Comparative review of algorithms and methods for chemical-shift-encoded quantitative fat-water imaging

Pierre Daudé, Tangi Roussel, Thomas Troalen, Patrick Viout, Diego Hernando, Maxime Guye, Frank Kober, Sylviane Confort Gouny, Monique Bernard, Stanislas Rapacchi

Received: 12 January 2023 | Revised: 19 August 2023 | Accepted: 21 August 2023

DOI: 10.1002/mrm.29860

RESEARCH ARTICLE

Magnetic Resonance in Medicine

Comparative review of algorithms and methods for chemical-shift-encoded quantitative fat-water imaging

Pierre Daudé[1,2,3] | **Tangi Roussel**[1,2] | **Thomas Troalen**[4] | **Patrick Viout**[1,2] | **Diego Hernando**[5,6] | **Maxime Guye**[1,2] | **Frank Kober**[1,2] | **Sylviane Confort Gouny**[1,2] | **Monique Bernard**[1,2] | **Stanislas Rapacchi**[1,2]

[1]Aix-Marseille Univ, CNRS, CRMBM, Marseille, France

[2]APHM, Hôpital Universitaire Timone, CEMEREM, Marseille, France

[3]Cardiovascular Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA

[4]Siemens Healthcare SAS, Saint-Denis, France

[5]Radiology, University of Wisconsin-Madison, Madison, Wisconsin, USA

[6]Medical Physics, University of Wisconsin-Madison, Madison, Wisconsin, USA

Correspondence
Pierre Daudé, Centre de Résonance Magnétique Biologique et Médicale—UMR 7339, Faculté de Médecine, 27 Bd Jean Moulin, Marseille, France.
Email: pierre-daude@hotmail.fr

Funding information
Centre National de la Recherche Scientifique, Grant/Award Number: MITI-Thèses transverses; France Life Imaging (FLI), Grant/Award Number: ANR-11-INBS-0006

Abstract

Purpose: To propose a standardized comparison between state-of-the-art open-source fat-water separation algorithms for proton density fat fraction (PDFF) and R_2^* quantification using an open-source multi-language toolbox.

Methods: Eight recent open-source fat-water separation algorithms were compared in silico, in vitro, and in vivo. Multi-echo data were synthesized with varying fat-fractions, B_0 off-resonance, SNR and TEs. Experimental evaluation was conducted using calibrated fat-water phantoms acquired at 3T and multi-site open-source phantoms data. Algorithms' performances were observed on challenging in vivo datasets at 3T. Finally, reconstruction algorithms were investigated with different fat spectra to evaluate the importance of the fat model.

Results: In silico and in vitro results proved most algorithms to be not sensitive to fat-water swaps and B_0 offsets with five or more echoes. However, two methods remained inaccurate even with seven echoes and SNR = 50, and two other algorithms' precision depended on the echo spacing scheme ($p < 0.05$). The remaining four algorithms provided reliable performances with limits of agreement under 2% for PDFF and $6\,s^{-1}$ for R_2^*. The choice of fat spectrum model influenced quantification of PDFF mildly (<2% bias) and of R_2^* more severely, with errors up to $20\,s^{-1}$.

Conclusion: In promoting standardized comparisons of MRI-based fat and iron quantification using chemical-shift encoded multi-echo methods, this benchmark work has revealed some discrepancies between recent approaches for PDFF and R_2^* mapping. Explicit choices and parameterization of the fat-water algorithm appear necessary for reproducibility. This open-source toolbox further enables the user to optimize acquisition parameters by predicting algorithms' margins of errors.

KEYWORDS
chemical shift encoded imaging, fat spectrum, proton density fat fraction, water/fat separation

1 | INTRODUCTION

Chemical shift encoded MRI (CSE-MRI) techniques have become the reference for quantitative in vivo evaluation of fatty depots. CSE-MRI of water and fat signals facilitates tissue and organ fat quantification. CSE-MRI[1] acquires images at multiple TEs when fat/water signals accrue different phase. The estimated proton-density fat fraction (PDFF), which is the ratio between MR visible protons from fat over all MRI visible protons (water and fat), can be computed from these images. PDFF quantification has been established as an accurate non-invasive biomarker[2] to assess tissue adiposity in the liver,[3] bone marrow,[4] and other organs.[5] Each application holds a specific range of PDFF but also a different type of fat composition. Also from CSE-MRI and complementary to PDFF, the quantification of R_2^* decay is another biomarker of interest to further probe iron overload[6] and hemorrhage.[7] The accuracy and precision of quantitative parameters PDFF and R_2^* is increasingly challenged, demanding precision down to a few percent of PDFF or tens of s^{-1} of R_2^* during follow-up,[8–10] to differentiate the type of adipose tissue[11] or to monitor cohorts.[12] Recently, groups of experts such as the PDFF Quantitative Imaging Biomarkers Assessment (QIBA) group and the ISMRM Quantitative MR Study group provided consensus guidelines to assess new quantitative MR methods,[13,14] guidelines that this study abides. The QIBA group realized a meta-analysis[15] on hepatic PDFF (PDFF $\leq$ 50%) and assessed its reproducibility coefficient at 4.12% across experimental set-ups and its repeatability coefficient at 2.99% using the same experimental condition. Moreover, one fat-water phantom traveled to multiple sites in another QIBA study[16,17] and demonstrated reproducible PDFF measurements across various MRI vendors' solutions.

Nowadays, to obtain these quantitative PDFF and R_2^* biomarkers, several CSE-MRI based methods for fat-water signal separation have been developed. Thus, the purpose of this study is to establish and compare the quantitative performances of these methods, their bias for accuracy, and limits of agreement for precision, within a reproducible research framework.[18] Relevantly, more than a decade ago, the ISMRM 2012 fat-water MRI Workshop proposed to standardize PDFF as a quantitative imaging biomarker. Gathered algorithms were benchmarked on a multitude of in vivo datasets[19] and a MATLAB algorithm toolbox was developed and disseminated. It provided standardization for the input/output formats of algorithms and facilitated their comparison.

Unfortunately, most studies only offered an evaluation of a discrete and limited range of PDFF, often under 50% for mixed fat-water samples. For a wider evaluation, numerical simulations can complement experimental comparison to explore the full range of PDFF more extensively, validate, and compare methods in many more scenarios. For instance, a Python open-source framework[20] was developed to explore optimal acquisition parameters according to the number of peaks resolved in water-fat signal model. However, this framework considered only a single algorithm, and focused on acquisition parameters. Even recently, algorithms continue to emerge,[21] with a diversity of methods to solve fat-water signal separation. Whether fitting for magnitude, complex or hybrid data, algorithms can be based on least-square iterative approaches,[22,23] graph-cut,[24–29] region growing,[30–32] and, more recently, deep learning[33–36] approaches. In addition, further image processing continues to build on these algorithms, whether to evaluate fatty acids composition,[37,38] quantitative susceptibility mapping,[29] or temperature mapping.[39] Thus, there is value to provide independent and reproducible performance evaluation of a set of previously proposed algorithms. Finally, to assess CSE-MRI algorithms performances, a numerical toolbox should remain open-source for facilitating continuous comparison and possible extensions and support the two major programming languages (Python and MATLAB) used by MRI researchers for prototyping new methods. Therefore, the purpose of this work was to develop a multilanguage (MATLAB and Python) numerical toolbox and assess the performances of open-source state-of-the-art fat-water reconstruction methods for fat-water swapped voxels (FWswaps) and PDFF and R_2^* quantification.

2 | METHODS

2.1 | Open-source algorithms

An open-source toolbox available both in Python and MATLAB (https://github.com/pdaude/CREAM_PDFF) was implemented to assess and numerically compare the performances of recent and novel open-source fat-water separation algorithms (Table 1). The already available algorithms include: Hernando et al[25] original graph-cut method (Hernando-GC), the ISMRM 2012 challenge winner leveraging quadratic pseudo-Boolean optimization graph-cut (Fatty-Riot-GC[26]), the multi-scale approach graph-cut (MSGCA-B[27]), enhanced later with spatial smoothing (MSGCA-A[24]), the globally optimal surface estimation (GOOSE[28]), the variable layer graph-cut (VLGCA[29]), a region-based approach (B0-NICE[32]), and an IDEAL constrained estimation (IDEAL-CE[40]).

TABLE 1 Summary of evaluated state-of-the-art open-source fat-water reconstruction algorithms with specific features, corresponding references, and code repository links

Reference	Referred as	Method	Code	2D/3D	Spectrum choice	Echo spacing	Code acceleration	Year	Code repository
Boehm et al.[29]	VLGCA[1]	Graph-cut	MATLAB	2D	Free	Uniform	Parallel computing	2021	https://gitlab.com/christofboehm/fieldmap-graph-cut
Bydder et al.[40]	IDEAL-CE[2]	Region-based	MATLAB	3D	Model Constrained	Free	GPU computing	2020	https://github.com/marcsous/pdff
Andersson et al.[24]	MSGCA-A[3]	Graph-cut	MATLAB	3D	Free	Uniform	No	2018	https://github.com/Snubben-B/FW-Recon-Spatial-Smoothing
Cui et al.[28]	GOOSE[4]	Graph-cut	MATLAB	2D	Fixed[a]	Uniform/Free	No	2015	https://research.engineering.uiowa.edu/cbig/content/goose
Berglund et al.[27]	MSGCA-B[5]	Graph-cut	Python	3D	Free	Uniform	No	2017	https://github.com/bretglun/fwqpbo
Liu et al.[32]	B0-NICE[6]	Region-based	MATLAB	3D	Fixed[b]	Uniform	No	2015	https://www.mathworks.com/matlabcentral/fileexchange/48313-b0-mapping-b0-nice
Berglund et al.[26]	Fatty-Riot-GC[7]	Graph-cut	MATLAB	3D	Free	Uniform	No	2012	https://github.com/welcheb/fw_i3cm1i_3pluspoint_berglund_QPBO
Hernando et al.[25]	Hernando-GC[8]	Graph-cut	MATLAB	2D	Free	Uniform/Free	No	2012	https://www.ismrm.org/workshops/FatWater12/data.htm

Note: (1) VLGCA: Variable-Layer Graph-Cut Algorithm. (2) IDEAL-CE: Iterative Decomposition of water and fat with Echo Asymmetry and Least-squares estimation Constraint Extended. (3/5) MSGCA-A/B Multi Scale Graph Cut Algorithm. (4) GOOSE: GlObally Optimal Surface Estimation. (6) B0-NICE: B_0 off-resonance mapping by Non-Iterative Correction of phase-Errors. (7/8) GC: Graph Cut.

[a] ISMRM 2012 challenge spectrum.

[b] Their own spectrum.

2.2 | Algorithm standardization

Algorithms were standardized, building upon ISMRM fat-water toolbox input MATLAB structure, with the addition of the voxel dimension. Toward generalization of applications, all algorithms were adapted to accommodate an input fat spectrum in their models. Output structure from algorithms comprised of algorithms' parameters for reproducible research, employed model fat spectrum, and maps of fat, water, R_2^*, B_0, PDFF, and the voxel-wise sum of square error. For all graph-cut algorithms, the discretization of B_0 off resonance steps was set to 2 Hz.

