FULL PAPER

Automated characterization of noise distributions in diffusion MRI data

Samuel St-Jean[1] | Alberto De Luca[1] | Chantal M. W. Tax[2] | Max A. Viergever[1] | Alexander Leemans[1]

[1]Image Sciences Institute, Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands

[2]Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, United Kingdom

Correspondence
Samuel St-Jean, Image Sciences Institute, Department of Radiology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands
Email: samuel@isi.uu.nl

Funding information
Fonds de recherche du Québec - Nature et technologies (FRQNT) (Dossier 192865); Netherlands Organization for Scientific Research (NWO), Grant/Award Number: VIDI 639.072.411; Netherlands Organization for Scientific Research (NWO), Grant/Award Number: Rubicon 680-50-1527; Wellcome Trust grant 096646/Z/11/Z

Purpose: To understand and characterize noise distributions in parallel imaging for diffusion MRI.

Theory and Methods: Two new automated methods using the moments and the maximum likelihood equations of the Gamma distribution were developed. Simulations using stationary and spatially varying non-central chi noise distributions were created for two diffusion weightings with SENSE or GRAPPA reconstruction and 8, 12 or 32 receiver coils. Furthermore, MRI data of a water phantom with different combinations of multiband and SENSE acceleration were acquired on a 3T scanner along with noise-only measurements. Finally, an in vivo dataset was acquired at 3T using multiband acceleration and GRAPPA reconstruction. Estimation of the noise distribution was performed with the proposed methods and compared with 3 other existing algorithms.

Results: Simulations showed that assuming a Rician distribution can lead to misestimation in parallel imaging. Results on the acquired datasets showed that signal leakage in multiband can lead to a misestimation of the parameters. Noise maps are robust to these artifacts, but may misestimate parameters in some cases. The proposed algorithms herein can estimate both parameters of the noise distribution, are robust to signal leakage artifacts and perform best when used on acquired noise maps.

Conclusion: Misestimation of the correct noise distribution can hamper further processing such as bias correction and denoising, especially when the measured distribution differs too much from the actual signal distribution e.g., due to artifacts. The use of noise maps can yield more robust estimates than the use of diffusion weighted images as input for algorithms.

KEYWORDS

Diffusion MRI, Noise estimation, Parallel acceleration, Gamma distribution, GRAPPA, SENSE

1 | INTRODUCTION

Diffusion magnetic resonance imaging (dMRI) is a non invasive imaging technique which allows probing microstructural properties of living tissues. Advances in parallel imaging techniques (1, 2), such as accelerated acquisitions (e.g., partial k-space (3), multiband imaging (4, 5) and compressed sensing (6, 7)), have greatly reduced the inherently long scan time in dMRI. While it is known that the noise distribution found in magnitude dMRI data depends on the reconstruction algorithm used and the number of channels in the receiver coils (8, 9), noise correlation effects in adjacent channels change the noise distribution from its theoretical formulation. New acquisition methods and pulse sequences in dMRI are also pushing the limits of spatial resolution while reducing scan time (10), which also affects the signal distribution in ways that are challenging to model. Estimation of signal distributions deviating from theoretical cases is challenging and oftentimes requires information such as coil sensitivities or reconstruction matrices. This information may not be recorded at acquisition time or is even not available from the scanner, making techniques relying on these parameters difficult to apply in practice.

With the recent trend towards open data sharing and large multicenter studies using standardized protocols (11, 12), differences in hardware, acquisition or reconstruction algorithms may inevitably lead to different signal distributions. This may affect large scale longitudinal studies investigating neurological changes due to these "scanner effects" (13) as the acquired data may be fundamentally different across sites in terms of statistical properties of the signal. Algorithms have been developed to mitigate these potential differences (14, 15), but characterization of the signal distribution from various scanners is challenging due to the black box nature of the acquisition process, especially in routine clinical settings. While some recent algorithms for dMRI are developed to include information about the noise distribution (16, 17), there is no method, to the best of our knowledge, providing a fully automated way to characterize the noise distribution using information from the magnitude data itself only. Due to this gap between the physical acquisition process and noise estimation theory, noise distributions are either assumed as Rician (with parameter σ_g related to the standard deviation) or noncentral chi (with fixed degrees of freedom N) and concentrate in estimating the noise standard deviation σ_g (18–20). This assumption inevitably leads to misestimation of the true signal distribution and may impact subsequent steps, for instance, bias correction (21), denoising (22) or diffusion model estimation (17, 23, 24). In this work, we propose to estimate the parameters σ_g and N from either the magnitude data or the acquired noise maps by using a change of variable to a Gamma distribution $Gamma(N, 1)$ (19), whose first moments and maximum likelihood equations directly depend on N. This makes the proposed method fast and easy to apply to existing data without additional information, while being robust to artifacts by rejecting outliers of the distribution.

Preliminary results of this work have been presented at the annual meeting of the MICCAI (25). This manuscript now contains additional theory, simulations including signal correlations and parallel acceleration, and experiments on phantoms and in vivo datasets acquired with parallel and multiband acceleration. As example applications, we perform bias correction and denoising on the in vivo dataset using the estimated distribution derived with each algorithm.

2 | THEORY

In this section, we introduce the necessary background on the Gamma distribution, its moments and maximum likelihood equations. Expressing the signal with a Gamma distribution highlights equations which can be solved to estimate parameters σ_g and N.

2.1 | Probability distribution functions of MRI data

To account for uncertainty in the acquisition process, the complex signal measured in k-space by the receiver coil array can be modeled with a separate additive zero mean Gaussian noise for each channel with identical variance σ_g^2 (26). After magnitude reconstruction, the resulting noise distribution can be represented by a Rician or noncentral chi distribution with its parameters depending on the employed reconstruction algorithm (8). Even though the magnitude signal model is still valid nowadays, the use of image filters (8), acceleration methods subsampling k-space (1, 2, 27) and spatial correlation between coil elements (8, 28) influence, amongst other factors, the parameters of the resulting signal distribution. This can be modeled by considering spatially varying degrees of freedom N_{eff} and standard deviation σ_{eff} (also called the *effective* values) and we generally have $N_{eff} \leq N$, (8, 28).

In the general case, the magnitude signal intensity m can be modeled as a noncentral chi distribution, parametrized by the (unknown) noiseless signal value η, number of coils N and noise standard deviation σ_g. The noncentral chi distribution includes the Rician ($N = 1$), the Rayleigh ($N = 1, \eta = 0$) and the central chi distribution ($\eta = 0$) as special cases (8). The probability density function (pdf) of the central chi distribution is given by

$$pdf(m|\eta = 0, \sigma_g, N) = \frac{m^{2N-1}}{2^{N-1}\sigma_g^{2N}\Gamma(N)} \exp\left(\frac{-m^2}{2\sigma_g^2}\right) dm, \quad [1]$$

where $\Gamma(x)$ is the Gamma function. With a change of variable introduced by (19), Eq. (1) can be rewritten as a Gamma distribution $Gamma(N, 1)$ with $t = m^2/2\sigma_g^2$, $dt = m/\sigma_g^2 dm$ which has a pdf given by

$$pdf(t|N) = \frac{1}{\Gamma(N)} t^{N-1} \exp(-t) dt. \quad [2]$$

Eq. (2) only depends on N, which can be estimated from the sample values.

