NMR IN BIOMEDICINE
NMR Biomed. 2008; **21**: 1030–1042
Published online 1 September 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI:10.1002/nbm.1271

Spectrum separation resolves partial-volume effect of MRSI as demonstrated on brain tumor scans

Yuzhuo Su,[1] Sunitha B. Thakur,[2] Sasan Karimi,[3] Shuyan Du,[4] Paul Sajda,[4] Wei Huang[2] and Lucas C. Parra[1]*

[1]Department of Biomedical Engineering, The City College of the City University of New York, New York, NY, USA
[2]Department of Medical Physics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA
[3]Department of Radiology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA
[4]Department of Biomedical Engineering and Department of Radiology, Columbia University, New York, NY, USA

Received 2 March 2007; Revised 16 February 2008; Accepted 20 February 2008

ABSTRACT: Magnetic resonance spectroscopic imaging (MRSI) is currently used clinically in conjunction with anatomical MRI to assess the presence and extent of brain tumors and to evaluate treatment response. Unfortunately, the clinical utility of MRSI is limited by significant variability of *in vivo* spectra. Spectral profiles show increased variability because of partial coverage of large voxel volumes, infiltration of normal brain tissue by tumors, innate tumor heterogeneity, and measurement noise. We address these problems directly by quantifying the abundance (i.e. volume fraction) within a voxel for each tissue type instead of the conventional estimation of metabolite concentrations from spectral resonance peaks. This 'spectrum separation' method uses the non-negative matrix factorization algorithm, which simultaneously decomposes the observed spectra of multiple voxels into abundance distributions and constituent spectra. The accuracy of the estimated abundances is validated on phantom data. The presented results on 20 clinical cases of brain tumor show reduced cross-subject variability. This is reflected in improved discrimination between high-grade and low-grade gliomas, which demonstrates the physiological relevance of the extracted spectra. These results show that the proposed spectral analysis method can improve the effectiveness of MRSI as a diagnostic tool. Copyright © 2008 John Wiley & Sons, Ltd.

KEYWORDS: magnetic resonance spectroscopic imaging (MRSI); non-negative matrix factorization (NMF); tumor grade classification; brain tumor

INTRODUCTION

Altered metabolic activity in cancerous tissue leads to abnormal metabolite concentrations, which are reflected in abnormal spectral profiles recorded with magnetic resonance spectroscopic imaging (MRSI) (1–3). Unfortunately, the significant variability of *in vivo* clinical spectra limits the diagnostic potential of MRSI. Our hypothesis is that the mixture of different tissue types within a given voxel – also called partial-volume effect – leads to increased variability in large voxel spectra compared with spectra obtained from homogeneous tissue. The resonance spectrum of such a mixed voxel can be described as a linear combination of spectra from different constituent tissue types. To validate this hypothesis, we aim to show that spectral variability can be reduced by representing each voxel's spectrum as a linear combination of constituent tissue types, each with a consistent spectrum across many voxels. This modeling process, which we here call 'spectrum separation', estimates the abundance (or partial volume fraction) of each tissue type within each voxel as illustrated in Fig. 1. Specifically, the model explains the observed spectra, $\boldsymbol{X}$, as a product of abundances, $\boldsymbol{A}$, with constituent tissue spectra, $\boldsymbol{S}$, and additive measurement noise, $\boldsymbol{N}$:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{N} \tag{1}$$

The columns in matrix $\boldsymbol{A}$ represent the abundance of the constituent tissue, and the rows in matrix $\boldsymbol{S}$ represent their corresponding spectra. The abundance matrix $\boldsymbol{A}$ has M columns (one for each constituent) and N rows (one for each voxel). $\boldsymbol{X}$ and $\boldsymbol{S}$ have L columns (one for each resonance band). Eqn (1) represents a linear superposition, which accurately reflects the superposition of the resonance signal under the assumption of a homogeneous field. This holds for absorption spectra, i.e. the real values of the Fourier transformation of the observed free induction decay (FID) (4). There are two types of methods for performing this factorization:

**Correspondence to:* L. C. Parra, Department of Biomedical Engineering, The City College of New York, CUNY, 138th Street and Convent Avenue, New York, NY 10031, USA.
E-mail: parra@ccny.cuny.edu
Contract/grant sponsor: MSKCC-CCNY partnership; *contract/grant number:* NIH/NCI U56 CA96299-0.

Abbreviations used: AUC, area under the ROC curve; Cho, choline; CNI, Cho-to-NAA index; Cr, creatine; CV, coefficient of variation; FID, free induction decay; FLAIR, fluid-attenuated inversion recovery; HGG, high-grade glioma; LGG, low-grade glioma; MRSI, magnetic resonance spectroscopic imaging; NAA, *N*-acetyl aspartate; NMF, non-negative matrix factorization; PCA, principal component analysis; ROC, receiver operating characteristic; SNR, signal-to-noise ratio.

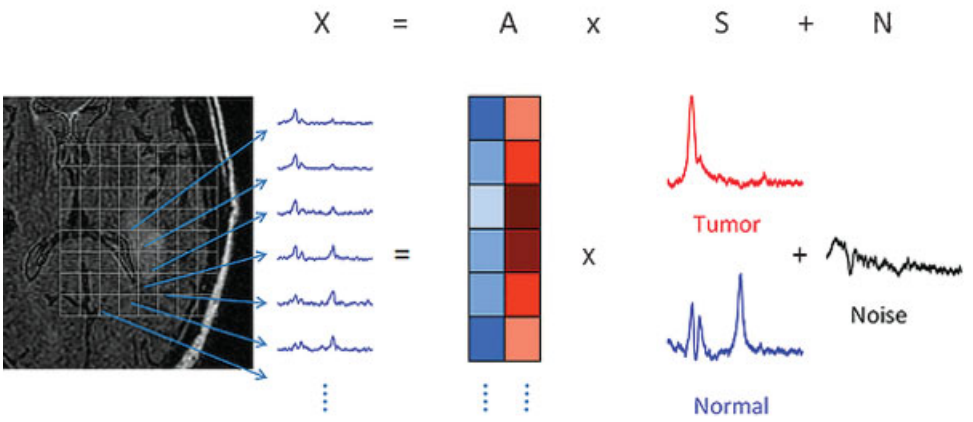

Figure 1. Sketch of spectrum separation approach. Spectra of multiple voxels, ***X***, are simultaneously analyzed and decomposed into constituent spectra, ***S***, and the corresponding abundance distributions, ***A***. The extracted constituent spectra can be identified by comparing them with known spectra of individual tissue types (normal tissue and tumor tissue in this study).

model-based methods (5–7) and statistics-based methods (8,9). Model-based methods such as variable projection (VARPRO) (10) and LCModel (11) impose an explicit or parametric model on ***S*** with prior biochemical knowledge, and consider one voxel at a time. The optimal parameters and abundances are chosen such that eqn (1) is satisfied with the smallest possible noise ***N***. Although these model-fitting methods can quantify metabolite concentrations from the MRS data, they are also very sensitive to noise and often require assumptions on the biochemical composition of the tissue. By analyzing individual spectra, these methods fail to exploit the potential benefits of averaging that implicitly occurs when multiple spectra are simultaneously fitted. Statistics-based techniques, such as principal component analysis (PCA) (12–14) and independent component analysis (14–16), instead use all spectra simultaneously to extract constituent components. Thus, they can potentially exploit the statistical structure of multi-voxel spectra to solve for all rows in eqn (1) simultaneously. Instead of making explicit assumptions about the structure of ***S***, these techniques assume statistical properties such as co-variation and independence. One of the problems with these earlier approaches, however, is the common assumption that constituent spectra are orthogonal, as there are many cases where the constituent spectra, because of overlapping peaks (e.g. lipid and lactate), can be highly correlated, and therefore an orthogonality or independence assumption is incorrect.

We previously proposed (17) the use of a statistics-based algorithm, called the non-negative matrix factorization (NMF) algorithm by Lee and Seung (18), for performing the factorization in eqn (1). This algorithm exploits the fact that both abundances and spectra can only take on positive values. Negative values in the data are ascribed to the noise term. The advantage of the resulting estimation process is that no reference spectra are required. Instead, the algorithm can adapt to the heterogeneous spectra observed for different tumors. When applied to brain MRSI of patients with tumors, the algorithm extracts spectral profiles and their spatial distributions consistent with different tissue types such as necrotic and proliferative regions, normal brain, and skull (17).

