equal to $(4\,\text{trace}(\mathbf{S}) + 2 - d_C)$. The number $d_{C,i}$, $i = 1, 2, 3$ is the number of hard constraints (reflecting the prior knowledge about the model parameters) imposed for the model M_i in AMARES.

Several values of λ can be used in (3), yielding slightly different model selection criteria:

- **GAIC:** for $\lambda = 2$, (3) is known as the Generalized version of Akaike's Information Criterion [7];
- **GBIC:** for $\lambda = \log(N)$, (3) corresponds to the Generalized Bayesian Information Criterion [10];
- **GIClog:** $\lambda = \log(\log(N))$ has also been proposed for autoregressive model order determination [5];
- **GAICC:** a generalized corrected version of AIC [6] is $\min \quad RSS_i + \left(N + \frac{2N * (d_{\text{eff},i} + 1)}{(N - d_{\text{eff},i} - 2)}\right) \hat{\sigma_e}^2$.

For each of these criteria, the best model was given as the one with the lowest criterion value.

C. Application on Monte-Carlo simulations

The four different criteria were applied on simulated ^{13}C signals with resonances of glycogen and α- and β-glucose. Signals were simulated with one, two, as well as three Lorentzians at the resonance frequency of glycogen, but with different dampings. Each of these simulation signals can be approximated as a model M_1, as well as a multi-exponential model M_2 or M_3. The amplitude a_4 of β-glucose was set to be 4/10 of the total amplitude of glycogen, unless mentioned otherwise. The amplitude of α-glucose was set to be $a_5 = 2/3a_4$, which corresponds to the equilibrium situation of D-glucose in solution.

Simulated GLY-peaks: Assume a signal with 2 components at the same frequency. Intuitively, the model selection method is expected to have a higher performance when the damping difference is larger or the amplitude of the broad component, relative to that of the small component, is larger. The influence of the difference between the dampings as well as the amplitude ratio was tested by considering four different combinations of dampings (linewidths lw) and amplitudes (a) (items s1-s4 below). Additionally, in order to investigate whether AMARES was able to detect three components of glycogen, two simulations with 3 peaks at the same frequency were constructed (items s5-s6).

s1) Large difference between the dampings ($lw_2 \gg lw_1$) and a relatively large amplitude a_2 ($a_2/a_1 = 1$): $lw_1 = 85Hz, lw_2 = 400Hz; a_1 = a_2 = 0.5 * 10^4$.

s2) $lw_2 \gg lw_1$, $a_2/a_1 \approx 0$: $lw_1 = 85Hz, lw_2 = 400Hz; a_1 = 0.9 * 10^4, a_2 = 0.1 * 10^4$.

s3) Small damping difference $lw_2 > lw_1$, and $a_2/a_1 = 1$: $lw_1 = 63Hz, lw_2 = 148Hz; a_1 = a_2 = 0.5 * 10^4$.

s4) $lw_2 > lw_1$, $a_2/a_1 \approx 0$: $lw_1 = 63Hz, lw_2 = 148Hz; a_1 = 0.9 * 10^4, a_2 = 0.1 * 10^4$.

s5) $lw_1 = 63Hz, lw_2 = 148Hz, lw_3 = 637Hz; a_1 = 0.4 * 10^4, a_2 = a_3 = 0.3 * 10^4$.

s6) $lw_1 = 63Hz, lw_2 = 148Hz, lw_3 = 637Hz; a_1 = 0.4 * 10^4, a_2 = a_3 = 0.3 * 10^4, a_4 = 0$.

Noise realizations: The criteria were tested by comparing the estimated model order with the true model order. By considering 500 noise realizations this can be done statistically, using the Wilcoxon signed rank test of equality of medians. This test returns the significance of testing that the median difference between two sample series is zero. M_1 is a statistically a better model than M_2 if the median criterion value of M_1 is lower than that of M_2 under a significance level of 0.05. For each group of 50 signals, the Wilcoxon test is applied, resulting in 10 significance tests for each SNR. (The SNR is computed as in [2].)

D. Application on experimental signals

From the results of the Monte-Carlo simulations, the optimal criterion was selected, in order to apply it to the experimental glycogen signals. Quantification was carried out, using M_1, M_2 and M_3, on signals of perfusion livers of starved and fed rats, at different noise levels. Results were compared with those obtained by [2].

