IEEE Journal of Translational Engineering in Health and Medicine

Received 25 September 2013; revised 20 January 2014; accepted 4 February 2014; Date of publication 3 March 2014; date of current version 19 March 2014.

Digital Object Identifier 10.1109/JTEHM.2014.2309333

In Vivo Brain Magnetic Resonance Spectroscopy: A Measurement of Biomarker Sensitivity to Post-Processing Algorithms

DANIEL COCUZZO[1], ALEXANDER LIN[2], (Member, IEEE), PETER STANWELL[3], CAROLYN MOUNTFORD[2,4], AND NIRMAL KESHAVA[5], (Member, IEEE)

[1]Department of Computer Science, Stanford University, Palo Alto, CA 80523, USA
[2]Centre for Clinical Spectroscopy, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215, USA
[3]School of Health Sciences, University of Newcastle, Callaghan, NSW 2308, Australia
[4]Centre for MR in Health, Faculty of Health, University of Newcastle, Callaghan, NSW 2308, Australia
[5]Department of Research and Development Information, AstraZeneca Pharmaceuticals, Waltham, MA 02451, USA

CORRESPONDING AUTHOR: N. KESHAVA (nkeshava67@gmail.com)

This work was supported by the Department of Defense under Grant W81XWH-10-1-0785.

ABSTRACT Clinical translation of reported biomarkers requires reliable and consistent algorithms to derive biomarkers. However, the literature reports statistically significant differences between 1-D MRS measurements from control groups and subjects with disease states but frequently provides little information on the algorithms and parameters used to process the data. The sensitivity of *in vivo* brain magnetic resonance spectroscopy biomarkers is investigated with respect to parameter values for two key stages of post-acquisitional processing. Our effort is specifically motivated by the lack of consensus on approaches and parameter values for the two critical operations, water resonance removal, and baseline correction. The different stages of data processing also introduce varying levels of uncertainty and arbitrary selection of parameter values can significantly underutilize the intrinsic differences between two classes of signals. The sensitivity of biomarkers points to the need for a better understanding of how all stages of post-acquisitional processing affect biomarker discovery and ultimately, clinical translation. Our results also highlight the possibility of optimizing biomarker discovery by the careful selection of parameters that best reveal class differences. Using previously reported data and biomarkers, our results demonstrate that small changes in parameter values affect the statistical significance and corresponding effect size of biomarkers. Consequently, it is possible to increase the strength of biomarkers by selecting optimal parameter values in different spectral intervals. Our analyses with a previously reported data set demonstrate an increase in effect sizes for wavelet-based biomarkers of up to 36%, with increases in classification performance of up to 12%.

INDEX TERMS Magnetic resonance spectroscopy, biomarkers, sensitivity, neuroimaging, statistical significance.

I. INTRODUCTION

Successful use of biomarkers to accurately monitor changes in health requires differences to be consistently measured in clinical settings. To be reliably used, however, the end-to-end system of measurement and the subsequent post-processing should be accurately characterized to assure that prospective biomarkers are robust to variability introduced at every stage. An accurate accounting of system-wide variance can maximize the use of intrinsic physiological differences, highlight specific sources of variance, and predict the performance of biomarkers in clinical environments. Here we report our effort to characterize the sensitivity of biomarkers reported previously in [1] from in vivo magnetic resonance spectroscopy (MRS), by measuring their change when key post-acquisitional processing parameters are varied. Our work provides a foundation for comprehensive statistical modeling

that can forecast and predict overall system performance as well as identify the sensitivities of prospective biomarkers through targeted trade studies. Efforts in other disciplines have capitalized on defining parameterized, end-to-end models [2], [3] and have become critical components for advancing technologies from demonstrations to implemented systems.

A. CLINICAL TRANSLATION OF MRS

Magnetic resonance imaging has been widely used for clinical diagnosis by providing a non-invasive and non-ionizing method of visualizing body tissue structure and the alterations brought about by disease. More recently, advanced MR techniques have provided additional functional information, particularly in the brain, yielding measures of diffusion, blood flow, and metabolism. Specifically, MRS has been demonstrated to provide a non-invasive means of measuring brain biochemistry and by doing so provides a ''virtual biopsy'' to monitor a range of neurological diseases such as brain cancer, Alzheimer's disease, brain injury, and chronic pain [4]. Most importantly, the biochemical changes observed by MRS often precede morphological and other functional changes, thus providing earlier diagnosis and more effective therapeutic monitoring [5]. In clinical practice, changes in the spectral pattern or semi-quantitative approaches are used [6]. While the current approach is effective for diseases with gross metabolic changes, such as brain tumors [7], it is less effective for more subtle metabolic changes that occur in diseases such as schizophrenia [8], autism [9], and concussion [10]. In these neurological disorders, improvements in sensitivity would have a significant impact on the diagnostic use of MRS. Therefore, there is an important need to comprehensively understand how the tools for measurement and analysis can expose and manage these small differences for clinical implementation.

In this study, we have focused on a specific neurological condition: chronic pain in spinal cord injury (SCI). Recent studies have shown that SCI patients suffer from a high incidence of chronic pain after their injury [11]. Numerous studies have shown neurochemical alterations in chronic pain using MRS, such as increases in glutamate (Glu) and gamma-amino butyric acid (GABA) [12]. One of the primary objectives in understanding the central mechanism of SCI-related pain is the delineation of the differences in neurochemistry between those individuals with and without spinal cord injury to identify biomarkers to better understand the underlying pathophysiological changes that occur in this disorder.

B. POST-ACQUISITIONAL PROCESSING APPROACHES

Post-acquisitional processing of MRS signals consists of the algorithms applied to individual MRS signals, whereby different distortions are successively removed from signals and biomarkers are identified to enable a numerical inference on one signal or between two or more cohorts. Jiru [13] published a review of a subset of algorithms that comprise post-acquisitional processing for MRS signals, describing the origin of the distortion to be mitigated as well as some of the technical approaches used for corrections. The wide range of effects of post-acquisitional processing for biological material was described in [14].

The key analytical post-acquisitional processing platforms for MRS have focused on operations enabling the estimate of absolute and relative chemical concentrations from individual signals, where the two main approaches are time-domain and frequency-domain methods. Time domain methods have the advantage of easily managing missing data points [15] as well as affording greater flexibility to the model function adopted in contrast to frequency domain methods [16]. Frequency domain methods have the advantage of being naturally suited to frequency selective analysis which can reduce the number of model parameters, hence reducing computational demands.

A popular technique for time-domain analysis is MRUI-AMARES [17], [18] (Magnetic Resonance User Interface using Advanced Method for Accurate, Robust and Efficient Spectral fitting), an interactive, time-domain method which fits a parameterized model function to the data by minimizing a cost function [19]. AMARES is typically preceded by algorithms to mitigate distortions. Another time-domain approach in MRUI, frequently used, is HSVD [20] which performs a subspace-based inversion of individual MRS spectra to recover the least-squares-based estimates of resonance parameters, presumably also following the application of algorithms to remove different distortions. The most popular frequency-domain technique is LCModel (Linear Combination Model) [21] which uses a linear combination of model spectral basis sets composed of individual metabolites to fit in vivo data in the frequency domain. In contrast to AMARES, LCModel does not utilize algorithms before its application to eliminate distortions, choosing instead to estimate key values during optimization. Many of these inference techniques are accompanied by Cramer-Rao bounds to provide an overall measure of the precision of the algorithm, but they do not fold in broader dependencies, such as preceding algorithms or model-order selection.

While there are inherent differences in the mathematical approaches used by these parameter estimation algorithms, there are also major user-oriented issues as well. It is likely that these practical issues have driven the popularity of LCModel, which requires the selection of an appropriate basis set (dependent on the data acquisition parameters such as field strength, echo time, localization sequence) and optional loading of the FID with a water reference. Although there is the possibility for additional user interaction, little more is required for obtaining relative metabolite concentrations from in vivo MRS data. The drawback to such a method is that less is known about what specific post-processing routines were employed to obtain those measurements. In contrast, MRUI-AMARES is interactive, allowing the user full control over all the post-processing routines. The drawback to this level of detail is that it requires user input at each step, which results in user-dependencies that can lead to greater

IEEE Journal of Translational Engineering in Health and Medicine

variability of measurements [22], [23]. Moreover, there can be a number of issues regarding the sensitivity of the results from these approaches due to a variety of user-determined parameters, such as the number of resonance modes in the models, the number of spectra in basis sets, prior knowledge, etc.

C. PREVIOUS EFFORTS

We build upon earlier attempts to gauge the sensitivity of MRS measurements to different factors. The effect of different amounts of prior knowledge and windowing was previously explored by [24] to determine how different factors affect the precision of quantitation for 15 metabolites when recorded from repeated scans collected successively from a single subject. The quantitation was undertaken using a non-linear fitting algorithm and demonstrated that changes in processing techniques at different stages, as well as different amounts of prior knowledge, yielded different mean values and standard deviations for estimates of metabolite concentrations. For many chemical biomarkers and processing stages, the differences were statistically significant. Further, the inaccurate modeling of resonances outside the spectral regions of interest could negatively impact estimates of concentration within the region of interest. Similarly, the application of rectangular and exponential windows of different lengths to the time series demonstrated significant changes to estimates of metabolite levels. Subsequent efforts [25] demonstrated an increase in precision when field-specific prior knowledge was utilized, but significant differences were observed during different acquisition conditions, and optimal parameter values for windowing and model function selection remained unspecified.

Kreis addressed issues of spectral quality in proton MR spectroscopy by examining the relative utility of different definitions of signal quality, namely SNR, in time-domain as well as frequency domain, as part of a broader attempt to characterize a variety of artifacts in different stages of MRS acquisition and processing [26]. More importantly, an initial attempt was made to categorize and concatenate the different sources of error that are encountered during absolute chemical quantitation, ranging from hardware, acquisition, and processing and analysis. Somorjai [27] demonstrated the sensitivity of classification accuracy to post-acquisitional processing, and in particular to strategies for normalization of spectra. He further demonstrated the importance of considering how MRS spectra are processed after acquisition, albeit with *ex vivo* brain tissue samples [28], with a conclusion that the mathematical normalization of MRS spectra greatly affects classification performance.

Assessments of quantitation precision based on baseline removal were investigated by Elster and colleagues [29] where they used semi-parametric models as a mechanism to address uncertainty in making quantitative measurements with MRS. They also addressed efforts to improve Cramer-Rao bounds when using metabolite basis sets to avoid overly optimistic estimates of precision [30]. Sorensen [31] evaluated MRS in comparison to other imaging modalities for its comparative effectiveness as a cancer imaging biomarker both for diagnosis as well as for evaluating response-to-therapy. The conclusion was that the lack of standardization, and the resulting lack of reproducibility, has limited its acceptance, and that significant effort is required to determine the precise technique to be used as well as how it should be utilized in a clinical setting.