2.3 | Monte Carlo simulation

To numerically evaluate the algorithms' performances, synthetic CSE-MRI volumes were modeled as:

$$y = \left(W + F \sum_{m=1}^{m} \alpha_m e^{i\omega_m t} \right) e^{i\varphi_0 + (i2\pi f_0 - R_2^*)t} + \eta(t) \qquad (1)$$

with W and F corresponding to relative water and fat absolute signal, f_0 the off-resonance frequency, $R_2^* = 50$, 150 or 300 s^{-1} the transverse decay rate, $\varphi_0 = 30°$ the initial phase which holds only for low flip angles as demonstrated by Wang et al.,[41] $\eta(t)$ the complex gaussian noise and α_m, $e^{i\omega_m}$ the relative amplitude and frequency offset of a subcutaneous fat spectrum respectively. Considering a 3T scanner field strength (the exact value was 2.89T, which was further employed in this study), virtual CSE-MRI volumes were synthesized as follows: PDFF varied from 0 to 100% with 1% step, f_0 was uniformly distributed from −300 Hz to 300 Hz[42] with 6 Hz step, and the third dimension consisted in 100 repetitions varying according to a Gaussian noise $n(t)$ which was added to obtain SNR = 10, 50, 100 for the first TE. Synthetic volumes intensity was normalized based on 99% of the maximum of the first echo. To avoid border effects due to spatial regularization, a five pixels padding was added for each CSE-MRI volume. Different numbers of TEs (NTE = 3, 5, 7, 9) and echo spacing schemes were considered.

2.4 | Acquisition parameter: echo spacing schemes

The impact of three schemes for TE sampling was also considered in this study. Only TE schemes using single-TR monopolar readout gradients (often termed "flyback" echo trains) were considered to avoid the additional issue of gradients distortions. Realistic physical acquisition constraints were also included since they influence possible echo-times sampling in their minimum TE and echo-spacing, depending on the magnetic field strength, sampling dwell time and resolution. Additionally, most compared algorithms were, in their design, constrained to uniform echo spacing imaging. Thus, in proposed toolbox a function was developed to automatically calculate uniform TE schemes based on acquisition constraints (minimum TE and echo-spacing) for any number of echoes. Three schemes were evaluated: first, alternating in and out of phase echoes; second, minimum TEs constrained by acquisitions (further referred as "realistic minimal"); and third, the uniform IDEAL echo-shift formulated by Pineda[43] for three echoes was generalized to offer realistic IDEAL echo spacing, abiding to both acquisition constraints and the following extended-IDEAL criterion:

$$\Delta\theta_{k,j} = \frac{2k}{N}\pi \text{ with } k \in [\,1, N] \text{ and } \gcd(k, N) = 1, \quad (\gcd = \text{greatest common divisor}). \qquad (2)$$

This equation enables to provide the shortest angular step that leads to a uniform N-sampling of the unit circle of fat-water phase, given hardware constraints (that prevent the shortest $\frac{2\pi}{N}$ steps). Demonstration of this criterion is detailed in Data S1.

In practice, the extended IDEAL echo spacing will obtain shorter TEs for prime or odd number of echoes than for even number of echoes. As such, the three TEs schemes were evaluated for 3, 5, 7, and 9 echoes each. Hardware constraints were defined for a 3T acquisition with minimum TE $TE_{min} = 0.98$ ms and minimum echo spacing $\Delta TE_{min} = 1.68$ ms.

2.5 | Fat spectrum library

The CSE-MRI signal model (Eq. 1) relies on a multi peak fat spectrum model with established relative amplitudes and chemical shifts corresponding to triglycerides.[23,44] The choice of a multi-peak fat spectrum has been shown to reduce PDFF bias estimation compared to single fat peak spectrum model[45] but remains highly variable in the implementations. To extend from the ISMRM 2012 challenge, which benchmarked algorithms with a single human fat spectrum, this study pursued to benchmark the algorithms' sensitivity to changing the selected human fat NMR spectra.[46] In the toolbox, fat NMR spectra can be described either with a list of peaks (pairs of chemical shifts and amplitudes) using a generic triglyceride model,[47] with only three parameters: the number of double bonds (ndb), the number of interrupted methylene double bonds (nmidb), and the chain length (CL). In addition, a documented human fat spectrum[37,47,48] library was implemented in the toolbox. This simplified model

enabled us to translate gas-chromatography measurements of fatty acids composition to NMR spectrum signal. Thus, any spectrum could easily be associated with each algorithm. Unless specified, synthetic CSE volumes were modeled with the subcutaneous fat spectrum (CL = 17.29, ndb = 2.69, nmidb = 0.58). To probe spectra influence, synthetic signals (NTE = 9) were simulated with a peanut oil spectrum but processed with two very different spectra: either the same calibrated peanut oil spectrum or with the ISMRM 2012 challenge spectrum. Moreover, the effect of temperature T on the relative shift between water peak and the fat spectrum ΔCS_{wf} was also included in the toolbox using the formula[49]:

$$\Delta CS_{wf}[\text{ppm}] = 3.748 - 0.0108 * T\ [^{\circ}C]$$

2.6 | In vitro: fat-water phantom

Eight fat-water phantoms of 50 mL were prepared with different fat fractions of peanut oil. Recipes and protocols from Hines et al[50] and Bush et al[51] were followed to obtain stable and homogeneous samples. A water solution was prepared containing deionized water, 15 mM sodium dodecyl sulfate (Sigma Aldrich, St Louis, MO), 3.75 mM of sodium azide (Sigma Aldrich, St Louis, MO), 0.185 mM of $MnCl_2$ (Sigma Aldrich, St Louis, MO) to shorten $T_1^{H_2O}$ and 1.0% w/v agarose (Sigma Aldrich, St Louis, MO). High fat fraction vials (>70%) were not solidified due to a lack of oil-soluble surfactant and required shaking for homogenization prior to experiments. The volume percentages of oil in the phantoms were targeted at 0%, 10%, 20%, 40%, 60%, 80%, 90%, and 100% and their respective reference PDFF values were calculated with magnetic resonance spectroscopy (MRS) at 0%, 7.4%, 19.6%, 39.7%, 62.3%, 77.2%, 85.3%, and 100% (Figure S1). The room temperature was input to algorithms through the toolbox to correct for chemical shift variations. Reference R_2^* values were calculated as the mean value of the R_2^* values obtained from the five algorithms with minimal bias from the Monte Carlo study, using data with NTE = 9 and in/out of phase echo spacing.

Imaging and spectroscopy data were acquired in a 3T MRI system (Magnetom Vida, Siemens Healthcare, Erlangen Germany, exact field strength 2.89T) using a 20-channel head coil. Phantoms vials were aligned along the static magnetic field direction, and iterative shimming was performed prior to the acquisitions. Using a monopolar 3D spoiled gradient echo sequence, $1.8 \times 1.8 \times 2.5\ mm^3$, FOV = $172 \times 288 \times 140\ mm^3$, flip angle (FA) = 6°, bandwidth (BW) = 1563 Hz/px, TR = 17 ms, and TEs similar to the Monte Carlo simulation (realistic minimal: TE1/ΔTE = 0.98/1.68 ms; in-/out of phase: TE1/ΔTE = 1.19/3.58 ms; IDEAL: TE1 = 0.98 ms; ΔTE = 3.18, 1.91, 1.70, 1.86 ms for 3, 5, 7, and 9 echoes, respectively). Spectroscopy data were acquired using a multi-echo STEAM sequence with 1024 points, TR = 2000 ms and four TEs: 12, 24, 48 and 64 ms for each phantoms vial.

Fitting of spectroscopy data was performed using a Linear Combination model implemented in FSL-MRS[52] version 1.1.10, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Briefly, basis spectra were fitted to the complex valued spectrum in frequency domain with Voigt line shape.[53,54] The basis spectra were shifted and broadened with parameters fitted to the data grouped in two metabolites groups (water and lipids). A complex polynomial baseline was also concurrently fitted (order = 3). Model fitting was performed using the Metropolis-Hastings algorithm.

2.7 | In-vitro: multi-site fat-water phantom MRI data

To explore the reproducibility of algorithms across sites, field strength and protocols, a complementary experimental evaluation was performed on a publicly available dataset.[16,55] It included six echoes CSE-MRI acquisitions of a standardized fat-water phantom obtained across six sites at two field strengths (1.5T and 3T), with three different vendors (GE Healthcare, Philips and Siemens Healthineers) and with two different protocols (echo spacing schemes: in-/out of phase and realistic minimal). The fat-water phantom consisted of 11 vials with oil to water concentrations of: 0%, 2.6%, 5.3%, 7.9%, 10.5%, 15.7%, 20.9%, 31.2%, 41.3%, 51.4%, 100%. Following the original multi-site study,[16] bipolar acquisitions were corrected using the same method and the six-peak peanut oil spectrum model was corrected for room temperature prior to the fat-water separation processing.

2.8 | In vivo imaging

For practical demonstration purposes, algorithms were evaluated on challenging in vivo anatomies including the supraclavicular, sacral, and liver regions. Three healthy volunteers were recruited after informed consent. MR images were acquired with the same 3T MRI system using the vendor thoracic coil array and the spine coil array, totaling up to 32 channels together. Imaging the supraclavicular body fat, that contains white and brown fat, the bone marrow fat in the sacrum region, and the liver were considered challenging applications due to a large range of B_0 inhomogeneity, limited SNR ranges and various spatial distributions of adipose tissues.

Two supraclavicular CSE-MRI data sets were acquired using a coronal 3D spoiled gradient echo sequence, with (A) a spatial resolution of $1.88 \times 1.88 \times 3\,mm^3$, nine echoes (TE1/ΔTE = 0.73/1.44 ms), FOV = $360 \times 247 \times 192\,mm^3$, FA = 6°, TR = 11 ms, BW = 1532 Hz/px for all echoes and (B) a spatial resolution of $1.67 \times 1.67 \times 3\,mm^3$ six echoes (TE1/ΔTE = 1.23/1.23 ms), FOV = $319 \times 319 \times 192\,mm^3$, FA = 6, and BW = 1240 Hz/px for all echoes. Sacrum CSE-MRI was also acquired using 2D spoiled gradient echo sequence with in-plane resolution of $1.8 \times 1.8\,mm^2$, seven echoes (TE1/ΔTE = 1.2/1.54 ms), FOV = $287 \times 287\,mm^2$, slice thickness 6 mm, FA = 8°, TR = 14 ms, BW = 744 Hz/px for all echoes. Finally, liver CSE-MRI was acquired using a transversal 3D spoiled gradient echo sequence with a spatial resolution of $2.25 \times 2.25 \times 2.3\,mm^3$, seven echoes (TE1/ΔTE = 0.93/1.53 ms), FOV = $360 \times 288 \times 147\,mm^3$, FA = 5°, TR = 11 ms, BW = 1563 Hz/px for all echoes and CAIPIRINHA[56] acceleration rate 4 (2 × 2).

2.9 | Evaluation metrics and statistical analysis

The comparison followed the guidelines provided by the QIBA[13] recommendations. From Monte-Carlo simulation, algorithms results were evaluated based on their bias, and limits of agreement (LOA) for each model parameters (PDFF, B_0, R_2^*). The percentage of FWswaps was quantified. A voxel was classified as an FWswap if the estimated PDFF ($PDFF_E$) abided by the following criteria:

$$\|PDFF_E - PDFF_{\text{True}}\| \geq 10\% \text{ and}$$
$$\|PDFF_E - (100 - PDFF_{\text{True}})\| \leq 10\%$$

FWswaps were excluded from the analysis for linearity and bias.

