Chemical Shift–Based Water/Fat Separation: A Comparison of Signal Models

Diego Hernando,[1,2]* Zhi-Pei Liang,[1,2] and Peter Kellman[3]

Quantitative water/fat separation in MRI requires careful modeling of the acquired signal. Multiple signal models have been proposed in recent years, but their relative performance has not yet been established. This article presents a comparative study of 12 signal models for quantitative water/fat separation. These models were selected according to three main criteria: magnitude or complex fitting, use of single-peak or multipeak fat spectrum, and modeling of T_2^* decay. The models were compared based on an analysis of the bias and standard deviation of their resulting estimates. Results from theoretical analysis, simulation, phantom experiments, and in vivo data were in good agreement. These results show that (a) complex fitting is uniformly superior to magnitude fitting, (b) multipeak fat modeling is able to remove the bias present in single-peak fat modeling, and (c) a single-T_2^* model performs best over a range of clinically relevant signal-to-noise ratios (SNRs) and water/fat ratios. Magn Reson Med 64:811–822, 2010. © 2010 Wiley-Liss, Inc.

Key words: fat-water imaging; fat quantification; Dixon imaging; multi-peak fat; T_2^* measurement

The ability to quantitatively measure fat content in tissues has multiple important applications in MRI, including studies of bone marrow (1), breast (2), muscle (3), brain (4), liver (5,6), and heart (7–9). In recent years, chemical shift–encoded water/fat separation methods have become increasingly popular for quantitative fat measurement. This popularity is largely due to the ability of chemical shift–encoded methods to overcome the limitations of alternative techniques: lack of spatial information in single-voxel spectroscopy, sensitivity to amplitude of static field (B_0) and amplitude of radiofrequency field inhomogeneities in conventional fat saturation, or loss of SNR and inherent T_1-weighting in short-tau inversion recovery (10–12).

There are four key issues with chemical shift–encoded water/fat separation. First, the presence of large B_0 magnetic field inhomogeneities can result in large errors in water/fat separation if the B_0 effects are not adequately addressed (13,14). Second, the commonly used spoiled gradient echo sequences may result in considerable residual T_1weighting, typically leading to bias (overestimation) in the estimated fat component, which has a shorter T_1 than the water component (15). Third, noise also results in bias in the estimation of the minority component of the signal (whether it is water or fat), particularly in cases where the minority component is very small compared to the majority component, i.e., fat fractions (FFs) close to 0% or 100% (15). Fourth, inaccurate modeling of the acquired chemical shift–encoded signal also results in considerable bias in fat quantification (6,16,17).

Complications due to B_0 field inhomogeneities, T_1 bias, and noise have been thoroughly addressed in the literature. Field inhomogeneities can be corrected by region-growing or regularized estimation methods (14,18–22). T_1 bias in spoiled gradient echo acquisitions can be avoided by using a small flip angle or corrected by using a dual-flip-angle acquisition (15). Noise bias also can be corrected effectively by using magnitude discrimination or phase-constrained reconstruction (15), or by using a look-up table bias correction over a region of interest (23). However, signal modeling for quantitative water/fat separation remains largely an unresolved issue. Specifically, there are three key decisions to make when modeling the acquired signal: use of magnitude or complex fitting, use of single-peak or multipeak fat modeling, and modeling of the signal T_2^* decay. These alternatives can be summarized as follows:

- *Magnitude vs. complex fitting.* Fitting the magnitude of the signal has been proposed as a means of simplifying the estimation since it removes the effects of field inhomogeneity (16,24,25). However, magnitude fitting has several well-known drawbacks, such as the nongaussian distribution of the noise in magnitude MR images, and an inability to correctly detect FFs above 50%.
- *Single peak vs. multipeak fat models.* The basic single-peak signal model ignores the presence of multiple spectral peaks in the fat signal, which leads to bias in quantification. This bias can be overcome by using a more sophisticated, multipeak fat model, where the relative amplitudes of the different fat peaks can either be precalibrated or autocalibrated (6,17,26).
- *Modeling of T_2^* decay.* In general, the amplitudes of the water and fat components of the signal will decrease with echo time (TE) due to T_2^* decay. It has been shown that ignoring this decay may result in considerable bias, and a number of groups have developed methods for including T_2^* in the model. In general, water and fat will have different T_2^* decays (and even the different fat peaks will have different decays, although this is typically ignored as it would result in significant complication in the estimation (12)), so these should be estimated separately, adding two nonlinear parameters to the estimation (26–28). As a simplification of

[1]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[2]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[3]Laboratory of Cardiac Energetics, National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA

Grant sponsor: NIH; Grant number: P41-RR023953-01, P41-EB001977-21; Grant sponsor: NSF; Grant number: CBET-07-30623.

*Correspondence to: Diego Hernando, Beckman Institute for Advanced Science and Technology, 405 N. Mathews Ave, Urbana, IL 61801. E-mail: dhernan2@illinois.edu

Received Month 2010; accepted Month 2010.
DOI 10.1002/mrm.22455
Published online 30 June 2010 in Wiley Online Library (wileyonlinelibrary.com).

this general model, a single T_2^* has been proposed for both water and fat (29,30). Intermediate models have also been proposed, where the decay rates of water and fat are different, but the difference is assumed known (16).

In this article, we present a comparative analysis of multiple models based on the alternatives described above. The analysis focuses on two key properties of the estimates for each model: bias and standard deviation. These properties capture the behavior of different models regarding model mismatch (bias) and noise sensitivity (standard deviation). The analysis is based on theoretical properties of the different models, simulations, and phantom data. Additionally, the conclusions derived from this analysis are verified qualitatively with an in vivo dataset.

The results presented in this article can be viewed as an extension of previous studies that focused on a subset of the models considered here. For instance, in Yu et al. (29), the authors proposed to include a single T_2^* decay in the signal model and showed its advantage over a no-decay model. In Bydder et al. (16), the authors performed a comparison of different magnitude-based fitting models, with an emphasis on modeling the T_2^* decay. In Reeder et al. (6) and Yu et al. (17), the authors demonstrated the advantages of multipeak fat modeling relative to single-peak fat modeling. In Chebrolu et al. (28), the authors compared the use of a single T_2^* or different T_2^*s for water and fat in the context of complex fitting. Relative to these works, this article includes a more comprehensive set of models, as well as a thorough analysis ranging from the theoretical characterization of different models to phantom and in vivo results.

MATERIALS AND METHODS

Signal Models

A total of 12 signal models were evaluated by considering the main alternatives proposed in the literature, in terms of the following choices:

- *Magnitude and complex fitting (two alternatives).* All the models were implemented in two versions: fitting the magnitude of the acquired data and fitting the complex-valued data.
- *Single-peak and multipeak fat models (two alternatives).* In addition to the model based on a single-peak fat spectrum, a six-peak fat model was calibrated and implemented. Even though a larger number of fat peaks can be found by spectroscopy, six fat peaks is the most that has been used in the water/fat separation literature. If more peaks are used, some of the peaks will have very similar frequencies, resulting in considerably more unstable calibration.
- *Modeling of $R_2^* = 1/T_2^*$ decay (three alternatives).* No-decay, one-decay (with a single R_2^*, common for water and fat), and two-decay (with independent $R_{2,W}^*$ and $R_{2,F}^*$ for water and fat) models were implemented. Intermediate models, e.g., where water and fat have different decay rates but the difference $R_{2,F}^* - R_{2,W}^*$ is assumed known, were not considered in this work due in part to the difficulty of estimating this difference and also to the dependence of the apparent difference $R_{2,F}^* - R_{2,W}^*$ on whether a single-peak or multipeak fat model is used (16). Similarly, even more sophisticated fat models, where different peaks within the fat spectrum have different decay rates, were also not considered due to their significant increase in complexity and noise sensitivity.

All model fitting was done voxel by voxel in MatLab (The MathWorks, Natwick, MA) using a standard gradient-based least-squares fitting procedure (lsqnonlin). Magnitude fitting was also least squares (even though the noisy magnitude data do not follow a gaussian distribution) because this is a good approximation (except at very low SNRs), and this is the approach taken in practice in most previous works (16).

Theoretical Analysis

The well-known Cramér-Rao lower bound (CRLB) provides a bound on the variance of any unbiased estimator. In the context of water/fat separation, it has been used to effectively characterize different models and acquisition strategies (29,31). In this work, we computed the CRLB for each of the 12 models under consideration for a range of water/fat ratios, assuming a fixed set of eight TEs: $1.43 + (n-1)2.23$ ms, for $n = 1, \ldots, 8$ (the same TEs used in the simulations and phantom experiments). Computation of the CRLB for the complex fitting models was done as described in Scharf et al. (32), and for the magnitude-fitting models as described in Karlsen et al. (33).