Our goal is to numerically assess the sensitivity of MRS biomarkers to processing algorithms and parameters and thereby add rigor to biomarker discovery and validation. To achieve this goal we employ a unique and novel algorithmic platform by which analyses of spectra can be undertaken to identify differences between two or more groups and thus gauge the sensitivity of these differences. Algorithms at each stage of processing are parameterized and enable performance comparisons over ranges of values. The platform can also extract both pattern recognition-based and model-based features that provide complementary approaches to interpreting MRS data.

We reported preliminary results of biomarker sensitivity to different parameter values for water resonance removal in [32]. This paper significantly expands on that effort by also considering baseline removal algorithms, as well as their interaction. We also quantify the impact of optimizing post-processing algorithm parameter values on classification rates.

II. METHODS AND PROCEDURES

A. PARTICIPANTS AND RECRUITMENT

Here we use the same data set previously reported [1] to investigate a completely separate, but equally important, topic: how the sensitivity of different stages of post-acquisitional processing affects biomarker discovery and clinical translation. The participants in [1] consisted of two groups, a SCI group ($n = 10$) with a mean age of 36.4 years ($SD = 10.4$ years) and an able-bodied control group ($n = 10$), that were age (± 5 years) and gender matched with a mean age of 27.4 years ($SD = 6.8$ years) ($p>0.05$). The SCI participants had injuries sustained more than 12 months prior to testing (mean = 5.1 years, $SD = 4.3$ years) and they were all diagnosed as having complete thoracic lesions using the American Spinal Injury Association (ASIA) impairment scale as assessed at specialist spinal injury units. For SCI with pain the mean duration of condition was 63.6 months since injury ($SD = 49.6$), and for SCI without pain the mean duration of condition 58.2 months since injury ($SD = 59.7$) ($p > 0.05$). Exclusion criteria for the SCI-group included incomplete SCI and/or tetraplegia, history of psychopathology, previous pain syndromes prior to their injury, epilepsy, or the presence of non MR-compatible devices. Other information, such as history of traumatic brain injury (TBI), current metallic implants and medications were also recorded to standardize result reporting. The study had Institutional Review Board approval and had the equivalence of HIPPA compliance.

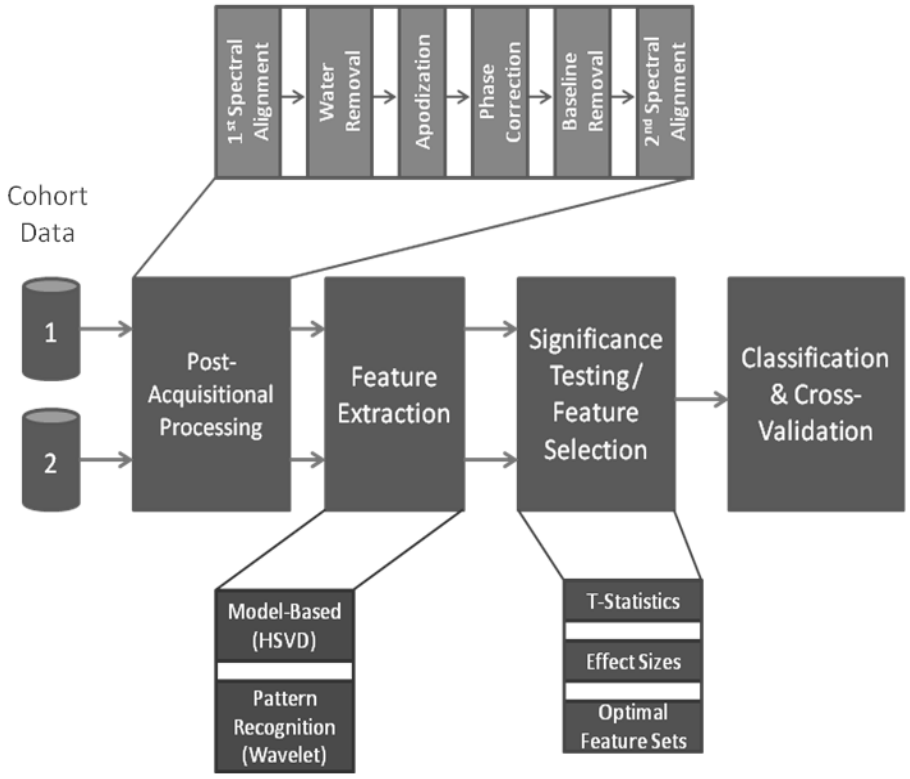

FIGURE 1. **Processing pipeline architecture describing steps of post-acquisitional processing for individual spectra as well as different types of feature extraction approaches. Cohorts of MRS measurements are subsequently analyzed with significance testing and feature selection to identify the best combination of features to be submitted to classification and cross-validation algorithms.**

B. MAGNETIC RESONANCE

MR spectra were acquired on a 3 Tesla Siemens Trio scanner (Siemens AG, Erlangen, Germany), equipped with a standard quadrature head coil. Localizer images were initially acquired in three orthogonal planes (transverse, sagittal, and coronal) and were used for placing the collection voxel within the discrete anatomic locations of the left thalamus (THA). MR spectra were acquired using the stimulated echo acquisition mode method with single volume selection (TR/TE 3000/20ms; 2048 FID points, bandwidth 2 kHz; 6 cm^3 voxel). Before each acquisition, localized shimming was performed to optimize final spectral resolution. Adjustments of zero- and first-order shim gradients using the automatic B_0-field mapping technique supplied by the vendor (Siemens AG, Erlangen, Germany) were performed, followed by manual adjustment of zero-order shim gradients to achieve a resulting peak width of water at half-maximum that was 16 Hz or better for all voxels. The WET-technique was used to selectively suppress the water component of each FID signal [33].

C. THEORY

In vivo MR spectroscopy signals of the brain are a superposition of the resonances from several brain metabolites, in addition to a variety of distortions and artifacts introduced during acquisition and from certain macromolecules. The classical model for MRS signals is:

$$x(n) = \sum_{q=1}^{Q} \rho_q \lambda_q^t + \epsilon(t_n), \qquad \lambda_q^t = e^{(\alpha_q + i\omega_q)t_n} \tag{1}$$

where $n = 0, \ldots, N-1$ samples, Q is the total number of sinusoidal components, $(\rho_q, \alpha_q, \omega_q)$ are the complex amplitude (includes phase), damping, and frequency of the qth component, $\epsilon(t_n)$ is the noise contribution, and $t_n = n\Delta t$, where Δt is the sampling rate [20]. This linear model with Lorentzian line-shapes and Gaussian noise has been the basis for most model-based processing of MRS signals [34].

D. ALGORITHMS

We utilize a comprehensive suite of MRS biomarker discovery and statistical classification algorithms previously employed in [1]. The overall processing pipeline architecture is shown in Fig. 1, consisting of consecutive stages of fully parameterized post-processing algorithms that serially mitigate distortions in MRS signals, followed by feature extraction, significance testing, feature selection, and classification algorithms. We examine two of the post-acquisitional processing steps, baseline removal and water resonance removal, to determine how sensitive prospective biomarkers are to changes in their respective parameter values. These steps were selected because both

procedures utilized parameterized algorithms for mitigating distinct distortions. Additionally, the literature reflects little consensus on optimal ways of implementing or using these algorithms. The specific algorithms in each of these steps were the ones used in [1], but the same methods for evaluating sensitivity can be adapted for any other algorithms.

One of the main challenges facing the clinical translation of MRS is extracting features that are directly attributable to specific pathologies while also providing specificity between different disease states. As depicted in Fig. 1, we consider two complementary types of features. Model-based features are derived mathematically from differential equations describing magnetic resonance and are the estimated parameter values for specific chemical resonances (e.g., N-acetyl aspartate, creatine, choline, etc.) that comprise brain MRS spectra. Deviations in specific chemical concentrations have been associated with different types of brain injuries [35] and neurodegeneration [36].

Model-based features, however, can be limited in their ability to capture changes in all parts of the MRS spectrum, particularly in locations where chemical resonances are unassigned [37]. An alternative is a machine learning-based approach that objectively evaluates consecutive, adjacent intervals in an entire MRS spectrum, independent of their proximity to major chemical resonances. Though the feature values do not directly correspond to specific chemical levels and are less attributable to neurophysiological changes, previous efforts have reported that combinations of features can be diagnostic [38].

Since the publication of the results in [1], the processing pipeline has been augmented to obtain greater accuracy in removing distortions and more flexibility in deriving features that can discriminate the two classes. As a consequence of these upgrades, the biomarkers and resulting effect sizes we examine here are not identical to those reported in [1], although in many spectral intervals they are very similar. Nevertheless, our investigation utilizes the same core algorithms and the identical parameters used in [1].

1) Water Resonance Removal

Raw proton MR spectra of the brain contain a dominant water resonance structure whose intensity is multiple orders of magnitude larger than the brain metabolites that we are interested in analyzing. While most of the water signal is suppressed at time of acquisition using narrow band radiofrequency chemical-shift selective (CHESS) pulses followed by spoiling gradients, the actual degree of water suppression depends on *in vivo* shim and other factors. Hence, there is still a remaining water signal thus facilitating the need for a water removal algorithm. The processing pipeline implements a model-based water removal algorithm that estimates and removes the modes that comprise the resonance structure of water. The implementation described by [20] exploits the state-space formalism for harmonic retrieval discussed in [39]. For each individual FID, $\mathrm{x_i}$, we form the Hankel data matrix,

$$\mathrm{X_i} = \begin{bmatrix} \mathrm{x_i}(0) & \mathrm{x_i}(1)\cdots & \mathrm{x_i}(\mathrm{M}-1) \\ \mathrm{x_i}(1) & \mathrm{x_i}(2)\cdots & \mathrm{x_i}(\mathrm{M}) \\ \vdots & \ddots & \vdots \\ \mathrm{x_i}(\mathrm{L}-2) & \mathrm{x_i}(\mathrm{L}-1)\cdots & \mathrm{x_i}(\mathrm{N}-2) \\ \mathrm{x_i}(\mathrm{L}-1) & \mathrm{x_i}(\mathrm{L})\cdots & \mathrm{x_i}(\mathrm{N}-1) \end{bmatrix} \tag{2}$$

where the total number of data points $N = L + M$; we have chosen $L = M = N/2 = 1024$ for our implementation. Next, we compute the singular value decomposition (SVD) of each Hankel matrix,

$$\mathrm{X_i} = \mathrm{U_{X_i}} \mathrm{S_{X_i}} \mathrm{V_{X_i}^H} \tag{3}$$

where $\mathrm{U_{X_i}}$ is a $\mathrm{L} \times \mathrm{L}$ matrix of orthonormal output basis column vectors, $\mathrm{S_{X_i}}$ is a $\mathrm{L} \times \mathrm{M}$ diagonal matrix of the singular values of $\mathrm{X_i}$, and $\mathrm{V_{X_i}}$ is a $\mathrm{M} \times \mathrm{M}$ matrix of orthonormal input basis column vectors. Resonance parameters are estimated by optimizing the projection of model functions against the K-dimensional subspaces using the K leading columns of $\mathrm{U_{X_i}}$ [40]. Subsequently, a subset of water-based modes is identified as those resonances whose frequencies reside within a pre-determined spectral interval at $\pm Q$ Hz surrounding the known resonance location of water.