As an indication, the computational times of the algorithms were recorded and reported as seconds per slice. Computations were performed in MATLAB R2019b and Python version 3.7.5 with a computer equipped with a GPU (Nvidia Quadro P5000, 16 Gb) and 40 CPUs (Intel Xeon e5-2630 version 4, 2.20 GHz). Statistical analysis was conducted using R (version 3.6.3).[57]

3 | RESULTS

3.1 | Numerical simulations

Expectedly, errors in PDFF and B_0 were connected (Figure 1): a bias in B_0 off-resonance field map estimation could lead to either FWswaps or PDFF bias. With only three echoes, five algorithms suffered from either FWswaps ($FWswaps_{\text{B0-NICE}} = 14\%$, $FWswaps_{\text{Fatty-Riot}} = 11\%$, and $FWswaps_{\text{IDEAL\_CE}} = 4\%$) or PDFF bias, related to B_0-inhomogenities, as shown from Hernando-GC results. MSGCA-A/B and VLGCA were the only algorithms with robust results using three echoes (FWswaps <1% and $PDFF_{\text{AbsError}} \leq 2.8\%$).

For NTE ≥5 and $R_2^* = 50\ s^{-1}$, most algorithms were robust to FWswaps (FWswaps <1%), apart from GOOSE and B0-NICE, each with 5% FWswaps that also corresponded to large B_0 absolute errors: B0-NICE = 43.3 Hz, GOOSE = 13.8 Hz. Other algorithms with no swaps provided B_0 absolute errors on the order of the 2 Hz graph-cut B_0 discretization for NTE = 7: VLGCA = 2 Hz, Hernando-GC and Fatty-Riot-GC = 0.9 Hz, IDEAL-CE = 0.5 Hz, MSGCA-A/B = 0.26 Hz. From the statistical analysis, PDFF measurement using any of the graph-cut approaches were found significantly correlated to B_0 offset-resonance ($p < 0.01$), albeit with a negligible linear correlation of 0.1% per 100 Hz. B0-NICE and GOOSE proved to remain highly biased with more echoes and were not further included in the quantitative analysis.

Unsurprisingly, the performance of the algorithms were significantly dependent on the number of echoes and echo sampling schemes (Figure S2). For NTE ≥ 5, VLGCA PDFF estimation was influenced by echo spacing schemes ($p < 0.0001$) with a preference for the in-/out of phase scheme. Increasing the number of TE (NTE = 7 and then 9) improved PDFF accuracy: absolute bias significantly decreased ($p < 0.0001$) for all algorithms (Figure S3). Considering the best NTE = 7 echo spacing scheme for each algorithm and SNR = 50, the remaining five algorithms provided accurate PDFF estimation with similar PDFF bias (bias < 0.15%, LOA < 2.6%) (Figure 2A). Increasing the number of TE significantly decreased ($p < 0.0001$) the LOA for all algorithms. At NTE = 5 and minimal echo spacing, five algorithms provided a low R_2^* mean bias ($<0.5\ s^{-1}$), whereas MSGCA-A suffered at low PDFF (<35%) from a large bias ($-48\ s^{-1}$) at high R_2^* ($\geq 150 s^{-1}$) (Figure 2B). These five algorithms achieved a LOA within 15% of targeted value, with the lowest precision for PDFF = 63%. Computation times for processing one repetition with NTE = 7 ranged from $T_{\text{B0-NICE}} = 1.8 \pm 0.2$ s, $T_{\text{IDEAL-CE}} = 1.95 \pm 0.02$ s, to $T_{\text{Fatty-Riot-GC}} = 51.3 \pm 18.0$ s and $T_{\text{GOOSE}} = 5455 \pm 7122$ s.

3.2 | In vitro experiments

The local experiments on our custom phantom allowed to validate results observed in the numerical study. Using the NTE = 9 results, B_0 off-resonance in the phantom was

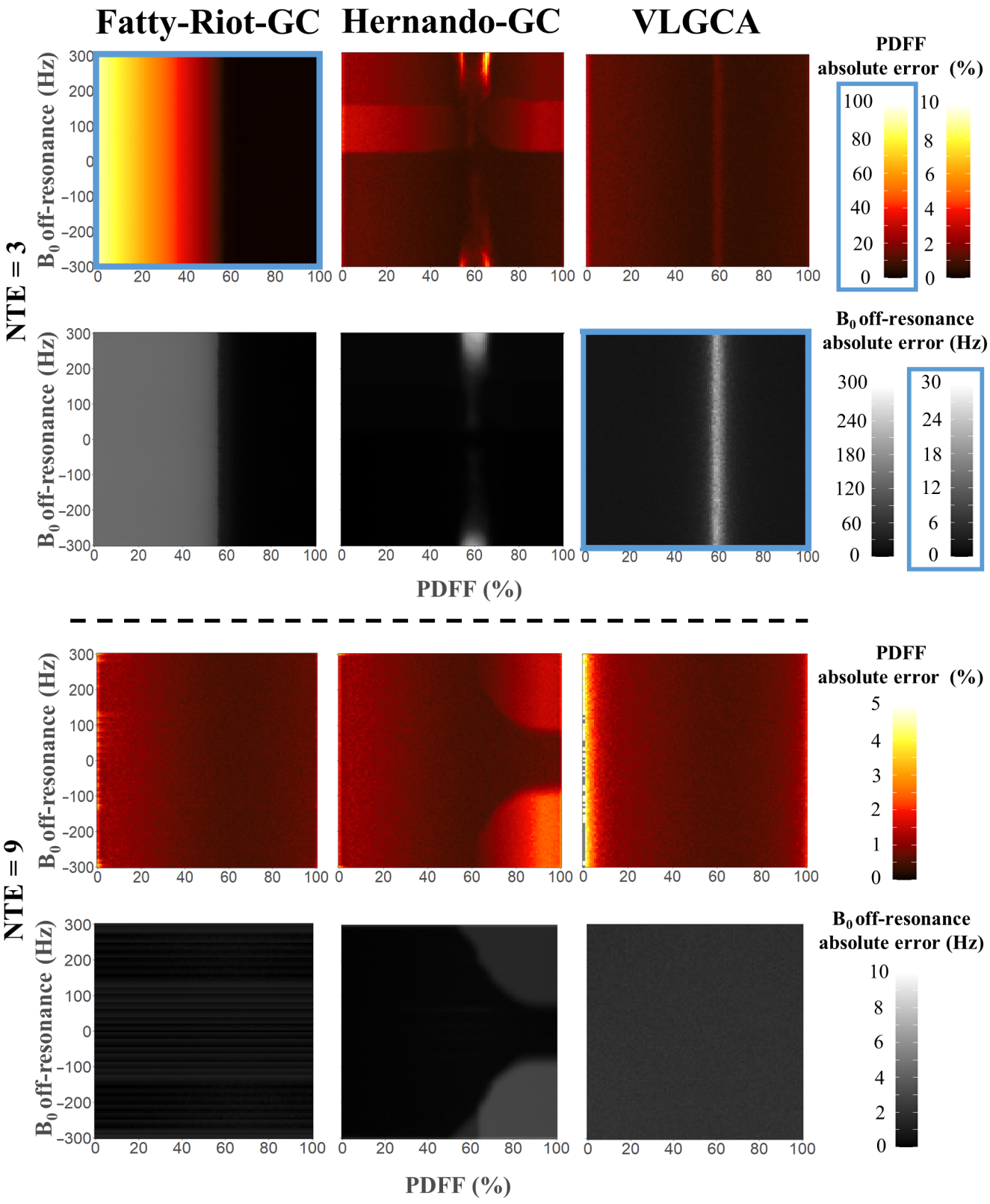


FIGURE 1 PDFF and B_0 off-resonance absolute error was measured on three algorithms with sensitivity to B_0 inhomogeneity: Fatty-Riot-GC, Hernando-GC, and VLGCA, using synthetic CSE-MRI volumes with SNR = 50 at three and nine echoes in-/out of phase echo spacing. PDFF and B_0 absolute error maps were averaged along the repetition axis. Bias in B_0 off-resonance field map estimation led to FWswaps with Fatty-Riot-GC, while Hernando-GC suffered from PDFF bias due to B_0 inhomogeneities. With nine echoes, algorithms provided more reliable quantitative maps. Maps with specific scales are displayed in blue boxes.

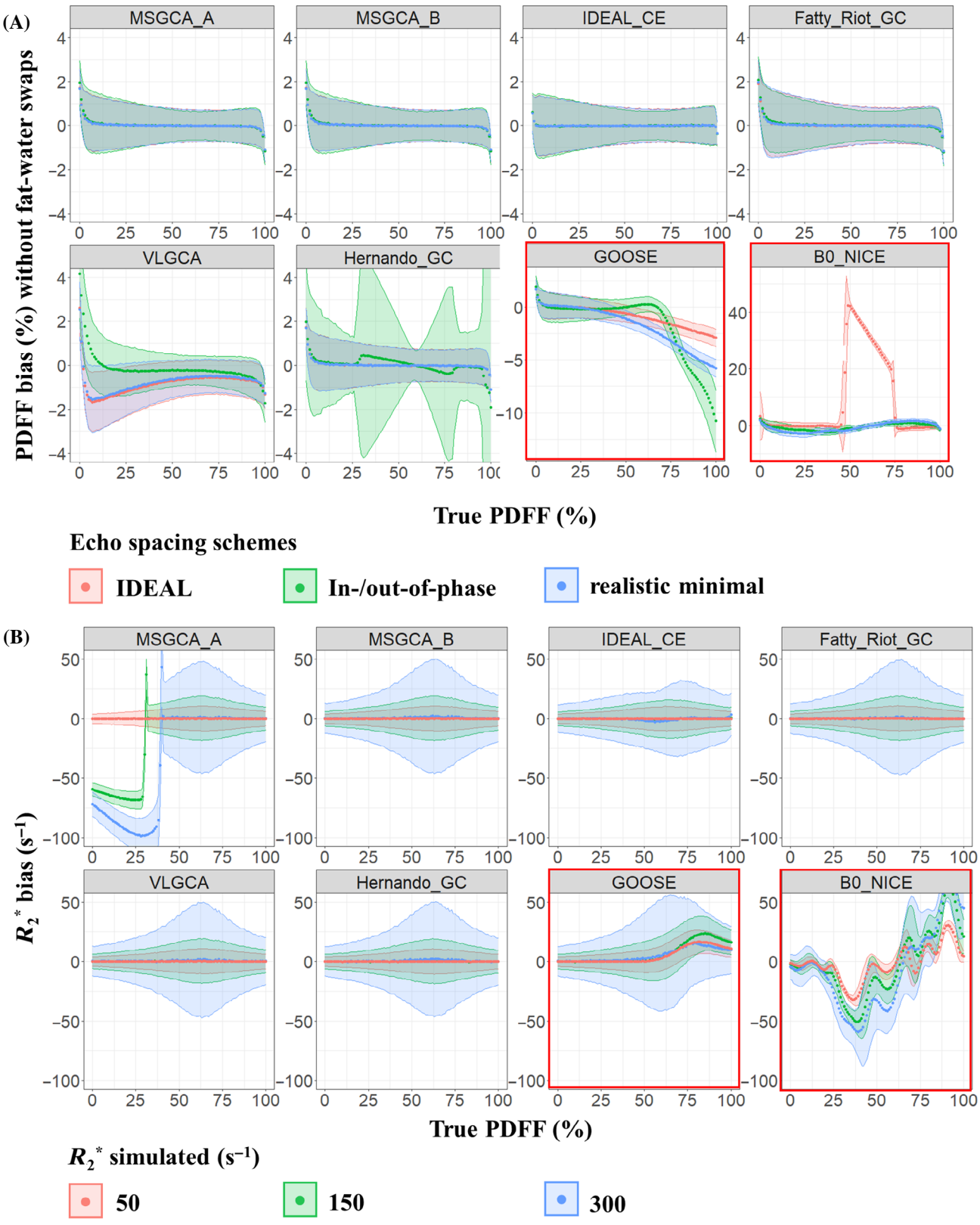


FIGURE 2 Comparison of PDFF (A) and R_2^* (B) bias of each algorithm over synthetic CSE-MRI volumes with NTE = 7 and SNR = 100 (A) or with NTE = 5, minimal echo spacing, and SNR = 50 (B). Mean and SD PDFF and R_2^* bias were averaged along the B_0 off-resonance and repetition axes and separated (in color) according to the echo spacing schemes (A) or simulated R_2^* (B). GOOSE and B0-NICE (in red square) were not further investigated due to highly biased results.