Simulations

Even though the CRLB provides an elegant characterization (bound) of the variance of any unbiased estimator, it does not capture the effects of bias (e.g., due to model mismatch) or the practical performance of a given estimator. To address these issues, chemical shift–encoded data were simulated, using the following model, which employs representative values measured in the phantom experiments (as will be described in the "Multipeak calibration" subsection):

- The fat signal has six peaks, at $-244.3, -221.7, -175.4, -119.3, -32.1$, and 34.0 Hz, with relative amplitudes $0.01 \cdot [9.45e^{-i\pi 0.181}, 64.66, 9.67e^{i\pi 0.046}, 2.26e^{-i\pi 0.567}, 2.22e^{-i\pi 0.244}, 8.83e^{-i\pi 0.089}]$. These were the same relative amplitudes obtained by precalibration on the phantom.
- Water and fat have different R_2^* values: $R_{2,W}^* = 42$ sec^{-1} and $R_{2,F}^* = 54$ sec^{-1}.
- Two sets of 21 true water/fat ratios were chosen: covering the range of water/fat ratios $[0.01, 100]$ (equally spaced on a logarithmic scale) and covering the range of FFs [0%, 100%] (equally spaced on a linear scale).

The simulations used the same TEs as the CRLB analysis. Noise was added to the simulated data, resulting in two different regimens: moderate SNR (SNR = 30) and high SNR (SNR = 100). A total of 1024 noisy instances were generated for each water/fat ratio and each SNR. These data

were then fitted with the 12 models described above, using all combinations of the following choices: magnitude or complex fitting (two options), single-peak or multipeak fat modeling (two options), and no-decay, one-decay, or two-decay models (three options).

Experimental Studies

Phantom Construction

A water/fat phantom was constructed based on the methods described elsewhere (34–36). The phantom consisted of 11 vials containing water/fat mixtures at FFs ranging from 0% to 100% in increments of 10%. In order to obtain stable emulsions with biologically relevant T_1 and T_2 parameters, appropriate volumes of vegetable oil were mixed in vials with a solution of saline, agarose (2% mass/volume concentration), $CuSO_4$ (0.5 mM), and sodium dodecyl sulfate (43 mM). The $CuSO_4$, agarose, and sodium dodecyl sulfate were dissolved in saline. The resulting saline solution was heated until it boiled for 30 seconds. The vegetable oil and saline solution were placed in a waterbath at 50°C and then carefully poured in the appropriate proportions into the 11 vials. After filling each vial, it was immediatly mixed by gentle inversion to obtain a homogeneous emulsion (35) and placed in ice for the gel to form. The vials up to 60% FF formed homogeneous gels. The vial at 70% FF did not gel, but the mixture remained homogeneous. The vials at 80% and 90% FFs contained severe inhomogeneities in the mixture and were not used for the quantitative study.

Data Acquisition

Data for quantitative evaluation were acquired on a Siemens Avanto (Siemens AG Medical Solutions, Erlangen, Germany) 1.5-T scanner, using a phased-array coil. Phantom experiments were performed with a spoiled gradient echo sequence, using an echo-train with monopolar readout: field of view = 36.0 cm × 14.3 cm; bandwidth = 977 Hz/pixel; matrix size = 256 × 102, eight and 32 TEs with spacing 2.23 ms and initial TE 1.43 ms. Separate acquisitions were used in order to obtain moderate SNR (SNR ≈ 30) and high SNR (SNR ≈ 90). The moderate SNR acquisition was performed with flip angle 8° and pulse repetition time 500 ms, and the high SNR acquisition was performed with flip angle 25° and pulse repetition time 2000 ms. These long pulse repetition time values were chosen in order to avoid T_1 bias (15), which results in a bias under 1% for the T_1 values of water and oil measured in the phantom (953 ms and 207 ms, respectively) (15). This choice of acquisition parameters was made purposely to isolate the desired component (signal modeling) from the other complicating factors involved in fat quantification.

The phantom data were acquired using a Monte Carlo strategy. In order to perform a quantitative evaluation of bias and standard deviation for the different fitting models, each eight-echo acquisition was performed 128 times. This allows us to derive statistics from the estimated parameters on a voxel-by-voxel basis. Therefore, the phantom experiments correspond closely to the analytical and simulation-based results.

Additionally, data for measuring T_1 and T_2 were acquired on the phantom from the same slice. For the T_1 measurements, an inversion-recovery sequence was used with an echo train with monopolar readout. This allowed water/fat separation at each inversion time (which ranged from 100 ms to 1000 ms). For the T_2 measurements, a spin-echo sequence was used, with TEs ranging from 11 to 200 ms.

In vivo imaging was performed on a healthy normal volunteer under a research protocol approved by our institutional review board, with written informed consent. Data were obtained using an electrocardiogram (ECG)-triggered spoiled gradient echo sequence, acquiring eight echoes with spacing 2.11 ms and initial TE 1.47 ms. Other parameters were fileld of view = 40.0 cm × 40.0 cm, bandwidth = 977 Hz/pixel, pulse repetition time = 18.03 ms, flip angle = 10°; matrix size = 256 × 156, with 13 views per segment. These imaging parameters result in <4% bias in fat amplitude estimation in the liver, given typical relaxation parameters of fat and liver water at 1.5T (37). For low FFs, this bias results in very small systematic errors in FF estimation: for instance, if the true FF is 3.00%, a 4% positive bias in fat signal amplitude relative to water will lead to a 3.12% estimated FF in the absence of noise (i.e., an error typically well below the noise level). An additional in vivo dataset with 16 echoes was obtained to calibrate the relative amplitudes of the fat peaks in vivo.

All images were reconstructed using SNR-scaled reconstruction (23). This allows convenient evaluation of SNR at each voxel. Multicoil data were combined prior to water/fat separation, using the eigenvector filter method described in Walsh et al. (38).

Multipeak Calibration

The 32-echo dataset was used for precalibration of the multipeak fat model (frequencies and relative amplitudes of the fat peaks). A six-peak model was used in this work. The calibration was performed in two steps:

1. From the vial containing 100% fat, the relative frequencies and relative amplitudes of six fat peaks were estimated by nonlinear least- squares fitting. Note that this step provides a good calibration of the relative amplitudes of the fat peaks (due to the absence of a water peak to interfere with the calibration of fat peaks near the water peak). However, their frequency shifts (with respect to the water peak) can only be estimated up to a common shift.
2. From the vial containing 50% fat: by fixing the relative frequencies and relative amplitudes of the fat peaks, we estimate the absolute frequency shift of the fat peaks (with respect to the water peak). This was done by estimating a common shift for the fat relative frequencies, using nonlinear least-squares fitting.

The calibration resulted in fat peaks at −244.3, −221.7, −175.4, −119.3, −32.1, and 34.0 Hz, with complex-valued relative amplitudes (where the phase of the main fat peak at −221.7 Hz is kept at zero) $0.01 \cdot [9.45e^{-i\pi 0.181}, 64.66, 9.67e^{i\pi 0.046}, 2.26e^{-i\pi 0.567}, 2.22e^{-i\pi 0.244}, 8.83e^{-i\pi 0.089}]$. The nonzero phase of the smaller peaks is likely due to the fact that even this relatively sophisticated six-peak fat model is only an approximation, and some of the calibrated peaks in fact consist of several peaks at nearby frequencies.

Multipeak precalibration for in vivo fitting was performed on a 16-echo in vivo dataset (6). The frequencies of the six fat peaks were kept constant, and their calibrated relative amplitudes were, respectively, $0.01 \cdot [7.98e^{-i\pi 0.142}$, 70.0, $8.38e^{i\pi 0.121}$, $1.52e^{-i\pi 0.892}$, $5.04e^{i\pi 0.112}$, $7.09e^{-i\pi 0.162}]$.

Comparison Strategy: Bias and Standard Deviation

We use the estimated bias and standard deviation to compare the performance of the different signal models. Intuitively, the bias reflects the model mismatch, whereas the standard deviation reflects the noise sensitivity. Two different quantities were considered: the estimated fat amplitude and the estimated FF. These were evaluated as follows:

- The fat amplitude was analyzed in terms of the standard deviation of the estimates, as well as the root mean squared error (RMSE), which is due to both bias and standard deviation. The resulting standard deviation was compared to the CRLB for each model. For the phantom data, averaged estimates (from eight measurements) obtained using a two-decay multipeak model on the 32-echo acquisition were used as the gold standard.
- The FF was analyzed in terms of mean ± standard deviation for each true FF (both in simulation and in the phantom data). Since the FFs were measured when building the phantom, the known (volume) FF was used as the gold standard for the observed (signal) FF (39).