2) Baseline Removal

Raw MR spectra exhibit a rolling baseline trend due to acquisition noise, as well as contributions from broadband resonance structures of lipids, macromolecules, and metabolites with low concentrations when using short echo time (TE) MRS. While long echo (TE > 50 ms) MRS does not suffer from this problem, it is also much more limited in the number of metabolites of interest. Short-echo MRS allows for the detection of important brain metabolites such as glutamate, myo-inositol, and glutathione, that have short T2 relaxation that are no longer present at long echo. Therefore, the need for baseline correction or removal is important. The processing pipeline implements a time-varying baseline removal filter that has been successfully used to eliminate baselines [41] and works by estimating the baseline value at each point, $b(n), n = 1, \ldots, N$ based on local estimates from the surrounding neighborhood:

$$\hat{b}_i(n) = cdf\left(\hat{x}_\mathrm{i}(\mathrm{k}) : \mathrm{n} - \frac{\mathrm{W}}{2} \le \mathrm{k} \le \mathrm{n} + \frac{\mathrm{W}}{2} | \mathrm{W}, \alpha\right), \quad n = 1, \ldots, N \tag{4}$$

$$\hat{r}_i(n) = \hat{x}_\mathrm{i}(\mathrm{n}) - \hat{b}_i(n), n = 1, \ldots, N \tag{5}$$

where $\hat{b}_i(n)$ estimates the value of the cumulative distribution function (cdf) of a processed signal, $\hat{x}_i(n)$, in a windowed region of width, W, around n at the value of $\alpha, 0 \le \alpha \le 1$. Larger values of W derive estimates from wider neighborhoods, while larger values of α presume signals with effectively larger amounts of noise.

From a clinical perspective, water removal and baseline removal are important as they can significantly alter the appearance of the spectrum as well as subsequent estimates

of chemical concentrations. In an ideal spectrum, the baseline should be flat with resonances resolved to the baseline and having no overlap with adjacent resonances, such that measurement of the peak area within a certain frequency would be selectively reflective of specific metabolites and is well-resolved to allow for integration of peak areas without spectral overlap. However, noise, macromolecules and lipids with short T2 relaxation times can induce broad resonances that affect the baseline, especially with short-echo time acquisitions as used in the data analyzed in this study. Establishing the appropriate baseline parameters can be challenging as the broad underlying lipid and macromolecular resonances may themselves be diagnostic and therefore aggressive removal of the baseline could potentially remove additional diagnostic biomarkers [42]. Similarly, water removal methods are employed to analyze the metabolites that are close to the water resonance at 4.7 ppm such as myo-inositol and the secondary resonances of glutamate, glutamine, and creatine. If water is removed too aggressively, the concentrations of those metabolites can be underestimated and vice versa.

3) Wavelet-Based Features

The wavelet-based feature extraction stage in Fig. 1 that was utilized in [1] is a wavelet decomposition technique based upon the well-known scaling and dilation equations that give rise to a basis:

$$\phi(t) = 2\sum c(k)\phi(2t-k) \tag{6}$$

$$\psi(t) = 2\sum d(k)\phi(2t-k) \tag{7}$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{8}$$

Here, $\{c, d\}$ are scaling and wavelet coefficient filter banks that operate upon a scaling function, $\phi(t)$, and corresponding wavelet function, $\psi(t)$. We utilized a Haar mother wavelet at a depth of 1. As shown in Fig. 2, when used to transform a function, the Haar scaling functions can isolate local spectral amplitudes, while the wavelet functions can characterize local fluctuations. The original wavelet transform method described in [1] was subsequently augmented as a stationary wavelet transform [43], introducing significantly more features by retaining all integer translations of basis functions at every scale. Although the resulting transform is no longer orthonormal, the overcomplete decomposition permits localized differences to be more precisely isolated.

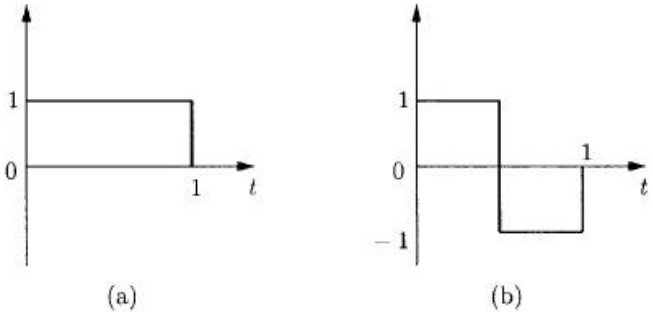

FIGURE 2. **(a) Haar scaling function; (b) Haar wavelet function.**

4) Model-Based Features

In contrast to the wavelet-based features, model-based feature extraction based on HSVD [20], [39] were implemented to directly estimate the four parameters for each chemical resonance that comprises MRS spectra: resonance amplitudes, frequencies, decays, and phases. Spectral estimation algorithms capitalize upon idealized models of the linear superposition of different chemical resonances. When estimated resonances are matched across multiple spectra the mode parameters become features that can be tested to find differences between groups that indicate altered brain biochemistry.

Similar to the water removal algorithm, the model-based HSVD technique estimates a specified number of resonance structures across the entire spectral range by computing the singular value decomposition of the Hankel data matrix, $X = USV^H$, of each individual FID and retaining the five leading columns of the left singular vector matrix, U_5 and the right singular matrix, V_S^H. The estimated model-based harmonic parameters for each of the five modes are obtained by performing a least-squares projection that maximizes the representation of the subspaces in U_5 and V_5^H by a Vandermonde matrix representing the sampling of a harmonic process given by Equation 1. Estimated resonance structures that are located within a tolerance ($\pm$0.1 ppm) to major metabolite peaks are matched accordingly. Spectral estimation approaches can be sensitive to noise, however, and are not guaranteed to identify the same modes in different signals.

5) Significance Testing

Candidate biomarkers are evaluated using a two-sided, equal variance (unpooled) student's t-test to compare the statistical significance of features ($p \leq 0.05$) and to estimate corresponding effect sizes. To ensure that significant wavelet-based features in [1] were not attributable to random noise effects, significant wavelet-based features were also required to exceed 5% of the NAA peak amplitude. For this analysis, no methods to correct for multiple comparisons were utilized, e.g., Bonferroni correction.

E. PERTURBATION OF ALGORITHM PARAMETERS

The results reported in [1] provide a set of default algorithm parameters and default biomarkers that can be used as a reference point to determine the sensitivity of the reported biomarkers in terms of their statistical significance and their effect size as a function of parameters for water removal and baseline removal. Table 1 summarizes the default parameter values that were utilized for water removal and baseline removal in [1] and the range of values that were considered for each parameter; the parameters in Table 1 correspond to those that appear above in the methods section. For each processing stage, the associated pair of parameters was varied and the statistical significance of resulting biomarkers was evaluated. A two-dimensional *heatmap* illustrates the resulting surface of statistical values generated by sweeping the parameters

IEEE Journal of Translational Engineering in Health and Medicine

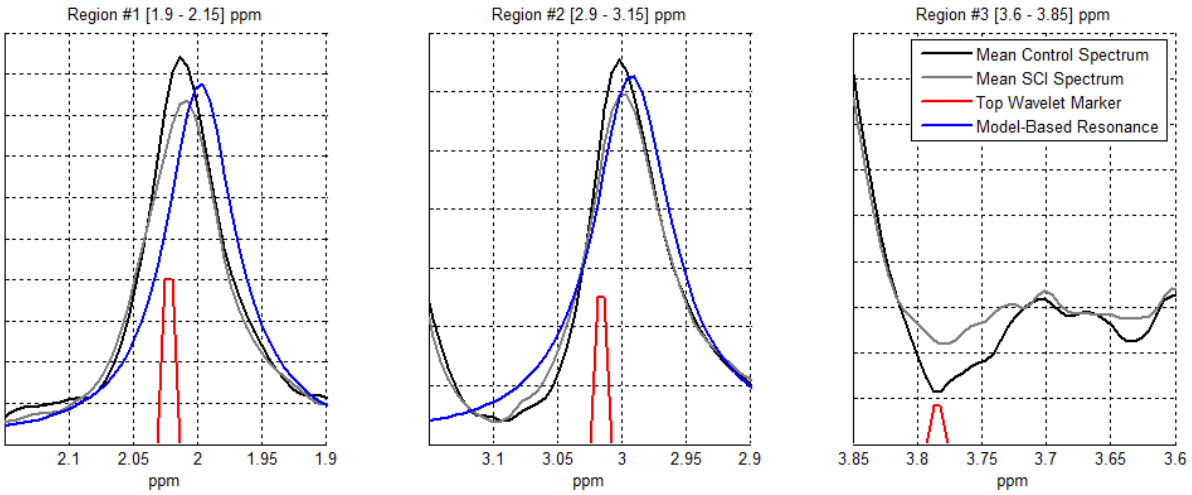

FIGURE 3. Wavelet-based and model based biomarkers obtained for Region 1, 2, and 3 respectively, using the default processing parameter values. Regions 1 and 2 correspond to NAA and creatine resonances, respectively; Region 3 does not contain an apparent chemical resonance.

TABLE 1. The default processing parameter value ranges and increments used to generate heatmap response surfaces of the water removal and baseline removal stages.

Processing Parameter Value Ranges & Increments						
Processing Stage	**Parameter**	**Default Value**	**Min Value**	**Max Value**	**Increment**	**# of Trials**
Water Removal	K	**10 terms**	5	15	1	11
	Q	**± 25 Hz**	5	60	5	12
Baseline Removal	W	**101 spectral points**	51	201	10	16
	α	**0.15**	0.05	0.4	0.05	8

over a range of values detailed in Table 1. The heatmaps are annotated to indicate the location of default parameter values (i.e., those used in [1]) and the values that achieved the maximum and minimum significance levels and effect sizes.

In addition to analyzing individual processing stages in isolation, we investigated the composite sensitivity of both stages by varying the processing parameters of the baseline removal and water removal stages concurrently. We evaluated 16,896 ($11 \times 12 \times 16 \times 8$) total trials that span all possible parameter values across the combined four parameters of these two stages, shown in Table 1, using a networked cluster of 32 CPU cores to achieve a feasible execution time. For the best and worst-performing trials, the parameter values, corresponding effect sizes, and percentage change in significance as compared to the biomarker performance at default parameter values are also reported.