found to span more than 300 Hz, from −220 to 150 Hz. Matching simulation results, VLGCA confirmed to have PDFF bias superior to 4% in the water vial, Hernando-GC in the B_0 off-resonance vials range shown in Figure 3A provide reliable PDFF measurement whereas with NTE = 3, Fatty-Riot-GC was still influenced by B_0 inhomogeneity resulting in a PDFF bias of 10% in the water vial instead of FWswaps. Confirming simulation results, algorithms proved to be robust to FWswaps (≥0.05%) with NTE ≥ 7. For PDFF quantification, all algorithms were independent to the echo spacing with NTE = 7 and provided a mean bias under 2.5% (Figure 3B). However, all algorithms overestimated the 40 and 60% vials. These vials were found to have very different water and fat R_2^* values ($R^*_{2fat} = 22$ and 25 s^{-1}, $R^*_{2water} = 47$ and 49 s^{-1} for 40 and 60%, respectively). Algorithms in our study assumed a single R_2^* value, and this hypothesis failed in these experiments. Numerical simulation of simulated data matching these characteristics confirmed a +3% bias from algorithms within the 40%–60% PDFF range (Figure S4). R_2^* absolute error corroborated numerical simulations (Figure S5) with a mean absolute error higher than 10 s^{-1} for B0-NICE, whereas it is lower than 4.5 s^{-1} for the seven other algorithms.

3.3 | In-vitro: multi-site fat-water phantom MRI data

Algorithms performances were dependent on the multi-echo acquisition type (monopolar, interleaved, bipolar). Using interleaved echoes acquisition (sites 3 and 4, both 1.5T), all algorithms failed to provide accurate PDFF quantification (Figure 4 and Table S1) with a mean absolute error superior to 7% for IDEAL-CE and to 12% for the other algorithms. Thus, the issue lies in the double acquisition and is independent of the algorithms. For bipolar acquisitions, even after correction for odd and even echoes (sites 5 and 6), IDEAL-CE suffered from FWswaps, with mean FWswaps of 34.5% and a mean bias of 23.7% for non-swapped pixels. Fatty Riot also exhibited swaps, with FWswaps superior to 15% for sites 2 and 3. The four other algorithms provided reliable PDFF measurement across sites, protocols, and field strength with a mean bias less than 2%. However, their PDFF measurements were significantly dependent, albeit with small biases, on the field strength (max 1.2% bias, $F = 5.7$, $p < 0.05$), vendor system (max 1.5% bias, $F = 4.2$, $p < 0.05$), and site (max 1.3% bias, $F = 3.8$, $p < 0.01$).

3.4 | In vivo experiments

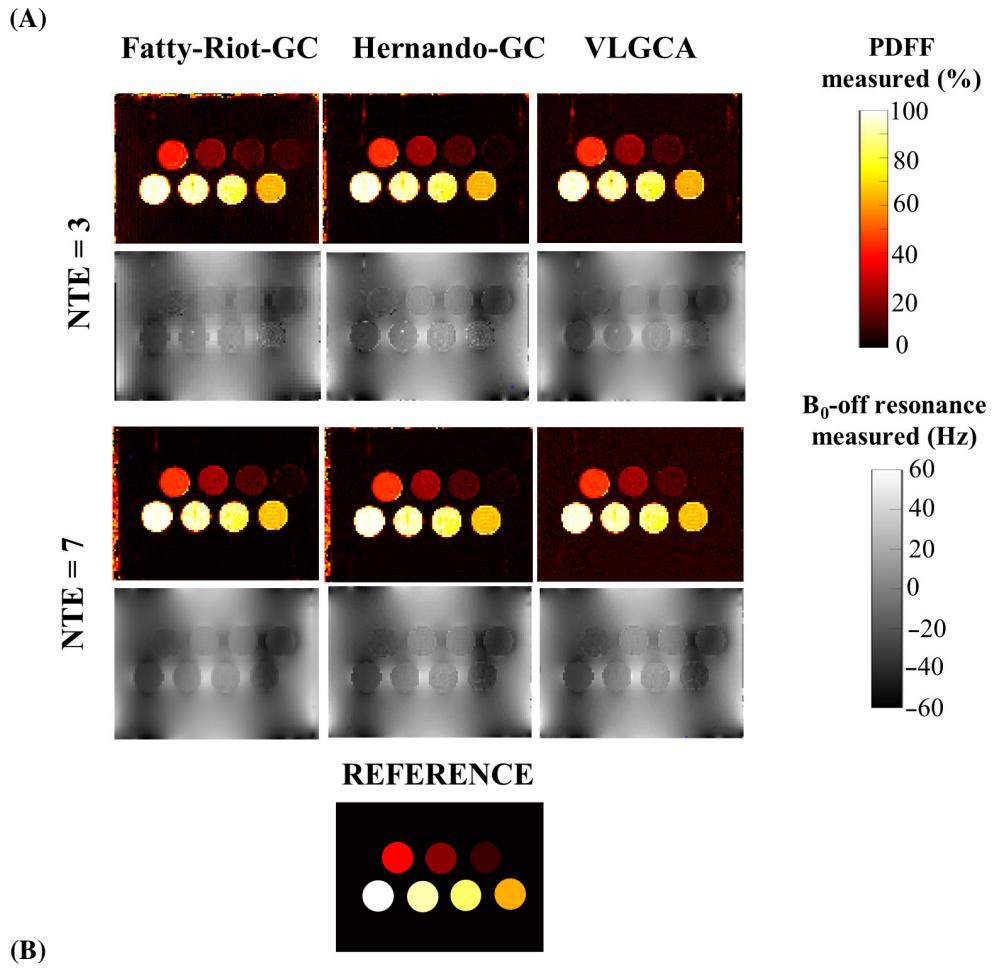
Large B_0 inhomogeneity variations were measured for the in vivo experiments, with frequency offsets up to $\Delta B_0 = 500$ and 350 Hz for supraclavicular and sacrum regions, respectively. From these challenging datasets (Figure 5), algorithms that were comparable in silico (IDEAL-CE, MSGCA-B, and Fatty-Riot-GC) demonstrated significant PDFF discrepancies in vivo: graph-cuts algorithms were misled by the strong B_0 gradient and resulted in FWswaps and B_0-dependent PDFF values. Eventually, apart from IDEAL-CE, most algorithms suffered from a spatial bias suggesting fat content in water tissues or fluids (as in the bladder: $PDFF_{IDEAL\text{-}CE} = 0.5 \pm 0.8\%$ vs. $PDFF_{MSGCA\text{-}A/MSGCA\text{-}B/Hernando\text{-}GC} = 18.2 \pm 10.2\%$ and $PDFF_{Fatty\text{-}Riot\text{-}GC=}90.9 \pm 2.7\%$, $PDFF_{VLGCA} = 91.4 \pm 2.7\%$). Interestingly, for most algorithms these spatial biases did not impact PDFF measurements in actual adipose tissues (as in the sacrum bone marrow: $PDFF_{IDEAL\text{-}CE} = 45.5 \pm 6.1\%$, $PDFF_{MSGCA\text{-}A} = 45.6 \pm 6.1\%$, $PDFF_{MSGCA\text{-}B} = 45.5 \pm 6.0\%$, $PDFF_{Hernando\text{-}GC} = 44.9 \pm 6.2\%$) but FWswaps led to elevated PDFF in 2 algorithms ($PDFF_{Fatty\text{-}Riot\text{-}GC} = 69.1 \pm 7.7\%$, $PDFF_{VLGCA} = 69.6 \pm 4.1\%$). For highly accelerated liver CSE-MRI, the challenge lied in its low SNR = 14, and the three algorithms shown provided similar performances with moderate parametric SDs in the liver ($SD_{IDEAL\text{-}CE} = 2.45\%/10.26\ s^{-1}$, $SD_{MSGCA\text{-}A=}$ $2.46\%/10.25\ s^{-1}$, $SD_{Fatty\text{-}Riot\text{-}GC=}$ $2.35\%/10.25\ s^{-1}$) for PDFF and R_2^*, respectively, confirming equivalent noise propagation in these algorithms.

3.5 | Influence of the spectrum model

From the synthetic magnitude CSE signals, it was confirmed that R_2^* decay might depend on the chosen fat spectrum model (Figure 6A). Processing synthetic data with a different spectrum (ISMRM 2012 challenge) than the one employed for simulation (peanut oil) revealed a significant bias in PDFF quantification ($p < 0.0001$) and R_2^* ($p < 0.0001$) with a maximum bias of 2.35% and 20.70 s^{-1}, respectively, at NTE = 9 and SNR = 100 (Figure 6). In vitro (Figure 6B) and in vivo (Figure 6C) experiments confirmed similar biases in practice, with mean PDFF and R_2^* differences of 1.22% and 12.22 s^{-1}, respectively, in supra-clavicular adipose tissue.

4 | DISCUSSION

In keeping with the community-driven standardization of MRI body fat and iron quantification, several recent fat-water separation algorithms were compared through an open-source toolbox for their reproducibility, precision, and accuracy. This benchmarking also included the influence of acquisition parameters (number of echoes



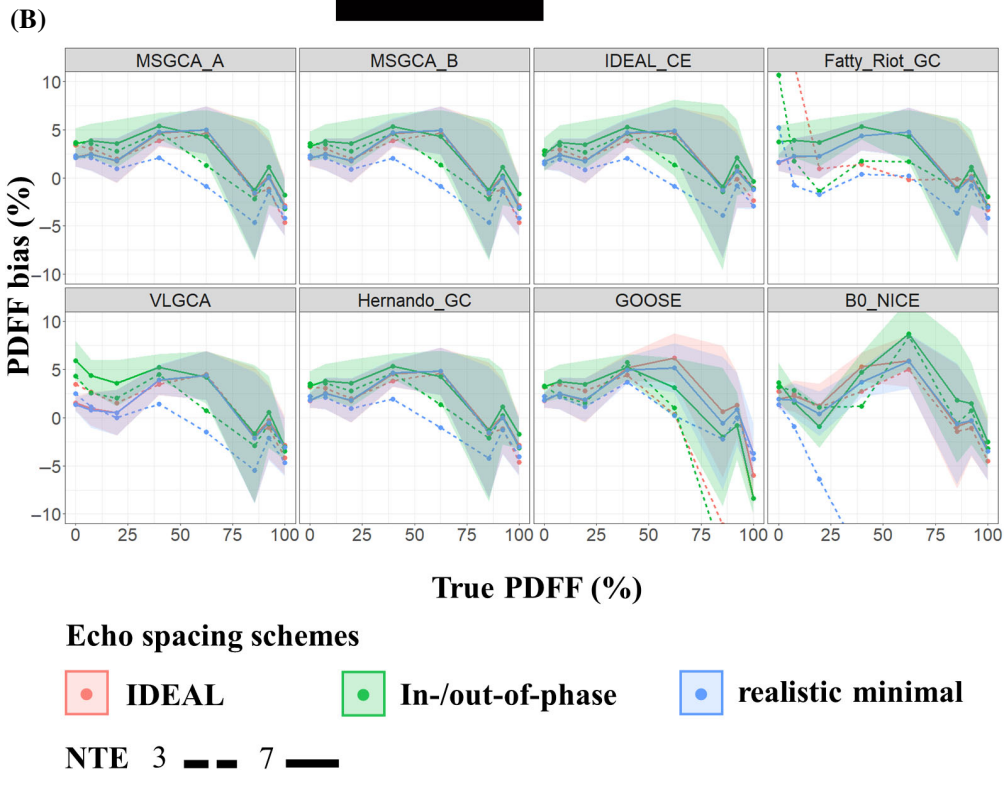


FIGURE 3 (A) PDFF measurement from in vitro experiments using three algorithms: Fatty-Riot-GC, Hernando-GC, and VLGCA with NTE = 3 and NTE = 7 and in-/out-of-phase echo spacing. (B) Comparison of PDFF bias of each algorithm in phantoms. Mean and SD PDFF bias were averaged plotted according to the echo spacing schemes (in color) and echo number (in line style). For clarity, only SD of PDFF bias for NTE = 7 have been plotted.