The noise performance of the amplitude estimates for different models was compared using the $\sqrt{\text{NSA}} = \sigma_n/\sigma_F$, where σ_n is the standard deviation of the noise in the acquired images and σ_F is the standard deviation of the fat amplitude estimates. $\sqrt{\text{NSA}}$ was used instead of number of signal averages (NSA) to reflect SNR more directly, and also because of its decreased range, in order to allow better visualization and comparison of the different models. NSA values for multipeak models were adjusted for the Euclidean norm of the fat signal model (since the relative amplitudes were normalized so that their magnitudes add up to 1, the Euclidean norm was <1, which needs to be accounted for when computing NSA).

Assumptions and Fixed Parameters

The present study includes several assumptions and constraints:

- Multipeak fat modeling is limited to a six-peak model, even though more peaks can be found by spectroscopy. Previous works have also used three-peak fat models for ease of calibration (17,22).
- This work uses a fixed set of TEs, similar to those used in previous works (16). The effect of varying the choice of TEs on the different models is not analyzed but is not expected to alter the conclusions of this work.
- For the computation of fat amplitude estimation errors using different models, the gold standard was an eight-averaged estimate from the 32-echo acquisition, using a complex, multipeak, two-decay model (i.e., the most sophisticated among all the models under consideration).

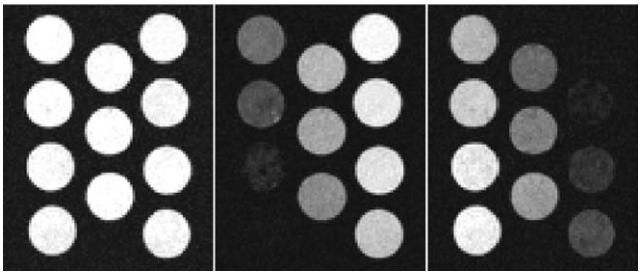

FIG. 1. Image of the phantom used in this study, including in-phase image (generated by combining the estimated water and fat images) and water/fat separated images.

- This work does not focus on the ability of different models to prevent water/fat swaps. Thus, the B_0 field map is first estimated using a spatially regularized formulation (22), and subsequently the different models are applied voxel by voxel. The complex fitting methods use the regularized field map as initial estimate at each voxel for the descent procedure. The magnitude fitting methods are initialized with the water/fat amplitudes obtained from the corresponding complex fitting methods (similarly to Yu et al. (40)).

RESULTS

Figure 1 shows the phantom setup used in this work, including an in-phase image, as well as separated water and fat images. The average estimated relaxation parameter values in the water component (water-only vial) were $T_{1,W} = 953$ ms and $T_{2,W} = 82$ ms; in the fat component (fat-only vial), these values were $T_{1,F} = 207$ ms and $T_{2,F} = 43$ ms. It must be noted that $T_{1,W}$ seemed to decrease in the mixed vials (e.g., it was measured to be 813 ms in the vial containing 50% fat) (41). However, this range of values does not affect the results of bias and standard deviation comparison as the sequence parameters were chosen to avoid T_1 weighting.

Results are shown in the form of sets of 12 plots/images (one for each of the 12 models under study). First, we examine the standard deviation of the fat amplitude estimates, without regard for bias. Figure 2 shows $\sqrt{\text{NSA}}$ for SNR = 100 and a range of water/fat ratios. Note that the maximum $\sqrt{\text{NSA}}$ attainable by any unbiased estimator is $\sqrt{8}$. Figure 2 includes both CRLB-based predictions and results on simulated data. CRLB and simulations provide similar results, but not equal, largely due to model mismatches (which are not accounted for in the CRLB). In other words, only the multipeak, two-decay signal models are fitting the correct model to the data.

Figure 3 shows similar results, but comparing the CRLB predictions with $\sqrt{\text{NSA}}$ based on the measured standard deviation for fat amplitude estimation in the actual phantom experiments. Note that the phantom results closely follow the simulations (shown in Fig. 2), with the largest difference arising in the magnitude fitting using a single peak and no decay, where the phantom estimates often converged to zero at low FFs, thus showing very low standard deviation (and very high NSA). Aside from that effect,

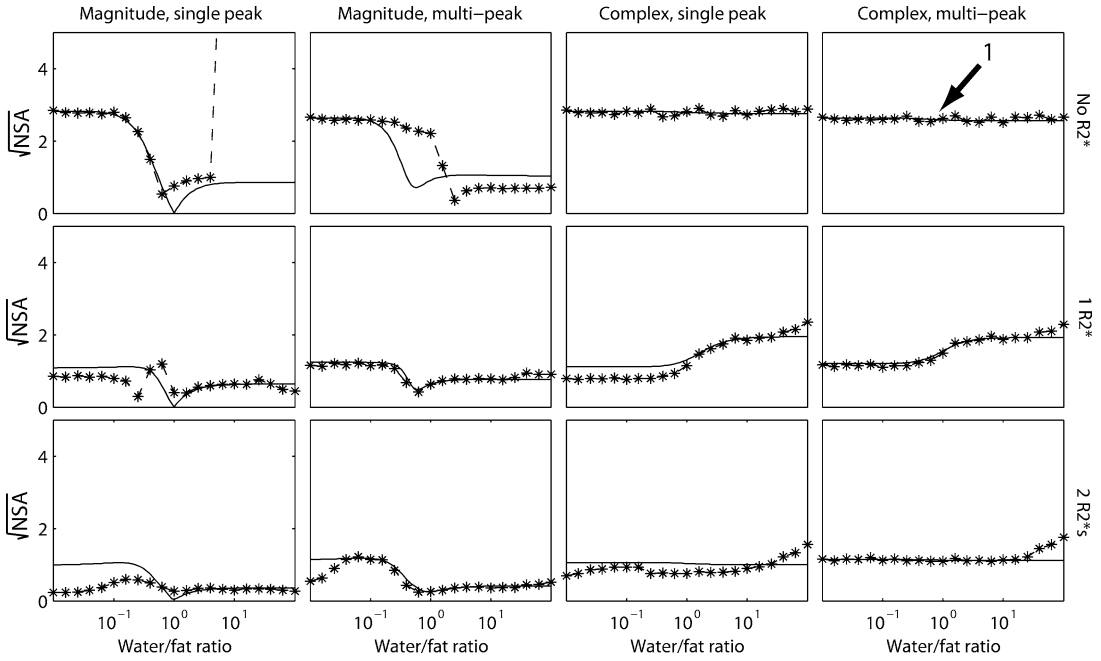

FIG. 2. Fat quantification $\sqrt{\text{NSA}}$ on simulated data for SNR = 100. The stars show the $\sqrt{\text{NSA}}$ values obtained by simulation (including model mismatch), and the solid line shows the $\sqrt{\text{NSA}}$ values predicted by CRLB analysis (without model mismatch). The arrow highlights the fact that the CRLB-based NSA provides a good indication for observed noise performance (particularly for complex-fitting methods), even in the presence of model mismatch.

magnitude-fitting models result in lower NSA than their complex-fitting counterparts, both in theory (CRLB) and in practice (simulations and phantom data).

