SPi	Journal Code				Article ID				Dispatch: 28.05.12	CE: Eroy, Carthne
	C	E	M		2	4	5	6	No. of Pages: 10	ME:

Research Article

Journal of CHEMOMETRICS

Received: 4 September 2011, Revised: 30 April 2012, Accepted: 11 May 2012, Published online in Wiley Online Library: 2012

(wileyonlinelibrary.com) DOI: 10.1002/cem.2456

Quantifying brain tumor tissue abundance in HR-MAS spectra using non-negative blind source separation techniques

Anca Ramona Croitor Sava[a,b], Carmen Marie Martinez-Bisbal[c,d], Diana Maria Sima[a,b], Jorge Calvar[e], Vicente Esteve[c,d], Bernardo Celda[c,d], Uwe Himmelreich[f] and Sabine Van Huffel[a,b]

Given high-resolution magic angle spinning (HR-MAS) spectra from several glial tumor subjects, our goal is to differentiate between tumor tissue types by separating the different sources that contribute to the profile of each spectrum. Blind source separation techniques are applied for obtaining characteristic profiles for necrosis, highly cellular tumor and border tumor tissue and providing the contribution (abundance) of each of these tumor tissue types to the profile of each spectrum. The problem is formulated as a non-negative source separation problem. Non-negative matrix factorization, convex analysis of non-negative sources and non-negative independent component analysis methods are considered. The results are in agreement with the pathology obtained by the histopathological examination that succeeded the HR-MAS measurements. Furthermore, an analysis to verify to which extent the dimension of the input space, the input features and the number of sources to be extracted from the HR-MAS data could influence the performance of the source separation is presented. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: blind source separation; HR-MAS; necrosis; cellularity; tumor tissue abundancy

1. INTRODUCTION

Ex vivo high-resolution magic angle spinning (HR-MAS) is a magnetic resonance (MR) spectroscopy technique frequently used for metabolic analysis of tissue samples. It provides significant biochemical information on the metabolites by exhibiting peaks at frequencies specific to the molecular composition of the tissue under investigation. Because HR-MAS MR allows the identification of an important number of metabolites, it has lately been extensively used for characterizing and diagnosing brain tumors [1]. Previous studies have reported that alterations in the metabolite concentrations are correlated with the brain tissue type [2]. Hence, each tissue type can be viewed as having a characteristic metabolic profile corresponding to the chemical composition of the tissue [3].

High-resolution magic angle spinning spectra are obtained from a small tissue biopsy sample; thus, these spectra reflect the local tissue characteristics. A significant variability within the HR-MAS spectral profiles belonging to the same brain tumor tissue type can be observed [4], reflecting the heterogeneity that is a characteristic of brain tumors. Thus, in glial brain tumors, there are frequently necrotic regions found at a microscale level along with viable tumor cells. Moreover, it may also be that the tumor infiltrates normal tissue at the border of the affected area. Hence, the tumor tissue under investigation might present contributions from three different tumor tissue types: necrosis, highly cellular tumor and border tissue that contain tumor and apparently healthy brain cells. The observed spectra are therefore a combination of different constituent subspectra, because the measured signal is the response to the stimulation of the entire tissue sample. The overall gain with which a tissue type contributes to a spectrum is proportional to its abundance, that is, its proportion in the entire mixed tissue sample. This concept can be summarized by describing the spectra available from m samples, which are stacked as n-dimensional row vectors in an $m \times n$ matrix X, where n is the number of observations (or data points) in each spectrum, as in [5]:

$$X = AS \tag{1}$$

* *Correspondence to: A. R. Croitor Sava, Department of Electrical Engineering (ESAT) Division SCD, Katholieke Universiteit Leuven, Leuven, Belgium*
E-mail: anca.croitor@esat.kuleuven.be

a *A. R. Croitor Sava, D. M. Sima, S. Van Huffel*
Department of Electrical Engineering (ESAT) Division SCD, Katholieke Universiteit Leuven, Leuven, Belgium

b *A. R. Croitor Sava, D. M. Sima, S. Van Huffel*
IBBT-K.U.Leuven Future Health Department, Leuven, Belgium

c *C. M. Martinez-Bisbal, V. Esteve, B. Celda*
Departamento de Química-Física, Facultad de Química, Universitat de Valencia, Valencia, Spain

d *C. M. Martinez-Bisbal, V. Esteve, B. Celda*
CIBER-BBN, ISC-III, Universitat de Valencia, Valencia, Spain

e *J. Calvar*
FLENI, Buenos Aires, Argentina

f *U. Himmelreich*
Biomedical MRI Unit/Molecular Small Animal Imaging Center, Department of Imaging and Pathology, Katholieke Universiteit Leuven, Leuven, Belgium

UNCORRECTED PROOF

Copyright © 2012 John Wiley & Sons, Ltd.

where S is a $k \times n$ matrix whose $k = 3$ rows are the unknown pure necrosis, highly cellular tumor and border tissue spectra (also referred to as sources). A, an $m \times k$ matrix, contains the amounts, or abundances, of the constituent pure tissue sources in each sample.

At present, histopathological analysis is the standard procedure routinely used to reveal the microheterogeneity of a tumor tissue [6]. Because histopathology is time and effort expensive, it is interesting to obtain a similar separation using HR-MAS information. When classifying HR-MAS spectra, a common procedure is to assign each spectrum to a certain tissue class without taking into consideration the tissue microheterogeneity.

The purpose of the study is to identify the sources of the different pure tissues types, S, in glial tumors and to estimate the amounts/abundances, A, of each source. This problem is formulated as a blind source separation (BSS) problem. BSS techniques have been previously considered when analyzing MR spectroscopic data because they have shown some interesting properties such as a straightforward interpretation and analysis of the data without the need for prior knowledge (no representative training set or representative models are required), a better interpretation of the complexity of the data by allowing one to resolve the partial volume effect caused by the heterogeneity of the brain tissue, decomposing the mixture of the spectra under investigation and easily interpretable results for the clinicians. Previous studies mostly refer to the analyses of *in vivo* single voxel MR spectroscopy (MRS) [7,8] and multi-voxel MR spectroscopic imaging brain data [9–11]. Because it is very difficult, whether not impossible, to verify if the extracted sources are indeed corresponding to spectra of pure tissue in the brain, an empirical validation of the results was proposed in all these studies. Although for healthy tissue and necrosis, it is possible to acquire spectra from voxels filled with only one tissue; for highly cellular tumor tissue, there is a weak probability to obtain such a direct validation. In the proposed study, several BSS techniques are applied to *ex vivo* HR-MAS data to obtain the pure tissue sources and quantify their abundance in the analyzed spectra. With HR-MAS, a straightforward validation of the results is available, because a histopathological analysis was performed post-HR-MAS acquisition on the same tissue samples. Histopathology, which is often considered the ground truth for brain tissue typing, provides detailed information on the mixture of tissues present in the data under investigation, as well the abundance of each tissue.