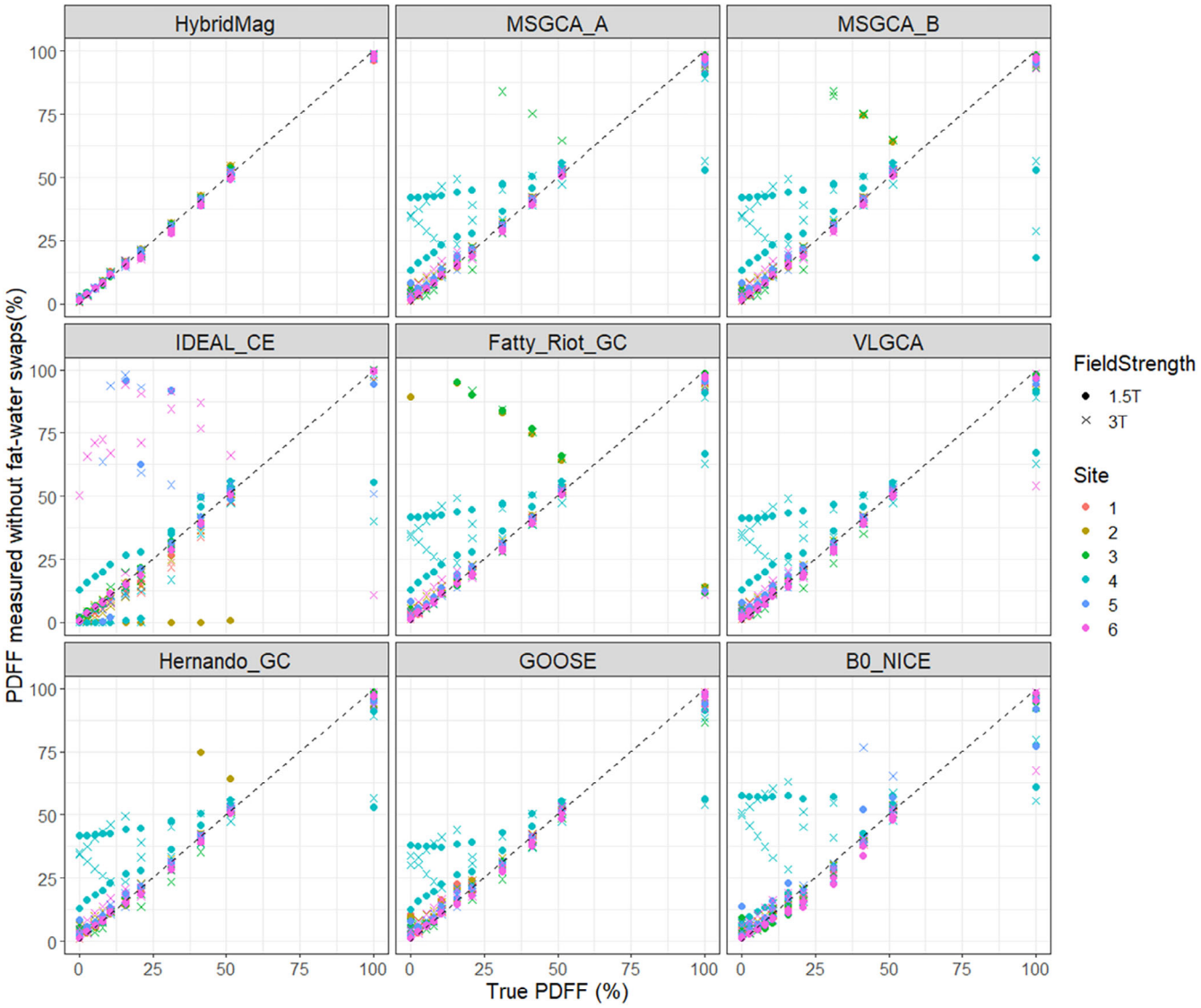


FIGURE 4 Linear regression analysis on multi-site fat-water phantom data. The targeted linearity response between true PDFF and estimated PDFF is represented in a black dotted line. Estimation errors correspond to deviations from this line. The HybridMag algorithm initially included in the original study was included as a reference. Sites 3 and 4 employed interleaved echoes schemes, which impacted algorithms quantification. Sites 5 and 6 employed bipolar readout gradients that impacted only IDEAL-CE quantification despite a preprocessing phase correction.

and echo spacing) to obtain more accurate quantitative maps.

4.1 | Open-source framework and reproducibility research

The proposed toolbox was developed in both Python and MATLAB programming languages to facilitate benchmarking of fat-water separation algorithms. It can be considered as an extension of the ISMRM fat-water toolbox[19] that currently only allows to compare MATLAB algorithms. Another addition is the capability to operate algorithms with various fat spectra from an exhaustive fat spectra library. This framework was made open-source (https://github.com/pdaude/CREAM_PDFF) to facilitate comparison with new upcoming methods.

This study did not aim to elect the optimal fat-water separation algorithm but rather to provide a practical evaluation framework for helping researcher in the selection of a fat-water separation algorithm. The choice for a specific algorithm remains multi-factorial and this study aims to set basis for further reproduction and adaption. First, the targeted application has to be defined including choice of bio-physical model (fat spectrum calibration), expected range of quantitative parameters (PDFF,

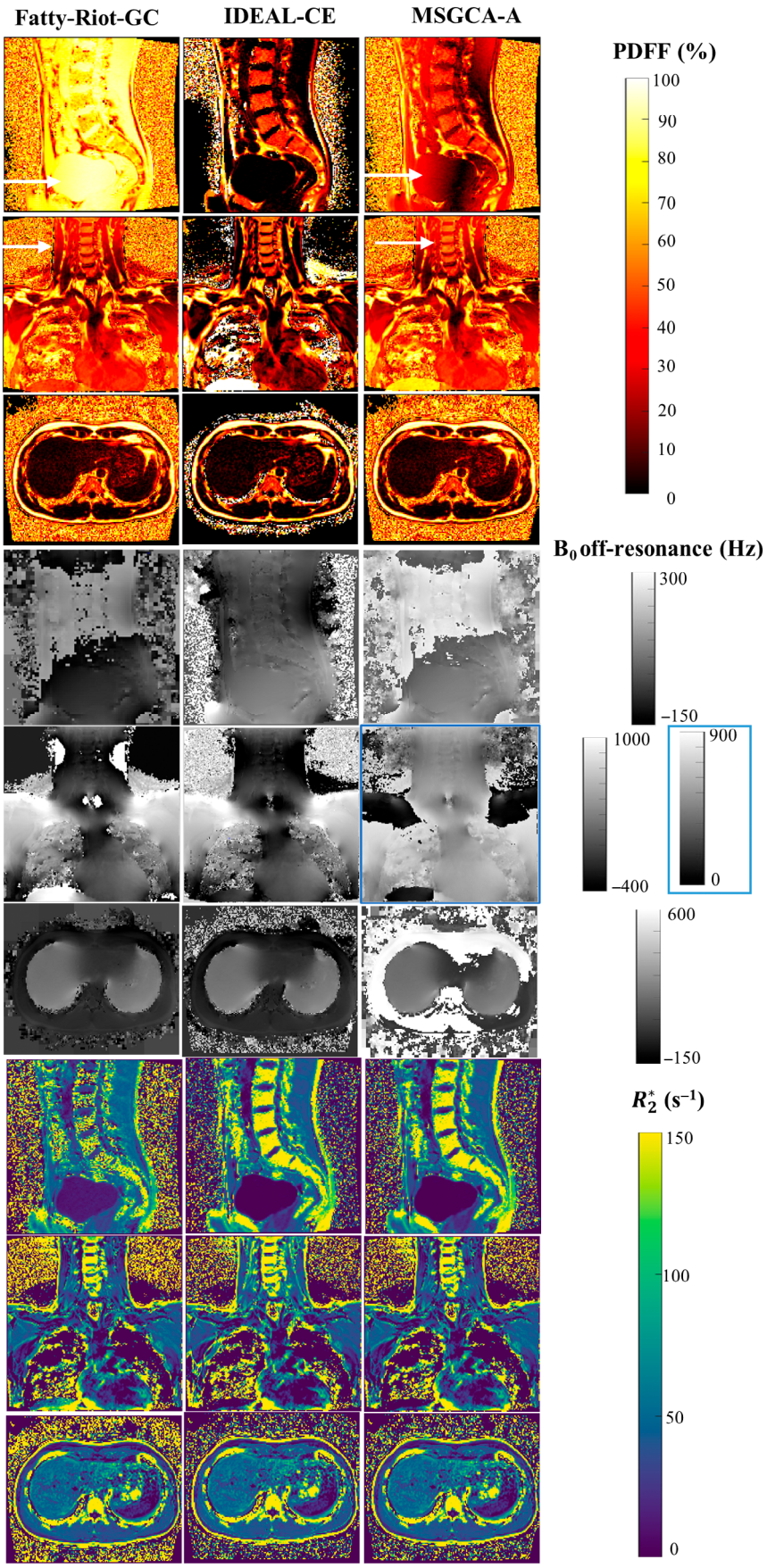


FIGURE 5 PDFF, R_2^*, and B_0 off-resonance quantification of three algorithms of equivalent performances so far: Fatty-Riot-GC, IDEAL-CE, and MSGGA-B over challenging in vivo datasets at 3T. PDFF overestimation or complete FWswaps were observed where large B_0 off-resonances were present. Using MSGCA-B and Fatty-Riot-GC as shown by white arrows, fat was erroneously quantified in regions where it was not expected such as inside the bladder or the neck muscles. Fatty-Riot-GC even led to a full volume FWswap in the pelvis due to a large off-resonance range.