2.2 | Parameter estimation using the method of moments and maximum likelihood

The method of moments

The pdf of $Gamma(\alpha, \beta)$ is defined as

$$pdf(x|\alpha, \beta) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp(-x/\beta) dx \quad [3]$$

and has mean μ_{gamma} and variance σ_{gamma}^2 given by

$$\mu_{gamma} = \alpha\beta, \; \sigma_{gamma}^2 = \alpha\beta^2. \quad [4]$$

Another useful identity comes from the sum of Gamma distributions, which is also a Gamma distribution (29) such that if $t_i \sim Gamma(\alpha_i, \beta)$, then

$$\sum_{i=1}^{K} t_i \sim Gamma\left(\sum_{i=1}^{K} \alpha_i, \beta\right) \quad [5]$$

From Eq. (4), we obtain that the mean and the variance of the distribution $Gamma(N, 1)$ are in fact equal and of value N. That is, we can estimate the Gaussian noise standard deviation σ_g and the number of coils N from the sample

moments of the magnitude images themselves, provided we can select voxels without any signal contribution where $\eta = 0$. Firstly, σ_g can be estimated from Eq. (4) as

$$\sigma_g = \frac{1}{\sqrt{2}}\sqrt{\frac{\sum_{v=1}^{V} m_v^4}{\sum_{v=1}^{V} m_v^2} - \frac{1}{V}\sum_{v=1}^{V} m_v^2}, \qquad [6]$$

where V is the number of identified noise only voxels and m_v the value of such a voxel, see Supplementary Material A.1 for the derivations. Once σ_g is known, N can be estimated from the sample mean of those previously identified voxels as

$$N = \frac{1}{V}\sum_{v=1}^{V} t_v = \frac{1}{2V\sigma_g^2}\sum_{v=1}^{V} m_v^2. \qquad [7]$$

Derivations of Eqs. (6) and (7) are detailed in Supplementary Material A.1.

Maximum likelihood equations for the Gamma distribution

Estimation based on the method of maximum likelihood yields two equations for estimating α and β. Rearranging the equations for a Gamma distribution will give Eq. (7) and a second implicit equation for N that is given by (30)

$$\log(\beta) + \psi(\alpha) = \frac{1}{V}\sum_{v=1}^{V} \log t_v, \qquad [8]$$

where $\psi(x) = \frac{d}{dx}\log(\Gamma(x))$ is the digamma function. For the special case $Gamma(N, 1)$, we can rewrite Eq. (8) as

$$\psi(N) = \frac{1}{V}\sum_{v=1}^{V} \log(m_v^2/2\sigma_g^2). \qquad [9]$$

Combining Eq. (7) and Eq. (9), we also have an implicit equation to find σ_g

$$\psi\left(\frac{1}{2V\sigma_g^2}\sum_{v=1}^{V} m_v^2\right) = \frac{1}{V}\sum_{v=1}^{V} \log(m_v^2/2\sigma_g^2) = \frac{1}{V}\sum_{v=1}^{V} \log(m_v^2) - \log(2\sigma_g^2). \qquad [10]$$

As Eqs. (9) and (10) have no closed form solution, they can be solved numerically e.g., using Newton's method. See Supplementary Material A.1 for practical implementation details.

3 | METHODS

3.1 | Automated and robust background separation

The equations we presented in Section 2.1 are only valid when $\eta = 0$ by construction and assume that each selected voxel m_v belongs to the same Gamma distribution. For simplicity, we assume that each 2D slice with the same spatial location belongs to the same statistical distribution throughout each 3D volume. Following a methodology similar to (19) and using Eq. (5), the sum of all DWIs can be used to separate the voxels belonging to the Gamma distribution $Gamma(KN, 1)$, where K is the number of acquired DWIs, from the voxels not in that specific distribution with a rejec-

tion step using the inverse cumulative distribution function[1](cdf). In the particular case $Gamma(KN, 1)$ at a probability level p, the inverse cdf is $icdf(\alpha, p) = P^{-1}(\alpha, p)$, where P^{-1} is the inverse lower incomplete regularized gamma function[2]. This relationship can be used to identify potential outliers, such as voxels which contain non background signal, by excluding any voxel m_v whose value does not fall between $\lambda_- = icdf(\alpha, p/2)$ and $\lambda_+ = icdf(\alpha, 1 - p/2)$, i.e., m_v is an outlier if $m_v < \lambda_-$ or $m_v > \lambda_+$.

To provide a better understanding of the change of variable $t = m^2/2\sigma_g^2$, Fig. 1 shows the histogram for a synthetic dataset at b = 3000 s/mm^2, which will be detailed later in Section 3.2. Voxels belonging to the background are easily separated in terms of the Gamma distribution after transformation, thus allowing estimation of parameters from voxels truly belonging to the noise distribution, see Supplementary Material A.2 and (25) for technical details. Our implementation of the proposed algorithm is freely available[3].

3.2 | Datasets and experiments

Synthetic phantom datasets

Two synthetic phantom configurations from previous dMRI challenges were used. The first simulations were based on the ISBI 2013 HARDI challenge[4] using phantomas (31). We used the given 64 gradient directions to generate two separate noiseless single-shell phantoms with either b = 1000 s/mm^2 or b = 3000 s/mm^2 and an additional b = 0 s/mm^2 volume. The datasets were then corrupted with Rician ($N = 1$) and noncentral chi noise profiles ($N = 4, 8$ and 12), both stationary and spatially varying, at a signal-to-noise ratio (SNR) of 30 according to

$$\hat{I} = \sqrt{\sum_{i=0,j=0}^{N} \left(\frac{I}{\sqrt{N}} + \tau\epsilon_i\right)^2 + (\tau\epsilon_j)^2}, \quad \text{where } \epsilon_i, \epsilon_j \sim \mathbb{N}(0, \sigma_g^2), \qquad [11]$$

where I is the noiseless volume, $\hat{I}$ is the resulting noisy volume, τ is a mask for the spatial noise pattern, $\mathbb{N}(0, \sigma_g^2)$ is a Gaussian distribution of mean 0 and variance $\sigma_g^2 = (\bar{m}/SNR)^2$ and $\bar{m} = (1/V \sum_{v=1}^{V} m_v)$ with m_v each voxel of the b = 0 s/mm^2 image inside the white matter. In the stationary noise case, τ is set to 1 so that the noise is uniform. For the spatially varying noise case, τ is a sphere with a value of 1 in the center up to a value of 1.75 at the edges of the phantom, thus generating a stronger noise profile outside the phantom than for the stationary noise case. Since all datasets are generated at SNR 30, the noise standard deviation σ_g is the same even though the b-value or number of coils N is different, but the magnitude standard deviation σ_{m_N} is lower than σ_g.

The second set of synthetic experiments is based on the ISMRM 2015 tractography challenge (32) which consists of 25 manually delineated white matter bundles. Ground truth data consisting of 30 gradient directions at either b = 1000 s/mm^2 or b = 3000 s/mm^2 and 3 b = 0 s/mm^2 images at a resolution of 2 mm isotropic was generated using fiberfox (33) without artifacts or subject motion. Subsequent noisy datasets were created at SNR 20 by simulating an acquisition with 8, 12 and 32 coils using the parallel MRI simulation toolbox[5] with SENSE (1) or GRAPPA (2) reconstructions with an acceleration factor of $R = 2$. The SENSE simulated datasets also included spatial correlations between coils of $\rho = 0.1$, increasing the spatially varying effective noise standard deviation σ_g and keeping the signal Rician distributed ($N = 1$). For the GRAPPA reconstructed datasets, 32 calibrating lines were sampled in the k-space

[1] The inverse cdf is also known as the quantile function.

[2] As there is no analytical solution to the inverse cdf of a Gamma distribution, one can use the function $gaminv(p, \alpha, \beta = 1)$ in Matlab or $InverseGammaRegularized(\alpha, 1 - p)$ in Mathematica to numerically estimate it.