Figures 4 and 5 show the standard deviation σ_F and the RMSE for fat amplitude estimation using the 12 models, both for the simulation (Fig. 4) and for the phantom data (Fig. 5). Note the close correspondence of simulation and phantom results for most models. Several of the magnitude-fitting models present a larger discrepancy between simulation and phantom data. We suggest that this

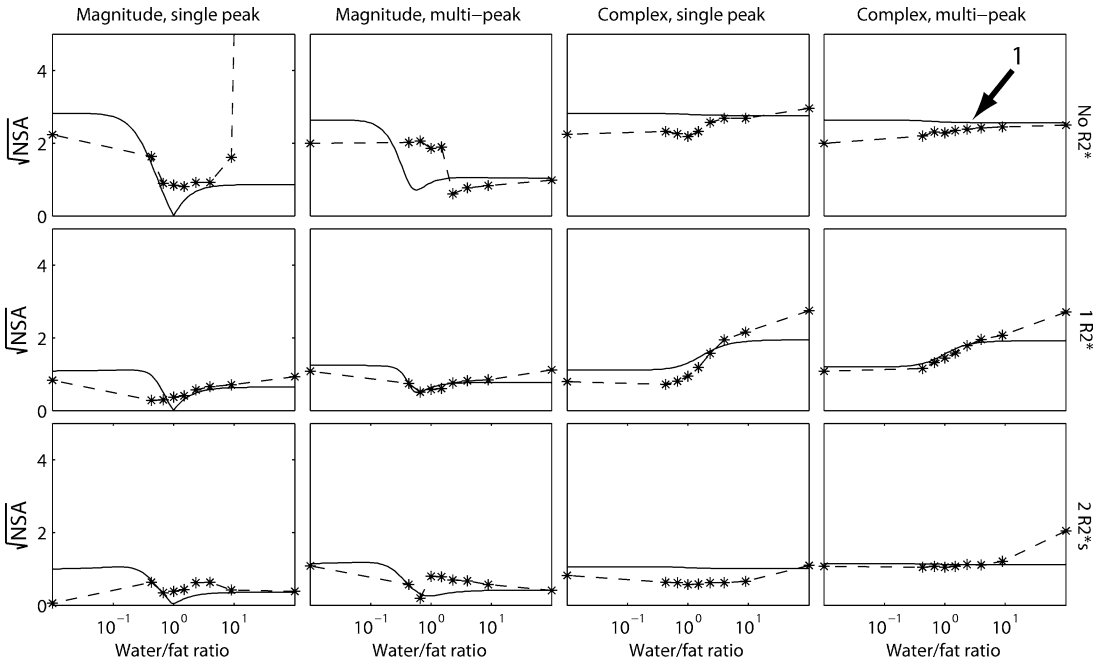

FIG. 3. Fat quantification $\sqrt{\text{NSA}}$ on phantom data for SNR = 90. The stars show the $\sqrt{\text{NSA}}$ values obtained from the phantom data (including model mismatch), and the solid line shows the $\sqrt{\text{NSA}}$ values predicted by CRLB analysis (without model mismatch). The arrow highlights the fact that the CRLB-based NSA provides a good indication for observed noise performance (particularly for complex-fitting methods), even in the presence of model mismatch.

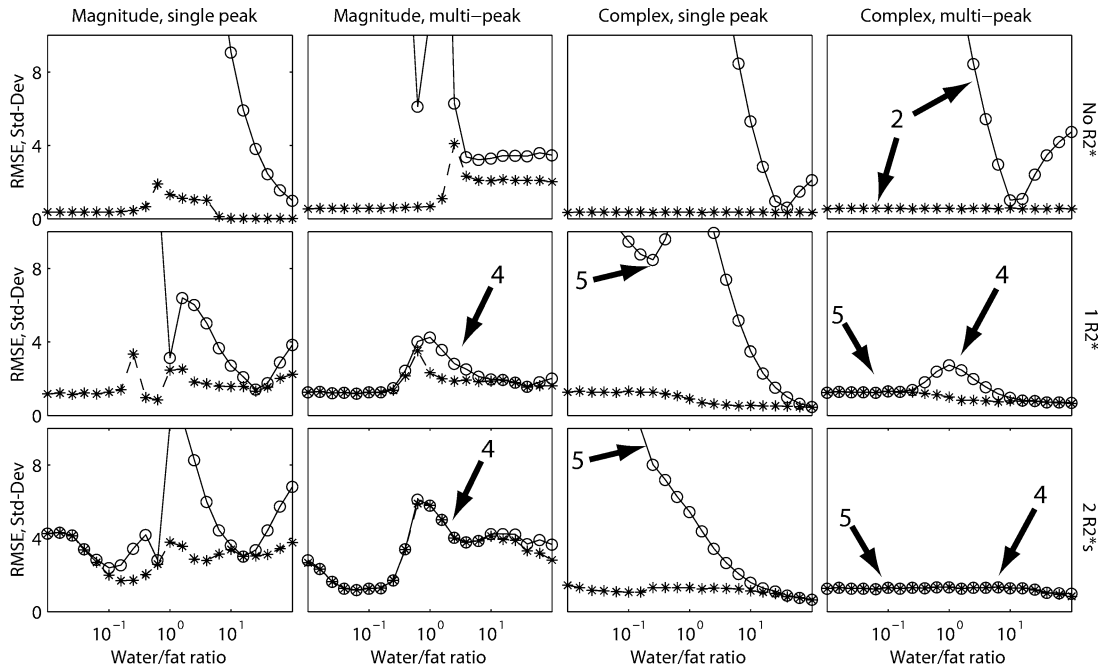

FIG. 4. Fat quantification standard deviation (stars) and RMSE (circles) on simulated data for high SNR (SNR = 100). Arrows with different labels highlight different aspects of these results: "2," in the presence of model mismatch, the bias component of the RMSE can be significantly larger than the standard deviation; "4," complex fitting generally results in better estimates (lower standard deviation and RMSE) compared to magnitude fitting; "5," for one- and two-decay complex fitting, multipeak models largely remove the bias present in single-peak models.

discrepancy might be due to residual model mismatches in the phantom case. A more detailed discussion of this effect will be deferred to the description of FF estimation results. The simpler models (e.g., without accounting for R_2^* or multipeak fat), produce significant bias in the estimation of fat amplitudes, resulting in RMSE much higher than σ_F. For these models, the bias dominates the errors. Therefore, an analysis of these based only on CRLB (or standard deviations) will not give an accurate assessment of the quality of the estimates.

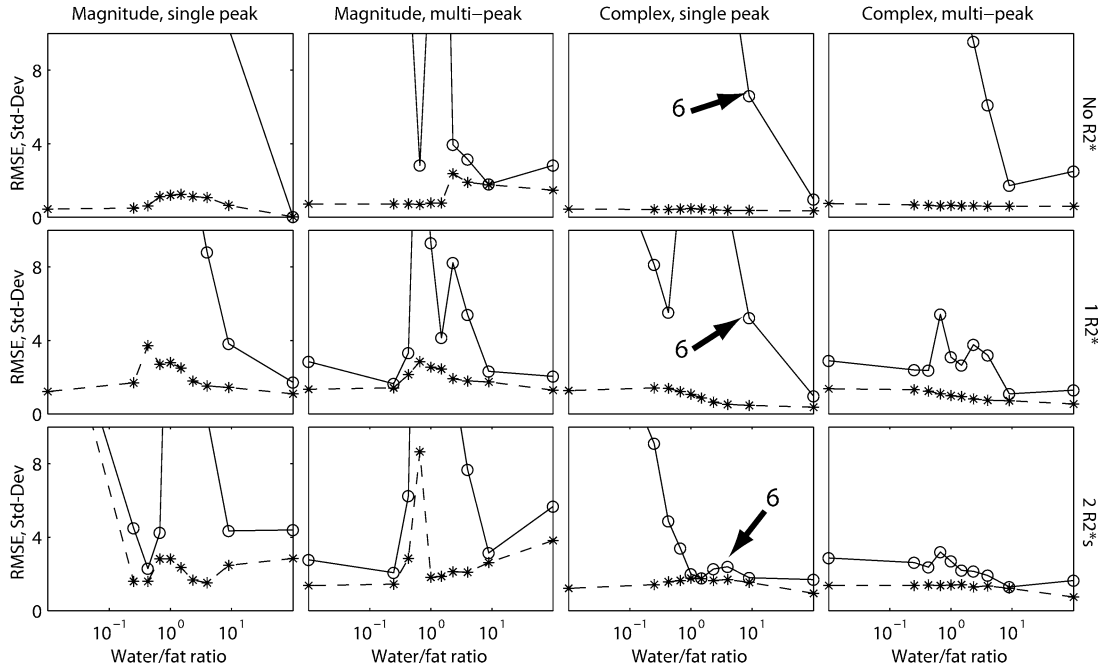

FIG. 5. Fat quantification standard deviation (stars) and RMSE (circles) on phantom data for high SNR (SNR = 90). The arrows highlight the bias incurred by single-peak fat modeling. For single-peak fat modeling, a two-decay model is able to reduce bias over a range of FFs by allowing the estimated fat signal to decay faster than the water signal, approximately accounting for the multipeak nature of the fat signal.