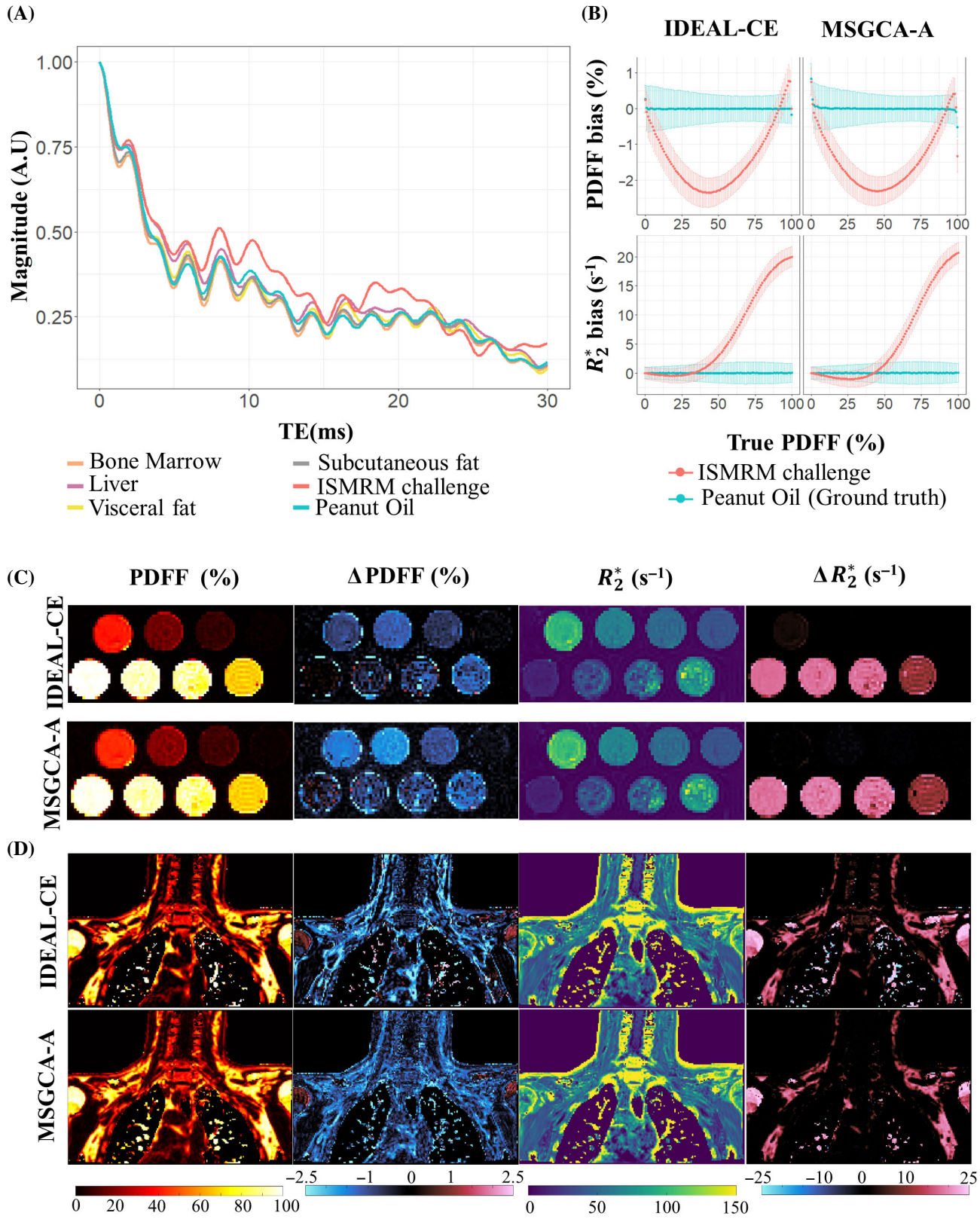


FIGURE 6 Influence of fat spectrum models on PDFF and R_2^* quantification with either an IDEAL algorithm (IDEAL-CE) or a graph-cut algorithm (MSGCA-A). (A) Synthetic magnitude CSE-MRI signal at PDFF 100% and T_2^*=20 ms with different spectrum depicted in the literature or acquired for this study. (B) PDFF and R_2^* bias when synthetic CSE-volume are processed with either the same spectrum (peanut oil) or with the ISMRM 2012 challenge one at 3T with nine echoes and realistic minimal echo-spacing. PDFF, R_2^* measurement when in vitro (C) or in vivo (D) experiments were processed with peanut oil spectrum at 3T with nine echoes with a minimal echo spacing scheme and the difference ΔPDFF, ΔR_2^* resulted when those data were processed with the ISMRM 2012 challenge spectrum.

R_2^*, B_0), SNR, acquisition constraints (sampling scheme and number of echoes), and required precision. For example, while algorithms accept different TE spacing, results show that some performances depended on this TE spacing (such as VLGCA and Hernando-GC). Second, the multi-scale evaluation appeared impactful in our study: when in silico results pointed toward three equivalent algorithms (MSGCA-B, IDEAL-CE, and Fatty-Riot-GC), in vitro results confirmed it and penalized one of them (confirmation with our custom phantom and penalization of IDEAL-CE over the multi-site dataset), whereas one algorithm (IDEAL-CE) stood out in in vivo challenges. Third, algorithms can be individually optimized for one application or one computing architecture. Proposed toolbox should facilitate such optimization leading to improved performances of certain algorithms that were implemented "as provided" in our study. Fourth, the computational time can impact the usability of the algorithms and might enter considerations for optimal choice. Eventually, this toolbox also estimates margins of errors of parameter quantifications which could help to further discriminate algorithms. These estimates can also further be compared to Cramér-Rao lower bounds as provided by Diefenbach et al.[20]

4.2 | Numerical simulations

Provided with synthesized data using only three echoes, most algorithms suffered from FWswaps or PDFF bias due to B_0 inhomogeneities (Figures 1 and S2) while five or seven echoes provided a significant improvement in accuracy and precision in PDFF measurement with almost no FWswaps(<1%). With NTE≥5, B_0 offset resonance field maps were correctly estimated by all algorithms apart from B0-NICE and GOOSE. Even when using five echoes, the PDFF precision for VLGCA, B0-NICE and GOOSE remained greatly dependent on the echo spacing (i.e., different echo sampling schemes). In, Hernando-GC with in-/out of phase echo spacing suffered from random failures of B_0 field map estimation over the repetition axes, leading to large PDFF SDs (Figure 2A). Eventually, for all other cases, MSGCA-A, MSGCA-B, IDEAL-CE, Hernando-GC, and Fatty-Riot-GC provided similar results, all suitable for a reliable PDFF quantification (Figure 2A).

Considering these latter five algorithms, echo spacing still influenced the precision of the R_2^* quantification (Figure S6). Considering moderate R_2^* values and a fixed number of TEs, R_2^* precision depended on the longest sampled TE. This conclusion does not, however, account for high R_2^* values that might occur in the presence of liver iron accumulation or bone marrow. Therefore, to quantify moderate R_2^* with a given number of echoes, in-/out of phase schemes should be preferred over IDEAL or minimal echo spacing. Nevertheless, with a fixed TR, minimal echo spacing, potentially allowing fitting more echoes, should be preferred to the other echo spacing schemes for R_2^* accuracy. Tighter echo spacing (i.e. "realistic minimal" echo spacing), associated with higher signal and more sampled echoes within the same TR, improves noise performances at the acquisition level. When evaluated over a realistic range of R_2^*, five algorithms (VLGCA, Hernando-GC, Fatty-Riot-GC, IDEAL-CE, and MSGCA-B) provided reliable quantification with low R_2^* bias ($<0.5\,s^{-1}$) (Figure 2B). For computation efficiency, MSGCA-A is implemented using only the decoupled estimation of B_0 and R_2^* maps, resulting in FWswaps and R_2^* bias if the assumed R_2^* used in the estimation of B_0 map is too far from the ground truth.

4.3 | In vitro experiments

The custom in vitro experiments enabled to validate the numerical findings. Notably, experimental results confirmed the necessity to acquire at least five echoes to avoid FWswaps and bias due to B_0 inhomogeneities with most algorithms (Figure 3). It also showed that algorithms provided a bias in the PDFF quantification which is lower than the repeatability coefficient (for NTE = 7, $PDFF_{bias} \leq 2.4\%$ vs. $PDFF_{repeatibility} \leq 2.99\%$). Using spectroscopy as a reference measurement, PDFF bias was higher than in the corresponding simulation for all algorithms especially for vials with 40% and 60% fat. This discrepancy was assumed to be due to the model assumption of a common single R_2^* value for fat and water when a dual R_2^* would have been more appropriate.[58] A numerical simulation with R_2^* values obtained from spectroscopy measurements confirmed this discrepancy (Figure S4). As most fat-water algorithms are built on this assumption and dual R_2^* modeling is inherently more sensitive to noise, separate R_2^* have not been investigated in this study, albeit they would be a valuable extension to the present study.

4.4 | In-vitro: multi-site fat-water phantom MRI data

Analyzing multi-site fat-water phantoms required a pre-processing step for correcting interleaved and bipolar echoes,[59,60] otherwise fat-water algorithms failed to provide realistic quantitative maps (Figure 4 and Table S1). In general, corrections for gradients distortions during acquisition, including imperfect gradient system transfer

function,[61] gradients spatial non-linearity and concomitant gradients[59] during echo trains, correspond to an image processing step performed prior to fat-water separation.[62] Thus, these corrections were not considered in this study. As suggested in the original study,[16] the PDFF biases found across sites could be due to the variability of temperature which might affect PDFF quantification.[49,63] Eventually, Fatty-Riot-GC and IDEAL-CE exhibited more FWswaps and bias compared to custom in vitro experiments.

4.5 | In vivo experiments

As observed in phantoms, challenging in vivo data revealed disparities between algorithms (IDEAL-CE, MSGCA-B, Fatty-Riot-GC). These differences may be due to rapid and large variations of B_0, which breaks the stringent constraint of a smooth field map assumed by these algorithms (Figure 5). However, algorithms provided similar performances and showed to be highly resilient to challenging low SNR in vivo data (SNR = 14), with PDFF and R_2^* measurements SD in liver of 2.4% and 11 s^{-1}, which was verified by simulation (Figure S7).

4.6 | Influence of spectrum model

Interestingly, processing data with an inadequate spectrum led to a non-negligible bias in PDFF and R_2^* (Figure 6). The in vitro and in vivo experiments confirmed this difference, as shown in Figure 6C,D. Thus, the choice of a relevant spectrum remains essential for characterizing fat deposits with different fatty acid composition. This consideration might even be more important in applications such as CSE-MRI of bone marrow. The sensitivity of PDFF quantification alone to multi peak fat spectrum calibration has been previously explored in the liver on non-alcoholic steatohepatitis patients[64] or using synthetic CSE data with low PDFF (≤40%),[45] six echoes and a graph-cut algorithm. Within this range, the choice of the number of spectrum peaks was found not to significantly impact PDFF. Our results extend these findings, demonstrating that with more echoes (NTE = 9) and a high SNR a small PDFF bias (<2%) can arise within the PDFF range of 20–80%, depending on the selected fat spectra which is however lower than the repeatability coefficient ($PDFF_{repeatability}$ ≤2.99%). But more importantly, the choice of spectra significantly influenced the R_2^* bias (12.22 s^{-1}), which might alter the evaluation of iron content in organs such as liver or bone marrow. R_2^* enables the measurement of liver iron concentration (LIC)[65–67] and the R_2^*-LIC calibration has been studied[68,69] across field strengths, centers, and vendors, but the variation in fat composition could also stand as an interesting element to consider. From our results, the fat-composition R_2^* bias would correspond to a variation of 0.17 mg/g ($LIC_{2.89T} = -0.09 + 1.387 * 10^{-2}\ R_2^*$).

4.7 | Algorithms running time

Apart from GOOSE, the algorithms' running times on our computer were within seconds to minutes of processing time per slice. Considering the limited computational optimization of open-source implementations (in MATLAB or Python), the eight remaining algorithms are all potentially suitable for a clinical routine PDFF quantification upon code optimization. We noted that algorithms developed after the ISMRM 2012 challenge (MSGCA-A/B, IDEAL-CE, B0-NICE, VLGCA) were faster than the earlier ones (Hernando-GC, Fatty-Riot-GC).