[3] https://github.com/samuelstjean/autodmri

[4] http://hardi.epfl.ch/static/events/2013_ISBI/

[5] https://mathworks.com/matlabcentral/fileexchange/36893-parallel-mri-noisy-phantom-simulator

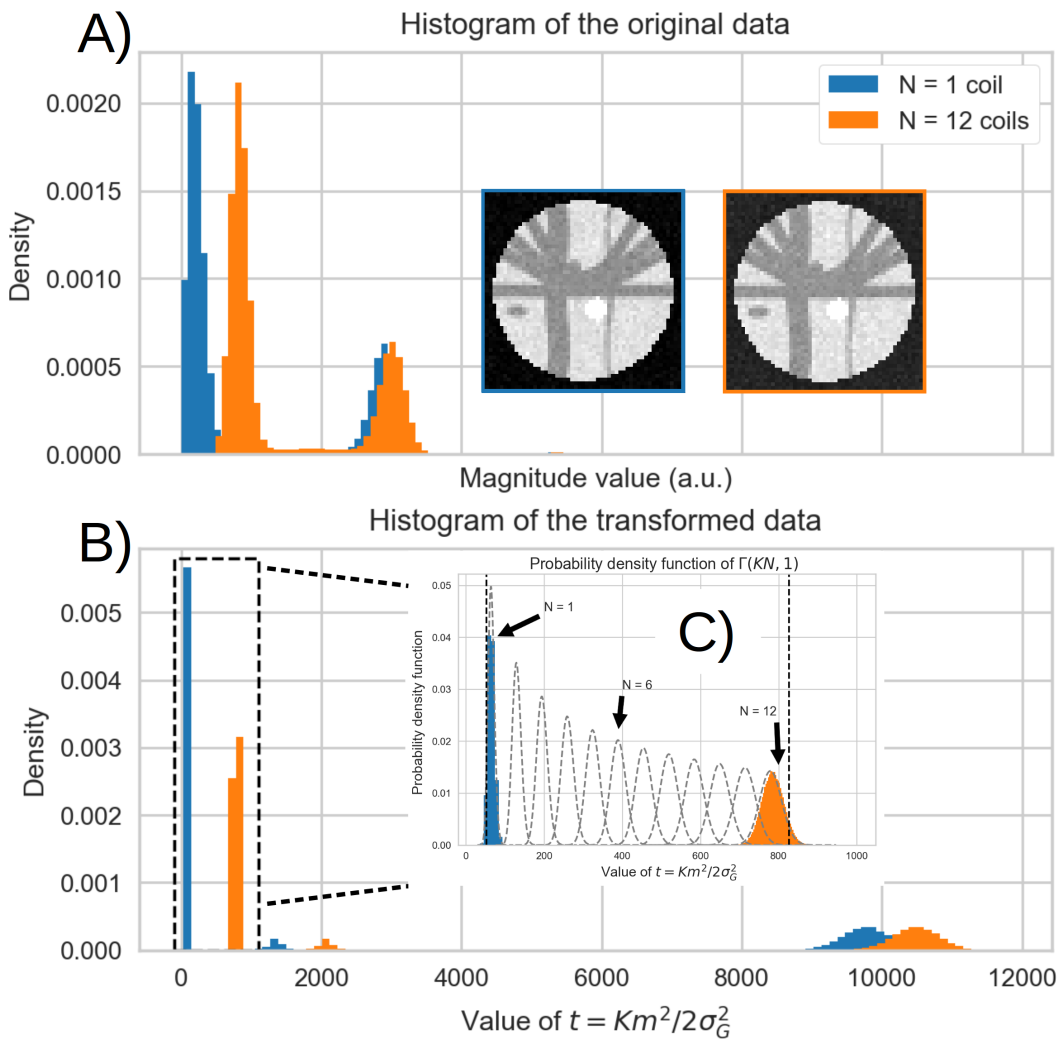

FIGURE 1 Histogram of the synthetic data at b = 3000 s/mm^2 **A)** before the change of variable to a Gamma distribution and **B)** after the change of variable to a Gamma distribution for $N = 1$ and $N = 12$. Summing all K DWIs together separates the background voxels from the rest of the data, which follows a Gamma distribution $Gamma(KN, 1)$ by construction. In **C)**, a view of the left part from **B)** with the theoretical histograms of Gamma distributions from $N = 1$ up to $N = 12$. The black dotted lines represent the lower bound λ_- to the upper bound λ_+, with $p = 0.05$, $N_{min} = 1$ and $N_{max} = 12$. This broad search covers the background voxels in both cases while excluding remaining voxels which do not belong to the distribution $Gamma(KN, 1)$.

center, neglecting spatial correlations ($\rho = 0$) as it is a k-space method (34). The resulting effective values of N and σ_g will be both spatially varying. We additionally generated 33 synthetic noise maps per dataset by setting the underlying signal value to $\eta = 0$ and performing the reconstruction using the same parameters as the DWIs. All generated datasets are available online (35).

Acquired phantom datasets

We acquired images of a bottle of liquid on a 3T Philips Ingenia scanner using a 32 channels head coil with a gradient strength of 45 mT/m. We varied the SENSE factor from $R = 1, 2$ or 3 and multiband acceleration factors from no multiband, $MB = 2$ or $MB = 3$ while fixing remaining acquisition parameters to investigate their influence on the resulting signal distributions, resulting in 9 different acquisitions. The datasets consist of 5 b = 0 s/mm^2 volumes and 4 shells with 10 DWIs each at b = 500 s/mm^2, b = 1000 s/mm^2, b = 2000 s/mm^2 and b = 3000 s/mm^2 with a voxel size of 2 mm isotropic and TE / TR = 135 ms / 5000 ms, Δ/δ = 66.5 ms / 28.9 ms. Six noise maps were also acquired during each of the experiments by disabling the RF pulse and gradients of the sequence. The acquired phantom datasets are also available (35).

In vivo dataset

One dataset from the 2017 MICCAI harmonization challenge[6](14) was used, consisting of 14 b = 0 s/mm^2 volumes and 3 shells with 60 DWIs each at b = 1200 s/mm^2, b = 3000 s/mm^2 and b = 5000 s/mm^2 with a voxel size of 1.5 mm isotropic acquired on a 3T Siemens Prisma scanner with a pulsed-gradient spin-echo echo-planar imaging (PGSE-EPI) sequence with gradient strength of 80 mT/m. Multiband acceleration $MB = 3$ was used with GRAPPA parallel imaging with R = 2 and a sum of squares reconstruction employing a 32 channels head coil. Other imaging parameters were TE / TR = 80 ms / 4500 ms, Δ/δ = 38.3 ms / 19.5 ms, bandwidth of 1476 Hz/pixel and partial Fourier 6/8.

Noise estimation algorithms for comparison

To assess the performance of the proposed methods, we used three other noise estimation algorithms previously used in the context of dMRI. Default parameters were used for all of the algorithms as done in (25). The local adaptive noise estimation (LANE) algorithm (20) is designed for noncentral chi signal estimation, but requires *a priori* knowledge of N. Since the method works on a single 3D volume, we only use the b = 0 s/mm^2 image for all of the experiments to limit computations as the authors concluded that the estimates from a single DWI are close to the mean estimate. We also use the Marchenko-Pastur (MP) distribution fitting on the principal component analysis (PCA) decomposition of the diffusion data, which is termed MPPCA (18). Finally, we also compare to the Probabilistic Identification and Estimation of Noise (PIESNO) (19), which originally proposed the change of variable to the Gamma distribution that is at the core of our proposed method. PIESNO requires knowledge of N (which is kept fixed by the algorithm) to iteratively estimate σ_g until convergence by removing voxels which do not belong to the distribution $Gamma(N, 1)$ for a given slice. We set $p = 0.05$ and $I = 50$ for the initial search of σ_g in PIESNO and our proposed method, with additional parameters set to $N_{min} = 1$ and $N_{max} = 12$ for all cases. When estimating distributions from noise maps, we compute values in small local windows of size $3 \times 3 \times 3$. To the best of our knowledge, ours is the first method which estimates both σ_g and N jointly without requiring any prior information about the reconstruction process of the MRI scanner. Because PIESNO and LANE both *require* knowledge of the value of N, we set the correct value of N for the spatially varying noise phantom experiments and $N = 1$ for the remaining experiments as suggested by (20) when N is unknown. We quantitatively assess the performance of each method on the synthetic datasets by measuring the

[6] https://www.cardiff.ac.uk/cardiff-university-brain-research-imaging-centre/research/projects/cross-scanner-and-cross-protocol-diffusion-MRI-data-harmonisation

standard deviation of the noise and the relative error inside the phantom against the known value of σ_g, computed as

$$\text{relative error} = 100 \times \left(\sigma_{g_{estimated}} - \sigma_{g_{true}}\right) / \sigma_{g_{true}}. \quad [12]$$

As PIESNO and our proposed method estimate a single value per slice whereas MPPCA and LANE provide estimates from small spatial neighborhood, we report the mean value and the standard deviation estimated inside the phantom on each slice. In addition, we report N as estimated by our proposed method.

