Received: 23 August 2019 | Revised: 19 November 2019 | Accepted: 27 November 2019

DOI: 10.1002/mrm.28135

FULL PAPER

Magnetic Resonance in Medicine

Performance of compressed sensing for fluorine-19 magnetic resonance imaging at low signal-to-noise ratio conditions

Ludger Starke[1] | **Andreas Pohlmann**[1] | **Christian Prinz**[1] | **Thoralf Niendorf**[1,2] | **Sonia Waiczies**[1]

[1]Berlin Ultrahigh Field Facility (B.U.F.F.), Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany

[2]Experimental and Clinical Research Center, a joint cooperation between the Charité Medical Faculty and the Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany

Correspondence
Ludger Starke, Berlin Ultrahigh Field Facility (B.U.F.F.), Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Robert-Rössle-Strasse 10, 13125 Berlin, Germany.
Email: Ludger.Starke@mdc-berlin.de

Funding information
Deutsche Forschungsgemeinschaft, Grants/Award Numbers: DFG WA2804, DFG PO1869, SFB 1365; H2020 European Research Council, Grants/Award Numbers: 743077 (ThermalMR)

Purpose: To examine the performance of compressed sensing (CS) in reconstructing low signal-to-noise ratio (SNR) ^{19}F MR signals that are close to the detection threshold and originate from small signal sources with no a priori known location.

Methods: Regularization strength was adjusted automatically based on noise level. As performance metrics, root-mean-square deviations, true positive rates (TPRs), and false discovery rates were computed. CS and conventional reconstructions were compared at equal measurement time and evaluated in relation to high-SNR reference data. ^{19}F MR data were generated from a purpose-built phantom and benchmarked against simulations, as well as from the experimental autoimmune encephalomyelitis mouse model. We quantified the signal intensity bias and introduced an intensity calibration for in vivo data using high-SNR ex vivo data.

Results: Low-SNR ^{19}F MR data could be reliably reconstructed. Detection sensitivity was consistently improved and data fidelity was preserved for undersampling and averaging factors of $\alpha = 2$ or $= 3$. Higher α led to signal blurring in the mouse model. The improved TPRs at $\alpha = 3$ were comparable to a 2.5-fold increase in measurement time. Whereas CS resulted in a downward bias of the ^{19}F MR signal, Fourier reconstructions resulted in an unexpected upward bias of similar magnitude. The calibration corrected signal-intensity deviations for all reconstructions.

Conclusion: CS is advantageous whenever image features are close to the detection threshold. It is a powerful tool, even for low-SNR data with sparsely distributed ^{19}F signals, to improve spatial and temporal resolution in ^{19}F MR applications.

KEYWORDS

compressed sensing, experimental autoimmune encephalomyelitis, fluorine-19, magnetic resonance, neuroinflammation, signal bias

1 | INTRODUCTION

Fluorine (^{19}F) MRI has been commonly used in cell tracking and molecular imaging.[1-4] Its widespread applicability, along with the challenge that ^{19}F is only sparsely available in vivo, has motivated the development of strategies that enhance signal-to-noise ratio (SNR) efficiency, for example, more sensitive hardware[5-7] and more time-efficient pulse sequences.[8-10] Digital signal processing, particularly compressed sensing (CS),[11-13] is a new avenue to boost detection performance.

The first application of CS to ^{19}F MR methods was in chemical shift imaging.[14-16] When applied to ^{19}F MRI, the potential of CS to improve acquisition time efficiency was demonstrated.[17,18] Under optimal conditions, CS improved SNR efficiency by a factor of ~8.5.[17,18] At SNR = 58, even small features consisting of fewer than 5 voxels were correctly recovered from 32-fold undersampled phantom experiment data.[17] However, lower SNRs (8 and 14) resulted in a loss of these features in all CS reconstructions.[17] In a phantom experiment, reinvesting the time saved by undersampling into increased averaging lowered the detection threshold, when compared to fully sampled data at equal acquisition time.[18] Application of CS to in vivo ^{19}F MR was previously shown in situations where the location of ^{19}F signals in vivo was known: localized wound inflammation and transplantation of labeled pancreatic islets.[17,18] In the latter, 4-fold undersampling lowered the detection threshold, but also introduced false positives.[18] Regarding reconstruction of the correct signal intensity (SI), CS was found to introduce a downward bias of more than 50% in phantom experiments.[17]

From this rich body of literature, it is not clear how well CS performs under more challenging conditions (i.e., when studying small features close to detection thresholds). It is these conditions which necessitate sensitivity improvements most to boost detection and make high resolutions viable. SNR efficiency does not convey information regarding edge preservation,[19,20] and image defects are generally not captured by known noise distributions.[17] Given that CS is a nonlinear algorithm,[11-13] results cannot be extrapolated from experiments studying higher SNRs, and reconstruction methods must be compared at equal acquisition time.[21] This is particularly relevant for quantitative ^{19}F MR studies and raises a fundamental question under which conditions CS improves detection performance without compromising data fidelity. To draw valid conclusions about performance at low SNRs, a large number of reconstructions, signal distributions, and sampling patterns must be investigated to exclude random effects.

Unknown signal locations introduce additional challenges. The measured information and sparsity constraints need to be balanced in CS. For low-SNR data, case-by-case optimization based on visual inspection is not feasible. Given that the optimal choice depends on data scaling, noise level, undersampling factor, and image sparsity,[22] adopting a value optimized for a reference data set is problematic. To address this issue, we propose automatic regularization strength selection following Morozov's discrepancy principle.[23,24]

Studying the distribution of ^{19}F-labeled inflammatory cells in the central nervous system (CNS) during the disease course of experimental autoimmune encephalomyelitis (EAE) is 1 of many biomedical applications that could benefit greatly from improved time efficiency,[1,5,25-28] where addressing the aforementioned issues is necessary to conclude whether CS will be beneficial. EAE is a model of multiple sclerosis characterized by inflammatory lesions in the CNS,[29] which appear in arbitrary locations especially in the cerebrum, but can also follow fine anatomical structures, such as white matter tracts, in the cerebellum.[6] ^{19}F nanoparticles (NPs) are applied intravenously in animal models to study the distribution of inflammatory cell lesions in vivo.[17-21]

To study the performance of CS for ^{19}F MRI applications that include low-SNR conditions, we compared undersampled and CS reconstructed data with conventional Fourier and denoised reconstructions of fully sampled data. Reliability of automatic regularization strength selection and dependence of CS performance on undersampling factor and noise level was investigated in digital and MR phantom experiments. Fully sampled ^{19}F MR data of inflammation in EAE were acquired from ex vivo tissue phantoms and in vivo animals to allow preparation of 2000 k-spaces with different undersampling patterns and noise realizations by retrospective undersampling. The 2 setups corresponded to different EAE cohorts. In a third cohort, prospectively undersampled in vivo data were acquired with a genuine CS sequence. For performance assessment, we examined the root-mean-square deviation (RMSD) from the reference as a universal metric of image quality, the number of recovered features or true positive rates (TPRs) to assess detection performance, false discovery rates (FDRs) to quantify data fidelity, and the relative SI deviation from the reference to investigate bias effects. Finally, we propose a scheme for signal-level–specific intensity calibration using the ex vivo tissue collection as calibration data and assess the method's performance for in vivo measurements.

2 | METHODS

Simulations, data preparation, image reconstruction, and analysis were performed in MATLAB (R2018a; The MathWorks, Inc., Natick, MA), using pvmatlab for data import (Buker Biospin MRI, Ettlingen, Germany).

2.1 | Compressed sensing, denoising, and automatic regularization strength selection

CS reconstructions were computed identically for all setups, and denoised images were computed by applying the same

algorithm to fully sampled data. ^{19}F MR images are generally sparse, with the main component being background. It has been shown that ^{19}F chemical shift Imaging reconstructions enforcing strict data consistency and using only regularization with the ℓ^1-norm of the image, $\|r\|_1 = \sum_i |r_i|$, necessitate additional postprocessing attributable to spike artifacts.[14,15] Similar to previous work,[17] we therefore opted for relaxed data consistency, the common choice for noisy data,[30] and also added a total variation constraint to ensure robustness at low SNRs and effective noise reduction. The CS reconstruction $\hat{r}$ was computed as (Equation 1):

$$\hat{r} = \text{argmin}_r \left\| F_u r - y \right\|_2^2 + \lambda \left(\|r\|_1 + \text{TV}_{\text{iso}}(r) \right), \quad (1)$$

where F_u denotes the undersampled Fourier transform, y the measured k-space data, $\text{TV}_{\text{iso}}(r)$ the isotropic total variation,[31] and λ the regularization strength controlling the balance between the regularization terms and the data and thereby the data consistency $\varepsilon = \left\| F_u \hat{r} - y \right\|_2^2$. The used regularization terms have the advantage of being easily interpretable compared to wavelet constraints commonly utilized in anatomical MRI.[32] Equation 1 was solved using an implementation of the accelerated alternating direction method of multipliers[33] with fast Fourier transform–based exact inversion exploiting the circulant matrix structure that arises from periodic boundary conditions for the total variation.[33,34]

To determine the desired data consistency, we applied Morozov's discrepancy principle.[23,24] The data consistency of the reconstruction is matched with the accuracy of the data itself. Given noise standard deviation σ and n sampled complex data points, the expected deviation of the data from true, but unknown, image r_t is $\varepsilon' \left\langle \left\| F_u r_t - y \right\|_2^2 \right\rangle = 2n\sigma^2$. Multiple reconstructions are performed until a value of λ is found for which $\varepsilon \approx \eta\varepsilon'$ with η specifying the desired ratio of ε and ε'.[24,35] To this end, we used the Illinois algorithm.[36]

We performed simulations based on a digital phantom to determine a suitable value of η. The [128 × 128] pixel phantom consisted of 8 circular features (each with a diameter of 4.9 pixels), with SI (1/8, 2/8, …, 8/8) forming a larger circle. Partial volume effects were emulated by creating the digital phantom image 20 times more resolved, before downsampling. k-Space data were generated by adding complex Gaussian noise to the digital phantom before applying the Fourier transform and, if appropriate, an undersampling mask in 1 dimension (see "Data Preparation" section). We examined 5 different noise levels ($\sigma = 0.01$ to $\sigma = 0.2$ with 4-fold undersampling ($f_{\text{us}} = 1/4$) and 5 different undersampling factors ($f_{\text{us}} = 1$ to $f_{\text{us}} = 1/8$) with $\sigma = 0.04$. For each condition and 20 values of η between 0.9 and 1.05, we performed 40 reconstructions of data with individual noise realizations and undersampling masks. As a metric of image quality, we computed the mean and standard deviation of the RMSD from the reference.

2.2 | MR measurements and experimental setups

We used 3 MR measurement setups to examine CS performance. All MR experiments were carried out on a 9.4T MR system (BioSpec 94/20, G_{max} = 440 mT/m, slew rate = 3440 mT/m/ms, B-GA12, PV6.1 software, Bruker BioSpin MRI; Bruker Corporation, Billerica, MA), using an ^{19}F/^{1}H mouse head volume radiofrequency (RF) coil.[5] Animal experiments were conducted in accord with procedures approved by the Animal Welfare Department of the State Office of Health and Social Affairs Berlin (LAGeSo, Ref0127-16) and conformed to guidelines to minimize discomfort to animals (86/609/EEC).