The goal of this study was to demonstrate the physiological relevance of the NMF algorithm for routine clinical brain tumor scans. First, we validated the physical interpretation of abundances, ***A***, as volume fraction in a phantom study by comparing the extracted abundance values with the expected values. Next, to confirm the physiological and clinical relevance of the extracted spectra, ***S***, we correlated the analysis results with pathologically proven tumor grades from 20 patients and showed an improved correlation of choline (Cho) and *N*-acetyl aspartate (NAA) peak areas with tumor grade. The proposed decomposition of the data can be expected to coincide with constituent spectra and abundance only if the imaging process and spectral analysis satisfy the linear model [1]. In practice, and in particular for our clinical scans, field inhomogeneities lead to non-linear distortions of the data such as phase distortions and frequency shift. This study therefore explores the limits of the linear model for varying signal-to-noise ratio (SNR) on simulated data, and compares these limits with the SNR of 1.5 T routine clinical scans.

A problem of multi-voxel methods is that they assume constituent spectra to be the same across voxels, thus requiring careful correction of deviations such as phase errors and frequency shifts. In addition, known tumor heterogeneity may lead to spectra that vary across voxels. We intend to use the proposed multi-voxel analysis method to capture the variability that is due to volume fraction, and consequently improve the diagnostic potential of MRSI despite known tumor heterogeneity.

DOI: 10.1002/nbm

EXPERIMENTAL

Data acquisition protocol

This study analyzed data from the routine clinical patient population at Memorial Sloan-Kettering Cancer Center (MSKCC). Institutional review board waiver was obtained to retrospectively analyze clinical proton MRSI data and examine the medical records. Thirty-two patients with pathologically proven brain gliomas were evaluated. Of these, we selected 20 MRSI scans with sufficient data quality (see section on voxel selection). The MRSI data were collected using a 1.5 T GE scanner (General Electric Medical Systems, Milwaukee, WI, USA) with a long echo time ($TE = 144$ ms, $TR = 1000$ ms) three-dimensional PRESS sequence with water suppression. The voxel size of the data acquired with MRSI ranges from 1 cm^3 down to 0.5 cm^3. The acquisition or phase-encoding volume typically covers $8 \times 8 \times 8$ voxels, which results in 8.5 min scanning time at a 1 cm^3 nominal voxel size. Pre-contrast fluid-attenuated inversion recovery (FLAIR) images (3 mm thickness/0 mm spacing) were used as scouts for placement of the MRSI excitation volume. FLAIR and T_1 post-contrast MR images were collected in alignment with the MRSI study so that they could be combined for diagnosis and tumor segmentation. The final pathological diagnoses were retrospectively confirmed for all patients. Of the 20 patients (all of whom had received some form of treatment before the scan), 10 had low-grade (WHO grade I–II) gliomas (LGGs) including WHO grade II astrocytoma, oligodendrioglioma, and oligoastrocytoma, and the other 10 had high-grade (WHO grade III–IV) gliomas (HGGs) including anaplastic astrocytoma, anaplastic oligodendrioglioma, and glioblastoma. This categorization was used as truth data for the classification.

Data preprocessing

The GE spectrum analysis software (Functool) does not output the computed spectra numerically. For our analysis it was therefore necessary to redevelop the corresponding data-conditioning routines starting with the raw time-domain data (GE P-files). These routines, written in MATLAB, filtered residual water signal, performed phase correction, and corrected for frequency shifts due to an inhomogeneous magnetic field. The results of this processing are equivalent to those obtained with Functool.

The MRSI data were preprocessed automatically and identically for all datasets. This resulted for each patient scan in $8 \times 8 \times 8$ ($N = 29$) or $16 \times 8 \times 8$ ($N = 3$) spectra with 1024 points covering 1 kHz spectral width.

Water filtering. The most critical step in preprocessing is to remove the low frequency signals due to water, which tend to overwhelm, in particular, Cho and creatine (Cr) spectral lines. We used a third-order Butterworth high-pass filter with 75 Hz cutoff applied forward and backward in time to avoid phase distortions (MATLAB filtfilt function).

Zero filling. The time-domain FID data typically contain 512 samples at 1 kHz sampling frequency. To increase spectral resolution, we increased the length of the signal by appending zero values to a length of 1024.

Line broadening. To reduce the effect of noise, one can smooth the spectrum by windowing the data in the time domain. We used exponential windowing, resulting in 3 Hz line broadening, which compromises between noise and resolution.

Frequency decomposition. After these steps, the frequency-domain data were recovered from the three-dimensional phase-encoded FID time-domain sequences with a four-dimensional Fourier transformation. The only remaining step that has to be validated is a potential phase-encoding offset, which would manifest itself in spatial misalignment of the spectra with the FLAIR images.

Phase correction. Phase distortions are particularly problematic, as they will lead to negative values in the absorption spectra, violating the main assumption of the NMF algorithm. More importantly, if the phase is not corrected to give the same effective phases across voxels, the assumption that the constituent spectra are the same across voxels is violated. To ensure positive spectra, one has to determine a separate phase factor for each voxel. The noise levels and phase distortions due to an inhomogeneous field are significant in these clinical data. Moreover, at long *TE* values, an inverted lipid/lactate peak at 1.3 ppm may be present with a 180° phase that cannot be corrected. So, instead of the conventional zero-order and first-order phase correction, we opted to use the absolute value of the spectrum rather than the real (absorption) spectrum.

Frequency alignment. The spectrum separation method operates simultaneously on multiple spectra and assumes that the constituent spectra coincide across voxels. To determine potential frequency shifts, we cross-correlated the resulting absolute spectra and adjusted frequency shifts individually for each voxel by no more than 0.3 ppm to maximize correlation in the frequency range 1.6–3.6 ppm. This method worked well on the present data with SNR >4 dB. Baseline correction or modeling was not required.

Spectrum separation with NMF algorithm

The proposed model of eqn (1) interprets matrix $\boldsymbol{A}$ as abundance, which therefore only takes on non-negative values. In addition, as the constituent spectra, $\boldsymbol{S}$, represent

DOI: 10.1002/nbm

amplitudes of resonances, in theory the smallest resonance amplitude is zero, corresponding to the absence of resonance at a given frequency. The factorization of eqn (1) is therefore constrained by, $\boldsymbol{A} \geq 0$ and $\boldsymbol{S} \geq 0$. The basic idea of the NMF algorithm is to maximize the likelihood of observing $\boldsymbol{X}$ given non-negative $\boldsymbol{A}$ and $\boldsymbol{S}$. Assuming Gaussian noise, $\boldsymbol{N}$, the log-likelihood is simply a quadratic function which is to be maximized with respect to $\boldsymbol{A}$ and $\boldsymbol{S}$ subject to the positivity constraint. A corresponding gradient ascent algorithm is converted into a multiplicative update algorithm by the appropriate choice of gradient step sizes (17,28). After preprocessing, the absolute spectra were analyzed using the NMF algorithm (17). This process extracts spectra directly from the data, and hence no reference spectra are required. The only model parameter of NMF that needs to be determined is the dimensionality of the matrices $\boldsymbol{A}$ and $\boldsymbol{S}$, namely the number of constituent spectra to recover. Conventional subspace analysis (PCA) was used to determine the number of constituents (29). On the present data, PCA suggested the use of at most two or three spectra. Whenever a third constituent was used, it represented primarily residual water (18 out of 20). When only two constituent spectra were used, water activity was ascribed to residual noise. In the cases with clear lactate/lipid peaks, the third constituent corresponds to high-lipid tumor region (2 out of 20), provided that the corresponding frequency range is included. In all cases, the spectrum with highest Cho-to-NAA index (CNI) was assigned to tumor tissue, and the spectrum with lowest CNI was assigned to normal tissue. Using two rather than three spectra had little effect on the classification of tumor type and tumor location. For simplicity, we will report here results for two constituent spectra.

Voxel selection

The specific selection of voxels will affect the results of spectrum separation using the NMF algorithm. In particular, inclusion of noisy or distorted spectra (due to poor shimming) has a detrimental effect. In this study, we selected rectangular volumes within the excitation box. Only slices (horizontal, coronal and sagittal) that included obvious metabolite resonance peaks in the affected brain area were selected. Peak height had to be 4 dB above background for at least one voxel in the slice within the frequency ranges covering Cho (3.34–3.14 ppm), Cr (3.14–2.94 ppm), and NAA (2.22–1.82 ppm). Examples of spectra that would have been included or excluded are shown in Fig. 2. Also excluded were slices that intersected the skull with obvious distortions or lipid content. The separation algorithm requires multiple voxels for extraction of meaningful spectra, we only used those 'high-yield' datasets with at least 10% of voxels satisfying the above criteria. Of the 32 datasets, four with radiation necrosis and eight with 'low yield' were excluded, which resulted in 20 datasets for further analysis. Examples of the datasets with selected voxels are shown as colored areas in Fig. 3 (excitation box delineated in white).