Bias correction and denoising of the in vivo dataset

In a practical setting, small misestimation in the noise distribution (e.g., spatially varying distribution vs nature of the distribution) might not impact much the application of choice. We evaluate this effect of misestimation on the in vivo dataset with a bias correction and a denoising task. Specifically, we apply noncentral chi bias correction (21) on the in vivo dataset. The algorithm is initialized with a spherical harmonics decomposition of order 6 (36) as done in (22). The data is then denoised using the non local spatial and angular matching (NLSAM) algorithm with 5 angular neighbors where each b-value is treated separately (22). Default parameters of a spatial patch size of $3 \times 3 \times 3$ were used and the estimation of σ_g as computed by each method was given to the NLSAM algorithm. For MPPCA, LANE and PIESNO, a default value of $N = 1$ was used and the value of N as computed by the moments and maximum likelihood equations for the proposed methods.

4 | RESULTS

We show here results obtained on the phantoms and in vivo datasets. The first set of simulations uses a sum of square reconstruction with stationary and spatially varying noise profiles. The second set of simulations includes SENSE and GRAPPA reconstructions, resulting in both spatially varying signal distribution profiles. Finally, the distributions estimated by each algorithm for the in vivo dataset are used for a bias correction and denoising task.

4.1 | Synthetic phantom datasets

Simulations with a sum of squares reconstruction

Fig. 2 shows results from simulations with stationary and spatially varying noise profiles for all datasets. Estimation of σ_g is the most accurate for the proposed methods with an error of about 1%, followed by MPPCA making an error of approximately 5% and LANE of 20%. PIESNO error is increasing with the value of N, presumably due to misspecification in the signal distribution. Finally, Fig. 3 shows the estimated values of N by the proposed methods. Estimation generally follows the correct value, regardless of misestimation of σ_g.

Simulations with parallel imaging

Fig. 4 shows the estimated values of σ_g from a SENSE reconstruction and Fig. 5 shows the results for the GRAPPA reconstructed datasets. For SENSE, estimation using noise maps is the most precise for both proposed methods and PIESNO where the average error is around 0, followed by LANE when using DWIs as the input which results in 10% of overestimation. MPPCA generally underestimates σ_g by around 15% for data at b = 1000 s/mm^2 and 30% for data at b = 3000 s/mm^2, while results are similar at both b-values with the other methods. In the case of GRAPPA, MPPCA performs best when using DWIs as input for data at b = 1000 s/mm^2 and performs equally well at b = 3000 s/mm^2

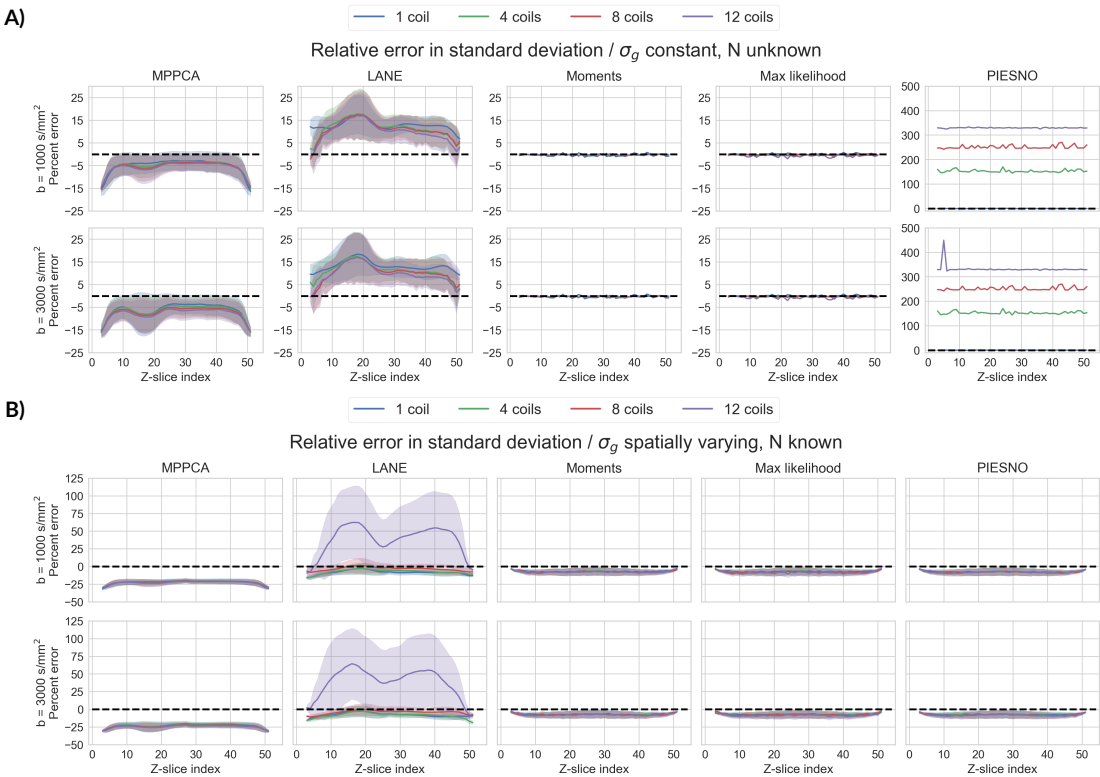

FIGURE 2 Percentage of error when the real value of N is known and σ_g is constant (in **A**)) and N is known with σ_g spatially varying (in **B**)) with the mean (solid line) and standard deviation (shaded area). For stationary σ_g, both MPPCA and LANE are stable in their estimation with increasing values of N. However, MPPCA always underestimates σ_g by approximately 10% and LANE overestimates σ_g by 15%. The proposed methodology using equations based on the moments and maximum likelihood recovers the correct value of σ_g in all cases with the lowest variance across slices, indicating that the estimated value of σ_g is similar in all slices as is expected. The same behavior is observed for PIESNO when $N = 1$, but the estimated σ_g is larger than the correct value by two to three times when N is misspecified. All methods underestimate spatially varying σ_g, except for LANE with $N = 12$ which overestimates it instead. On average, all methods are tied at around 5% of error with MPPCA reaching approximately 25% of error. Of interesting note, the proposed methodologies are tied with PIESNO when the correct value of N is given to the latter, but do not require an estimate of N, which is now an output instead of a prerequisite.