III. RESULTS

A. DEFAULT PARAMETERS AND BIOMARKERS

Fig. 3 illustrates three distinct spectral intervals in which wavelet-based biomarkers had statistically significant differences ($p \leq 0.05$) that discriminated healthy subjects from those having spinal cord injury. Each significant feature occurred within a contiguous cluster of several adjacent, statistically significant wavelet-based features, revealing an extended interval of structural difference. In order to maximize the discrimination between classes, the specific features chosen in Fig. 3 had the highest effect size in the cluster. Table 2 shows the respective locations, p-values, and effect sizes of these biomarkers from the three chosen regions. Region 1 & Region 2 span spectral intervals in which prominent resonances, N-acetyl aspartate (NAA) and creatine (Cr) respectively, are clearly present. Region 3 spans a spectral interval in which the mean intensity spectrum for each class

TABLE 2. Wavelet biomarker statistics obtained using the default processing parameters. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$.

Default Wavelet-Based Biomarkers and Statistics			
Region	**PPM Location**	**|Effect-Size|**	**P-Value**
1	2.02	1.68**	0.002
2	3.02	1.29*	0.013
3	3.78	2.20**	0.000

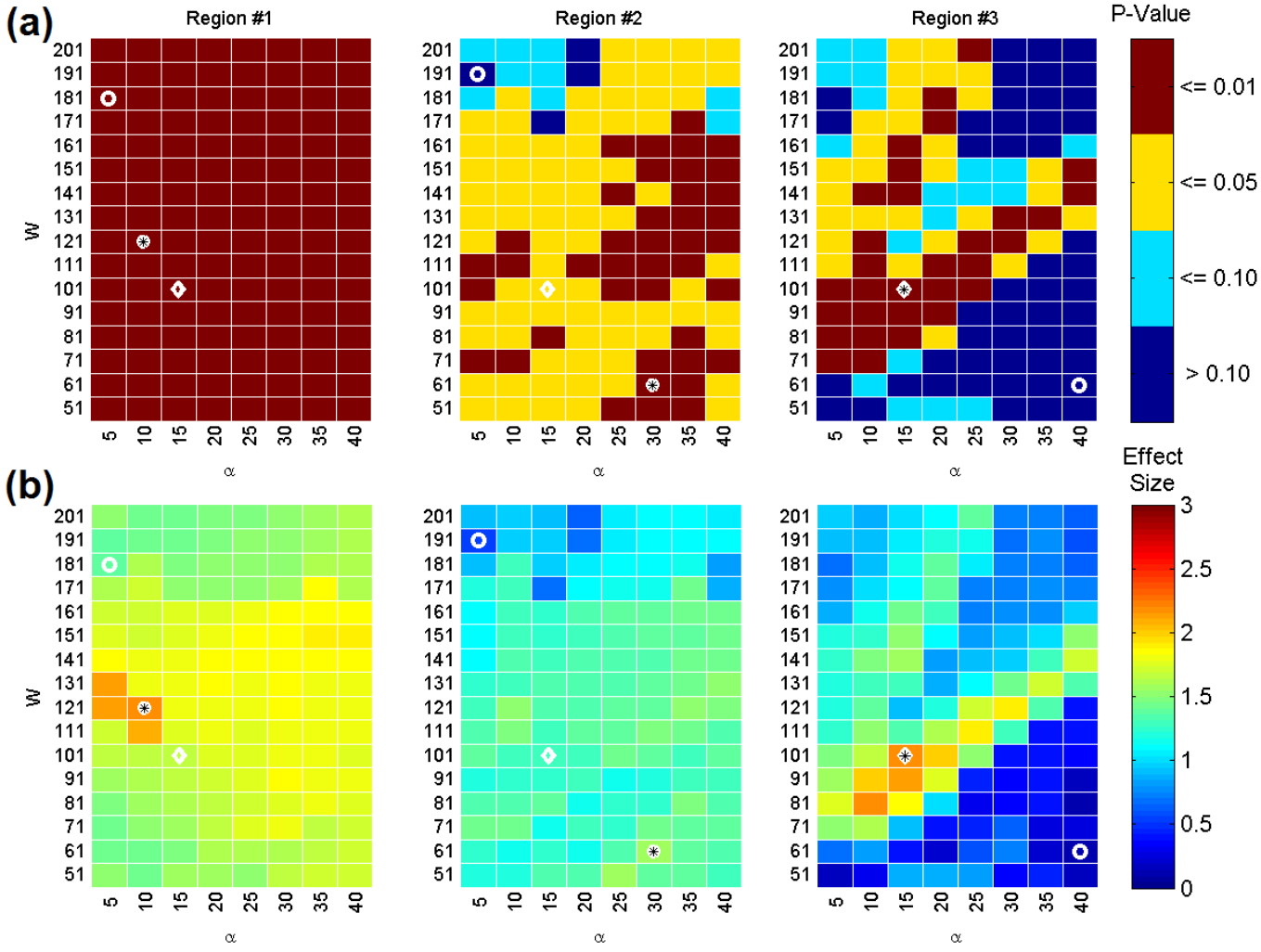

FIGURE 4. **Heatmap visualizations of wavelet-based biomarker statistics across baseline removal sensitivity trials for Regions 1, 2, and 3, corresponding to 2.02, 3.02, and 3.78 ppm respectively. The default parameter value trials are denoted with a diamond; the best and worst performing trials are denoted by an asterisk and circle respectively. (a) P-Value heatmap visualizations for wavelet biomarkers in Regions 1, 2, and 3. (b) Effect size heatmap visualizations for wavelet biomarkers in Regions 1, 2, and 3.**

exhibits a clear distinction between the two classes but for which there is no discernible resonance to which the difference can be attributed.

In addition to wavelet-based biomarkers, the ensemble mean of estimated model-based resonance structures obtained using the HSVD algorithm are also shown in Fig. 3, for Regions 1 (NAA) and 2 (Cr) respectively. No apparent resonance structure is observable in Region 3, and therefore it is omitted from our analysis of model-based biomarkers. The spectral locations and statistics of the model-based biomarkers derived from the resonance structures are listed in Table 3. From the set of model-based biomarkers obtained using the default processing parameters, only the resonance decay biomarker of creatine in Region 2 achieves statistical significance.

The biomarkers identified using the default processing parameters, both wavelet and model-based, serve as a reference to evaluate alternative parameter selection scenarios and their impact on the statistical significance of specific biomarkers. We investigated two such scenarios using the baseline removal and the water removal stages.

Baseline Removal Sensitivity

The heatmaps in Fig. 4 describe different strata of P-values (upper) and effect sizes (lower) observed for all values of (W, α) for the wavelet-based biomarkers identified with default baseline parameter values in Regions 1, 2, and 3. In Region 1, the wavelet-based biomarker remains statistically significant ($p < 0.01$) for all values of (W, α), and the maximum effect size (asterisk) occurs at $(W = 121, \alpha = 10)$, which is slightly higher than that of the default (diamond) parameters $(W = 101, \alpha = 15)$. The lowest effect size (circle) occurs, however, for $(W = 181, \alpha = 5)$. For Region 2 and 3, significance levels vary dramatically for different values of (W, α), with p-values going both below 0.01 and also exceeding 0.05. Likewise, maximum and minimum effect sizes at values of (W, α) different from the default values. In Region 3, the default parameters yield the greatest significance and the largest effect sizes, but small

TABLE 3. Model-based biomarker statistics obtained using the default processing parameters. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$.

Default Model-Based Model-Based Bioarkers and Statistics				
Region	**PPM Location**	**Feature Type**	**|Effect-Size|**	**P-Value**
1	2.02	Res. Amp	0.31	0.522
		Res. Area	0.5	0.301
		Res. Decay	0.89	0.076
2	3.02	Res. Amp	0.66	0.179
		Res. Area	0.48	0.323
		Res. Decay	1.01*	0.046

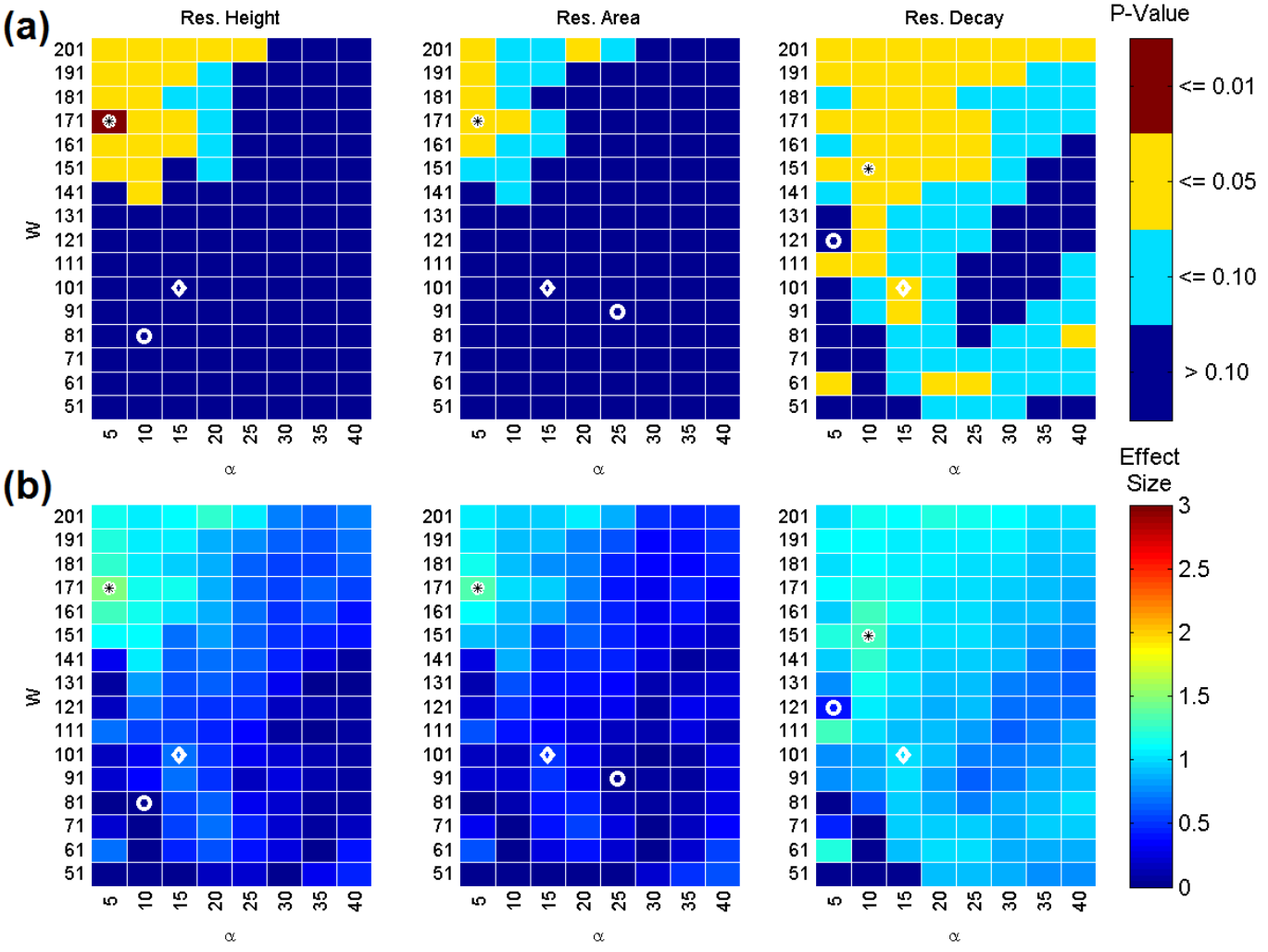

FIGURE 5. Heatmap visualizations of model-based biomarker statistics across baseline removal sensitivity trials for Region, 2, corresponding to the creatine peak at approximately 3.02 ppm. The default parameter value trials are denoted with a diamond; the best and worst performing trials are denoted by an asterisk and circle respectively. (a) P-Value heatmap visualizations for model-based biomarkers (resonance height, area, and decay) in Region 2. (b) Effect size heatmap visualizations for model-based biomarkers in Region 2.

deviations in parameter values can cause precipitous changes in both quantities. Looking across all three regions, the set of parameter values where the biomarkers are not significant ($p > 0.05$) do not coincide.