2.2.1 | Setup 1: capillary tube phantom

Eight concentrically-arranged capillary tubes (inner diameter, 1.15 mm) submerged in water (placed horizontally in the magnet bore) were filled with (1/8, 2/8, … 8/8) × 33% trifluoroethanol diluted in water and 1 mmol/L of gadolinium (Magnevist; Bayer Vital, Leverkusen, Germany). Axial 2D images were acquired using rapid acquisition with relaxation enhancement (RARE; flip-back module, echo train length [ETL] = 8, TR = 600 ms, TE = 12 ms, [30 × 30] mm2 field of view [FOV], 128 × 128 matrix). To construct data with different noise levels, 6 slice thicknesses were used (0.1, 0.25, 0.5, 1, 2.5, and 5 mm) and 64 repetitions/slice acquired. The digital phantom described in the preceding section was designed to conform to this phantom such that the MR experiment could be replicated by simulations.

2.2.2 | Setup 2: ex vivo EAE tissue

EAE was actively induced in female 3-month-old SJL/J mice ($n = 5$), by subcutaneously immunizing with the CNS antigen proteolipid protein ($\text{PLP}_{139–151}$, 250 μg/animal) emulsified with Mycobacterium Tuberculosis H37RA (800 μg/animal in 100 μL of complete Freund's adjuvant).[37] Pertussis toxin (1.25 ng/μL in 200 μL of phosphate-buffered saline [PBS]) was administered intraperitoneally on days 0 and 2. Animals were weighed daily, and a neurological scoring was performed daily to assess the EAE symptoms. Adding up scoring points (righting reflex weakness 0.5, tail paresis 0.5, tail paralysis 1, unilateral hindlimb paresis 0.5, bilateral hindlimb paralysis 1, unilateral forelimb paresis 0.75, and bilateral forelimb paralysis 1.5) results in the final EAE score.

High-fluorine-content Pluronic-PFCE NPs consisting of perfluoro-15-crown-5-ether (PFCE; Fluorochem Ltd, Hadfield, UK), emulsified in Pluronic F-68 (Sigma-Aldrich,

Steinheim, Germany), were prepared (1.2 M) and characterized as previously described.[26,38] NPs containing 5 μmol of PFCE were administered daily from day 5 after immunization until the end of the experiment.[6] Animals were sacrificed on day 10 using a lethal dose of anesthetic and transcardially perfused (20 mL of PBS and 20 mL of 4% paraformaldehyde [PFA] fixative). Tissue was prepared for ex vivo MRI as described previously[6] and secured within 15-mL tubes filled with 4% PFA, keeping the CNS as well as draining lymph nodes in situ. A cylindrical cap (diameter 3 mm, height 1 mm) containing NPs (2 mM of PFCE in 2% agarose) was fixed outside of the tube as an external standard. ^{19}F MRI was conducted using 3D-RARE (flip-back module, ETL = 40, TR = 800 ms, TE = 4.1 ms, [40 × 16 × 16] mm^3 FOV, 100 × 40 × 40 matrix, 25 repetitions, each repetition = 6 averages). For ^{1}H MR, 3D-RARE (flip-back module, ETL = 8, TR = 1000 ms, TE = 5.6 ms, [40 × 16 × 16] mm^3 FOV, 241 × 96 × 96 matrix, 1 average) was used.

2.2.3 | Setup 3: in vivo EAE model, measurements with retrospective undersampling

In a second cohort of EAE mice (n = 5), 11 fully sampled data sets were acquired between days 10 and 14 after induction. EAE was induced as described for setup 2. Mice were anesthetized using a mixture of isoflurane (0.5–1.5%) in pressurized air and oxygen as inhalation narcosis. The animal was placed on a warm animal MR bed and supplied with oxygen and air. Respiration was monitored using a respiration pad and temperature by a rectal probe. To distinguish NP ^{19}F signals from isoflurane, excitation/refocusing RF pulses were restricted to a bandwidth of 3000 Hz. 3D MRI was performed as in setup 2, keeping parameters including resolution the same, but reducing FOV and matrix size in the frequency-encoding direction to [45 × 16 × 16] mm^3 and 112 × 40 × 40 or 270 × 96 × 96 for ^{19}F or ^{1}H imaging, respectively. In vivo, only 25 repetitions could be acquired. The PFCE NP cap used in setup 2 was positioned on the tooth bar.

2.2.4 | Setup 4: in vivo EAE model, measurements with prospective undersampling

In a third cohort of EAE mice (n = 4), 4 data sets with prospective undersampling were acquired on days 12 to 14 after EAE induction (as described for setups 2 and 3). We developed a 2D-RARE CS protocol for ^{19}F MRI (flip-back module, ETL = 40, TR = 1020 ms, TE = 5.1 ms, [20 × 20] mm^2 FOV, 128 × 128 matrix, 3.2-mm slice thickness, 0.4-mm slice gap, 6 slices). Two hundred ninety-six, 592, and 1184 averages were acquired with no, 2-fold, and 4-fold undersampling, respectively (acquisition time = 20 minutes). The fully sampled measurement was repeated 4 times as a reference (80 minutes), and a pure noise scan was acquired to determine the noise level.[39] For ^{1}H MR, 2D-RARE (no flip-back module, ETL = 4, TR = 1000 ms, TE = 5.5 ms, [20 × 20] mm^2 FOV, 256 × 256 matrix, 0.8 mm slice thickness, 0.4 mm slice gap, 18 slices, 8 averages) was used.

2.3 | Undersampling and data preparation

Undersampling patterns were generated based on polynomial variable-density distributions adapted from Zijstra et al.[40] For setups 1 to 3, undersampled data were created by retrospectively undersampling k-space data in the phase-encoding direction(s). For setup 4, undersampling was implemented prospectively in the data acquisition. Ten percent of the sampled k-space lines were assigned deterministically to the k-space center. The remaining lines were drawn randomly. For the 2D acquisitions (setups 1 and 4), undersampling masks were sampled from the 1D distribution (Equation 2):

$$p\left(k_x\right) \propto \left(1-\left|k_x\right|\right)^{\frac{3}{2}}, \tag{2}$$

where $k_x \in [-1, 1]$ denotes the x-position of the k-space line. Similarly, a 2D distribution was used to undersample the 3D data of setups 2 and 3 (Equation 3):

$$p\left(k_x, k_y\right) \propto \max\left(1-\sqrt{k_x^2+k_y^2}, 0\right)^{\frac{3}{2}}, \tag{3}$$

where k_x, $k_y \in [-1, 1]$ denote x and y positions of the sampled lines. Both distributions are weighted toward lower spatial frequencies by choosing an exponent >1.

We compared Fourier and denoised reconstructions of fully sampled data with CS reconstructions of undersampled data at equal scan time. For setups 1 to 3, multiple repetitions were averaged to offset the accelerated data acquisition (Figure 1A). Averaging was performed on the k-space data before image reconstruction. The factor α denotes undersampling and averaging: $\alpha = 1$ signifies fully sampled data ($f_{us} = 1$); $\alpha = 3$ signifies 3-fold undersampling ($f_{us} = 1/3$) with 3-fold averaging. Table 1 summarizes the data preparation for all setups.

2.3.1 | Setup 1

For each slice thickness, the SI of the tube with the highest ^{19}F signal was determined. Because of slice profile imperfections, the SI values are not simply proportional to the slice thickness. The data were scaled relative to each other by assigning the same SI to the tube with the highest ^{19}F signal. Twelve noise-level conditions were examined (Table 1).