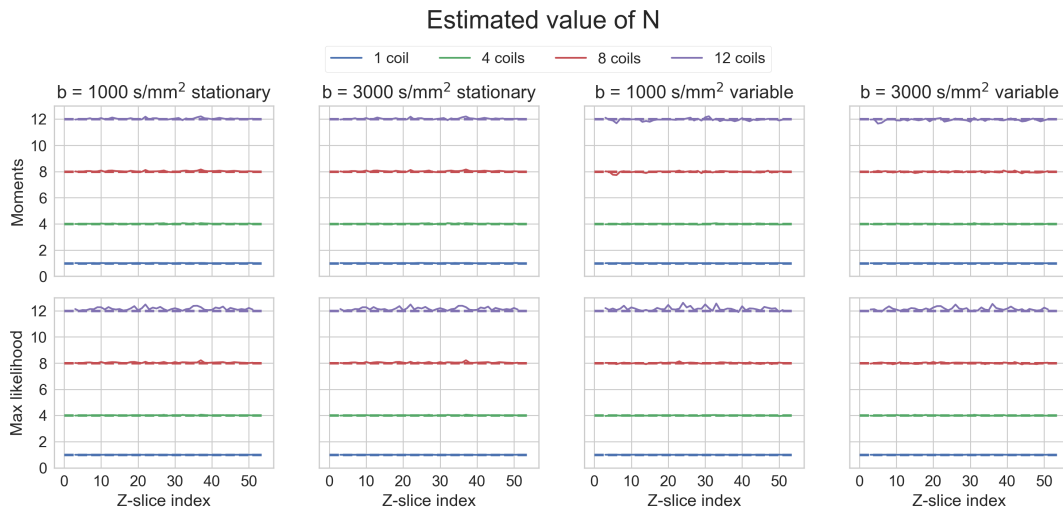

FIGURE 3 Estimated value of N using equations from the moments (top) and with maximum likelihood (bottom) for the proposed methods. Even for the spatially variable case where σ_g is slightly underestimated, the estimated values of N are stable and correspond to the real values used in the synthetic simulations in every case.

as the proposed methods with an average error of about 20%. In this case, LANE overestimates σ_g by around 50% when using DWIs, but performs similarly to MPPCA when estimating σ_g from the noise maps. The error of PIESNO increases with N as seen in Fig. 5 panel **C)**. Finally, Fig. 6 shows the estimated values of N_{eff} using the datasets from Figs. 4 and 5 by the proposed methods. For the SENSE case, values of N are close to 1 as expected. In the case of GRAPPA, using the noise maps as input leads to spatial misestimation as the correct parameters depend on the true signal value η. For simulations using 8 and 12 coils, estimated values of N are, in general, following the expected values.

4.2 | Acquired phantom datasets

Fig. 7 shows the estimated values of σ_g for all methods with a SENSE acceleration of rate $R = 1, 2$ and 3 with multiband imaging at acceleration factors of $MB = 2$, $MB = 3$ or deactivated in panel **A)**. Results show that σ_g increases with R and is higher when $MB = 3$ for R fixed, even if in theory σ_g should be similar for a given R and increasing MB. Panel **C)** shows the estimated values of σ_g when using noise maps as the input for $R = 3$ and $MB = 3$. As in the synthetic experiments, MPPCA and LANE have the lowest estimates for σ_g with PIESNO and the proposed methods estimating higher values. Since the correct value is unknown, a reference sample slice of a noise map is also shown. When compared to values from the measured noise map, estimated values of σ_g are approximately fivefold lower for MPPCA, four times lower for LANE and around half for the other methods. Estimation on the noise maps yields a value of around $N = 1$ for both proposed methods as seen in panels **B)** and **E)**, irrespective of the acceleration used. In the case of estimation using the DWIs, the range of estimated values is larger and increases at acceleration factors of $MB = 2$ and $R = 2$ or 3.