To compare the spectra extracted with the NMF algorithm with conventional analysis, we also selected representative raw voxel spectra obtained following current conventional practice. This selection was based on FLAIR intensity enhancement and Cho/NAA peak-area ratios. Again, only spectra with sufficient SNR were selected (>4 dB). Spectra from areas with obvious intensity enhancements were selected as tumor spectra. For normal brain, we selected only voxels far removed (>2 cm) from areas of increased intensity. Because signal quality varies among different datasets, different numbers of voxels were selected for normal ($n = 17 \pm 11$) and tumor ($n = 15 \pm 12$) spectra for different datasets. Within this selection, we also chose one example with an extreme Cho/NAA peak-area ratio as an 'extreme' spectrum. This corresponds to current clinical practice, which considers the most extreme Cho/NAA ratio as an indicator of tumor malignancy. Finally, for each subject, 'average' spectra were computed as the mean across these selected voxels for tumor and normal tissue.

Spectrum varability

To quantify the reduction in variability, we measured the coefficient of variation (CV) of the Cho/NAA ratio. The statistic was computed for Cho and NAA on the basis of maximum peak areas of the extracted spectra. To normalize for the arbitrary scale of the spectra, we used the Cr peak area. Specifically, the peak area of Cr, Cho, and NAA were scaled to sum to one. Hence, NAA and Cho were given in arbitrary units that are insensitive to scale. Note that this operation preserves linear separability of the data (separating lines in Fig. 4 remain lines for any such scaling transformation).

Overlay construction

To present the result of the NMF separation algorithm, we merged the FLAIR image intensity with the abundance estimates, $\boldsymbol{A}$, encoded in color. Examples of this visualization method are shown in Fig. 3 and are compared with the conventional spectral display. In this overlay, the abundance of the tumor spectrum for a given voxel is represented as a color between blue (no tumor) and red (tumor). Color saturation shows how well the spectrum $\boldsymbol{X}$ is represented by the two constituents $\boldsymbol{S}$ [measured as goodness of fit, $0 \leq r^2 \leq 1$, with $r^2 = 1 - \sum(X - AS)^2 / \sum X^2$], and is hence an indirect indication of SNR. Color therefore is only visible in the region of interest inside the signal acquisition volume (white PRESS box). Brightness represents FLAIR intensity as in a conventional MR image.

DOI: 10.1002/nbm

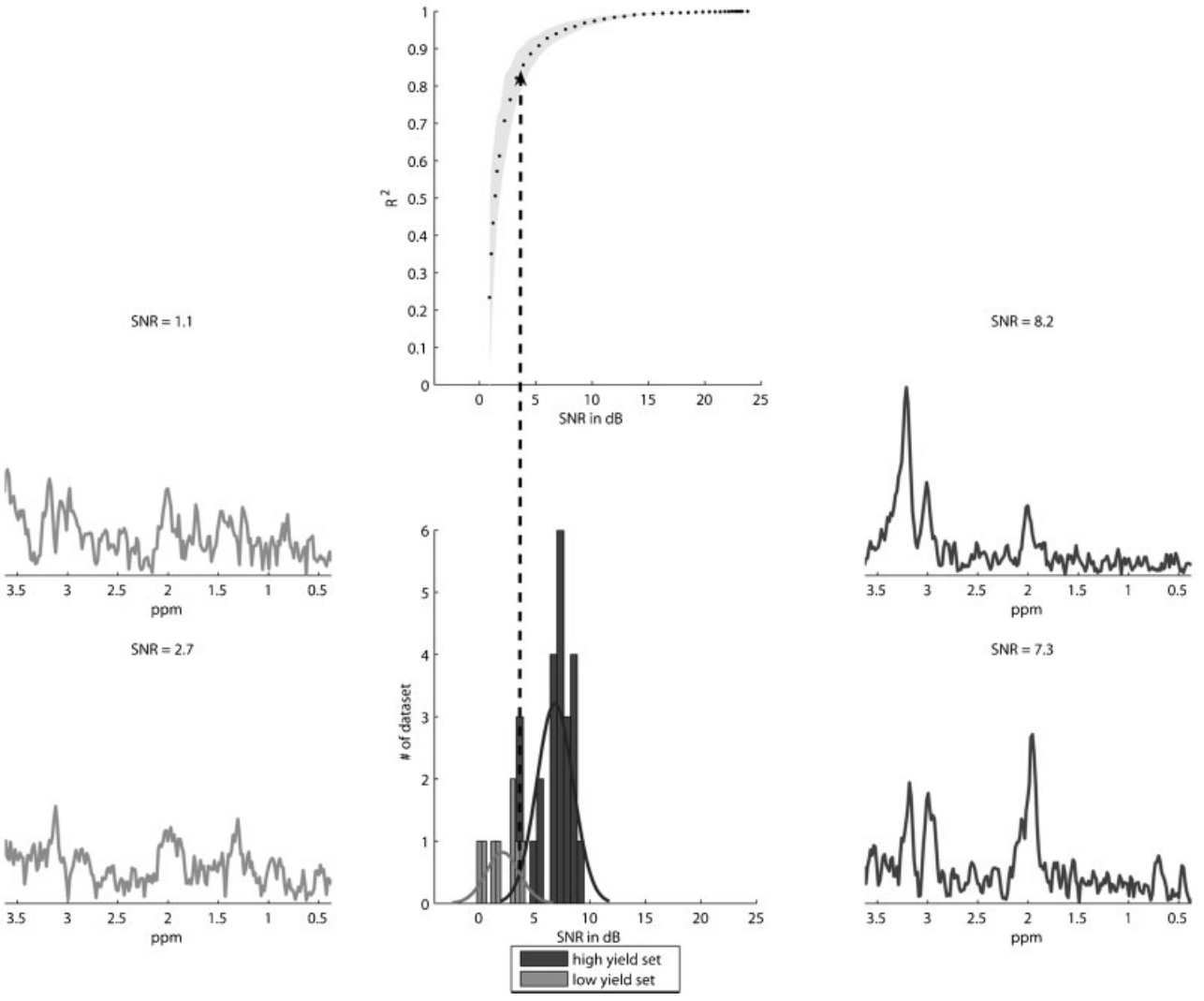

Figure 2. Linearity validation. Top graph in the center panel shows the fit of the linear model to simulated data for varying noise levels. Accuracy is measured as R^2 value capturing how well the linear model approximates the data ($R^2 = 1$ corresponds to a perfect match). Black dots represent the mean R^2 from multiple repeats ($N = 100$) with random noise. Shading around the dots are the 95% confidence intervals. Bottom graph in the center panel shows the distribution of data quality for the 32 available clinical datasets. The datasets are divided into two groups depending on the number of useful voxels, i.e. 24 datasets with a 'high yield' of useful voxels (four datasets with radiation necrosis were excluded for further analysis) and eight with a 'low yield' of voxels (see section on voxel selection). Evidently these two datasets also differ in SNR. In particular, all useful datasets have an SNR above 4 dB, suggesting that the linear model may be sufficient with mean $R^2 = 0.85$. The 'low-yield' datasets were excluded from the current study. Left panel shows two examples of spectra from 'low-yield' datasets with SNR below 4 dB. Right panel shows two examples of spectra from 'high-yield' datasets with SNR above 4 dB.

To validate this abundance-based overlay, we selected the central slice from each of the 20 cases, and a neuroradiologist labeled the voxels in the corresponding slice as 'normal', 'tumor', 'HGG', or 'LGG' on the basis of anatomical MRI information (pre-contrast sagittal and axial T_1-weighted, axial T_2-weighted, axial FLAIR, axial diffusion-weighted, and post-contrast T_1-weighted images in axial, sagittal, and coronal planes). These labels were then digitized and used as 'truth data' for the receiver operating characteristic (ROC) analysis (30).