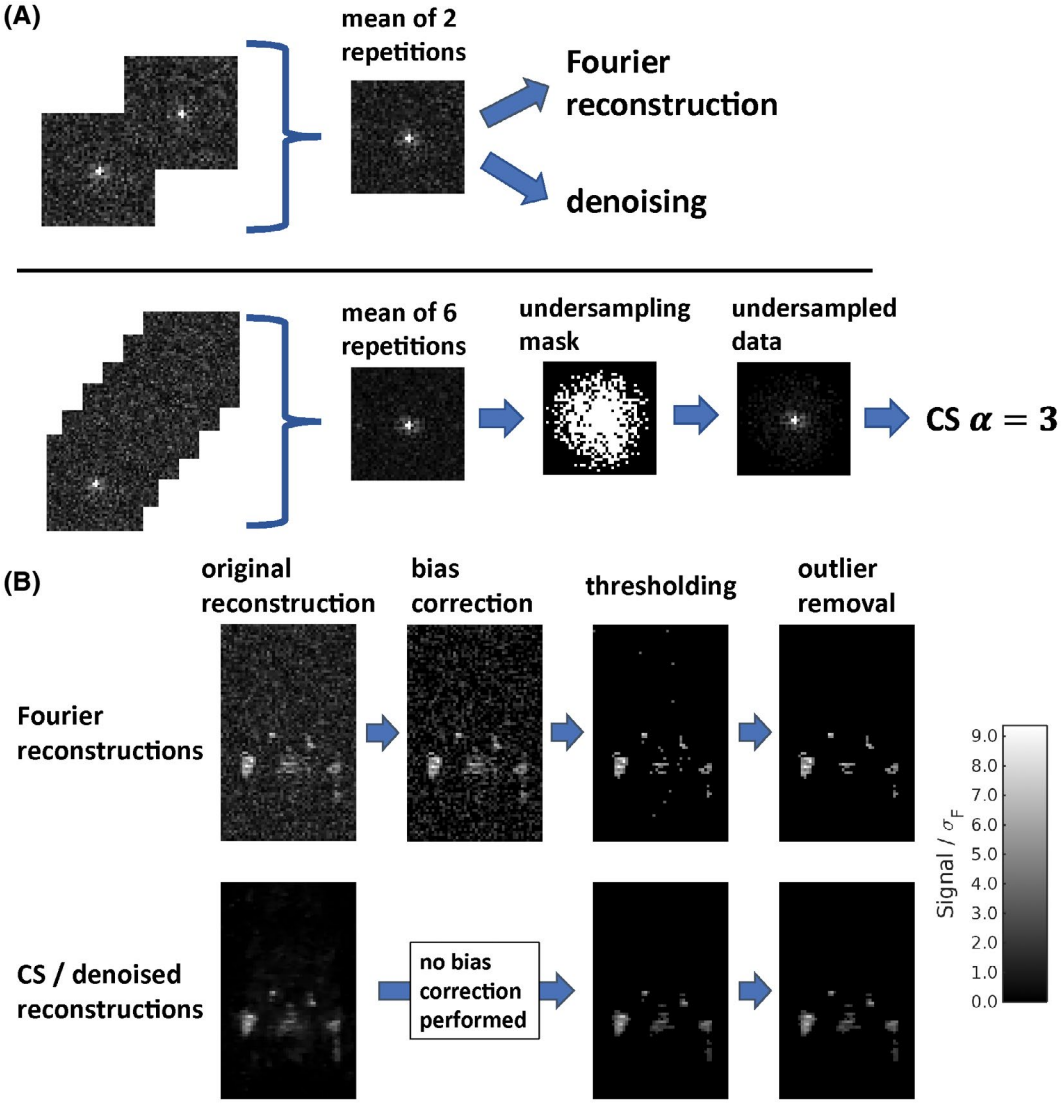

FIGURE 1 Data preparation and background subtraction. (A) In the MR experiments of setups 1 to 3, multiple repetitions of fully sampled data were acquired, which were then averaged to generate data with different noise levels. For the EAE experiments (setups 2 and 3), the acquisition of a single repetition took 3 minutes 12 seconds. Thus, for example, fully sampled data corresponding to a measurement time of 6 minutes 24 seconds was generated by averaging 2 randomly selected repetitions. These data were then either reconstructed using a simple Fourier transform or they were denoised. To generate 3-fold undersampled data for the same measurement time ($\alpha = 3$), 6 randomly selected repetitions were averaged before application of an undersampling mask focusing on the k-space center. Multiple noise realizations were obtained by selecting different repetitions. (B) Before further analysis, image voxels were classified as either signal or background. Here, the used multistep procedure is demonstrated in a slice of an exemplary Fourier reconstruction and CS reconstruction of in vivo data (setup 3). The slice is cropped in read direction to focus the illustration on the fluorine signal. For the Fourier reconstructions only, first the Rician noise bias was corrected. Second, the data were thresholded at 3.5σ (Fourier reconstructions) or 2σ (denoised and CS reconstructions), where σ denotes the noise standard deviation of the underlying k-space data. Last, remaining groups of <3 connected signal voxels were interpreted as outliers and removed from the data. Denoised reconstructions were treated identical to CS reconstructions. σ_F in the label of the color scale denotes the noise standard deviation of the Fourier reconstruction

These conditions signify different measurement times and are distinguished by peak SNR ($pSNR_F$, the SNR of the highest SI tube in the Fourier reconstruction). For each noise condition and each undersampling and averaging factor ($\alpha = 1$ to 8), 60 k-space data sets with individual undersampling masks and combinations of averaged repetitions were generated. Artificial data were generated based on the digital phantom with the same noise conditions, undersampling and averaging factors, and number of data sets. Thus, for both the capillary tube phantom and digital phantom, 5760 different k-spaces were reconstructed.

2.3.2 | Setups 2 and 3

In the EAE model, 5 measurement times (192–960 seconds) were investigated. Four CS undersampling and averaging factors were used (α = [2, 3, 4, 5]). For each

TABLE 1 Summary of acquired data, data preparation, and computed reconstructions

	Setup 1		Setup 2	Setup 3	Setup 4
	Capillary tube phantom	**Digital phantom**	**EAE ex vivo**	**EAE in vivo**	**EAE in vivo**
Dimension	2D	2D	3D	3D	2D
No. of individual datasets	1	1	5	11	4
Data in each dataset	6 slice thicknesses (0.1–5.0 mm) with 64 repetitions each	-	40 repetitions	25 repetitions	Fourier: 4 repetitions $\alpha = 2$: 1 rep. $\alpha = 4$: 1 rep.
Ground truth	Average of 64 repetitions, slice thickness 5 mm ($pSNR_F = 481.5$)	Digital phantom	Average of 40 repetitions (meas. time 128 min)	Average of 25 repetitions (meas. time 80 min)	Average of 4 repetitions (meas. time 80 min)
No. of investigated measurement times	12	12	5	5	1
Corresponding fully-sampled data	0.1 mm slice thickness, 1 average ($pSNR_F = 1.0$) to 2.5 mm slice thickness, 2 averages ($pSNR_F = 41.1$)	$pSNR_F = 1.0$ to $pSNR_F = 41.1$	1 repetition (192 sec) to 5 averaged repetitions (960 sec)	1 repetition (192 sec) to 5 averaged repetitions (960 sec)	1 repetition
Investigated acquisition and reconstruction methods	Fourier, denoised, CS $\alpha = 2$ to CS $\alpha = 8$	Fourier, denoised, CS $\alpha = 2$ to CS $\alpha = 8$	Fourier, denoised, CS $\alpha = 2$ to CS $\alpha = 5$	Fourier, denoised, CS $\alpha = 2$ to CS $\alpha = 5$	Fourier, denoised, CS $\alpha = 2$, CS $\alpha = 4$
No. of reconstructions for each measurement time and method	60	60	5	5	1
Overall no. of recon.	6480	6480	750	1650	16

$pSNR_F$ denotes the peak SNR of the Fourier reconstruction at equal measurement time and α the factor of undersampling and averaging.

measurement time and each acquisition and reconstruction method, 5 k-space datasets were built. For the 5 ex vivo and 11 in vivo data sets, this led to overall 625 and 1375 different k-spaces.

2.3.3 | Setup 4

No further data preparation was required because of prospective undersampling.

2.4 | Data analysis

Performance was assessed by comparing reconstructions with fully sampled high-SNR reference data. The average of 64 repetitions of the thickest slice (pSNR = 481.5) and the digital phantom itself were used in setup 1. In setup 2, a measurement time of 128 minutes and in setups 3 and 4 measurement times of 80 minutes were used (Table 1).

Each data set of setup 3 was considered to be independent because the distribution of the ^{19}F MR signal changed over time. Before further analysis, the ^{19}F MR signal from the reference cap and investigated sample was segregated and only the latter part included in the evaluation of performance (setups 2 and 3).

2.4.1 | Rician noise bias correction

Conventional Fourier-reconstructed MR magnitude images are biased because of noise effects.[41] The measured signal follows a Rician distribution.[42] For true but unknown signal, S_t, the expected measured signal is[43] (Equation 4):

$$< S >_{p(S|S_t)} = \sigma \sqrt{\frac{\pi}{2}} L_{1/2}\left(-\frac{S_t^2}{2\sigma^2}\right), \tag{4}$$

where σ denotes the standard deviation of the Gaussian noise in the real and imaginary image and $L_{1/2}$ a Laguerre polynomial.

This function cannot be analytically inverted to yield a function $S_t(\langle S \rangle)$. Thus, we inverted it by means of a lookup table to implement the noise correction.[41] Noise bias correction was applied only to Fourier reconstructions (Figure 1B).

2.4.2 | Background subtraction

All reconstructions were thresholded to distinguish signal and background voxels based on the noise standard deviation of the underlying k-space data σ (Figure 1B). The threshold was set to 3.5σ for Fourier reconstructions and to 2σ for CS and denoised reconstructions (reflecting reduced background noise level). Groups of <3 connected signal voxels were removed as outliers (Figure 1B).

2.4.3 | Performance metrics

As a general metric of image quality, we used the RMSD from the reference. To visualize image quality gains between methods, we computed the relative RMSD (RMSD divided by the average RMSD of the Fourier reconstructions at equal measurement time). For setups 2 and 3, also the standard deviation of the relative RMSD over all reconstructions with a given measurement time and reconstruction method was calculated.

In setup 1, the number of detected tubes was counted as a measure of detection performance. Tubes were counted as detected if more than half of the corresponding voxels were classified as signal. For setups 2, 3, and 4, we used TPRs as a statistical measure of sensitivity and FDRs as a measure of reliability (Equations 5 and 6):

$$\mathrm{TPR} = n_{\mathrm{TP}} / \left(n_{\mathrm{TP}} + n_{\mathrm{FN}} \right) \tag{5}$$

$$\mathrm{FDR} = n_{\mathrm{FP}} / \left(n_{\mathrm{TP}} + n_{\mathrm{FP}} \right), \tag{6}$$

where n_{TP}, n_{FN}, and n_{FP} denote the number of true positives, false negatives, and false positives, respectively. We used a sliding window approach to compute signal-level–specific TPRs (Figure 2) and FDRs: All reconstructions were scaled to σ_{F} (noise standard deviation of the Fourier reconstruction at equal scan time) to make results at different measurement times comparable. The TPR was computed for different levels of the reference signal, S_r, whereas the FDR was computed for different levels of the measured signal, S. Four hundred windows with a width of 0.5 σ_{F} and equally spaced centers between 0 and 8 σ_{F} were considered. To combine the results for different reconstructions, the weighted mean and standard deviation were computed with weights given by the number of true signal voxels for the TPR and the number of observed signal voxels for the FDR (Figure 2). For setup 4, only the weighted mean was computed because of the smaller sample size.

For setup 1, we determined the signal deviation (measured signal divided by reference signal, S/S_r) averaged over all true positive voxels. For setups 2 and 3, the signal deviation of true positive voxels in each reconstruction was computed for windows of the measured signal, S (Figure 2), whereas mean and standard deviation were computed as weighted by the number of true positives in the subinterval. ^{19}F concentration was determined in the reference measurements by comparing signal intensities in tissue with those in the NP cap.

2.4.4 | Signal intensity calibration

Signal deviations measured in setup 2 (ex vivo) were used to compute a signal-level and method-specific correction factor, which was then applied to calibrate the reconstructions of setup 3 (in vivo), as illustrated in Figure 3. A fifth-degree polynomial was fitted to the signal deviation observed in the reference setup, yielding a function $f_\beta\left(\frac{S}{\sigma_{\mathrm{F}}}\right)$ for every reconstruction method β smoothly approximating the deviation at S/σ_{F}. The corrected signal $S_i^{(c)}$ for voxel i of the in vivo data was then computed as (Equation 7):

$$S_i^{(c)} = \left(f_\beta \left(\frac{S_i}{\sigma_{\mathrm{F}}} \right) \right)^{-1} S_i. \tag{7}$$

3 | RESULTS

3.1 | Automatic regularization strength selection

Choice of parameter η strongly influenced the suppression of background noise and the rendering of low-SNR features, following qualitative assessment (Figure 4A). Dependence of image quality on η was quantified by computing the RMSD from the reference (Figure 4B,C). For 4-fold undersampling, image quality was barely affected at $0.93 < \eta < 0.98$ for various noise levels (Figure 4B). In this range, RMSD differences remained below random data variability. The same held true for $f_{\mathrm{us}} = 1/6$ and $f_{\mathrm{us}} = 1/8$ (Figure 4C). Pure denoising and $f_{\mathrm{us}} = 1/2$ results, however, depended more critically on η (Figure 4C), with optimal performance at $\eta = 0.979$ and $\eta = 0.963$, respectively. Thus, $\eta = 0.97$ was chosen for all further reconstructions.