IV. RESULTS

A. Monte-Carlo simulations

The results of the Monte Carlo simulations are presented in Table I and explained below.

- s1) For SNR$\geq$12, all criteria gave an almost perfect result. (See also Fig. 1.)
- s2) All criteria, except GBIC, resulted in the correct model order for SNR$\geq$64.
- s3) All criteria resulted in the correct model order for SNR$\geq$48, GAIC and GIClog even gave perfect results for SNR$\geq$24.
- s4) GBIC failed to select the correct model order, while all other criteria gave a good result for SNR$\geq$80.
- s5) GBIC once more had most problems to select the correct model order. For the other criteria also a high SNR$\geq$100 was needed to obtain 90% correct results.
- s6) In comparison to the previous item, similar results were obtained. Nevertheless, now also GBIC obtained 100% correctness at SNR=128, and also the performance of the other criteria increased slightly.

From the simulation results obtained here (and from more extensive experiments in [3, Chapter 7]), we observed that GAICC was usually less biased than GAIC and GIClog and overall showed to be the optimal criterion out of the considered ones for this problem. For the analysis on experimental multi-exponential glycogen signals, we decided to use GAICC for model order selection.

B. Experimental signals

1) Optimal model order: Analysis of the high SNR signals (sliding rule) in the starved state yielded that GAICC selected M_1 as the optimal model for the first pulse phase. Here, models M_2 and M_3 did not converge well (either two or three of the glycogen peaks had the same damping or one or more of the glycogen peaks had zero amplitude). At high SNR (sliding rule), model selection resulted in M_2 as the best model for all time points of the chase phase. Even at lower SNR (6 experiments) a multi-exponential signal detected as M_2 was the optimal model for all (8 out of 8) signals. Although for high SNR signals of the second pulse phase, M_2 was selected as optimal model for the last 5 time points (sliding rule), the analysis with M_3 also resulted in a better model than M_1 for the last 4 time points. Even at lower SNR (7 experiments) multi-exponential signals were detected, as M_2 had the lowest GAICC value for 6 out of 8 signals and M_3 had a lower GAICC value for one signal.

Also in the fed state, signals of the first pulse phase (sliding rule) were best modeled by only one Lorentzian.

TABLE I

RESULTS OF MONTE CARLO SIMULATIONS. FOR EACH SNR, WILCOXON SIGNED RANK TEST RESULTS FOR THE VALUES OF GAIC, GBIC, GAICC AND GICLOG ARE SHOWN AS THE NUMBER OF TIMES (OUT OF 10) THAT THE TRUE MODEL WAS CORRECTLY CHOSEN AS THE BEST MODEL.

		M_1 vs. M_2				M_2 vs. M_3	
Method	SNR	s1)	s2)	s3)	s4)	s5)	s6)
GAIC	8	8	0	1	0	0	0
	16	10	1	6	1	0	0
	32	10	4	10	4	0	0
	64	10	10	10	1	0	2
	128	10	10	10	10	10	10
GBIC	8	0	0	0	0	0	0
	16	9	0	0	0	0	0
	32	10	1	8	0	0	0
	64	10	1	10	0	0	0
	128	10	10	10	1	4	10
GAICC	8	7	0	2	0	0	0
	16	10	0	6	0	0	0
	32	10	2	10	2	0	0
	64	10	10	10	2	0	2
	128	10	10	10	10	10	10
GIClog	8	2	2	2	1	0	0
	16	10	1	8	1	0	0
	32	10	2	10	4	0	0
	64	10	10	10	4	2	2
	128	10	10	10	10	10	10

However, for high SNR (sliding rule) signals from the chase phase and second pulse phase, GAICC selected M_2 as the optimal model. Two components were found for the last 5 (out of 6) time points of the chase phase and for all 6 (out of 6) time points of the second pulse phase.

2) Time course of the linewidth: In the first pulse phase, signals were difficult to model with multiple Lorentzians (M_2 and M_3) without additional constraints on the linewidth. Even at high SNR (sliding rule) no appropriate fit was obtained for the starved as well as the fed state. The average estimates of the linewidths are tabulated in Table II.