Phantom study

To empirically validate the accuracy of the abundance estimates, we built a cylindrical phantom (Plexiglas) with two semi-cylindrical chambers, which were filled with solutions containing Cr and NAA at concentrations corresponding to ''normal brain tissue'' (NAA, 10.43 mM; Cr, 7.49 mM) and ''tumor tissue'' [NAA, 5.21 (=10.43/2) mM; Cr, 14.98 (=7.49 × 2) mM]. Note that, with this choice, we expected a factor 4 difference in

DOI: 10.1002/nbm

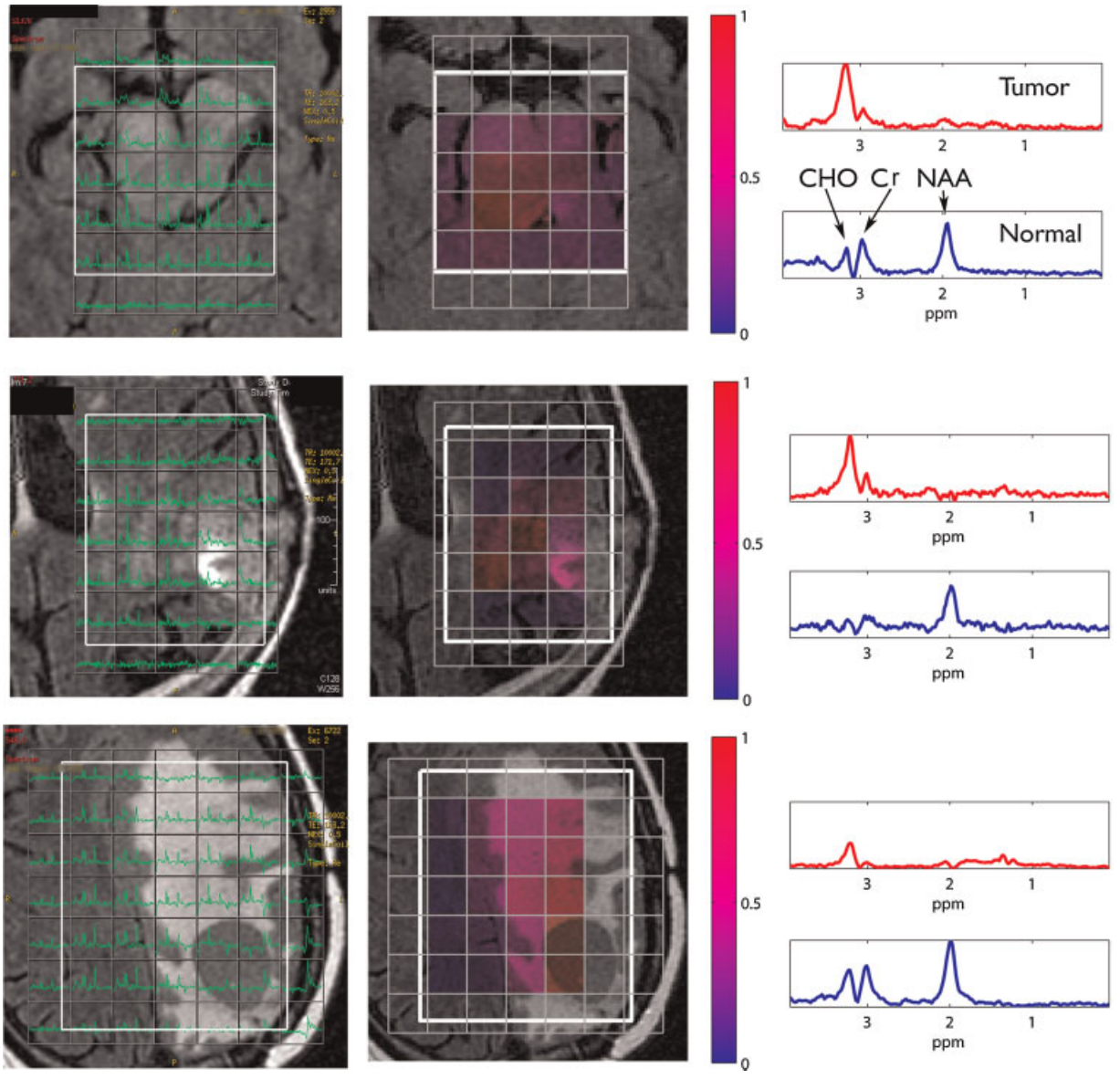

Figure 3. Comparison of conventional MRSI spectra (left) with the results of spectrum separation (center, right). The left panels show the conventionally processed MRSI multi-voxel spectra overlaid on to FLAIR images, with areas of hyperintensity indicating the abnormal region. The white box represents the region of radiofrequency excitation, which is smaller than the field of view of MRSI acquisition. The center panels show the tumor tissue abundance map – column in matrix ***A*** corresponding to the tumor spectrum – merged with the FLAIR image: the redder the area, the more abundant the tumor tissue. For a smooth spatial distribution, this color map has been interpolated between voxels at the same resolution as the FLAIR image. The right panels show the extracted constituent spectra corresponding to the normal tissue and tumor tissue. The constituent tumor spectrum revealed a typical pattern of high Cho concentration and low NAA concentration.

the Cr/NAA ratio between the spectra of the two chambers. Cho was not included because of poor stability in the solution. FLAIR and MRSI data were obtained ensuring partial coverage of voxels across the boundary of the chambers as shown in Fig. 5. Data were recorded using the same acquisition parameters as in the clinical scans and were processed in the same fashion, resulting in spectra with an SNR of 11.5 ± 0.7 dB ($n = 5$). All voxels inside the excitation box were selected to validate the accuracy of the abundance estimates. A sigmoid function was used to model the relationship between real abundance and the estimated abundance: $f(x) = [1 + \exp(-c - bx)]^{-1}$, where x represents the real abundance and f represents the abundance estimates. 95% confidence intervals were then calculated to quantify the accuracy of the abundance estimates using standard methods (25–27).

Simulation Studies

Simulated data for linearity validation. Nonlinear distortions originating from an inhomogeneous