An important constraint considered in this study is non-negativity. This solution is motivated by the nature of the NMR magnitude spectra, where one has to deal with non-negative signals. BSS approaches for non-negative sources have been recently successfully used in many applications where the sources to be separated are of non-negative nature, for example, biomedical imaging [12], analytic chemistry [13], hyperspectral imaging [14] and biomedical signals [10]. The way to exploit the non-negativity constraint in non-negative BSS (nBSS) techniques depends on other specific properties of the data under analysis. Therefore, numerous nBSS alternatives have been proposed. A class of nBSS methods utilizes the statistical property that the sources are mutually uncorrelated or independent. This class includes among others non-negative independent component analysis (nICA) [15–18] and Bayesian positive source separation [19]. Another class of nBSS methods is represented by implementations that require no assumption on source independence or uncorrelation. One such nBSS approach is the non-negative matrix factorization (NNMF). This method explicitly imposes source non-negativity and even mixing matrix non-negativity.

Compared with the previous studies on NMR spectroscopic data using BSS techniques [7–11], where only the performance of one method was reported; in this manuscript, a comparative study is proposed where three different methods based on different mathematical principles but having in common the non-negativity constraint are considered. NNMF, nICA and a convex analysis for blind separation of mixtures of non-negative sources (CAMNS) [20] are applied on the full magnitude HR-MAS spectra and on metabolic features extracted from the spectra. In this way, characteristic metabolic profiles for necrosis, highly cellular tumor and border tumor tissue are obtained, and the abundances of each tumor tissue to the profile of the spectra are provided. The study has a multifold purpose and evaluates

- which of the considered nBSS methods performs best;
- the nBSS methods' performance on the full magnitude HR-MAS signals versus on metabolic features extracted from the spectra;
- how the results are influenced by the choice of the parameters for the considered nBSS methods, mainly focusing on the number of samples, m, considered for the input matrix X and the number of sources to be calculated, k;
- whether the nBSS methods could provide results similar to histopathology, which could further be easily interpretable by the clinicians.

2. MATERIALS

2.1. Data acquisition

The HR-MAS spectra of brain tissue samples from 52 adult patients presenting a glial tumor with different degree of malignancy were considered. Three hospitals provided the samples used in this study: Hospital La Ribera-Alzira (Valencia), Hospital Clínico Universitario (Valencia) and Instituto FLENI (Buenos Aires), all participating as partners in the eTUMOUR project (Web Accessible MR Decision Support System For Brain Tumour Diagnosis And Prognosis, Incorporating *In Vivo* and *Ex Vivo* Genomic and Metabolomic Data. URL: http://www.etumour.net/). On the basis of the standard clinical diagnosis by histopathology, the data were split in three tumor subgroups: 27 grade IV glioblastoma tumor (GBM), 7 grade III glial tumor (GIII) and 18 grade II glial tumor (GII).

The tissue specimens were snap-frozen in liquid nitrogen and stored at $-80\,^{\circ}C$ until the time of spectroscopic analysis when HR-MAS data were acquired with a 1D PRESAT (pulse-and-acquire). HR-MAS experiments were conducted on 29/52 samples using a Bruker Avance DRX 600 spectrometer operating at a ^{1}H frequency of 600.13 MHz. The instrument was equipped with a 4-mm triple resonance $^{1}H/^{13}C/^{15}N$ HR-MAS probe with magnetic field gradients aligned with the magic angle axis. The 23/52 samples underwent HR-MAS experiments using a Bruker Avance DRX 500 spectrometer operating at a frequency of 500.13 MHz. The instrument was equipped with a 4-mm triple resonance $^{1}H/^{31}P/^{13}C$ HR-MAS probe. A single-pulse 90° pre-saturation experiment was acquired in all the samples. Number of transients was 256 collected into 32k data points for all the experiments. Water pre-saturation was used during 1 s along the recycling delay for solvent signal suppression. Spectral width

UNCORRECTED PROOF

 Copyright © 2012 John Wiley & Sons, Ltd.

was 8 and 10 kHz for 600 and 500 MHz NMR spectrometers, respectively. To keep the rotation sidebands out of the acquisition window, samples were spun at 5000 Hz in the DRX 600 spectrometer and at 4000 Hz in the DRX 500 spectrometer. Sample temperature was kept at 4 °C by using the cooling of the inlet gas pressures responsible for the sample spinning. Recycling delays were 1.5 s for all experiments. Each sample was introduced in a 4-mm ZrO_2 rotor fitted with a 50-mL cylindrical insert without any previous D_2O washing.

In the post-HR-MAS experiments, the tissue samples were recovered from the rotor, and 37 out of the 52 samples (22 GBM, 6 GIII and 9 GII) were submitted for quantitative histopathological examinations. This analysis revealed that the samples present different histopathological tissue properties reflected in a variable content of highly cellular tumor tissue, infiltrations with normal tissue and/or necrotic tissue. Additionally, for each sample, the percentage of each tissue type was calculated by measuring the total area of the biopsy sections within a sample and then delineating the regions of interest. The values are revealed in Table I. This information will be further considered in the validation of our HR-MAS data analysis results.

In a classical classification approach, each of the 52 HR-MAS spectra analyzed here would have been assigned to one and only one of the considered tumor subgroups without identifying the contributions from necrotic tissue and the infiltrations with normal tissue within each case. However, a strong correlation between the histopathological tissue properties and metabolic profiles coming from *ex vivo* HR-MAS has already been observed in glial tumors [4]. Therefore, for diagnosis purposes, it would be useful to decompose the HR-MAS spectra and identify the spectral sources reflecting pure highly cellular tumor tissue, border tissue (with infiltrations from normal brain tissue) and/or necrotic tissue and then determine their abundance in each spectrum. In particular, this will also address such difficult questions as the grading of glial tumors (e.g., necrosis is a marker for high grade tumors) or differentiating metastasis from glioblastoma (infiltration is a marker of glioblastoma).

2.2. Data processing

The complex-valued HR-MAS time-domain signals (i.e., before Fourier transformation to obtain spectra) were preprocessed as follows: signals were truncated from 8120 to the first 2048 points to reduce the computational load, the water components were removed by Hankel Lanczos singular value decomposition with partial reorthogonalization method [21] and baseline was corrected as follows. An apodization function, containing the broad baseline components, was computed by point-wise multiplying the signal in the time-domain with an exponentially decaying function. The resulting spectrum then mainly contained the broad baseline components, which were subsequently subtracted from the original spectrum, yielding the preprocessed signal. Magnitude spectra were then computed by taking the absolute value of the Fourier transformed time-domain signals. The spectra were then normalized (divided by the L2 norm of the spectrum between 0.5 and 4.2 ppm). Small variations in the chemical shift of peaks between HR-MAS experiments have been noticed. Therefore, all spectra were aligned with respect to the alanine (Ala) doublet at 1.47 ppm [22]. Further, we truncated the magnitude spectra to keep the frequency interval 0.25–4.2 ppm, where the contributions of the metabolites of interest are present. With this step, the signal is reduced from 2048 to 716 points.