4.8 | Choices of open-source algorithms

The list of selected open-source algorithms was arguably limited in our comparison study. Indeed, new approaches for solving the fat-water separation problem based on deep learning have recently emerged.[33,34,70] However, all currently available deep-learning algorithms were based on a fixed number of echoes or would need a specific training strategy and network modification to be compatible with our benchmark. Additionally, all algorithms in this study are based on the complex signal model, and it would be interesting to include algorithms using only the magnitude signal model, or hybrid methods, which have been developed to be more robust to phase errors and circumvent the field map estimation. However, to our knowledge, there were no open-source algorithms of this type documented in the literature at the time of this study. Finally, it would have also been of interest to compare the selected algorithms with commercial software which, in parts, are based on complex or hybrid magnitude-complex method. But the access to such methods remains limited, if not even possible.

4.9 | Possible extensions and new challenges

Our open-source toolbox was designed to be upgradeable to tackle new challenges, leaving room for multiple extensions that were beyond the scope of this study. Algorithms extensions based on external B_0 field map initialisation are of interest, such as methods incorporating a priori information from the scanner magnetic

field distribution.[62,71] In a prolonged perspective, some algorithms have started to include refined complex signal MR models designed for quantitative fatty acid composition parameter mapping.[37,38] A standard evaluation of such advanced algorithms performances could hold interest. Finally, a graphical user interface for this framework could also benefit from a user's perspective as it has already been done for QSM processing pipelines with the SEPIA[72] package.

5 | CONCLUSIONS

In accordance with standardization of MRI body fat and iron quantification, an open-source bi-language toolbox was developed to evaluate eight state-of-the-art open-source algorithms for fat-water separation. Leveraging this toolbox, a multi-scale evaluation of algorithms was demonstrated: first, algorithms' performances differed in silico, on numerical synthetic data; second, results were matched with in vitro experimental results from a custom phantom; third, complementary evaluation on multi-site phantom data probed algorithms resilience to various datasets. Finally, challenging in vivo datasets illustrated certain algorithms' failure cases. This framework sets basis for continued comparison of algorithms for fat-water separation and subsequent quantitative MRI as developments propose new avenues for refined adipose tissue characterization.

FUNDING INFORMATION

This project has received financial support from the CNRS through the MITI program and was performed within a laboratory member of France Life Imaging network (grant ANR-11-INBS-0006).

CONFLICT OF INTEREST STATEMENT

Thomas Troalen is an employee of Siemens Healthcare.

DATA AVAILABILITY STATEMENT

Data are available from the corresponding author on reasonable request. Code is freely available for download at https://github.com/pdaude/CREAM_PDFF.

ORCID

Pierre Daudé https://orcid.org/0000-0002-0565-8974
Tangi Roussel https://orcid.org/0000-0002-2389-2027
Thomas Troalen https://orcid.org/0000-0002-5764-4689
Diego Hernando https://orcid.org/0000-0002-0016-0317
Stanislas Rapacchi https://orcid.org/0000-0002-8925-495X

REFERENCES

1. Dixon WT. Simple proton spectroscopic imaging. *Radiology*. 1984;153:189-194. doi:10.1148/radiology.153.1.6089263
2. Reeder SB, Hu HH, Sirlin CB. Proton density fat-fraction: a standardized mr-based biomarker of tissue fat concentration. *J Magn Reson Imaging*. 2012;36:1011-1014. doi:10.1002/jmri.23741
3. Starekova J, Reeder SB. Liver fat quantification: where do we stand? *Abdom Radiol*. 2020;45:3386-3399. doi:10.1007/s00261-020-02783-1
4. Bainbridge A, Bray TJP, Sengupta R, Hall-Craggs MA. Practical approaches to bone marrow fat fraction quantification across magnetic resonance imaging platforms. *J Magn Reson Imaging*. 2020;52:298-306. doi:10.1002/jmri.27039
5. Idilman IS, Yildiz AE, Karaosmanoglu AD, Ozmen MN, Akata D, Karcaaltincaba M. Proton density fat fraction: magnetic resonance imaging applications beyond the liver. *Diagn Interv Radiol*. 2022;28:83-91. doi:10.5152/dir.2021.21845
6. Wood JC. Guidelines for quantifying iron overload. *Hematology*. 2014;2014:210-215. doi:10.1182/asheducation-2014.1.210
7. Triadyaksa P, Oudkerk M, Sijens PE. Cardiac T_2 * mapping: techniques and clinical applications. *J Magn Reson Imaging*. 2020;52:1340-1351. doi:10.1002/jmri.27023
8. Otto LAM, Froeling M, Eijk RPA, et al. Quantification of disease progression in spinal muscular atrophy with muscle MRI—a pilot study. *NMR Biomed*. 2021;34:e4473. doi:10.1002/nbm.4473
9. Ghosh A, Dutta K, Bhatt SP, et al. Dapagliflozin improves body fat patterning, and hepatic and pancreatic fat in patients with type 2 diabetes in North India. *J Clin Endocrinol Metabol*. 2022;107:e2267-e2275. doi:10.1210/clinem/dgac138
10. Lavynenko O, Abdul-Ghani M, Alatrach M, et al. Combination therapy with pioglitazone/exenatide/metformin reduces the prevalence of hepatic fibrosis and steatosis: the efficacy and durability of initial combination therapy for type 2 diabetes (EDICT). *Diabetes Obesity Metabolism*. 2022;24:899-907. doi:10.1111/dom.14650
11. Franz D, Diefenbach MN, Treibel F, et al. Differentiating supraclavicular from gluteal adipose tissue based on simultaneous PDFF and T_2 * mapping using a 20-echo gradient-echo acquisition: PDFF and T_2 * mapping using 20-echo GRE. *J Magn Reson Imaging*. 2019;50:424-434. doi:10.1002/jmri.26661
12. Hu HH, Hines CDG, Smith DL, Reeder SB. Variations in T2* and fat content of murine brown and white adipose tissues by chemical-shift MRI. *Magn Reson Imaging*. 2012;30:323-329. doi:10.1016/j.mri.2011.12.004
13. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology standards for quantitative imaging biomarkers. *Radiology*. 2015;277:813-825. doi:10.1148/radiol.2015142202
14. Weingärtner S, Desmond KL, Obuchowski NA, et al. Development, validation, qualification, and dissemination of quantitative MR methods: overview and recommendations by the ISMRM quantitative MR study group. *Magn Reson Med*. 2021;87:1184-1206. doi:10.1002/mrm.29084
15. Yokoo T, Serai SD, Pirasteh A, et al. Linearity, bias, and precision of hepatic proton density fat fraction measurements by using MR imaging: a meta-analysis. *Radiology*. 2018;286:486-498. doi:10.1148/radiol.2017170550
16. Hernando D, Sharma SD, Aliyari Ghasabeh M, et al. Multisite, multivendor validation of the accuracy and reproducibility of proton-density fat-fraction quantification at 1.5T and 3T using

a fat-water phantom: proton-density fat-fraction quantification at 1.5T and 3T. *Magn Reson Med*. 2017;77:1516-1524. doi:10.1002/mrm.26228
17. Hu HH, Yokoo T, Bashir MR, et al. Linearity and bias of proton density fat fraction as a quantitative imaging biomarker: a multicenter, multiplatform, multivendor phantom study. *Radiology*. 2021;298:640-651. doi:10.1148/radiol.2021202912
18. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533:452-454. doi:10.1038/533452a
19. Hu HH, Börnert P, Hernando D, et al. ISMRM workshop on fat-water separation: insights, applications and progress in MRI. *Magn Res Med*. 2012;68:378-388. doi:10.1002/mrm.24369
20. Diefenbach MN, Liu C, Karampinos DC. Generalized parameter estimation in multi-echo gradient-echo-based chemical species separation. *Quant Imaging Med Surg*. 2020;10:554-567. doi:10.21037/qims.2020.02.07
21. Bray TJP, Bainbridge A, Lim E, Hall-Craggs MA, Zhang H. MAGORINO: magnitude-only fat fraction and R2* estimation with Rician noise modeling. *Magn Res Med*. 2023;89:1173-1192. doi:10.1002/mrm.29493
22. Yu H, McKenzie CA, Shimakawa A, et al. Multiecho reconstruction for simultaneous water-fat decomposition and T2* estimation. *J Magn Reson Imaging*. 2007;26:1153-1161. doi:10.1002/jmri.21090
23. Yu H, Shimakawa A, McKenzie CA, Brodsky E, Brittain JH, Reeder SB. Multiecho water-fat separation and simultaneous R2* estimation with multifrequency fat spectrum modeling. *Magn Reson Med*. 2008;60:1122-1134. doi:10.1002/mrm.21737
24. Andersson J, Ahlström H, Kullberg J. Water-fat separation incorporating spatial smoothing is robust to noise. *Magn Reson Imaging*. 2018;50:78-83. doi:10.1016/j.mri.2018.03.015
25. 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*. 2009;63:79-90. doi:10.1002/mrm.22177
26. Berglund J, Kullberg J. Three-dimensional water/fat separation and T2* estimation based on whole-image optimization-application in breathhold liver imaging at 1.5 T. *Magn Res Med*. 2012;67:1684-1693. doi:10.1002/mrm.23185
27. Berglund J, Skorpil M. Multi-scale graph-cut algorithm for efficient water-fat separation: multi-scale graph-cut water/fat separation. *Magn Reson Med*. 2017;78:941-949. doi:10.1002/mrm.26479
28. Cui C, Wu X, Newell JD, Jacob M. Fat water decomposition using globally optimal surface estimation (GOOSE) algorithm. *Magn Reson Med*. 2015;73:1289-1299. doi:10.1002/mrm.25193
29. Boehm C, Diefenbach MN, Makowski MR, Karampinos DC. Improved body quantitative susceptibility mapping by using a variable-layer single-min-cut graph-cut for field-mapping. *Magn Reson Med*. 2021;85:1697-1712. doi:10.1002/mrm.28515
30. Peng H, Zou C, Cheng C, et al. Fat-water separation based on transition REgion extraction (TREE). *Magn Reson Med*. 2019;82:436-448. doi:10.1002/mrm.27710
31. Cheng C, Zou C, Liang C, Liu X, Zheng H. Fat-water separation using a region-growing algorithm with self-feeding phasor estimation: self-feeding phasor estimation for fat-water separation. *Magn Reson Med*. 2017;77:2390-2401. doi:10.1002/mrm.26297
32. Liu J, Drangova M. Method for B0 off-resonance mapping by non-iterative correction of phase-errors (B0-NICE): B0 mapping with multiecho data. *Magn Reson Med*. 2015;74:1177-1188. doi:10.1002/mrm.25497
33. Liu K, Li X, Li Z, et al. Robust water–fat separation based on deep learning model exploring multi-echo nature of mGRE. *Magn Reson Med*. 2021;85:2828-2841. doi:10.1002/mrm.28586
34. Jafari R, Spincemaille P, Zhang J, et al. Deep neural network for water/fat separation: supervised training, unsupervised training, and no training. *Magn Reson Med*. 2021;85:2263-2277. doi:10.1002/mrm.28546
35. Andersson J, Ahlström H, Kullberg J. Separation of water and fat signal in whole-body gradient echo scans using convolutional neural networks. *Magn Reson Med*. 2019;82:1177-1186. doi:10.1002/mrm.27786
36. Cho J, Park H. Robust water–fat separation for multi-echo gradient-recalled echo sequence using convolutional neural network. *Magn Reson Med*. 2019;82:476-484. doi:10.1002/mrm.27697
37. Leporq B, Lambert SA, Ronot M, Vilgrain V, Van Beers BE. Quantification of the triglyceride fatty acid composition with 3.0 T MRI. *NMR Biomed*. 2014;27:1211-1221. doi:10.1002/nbm.3175
38. Trinh L, Peterson P, Leander P, Brorson H, Månsson S. In vivo comparison of MRI-based and MRS-based quantification of adipose tissue fatty acid composition against gas chromatography. *Magn Reson Med*. 2020;84:2484-2494. doi:10.1002/mrm.28300
39. Poorman ME, Braškutė I, Bartels LW, Grissom WA. Multi-echo MR thermometry using iterative separation of baseline water and fat images. *Magn Reson Med*. 2019;81:2385-2398. doi:10.1002/mrm.27567
40. Bydder M, Ghodrati V, Gao Y, Robson MD, Yang Y, Hu P. Constraints in estimating the proton density fat fraction. *Magn Reson Imaging*. 2020;66:1-8. doi:10.1016/j.mri.2019.11.009
41. Wang X, Colgan TJ, Hinshaw LA, et al. T1-corrected quantitative chemical shift-encoded MRI. *Magn Reson Med*. 2020;83:2051-2063. doi:10.1002/mrm.28062
42. Matakos A, Balter JM, Cao Y. A robust method for estimating B0 inhomogeneity field in the liver by mitigating fat signals and phase-wrapping. *Tomography*. 2017;3:79-88. doi:10.18383/j.tom.2017.00003
43. Pineda AR, Reeder SB, Wen Z, Pelc NJ. Cramér-Rao bounds for three-point decomposition of water and fat. *Magn Reson Med*. 2005;54:625-635. doi:10.1002/mrm.20623
44. Reeder SB, Robson PM, Yu H, et al. Quantification of hepatic steatosis with MRI: the effects of accurate fat spectral modeling. *J Magn Reson Imaging*. 2009;29:1332-1339. doi:10.1002/jmri.21751
45. Wang X, Hernando D, Reeder SB. Sensitivity of chemical shift-encoded fat quantification to calibration of fat MR spectrum. *Magn Reson Med*. 2016;75:845-851. doi:10.1002/mrm.25681
46. Peterson P, Trinh L, Månsson S. Quantitative ^{1}H MRI and MRS of fatty acid composition. *Magn Reson Med*. 2021;85:49-67. doi:10.1002/mrm.28471
47. Hamilton G, Yokoo T, Bydder M, et al. In vivo characterization of the liver fat ^{1}H MR spectrum. *NMR Biomed*. 2011;24:784-790. doi:10.1002/nbm.1622
48. Ren J, Dimitrov I, Sherry AD, Malloy CR. Composition of adipose tissue and marrow fat in humans by ^{1} H NMR at 7 Tesla. *J Lipid Res*. 2008;49:2055-2062. doi:10.1194/jlr.D800010-JLR200