DOI: 10.1002/nbm

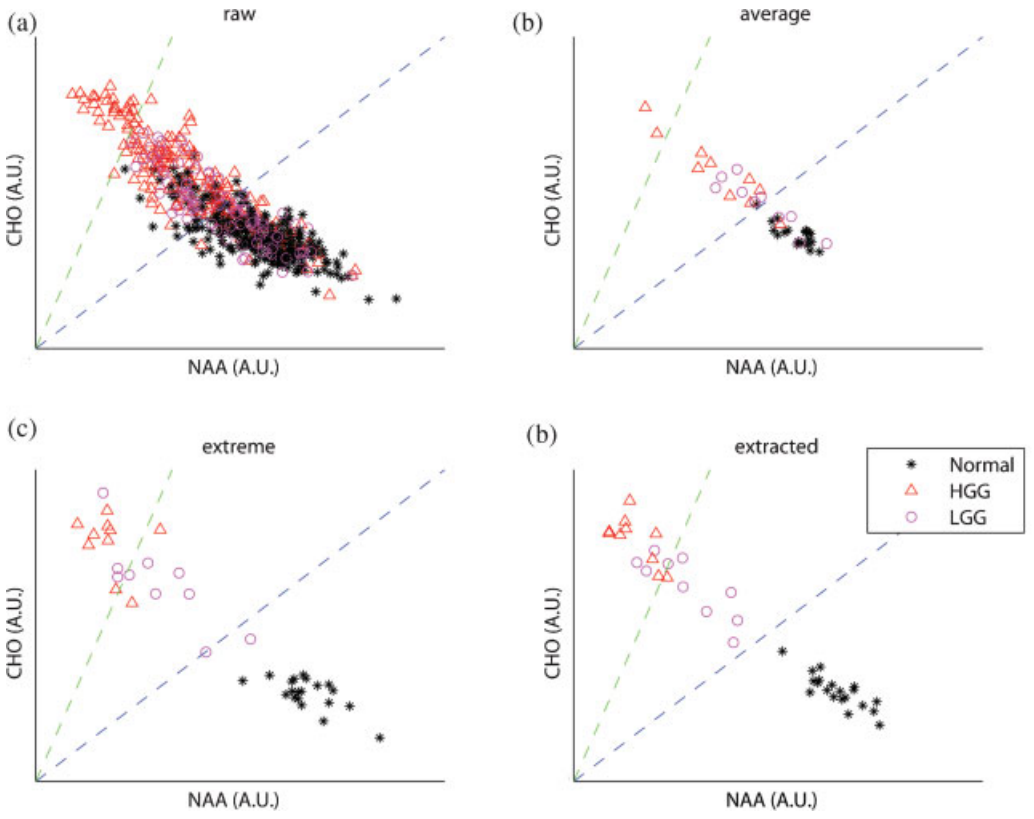

Figure 4. Validation of extracted spectra, $\boldsymbol{S}$, with pathologically proven tumor grades. (a) Cho vs NAA peak areas (in arbitrary units) for voxel spectra from all 20 patients, taken from the individual MRSI voxels located in the regions of FLAIR intensity enhancement and surrounding normal appearing areas before the NMF analysis. Points indicate maximum peak areas for spectra over the following frequency ranges: Cho 3.34–3.14 ppm and NAA 2.22–1.82 ppm. The colored symbols are for spectra from tumor voxel in the area of intensity enhancement. There is significant overlap in voxel spectral pattern between the HGG, LGG, and normal appearing regions. Dashed blue and green lines show the Cho/NAA ratio of 1 and 3, respectively. (b) Averaged Cho and NAA concentrations across voxels for each individual patient. Averaging across voxels reduces noise, but significant overlap remains. (c) Cho and NAA concentrations of the extreme spectra (see section on voxel selection) from all 20 patients. Compared with (b), the overlap is reduced and better separation is achieved. (d) The same type of plot of the constituent tumor and normal tissue spectra from all 20 patients after the NMF analysis. As with the averaged spectra, each of the 20 cases contributes two points corresponding to the extracted spectrum for normal and tumor tissue. The Cho–NAA patterns of the normal spectra are well separated from those of the tumor spectra. The separation is sufficient to distinguish tumor from normal tissue and HGG from LGG (Table 1) and matches very well with the clinical criteria based on Cho/NAA ratio (dashed blue and green lines).

field and the use of absolute spectra may lead to a violation of the linear model of eqn (1). Thus we tested the validity of the linear model assumption on simulated data (6,19). Idealized FID sequences were generated in the time domain ($N=512$) as the sum of $K=3$ exponentially damped complex sinusoids, which corresponded to the specific resonance peaks for H_2O, NAA and Cho:

$$S_n = \sum_{k=1}^{K} \alpha_k e^{i\varphi_k} e^{(-d_k + i2\pi f_k)t_n}, \quad n = 1, \ldots, N \qquad (2)$$

where s_n represents the n-th data point of the simulated signal, $\boldsymbol{i}$ represents the imaginary unit, the parameters α_k, φ_k, d_k, and f_k denote the amplitude, phase, damping factor, and frequency, respectively, and $t_n = n\Delta t$, with the sampling interval Δt.

Three 8×8 slices of simulated MRSI data containing H_2O, NAA, and Cho were generated. One slice had a Cho/NAA ratio of 2, the second had a Cho/NAA ratio of 1/2, and the third represented a mixed spectrum generated by adding the first two slices together. In order to mimic the real situation, uniform-distributed phase shifts ($\pm\pi$) and frequency shifts (± 5 Hz, one linewidth) were introduced variably from voxel to voxel when generating the time-domain data. This is comparable to previous simulation studies, which used phase shifts in the range $\pi/4$ to $\pi/3$ and frequency shifts of 1/2 to

DOI: 10.1002/nbm

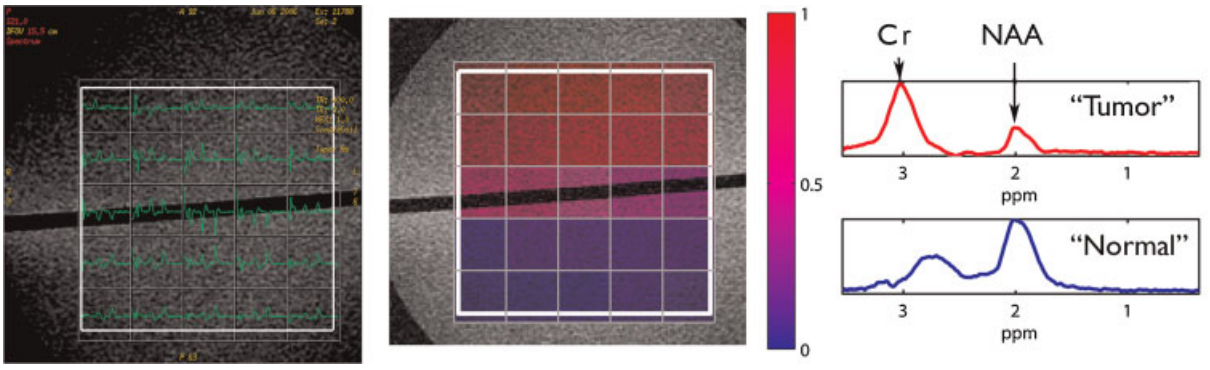

Figure 5. Results of spectrum separation on phantom study (display same as in Fig. 3). The two chambers were filled with Cr and NAA in a water solution at concentrations corresponding to ''tumor tissue'' (upper chamber) and ''normal brain tissue'' (lower chamber). The overlay in the center panel shows a good match between the estimated abundance ***A*** and the actual volume fraction (see also Fig. 6). This indicates that the method correctly decomposes the observed data into the constituent spectra corresponding to the different solutions with their specific metabolite concentrations. The Cr/NAA peak-area ratios for the two spectra differ by a factor of 4 as expected.

2 linewidth (20,21). Various levels of additive Gaussian-distributed white noise were introduced to evaluate the robustness of the preprocessing algorithm and resulting non-linearity as a function of SNR. Applying the inverse fast Fourier transformation brings the frequency data to the time domain. We applied to this simulated data the same preprocessing algorithms as were used on the clinical data. After preprocessing, the spectra from the first two slices were added together (assuming a linear model, we denote this S_{model}) and compared with the spectrum from the third slice (which was generated directly from a mixed spectrum, hence we call this S_{data}). The R^2 value was used to measure how well the linear addition model fits the mixed data. It is defined here as $R^2 = 1 - \sum(S_{data} - S_{\mathrm{model}})^2 / \sum S_{data}^2$. For each SNR level, R^2 values from 100 repetitions were used to assess the mean and 95% confidence intervals. The results of this analysis for varying SNR are shown in Fig. 2.

Simulated data for abundance estimates. We also used simulated data to explore the accuracy of the abundance estimates by testing the relationship between the actual volume fractions and the abundance estimates on simulated data. Two 8×8 slices of simulated MRSI data containing H_2O, NAA, and Cho were generated as described above, with one slice corresponding to a Cho/NAA ratio of 2, and a second slice to a ratio of 1/2. The third slice was generated by linearly mixing the first two slices with various volume fractions (selected at random) for each voxel. To assess the variance in estimation, this simulation was repeated 100 times with different randomly selected volume fractions and Gaussian noise at 11.5 dB. The simulated data were preprocessed and submitted to the NMF algorithm. The abundance estimation matrix, ***A***, was compared with the predefined volume fractions to evaluate the accuracy of the abundance estimates.

Simulated data for performance evaluation. To overcome the limitation of a small sample size, we simulated LGG and HGG cases as above with mean Cho/NAA ratios of 2 and 2.3, respectively (22–24). Variability across subjects and voxels was modeled by adding variability to these mean Cho/NAA ratios. Variance was adjusted to match the classification performance between LGGs and HGGs on the extreme spectra. These simulated data were used to estimate classification performance and CV on a larger sample.

RESULTS AND DISCUSSION

Linearity validation and data quality

Compared with our previous research study (17), the clinical scans of this study are characterized by significant background noise, phase distortions, and frequency shifts, indicating field inhomogeneities. We were able to use only 20 of the 28 available clinical scans, as outlined in the section on voxel selection. For these data, we developed data-conditioning routines to ensure proper filtering of residual water signal, as well as correcting for frequency shifts due to an inhomogeneous magnetic field. Phase correction proved difficult because of significant noise in the data. We opted therefore to operate on absolute spectra rather than the real-valued absorption spectra. Although this guaranteed positive spectra and resolved potential inconsistencies across voxels, it did introduce a non-linear distortion as a result of the absolute value operation. Fortunately, taking the absolute value of a sum of complex numbers is approximately a linear operation if the sum is dominated by a single element. Here, this means that linearity of the model can be approximately preserved if the resonance lines do not overlap and the noise is small. This is also true for 180° inverted peaks such as lactate.