Table I. Percentage of tissue contributions for each case according to the histopathological study

Necrosis	Tumor	Border	Collagen	Normal	Index
GBM					
80	20	0	0	0	2
0	0	0	0	100	3
10	70	0	20	0	4
0	0	5	0	95	5
100	0	0	0	0	6
0	0	100	0	0	7
10	90	0	0	0	8
0	90	0	10	0	9
0	90	0	10	0	10
0	90	0	10	0	11
75	0	0	25	0	12
0	50	0	50	0	13
0	20	80	0	0	14
20	50	0	30	0	15
0	0	10	90	0	16
25	75	0	0	0	21
25	75	0	0	0	22
100	0	0	0	0	23
30	70	0	0	0	24
0	0	100	0	0	25
20	80	0	0	0	26
30	70	0	0	0	27
GIII					
50	50	0	0	0	28
0	20	80	0	0	29
0	25	35	40	0	30
0	80	0	20	0	32
0	20	0	80	0	33
0	75	0	25	0	34
GII					
0	0	0	0	100	35
0	50	50	0	0	36
0	80	0	20	0	37
0	0	0	0	100	38
0	80	20	0	0	45
0	50	0	50	0	48
0	80	20	0	0	49
0	20	0	80	0	50
0	80	20	0	0	51

The index represents the number assigned at acquisition.

3. METHODS

3.1. Non-negative matrix factorization

The NNMF is a statistical technique that reveals hidden factors within a dataset of signals. Given a non-negative matrix X of size $m \times n$, NNMF finds two matrices A and S that minimize the function

$$\min f(A,S) = \frac{1}{2}\|X - AS\|_F^2, \text{ subject to } A, S \geq 0 \qquad (2)$$

where $\|\bullet\|_F$ is the Frobenius norm and $A, S \geq 0$ means that the elements in A and S are non-negative.

UNCORRECTED PROOF

Copyright © 2012 John Wiley & Sons, Ltd.

For this study, an NNMF implementation with sparseness constrains (NNMFSC) on the sources, S, is considered as in [23]. Thus, the following optimization problem is solved:

$$\min_{A,S}\left\{\|X-AS\|_F^2+\eta\|A\|_F^2+\beta\sum_{j=1}^{n}\|S(:,j)\|_1^2\right\}\text{ subject to }A,S\geq 0 \tag{3}$$

where $S(:,j)$ is the jth column vector of S, $\eta>0$ is a parameter meant to control the size of $\|A\|_F^2$ and $\beta>0$ balances the trade-off between the accuracy of the approximation and the sparseness of S. For this study, the results for $\eta=0.3$ and $\beta=0.1$ are presented, because for these values, we observed best performance when comparing a range of values for these parameters.

The NNMF is not a unique decomposition, because it may result in indeterminacy of the sources and of the mixing matrix. Thus, the algorithm proposed in [23] is an alternating non-negativity constrained least squares implementation. It begins with the initialization of A with non-negative values, and then several iterations (up to a maximum of 100) are performed till convergence [23]:

$$\min_{S}\left\|\begin{pmatrix}A\\ \sqrt{\beta}\,e_{1\times k}\end{pmatrix}S-\begin{pmatrix}X\\ 0_{1\times n}\end{pmatrix}\right\|_F^2\text{ subject to }S\geq 0 \tag{4}$$

$$\min_{A}\left\|\begin{pmatrix}S^T\\ \sqrt{\eta}\,I_k\end{pmatrix}A^T-\begin{pmatrix}X^T\\ 0_{kxm}\end{pmatrix}\right\|_F^2\text{ subject to }A\geq 0 \tag{5}$$

where k is the number of sources, $e_{1\times k}$ is a row vector with all components equal to one and $0_{1\times n}$ is a 0 vector. I_k is an identity matrix of $k\times k$, and $0_{k\times m}$ a zero matrix of size $k\times m$. Each sub-problem of the type [4] or [5] is solved with a non-negative linear least squares active set method [24].

3.2. Convex analysis of mixtures of non-negative sources

Another recently developed nBSS framework that accounts for sparsity is the CAMNS [20]. CAMNS is deterministic, requiring no source independence assumption, the premise that can be found in many existing (usually statistical) BSS frameworks. The development is based on a special assumption called local dominance. This assumption, closely related to sparsity, states that each source exhibits a non-zero value for a certain feature for which all the other sources have a zero value. Under local dominance, the BSS criterion becomes a convex analysis problem, namely finding the extreme points of an observation-constructed polyhedral set. This problem can be solved using linear programming and leads, in this setting, to a unique source separation with probability one. Thus, the source signals can be perfectly identified in a blind fashion [20].

Summarizing from [20], the CAMNS method first constructs from the data matrix X the mean vector $d=\frac{1}{m}1_m^TX$, then the mean centered matrix $U=X-1_md$ and the matrix C containing the first $k-1$ left singular vectors of U. Then, the algorithm proceeds at finding the extreme points of the polyhedral set Ω containing all non-negative sums between the mean vector d and the columns of C. Because of the local dominance assumption on the sources, it can be proven that the k extreme points of Ω are precisely the k sources that we seek. The extreme points of Ω can be computed efficiently, as described in [20].

3.3. Non-negative independent component analysis

Independent component analysis (ICA) is a special case of BSS method in which the goal is to find a linear representation of non-Gaussian data so that the sources are statistically independent, or as independent as possible. Through time, the method found many applications and implementations, and a nice review of the method is provided in [25,26]. Classically, having a matrix X, with ICA, we can decompose $X=AS$, with $X=(x_1, x_2,\ldots, x_m)$. S contains the source signals $S=(s_1, s_2,\ldots, s_k)^T$, and A is the mixing matrix.

Thus, one possible approach to the complexity of the spectra is to assume that the different tissue types are statistically independent. Following this principle, ICA has been previously considered in biomedical signal processing and analyses of single voxel MRS [7,8] and multi-voxel MR spectroscopic imaging data [9–11]. These studies have shown that ICA can provide an effective way to reduce the dimensionality of MR spectra. In this study, a non-negative implementation of ICA (nICA), which is based on a gradient algorithm operating on pre-whitened data, as suggested by [17], is considered. This approach is related to the "nonlinear PCA" algorithms investigated by Oja [27]. The independent sources are considered to be non-negative and well grounded, a concept proposed by Plumbley [15]. The algorithm is guaranteed to find a permutation of the well-grounded, non-negative sources. The idea behind this is that an orthogonal rotation of the whitened observation vector into non-negative outputs gives a positive permutation of the original sources. The proposed cost function minimum coincides with non-negative solutions [17].

3.4. Application to HR-MAS NMR data

3.4.1. The feature input space

The proposed nBSS techniques are applied on the magnitude HR-MAS spectra and on sets of features obtained from the spectra. For the spectra, $n=716$ points representing the magnitude HR-MAS signal in the region of interest between 0.25 and 4.2 ppm were used. For the feature case, two experiments are designed. One where the concentration of 19 most visible metabolite peaks (Table II) are considered ($n=19$) and one where the concentration of the most representative metabolites for characterizing histopathological tissue properties ($n=8$) [4,6] are used. Chemical shift regions were assigned on the basis of 2D NMR spectra and literature reference data [6,28,29]. Although visible, Ala and lactate were not added to the set because of their sensitivity to snap-freezing the samples, the temperature conditions and the unavoidable period of ischemia during the biopsy procedure, which significantly alters them when HR-MAS is used [2,30].