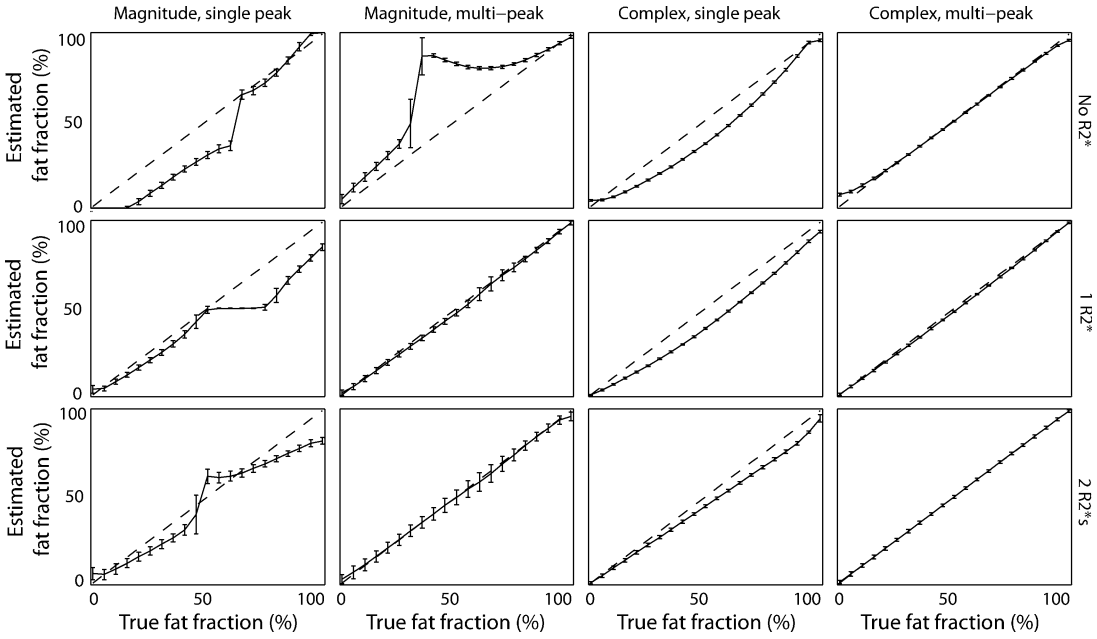

FIG. 6. Fat quantification FF on simulated data including standard deviations for high SNR. The dashed line shows the desired exact estimates.

Figures 6 and 7 show FF results (mean ± standard deviation) for simulated and phantom data for a range of true FFs between 0% and 100%. All single-peak models result in considerable bias. For the multipeak, no-decay model, the bias in fat amplitude estimation seems to be approximately compensated by the bias in water amplitude estimation, resulting in good estimates except at very low or very high FFs. Generally, complex-fitting models perform significantly better (smaller bias and standard deviation) than their magnitude-fitting counterparts. Furthermore, complex-fitting phantom results show better agreement with simulation results. Magnitude-fitting phantom results show somewhat different behavior (most notably an increased bias) with respect to the simulations. We hypothesize that the cause is the sensitivity of magnitude fitting to model mismatches. To test this hypothesis, we generated a second set of simulated data, where the multipeak (six-peak) fat model is not exactly correct, but

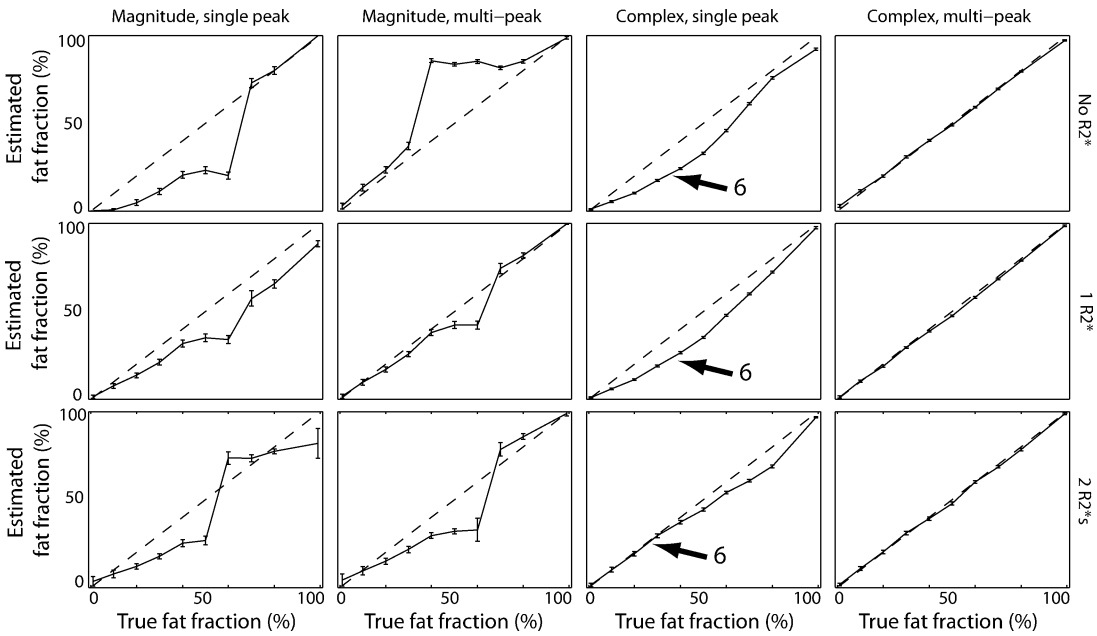

FIG. 7. Fat quantification FF on phantom data including standard deviations for high SNR (SNR = 90). The dashed line shows the desired exact estimates. The arrows highlight the differences among complex-fitting, single-peak fat models: improved estimates for low FF values are obtained with the two-decay model compared to no-decay and one-decay models All three complex-fitting, multipeak models show good performance in this case.

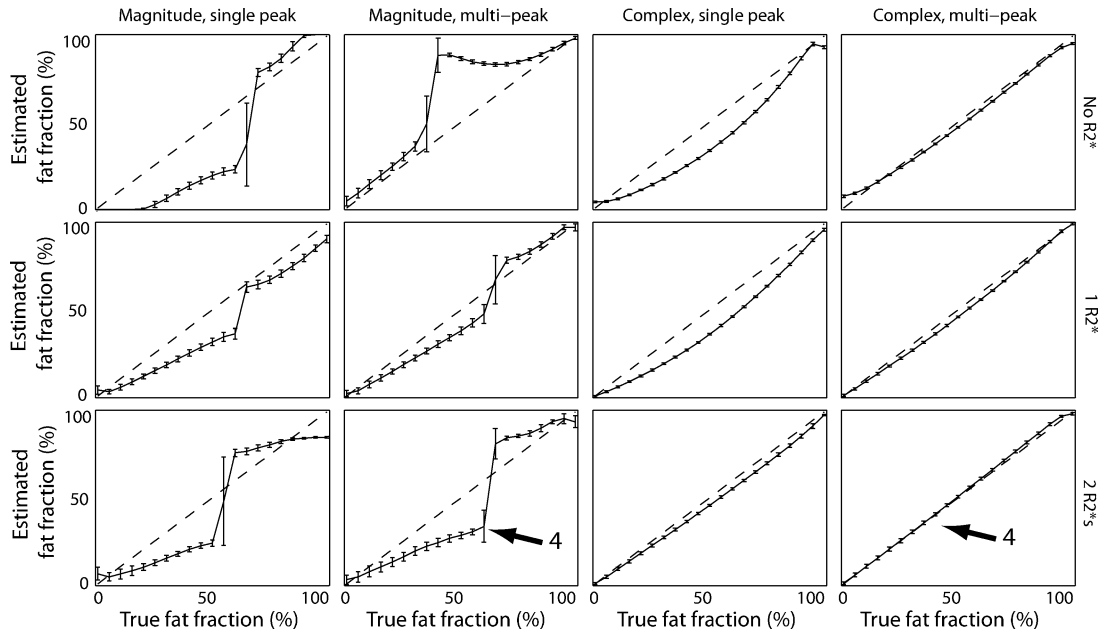

FIG. 8. Fat quantification FF on simulated data including calibration error for high SNR (SNR = 100). Magnitude-fitting models are more heavily perturbed by calibration error compared to complex fitting (see arrows).

instead the peaks at −175 Hz and −119 Hz were each split into two peaks separated by 10 Hz, with the same amplitude as the original peak. Noise was added to the resulting simulated data, as described in the "Materials and Methods," and the resulting signals were fitted using all 12 models, where the multipeak model still consisted of the original six peaks. The resulting FF plots are shown in Fig. 8. The complex-fitting results are similar to the ones shown in Fig. 6. However, the magnitude-fitting results have increased bias and standard deviation due to the model mismatch. These results correspond well with the observed phantom results (Fig. 7).