The presence of prominent, assigned peaks in Regions 1 and 2 made them candidates for an identical analysis using estimated model-based resonance parameters. Fig. 5 presents the heatmaps of model-based biomarker statistics in Region 2

TABLE 4. Wavelet-based biomarker statistics for baseline removal parameter sensitivity trials. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$.

Wavelet Biomarker Statistics - Baseline Removal Sensitivity Results										
		Default	Best Case Performance				Worst Case Performance			
Region	PPM	|Effect-Size|	W	α	|Effect-Size|	% Change	W	α	|Effect-Size|	% Change
1	2.02	1.68**	121	10	2.18**	30%	181	5	1.38	-37%
2	3.02	1.29*	61	30	1.56**	21%	191	5	0.56	-64%
3	3.78	2.20**	101	15	2.20**	0%	61	40	0.12	-95%

TABLE 5. Model-based biomarker statistics for baseline removal parameter sensitivity trials. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$.

Model-Based Biomarker Statistics - Baseline Removal Sensitivity Results											
			Default	Best Case Performance				Worst Case Performance			
Region	PPM	Type	|Effect-Size|	W	α	|Effect-Size|	% Change	W	A	|Effect-Size|	% Change
1	2.02	Res. Amp	0.31	111	30	0.88	186%	171	5	0.01	-96%
		Res. Area	0.50	111	30	1.03*	105%	201	5	0.02	-95%
		Res. Decay	0.89	201	5	1.26*	39%	101	5	0.36	-60%
2	3.02	Res. Amp	0.66	171	5	1.47**	123%	81	10	0.01	-99%
		Res. Area	0.48	171	5	1.33*	178%	91	25	0.01	-99%
		Res. Decay	1.01*	151	10	1.30*	28%	121	5	0.41	-60%

derived during baseline removal trials that exhibit meaningful results. Region 2 corresponds to the prominent creatine resonance, for which only the estimated resonance decay parameter was statistically significant using the default baseline removal parameters, as indicated by the diamond overlay at $(W = 101, \alpha = 0.15)$. Moreover, Fig. 5 demonstrates that a range of parameter values can make resonance height and area estimates statistically significant. The optimal parameter values, $(W = 171, \alpha = 0.05)$, indicate that wider estimation intervals with lower estimates of noise levels may be better for assessing the amplitude and area biomarkers corresponding to the creatine resonance in Region 2. The corresponding NAA resonance parameters estimated by HSVD for Region 1 only demonstrate statistically significant differences for the resonance decay biomarker, and therefore are not shown for the sake of brevity. The parameter values that reflect the least significance do not coincide across different biomarker types, nor are they consistent between biomarkers from Regions 1 and 2.

Tables 4 and 5 summarize the key heatmap parameter values from the baseline removal trials for both wavelet-based and model-based biomarkers, respectively, including the default, best-case and worst-performing parameter values (W, α) and effect sizes, and the percentage change from default parameter values. The amount of increase and decrease in effect size across the heatmap solution space was dramatically larger for model-based biomarkers. For both types of biomarkers, the effect size can be increased, but the optimal parameters differed for each region. In Region 3, the wavelet-based biomarker achieved the maximum effect size at the default values, which was purely coincidental, but the default values were not optimal for biomarkers in Regions 1 and 2.

For a subset of extreme values for (W, α), the desired resonances could not be consistently identified by HSVD, and therefore no biomarkers could be obtained. Consequently, these trials were omitted.

Water Removal Sensitivity

The heatmaps in Fig. 6 describe different strata of P-values (upper) and effect sizes (lower) observed for all values of (K, Q) for the wavelet-based biomarkers identified with default baseline parameter values in Regions 1, 2, & 3. In comparison to the heatmaps for baseline removal in Fig. 4, there is considerably less variation in P-values and effect sizes. Similar to Fig. 4, however, the optimal parameter values for water removal are different for Regions 1, 2, & 3. Fig. 7 conveys the corresponding heatmaps for the model-based biomarkers in Region 2

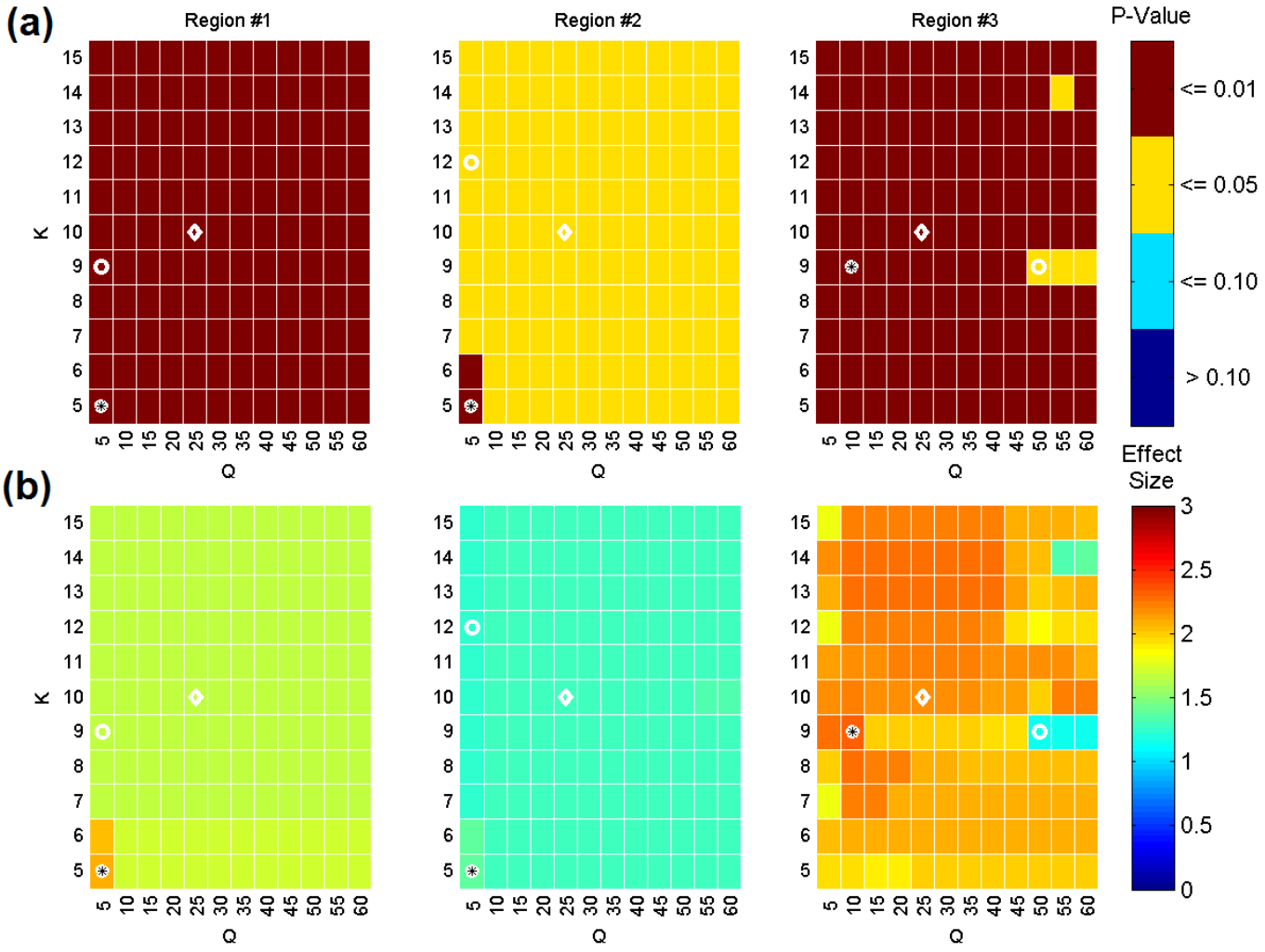

FIGURE 6. **Heatmap visualizations of wavelet-based biomarker statistics across water removal sensitivity trials for Regions 1, 2, and 3, corresponding to 2.02, 3.02, and 3.78 ppm respectively. The default parameter value trials are denoted with a diamond; the best and worst performing trials are denoted by an asterisk and circle respectively. (a) P-Value heatmap visualizations for wavelet biomarkers in Regions 1, 2, and 3. (b) Effect size heatmap visualizations for wavelet biomarkers in Regions 1, 2, and 3.**

TABLE 6. **Wavelet biomarker statistics for water removal parameter sensitivity trials. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$.**

Wavelet Biomarker Statistics - H20 Removal Sensitivity Results										
		Default	Best Case Performance				Worst Case Performance			
Region	PPM	|Effect-Size|	K	Q	|Effect-Size|	% Change	K	Q	|Effect-Size|	% Change
1	2.02	1.68**	5	5	2.07**	23%	9	5	1.65**	-20%
2	3.02	1.29*	5	5	1.39**	8%	12	5	1.24*	-11%
3	3.78	2.20**	9	10	2.31**	5%	9	50	1.15*	-50%

and, likewise, illustrates less variability than Fig. 5. Similar to baseline removal, model-based biomarkers for Region 1 remain statistically insignificant across water removal parameter values. However, as illustrated in Fig. 7, the response surface for the model-based creatine resonance decay biomarker in Region 2 demonstrated a limited range of parameter values that achieve statistical significance.