3.2 | Performance of CS in phantom experiments

Compared to Fourier reconstructions at the same measurement time, CS improved image quality and increased the number of detected features in phantom experiments (Figure 5A). No false positives occurred. At $\mathrm{pSNR_F} = 1.0$, the RMSD was

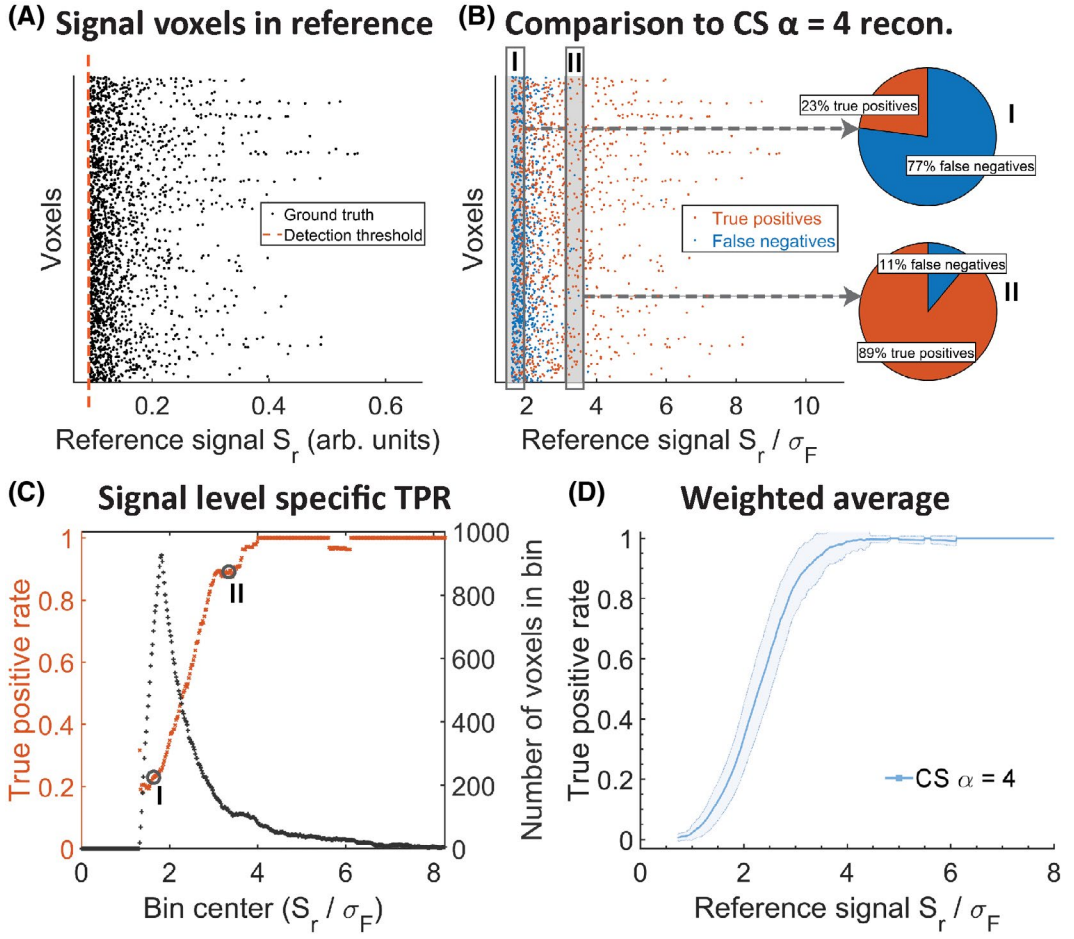

FIGURE 2 Computation of signal-level–specific TPRs. Based on data with a long scan time, a reference was computed for each data set. This reference typically contained a few hundred voxels above the detection threshold (A). The voxels are shown in an arbitrary order. A given reconstruction correctly detected a subset of these signal voxels (true positives), while other voxels are classified as background (false negatives; B). It is expected that the detection performance mainly depends on the SNR. Thus, the signal level is normalized by the noise standard deviation of the Fourier reconstruction at equal scan time σ_F to allow comparisons between different reconstruction methods and data sets. Considering only a subinterval of the signal range, a signal-level–specific TPR can be computed (I and II). Computing TPRs for a moving window yields a TPR curve summarizing the detection performance in a single reconstruction (C). The number of reference signal voxels in the window determines the reliability of the obtained results. Thus, it is used as a weight when averaging results for, for example, CS $\alpha = 4$ reconstructions from all data sets and investigated measurement times (d). The weighted standard deviation serves as a measure of variability. FDRs and SI deviations were computed analogously. arb. units = arbitrary units

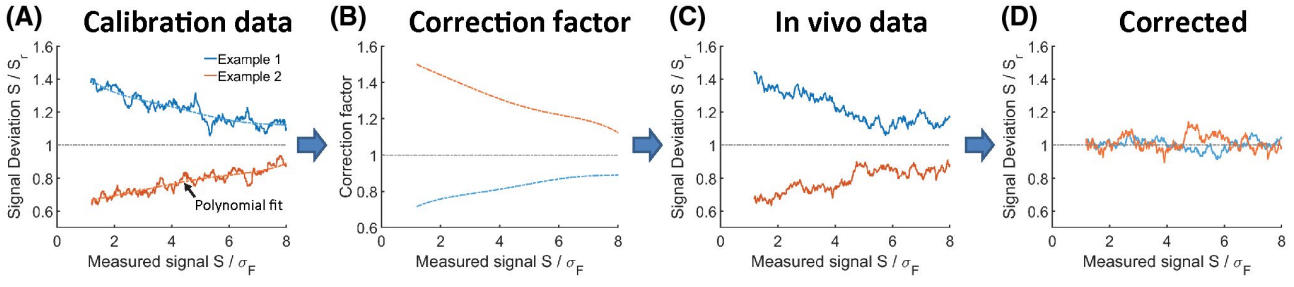

FIGURE 3 Signal intensity calibration. The figure shows artificial example data for 2 hypothetical reconstruction methods. (A) The intensity deviation is computed at different signal levels for a given reconstruction method based on a high-SNR reference. These reference calibration data should have a signal distribution similar to the in vivo data that need correction. The calibration data could be ex vivo data from a different animal cohort as shown in this article. For retrospective studies, this could also be ex vivo data acquired from the same animals postmortem. Even phantom or simulated data could be used. A smooth approximation of the observed deviation is achieved by fitting a fifth-degree polynomial function. (B) The inverse of this polynomial fit yields a method- and signal-level–specific correction factor. (C) Signal deviation observed in in vivo data reconstructed with the 2 methods. (D) Multiplication with the correction factor yields on average unbiased estimates

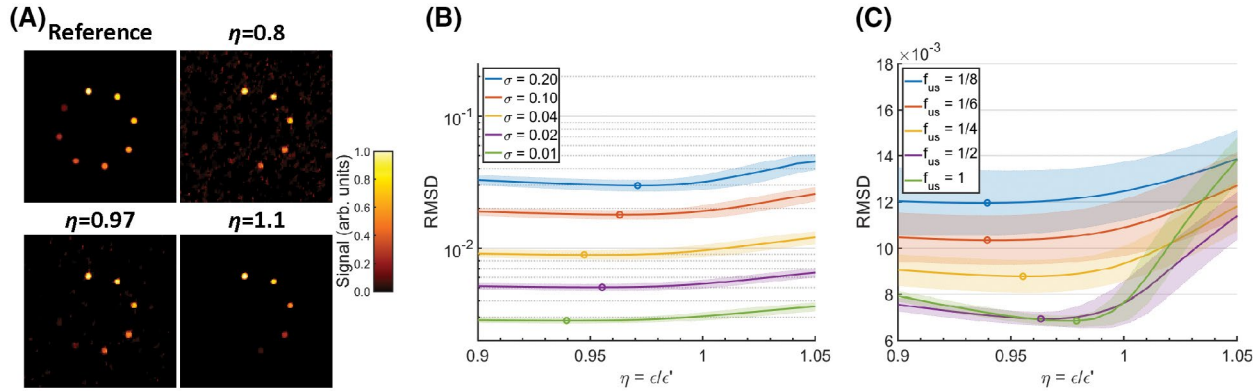

FIGURE 4 Automatic regularization strength selection. The discrepancy principle selects a value for the regularization strength by matching the data consistency of the reconstruction ε with the deviation of the data from the true, but unknown, image ε'. Their desired relation is specified by parameter $\eta = \varepsilon/\varepsilon'$. (a) Digital phantom reference with example reconstructions ($f_{us} = 1/4$, $\sigma = 0.1$). For $\eta = 0.8$, the image is only partially denoised and retains aliasing artifacts, making it difficult to distinguish between true features and noise. With $\eta = 0.97$, most features are well delineated and distinguishable from the background. With $\eta = 1.1$, the reconstruction is oversmoothed. Signal intensities are depressed and low-intensity features are missing. (b) RMSD from the reference for different values of η at 4-fold undersampling ($f_{us} = 1/4$) and 5 different noise standard deviations σ. The shaded area is given by the standard deviation of the metric over 40 reconstructions. The circle labels the minimal mean RMSD. (c) Analogous results for 5 different undersampling factors at $\sigma = 0.04$

equally high for all methods given that only few signal voxels were detected (Figure 5B). CS reconstructions were advantageous in the range of $pSNR_F = 1.5$ to 29.1. Between $pSNR_F = 2.1$ and $pSNR_F = 16.1$, CS reduced the RMSD by more than 20%. Whereas the largest reduction was achieved with $\alpha = 8$ (46%), most of the improvement was already achieved with $\alpha = 4$ (up to 45%; Figure 5B). At $pSNR_F = 1.5$, only $\alpha > 4$ offered an improvement >2%. At $pSNR_F = 29.1$, only $\alpha < 6$ was beneficial. Fourier reconstructions were only superior to CS at $pSNR_F = 41.1$, but here the absolute error in CS reconstructions was still small.

CS reconstructions consistently detected more image features (Figure 5C). Performance improved with increasing α though the difference between $\alpha = 6$, 7, or 8 was small (Figure 5C). The strongest improvement was achieved in conditions where few or no features were observed using conventional methods: at $pSNR_F = 3.0$ on average 5.1 tubes were detected with CS $\alpha = 8$, 4.3 with $\alpha = 4$, 2.8 with $\alpha = 2$, and only 0.1 in the Fourier reconstructions. At $pSNR_F = 6.2$, the difference between $\alpha = 8$ and $\alpha = 2$ was reduced to an average of 6.9 and 6.0 detected features. At $pSNR_F = 41.1$, all tubes were detected regardless of the method.

The signal was underestimated by up to 39% ($S/S_r = 0.61$) averaged over all true positive pixels in CS and denoised reconstructions (Figure 5D). The largest deviations were observed in noisy conditions ($pSNR_F = 2.1$–4.3) and with higher α factors. However, the trend was reversed at the lowest SNRs, and in some cases an overestimation occurred (CS $\alpha = 5$ at $pSNR_F = 1.0$ and denoised reconstructions at $pSNR_F = 2.1$). The average deviation dropped to <20% at $pSNR_F \geq 11.4$ for all α values. Fourier reconstructions also displayed an SI bias despite Rician noise bias correction. In contrast to CS, the signal was overestimated by up to 41% close to the detection threshold ($pSNR_F = 3.0$). For $pSNR_F \geq 8.8$, the average signal overestimation was below 5%.