In vivo liver imaging results are shown in Fig. 9. The SNR was approximately 20. The FF maps shown in Fig. 9 are provided to illustrate the differences in bias and standard deviation for the various signal models used for fat and water fitting. The low SNR of the fat in the liver region leads to a noise bias (15). Estimates of FF were calculated from the mean values of fat and water signal intensities within a circular region of interest rather than from the FF map, which is noisier. Furthermore, the complex fat images were filtered to improve the SNR. Using a 7×7 filter, yielded an SNR for fat signal of approximately 5 for the complex-fitting, multipeak, one-decay estimates, which results in noise bias error under 5%. All signal models are affected similarly by noise bias, which was not the objective of the paper. It must be noted that we do not have a ground truth for the in vivo data, but rather compare only the relative estimates of the different models. The single-peak models (with the exception of the complex-fitting, single-peak, two-decay model) result in lower FF estimates relative to the multipeak models. This is in good agreement with simulation and phantom results. Additionally, the two-decay estimates are noisier compared with the no-decay and one-decay models (with the exception of the magnitude-fitting, multipeak, one-decay model, which produces unstable results due to model mismatch).

Based on these results, we can highlight the following key observations (arrows are marked in the figures with the corresponding observation number):

1. Despite the model mismatch, the CRLB provides a useful approximation of the standard deviation obtained with the different models. However, the CRLB does not take model mismatch–related bias into account.
2. The bias component of the RMSE can be significantly larger than the standard deviation component.
3. The relative importance of the bias component with respect to the standard deviation component is a function of the SNR. This is shown in Fig. 10, where complex-fitting, multipeak fat models are compared. For low SNRs, the standard deviation component of the error, which is larger in the two-decay model, dominates the (approximate constant with SNR) bias component of the error, which is larger in the one-decay model.
4. Complex fitting results in better estimates than magnitude fitting. This is true for the standard deviation (as shown by the CRLB, simulation, and phantom results), as well as for the bias (as shown by the simulation and phantom results). Additionally, complex fitting is less sensitive to model mismatch.
5. Multipeak has significantly reduced bias error compared to single peak. Furthermore, single-peak models perform worse when there is more fat.
6. The no-decay models result in very large bias for fat amplitude estimation. For single-peak fat modeling, the two-decay model is needed in order to approximately account for the multipeak nature of the fat signal.

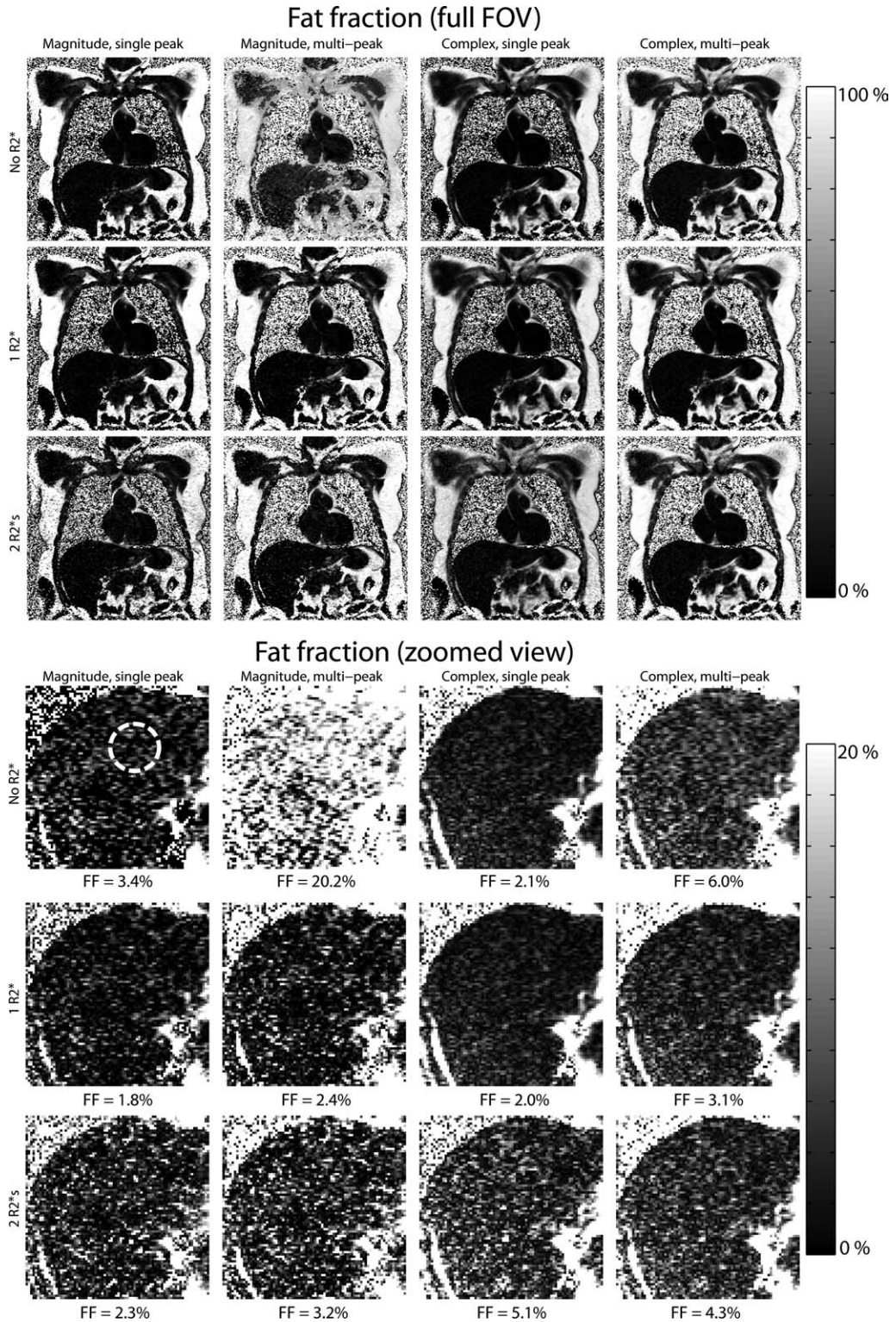

FIG. 9. In vivo liver FF estimates using all 12 models. (Top) Full field of view. (Bottom) Zoomed view on liver, with grayscale adapted to highlight the differences between the models. The quantified FF shown for each model was calculated using mean estimates for fat and water magnitudes over the shown region of interest (dashed circumference) after smoothing to improve SNR and reduce noise bias.

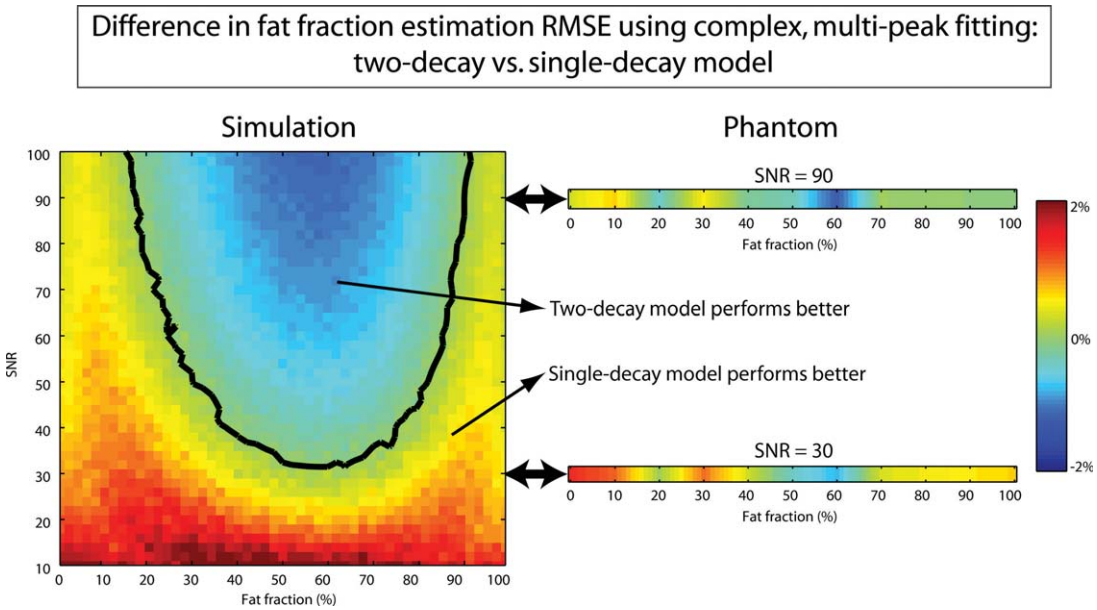

FIG. 10. Difference in RMSE for FF estimation with complex-fitting, multipeak models including two decay rates and a single decay rate. The contour shows the region where both models result in the same RMSE. For low SNR or FFs close to 0% or 100%, the one-decay model results in lower errors. For high enough SNR and FFs close to 50%, the two-decay model results in lower errors. There is some discrepancy between simulation and phantom results, particularly for FF = 20%, where in the phantom data the one-decay model resulted in higher bias than for neighboring values of FF, thus compensating for its reduced variance. Currently, we do not have an explanation for this effect.