The key heatmap values from the water removal trials are summarized in Tables 6 and 7 for both wavelet-based and model-based biomarkers, respectively, including the default, best-case and worst-case (**K**, **Q**) values and effect sizes, and the percentage change from default values. Generally, the range of effect sizes is smaller than that documented in Tables 4 and 5 for baseline removal. For wavelet-based biomarkers, the best case parameters values were identical

TABLE 7. Model-based biomarker statistics for water removal parameter sensitivity trials. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$.

Model-Based Biomarker Statistics - H20 Removal Sensitivity Results											
			Default	Best Case Performance				Worst Case Performance			
Region	PPM	Type	|Effect-Size|	K	Q	|Effect-Size|	% Change	K	Q	|Effect-Size|	% Change
1	2.02	Res. Amp	0.31	6	10	0.44	42%	13	5	0.03	-91%
		Res. Area	0.50	6	10	0.61	23%	12	5	0.07	-86%
		Res. Decay	0.89	15	50	1.02*	16%	15	5	0.54	-39%
2	3.02	Res. Amp	0.66	9	50	0.81	23%	5	25	0.32	-51%
		Res. Area	0.48	15	5	0.67	39%	5	25	0.14	-71%
		Res. Decay	1.01*	15	50	1.11*	10%	5	5	0.65	-36%

TABLE 8. Wavelet-based biomarker statistics for composite sensitivity trials. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$.

Wavelet Biomarker Statistics - Composite Sensitivity Trial Results														
		Default	Best Case Performance						Worst Case Performance					
Region	PPM	|Effect-Size|	W	α	K	Q	|Effect-Size|	% Change	W	α	K	Q	|Effect-Size|	% Change
1	2.02	1.68**	151	35	5	5	2.28**	36%	181	5	15	40	1.35	-19%
2	3.02	1.29*	61	30	5	5	1.75**	35%	191	5	15	35	0.47	-64%
3	3.78	2.20**	81	10	13	40	2.31**	5%	81	40	15	40	0.00	-100%

TABLE 9. Model-based biomarker statistics for composite sensitivity trials. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$.

Model-Based Biomarker Statistics - Composite Sensitivity Trial Results															
			Default	Best Case Performance						Worst Case Performance					
Region	PPM	Type	|Effect-Size|	W	α	K	Q	|Effect-Size|	% Change	W	α	K	Q	|Effect-Size|	% Change
1	2.02	Res. Amp	0.31	151	40	13	55	1.09*	254%	161	10	5	10	0.00	-100%
		Res. Area	0.50	141	40	13	55	1.2*	140%	181	5	5	10	0.00	-100%
		Res. Decay	0.89	201	5	7	10	1.26*	42%	61	35	13	5	0.00	-100%
2	3.02	Res. Amp	0.66	171	5	10	50	1.49**	126%	161	30	14	20	0.00	-100%
		Res. Area	0.48	171	5	10	50	1.36**	184%	61	15	15	15	0.00	-100%
		Res. Decay	1.01*	51	20	6	5	1.64**	62%	81	5	12	35	0.00	-100%

for biomarkers in Region 1 and Region 2 ($\mathbf{K = 5}$, $\mathbf{Q = 5}$), but different for Region 3 ($\mathbf{K = 9}$, $\mathbf{Q = 10}$), potentially indicating a dependence upon the proximity to the dominant water resonance.

Composite Sensitivity

Tables 8 and 9 summarize the results for wavelet-based and model-based biomarkers, respectively, when parameters for both water removal and baseline removal stages were optimized concurrently. As expected, the effect sizes were greater than those reported in Tables 4–7 and the optimal parameter values differed from those derived by optimizing individual stages. Tables 8 and 9 also illustrate that biomarkers in different spectral regions achieve their largest effect sizes using different sets of parameter values.

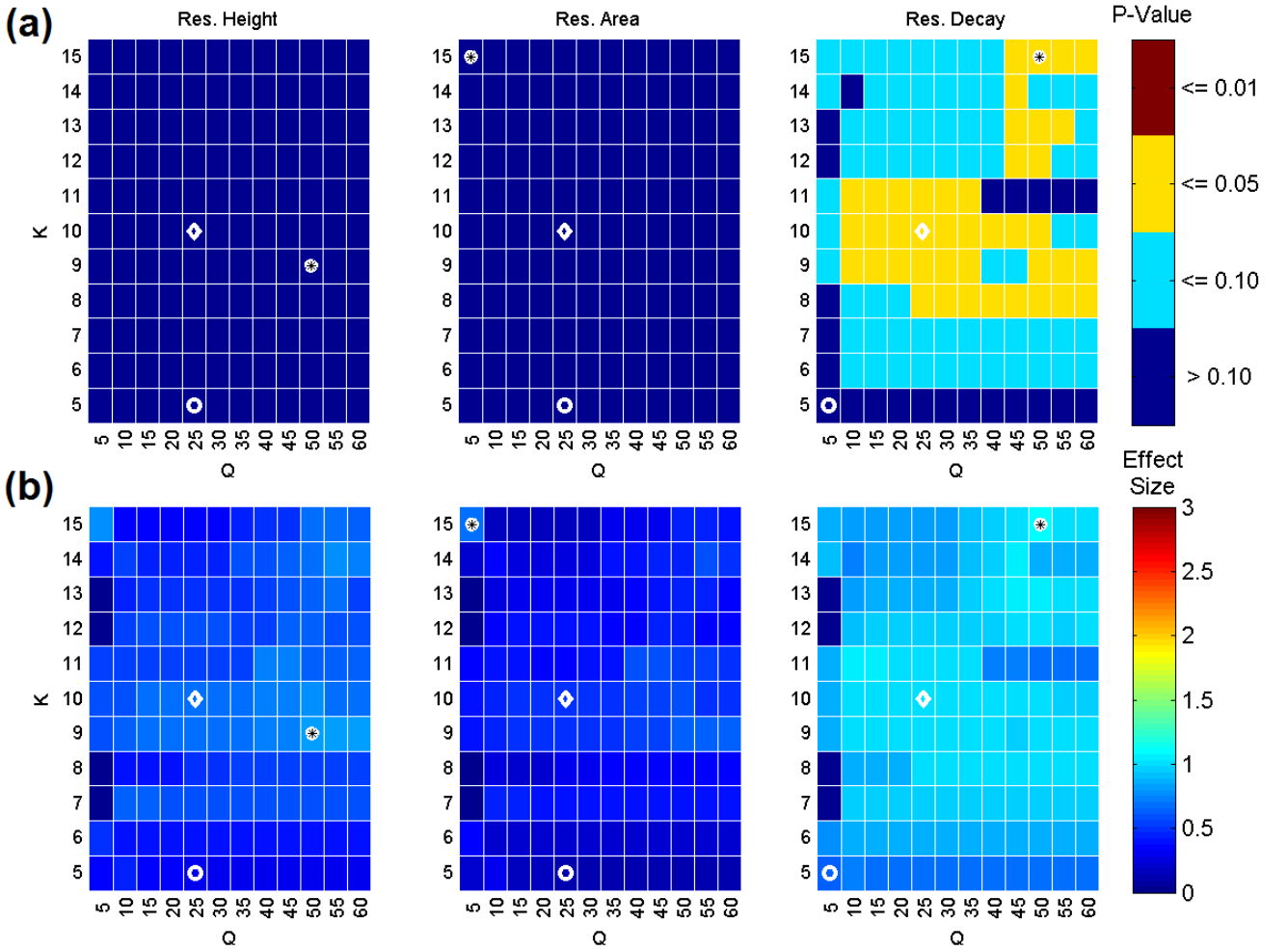

FIGURE 7. **Heatmap visualizations of model-based biomarker statistics across water removal sensitivity trials for Region, 2, corresponding to the creatine peak at approximately 3.02 ppm. The default parameter value trials are denoted with a diamond; the best and worst performing trials are denoted by an asterisk and circle respectively. (a) P-Value heatmap visualizations for model-based biomarkers (resonance height, area, and decay) in Region 2. (b) Effect size heatmap visualizations for model-based biomarkers in Region 2.**

To visually assess how different sets of parameter values impact the waveforms from which the features are extracted, Fig. 8 plots the mean waveforms from Region 3 for each class using different quartets of baseline removal and water removal parameters (W,α,K,Q) appearing in Table 8. The waveforms derived using the best-performing parameters (81, 10, 13, 40) and default parameters (101, 15, 10, 25) show little change in the proximity of the wavelet-based biomarker, but the class difference disappears when the worst-case parameters (81, 40, 15, 40) are used. It is worth noting that the best-case and worst-case parameter values are similar, with the major difference being the value of α, which is the percentage level at which the baseline value is estimated. In Region 3, $\alpha = 40$ erodes the difference in classes, but for Regions 1 & 2, best-case effect sizes were found using $\alpha = 35$ and $\alpha = 30$, respectively.

Classification

Table 10 reports the probability of correct classification (PCC) rates for the comparison between healthy control subjects and subjects with spinal cord injury using wavelet-based biomarkers. PCC values are given for the biomarkers derived using default baseline removal and water removal parameters as well as those achieved using the optimized parameters in Table 8. As in [1], three different classifiers were used. Three methods of estimating the PCC are provided. Apparent PCC represents the PCC achieved by testing classifiers on the data used to train them and provide an upper bound on PCC rates. Leave-One-Out (LOO) Cross-Validation PCC illustrates the average classification performance when the classifiers are tested on one out-of-sample datum. The 0.632+ LOO Bootstrap [44] provides a more rigorous estimate of PCC by performing a bootstrap operation on each trial of the LOO operation.

The results in Table 10 indicate equal or greater performance for the optimized biomarkers for several of the classifiers and cross-validation techniques.