DOI: 10.1002/nbm

To confirm whether the linear model is sufficient in the present context, we compared, using the simulated data, the absolute spectra obtained for mixed voxels with the sum of the absolute spectra of pure voxels. Figure 2 shows the result of the simulation for various SNRs. SNR was defined here as the power in the frequency range 1.8–2.2 ppm and 2.7–3.4 ppm over power in the range 0.6–1 ppm and 3.5–3.8 ppm (note that residual water will contribute to the noise estimate). These ranges were chosen to capture the Cho, Cr, and NAA signals. The black dots and the shaded areas represent the mean R^2 value and 95% confidence intervals at each SNR level. The graph indicates that the accuracy of the model depends on the noise levels. With poor SNR, the linearity assumption is violated, and the estimation for $\boldsymbol{A}$ and $\boldsymbol{S}$ may no longer be reliable. We measured the SNR also on our clinical data and found that the useful scans have an SNR of 4 dB or higher. For those SNR values, the linear model is a reasonable approximation, with mean $R^2 = 0.85$. Hence, the linear model may be appropriate for the absolute spectra in a subset of clinical data with SNR above 4 dB.

Spectrum separation of clinical MRS images of brain tumor

The NMF separation algorithm computes the abundance of each tissue type for each voxel. This information can be used to assess the spatial extent and infiltration of the tumor beyond the intensity enhancement region. Examples of the abundance estimates, $\boldsymbol{A}$, encoded as false-color images are shown in Fig. 3. These examples show the data for two patients with HGG (top two rows) and one patient with LGG (bottom row). In this overlay, the abundance of the tumor spectrum for a given voxel is represented as a color between blue (no tumor) and red (tumor). The figure also shows that the two extracted spectral profiles coincide with the standard clinical criteria for normal and tumor spectra, i.e. the spectrum with a large Cho and reduced NAA peak is considered to be a tumor, whereas a ratio of peak areas of 1/2 is considered normal (31,32).

The present method is similar to nosologic images (33), a previously proposed method for summarizing MRI and MRSI brain tumor data as a color-coded anatomical image. In that method, voxels are classified on the basis of MRSI and MRI data, and each tissue class is marked on an MR image as a separate color (e.g. HGG, LGG or necrosis, normal, etc.). In contrast, the present method only uses spectral information and reports continuous valued abundance estimates for tumor and normal tissue rather than discrete labels. It thus represents the spatial extent and infiltration of the tumor tissue. It provides an alternative view of the MRSI data to the radiologist, who can then combine this information with the anatomical MRI data.

Physiological relevance of the extracted spectra *S*: reduced cross-subject variability

Metabolite peak areas in conjunction with conventional MRI findings are used to determine the presence of tumor. A conventional quantitative criterion is the CNI, which measures the ratio of the Cho peak area (or height) to the area (or height) of the Cr or NAA peak (22). Disregarding other variables such as patients' age and locations of the voxels, a CNI slightly higher than 1 is generally considered abnormal but non-specific, and a CNI above 3 is likely to be a high-grade tumor. However, tumor heterogeneity, partial volume coverage, and measurement noise add significant variability to the spectra, making this quantitative criterion difficult to use (35,36).

To confirm the physiological relevance of the extracted spectra and demonstrate reduced variability, we looked at the Cho and NAA peak areas of the extracted spectra for the 20 available clinical cases with confirmed primary gliomas as shown in Fig. 4. Areas were normalized to be independent of an overall (arbitrary) scale (see Experimental). Figure 4a shows the original raw spectra of individually selected voxels (see Experimental) for all 20 cases. To show the effect that one may expect from (naively) combining multiple voxels, we show in Fig. 4b the spectra obtained after averaging over selected voxels and compare the averaged spectra with those obtained with the separation algorithm (Fig. 4d). Averaging reduces noise and hence some of the overlap, but is not sufficient to reliably distinguish between normal and tumor tissue. Evidently the overlap of averaged spectra is significantly larger than that of the extracted spectra. The origin of this reduced variability is twofold. First, assigning abundance values to each voxel captures and compensates for the variability due to partial volume. Second, computing a single spectrum that is applicable across many voxels reduces estimation variance due to measurement noise. To assess these effects, Fig. 4 also shows the results of spectra with extreme CNI (Fig. 4c). Peak areas from the extreme spectra – probably representing pure voxels with 100% volume fraction – show reduced overlap and improved separation between normal and tumor and between HGG and LGG. Reduced spectral variability manifests itself in better differentiation of the three tissue types after separation (HGG, LGG, and normal).

To quantify the reduction in variability, we measured the CV of CNI. Changes in CNI were not significant due to the small sample size. We therefore simulated data for 60 cases of HGG and LGG (see Experimental) and indeed found a significant reduction in CV (change in CV of 3.39 ± 0.09 for HGG, $p = 0.002$, and 4.07 ± 0.09 for LGG, $p = 0.0003$; $n = 120$). To judge whether this difference is clinically relevant, we now focus on the ability of the data to deliver a diagnosis on an individual subject basis. In general, the proposed method delivers multiple spectra and peak areas, and diagnosis may

DOI: 10.1002/nbm

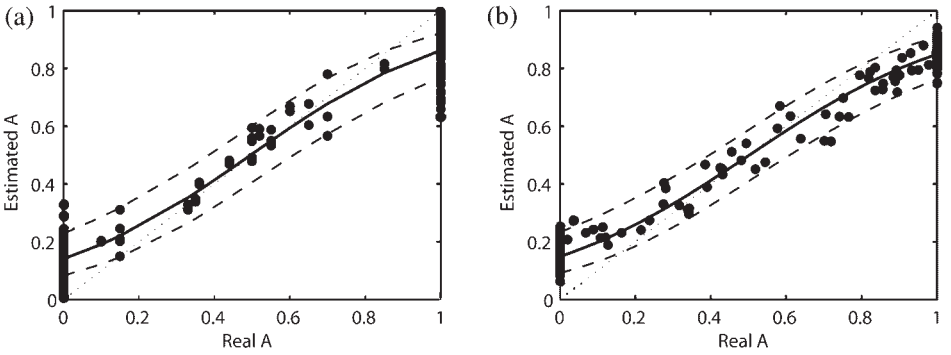

Figure 6. Validation of the estimated abundance, ***A***, in the phantom study (a) and simulated data (b). (a) The horizontal axis gives the actual volume fraction of the 'tumor-like' spectrum (real ***A***), which is calculated on the basis of the geometry of the phantom in Fig. 5. Pure voxels in the upper chamber corresponding to 'tumor-like' metabolite concentrations are given abundance values = 1, and pure voxels in the lower chamber have abundance values = 0. Voxels covering the chamber boundary have intermediate values based on the fraction of volume covered. The vertical axis gives the estimated abundance of 'tumor-like' spectrum as computed by the NMF algorithm (estimated ***A***). The dotted line indicates 100% accuracy of the estimation. Solid and dashed lines are the sigmoid fitting curve and the 95% confidence intervals, respectively. (b) The same plot as in (a) on simulated data with randomly chosen volume fractions and noise. The simulation was repeated 100 times to estimate the mean and confidence intervals. Sample points shown are a subset of these repetitions. The horizontal axis gives the predefined volume fractions. The vertical axis gives the abundance estimates computed by the NMF algorithm.

require a multivariate discrimination technique (33). On the present data, however, it was sufficient to consider the ratio between Cho and NAA peak areas. This specific linear discrimination criterion is captured by the CNI. The conventional figure of merit for classification, which captures sensitivity at various levels of specificity, is the area under the ROC curve (AUC) (30). ROC curves are shown in Fig. 7. Mean and standard deviation for the AUC were evaluated with a bootstrapping method (resampling from 20 cases with replacement) (37). The diagnostic specificity for distinguishing LGG from HGG is substantially increased from 0.83 to 0.89. To show a significant difference for this six point improvement in AUC would require at least 60 samples. This follows from a power analysis based on the test of DeLong *et al*. (38). We therefore simulated larger datasets with parameters and performance that match the clinical data and found a significant improvement in AUC ($p = 0.009$) with a sample size of 30 LGG and 30 HGG cases.

The first conclusion from Figs. 5 and 7 and Table 1 is that the extracted spectra are physiologically meaningful and can be given a clinically significant interpretation. Secondly, the reduced variability is reflected in improved diagnosis.

In this study, for simplicity, two constituent spectra were assigned either to normal or tumor tissue. As discussed above (see Experimental), for certain cases with clear lactate/lipid peaks (in two of the 20 subjects), one more component could be assigned and represented the high-lipid tumor region. Indeed, the inverted lipid peak is known to provide useful information for discrimination between HGG and LGG (33). For the present data, the inclusion of such a third component had no significant effect on discrimination performance (improvement from 0.89 to 0.92 in AUC). Future work with a larger subject population will consider algorithmic improvements that will capture the sign of inverted peaks.