Peak integration, a feature reduction method typical in NMR analysis, was used for extracting the concentration of the considered metabolites. Namely, the highest point in the area to be integrated was identified for each metabolite, then the area bounds were fixed for each metabolite individually to those ppm values at which peak slopes return to baseline but keeping symmetric intervals with respect to the highest point. This procedure aims at ensuring that the integrated values would measure mostly one metabolite and would minimize the contribution from small underlying signals from other molecules that would overlap.

UNCORRECTED PROOF

 Copyright © 2012 John Wiley & Sons, Ltd.

Table II. The list of main contributing metabolites and their chemical shift regions used for integration

Metabolite name	Metabolite abbreviation	Integration	Experiment $n=19$	Experiment $n=8$
Lipid	"Lip1"	0.84–0.91	X	X
Lipid	"Lip2"	1.255–1.285	X	X
Lipid	"Lip3"	2.79–2.81	X	X
N-acetyl-aspartate	"Naa"	2.0–2.015	X	X
Creatine	"Cr"	3.01–3.03	X	X
Choline	"Cho"	3.19–3.205	X	X
Phosphocholine	"Pcho"	3.21–3.22	X	X
Glycero-Pcho	"GPCho"	3.225–3.235	X	X
Glutamate	"Glu"	2.1–2.12	X	
Glutamine	"Gln"	2.12–2.13	X	
Glutamate	"Glu2"	2.3–2.33	X	
Glutamine	"Gln2"	2.41–2.44	X	
Taurine	"Tau"	3.245–3.265	X	
Myo-inositol	"Myo1"	3.49–3.525	X	
Glycine	"Gly"	3.53–3.55	X	
Myo-inositol	"Myo2"	3.58–3.61	X	
Glutamine + glutamate	"Glx"	3.76–3.78	X	
Alpha glucose	"Glcalf"	3.83–3.86	X	
Beta glucose	"Glcbet"	3.86–3.89	X	

UNCORRECTED PROOF

3.4.2. *Parameters to be considered*

The study also questions the robustness of the considered algorithms when they are applied to slightly different datasets, that is, number of samples, *m*, considered for the input matrix *X* (Eq. 1). Thus, firstly, we consider for building the matrix *X* the spectra from 27 measured GBM cases, the class for which histopathological results reported the highest heterogeneity; additionally, this class comprised representative samples (i.e., samples of a single, unmixed, tissue type) for the three considered histopathological tissue properties. Afterwards, we consider as input matrices *X* spectra coming from the available GII and GIII glial tumor subgroups separately. Finally, the whole data set of 52 cases (GII, GIII, and GBM) is considered together as a single dataset.

For every experiment, the number of the sources, *k*, to be extracted has to be chosen. How the value of *k* influences the results and the performance of the algorithms is also investigated. A logical assumption is to choose *k* equal to the number of histological groups, if one presumes that the inter-patient variability is not too large. Thus, for the GBM group, if one requires *S* to have three sources, then these sources should ideally represent the pure sources for necrotic, highly cellular and border tumor tissue. For GII and GIII glial tumor data sets, because the samples mainly contain highly cellular tumor tissue and border tumor tissue, we started from $k=2$. When considering all 52 cases, a *k* value up to 5 was tested. This choice is motivated by the high inter-patient variability, because in addition to the histopathological tissue variability, the samples present different grades of tumor aggressiveness.

3.4.3. *Performance measurement*

Firstly, to evaluate the accuracy of the proposed nBSS methods in identifying the pure tumor tissue sources, a measure of the separation quality is proposed. The nBSS results are validated by comparing them against the histology findings (the standard reference on which diagnosis is based nowadays). To this aim, the spectra coming from samples classified as pure highly cellular tumor tissue, infiltrations with normal tissue and necrotic tissue, respectively (more than 90% of the sample area is covered by one tissue type; Figure 1), are extracted, and their mean is further considered as reference HR-MAS spectra. Then, the correlation coefficient, $\boldsymbol{\rho}$, between the sources obtained with NNMFSC, CAMNS and nICA, respectively, and the reference HR-MAS tissue models is computed:

$$\rho = \frac{\mathrm{cov}(M_k, S_k)}{\sigma_M \sigma_S} \tag{6}$$

where *cov* stands for covariance, M_k for the reference HR-MAS tissue models, S_k for the obtained sources and σ is the standard deviation. The correlation coefficient takes a value between

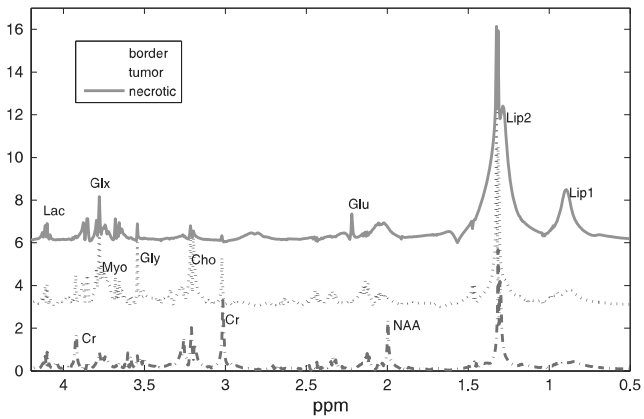


Figure 1. HR-MAS spectra profiles of GBM tissue samples classified according to subsequent histopathological examination as pure highly cellular tumor tissue, infiltrations with normal tissue (border tissue) and necrotic tissue. The illustrated spectra present more than 90% of the sample area covered by one tissue type.

Copyright © 2012 John Wiley & Sons, Ltd.

−1 and 1, where a value close to −1 indicates a negative correlation, close to 0 indicates that sources are uncorrelated and close to 1 that the sources are highly correlated with the reference tissue.

Secondly, with the mixing coefficients representing the contribution (abundance) of each source to the spectra, we can assign each case, on the basis of the highest abundance, to the predominant tissue type class. Additionally, for each case, the percentage of necrosis, tumor cell density and infiltrations with normal tissue is computed. For this, the source with the smallest abundance for each case was set to zero, and then, the percentage of the other two sources is computed. This choice was motivated by the glial brain tumor growth behavior, where necrotic regions are surrounded by highly cellular tumor cells, and infiltrations from normal tissue are present at the border of the affected area. Thus, as observed also in the histopathological study, necrotic and border tissue are not present simultaneously in the analyzed biopsy samples. Moreover, this approach will allow us to verify if the abundance matrix obtained with the considered BSS methods is correct with respect to the tumor growth behavior. The results are then compared with the histopathological study, and the mean squared error (MSE) for the percentage assignment is computed as

$$MSE = \frac{1}{k}\sum_{1}^{k}\left(\overset{\%}{S}_k - \overset{\%}{S}_k^{\text{his}}\right)^2 \tag{7}$$

where $\overset{\%}{S}_k$ is the estimated percentage, $\overset{\%}{S}_k^{\text{his}}$ is the percentage according to the histological analysis and $k = 3$.