IV. DISCUSSION

The main contribution of our effort is to introduce algorithmic techniques for measuring the sensitivity of MRS

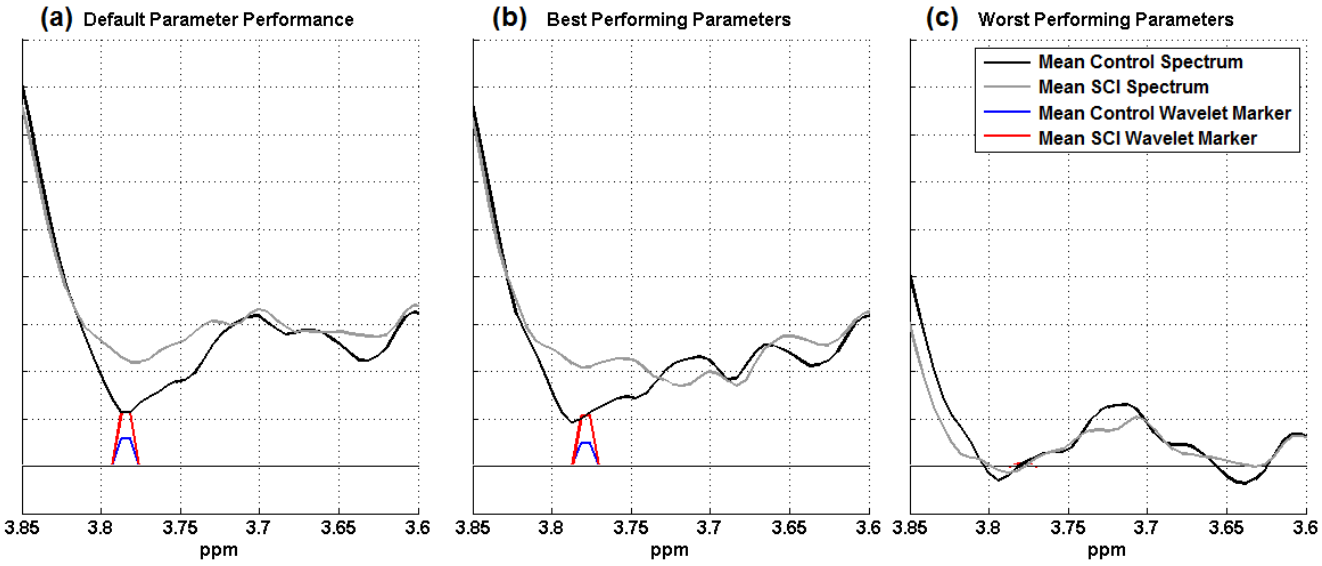

FIGURE 8. Visualization of the mean class spectra and corresponding wavelet-based biomarker intensity for Region 3 for different sets of water removal and baseline removal parameters. These results correspond to the (a) default processing parameters, (b) best-performing processing parameters, and (c) worst performing processing parameters. Specific parameter values and effect sizes are listed in Table 2 & Table 8.

TABLE 10. Classification rates for default MRS biomarkers and optimized wavelet-based biomarkers based on the parameters reported in Table 2 & Table 8.

Percent Correct Classification (PCC) Rates						
	Default Biomarker Values			Optimized Biomarker Values		
Classification Method	Linear	KNN(1)	SVM	Linear	KNN(1)	SVM
Apparent PCC	0.90	1.00	0.85	0.90	1.00	0.90
LOO Cross-Validation PCC	0.75	0.90	0.80	0.90	0.90	0.90
0.632+ LOO Bootstrap PCC	0.77	0.86	0.77	0.87	0.92	0.88

biomarkers that can accelerate the clinical translation of MRS biomarkers. While previous efforts have investigated sensitivity and reproducibility of measurements and individual algorithms operations, our effort is the first to focus specifically on the sensitivity of MRS biomarkers derived from a parameterized pipeline. More specifically, we have measured how varying parameters at distinct processing stages can either erode or strengthen the prospective biomarkers found in [1], and that it is possible to identify parameters that maximize the ability of biomarkers to discriminate between groups.

Our approach is especially useful in clinical studies with low N like [1] where important insights cannot be derived through simple significance testing. By interrogating the results in [1], Figs. 4 and 6 demonstrated that the biomarkers in [1] were more sensitive to baseline removal than water removal and also that the biomarkers in the three regions showed different levels of stability to baseline removal. As shown in Fig. 4, the biomarker in Region 1 did not vary in significance or effect size to baseline removal, but biomarkers in Region 2 and 3 did. Fig. 8 and Table 8 illustrate how an extended interval of structural difference containing a prospective biomarker can be completely eroded. Insights such as these could provide important hypotheses for future experiments about the stability of prospective biomarkers as well as the design of novel algorithms that maximize the strength of specific biomarkers.

Our effort to establish novel metrics for MRS algorithms can also enable equitable comparisons between algorithms and unambiguously identify gaps and requirements for new algorithms. This is especially true for emerging technologies, such as MRS, where there is no consensus on standardized methods for clinical acquisition, post-processing, and analysis.

As discussed in [1], the mechanisms of pain and spinal cord injury, along with the broader range of neurological disorders, remain poorly understood. The need to identify the underlying pathophysiology as well as targets for drug discovery requires a robust understanding of results from clinical experiments. The results documented in [1] both confirmed existing hypotheses but also generated new ones.

The current sensitivity analyses are especially clinically relevant given the importance of NAA and Cr in diagnosis. NAA is a metabolite that is produced within the axons and transported down the axons such that the relative concentration of this metabolite is proportional to the number of functioning brain cells within the region of interest measured by MRS [46]. In the context of SCI, changes in NAA are clinically significant as it demonstrates that injury downstream of the brain can have cerebral effects whether it be to chronic pain or deafferation [1]. The improved sensitivity provided by this study further supports this finding and more importantly demonstrates that post-processing methods can be optimized to better characterize this important metabolite. This has significant ramifications across a broad range of diseases that affect the CNS as neurodegenerative disorders (such as Alzheimer's dementia), psychiatric disorders, and traumatic brain injury have all shown changes in NAA.

Similarly, the improved sensitivity of region 2, which contains the resonances of Cr as also clinically important as Cr is not only reflective of brain energetics [47] but it is also used for spectral normalization based on the assumption that Cr levels are relatively stable. This assumption has been challenged across a number of diseases, most notably in brain injury where Cr has recently been shown to be decreased [48].

Region 3 contains the important neurotransmitter related metabolites of glutamate and glutamine (collectively described as Glx). Given that pain is frequently observed with glutaminergic processes, the ability to improve the characterization of this spectral region is important. Particularly as several metabolites co-resonate within this region, increased spectral sensitivity would improve the diagnostic specificity of the spectral measures.

Beyond the specific improvements cited in this particular study, the overall paradigm of the optimization of post-processing methods for improved biomarker sensitivity has implications across many other diseases. For example, the ability to select specific algorithm parameters that highlight activity in associated intervals of the MRS spectrum can be advantageous. Future studies can benefit from directly including the selection of optimal algorithm parameters as part of the experimental plan. Furthermore, the improvement in sensitivity will allow for studies in diseases with more subtle changes thereby extending the use of MRS into diseases that conventional post-processing methods would not have provided diagnostic results.

Our results are limited by the fact that the data we analyzed does not control all variables that might contribute to a tightly controlled and calibrated demonstration of system sensitivity. However, appropriate measures were taken to ensure uniformity in the data collection process across the entire cohort of subjects, as described in [1]. The data set was analyzed retrospectively and is representative of appropriately matched cohorts collected for the discovery of biomarkers where no a priori information about the disease state informs which chemicals or spectral intervals are more relevant than others. Future efforts might augment the current approach by employing cohorts where the expected differences are well understood and parameter trades can ascertain how well the differences can be revealed.

ACKNOWLEDGMENT

The authors would like to acknowledge the contributions of Laura Mariano at Draper Laboratory and Dr. Saad Ramadan.

REFERENCES

[1] P. Stanwell *et al.*, "Neuro magnetic resonance spectroscopy using wavelet decomposition and statistical testing identifies biochemical changes in people with spinal cord injury and pain," *Neuroimage*, vol. 53, no. 2, pp. 544–552, 2010.

[2] J. P. Kerekes and J. E. Baum, "Hyperspectral imaging system modeling," *MIT Lincoln Lab. J.*, vol. 14, no. 1, pp. 117–130, 2003.

[3] J. C. Leachtenauer, W. Malila, J. Irvine, L. Colburn, and N. Salvaggio, "General image-quality equation: GIQE," *Appl. Opt.*, vol. 36, no. 32, pp. 8322–8328, 1997.

[4] C. E. Mountford, P. Stanwell, A. P. Lin, S. Ramadan, and B. D. Ross, "Neurospectroscopy: The past, present, and future," *Chem. Rev.*, vol. 110, no. 5, pp. 3060–3086, 2010.

[5] A. Lin, B. D. Ross, K. Harris, and W. Wong, "Efficacy of proton magnetic resonance spectroscopy in neurological diagnosis and neurotherapeutic decision making," *NeuroRx*, vol. 2, no. 2, pp. 197–214, 2005.

[6] A. Lin, T. Tran, S. Bluml, S. Merugumala, H. J. Liao, and B. D. Ross, "Guidelines for acquiring and reporting clinical neurospectroscopy," *Seminal Neurol.*, vol. 32, no. 4, pp. 432–453, 2012.

[7] A. P. Lin and B. D. Ross, "Short-echo time proton MR spectroscopy in the presence of gadolinium," *J. Comput. Assist. Tomography*, vol. 25, no. 5, pp. 705–712, 2001.

[8] S. Brugger, J. M. David, S. Leucht, and J. M. Stone, "Proton magnetic resonance spectroscopy and illness stage in schizophrenia—A systematic review and meta-analysis," *Biol. Psychiatry*, vol. 69, no. 5, pp. 495–503, 2011.

[9] Y. Aoki, K. Kasai, and H. Yamasue, "Age-related change in brain metabolite abnormalities in autism: A meta-analysis of proton magnetic resonance spectroscopy studies," *Translational Psychiatry*, Jan. 2012.

[10] A. Gardner, G. L. Iverson, and P. Stanwell, "A systematic review of proton magnetic resonance spectroscopy findings in sport-related concussion," *J. Neurotrauma*, vol. 31, no. 1, pp. 1–18, Jan. 2014.

[11] P. J. Siddall and J. D. Loeser, "Pain following spinal cord injury," *Spinal Cord*, vol. 39, no. 2, pp. 63–73, Feb. 2001.

[12] R. E. Harris and D. J. Clauw, "Imaging central neurochemical alterations in chronic pain with proton magnetic resonance spectroscopy," *Neurosci. Lett.*, vol. 520, no. 2, pp. 192–196, 2012.

[13] F. Jiru, "Introduction to post-processing techniques," *Eur. J. Radiol.*, vol. 67, no. 2, pp. 202–217, 2008.

[14] E. J. Delikatny, W. E. Hull, and C. E. Mountford, "The effect of altering time-domains and window functions in two-dimensional proton COSY spectra of biological specimens," *J. Magn. Reson.*, vol. 94, no. 3, pp. 563–573, 1991.

[15] A. van den Boogaart, M. Ala-Korpela, F. A. Howe, L. M. Rodrigues, M. Stubbs, and J. R. Griffiths, "Magnetic resonance spectroscopy data analysis: Time or frequency domain?" *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 2, no. 3, pp. 479–482, 1994.

[16] C. Elster, A. Link, F. Schubert, F. Seifert, M. Walzel, and H. Rinneberg, "Quantitative MRS: Comparison of time domain and time domain frequency domain methods using a novel test procedure," *Magn. Reson. Imag.*, vol. 18, no. 5, pp. 597–606, 2000.

[17] A. Naressi, C. Couturier, I. Castang, R. de Beer, and D. Graveron-Demilly, "Java-based graphical user interface for MRUI, a software package for quantitation of in vivo/medical magnetic resonance spectroscopy signals," *Comput. Biol. Med.*, vol. 31, no. 4, pp. 269–286, 2001.

[18] D. Stefan *et al.*, "Quantitation of magnetic resonance spectroscopy signals: The jMRUI software package," *Meas. Sci. Technol.*, vol. 20, no. 10, p. 104035, 2009.