For signals in the starved state, there were a few clear differences between the linewidths of the peaks. In the chase phase, M_2 resulted in two Lorentzians, of which the first one had a smaller linewidth than the Lorentzian of M_1, while the second one was 4 to 5 times broader. The same observation was made for the linewidths of the first two Lorentzians of M_3, while a third Lorentzian had a slightly higher linewidth than the second one. For signals in the second pulse phase, a similar pattern was observed, although there was a stronger difference between the second and third peak of M_3. The linewidth estimates of the first two components of M_2 and M_3 were lower than the estimates of the chase phase. Presumably newly synthesized ^{13}C-1 glycogen is characterized by a narrower line.

In the fed state, differences between the small and the broad peaks of M_2 were less expressed. We observed that the peaks were smaller than those in the starved state. In the chase phase, the linewidth of the second component of M_2 showed a slightly increasing time course for the signals in the starved as well as in the fed state.

3) Time course of the signal intensity (amplitude): During both pulse phases and both states (starved and fed), the total signal intensity was linearly increasing in time, for all the selected models. In the chase phase, it was linearly

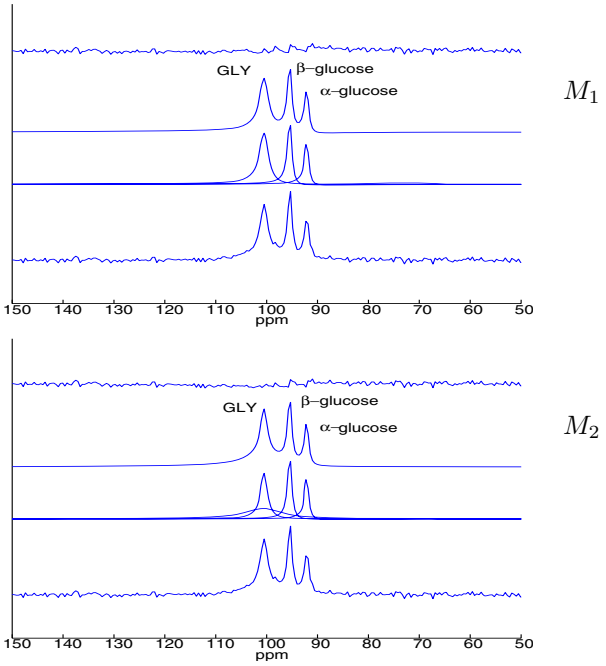

Fig. 1. Quantification result of AMARES on a simulated glycogen ^{13}C NMR signal at SNR=24. The true signal contained two glycogen peaks and a β- and α-glucose peak and was best modeled by M_2. The glycogen peaks had equal amplitudes and largely differing linewidths. Each figure contains respectively (from bottom to top) the original signal, the different model components, the model and the residue.

TABLE II

AVERAGES AND THEIR STANDARD DEVIATIONS OF THE ESTIMATES OF THE LINEWIDTH USING MODELS M_1, M_2 AND M_3 FOR ALL SIGNALS FOR WHICH AN APPROPRIATE FIT WAS OBTAINED. THE FIRST THREE COLUMNS DENOTE THE PHYSIOLOGICAL STATE OF THE RAT, THE MODEL USED AND THE NUMBER OF GLYCOGEN PEAKS. THE LAST THREE COLUMNS MENTION THE ESTIMATED LINEWIDTHS.

state	model	k	linewidth (Hz)		
			pulse phase 1	chase phase	pulse phase 2
starved	M_1	1	70.6 ± 1.2	107.0 ± 3.6	79.7 ± 1.0
	M_2	1		84.4 ± 3.6	67.7 ± 1.8
		2		397.9 ± 58.5	219.7 ± 45.8
	M_3	1		79.6 ± 1.0	56.9 ± 5.6
		2		301.4 ± 86.6	92.8 ± 7.2
		3		363.9 ± 1.8	369.5 ± 54.4
fed	M_1	1	53.3 ± 5.0	72.4 ± 3.0	70.7 ± 1.5
	M_2	1		60.8 ± 2.3	60.5 ± 1.8
		2		126.2 ± 24.8	183.7 ± 13.8

decreasing, as expected [1]. The use of multi-exponential signals in the chase phase and the second pulse phase yields larger values for the total signal intensities, which is underestimated by M_1. (See also Chapter 7 of [3].)