3.3 | Performance of CS in simulations of a digital phantom

The results obtained from the experimental phantom setup could be accurately replicated in simulations (Figure 5E–H). It was only at high SNRs, especially at $pSNR_F = 41.1$, that CS performed better in simulations compared to measured MR data: This was the case for relative RMSD (Figure 5F) and signal deviation (Figure 5H), but not signal detection (Figure 5G).

3.4 | Performance of CS in animal experiments

We next studied CS performance using retrospective undersampling in the EAE animal model. In this animal model, some of the ^{19}F signals are close to the detection threshold as shown from representative reconstructions of both ex vivo (Figure 6A) and in vivo (Figure 6E) data. Additional reconstructions can be found under: central.xnat.org/data/projects/CSperf_19F-MRI. Compared to Fourier and denoised reconstructions, more true-signal voxels were detected with CS; already, CS $\alpha = 2$ identified features that would have remained undetected using conventional approaches. Although more true positive ^{19}F signals were uncovered with increasing α, more false positives also emerged; these generally had

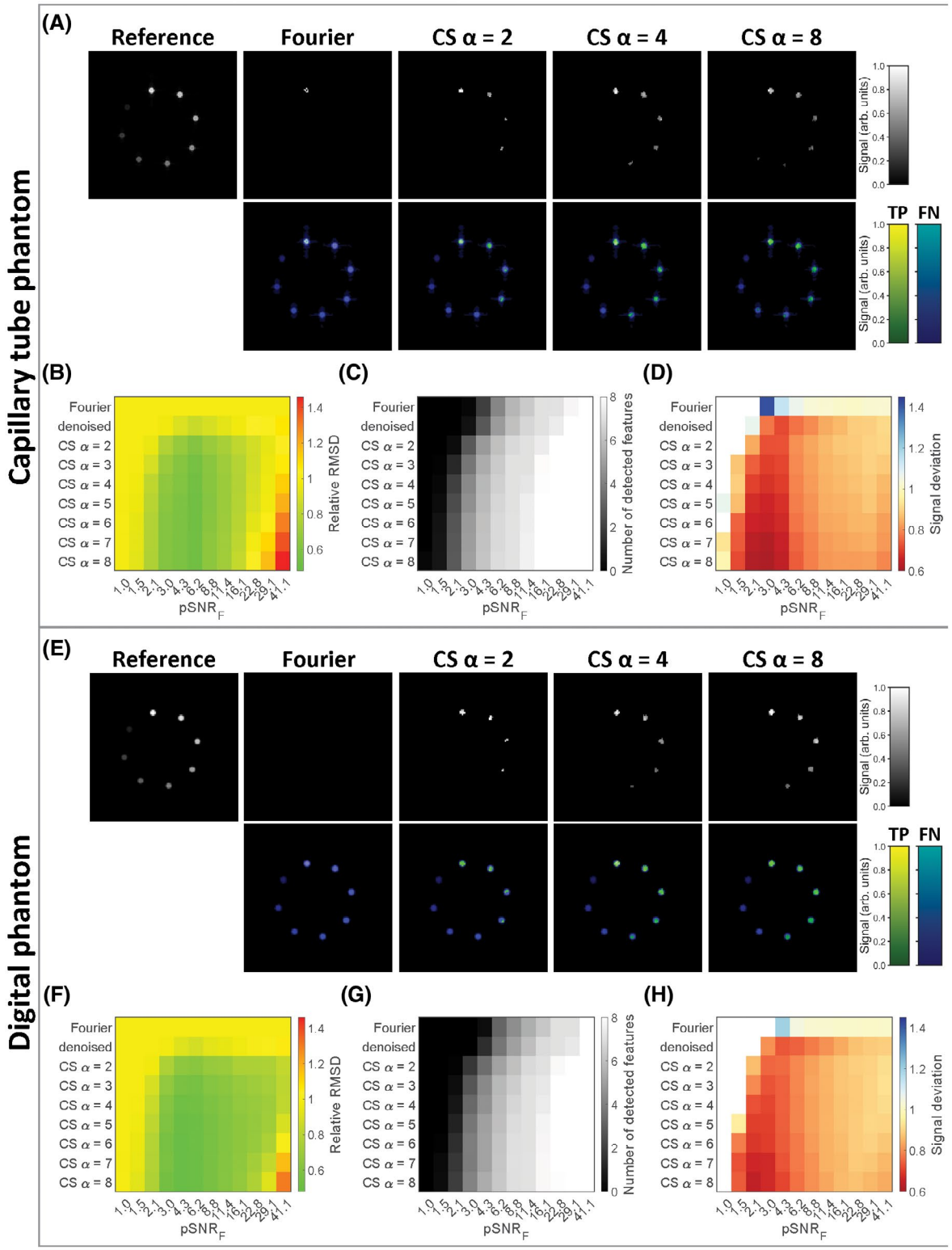

FIGURE 5 Phantom experiments (setup 1). (A) to (D) show results for the capillary tube phantom MR data and (F) to (H) for the digital phantom simulations. (A,E) Example reconstructions. The second line depicts the same reconstructions as the first with true positives (TPs) in green/yellow and false negatives (FNs) in blue/turquoise. No false positives occurred in these examples. (B,F) RMSD from the reference relative to the RMSD of the Fourier reconstruction at equal scan time. $pSNR_F$ denotes the peak SNR in the Fourier reconstruction at equal scan time. (C,G) Average number of detected features. A tube was counted as detected if half of its voxels were classified as signal. (D,H) Signal deviation averaged over all true positive voxels. The deviation was calculated as measured signal divided by reference signal

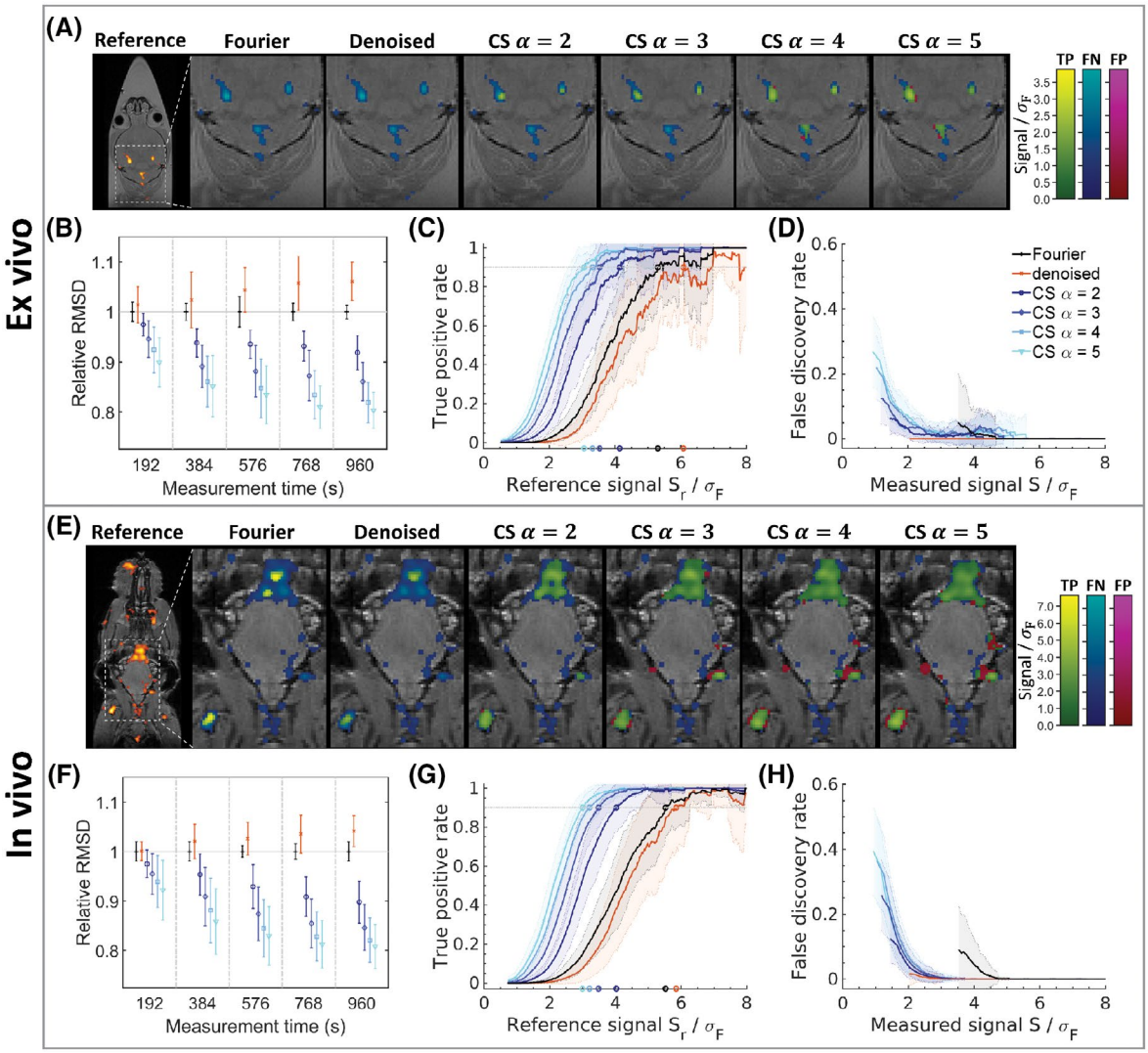

FIGURE 6 Detection performance for imaging inflammation in the EAE mouse model (setups 2 and 3). (A) to (D) show results for the ex vivo condition (setup 2) and (E) to (H) for the in vivo condition (setup 3). A legend is displayed in (D). (A,E) Comparisons of the different acquisition and reconstruction methods for exemplary slices. The reconstructions in both subfigures correspond to a measurement time of 16 minutes (960 seconds). A whole view of the slice with ^{19}F MR signal is shown in the reference image (far left) in the standard red hot color scale. True positives (TP) shown in green/yellow, false negatives (FN) in blue/turquoise, and false positives (FP) in red/violet were calculated in accord with the reference. (B,F) RMSD from the reference relative to the RMSD of the Fourier reconstruction at equal measurement time. Bars show the standard deviation of the measured improvement. (C,G) True positive rates (TPRs). Circles indicate the signal level where a TPR of 90% is reached. σ_F is the noise standard deviation of the Fourier reconstruction at equal measurement time and S_r the signal level in the reference. The shaded area marks the weighted standard deviation (see Figure 2). (D,H) FDRs. S denotes the measured signal level. Panels (B) to (D) and (F) to (H) summarize results from 750 and 1650 reconstructions, respectively

intensities close to the detection threshold and were localized at the edges of true ^{19}F features. This blurring was more pronounced in the in vivo data (Figure 6E).

Image quality was clearly improved in CS reconstructions compared to Fourier reconstructions, as shown by a consistently reduced relative RMSD (Figure 6B,F). This improvement increased with growing α. Denoising impaired image quality. For all measurement times except 192 seconds, image-quality improvements exceeded the variability (i.e., the standard deviation of the relative RMSDs). At measurement times longer than 384 seconds (ex vivo) and 768 seconds (in vivo), a differentiation in image-quality improvement appeared between $\alpha = 2$ and $\alpha = 5$.