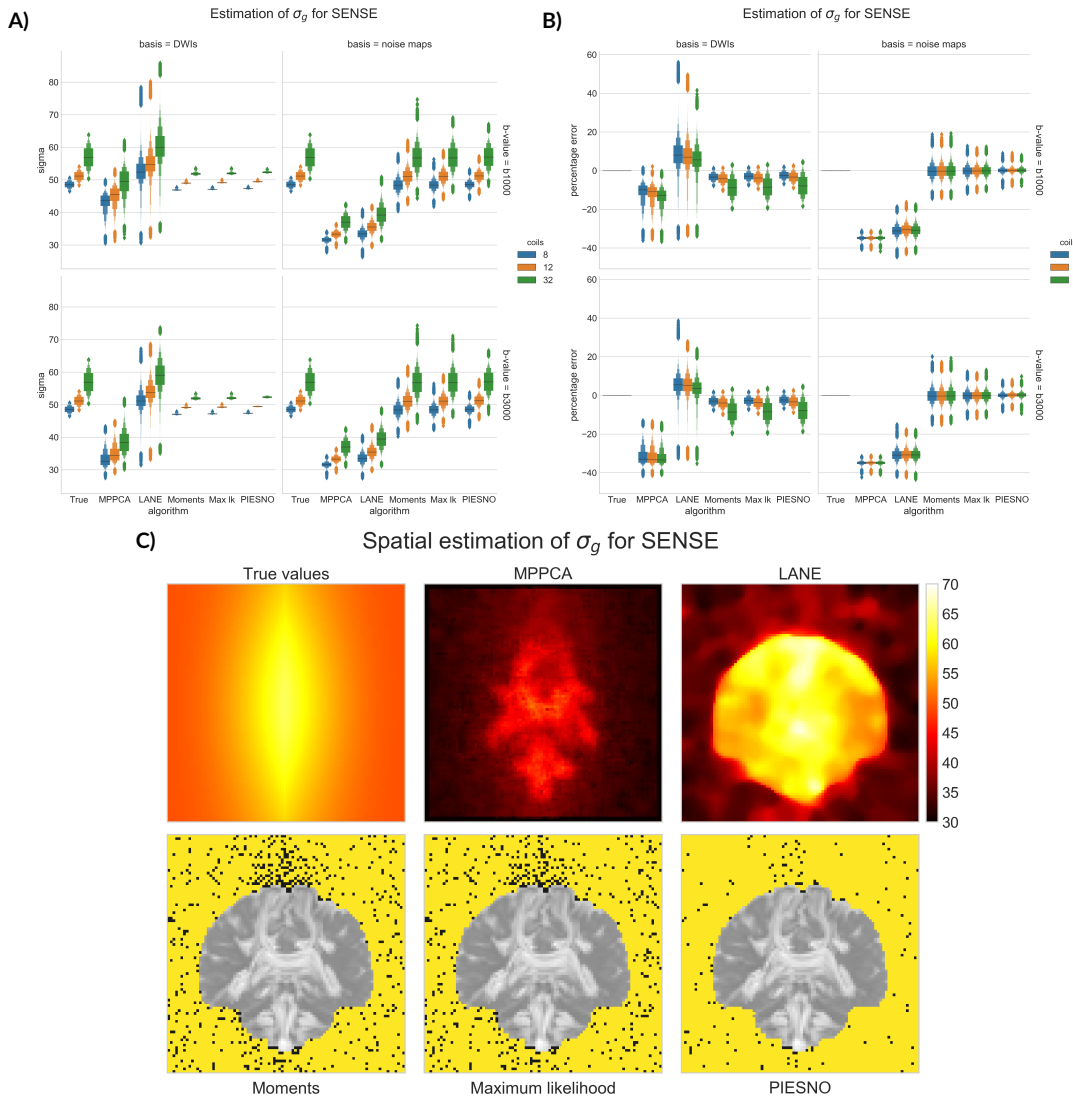

FIGURE 4 Estimation of the noise standard deviation σ_g (in **A**) and the percentage error (in **B**) inside the phantom only for each method using a SENSE reconstruction with 8, 12 or 32 coils. The left columns (basis = DWIs) shows estimation using all of the DWIs, while the right column (basis = noise maps) shows the estimated values from synthetic noise maps. Results for b = 1000 s/mm^2 are on the top row, while the bottom row shows results for the b = 3000 s/mm^2 datasets. MPPCA is generally underestimating σ_g while LANE instead overestimates when using DWIs and underestimates σ_g when using noise maps and knowing the correct value of $N = 1$. The moments and maximum likelihood equations (the proposed methodologies) and PIESNO are performing similarly, but PIESNO requires knowledge of the correct value of $N = 1$. Estimation is also more precise for the three methods using the Gamma distribution (Moments, Maximum likelihood and PIESNO) than those using local estimations (MPPCA and LANE) and closest to the true values when using noise maps. Figure **C)** shows the spatially estimated values of σ_g using the b = 3000 s/mm^2 dataset with 32 coils for a single slice. The top row shows the results from the true distribution and local estimation as done by MPPCA and LANE. The general trend shows that even though MPPCA and LANE misestimate σ_g, they still follow the spatially varying pattern (lower at edges with the highest intensity near the middle) from the correct values. In the bottom row, voxels identified as belonging to the same distribution Gamma$(N, 1)$ are overlaid in yellow over the sum of all DWIs. Note how voxels containing signal from the DWIs are excluded by all three methods.

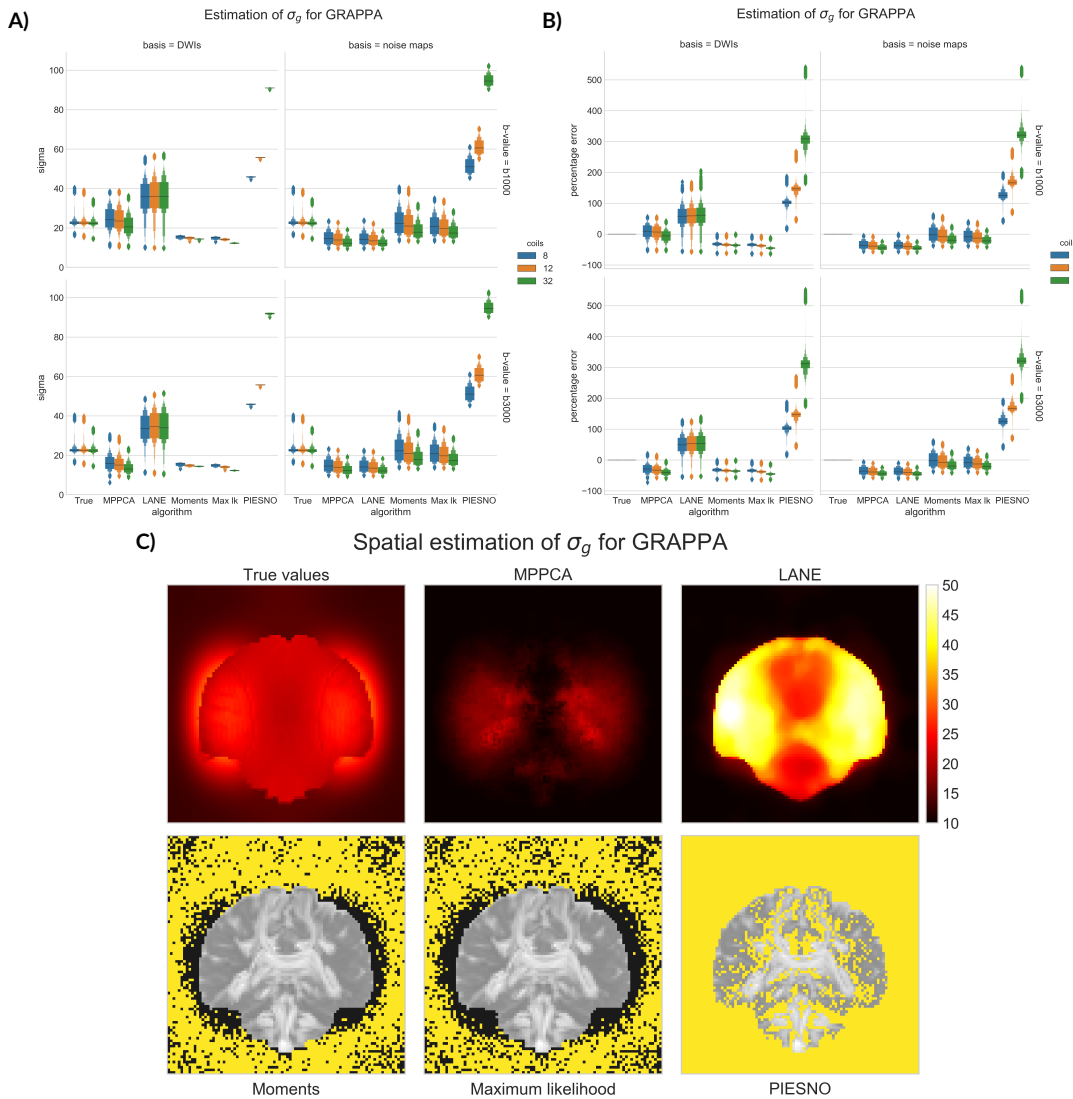

FIGURE 5 Estimation of the noise standard deviation σ_g (in **A**) and the percentage error (in **B**) inside the phantom only for each method using a GRAPPA reconstruction with 8, 12 or 32 coils, using the same conventions as Fig. 4. Results are similar to the SENSE experiments with the exception of MPPCA being more precise than the compared methods for the b = 1000 s/mm^2 case but still underestimating σ_g for the b = 3000 s/mm^2 case. Results using LANE are similar with increasing number of coils when assuming $N = 1$, while the estimated value from PIESNO also increases with the number of coils as previously seen in Fig. 2. Estimation from noise maps using the moments or maximum likelihood equations is the most precise in all cases. Figure **C)** shows the spatially estimated values of σ_g using the b = 3000 s/mm^2 dataset with 32 coils for a single slice The top row shows the true value of σ_g and the spatial estimation from MPPCA and LANE. There is once again a misestimation for both method while following the correct spatially varying pattern. In the bottom row, voxels identified as belonging to the same distribution Gamma(N, 1) are overlaid in yellow over the sum of all DWIs. Note how PIESNO mistakenly selects some low intensity voxels belonging to the gray matter, in addition to all of the voxels in the background, which causes an overestimation of σ_g with a fixed value of $N = 1$. Both proposed methods instead select voxels with small variations in intensity as belonging to the same distribution without mistakenly selecting gray matter voxels.