Comparison of abundance estimates, *A*, with anatomical MRI data

To quantitatively evaluate the abundance estimates, we performed an ROC study using 'truth labels' for each voxel obtained from a neuroradiologist who had access to all available anatomical MRI data (see Experimental). The AUC from the 20 cases is 0.90 ± 0.13 (mean $\pm$ SD) for discriminating normal and tumor voxels and 0.91 ± 0.09 for discriminating HGG and LGG voxels.

Validation of abundance estimates, *A*, as volume fraction on a phantom

Having estimates of tumor abundance for each voxel suggests that it may be possible to detect the presence of a malignant tumor even in low volume fractions at which the present method based on CNI fails: a small volume fraction within a voxel reduces CNI for an otherwise highly malignant tumor. Therefore, accurate abundance estimates, ***A***, could be used for tumor detection in this scenario of low volume fraction. This

DOI: 10.1002/nbm

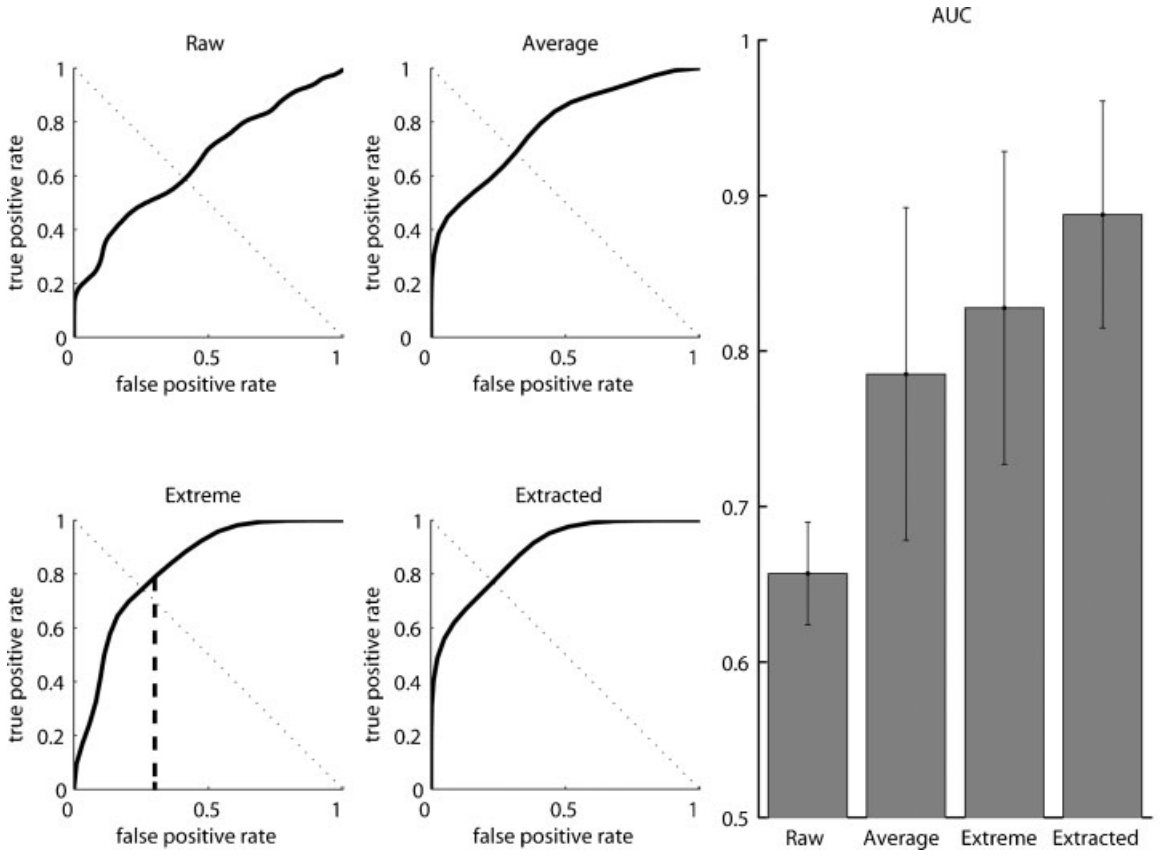

Figure 7. Classification of performance between HGG and LGG using CNI criterion. The four panels on the left show ROC curves estimated using a bootstrapping procedure. CNI diagnostic criterion (CNI = 3) is indicated as straight dashed line for the extreme spectra following clinical convention. Mean AUC for all four conditions is summarized on the right. Error bars indicate the standard deviation of the bootstrapping procedure. Diagnostic performance on these data is improved substantially by using extracted spectra (AUC = 0.89) as compared with the current clinical practice of using extreme spectra (AUC = 0.83).

should result in increased detection *sensitivity* in addition to the improvement in *specificity* resulting from reduced spectral variability of estimates, $\boldsymbol{S}$. This is particularly important given that diagnosis and treatment are typically based on the most malignant tumor, which may at first only be present in small volume fractions.

To validate the accuracy of the abundance estimates, we show in Fig. 5 the result from the phantom study. The NMF algorithm recovered spectral profiles with peak areas that correspond to the concentration in each solution. The 'normal' spectrum shows a displacement of the Cr peak, which would have been expected at 3 ppm. This shift is evident also in the original spectrum of the lower chamber (see Fig. 5). As the algorithm is adaptive, it has no difficulty in extracting this unexpected spectral profile. This shows that the method correctly decomposes the observed data into constituent spectra corresponding to the different solutions with their specific metabolite concentrations.

To quantify the accuracy of the abundance estimates, we show in Fig. 6a the relationship between the actual volume fraction covered by each voxel (real $\boldsymbol{A}$), as determined from the geometry of the overlay shown in Fig. 5, and the abundance estimates from the NMF algorithm (estimated $\boldsymbol{A}$). The result shows that the values are better estimated for intermediate-ranged abundances than for high and low abundance estimates, where one can see an upward and downward bias. We hypothesize

Table 1. Discriminability (AUC) for raw, average, extreme and extracted spectra (Fig. 4)

Comparison	Raw voxel spectra	Average spectra across voxels	Extreme voxel spectra	Extracted constituent spectra
Normal vs tumor	80%	90%	100%	100%
LGG vs HGG	67%	78%	83%	89%

DOI: 10.1002/nbm

that the origin of this estimation bias is a violation of the linearity assumption that is required for spectrum separation; the large confidence interval is due to noise. To test this hypothesis, we show in Fig. 6b the relationship between the actual volume fractions and the abundance estimates on simulated data with an SNR of 11.5 dB ($R^2 = 0.98$) comparable to the phantom data. The figure combines the results from multiple runs (100 repetitions, each with different randomly chosen volume fractions and Gaussian-distributed white noise). The result shows a similar pattern to the one in Fig. 6a. The same simulation study performed at 4 dB ($R^2 = 0.85$), which is the minimum required for the clinical data, shows a comparable bias with an increase in the confidence intervals. This indicates that the poor abundance estimates in high and low volume fractions are due to a systematic bias. A modified NMF separation algorithm, such as in (34), may be able to fix the bias because it does not require the non-negativity constraint. In fact, the bias may already be corrected, as it can be predicted from simulations. In contrast, the large confidence interval cannot be improved upon, as it results from the noise of the measurements. This noise is already evident in the pure voxel spectra, as shown in the left panel of Fig. 5.

Finally note that, confronted with noise (and distortions), the decomposition returns the correct spectra $\boldsymbol{S}$ (as reflected by correct peak ratios in Fig. 5) but makes a biased estimate on abundances $\boldsymbol{A}$. The linear decomposition permits $\boldsymbol{A}$ to be adjusted to every voxel, but it is forced to use the same $\boldsymbol{S}$ across all voxels. Hence, variability across voxels due to noise and distortions can only be captured by $\boldsymbol{A}$, leaving the estimate for $\boldsymbol{S}$ largely unaffected.

CONCLUSIONS

We have shown in the phantom study that the method correctly decomposes the observed data into constituent spectra corresponding to the different solutions with their specific metabolite concentrations. It validated the interpretation of abundance estimates, $\boldsymbol{A}$, as partial volume fraction and established bias and confidence intervals for its estimates. In addition, we have confirmed the physiological and clinical relevance of the extracted spectra, $\boldsymbol{S}$, by correlating the analysis results with pathologically proven tumor grades from 20 patients. Despite known tumor heterogeneity, we have shown improved correlation of tumor grade with spectral patterns (Cho concentration versus NAA concentration), supporting our hypothesis that some variability is due to the partial-volume effect. We have quantified the limitations of the method and found that a minimum SNR of 4 dB is required for at least a fraction of relevant voxels. Taken together, the results indicate that MRSI in combination with the proposed spectrum separation method may be useful in defining tumor margins for treatment planning of radiation therapy (39) or surgical resection.