4. RESULTS

4.1. Source detection performance

For each experiment conducted in this study, a set of source components, S, has been obtained. All three considered methods, nICA, CAMNS and NNMFSC, provide consistent results and mostly similar sources. In Figure 2, the sources extracted from the GBM group, for $n = 8$ metabolic features and $k = 3$, are shown. For this test, we observe that the mutual correlation coefficients between the similar sources obtained with CAMNS, NNMFSC and nICA are very high for necrotic and tumor tissue types (over 0.95). The border tissue source obtained with nICA presents more contributions from tLips and tCho compared with the source obtained with the other two methods.

As one aspect of the study is to analyze to which extent the feature input space could influence the results, Table III presents the results of the experiments conducted on the GBM group when working on the whole magnitude spectra ($n = 716$) or on a reduced feature input space ($n = 8$ or $n = 19$, as defined in Section 3). The accuracy of each method in extracting pure tissue sources, computed by means of correlation coefficients between the obtained nBSS sources and the reference spectra, is illustrated.

Table III. The correlation between the obtained tissue sources and the reference tissue models for different input feature spaces

	$n = 716$	$n = 19$	$n = 8$
NNMFSC			
Necrotic	0.97	0.97	0.99
Tumor	0.69	0.68	0.72
Border	0.91	0.92	0.92
CAMNS			
Necrotic	0.98	0.97	0.99
Tumor	0.62	0.61	0.71
Border	0.89	0.54	0.88
nICA			
Necrotic	0.81	0.93	0.98
Tumor	0.68	0.67	0.73
Border	0.83	0.51	0.68

All methods identify with a very high accuracy the necrotic tissue pure source, as the correlation with the reference model is mostly higher than 0.95 for all the considered feature input spaces. The sources obtained on the full magnitude spectra ($n = 716$), as well as on the extracted metabolic features, are clearly separated and are in agreement with the conclusions drawn in the literature [6]. We observed that the obtained necrotic tumor tissue source is characterized by elevated peaks of lipids (Lip1 and Lip2), while the rest of the metabolites are present in very low concentrations.

For extracting the border tissue source, NNMFSC outperforms the other two methods. With NNMFSC, the correlation with the histopathology reaches more than 0.9 out of 1. The sources present characteristic high peaks of NAA and Cr. The border tissue source obtained with CAMNS when applied on full spectra reveals some contributions from tCho as well as all high Gly concentrations, which would be typical for a highly cellular tumor tissue. With nICA, the border tissue source obtained on the full magnitude spectra presents contributions from tLips and tCho. This phenomenon is noticeable also on the basis of eight features (Figure 2), where the estimated source for border tissue appears in fact as a mixture of tumor, necrosis and border tissue, although with a predominant contribution from border tissue.

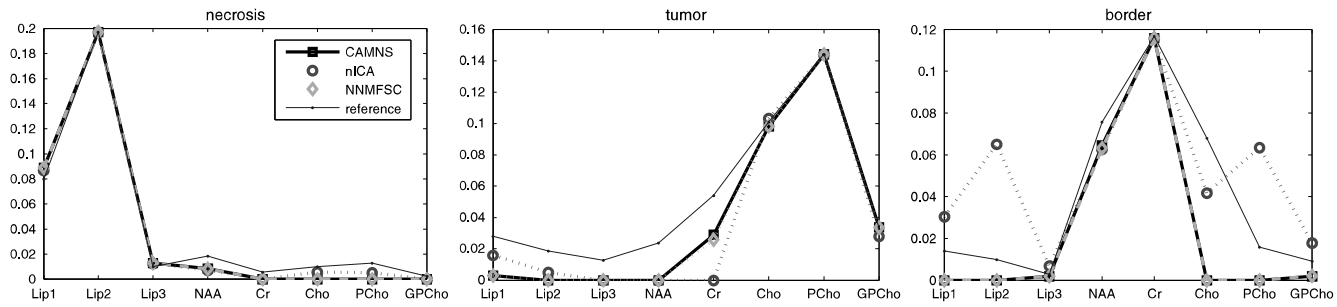


Figure 2. Tumor tissue source profiles identified with CAMNS, nICA and NNMFSC in the case of $n = 8$ metabolic features and the corresponding reference models for necrosis, highly cellular tumor and border tissue.

Colour online, B&W in print

UNCORRECTED PROOF

Copyright © 2012 John Wiley & Sons, Ltd.

Colour online, B&W in print

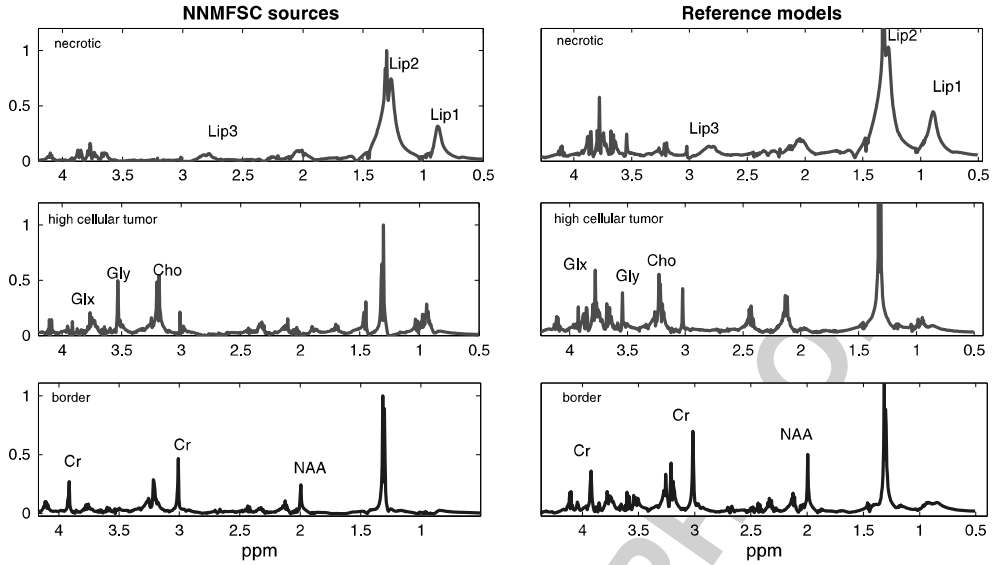


Figure 3. Tumor tissue sources obtained with NNMFSC when applied on the magnitude spectra (left column). Right column plots illustrate reference tissue samples, as indicated by histopathology.

All methods are less accurate in extracting a source that would correspond to a spectrum of pure tumor tissue in the brain. With CAMNS, the tumor tissue source presents levels of Cr that are higher than what is concluded in the literature [11,31,32], and with nICA, Lips are present in the source. Again, overall best performance is obtained with NNMFSC. As one can see in Figure 3, Ala and tCho metabolic components are more elevated in the tumor compared with the other sources.