V. CONCLUSIONS

Several model order selection criteria from information theory were applied for selecting the correct model order of multi-exponential MRS signals. Models were obtained with three different model orders by the quantification algorithm AMARES. Simulation results indicated that the results depend on the SNR and the differences in damping and intensity of the multiple glycogen peaks; especially at medium to high SNR or in favorable situations of realistic glycogen peaks with equal intensities and large damping differences, the correct model order was found. High SNR is crucial in order to have convergence of the algorithm when multiple overlapping peaks are used to approximate the signal. Based on this work one could set up a model with a broad and narrow component and apply this prior knowledge when fitting the glycogen signal at lower SNR rather than using one peak, which will underestimate the amount of glycogen.

GAICC as optimal criterion for simulations was applied to experimental glycogen signals. In general, the conclusions of [1] were strengthened by the implementation of a model order selection method. It is confirmed that large linewidths are needed to correctly fit the glycogen signal. These large linewidths are due to the slow motion of large particles, such as the glycogen macromolecule. For signals with convergent models, the analysis with multi-exponential models is needed to approximate glycogen signals during a pulse-chase experiments and to capture the time course of the parameter estimates as accurately as possible. In addition, results of the information criteria are obtained very fast, which is in contrast to cross-validation and bootstrapping techniques.

ACKNOWLEDGMENTS

This research work was carried out at the ESAT laboratory and the Biomedical NMR Unit, University Hospital of the Katholieke Universiteit Leuven, in the framework of the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian Federal Science Policy Office (IUAP Phase IV-02 and IUAP Phase V-22), the EU funded projects BIOPATTERN (EU network of excellence; contract no. 508803), eTUMOUR (FP6 integrated project; contract no. 503094), and Healthagents (IST-2004-27214), the Concerted Action Project AMBIORICS of the Flemish Community. AD research funded by a Ph.D grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). LVH was a postdoctoral researcher with the National Fund for Scientific Research FWO - Flanders.

REFERENCES

[1] N. Bergans. *Biochemical and ^{13}C NMR spectroscopy study of pathways involved in glycogen turnover*. PhD thesis, Faculty of Medicine, K.U. Leuven, Belgium, 2003.

[2] N. Bergans, T. Dresselaers, L. Vanhamme, P. Van Hecke, S. Van Huffel, and F. Vanstapel. Quantification of the glycogen $^{13}C-1$ NMR signal during glycogen synthesis in perfused rat liver. *NMR Biomed.*, 16:36–46, 2003.

[3] Andy Devos. *Quantification and Classification of Magnetic Resonance Spectroscopy Data and Applications to Brain Tumour Recognition*. PhD thesis, Faculty of Engineering, K.U. Leuven, Belgium, 2005.

[4] T. Gasser, L. Sroka, and C. Jennen-Steinmetz. Residual variance and residual pattern in non-linear regression. *Biometrika*, 73(3):625–633, 1986.

[5] E.J. Hannon and B.G. Quinn. The determination of the order of autoregression. *J. Roy. Stat. Inst. B*, 41:190–195, 1979.

[6] C.M. Hurvich and C.L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76:297–307, 1989.

[7] S. Konishi and G. Kitagawa. Generalised information criteria in model selection. *Biometrika*, 83(4):875–890, 1996. in Chapter 7 of [3]

[8] J.E. Moody. The effective number of parameters: an analysis of generalization and regularization in non-linear learning systems. In *Advances in Neural Information Processing Systems 4, Proceedings of the 1991 NIPS Conference*, pages 847–854.

[9] K. Overloop, P. Van Hecke, F. Vanstapel, H. Chen, S. Van Huffel, A. Knijn, and D. van Ormondt. Evaluation of signal processing methods for the quantification of a multi-exponential signal: the glycogen ^{13}C-1 NMR signal. *NMR Biomed.*, 9:315–321, 1996.

[10] G. Schwarz. Estimating the dimension of a model. *Ann. Stat.*, 6, 1978.

[11] L. Vanhamme. *Advanced time-domain methods for nuclear magnetic resonance spectroscopy data analysis*. PhD thesis, Faculty of Engineering, K.U. Leuven, Belgium, 1999.