Acknowledgements

We acknowledge Dr Dikoma Shungu from Weill Medical College of Cornell University for his help with preprocessing and technical details at various stages. Dr Truman Brown from Columbia University and Wei Yuan from City College of New York are gratefully acknowledged for help with the phantom study. This research was supported by the MSKCC-CCNY partnership grant (NIH/NCI U56 CA96299-0).

REFERENCES

1. Pirzkall A, Nelson SJ, McKnight TR, Takahashi MM, Li X, Graves EE, Verhey LJ, Wara WW, Larson DA, Sneed PK. Metabolic imaging of low-grade gliomas with three-dimensional magnetic resonance spectroscopy. *Int J Radiat Oncol Biol Phys* 2002; **53**(5): 1254–1264.
2. Moller-Hartmann W, Herminghaus S, Krings T, Marquardt G, Lanfermann H, Pilatus U, Zanella FE. Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions. *Neuroradiology* 2002; **44**(5): 371–381.
3. Howe FA, Barton SJ, Cudlip SA, Stubbs M, Saunders DE, Murphy M, Wilkins P, Opstad KS, Doyle VL, McLean MA, Bell BA, Griffiths JR. Metabolic profiles of human brain tumors using quantitative *in vivo* ^{1}H magnetic resonance spectroscopy. *Magn Reson Med* 2003; **49**(2): 223–232.
4. Liang Z, Lauterbur P. Mathematical fundamentals. In *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*, Liang Z, Lauterbur P (eds). IEEE Press: New York, 2000; pp 13–51.
5. Mierisova S, Ala-Korpela M. MR spectroscopy quantitation: a review of frequency domain methods. *NMR Biomed* 2001; **14**(4): 247–259.
6. Vanhamme L, Sundin T, van Hecke PV, van Huffel SV. MR spectroscopy quantitation: a review of time-domain methods. *NMR Biomed* 2001; **14**(4): 233–246.
7. in 't Zandt H, van Der Graaf M, Heerschap A. Common processing of *in vivo* MR spectra. *NMR Biomed* 2001; **14**(4): 224–232.
8. Hagberg G. From magnetic resonance spectroscopy to classification of tumors. A. review of pattern recognition methods. *NMR Biomed* 1998; **11**(4–5): 148–156.
9. el-Deredy W. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. *NMR Biomed* 1997; **10**(3): 99–124.
10. van der Veen JW, de Beer R, Luyten PR, van Ormondt D. Accurate quantification of *in vivo* ^{31}P NMR signals using the variable projection method and prior knowledge. *Magn Reson Med* 1988; **6**(1): 92–98.
11. Provencher SW. Automatic quantitation of localized *in vivo* ^{1}H spectra with LCModel. *NMR Biomed* 2001; **14**(4): 260–264.
12. Stoyanova R, Brown TR. NMR spectral quantitation by principal component analysis. *NMR Biomed* 2001; **14**(4): 271–277.
13. Kuesel AC, Stoyanova R, Aiken NR, Li CW, Szwergold BS, Shaller C, Brown TR. Quantitation of resonances in biological ^{31}P NMR spectra via principal component analysis: potential and limitations. *NMR Biomed* 1996; **9**(3): 93–104.
14. Menze BH, Lichy MP, Bachert P, Kelm BM, Schlemmer HP, Hamprecht FA. Optimal classification of long echo time *in vivo* magnetic resonance spectra in the detection of recurrent brain tumors. *NMR Biomed* 2006; **19**(5): 599–609.

DOI: 10.1002/nbm

15. Nuzillard D, Bourg S, Nuzillard J. Model-free analysis of mixtures by NMR using blind source separation. *J. Magn Reson* 1998; **133**(2): 358–363.
16. Ladroue C, Howe FA, Griffiths JR, Tate AR. Independent component analysis for automated decomposition of *in vivo* magnetic resonance spectra. *Magn Reson Med* 2003; **50**(4): 697–703.
17. Sajda P, Du S, Brown TR, Stoyanova R, Shungu DC, Mao X, Parra /SNM> 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**(12): 1453–1465.
18. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature* 1999; **401**(6755): 788–791.
19. Laudadio T, Pels P, De Lathauwer L, Van Hecke P, Van Huffel S. Tissue segmentation and classification of MRSI data using canonical correlation analysis. *Magn Reson Med* 2005; **54**(6): 1519–1529.
20. Brown TR, Stoyanova R. NMR spectral quantitation by principal-component analysis. II. Determination of frequency and phase shifts. *J. Magn Reson* 1996; **112**(1): 32–43.
21. Witjes H, Melssen WJ, in 't Zandt HJ, van der Graaf M, Heerschap A, Buydens LM. Automatic correction for phase shifts, frequency shifts, and lineshape distortions across a series of single resonance lines in large spectral data sets. *J. Magn Reson* 2000; **144**(1): 35–44.
22. McKnight TR, Noworolski SM, Vigneron DB, Nelson SJ. An automated technique for the quantitative assessment of 3D-MRSI data from patients with glioma. *J. Magn Reson Imaging* 2001; **13**(2): 167–177.
23. McKnight TR, von dem Bussche MH, Vigneron DB, Lu Y, Berger MS, McDermott MW, Dillon WP, Graves EE, Pirzkall A, Nelson SJ. Histopathological validation of a three-dimensional magnetic resonance spectroscopy index as a predictor of tumor presence. *J. Neurosurg* 2002; **97**(4): 794–802.
24. Park I, Tamai G, Lee MC, Chuang CF, Chang SM, Berger MS, Nelson SJ, Pirzkall A. Patterns of recurrence analysis in newly diagnosed glioblastoma multiforme after three-dimensional conformal radiation therapy with respect to pre-radiation therapy magnetic resonance spectroscopic findings. *Int J Radiat Oncol Biol Phys* 2007; **69**(2): 381–389.
25. MuCullagh P, Nelder JA. *Generalized Linear Models*. Chapman & Hall: London, 1990.
26. Dobson AJ. *An Introduction to Generalized Linear Models*. CRC Press: Boca Raton, FL, 1990.
27. Collett D. *Modelling Binary Data*. Chapman & Hall/CRC Press: London, 2002.
28. Sajda P. Machine learning for detection and diagnosis of disease. *Annu Rev Biomed Eng* 2006; **8**: 537–565.
29. Jolliffe IT. *Principal Component Analysis*. SpringerVerlag: New York, 1986.
30. Mason SJ, Graham NE. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society* 2002; **128**: 2145–2166.
31. Shimizu H, Kumabe T, Shirane R, Yoshimoto T. Correlation between choline level measured by proton MR spectroscopy and Ki-67 labeling index in gliomas. *AJNR Am J Neuroradiol* 2000; **21**(4): 659–665.
32. Furuya S, Naruse S, Ide M, Morishita H, Kizu O, Ueda S, Maeda T. Evaluation of metabolic heterogeneity in brain tumors using ^{1}H-chemical shift imaging method. *NMR Biomed* 1997; **10**(1): 25–30.
33. De Edelenyi FS, Rubin C, Esteve F, Grand S, Decorps M, Lefournier V, Le Bas JF, Remy C. A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images. *Nat Med* 2000; **6**(11): 1287–1289.
34. Ding C, Li T, Jordan M. *Convex and Semi-Nonnegative Matrix Factorizations*. Lawrence Berkeley National Laboratory, University of: California: Berkeley, 2006.
35. Nelson SJ. Multivoxel magnetic resonance spectroscopy of brain tumors. *Mol Cancer Ther* 2003; **2**(5): 497–507.
36. Howe FA, Opstad KS. ^{1}H MR spectroscopy of brain tumours and masses. *NMR Biomed* 2003; **16**(3): 123–131.
37. Polikar R. Bootstrap-inspired techniques in computational intelligence: ensemble of classifiers for incremental learning, data fusion and missing feature analysis. *IEEE Signal Processing Magazine* 2007; **24**(4): 59–72.
38. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**: 837–845.
39. Graves EE, Pirzkall A, Mcknight TR, Vigneron DB, Larson DA, Verhey LJ, Mcdermott M, Chang S, Nelson SJ. Use of proton magnetic resonance spectroscopic imaging data in planning focal radiation therapies for brain tumors. *Image Anal Stereol* 2002; **21**: 69–76.

DOI: 10.1002/nbm