Looking at the overall performance of the considered methods, results show that with NNMFSC, the obtained sources are the closest to the reference tissue models. Additionally, best overall accuracy for NNMFSC, as shown in Table III, is obtained when considering an input feature space consisting only of the most representative metabolites ($n = 8$), followed closely by the full magnitude spectra input space.

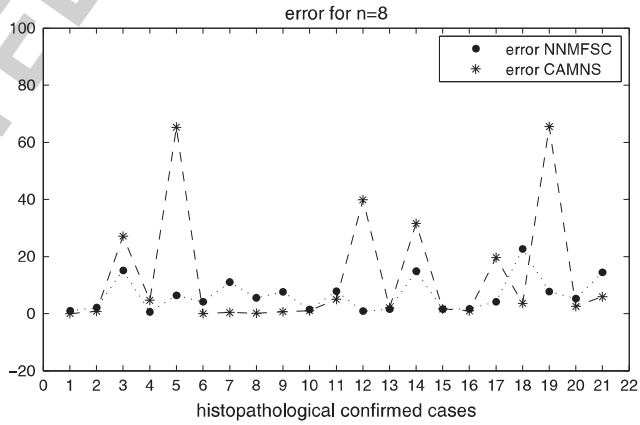


Figure 4. The MSE, computed as in Eq. 7, obtained with NNMFSC and CAMNS, respectively, on feature vectors ($n = 8$). The error corresponding to nICA (not shown) is higher overall.

4.2. Interpretation of the abundance matrix

For each spectrum, we obtain the mixing coefficients representing the contribution (abundance) of each source within the spectrum. Thus, we can identify for each case, on the basis of the highest abundance, the predominant tissue class, as well as to compute the approximate contribution from each tissue type. These mathematical results were compared with the histopathological results. For this comparison, only the histopathological confirmed cases could be considered (22GBM, 6 GIII and 9 GII). NNMFSC can accurately identify the predominant tissue class for 19 out of the 22 histopathologically confirmed GBM cases. With CAMNS and nICA, 16 and 17, respectively, out of 22 histopathologically confirmed GBM cases, were correctly assigned to the predominant tissue class. These results were also reflected in the percentages of contribution of each source to the cases, as a good agreement was obtained between histology and the methods proposed in this study. The MSE is mostly below 10 when considering the NNMFSC algorithm (Figure 4).

Although the predominant tissue within a sample was identified with a high accuracy, the algorithm is less accurate in identifying the minor sources within a mixture of sources. The differences between the histopathological findings and the BSS results are mainly for the cases for which histopathology reports only one source. This is mainly due to the fact that with NNMFSC, the abundance matrix always reports a mixture between at least two sources. Moreover, histopathology reports and quantifies the presence of collagen in our samples. Still, because collagen is not HR-MAS MR visible, this tissue was not considered in our study, and indeed this can represent another drawback in accurately estimating the minor source within a mixture of sources when HR-MAS spectroscopy is used. As suggested by histopathology, our study reports also no mixture between necrosis and border tissue, while tumor tissue is mostly mixing with necrosis.

4.3. Choosing the number of sources *k*

In Figure 5, the results obtained on the dataset represented by all 52 glial tumors are presented when NNMFSC is used for different k values ($k = 2,\ldots,5$). For $k = 2$, the necrotic tissue source is

UNCORRECTED PROOF

Copyright © 2012 John Wiley & Sons, Ltd.

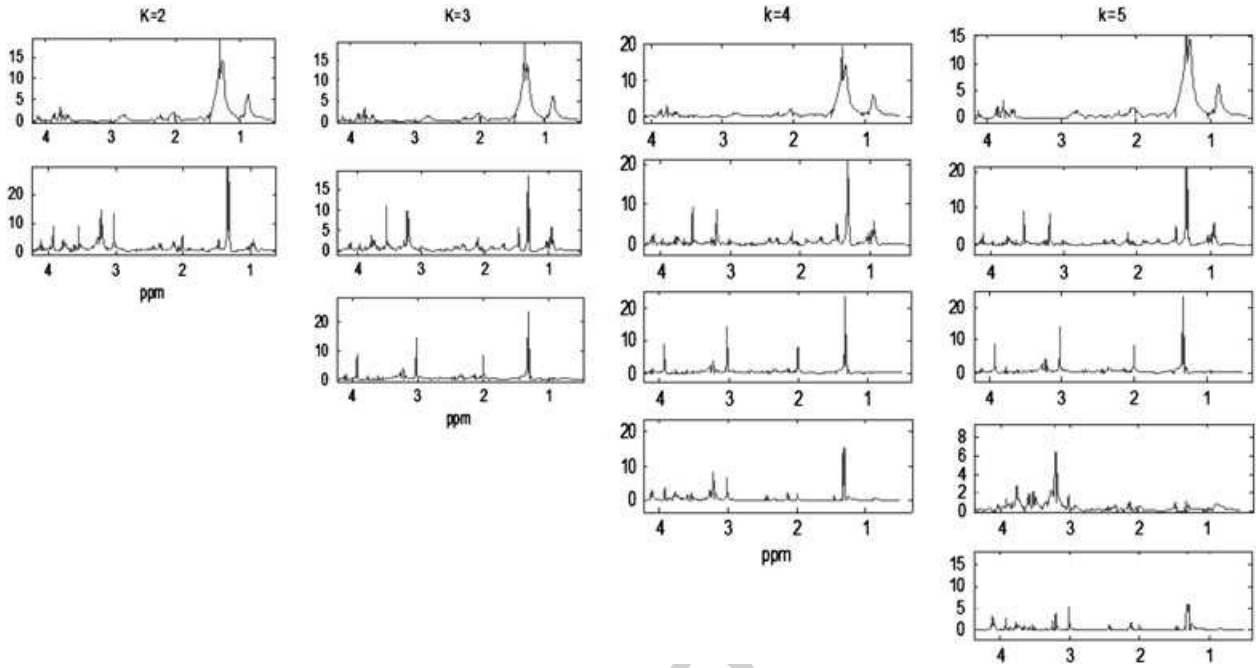


Figure 5. NNMFSC components calculated for the full dataset of 52 magnitude spectra for $k = 2$, 3, 4 and 5.

accurately detected. The second source presents characteristics that reveal a mixture between border and highly cellular tumor tissue. For $k = 3$, the three obtained sources present a high correlation with the reference models for pure necrotic, highly cellular and border tumor tissue, respectively. For $k > 3$, necrotic and border tissue remain relatively stable, while the other sources cannot be clinically or histopathologically interpretable. First source detectable is the one corresponding to necrotic tissue. Its profile does not change radically for different k values. Thus, adding another source does not influence the results. For $k = 3$, all the obtained sources present a histopathological meaning. These results confirm our assumption that the best k value is the one that reflects the number of histopathological sources present in the data. By choosing a smaller k, one would obtain impure sources that would present contribution from more than one tissue type, while by choosing a higher k, one would obtain sources that do not have a histopathological interpretation.