Detection sensitivity, estimated from signal-level–specific TPRs, improved with increasing α (Figure 6C,G). Whereas TPRs for all reconstructions converged to 1 at high signal levels (reference signal $S_r \geq 6\sigma_F$), TPRs for CS reconstructions were consistently superior to other reconstructions at lower signal levels. For voxels with $S_r < 4\sigma_F$ (ex vivo) or $S_r < 4.5\sigma_F$ (in vivo), improvements exceeded

the variability. CS TPRs increased with increasing α, yet the differences were smaller than those compared to Fourier or denoised reconstructions. Lower TPRs were observed for denoised than Fourier reconstructions. Supporting Information Table S1 shows the reference signal levels at which a TPR of 90% was reached. In vivo, this level was at $5.5\sigma_F$ (Fourier), $4.0\sigma_F$ (CS $\alpha = 2$), and $3.0\sigma_F$ (CS $\alpha = 5$). The achieved improvement in detection performance with CS would translate to a 1.9- and 3.4-fold reduced measurement time compared to a Fourier reconstruction for $\alpha = 2$ and $\alpha = 5$, respectively.

FDRs were generally lower in ex vivo (Figure 6D) compared to in vivo data (Figure 6H). In all cases, the highest levels were reached just above the background detection threshold, and FDRs increased with increasing α: In vivo FDRs reached a maximum of 9%, 12%, and 40% for Fourier, CS $\alpha = 2$, and CS $\alpha = 5$ reconstructions, respectively. Values for all reconstruction methods are shown in Supporting Information Table S1. FDRs dropped with increasing signal level. In vivo CS and denoised reconstructions FDRs converged to 0 at signal levels greater than $3\sigma_F$ and for Fourier reconstructions FDRs approached 0 at $S \geq 4.7\sigma_F$ (Figure 6H). An overlap of the measured FDR variation was observed between all CS reconstructions.

3.5 | Performance of CS with prospective undersampling

In vivo experiments with prospective undersampling confirmed the above results. Compared to conventional Fourier reconstructions, the number of detected ^{19}F signal voxels was greatly enhanced with CS $\alpha = 2$ (Figure 7A). Slight blurring was only present at the edges of true positive features. The number of

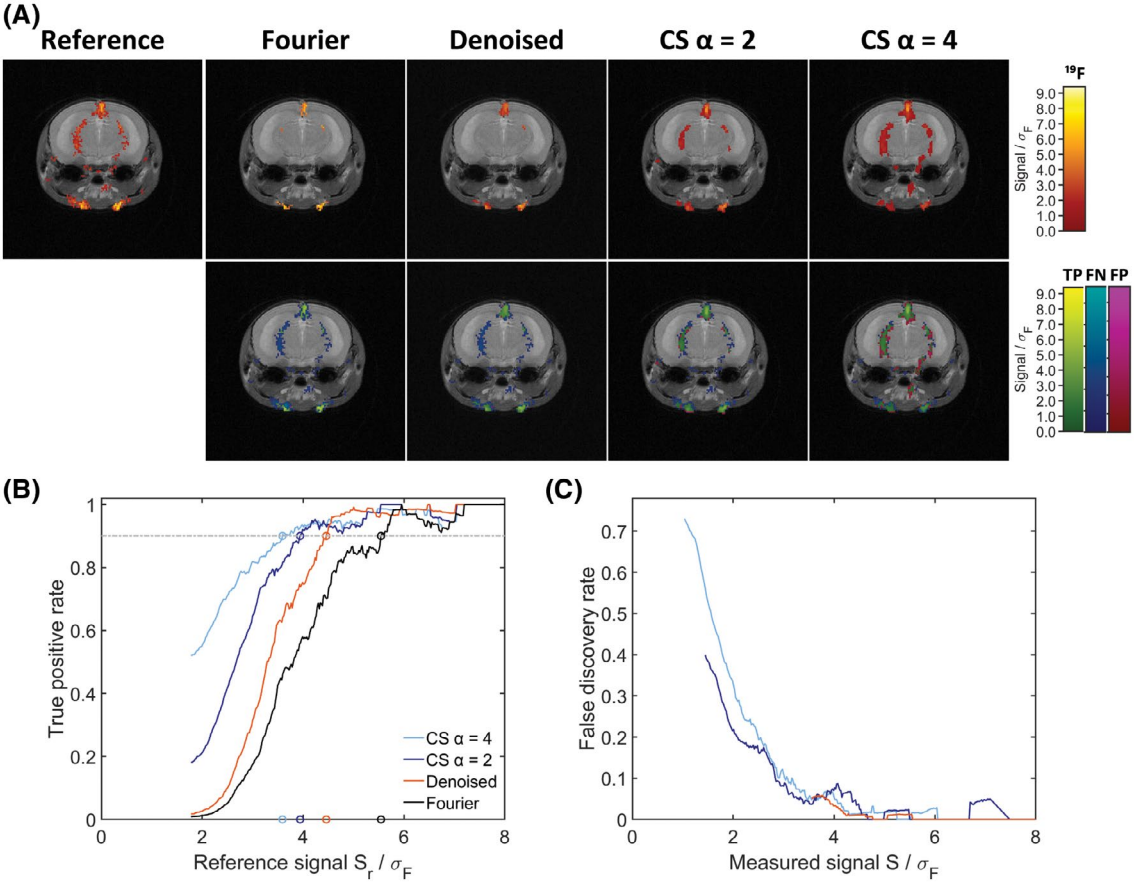

FIGURE 7 Detection performance for ^{19}F MRI in the EAE mouse brain using prospective undersampling. (A) Representative slice comparing conventional Fourier and CS reconstructions (20 minutes' measurement time) with the reference (Fourier reconstruction, 80 minutes' measurement time). The first row shows an overlay of the ^{19}F signal (red hot), normalized by the noise standard deviation of the Fourier reconstruction σ_F, on an anatomical image. The second row shows true positives (TP) in green/yellow, false negatives (FN) in blue/turquoise, and false positives (FP) in red/violet. Panels (B) and (C) show quantitative results for all 4 data sets acquired with prospective undersampling. (B) TPRs computed for different levels of the reference signal S_r using a sliding window approach. Circles indicate the signal level where a TPR of 90% is reached, and σ_F is the noise standard deviation of the Fourier reconstruction at equal measurement time. (C) FDRs computed for different levels of the measured signal S

true positives was further increased with $\alpha = 4$, but so was the blurring effect. CS showed improved TPRs at all signal levels (Figure 7B). A TPR of 90% was reached at $5.5\sigma_F$ (Fourier), $4.5\sigma_F$ (denoised), $3.9\sigma_F$ (CS $\alpha = 2$), and $3.6\sigma_F$ (CS $\alpha = 4$). FDRs in CS reconstructions, on the other hand, were elevated at low signal levels, especially for $\alpha = 4$ (Figure 7C). All reconstructions of prospectively undersampled data can be found under: central.xnat.org/data/projects/CSperf_19F-MRI.

3.6 | Signal intensity bias in animal experiments

Deviation of measured signal from reference signal was computed over an SI range for the EAE data (Figure 8). The SI bias ex vivo and in vivo (Figure 8) showed patterns similar to the observations in phantoms (Figure 5D). Fourier reconstructions overestimated SI by 10% ($S = 8\sigma_F$) to 35% ($S = 3.5\sigma_F$; Figure 8A); denoised (Figure 8B), and CS (Figure 8C–F) reconstructions underestimated the ^{19}F signal. CS reconstructions underestimated the signal by ~30% and denoised reconstructions by ~40% (Figure 8B–F). For all reconstructions, deviations increased with decreasing SI. The largest downward deviations were observed in CS $\alpha = 4$ and $\alpha = 5$. In vivo and ex vivo results showed a similar bias, except for CS reconstructions at lower SI. All CS reconstructions showed a similar bias at signal levels $S < 3\sigma_F$. At $S < 4\sigma_F$, the signal bias in CS reconstructions was up to 10% smaller in ex vivo than in in vivo data.

Using the calibration method for correcting in vivo images (Figure 3), bias was reduced for both conventional (Figure 8A,B) and CS (Figure 8C–F) reconstructions. Average signal deviation did not exceed 10% in the corrected reconstructions. Variability remained unchanged, and the best accuracy was achieved in cases where ex vivo and uncorrected in vivo data showed similar biases. In corrected Fourier (Figure 8A), denoised (Figure 8B), and CS $\alpha = 2$ (Figure 8C) in vivo data, deviations did not exceed the result variability at any signal level. For CS $\alpha = 3$, 4, and 5 (Figure 8D–F), correction reduced bias in the in vivo data to <5% for measured signals $S > 5.5\sigma_F$. For lower signal levels, the larger discrepancies between ex vivo and in vivo data resulted in signal deviations of ~10% in the in vivo data following correction.

Nearly all voxels in Fourier reconstructions show an overestimated ^{19}F signal, whereas most voxels in CS reconstructions show an underestimated signal before correction (Figure 9B, upper panel). Correction resulted in a more balanced distribution of the observed SI around the correct value (Figure 9B, lower panel). When studying NP concentrations measured in the reference data (Figure 9A), we observed that for both in vivo and ex vivo conditions, the occurrence

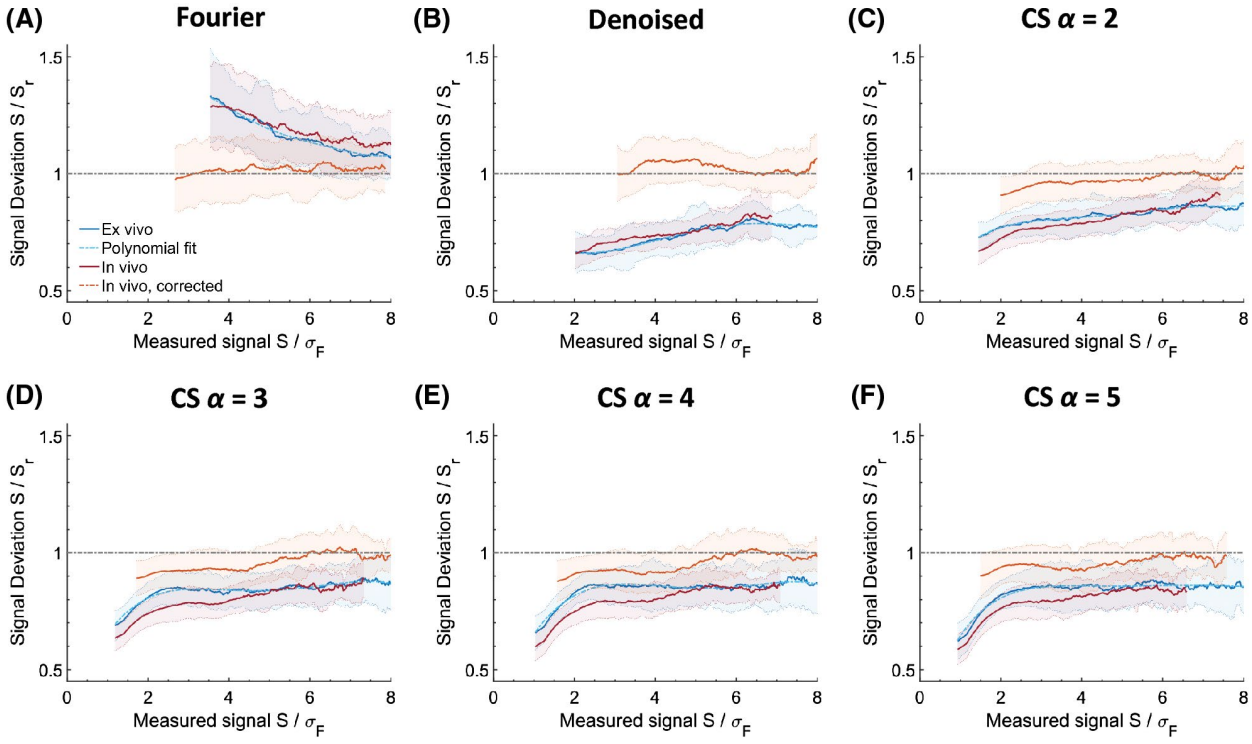

FIGURE 8 Signal intensity bias for imaging inflammation in the EAE mouse model (setups 2 and 3). (A–F) Average signal deviations for Fourier, denoised, and CS reconstruction. α denotes the factor of undersampling and averaging. Each subfigure gathers results for 1 acquisition and reconstruction method and contains the signal deviation measured in the ex vivo condition (setup 2), a polynomial fit of this measurement, the signal deviation measured in vivo (setup 2), and results for the corrected in vivo data (see also Figure 3 for an illustration of the calibration method). A legend is shown in (A), and the shaded areas mark the variability of the results