49. Hernando D, Sharma SD, Kramer H, Reeder SB. On the confounding effect of temperature on chemical shift-encoded fat quantification. *Magn Reson Med*. 2014;72:464-470. doi:10.1002/mrm.24951
50. Hines CDG, Yu H, Shimakawa A, McKenzie CA, Brittain JH, Reeder SB. T_1 independent, T_2* corrected MRI with accurate spectral modeling for quantification of fat: validation in a fat-water-SPIO phantom. *J Magn Reson Imaging*. 2009;30:1215-1222. doi:10.1002/jmri.21957
51. Bush EC, Gifford A, Coolbaugh CL, Towse TF, Damon BM, Welch EB. Fat-water phantoms for magnetic resonance imaging validation: a flexible and scalable protocol. *JoVE*. 2018;139:57704. doi:10.3791/57704
52. Clarke WT, Stagg CJ, Jbabdi S. FSL-MRS: an end-to-end spectroscopy analysis package. *Magn Reson Med*. 2021;85:2950-2964. doi:10.1002/mrm.28630
53. Marshall I, Higinbotham J, Bruce S, Freise A. Use of Voigt lineshape for quantification of in vivo 1H spectra. *Magn Reson Med*. 1997;37:651-657. doi:10.1002/mrm.1910370504
54. Simchick G, Zhao R, Hamilton G, Reeder SB, Hernando D. Spectroscopy-based multi-parametric quantification in subjects with liver iron overload at 1.5T and 3T. *Magn Reson Med*. 2022;87:597-613. doi:10.1002/mrm.29021
55. Hernando D, Sharma S, Aliyari M, et al. Multi-site fat-water phantom Mri data. 2016. doi:10.5281/ZENODO.48266
56. Breuer FA, Blaimer M, Mueller MF, et al. Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA). *Magn Reson Med*. 2006;55:549-556. doi:10.1002/mrm.20787
57. R Core Team. R: A Language and Environment for Statistical Computing. 2020 https://www.R-project.org/.
58. Reeder SB, Bice EK, Yu H, Hernando D, Pineda AR. On the performance of T2* correction methods for quantification of hepatic fat content. *Magn Reson Med*. 2012;67:389-404. doi:10.1002/mrm.23016
59. Colgan TJ, Hernando D, Sharma SD, Reeder SB. The effects of concomitant gradients on chemical shift encoded MRI: effects of CGs on CSE-MRI. *Magn Reson Med*. 2017;78:730-738. doi:10.1002/mrm.26461
60. Ruschke S, Eggers H, Kooijman H, et al. Correction of phase errors in quantitative water-fat imaging using a monopolar time-interleaved multi-echo gradient echo sequence: phase error correction in time-interleaved water-fat imaging. *Magn Reson Med*. 2017;78:984-996. doi:10.1002/mrm.26485
61. Daudé P, Troalen T, Mackowiak ALC, et al. Trajectory correction enables Free running cardiac DIXON at 3T. In: 2022 ISMRM-ESMRMB & ISMRT 31st Annual Meeting. London, UK. 2022. Abstract 1646
62. Diefenbach MN, Ruschke S, Eggers H, Meineke J, Rummeny EJ, Karampinos DC. Improving chemical shift encoding-based water-fat separation based on a detailed consideration of magnetic field contributions. *Magn Reson Med*. 2018;80:990-1004. doi:10.1002/mrm.27097
63. Navaratna R, Zhao R, Colgan TJ, et al. Temperature-corrected proton density fat fraction estimation using chemical shift-encoded MRI in phantoms. *Magn Reson Med*. 2021;86:69-81. doi:10.1002/mrm.28669
64. Hong CW, Mamidipalli A, Hooker JC, et al. MRI proton density fat fraction is robust across the biologically plausible range of triglyceride spectra in adults with nonalcoholic steatohepatitis. *J Magn Reson Imaging*. 2018;47:995-1002. doi:10.1002/jmri.25845
65. Wood JC, Enriquez C, Ghugre N, et al. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. *Blood*. 2005;106:1460-1465. doi:10.1182/blood-2004-10-3982
66. Hernando D, Cook RJ, Qazi N, Longhurst CA, Diamond CA, Reeder SB. Complex confounder-corrected R2* mapping for liver iron quantification with MRI. *Eur Radiol*. 2021;31:264-275. doi:10.1007/s00330-020-07123-x
67. Jhaveri KS, Kannengiesser SAR, Ward R, Kuo K, Sussman MS. Prospective evaluation of an R2* method for assessing liver iron concentration (LIC) against FerriScan: derivation of the calibration curve and characterization of the nature and source of uncertainty in the relationship: uncertainty in the FerriScan-R2* curve. *J Magn Reson Imaging*. 2019;49:1467-1474. doi:10.1002/jmri.26313
68. Hernando D, Zhao R, Yuan Q, et al. Multi-center, multi-vendor reproducibility and calibration of MRI-based R2* for liver iron quantification. *Blood*. 2021;138:2010. doi:10.1182/blood-2021-148803
69. Hernando D, Zhao R, Yuan Q, et al. Multicenter reproducibility of liver iron quantification with 1.5-T and 3.0-T MRI. *Radiology*. 2022;306(2):e213256. doi:10.1148/radiol.213256
70. Goldfarb JW, Craft J, Cao JJ. Water-fat separation and parameter mapping in cardiac MRI via deep learning with a convolutional neural network: MR water-fat separation. *J Magn Reson Imaging*. 2019;50:655-665. doi:10.1002/jmri.26658
71. Sharma SD, Artz NS, Hernando D, Horng DE, Reeder SB. Improving chemical shift encoded water–fat separation using object-based information of the magnetic field inhomogeneity. *Magn Reson Med*. 2015;73:597-604. doi:10.1002/mrm.25163
72. Chan KS, Marques JP. SEPIA—susceptibility mapping pipeline tool for phase images. *Neuroimage*. 2021;227:117611. doi:10.1016/j.neuroimage.2020.117611

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

DATA S1. Computation for all IDEAL TE values using arbitrary numbers of TE.

TABLE S1. Table of linear regression analysis on multi-site fat-water phantom data.

FIGURE S1. Magnitude spectra reference for the quantification of proton-density fat fraction (PDFF) of the phantoms using STEAM sequence at the first echo (TE = 12 ms). Targeted and measured PDFF values are reported for each spectrum. Of note, the 80% and 90% vials had to be shaken prior to acquisitions to restore the emulsion homogeneity, which lasted few hours afterwards.

FIGURE S2. PDFF was measured by Fatty-Riot-GC and Hernando-GC over synthetic CSE-MRI volumes with SNR = 50 at 3 and 7 echoes. PDFF maps were averaged along repetition axis and separated according to the echo spacing scheme (IDEAL, IN-/OUT-OF-PHASE,

MINIMAL). Fat-water swaps were visible at 3 T with Fatty-Riot-GC (inside the dash box) while bias appeared on the PDFF map processed by Hernando-GC due to B_0 inhomogeneities (shown by arrows). With 7 echoes, both algorithms provided more reliable quantitative maps.

FIGURE S3. Absolute error for the PDFF quantification of simulated datasets, averaged over repetitions, the whole range of B_0 off-resonance values with a SNR = 50 and realistic minimal echo spacing scheme.

FIGURE S4. PDFF bias estimated from a simulation with R_2^* values that differed between fat and water ($R^*_{2\text{fat}} = 20\,\text{s}^{-1}$, $R^*_{2\text{water}} = 50\,\text{s}^{-1}$), corresponding to cases similar to our vials 40% and 60%. Mean and standard deviation bias were averaged along the B_0 off-resonance and repetition axes. Parameters were NTE = 7 (MINIMAL) and SNR = 100 with a peanut oil spectrum as input for both simulation and algorithms. A clear overestimation of ∼3% is observed from both algorithms IDEAL-CE and MSGCA-A.

FIGURE S5. Comparison of R_2^* bias of each algorithm in phantoms with NTE = 7.

FIGURE S6. Comparison of R_2^* bias of each algorithm over synthetic CSE-MRI volumes with NTE = 7 and SNR = 100. Mean and standard deviation R_2^* bias were averaged along the B_0 off-resonance and repetition axes and separated (in color) according to the echo spacing schemes. GOOSE and B0-NICE (in red square) were not further investigated due to highly biased results.

FIGURE S7. PDFF bias for the six selected algorithms depending on the input SNR of the simulated datasets with NTE = 5 and realistic minimal echo spacing. All algorithms proved highly resilient to low SNR (SNR = 10), with limited noise propagation in the eventual PDFF quantification.

How to cite this article: Daudé P, Roussel T, Troalen T, et al. Comparative review of algorithms and methods for chemical-shift-encoded quantitative fat-water imaging. *Magn Reson Med*. 2023;1-19. doi: 10.1002/mrm.29860