These results are also confirmed by similar experiments performed when considering each glial tumor subgroup separately. For the GBM dataset, the most meaningful sources were also obtained when considering $k = 3$. For the GII and GIII subgroups, best performance is given for $k = 2$. This can be explained by the fact that the necrotic tissue type is not represented in these subgroups, as revealed by the histopathological analyses (Table I). The two sources extracted from GII and GIII groups, respectively, are highly correlated with border and highly cellular tumor tissue models.

Additionally, a relatively high similarity can be observed between the sources obtained with NNMFSC on the different tumor groups or on all 52 spectra. See Figure 6, where plots displaying the sources extracted from all cases and from each glial tumor subgroup, respectively, are illustrated. The metabolite feature space is represented by the eight most representative metabolites. F6

Thus, the performance of the algorithm is not influenced by the number of available cases used as input matrix. Nevertheless, the method will always identify only those histopathological classes that are well represented in the data.

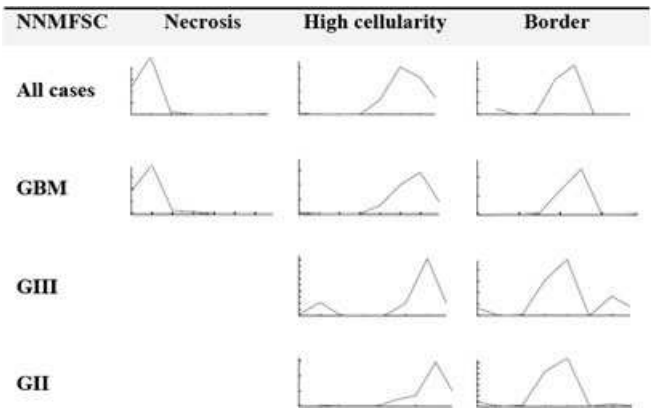


Figure 6. NNMFSC components calculated for eight metabolic features for GII, GIII, GBM and all glial tumor datasets.

UNCORRECTED

Colour online, B&W in print

 Copyright © 2012 John Wiley & Sons, Ltd.

5. DISCUSSION

In this study, nBSS methods are used for obtaining characteristic profiles for each tissue subtype and their abundance within a spectrum. The interpretation of the results is strongly supported by a histopathological study. We show that NNMFSC, CAMNS and nICA can reliably answer the problem of source separation when analyzing HR-MAS data. Still, NNMFSC overall outperforms the other two methods, followed very closely by CAMNS. The similarity in the results obtained with NNMFSC and CAMNS could be due to the fact that the methods rely on the same principles, that is, they do not assume any independence or correlation, but sparsity is imposed. Still, with CAMNS, the assumption of sparsity (namely the local dominance property) is too strong in the sense that when a source such as border and highly active tumor exhibits a non-zero value for a certain feature, all the other sources must have a zero value. On the other hand, this assumption is appropriate for the necrotic tissue source, which is characterized by high levels of lipids, while the other metabolic components are mostly embedded in noise and therefore are close to zero. With NNMFSC, a trade-off between the accuracy of the approximation and the sparseness of S is allowed. This is mainly important because for some metabolites, it can happen that they exhibit peaks of different concentration in more than one source.

For nICA, one imposes a strong assumption of independence and consequently uncorrelation. Taking into consideration the nature of HR-MAS data, where correlations in the metabolic profiles may still be present to some degree between the spectra from different tissue types, such an approach is not the most suitable for our problem, as could be concluded from the results. In reality, it would be hard to identify a 100% pure highly cellular tissue in glial tumors. This tissue type will mostly exhibit some contribution from normal tissue or some necrotic cells. Moreover, the metabolites describing the histopathological tissue properties are correlated as shown in [4]. Therefore, the extracted tumor tissue source would always correlate to some extent to the other two histopathological tissue types.

It has been observed that once a source is well represented in the data, it is more prone to be correctly identified by the proposed nBSS algorithms and is less sensitive to the k value. These conclusions are in agreement with [9]. On the other hand, the overall best performance in identifying all sources, even the sources that are not very well represented in the input data, is obtained for $k = 3$. The dimension of the input space (the number of spectra used for extracting the sources) is not influencing the results as proven by the experiments described in Section 4.

With respect to the best input feature space to be considered for best separation between necrotic, highly cellular tumor tissue and infiltrations, our results are in agreement with the literature and show that NAA, Cr, Lips and tCho group are most representative for solving this classification problem. Adding extra information will not bring any added value in the performance of the considered nBSS methods. Additionally, by increasing the dimensionality of the matrices (case $n = 19$ and $n = 716$), the problem becomes more expensive and also prone to provide a solution that is less meaningful for the given problem. A reduction of the input feature space to the most representative features would bring a twofold advantage. On one side, it reduces the computational time, and secondly, we avoid bringing irrelevant information into the problem.

Using the mixing coefficients, we can compute the percentage of necrosis, tumor cellularity and infiltrations with normal tissue for each case. This approach presents the advantage that the results are closer in what concerns their interpretation to the histopathological output and to the clinical interpretation.

6. CONCLUSIONS

The nBSS methods proposed for obtaining characteristic profiles for each tissue subtype and the abundance of each source within a spectrum can reliably answer the problem of source separation when analyzing HR-MAS data. This finding can provide relevant additional information for a better interpretation and classification of brain tumor tissue in *ex vivo* high-resolution MRS. Furthermore, brain tumor tissue classification problems arising from *in vivo* MRS can benefit from a similar approach.

UNCORRECTED PROOF

Acknowledgements

Research Council KUL: GOA MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), PFV/10/002 (OPTEC), IDO 08/013 Autism, IOF-KP06/11 FunCopt, several PhD/postdoc & fellow grants. Flemish Government: FWO: PhD/postdoc grants, projects: FWO G.0302.07 (SVM), G.0341.07 (Data fusion), G.0427.10N (Integrated EEG-fMRI), G.0108.11 (Compressed Sensing) research communities (ICCoS, ANMMM); IWT: TBM070713-Accelero, TBM070706-IOTA3, TBM080658-MRI (EEG-fMRI), PhD Grants; IBBT. Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, "Dynamical systems, control and optimization", 2007–2011); ESA AO-PGPF-01, PRODEX (CardioControl) C4000103224. EU: RECAP 209G within INTERREG IVB NWE programme, EU HIP Trial FP7-HEALTH/2007–2013 (no. 260777) Neuromath (COST-BM0601), (FP6-MC-RTN-035801), EU: eTUMOUR (FP6-2002-LIFESCIHEALTH 503094). Other: BIR&D Smart Care. Ministerio Ciencia e Innovación de España projects: SAF2007-65473; SAF2007-29393-E; SAF2007-29394-E; SAF2007-29455-E; SAF2008-00270. Daniel Monleon is acknowledged. Diana M. Sima is a postdoctoral fellow of the Foundation for Scientific Research Flanders.