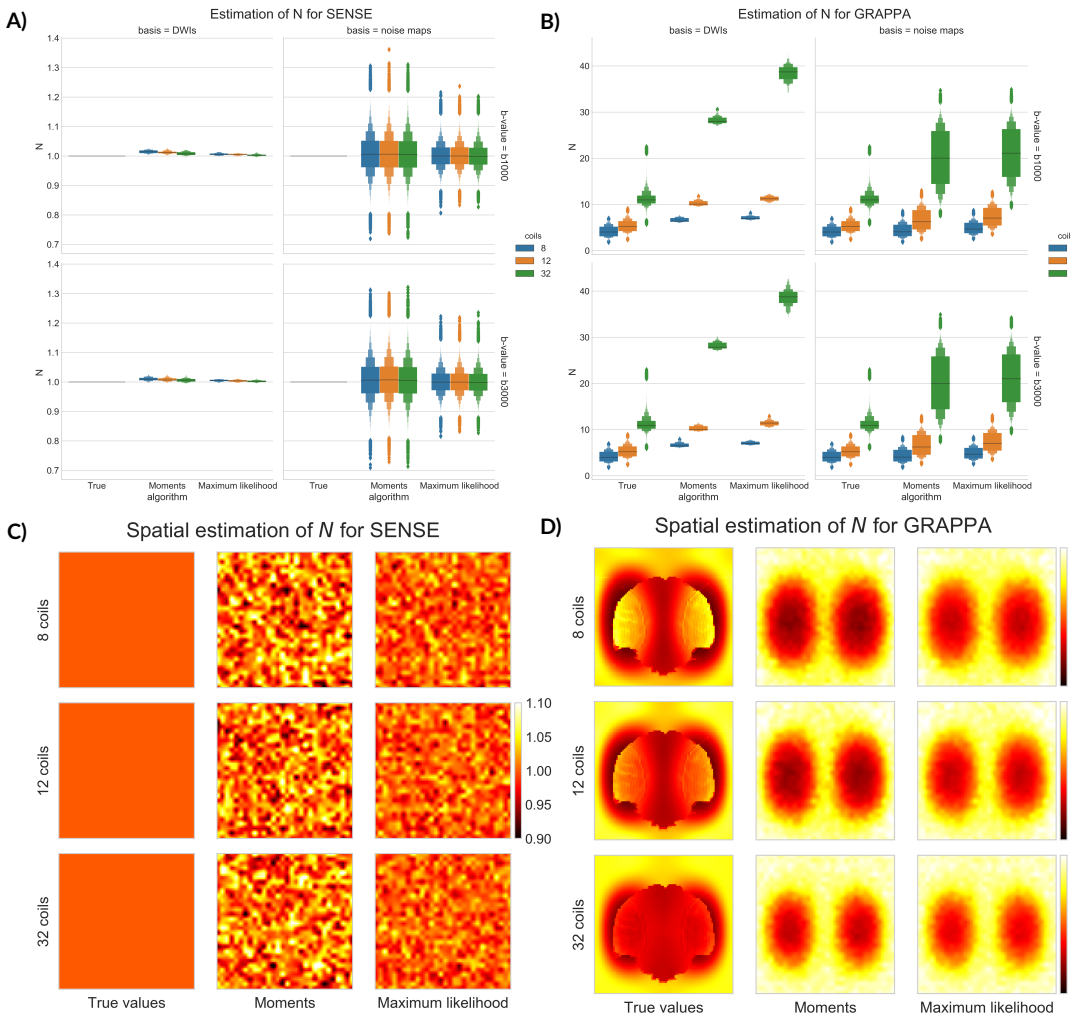

FIGURE 6 The estimated values of N for SENSE (left column) and GRAPPA (right column) for the b = 1000 s/mm^2 (first row of boxplots) and b = 3000 s/mm^2 (second row of boxplots) datasets. The left column shows results computed from the automatically selected background voxels (basis = DWIs), while the right column shows local estimation using noise maps (basis = noise maps). In **A)** and **B)**, the boxplot of N inside the phantom for the SENSE/GRAPPA algorithm with a spatial map of N shown in **C)** and **D)** computed using the noise maps from the b = 3000 s/mm^2 datasets. For the SENSE case, the true value is a constant $N = 1$ by construction and the estimated values by both algorithms are on average correct with the maximum likelihood equations having the lowest variance. For the GRAPPA case, values of N vary spatially inside the phantom and depend on the per voxel signal intensity, just as σ_g does in Fig. 5. This leads to overestimation by both methods as they only consider background voxels with the best estimation obtained when using the noise maps. However, the spatially varying pattern can not be fully recovered as the correct value of N depends on the true signal intensity η in each voxel, which is not present when collecting noise only measurements.

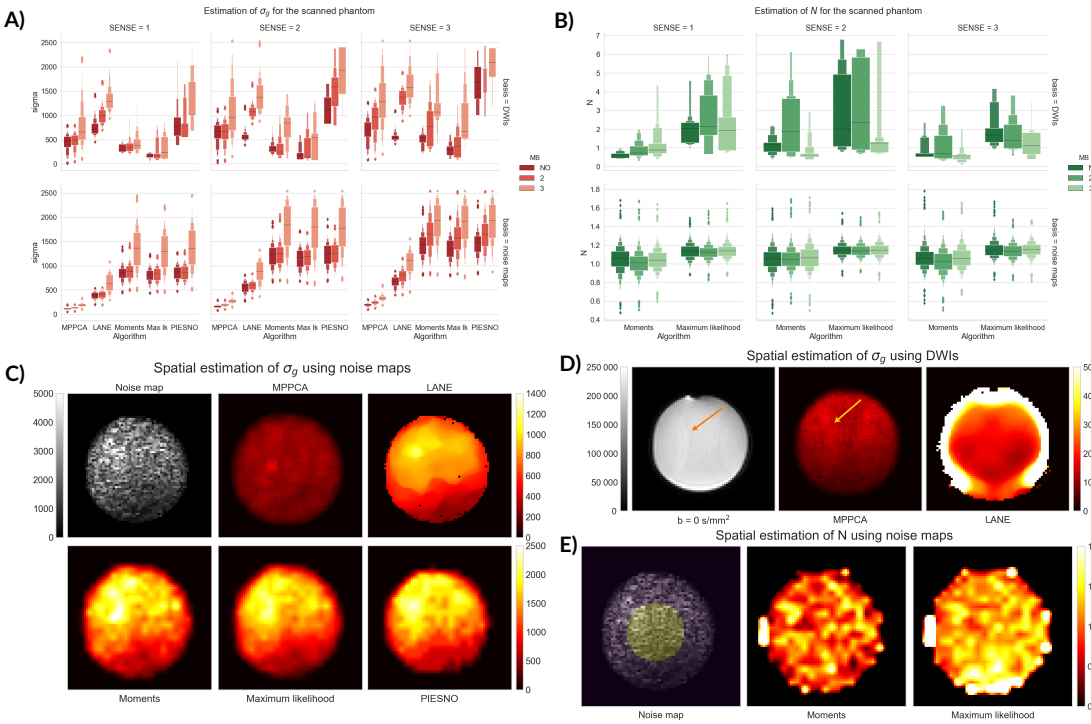

FIGURE 7 Estimation of noise distributions for the scanned phantom datasets inside a small ROI. Large outliers above the 95th percentile were removed to not skew the presented boxplots. In **A)**, the estimation of the noise standard deviation σ_g for each method using DWIs (top row) and using noise maps (bottom row). Each column shows an increasing SENSE factor, where σ_g increases (according to theory) with the square root of the SENSE factor. The different hues show an increasing multiband factor, which should not influence the estimation of σ_g. For the case $MB = 3$, there may be signal leakage to adjacent slices, which would increase the measured values of σ_g even when the estimation uses only noise maps. In **B)**, boxplots for the values of N estimated by both proposed methodologies for the experiments shown in **A)**. Estimated values using noise maps are always close to 1 on average while estimations using DWIs seems to be affected by the possible signal leakage inherent to the use of multiband imaging. In **C)**, an axial slice of a noise map and estimated values of σ_g by all methods for the case $R = 3$ and $MB = 3$, which is the highest rate of acceleration from all of the investigated cases. Note the different scaling between the top and bottom row as MPPCA and LANE estimates of σ_g are two to three times lower than other methods. In **D)**, a b = 0 s/mm^2 image of the phantom and spatially estimated values of σ_g for MPPCA and LANE. Note how some signal leakage (orange arrows) is affecting the b = 0 s/mm^2 volume due to using $MB = 3$. In **E)**, location of the spherical ROI used for the boxplots overlaid on a noise map and spatially estimated values of N for both proposed methods. As less voxels are available near the borders of the phantom, estimating the noise distributions parameters results in lower precision.