7. For multipeak fat modeling, the two-decay model typically results in lower bias than the one-decay model, but the increased standard deviation results in higher errors except at high SNR and FFs close to 50%. For SNRs <30, the increased standard deviation in the two-decay model dominates the improvement in bias with respect to the one-decay model. This is in good agreement with Chebrolu et al. (28) and is demonstrated in Fig. 10 with simulation and phantom results for a range of SNR and FF values.

DISCUSSION

We have performed a systematic comparison of signal models for water/fat separation from chemical shift–encoded acquisitions. The analysis was based on comparing the bias and standard deviation resulting from the different models. This study can be viewed as an extension of previous work, e.g., where the standard deviation was studied for different acquisition strategies using the CRLB (31), or different sets of models were compared empirically (6,16).

The present study has several limitations. First, the study assumes that the signal phase is reliable. Under these conditions, complex fitting is uniformly superior to magnitude fitting. In the presence of phase distortions (e.g., due to eddy currents), magnitude fitting (16) or a mixed approach (40) may become more attractive. However, phase distortions were not found to be significant in our experimental data. Similarly, ghosting dut to motion may complicate the fitting, but it was not observed in our in vivo data. Second, the study assumes that a suitable calibration is available for multipeak fat models. Third, in order to limit the complexity of the study, we fixed several parameters such as the choice of TEs. The present set of eight TEs allowed stable application of even the more complicated, two-decay models. Using fewer TEs (e.g., four) is expected to result in increased noise sensitivity, particularly in the more sophisticated two-decay models. This choice was made to approximately follow the usual sets of TEs in recent fat quantification literature (16). Fourth, this study does not take computation time into account. Generally, increasing the number of parameters (especially nonlinear parameters) in a model will result in increased computation. For instance, computation times to process 1024 voxels with the three complex multipeak models (no-decay, one-decay, and two-decay models), were 8.9, 9.6, and 16.4 sec, respectively, in our nonoptimized MatLab (MathWorks) implementation.

Multipeak fat modeling has been shown in this and previous work to result in reduced bias in fat quantification relative to single-peak fat modeling (6). However, the present results seem to indicate that even the six-peak fat model with separate R_2^* decays for water and fat does not completely describe the fat signal. This residual model mismatch appears in two ways: (a) the multiple fat peaks are not all in phase in the calibration, and (b) magnitude fitting contains significant bias. However, incorporating more peaks into the model results in more difficult calibration due to the complication of calibrating peaks with very similar resonant frequencies.

The decay constant R_2^* for each species can be approximated as a combination of an intrinsic component due to spin-spin interactions and an extrinsic component due to field inhomogeneities and susceptibility effects: $R^*_{2,W} = R_{2,W} + R'_2$ and $R^*_{2,F} = R_{2,F} + R'_2$, where $R_{2,W} = 1/T_{2,W}$, $R_{2,F} = 1/T_{2,F}$, $R'_2 \sim \gamma \Delta B$, and ΔB is the amount of B_0 field

variation within the voxel (42,43). Thus, $R^*_{2,W}$ and $R^*_{2,F}$ will generally be different, which is observed in the phantom data, using a multipeak, two-decay model, where the estimated difference was $R^*_{2,F} - R^*_{2,W} \approx 12$ sec^{-1}. This is in good agreement with the T_2 relaxation parameters measured in the phantom (using a spin-echo sequence with varying TEs), where $T_{2,W} \approx 82$ ms, and $T_{2,F} \approx 43$ ms, resulting in $R_{2,F} - R_{2,W} \approx 11$ sec^{-1}. Furthermore, according to this approximation, the difference $R^*_{2,F} - R^*_{2,W} = R_{2,F} - R_{2,W}$ can be approximately known a priori if $T_{2,W}$ and $T_{2,F}$ are assumed known. However, it has been suggested that $R^*_{2,F}$ and $R^*_{2,W}$ may behave differently, e.g., as a function of iron concentration (28). If a single-peak fat model is used, the apparent $R^*_{2,F}$ will be higher as it has to account for the dephasing due to interference between multiple fat peaks at frequencies near the the main peak. Moreover, assuming that all the fat peaks share a single $R_{2,F}$ (or $R^*_{2,F}$) is also an approximation, but estimating independent decay constants for each fat peak would result in greatly increased computational complexity and noise sensitivity, likely making it impractical. Furthermore, if the relative differences between the decay rates of the different fat peaks can be assumed known a priori, this information can also be incorporated into the model.

According to our results, finding the optimal model for water/fat separation reduces to a choice between complex, multipeak fitting including either two decays ($R^*_{2,W}, R^*_{2,F}$) or a single decay rate R^*_2. This choice presents a clear tradeoff of bias and standard deviation: the two-decay model can represent the acquired signal more accurately (reduced bias), but the estimation of an additional decay rate increases the noise sensitivity (increased standard deviation). This increased standard deviation is particularly significant in the estimates of the "minority" component of the signal: in the one-decay model, the minority component "gets to share" the R^*_2 parameter of the majority component, resulting in very stable (although somewhat biased) estimates of the minority component. In the two-decay model, estimation of the decay parameter for the minority component must be done independently, resulting in noisy decay rate estimates and in turn noisy amplitude estimates. As shown in Fig. 10, the choice between one or two decays depends on the SNR and the (expected) true FF. In several important applications, low FFs (e.g., 0–20%) are expected (5,6,9,28), which makes the one-decay model preferable unless very high SNR can be achieved.

CONCLUSIONS

This article has presented a comparative study of 12 signal models for quantitative water/fat separation. The models were compared based on an analysis of the bias and standard deviation of their resulting estimates. Results from theoretical analysis, simulation, phantom experiments, and in vivo data are in good agreement. This study shows that complex fitting is uniformly superior to magnitude fitting, multipeak fat modeling is able to remove the bias present in single-peak fat modeling, and a single-T^*_2 model performs best over a range of clinically relevant SNRs and water/fat ratios.

ACKNOWLEDGEMENTS

This research was supported in part by the Intramural Research Program of the NIH; National Heart, Lung, and Blood Institute; and a Cooperative Research and Development Agreement between the National Heart, Lung, and Blood Institute and Siemens Medical Solutions.