REFERENCES

1. Martinez-Bisbal MC, Marti-Bonmati L, Piquer J, Revert A, Ferrer P, Llacer JL, Piotto M, Assemat O, Celda B. 1H and 13C HR-MAS spectroscopy of intact biopsy samples ex vivo and in vivo 1H MRS study of human high grade gliomas. *NMR Biomed.* 2004; **17**(4): 191–205.
2. Opstad KS, Wright AJ, Bell BA, Griffiths JR, Howe FA. Correlations between in vivo (1) H MRS and ex vivo (1) H HRMAS metabolite measurements in adult human gliomas. *J. Magn. Reson. Imaging* 2010; **31**: 289–297.
3. Devos A. Quantification and classification of magnetic resonance spectroscopy data and applications to brain tumour recognition. *PhD Thesis*, Faculty of Engineering, K.U.Leuven (Leuven, Belgium). 2005.
4. Croitor Sava AR, Martinez-Bisbal MC, Van Huffel S, Cerda JM, Sima DM, Celda B. Ex vivo high resolution magic angle spinning metabolic profiles describe intratumoral histopathological tissue properties in adult human gliomas. *Magn. Reson. Med.* 2011; **65**: 320–328.
5. Sajda P, Du S, Brown TR, Shungu RSDC, Mao X, Parra LC. Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain. *IEEE Trans. Med. Imaging* 2004; **23**: 1453–1465.
6. Cheng LL, Chang LW, Louis DN, Gonzalez G. Correlation of high-resolution magic angle spinning proton magnetic resonance spectroscopy with histopathology of intact human brain tumor specimens. *Cancer Res.* 1998; **58**: 1825–1832.

Copyright © 2012 John Wiley & Sons, Ltd.

7. Lee Y, Huang Y, El-Deredy W, Lisboa PJ, Ards C, Harris P. Robust methodology for the discrimination of brain tumours from in vivo magnetic resonance spectra. *IEE Proc.: Sci. Meas. Technol.* 2000; **147**: 309–314.
8. Ladroue C, Tate AR, Howe FA, Griffiths JR. Independent component analysis. For automated decomposition of in vivo magnetic resonance. *Magn. Reson. Med.* 2003; **50**: 697–703.
9. Szabo de Edelenyi F, Simonetti AW, Postma G, Huo R, Buydens LMC. Application of independent component analysis to ^{1}H MR spectroscopic imaging exams of brain tumours. *Anal. Chim. Acta* 2005; **544**: 36–46.
10. Su Y, Thakur SB, Karimi S, Du S, Sajda P, Huang W, Parra LC. Spectrum separation resolves partial-volume effect of MRSI as demonstrated on brain tumor scans. *NMR Biomed.* 2008; **21**: 1030–1042.
11. Wright AJ, Fellows G, Byrnes TJ, Opstad KS, McIntyre DJO, Griffiths JR, Bell BA, Clark CA, Barrick TR, Howe FA. Pattern recognition of MRSI data shows regions of glioma growth that agree with DTI markers of brain tumor infiltration. *Magn. Reson. Med.* 2009; **62**: 1646–1651.
12. Nuzillard D, Nuzillard JM. Application of blind source separation to 1-D and 2-D nuclear magnetic resonance spectroscopy. *IEEE Trans. Signal Process. Lett.* 1998; **5**(8): 209–211.
13. Malinowski ER. *Factor Analysis in Chemistry*. John Wiley: New York, 2002.
14. Nascimento JMP, Dias JMB. Does independent component analysis play a role in unmixing hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* 2005; **43**(1): 175–187.
15. Plumbley MD. Conditions for nonnegative independent component analysis. *IEEE Signal Process. Lett.* 2002. **9**(6): 177–180.
16. Plumbley MD. Algorithms for non-negative independent component analysis. *IEEE Trans. Neural Netw.* 2003; **14** (3): 534–543.
17. Oja E, Plumbley MD. Blind separation of positive sources by globally convergent gradient search. *Neural Comput.* 2004; **16**: 1811–1825.
18. Yuan Z, Oja E. A FastICA algorithm for nonnegative independent component analysis. *Proceedings of the Fifth International Symposium on Independent Component Analysis and Blind Signal Separation, ICA'04*, Granada, Spain, 2004.
19. Moussaoui S, Brie D, Mohammad-Djafari A, Carteret C. Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling. *IEEE Trans. Signal Process.* 2006; **54**(11): 4133–4145.
20. Cha H, Ma WK, Chi CY, Wang Y. A convex analysis framework for blind separation of non-negative sources. *IEEE Trans. Signal Process.* 2008; **56** (10): 5120–5134.
21. Laudadio T, Mastronardi N, Vanhamme L, Van Hecke P, Van Huffel S. Improved Lanczos algorithms for blackbox MRS data quantitation. *J. Magn. Reson.* 2002; **157**: 292–297.
22. Poullet JB, Martinez-Bisbal MC, Valverde D, Monleon D, Celda B, Arus C, Van Huffel S. Quantification and classification of high-resolution magic angle spinning data for brain tumor diagnosis. *Conf Proc IEEE Eng Med Biol Soc* 2007; 5407–5410.
23. Kim H, Park H. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* 2007; **23–12**: 1495–1502.
24. Kim H, Park H. Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM J. Matrix Anal. Appl.* 2008; **30**: 713–730.
25. Hyvärinen A, Karhunen J, Oja E. *Independent Component Analysis*. John Wiley & Sons, 2001.
26. Amari SI, Cichocki A. *Adaptive Blind Signal and Image Processing*. John Wiley & Sons, Inc., 2002.
27. Oja E. Nonlinear PCA criterion and maximum likelihood in independent component analysis. *Proceedings of International Workshop on Independent Component Analysis and Blind Signal Separation, ICA'99*, Aussois, France, 1999; 143–148.
28. Himmelreich U, Accursso R, Malik R, Dolenko B, Somorjai RL, Gupta RK, Gomes L, Mountford CE, Sorrell TC. Identification of Staphylococcus aureus brain abscesses: rat and human studies using H-1 MR spectroscopy. *Radiology* 2005; **236**: 261–270.
29. Himmelreich U, Somorjai RL, Dolenko B, Lee OC, Daniel HM, Murray R, Mountford CE, Sorrell TC. Rapid identification of Candida species by using nuclear magnetic resonance spectroscopy and a statistical classification strategy. *Appl. Environ. Microbiol.* 2003; **69**: 4566–4674.
30. Opstad KS, Bell BA, Griffiths JR, Howe FA: An assessment of the effects of sample ischaemia and spinning time on the metabolic profile of brain tumour biopsy specimens as determined by high-resolution magic angle spinning (1)H NMR. *NMR Biomed.* 2008; **21**: 1138–1147.
31. Majos C, Julia-Sape M, Alonso J, Serrallonga M, Aguilera C, Acebes JJ, Arus C, Gili, K. Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE. *Am. J. Neuroradiol.* 2004; **25**: 1696–1704.
32. Tong Z, Yamaki T, Harada K, Houkin K. In vivo quantification of the metabolites in normal brain and brain tumors by proton MR spectroscopy using water as an internal standard. *Magn. Reson. Imaging* 2004; **22**: 735–742.

UNCORRECTED PROOF

 Copyright © 2012 John Wiley & Sons, Ltd.