[19] L. Vanhamme, A. van den Boogaart, and S. Van Huffel, "Improved method for accurate and efficient quantification of MRS data with use of prior knowledge," *J. Magn. Reson.*, vol. 129, no. 1, pp. 35–43, 1997.

[20] W. W. F. Pijnappel, A. Van Den Boogart, R. De Beer, and D. Van Ormondt, "SVD-based quantification of magnetic resonance signals," *J. Magn. Reson.*, vol. 97, no. 1, pp. 122–134, 1992.

[21] S. W. Provencher, "Estimation of metabolite concentrations from localized in vivo Proton NMR spectra," *Magn. Reson. Med.*, vol. 30, no. 6, pp. 672–679, 1993.

[22] M. Kanowski, J. Braun, J. Bernarding, and C. Tempelmann, "Quantitation of simulated short echo time 1H human brain spectra by LCModel and AMARES," *Magn. Reson. Med.*, vol. 51, no. 5, pp. 904–912, 2004.

[23] P. Mandal, "In vivo proton magnetic resonance spectroscopic signal processing for the absolute quantitation of brain metabolites," *Eur. J. Radiol.*, vol. 81, no. 4, pp. 653–664, 2012.

[24] R. Bartha, D. J. Drost, and P. C. Williamson, "Factors affecting the quantification of short echo in-vivo 1H MR spectra: Prior knowledge, peak elimination, and filtering," *NMR Biomed.*, vol. 12, no. 4, pp. 205–216, 1999.

[25] R. Bartha, D. J. Drost, R. S. Menon, and P. C. Williamson, "Comparison of the quantification precision of human short echo time ^{1}H spectroscopy at 1.5 and 4.0 Tesla," *Magn. Reson. Med.*, vol. 44, no. 2, pp. 185–192, 2000.

[26] R. Kreis, "Issues of spectral quality in clinical ^{1}H-magnetic resonance spectroscopy and a gallery of artifacts," *NMR Biomed.*, vol. 17, no. 6, pp. 361–381, 2004.

[27] R. L. Somorjai *et al.*, "Classification of 1H MR spectra of human brain neoplasms: The influence of preprocessing and computerized consensus diagnosis on classification accuracy," *J. Magn. Reson. Imag.*, vol. 6, no. 3, pp. 437–444, 1996.

[28] A. E. Nikulin, B. Dolenko, T. Bezabeh, and R. L. Somorjai, "Near-optimal region selection for feature space reduction: Novel preprocessing methods for classifying MR spectra," *NMR Biomed.*, vol. 11, nos. 4–5, pp. 209–216, 1998.

[29] C. Elster, F. Schubert, A. Link, M. Walzel, F. Seifert, and H. Rinneberg, "Quantitative magnetic resonance spectroscopy: Semi-parametric modeling and determination of uncertainties," *Magn. Reson. Med.*, vol. 53, no. 6, pp. 1288–1296, 2005.

[30] H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. Van Ormondt, and D. Graverson-Demilly, "Time-domain semi-parametric estimation based on a metabolite basis set," *NMR Biomed.*, vol. 18, no. 1, pp. 1–13, 2005.

[31] G. Sorensen, "Magnetic resonance as a cancer imaging biomarker," *J. Clin. Oncol.*, vol. 24, no. 20, pp. 3274–3281, 2006.

[32] D. Cocuzzo and N. Keshava, "Sensitivity of biomarkers to post-acquisitional processing parameters for in vivo brain MR spectroscopy signals," in *Proc. IEEE Int. Symp. Biomed. Imag. Nano Macro*, Apr. 2011, pp. 1670–1675.

[33] R. Ogg, P. B. Kingsley, and J. S. Taylor, "WET, a T1- and B1-insensitive water-suppression method for in vivo localized ^{1}H NMR spectroscopy," *J. Magn. Reson.*, vol. 104, no. 1, pp. 1–10, 1994.

[34] R. de Beer, D. van Ormondt, and W. Pijnappel, "Quanitification of 1-D and 2-D magnetic resonance time domain signals," *Pure Appl. Chem.*, vol. 64, no. 6, pp. 815–823, 1992.

[35] A. P. Lin, H. J. Liao, S. K. Merugumala, S. P. Prabhu, W. P. Meehan III, and B. D. Ross, "Metabolic imaging of mild traumatic brain injury," *Brain Imag. Behavior*, vol. 6, no. 2, pp. 208–223, 2012.

[36] W. R. Martin, "MR spectroscopy in neurodegenerative disease," *Molecular Imag. Biol.*, vol. 9, no. 4, pp. 196–203, 2007.

[37] W. K. Oakden and M. D. Noseworthy, "Propylene glycol is essential in the LCModel basis set for pediatric 1H-MRS," *J. Comput. Assist. Tomography*, vol. 29, no. 1, pp. 136–139, 2005.

[38] T. Bezabeh *et al.*, "Statistical classification strategy for proton magnetic resonance spectra of soft tissue sarcoma: An exploratory study with potential clinical utility," *Sarcoma*, vol. 6, no. 3, pp. 97–103, 2002.

[39] S. Y. Kung, K. S. Arun, and B. V. D. Rao, "State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem," *J. Opt. Soc. Amer.*, vol. 73, no. 12, pp. 1799–1811, 1983.

[40] H. Barkhuijsen, R. De Beer, and D. Van Ormondt, "Improved algorithm for noniterative time domain model fitting to exponentially damped magnetic resonance signals," *J. Magn. Reson.*, vol. 73, no. 3, pp. 553–557, 1987.

[41] A. Willse *et al.*, "Identification of major histocompatibility complex-regulated body odorants by statistical analysis of a comparative gas chromatography/mass spectrometry experiment," *Anal. Chem.*, vol. 77, no. 8, pp. 2348–2361, 2005.

[42] I. Mader *et al.*, "Proton MR spectroscopy with metabolite-nulling reveals elevated macromolecules in acute multiple sclerosis," *Brain*, vol. 124, pp. 953–961, May 2001.

[43] J. Fowler, "The redundant discrete wavelet transform and additive noise," *IEEE Signal Process. Lett.*, vol. 12, no. 9, pp. 629–632, Sep. 2005.

[44] B. Efron and R. Tibshirani, "Improvements on cross-validation: The .632+ bootstrap method," *J. Amer. Statist. Assoc.*, vol. 92, no. 438, pp. 548–560, 1997.

[45] B. Efron, "Large-scale simultaneous hypothesis testing: The choice of a null hypothesis," *J. Amer. Statist. Assoc.*, vol. 99, no. 465, pp. 96–104, 2004.

[46] J. R. Moffett, B. Ross, P. Arun, C. N. Madhavarao, and A. M. Namboodiri, "N-acetylaspartate in the CNS: From neurodiagnostics to neurobiology," *Progr. Neurobiol.*, vol. 81, no. 2, pp. 89–131, Feb. 2007.

[47] G. J. Kemp, "Non-invasive methods for studying brain energy metabolism: What they show and what it means," *Develop. Neurosci.*, vol. 22, nos. 5–6, pp. 418–428, 2000.

[48] R. Vagnossi *et al.*, "Decrease in N-acetylaspartate following concussion may be coupled to decrease in creatine," *J. Head Trauma Rehabil.*, vol. 28, no. 4, pp. 284–292, 2013.

DANIEL COCUZZO is a graduate student with the Department of Computer Science at Stanford University, focusing on large scale distributed systems, security, and data mining. Previously, he was an engineer at Draper Laboratory in Cambridge, MA. Dan earned his B.S. in electrical engineering at Northeastern University in Boston, MA.

ALEXANDER LIN (M'10) received his Ph.D. degree in biochemistry and molecular biophysics from the California Institute of Technology, Pasadena, CA, USA, in 2009.

He is currently an Instructor with the Department of Radiology, Harvard Medical School, and a Research Associate with the Brigham and Women's Hospital, Boston, MA, USA. He is also a Visiting Research Associate with the Center of MR Research, University of Illinois at Chicago. Prior to this position, he was a Senior Research Associate and the Director of clinical services with the Huntington Medical Research Institutes and a Research Fellow with the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA.

Dr. Lin's primary research interests include the clinical applications of magnetic resonance spectroscopy. He was awarded the Young Investigator's Award in 2003 for his work in 13C spectroscopy in Alzheimer's disease and the NHLBI Fellows Research Award for his work in strain mapping of the carotid arteries in 2007. More recently, he was awarded the Partners in Excellence Award for his mentorship of students. He has authored over 36 papers, eight book chapters, and numerous conference abstracts. He has been a Reviewer for the IEEE International Symposium on Biomedical Imaging since 2010 and Grant Reviewer for the Department of Defense, National Institutes of Health. He is a member of numerous professional organizations, including the International Society of Magnetic Resonance in Medicine and Radiological Society of North America, and a Reviewer for multiple journals, including *Radiology, Neuroimage, Neurosurgery, NMR in Biomedicine*, and *Journal of Magnetic Resonance*.

IEEE Journal of Translational Engineering in Health and Medicine

PETER STANWELL is currently an Associate Professor with the Faculty of Health and Medicine, the University of Newcastle, Australia. His research interests center on developing medical imaging techniques to address clinically relevant needs, including ischemic stroke imaging, traumatic brain injury, sport-related concussion, and radiation oncology.

CAROLYN MOUNTFORD received the M.Sc. degree in biophysics and the Ph.D. degree in biochemisty from Somerville College Oxford. She is currently a Professor of radiology with the University of Newcastle, Australia, and a Full Professor of radiology with the Harvard Medical School, USA. She is the Director with the Centre for Magnetic Resonance in Health, Newcastle, and the Director with the Centre for Clinical Spectroscopy, Brigham and Women's Hospital, Boston. She is a fellow of the International Society of Magnetic Resonance in Medicine.

NIRMAL KESHAVA (M'00) received the B.S. degree in electrical engineering from UCLA, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, in 1989 and 1997, respectively.

He was with Bellcore from 1989 to 1992, the MIT Lincoln Laboratory from 1997 to 2004, and the Charles Stark Draper Laboratory from 2004 to 2013. He is currently a Senior Principal Informatics Scientist with the Research and Development Information Organization, AstraZeneca Pharmaceuticals, Waltham, MA, USA. His responsibilities include developing analytics for big data challenges across all areas of drug development.

Dr. Keshava has been the Principal Investigator on research programs focusing on neuroimaging, psychology, and neurophysiology. He has been a Reviewer for the IEEE Transactions on Signal Processing and the IEEE Transactions on Geoscience and Remote Sensing. In 2011, he was an invitee of the National Academy of Engineering Frontiers of Engineering Symposium and is an invitee for the 2014 National Academy of Engineering Indo-American Frontiers of Engineering Symposium. He is also organizing a special session on bioinformatics for the 2014 IEEE High Performance Extreme Computing Conference.