REFERENCES

1. Schellinger D, Lin CS, Lim J, Hatipoglu HG, Pezzullo JC, Singer AJ. Bone marrow fat and bone mineral density on proton MR spectroscopy and dual-energy x-ray absorptiometry: their ratio as a new indicator of bone weakening. AJR Am J Roentgenol 2004;183:1761–1765.
2. Agannathan NR, Singh M, Govindaraju V, Raghunathan P, Coshic O, Julka PK, Rath GK. Volume localized in vivo proton MR spectroscopy of breast carcinoma: variation of water-fat ratio in patients receiving chemotherapy. NMR Biomed 1998;11:414–422.
3. Boesch C, Machann J, Vermathen P, Schick F. Role of proton MR for the study of muscle lipid metabolism. NMR Biomed 2006;19:968–988.
4. Murphy PS, Rowland IJ, Viviers L, Brada M, Leach MO, Dzik-Jurasz AS. Could assessment of glioma methylene lipid resonance by in vivo (1)H-MRS be of clinical value? Br J Radiol 2003;76:459–463.
5. Hussain HK, Chenevert TL, Londy FJ, Gulani V, Swanson SD, McKenna BJ, Appelman HD, Adusumilli S, Greenson JK, Conjeevaram HS. Hepatic fat fraction: MR imaging for quantitative measurement and display–early experience. Radiology 2005;237:1048–1055.
6. Reeder SB, Robson PM, Yu H, Shimakawa A, Hines CDG, McKenzie CA, Brittain JH. Quantification of hepatic steatosis with MRI: the effects of accurate fat spectral modeling. J Magn Reson Imaging 2009;29:1332–1339.
7. Sharma S, Adrogue JV, Golfman L, Uray I, Lemm J, Youker K, Noon GP, Frazier OH, Taegtmeyer H. Intramyocardial lipid accumulation in the failing human heart resembles the lipotoxic rat heart. FASEB J 2004;18:1692–1700.
8. McGavock JM, Lingvay I, Zib I, Tillery T, Salas N, Unger R, Levine BD, Raskin P, Victor RG, Szczepaniak LS. Cardiac steatosis in diabetes mellitus: a ^{1}H-magnetic resonance spectroscopy study. Circulation 2007;116:1110–1112.
9. Liu CY, Redheuil A, Ouwerkerk R, Lima J, Bluemke D. Myocardial fat quantification using 2D Dixon MRI: feasibility study. J Cardiovasc Magn Reson 2009;11(suppl 1):107.
10. Reeder SB, Wen Z, Yu H, Pineda AR, Gold GE, Markl M, Pelc NJ. Multicoil Dixon chemical species separation with an iterative least squares estimation method. Magn Reson Med 2004;51:35–45.
11. Reeder SB, Markl M, Yu H, Hellinger JC, Herfkens RJ, Pelc NJ. Cardiac CINE imaging with IDEAL water-fat separation and steady-state free precession. J Magn Reson Imaging 2005;22:44–52.
12. Ma J. Dixon techniques for water and fat imaging. J Magn Reson Imaging 2008;28:543–558.
13. Glover GH, Schneider E. Three-point Dixon technique for true water/fat decomposition with B_0 inhomogeneity correction. Magn Reson Med 1991;18:371–383.
14. Yu H, Reeder SB, Shimakawa A, Brittain JH, Pelc NJ. Field map estimation with a region growing scheme for iterative 3-point water-fat decomposition. Magn Reson Med 2005;54:1032–1039.
15. Liu CY, McKenzie CA, Yu H, Brittain JH, Reeder SB. Fat quantification with IDEAL gradient echo imaging: correction of bias from T1 and noise. Magn Reson Med 2007;58:354–364.
16. Bydder M, Yokoo T, Hamilton G, Middleton MS, Chavez AD, Schwimmer JB, Lavine JE, Sirlin CB. Relaxation effects in the quantification of fat using gradient echo imaging. Magn Reson Imaging 2008;26:347–359.
17. Yu H, Shimakawa A, McKenzie CA, Brodsky EK, Brittain JH, Reeder SB. Multiecho water-fat separation and simultaneous R2* estimation with multifrequency fat spectrum modeling. Magn Reson Med 2008;60:1122–1134.
18. Szumowski J, Coshow WR, Li F, Quinn SF. Phase unwrapping in the three-point Dixon method for fat suppression MR imaging. Radiology 1994;192:551–561.
19. Xiang QS, An L. Water-fat imaging with direct phase encoding. J Magn Reson Imaging 1997;7:1002–1015.
20. Lu W, Hargreaves BA. Multiresolution field map estimation using golden section search for water-fat separation. Magn Reson Med 2008;60:236–244.

21. Huh W, Fessler JA, Samsonov AA. Water-Fat decomposition with regularized field map. In Proceedings of the 16th annual meeting of ISMRM, Toronto, Canada, 2008; p 1382.
22. Hernando D, Kellman P, Haldar JP, Liang ZP. Robust water/fat separation in the presence of large field inhomogeneities using a graph cut algorithm. Magn Reson Med 2010;63:79–90.
23. Kellman P, McVeigh ER. Image reconstruction in SNR units: a general method for SNR measurement [published erratum in Magn Reson Med 2007;58:211–212]. Magn Reson Med 2005;54:1439–1447.
24. O'Regan DP, Callaghan MF, Wylezinska-Arridge M, Fitzpatrick J, Naoumova RP, Hajnal JV, Schmitz SA. Liver fat content and T2*: simultaneous measurement by using breath-hold multiecho MR imaging at 3.0 T—feasibility. Radiology 2008;247:550–557.
25. Yokoo T, Bydder M, Hamilton G, Middleton MS, Gamst AC, Wolfson T, Hassanein T, Patton HM, Lavine JE, Schwimmer JB, Sirlin CB. Nonalcoholic fatty liver disease: diagnostic and fat-grading accuracy of low-flip-angle multiecho gradient-recalled-echo MR imaging at 1.5 T. Radiology 2009;251:67–76.
26. Ma J, Wehrli FW, Song HK, Hwang SN. A single-scan imaging technique for measurement of the relative concentrations of fat and water protons and their transverse relaxation times. J Magn Reson 1997;125: 92–101.
27. Glover GH. Multipoint Dixon technique for water and fat proton and susceptibility imaging. J Magn Reson Imaging 1991;1: 521–530.
28. Chebrolu VV, Hines CD, Yu H, Pineda AR, Shimakawa A, McKenzie C, Brittain JH, Reeder SB. Independent estimation of T2* for water and fat for improved accuracy of fat quantification. In Proceedings of the 17th annual meeting of ISMRM, Honolulu, HI, 2009; p 2847.
29. Yu H, McKenzie CA, Shimakawa A, Vu AT, Brau ACS, Beatty PJ, Pineda AR, Brittain JH, Reeder SB. Multiecho reconstruction for simultaneous water-fat decomposition and T2* estimation. J Magn Reson Imaging 2007;26:1153–1161.
30. Hernando D, Haldar JP, Sutton BP, Ma J, Kellman P, Liang ZP. Joint estimation of water/fat images and field inhomogeneity map. Magn Reson Med 2008;59:571–580.
31. Pineda A, Reeder S, Wen Z, Pelc NJ. Cramér-Rao bounds for three-point decomposition of water and fat. Magn Reson Med 2005;54: 625–635.
32. Scharf LL, McWhorter LT. Geometry of the Cramér-Rao bound. In Proceedings of the IEEE Sixth SP Workshop on Statistical Signal and Array Processing, Victoria, BC, Canada, 1992. pp 5–8.
33. Karlsen OT, Verhagen R, Vovée WMMJ. Parameter estimation from Rician distributed data sets using a maximum likelihood estimator: application to T1 and perfusion measurements. Magn Reson Med 1999;41:614–623.
34. Bernard CP, Liney GP, Manton DJ, Turnbull LW, Langton CM. Comparison of fat quantification methods: a phantom study at 3.0T. J Magn Reson Imaging 2008;27:192–197.
35. Hines CD, Yu H, Shimakawa A, McKenzie CA, Chebrolu VV, Brittain JH, Reeder SB. Validation of fat quantification with T2* correction and accurate spectral modeling in a novel fat-water-iron phantom. In Proceedings of the 17th annual meeting of ISMRM, Honolulu, HI, 2009; p 2707.
36. Hines CD, Yu H, Shimakawa A, McKenzie CA, Brittain JH, Reeder SB. T1 independent, T2* corrected MRI with accurate spectral modeling for quantification of fat: validation in a fat-water-SPIO phantom. Magn Reson Med 2009;30:1215–1222.
37. de Bazelaire CM, Duhamel GD, Rofsky NM, Alsop DC. MR imaging relaxation times of abdominal and and pelvic tissues measured in vivo at 3.0T: preliminary results. Radiology 2004;230:652–659.
38. Walsh DO, Gmitro AF, Marcellin MW. Adaptive reconstruction of phased array MR imagery. Magn Reson Med 2000;43:682–690.
39. Reeder SB, Hines CD, Yu H, McKenzie C, Brittain JH. On the definition of fat-fraction for in vivo fat quantification with magnetic resonance imaging. In Proceedings of the 17th annual meeting of ISMRM, Honolulu, HI, 2009; p 211.
40. Yu H, Shimakawa A, Reeder SB, McKenzie CA, Brittain JH. Magnitude fitting following phase sensitive water-fat separation to remove effects of phase errors. In Proceedings of the 17th annual meeting of ISMRM, Honolulu, HI, 2009; p 461.
41. Hu HH, Nayak KS. Change in the proton T1 of fat and water in mixture. Magn Reson Med 2010;63:494–501.
42. Haacke EM, Brown RW, Thompson MR, Venkatesan R. Magnetic resonance imaging: physical principles and sequence design, New York: Wiley-Liss; 1999; p 913.
43. Bernstein MA, King KF, Zhou XJ. Handbook of MRI pulse sequences, Burlington, MA: Academic Press; 2004; p 1017.