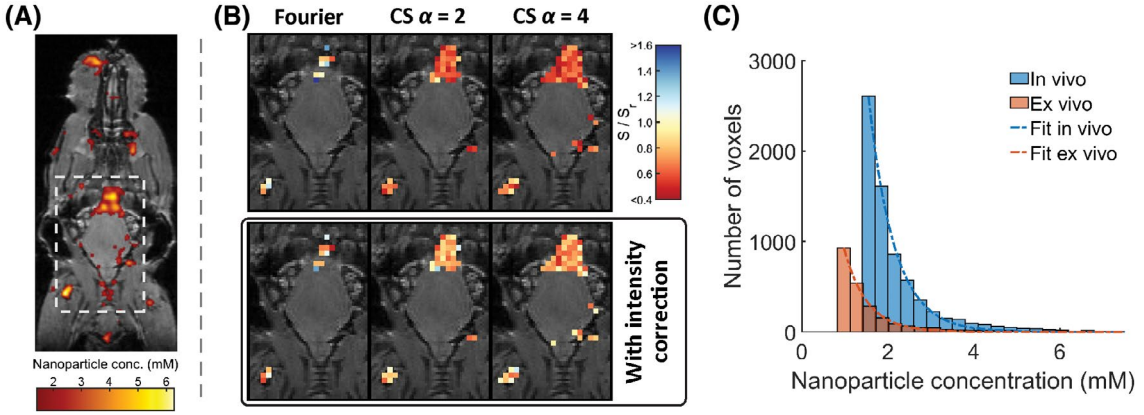

FIGURE 9 Signal intensity calibration example and NP concentration rate of occurrence. (A) Nanoparticle concentration measured in the reference (in vivo). (B) Deviation of the reconstructed SI from the reference before correction (upper panel) and following correction (lower panel). Only true positive voxels are shown, and the cutout is marked in (A). Blue color indicates overestimation, and red color indicates underestimation. (C) Histogram of NP concentrations measured in the references of ex vivo and in vivo data sets (setups 2 and 3). The dashed lines show fits of the negative exponential distribution with rate parameters $\varphi = 1.89\text{mM}^{-1}$ and 1.68mM^{-1} for ex vivo and in vivo data, respectively

rates closely followed a negative exponential distribution that decayed with increasing signal level (Figure 9C).

4 | DISCUSSION AND CONCLUSION

Our study shows that CS can be successfully applied to reconstruct small, low-SNR ^{19}F MR features without an a priori known location. The automatic regularization strength selection we propose reconstructed high-quality images without the need of any manual input. We demonstrated that CS improves detection performance and provides reliable information, not only for high-SNR signals, but even for small image features close to the detection threshold. Although previous work had shown that CS has potential for ^{19}F MRI, our results constitute the first thorough validation of CS for a challenging ^{19}F MRI application. In order to maintain reliability, the undersampling factor and background subtraction threshold need to be chosen with care. Negative SI bias is inherent to CS[35,44] and poses a challenge for quantitative ^{19}F MRI.[17,45] However, and quite unexpectedly, Fourier reconstructions also showed an SI bias with a similar magnitude. The intensity calibration we propose provides a strategy to obtain close to unbiased results, yet it necessitates additional MR reference measurements and postprocessing.

We observed that CS achieved a favorable trade-off between undersampling losses and averaging gains whenever image features were close to the detection threshold. The undersampling patterns that we used focus on the low spatial frequencies, which contain the bulk of contrast information. Allocating a larger share of acquisition time to these data points and thus measuring them with higher SNR lifts features above the noise background. Missing high-spatial-frequency information did not introduce false positives in phantom experiments, even with 8-fold undersampling. This is probably attributable to the extreme sparsity of the images. In the animal model data, reduced sparsity and more complex-shaped features led to slight blurring at the feature edges for $\alpha > 2$. The blurring increased with the degree of undersampling. Higher FDRs in the in vivo compared to ex vivo experiments can be attributed to movement artifacts and shorter reference data measurement times. Given that the detection performance consistently increased with higher α, operators must decide on a balance between sensitivity and data fidelity. For in vivo studies similar to the present one, we recommend 2- to 3-fold undersampling. It should be noted that FDRs dropped to 0 or close to 0 rapidly with increasing SI, so that higher background subtraction thresholds can always be chosen post hoc to obtain more conservative results. At SNRs that were high enough for all features to be detected regardless of the method, only minor differences in image quality were observed between Fourier and CS reconstructions. These SNRs are practically not reached in challenging in vivo ^{19}F MR applications.

The large number of reconstructions (>2000) of ex vivo and in vivo measurements leading to these conclusions guarantees that the observed improvment is not attributable to randomness. Our in vivo experiments with prospective undersampling showed good agreement with results from retrospectively undersampled data and demonstrate the applicability of CS for preclinical imaging studies. The automatic regularization strength selection we used provides a method to optimize CS reconstructions based on objective criteria that yields high-quality images. Because of the small matrix typical of ^{19}F MR images, the increased reconstruction time does not pose practical problems. Required knowledge of the noise level can easily be made available by performing a pure noise scan,[39] which should also be executed to facilitate data analysis. It is a recognized problem of CS-MRI that regularization parameter

values are not reported or the choice is not sufficiently substantiated.[32] Automatic selection provides independence from manual adjustments, making CS more accessible to nonexpert users and improving the reproducibility and comparability of results.

Contrary to our expectation,[17] we observed a positive SI bias in Fourier reconstructions despite Rician noise bias correction. This bias is explained by the exponentially decaying rate of occurrence of NP concentrations (Figure 9C): Assuming a negative exponential prior distribution over the true SI (S_t) with rate parameter φ and a simplified Gaussian forward model, the expected true SI for measured SI S is $<S_t>_{p(S_t|S)}=S-\frac{\varphi\sigma^2}{2}$, which is smaller than S itself (see the Appendix for details). In CS and denoised reconstructions, this effect reduced the signal underestimation. Our proposed intensity calibration corrected the systematic deviations. Therefore, high-quality ex vivo scans of comparable signal distributions are recommended for in vivo applications utilizing CS and ^{19}F MR for quantitative analyses. Taking not only the SI, but also feature size into account could further improve performance and applicability of the method. This method could potentially also be used in human in vivo studies, but would require long ex vivo reference MR measurements of similar tissue (e.g., from biopsy or autopsy material). Given that similar bias effects would pose a major challenge for future ^{19}F MR studies, further detailed investigations are required to seek and correct discrepancies as early as possible.

Another avenue for future CS research is to offset motion artifacts and long-term drift effects.[46] Although the increased averaging reduces the magnitude of motion artifacts, especially in cardiac and abdominal applications, shortened scan times would reduce the influence of signal drift. However, investigating these highly circumstantial effects would necessitate proper controls, which is challenging, given that parallel acquisitions of reference measurements, even motion correction data, introduce a technical hurdle.

We expect the demonstrated improvement of signal detection without impaired data fidelity or necessarily compromised quantification to hold for other ^{19}F MRI preclinical applications with similar signal distributions. Besides lowering the detection threshold, increased ^{19}F sensitivity can be translated into shorter measurement times or improved spatial resolution. CS will be particularly useful in studies that involve small amounts of ^{19}F, as is the case in pharmacokinetic studies[47] as well as studies involving ^{19}F-target–specific theranostic nanoparticles,[48] ^{19}F/^{1}H MR smart probes,[49,50] or ^{19}F-labeled cells administered as therapies for tumor disease.[2,51]

ACKNOWLEDGMENTS

This study was funded, in part, by the Deutsche Forschungsgemeinschaft to S.W. and A.P. (DFG WA2804, DFG PO1869). This work was funded in part (Thoralf Niendorf, Sonia Waiczies, Andreas Pohlmann) by the German Research Foundation (Gefoerdert durch die Deutsche Forschungsgemeinschaft (DFG), Projektnummer 394046635, SFB 1365, *RENOPROTECTION*. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project number 394046635, SFB 1365, *RENOPROTECTION*). This project has received funding in part (Thoralf Niendorf) from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No 743077 (ThermalMR).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
We thank Stefanie Münchberg for preparation of nanoparticles and assistance with animal handling and Jason Millward (both from the Berlin Ultrahigh Field Facility (B.U.F.F.) at the Max-Delbrueck Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany) for preparation of ex vivo tissue samples.

ORCID

Ludger Starke https://orcid.org/0000-0002-0359-0101
Andreas Pohlmann https://orcid.org/0000-0002-8572-2568
Christian Prinz https://orcid.org/0000-0003-3330-031X
Thoralf Niendorf https://orcid.org/0000-0001-7584-6527
Sonia Waiczies https://orcid.org/0000-0002-9916-9572

REFERENCES

1. Ahrens ET, Flores R, Xu H, Morel PA. In vivo imaging platform for tracking immunotherapeutic cells. *Nat Biotechnol*. 2005; 23:983–987.
2. Ahrens ET, Helfer BM, O'Hanlon CF, Schirda C. Clinical cell therapy imaging using a perfluorocarbon tracer and fluorine-19 MRI. *Magn Reson Med*. 2014;72:1696–1701.
3. Flogel U, Ahrens E. *Fluorine Magnetic Resonance Imaging*. New York, NY: Pan Stanford; 2016.
4. Ruiz-Cabello J, Barnett BP, Bottomley PA, Bulte JW. Fluorine (19F) MRS and MRI in biomedicine. *NMR Biomed*. 2011;24:114–129.
5. Waiczies H, Lepore S, Drechsler S, et al. Visualizing brain inflammation with a shingled-leg radio-frequency head probe for 19F/1H MRI. *Sci Rep*. 2013;3:1280.
6. Waiczies S, Millward JM, Starke L, et al. Enhanced fluorine-19 MRI sensitivity using a cryogenic radiofrequency probe: technical developments and ex vivo demonstration in a mouse model of neuroinflammation. *Sci Rep*. 2017;7:9808.
7. Waiczies S, Rosenberg JT, Kuehne A, et al. Fluorine-19 MRI at 21.1 T: enhanced spin–lattice relaxation of perfluoro-15-crown-5-ether and sensitivity as demonstrated in ex vivo murine neuroinflammation. *MAGMA*. 2019;32:37–49.
8. Mastropietro A, De Bernardi E, Breschi GL, et al. Optimization of rapid acquisition with relaxation enhancement (RARE) pulse sequence parameters for (1)(9)F-MRI studies. *J Magn Reson Imaging*. 2014;40:162–170.
9. Constantinides C, Maguire M, McNeill E, et al. Fast, quantitative, murine cardiac 19F MRI/MRS of PFCE-labeled progenitor stem cells and macrophages at 9.4T. *PLoS ONE*. 2018;13:e0190558.
10. Faber C, Schmid F. Pulse sequence considerations and schemes. In: Flögel U, Ahrens E, eds. *Fluorine Magnetic Resonance Imaging*. New York, NY: Pan Stanford; 2016:1–28.

11. Candes EJ, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inf Theory*. 2006;52:489–509.
12. Donoho DL. Compressed sensing. *IEEE Trans Inf Theory*. 2006;52:1289–1306.
13. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med*. 2007;58:1182–1195.
14. Kampf T, Fischer A, Basse-Lüsebrink TC, et al. Application of compressed sensing to in vivo 3D 19F CSI. *J Magn Reson*. 2010;207:262–273.
15. Kampf T, Sturm VJF, Basse-Lusebrink TC, et al. Improved compressed sensing reconstruction for 19F magnetic resonance imaging. *MAGMA*. 2019;32:63–77.
16. Fischer A, Basse-Lüsebrink T, Kampf T, et al. Improved sensitivity in 19F cellular imaging using non-convex compressed sensing. Paper presented at: Proceedings of the 17th Annual Meeting of the ISMRM; 2009; Honolulu, HI. Abstract 3377.
17. Zhong J, Mills PH, Hitchens TK, Ahrens ET. Accelerated fluorine-19 MRI cell tracking using compressed sensing. *Magn Reson Med*. 2013;69:1683–1690.
18. Liang S, Dresselaers T, Louchami K, Zhu C, Liu Y,Himmelreich U. Comparison of different compressed sensing algorithms for low SNR (19) F MRI applications-Imaging of transplanted pancreatic islets and cells labeled with perfluorocarbons. *NMR Biomed*. 2017;30. https://doi.org/10.1002/nbm.3776.
19. Sadykova D, James AP. Quality assessment metrics for edge detection and edge-aware filtering: a tutorial review. Paper presented at: Sixth International Conference on Advances in Computing, Communications and Informatics (ICACCI-2017), Manipal, India, 2017.
20. Yu Z, Wang J, Lu G. Optimized self-adapting contrast enhancement algorithm for wafer contour extraction. *Multimed Tools Appl*. 2019;78:32087–32108.
21. Zhang L, Xing M, Qiu C-W, et al. Resolution enhancement for inversed synthetic aperture radar imaging under low SNR via improved compressive sensing. *IEEE Trans Geosci Remote Sens*. 2010;48:3824–3838.
22. Ji S, Xue Y, Carin L. Bayesian compressive sensing. *IEEE Trans Sig Process*. 2008;56:2346–2356.
23. Morozov VA. On the solution of functional equations by the method of regularization. Paper presented at: Dokl Akad Nauk SSSR; 1966; Moscow, USSR.
24. Kilmer ME, O'Leary DP. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM J Matrix Anal Appl*. 2001;22:1204–1221.
25. Ahrens ET, Bulte JW. Tracking immune cells in vivo using magnetic resonance imaging. *Nat Rev Immunol*. 2013;13:755–763.
26. Waiczies H, Lepore S, Janitzek N, et al. Perfluorocarbon particle size influences magnetic resonance signal and immunological properties of dendritic cells. *PLoS ONE*. 2011;6:e21981.
27. Flögel U, Ding Z, Hardung H, et al. In vivo monitoring of inflammation after cardiac and cerebral ischemia by fluorine magnetic resonance imaging. *Circulation*. 2008;118:140–148.
28. Jacoby C, Borg N, Heusch P, et al. Visualization of immune cell infiltration in experimental viral myocarditis by (19)F MRI in vivo. *MAGMA*. 2014;27:101–106.
29. Lassmann H. Multiple Sclerosis Pathology. *Cold Spring Harb Perspect Med*. 2018;8:a028936.
30. Qaisar S, Bilal RM, Iqbal W, Naureen M, Lee S. Compressive sensing: from theory to applications, a survey. *J Commun Netw*. 2013;15:443–456.
31. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D*. 1992;60:259–268.
32. Jaspan ON, Fleysher R, Lipton ML. Compressed sensing MRI: a review of the clinical literature. *Br J Radiol*. 2015;88:20150487.
33. Goldstein T, O'Donoghue B, Setzer S, Baraniuk R. Fast alternating direction optimization methods. *SIAM J Imaging Sci*. 2014;7:1588–1623.
34. Chen M. On the solution of circulant linear systems. *SIAM J Numer Anal*. 1987;24:668–683.
35. Osher S, Yin W. *Sparse recovery via l1 and L1 optimization*. Los Angeles, CA: California Univ. Los Angeles Dept. of Mathematics; 2014.
36. Dowell M, Jarratt P. A modified regula falsi method for computing the root of an equation. *BIT Numer Math*. 1971;11:168–174.
37. Lepore S, Waiczies H, Hentschel J, et al. Enlargement of cerebral ventricles as an early indicator of encephalomyelitis. *PLoS ONE*. 2013;8:e72841.
38. Waiczies S, Lepore S, Sydow K, et al. Anchoring dipalmitoyl phosphoethanolamine to nanoparticles boosts cellular uptake and fluorine-19 magnetic resonance signal. *Sci Rep*. 2015;5:8427.
39. National Electrical Manufacturers Association. *Determination of signal-to-noise ratio (SNR) in diagnostic magnetic resonance imaging*. NEMA Standards Publication MS 1–2001. Arlington, VA: National Electrical Manufacturers Association; 2001.
40. Zijlstra F, Viergever MA, Seevinck PR. Evaluation of variable density and data-driven k-space undersampling for compressed sensing magnetic resonance imaging. *Invest Radiol*. 2016;51:410–419.
41. Henkelman RM. Measurement of signal intensities in the presence of noise in MR images. *Med Phys*. 1985;12:232–233.
42. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. *Magn Reson Med*. 1995;34:910–914.
43. Constantinides CD, Atalar E, McVeigh ER. Signal-to-noise measurements in magnitude images from NMR phased arrays. *Magn Reson Med*. 1997;38:852–857.
44. Becker SR. *Practical compressed sensing: modern data acquisition and signal processing*. Pasadena, CA: California Institute of Technology; 2011.
45. Hu S, Lustig M, Balakrishnan A, et al. 3D compressed sensing for highly accelerated hyperpolarized 13C MRSI with in vivo applications to transgenic mouse models of cancer. *Magn Reson Med*. 2010;63:312–321.
46. Vos SB, Tax CM, Luijten PR, Ourselin S, Leemans A, Froeling M. The importance of correcting for signal drift in diffusion MRI. *Magn Reson Med*. 2017;77:285–299.
47. Bolo NR, Hode Y, Macher JP. Long-term sequestration of fluorinated compounds in tissues after fluvoxamine or fluoxetine treatment: a fluorine magnetic resonance spectroscopy study in vivo. *MAGMA*. 2004;16:268–276.
48. Pan H, Myerson JW, Hu L, et al. Programmable nanoparticle functionalization for in vivo targeting. *FASEB J*. 2013;27:255–264.
49. Keliris A, Mamedov I, Hagberg GE, Logothetis NK, Scheffler K, Engelmann J. A smart 19F and 1H MRI probe with self-immolative linker as a versatile tool for detection of enzymes. *Contrast Media Mol Imaging*. 2012;7:478–483.
50. Bar-Shir A, Yadav NN, Gilad AA, van Zijl PC, McMahon MT, Bulte JW. Single 19F probe for simultaneous detection of multiple metal ions using miCEST MRI. *J Am Chem Soc*. 2015;137:78–81.
51. Ku M-C, Edes I, Bendix I, et al. ERK1 as a therapeutic target for dendritic cell vaccination against high-grade gliomas. *Mol Cancer Ther*. 2016;15:1975–1987.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

TABLE S1 Detection performance in the EAE model. Ex vivo and in vivo refer to setups 2 and 3, respectively. Rows 2 and 3 show the reference signal level at which a TPR of 90% is reached. Rows 4 and 5 show the FDR at signal levels just above the background subtraction threshold and rows 6 and 7 the overall FDR

How to cite this article: Starke L, Pohlmann A, Prinz C, Niendorf T, Waiczies S. Performance of compressed sensing for fluorine-19 magnetic resonance imaging at low signal-to-noise ratio conditions. *Magn Reson Med*. 2020;84:592–608. https://doi.org/10.1002/mrm.28135

APPENDIX SIGNAL INTENSITY BIAS IN FOURIER RECONSTRUCTIONS

Following Figure 9C, we assume that the SI rate of occurrence follows an exponential distribution. Thus, the prior distribution over the true signal intensities is (Equation A1):

$$p_0\left(S_t\right)=\varphi\exp\left(-\varphi S_t\right), \tag{A1}$$

with rate parameter φ. We consider a simplified forward model with additive Gaussian noise (Equation A2):

$$p\left(S|S_t\right)=N\left(S;S_t,\sigma\right)=\frac{1}{\sqrt{2\sigma^2}}\exp\left(-\frac{\left(S-S_t\right)^2}{2\sigma^2}\right), \tag{A2}$$

where S denotes the measured signal and σ the noise standard deviation. The posterior distribution follows from the Bayes theorem (Equation A3):

$$\begin{aligned} p\left(S_t|S\right)\propto p\left(S|S_t\right)p_0\left(S_t\right) \quad &|\propto\exp\left(-\frac{\left(S-S_t\right)^2}{2\sigma^2}-\varphi S_t\right)\\ &|\propto\exp\left(-\frac{S^2-2S_t\left(S-\frac{\varphi\sigma^2}{2}\right)}{2\sigma^2}\right)\\ &|\propto N\left(S_t;S-\frac{\varphi\sigma^2}{2},\sigma^2\right). \end{aligned} \tag{A3}$$

Thus, the expected true signal for measured signal S is (Equation A4):

$$<S_t>_{p(S_t|S)}=S-\frac{\varphi\sigma^2}{2}. \tag{A4}$$