4.3 | In vivo dataset

Fig. 8 shows in **A)** the estimated values of σ_g inside a brain mask and in **B)** the values of N computed by the proposed methods. Estimated values of σ_g vary by an order of magnitude between the different methods. In the case of MPPCA and LANE, the median of the estimates is higher than the reference b = 5000 s/mm^2 data, while PIESNO and the proposed methods estimate values close to 0. In **C)**, the top row shows the b = 5000 s/mm^2 volume and spatial maps of σ_g as estimated by MPPCA and LANE. The bottom row shows voxels identified as pure noise (in light purple) using the moments, the maximum likelihood equations and PIESNO. Ghosting artifacts are excluded, but presumably affect estimation using the entire set of DWIs shown in the top row. Fig. 9 shows in **A)** the signal intensity after applying bias correction (left column) and after denoising (right column) for each volume ordered by increasing b-value. The top row (resp. bottom row) shows the mean (resp. standard deviation) as computed inside a white and gray matter mask. The mean signal decays with increasing b-value as expected, but the standard deviation does not follow the same trend in the cases of MPPCA and LANE. After denoising, the mean signal and its standard deviation decays once again as for the original data. Panel **B)** shows the average DWI at a given b-value for the original dataset and after denoising using the noise distribution from each method. Finally, panel **C)** shows the absolute difference between the original and the denoised dataset obtained by each method. At b = 5000 s/mm^2, LANE removes most of the signal in the gray matter mistakenly due to overestimating σ_g. Other methods perform comparably well on the end result, despite values of σ_g of different magnitude.

5 | DISCUSSION

We have shown how a change of variable to a Gamma distribution $Gamma(N, 1)$ can be used to robustly and automatically identify voxels belonging only to the noise distribution. At each iteration, the moments (Eqs. (6) and (7)) and maximum likelihood equations (Eqs. (9) and (10)) of the Gamma distribution can be used to compute the number of degrees of freedom N and the Gaussian noise standard deviation σ_g relating to the original noise distribution. Voxels not adhering to the distribution are discarded, therefore refining the estimated parameters until convergence. One of the advantage of our proposed methods is that no *a priori* knowledge is needed from the acquisition or the reconstruction process itself, which is usually not stored or hard to obtain in a clinical setting. Results from Section 4.1 show that we can reliably estimate parameters from the magnitude data itself in the case of stationary distributions. For spatially varying distributions without parallel acceleration, the proposed methods achieve an average relative error of approximately 10% when estimating σ_g, which is equal or better than the other methods compared in this work. Estimated values of N are around the true values, even when σ_g is misestimated. While these experiments may still be considered to be simplistic when compared to modern scanning protocols where parallel acceleration is ubiquitous, they highlight that even textbook cases can lead to misestimation if the correct signal distribution is not taken into account. Practical tasks taking advantage of the signal distribution such as bias correction (37), noise floor removal (17), deep learning reconstruction with various signal distributions (38) or diffusion model estimation (16, 23, 24) may be tolerant, but not perform optimally, to some misestimation of the noise distribution. See e.g., (39) for discussions on the impact of noise bias correction on diffusion metrics in an ex vivo rat brain dataset.

Effects of misspecification of the noise distribution

Experiments with SENSE from Section 4.1 reveal that using a local estimation with noise maps provides the best estimates for the proposed methods and PIESNO. MPPCA and LANE perform better when using DWIs as the input rather than noise maps, but at the cost of a broader range of estimated values for σ_g and still underperform when

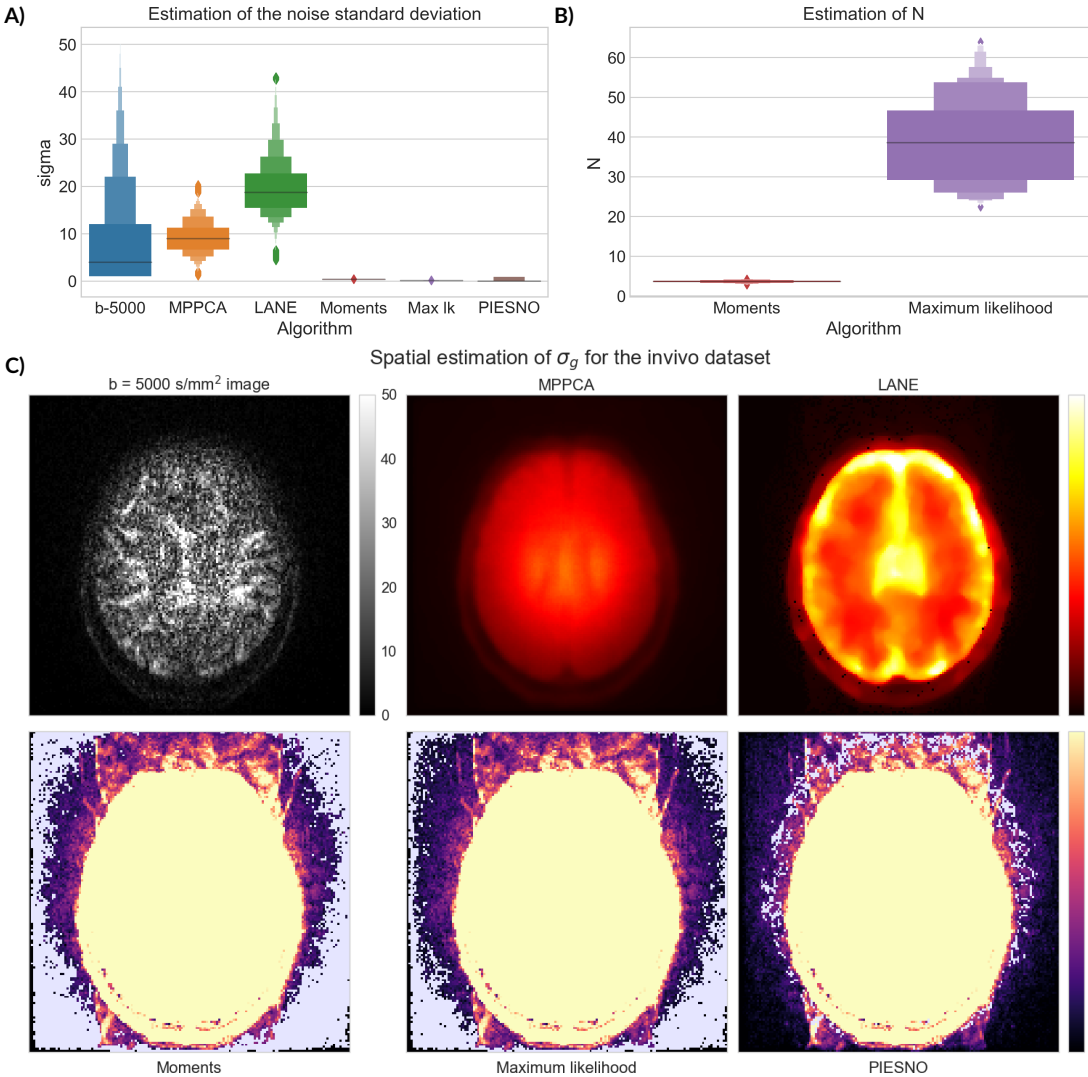

FIGURE 8 Estimation of noise distributions for the in vivo datasets. In **A)**, signal distribution of the original data and noise standard deviation σ_g for all methods, where data above the 99th percentile for the b = 5000 s/mm^2 volume and LANE were discarded. In **B)**, values of N as estimated using the moments (in red) and by maximum likelihood (in purple). In **C)** on the top row, a b = 5000 s/mm^2 volume and spatial estimation of σ_g as measured by MPPCA and LANE. On the bottom row, voxels identified as containing only noise by the moments, maximum likelihood and PIESNO overlaid on top of the sum of the b = 0 s/mm^2 volumes. Note how each algorithm identifies different voxels, while automatically ignoring voxels belonging to the data or contaminated with signal leakage from multiband imaging.

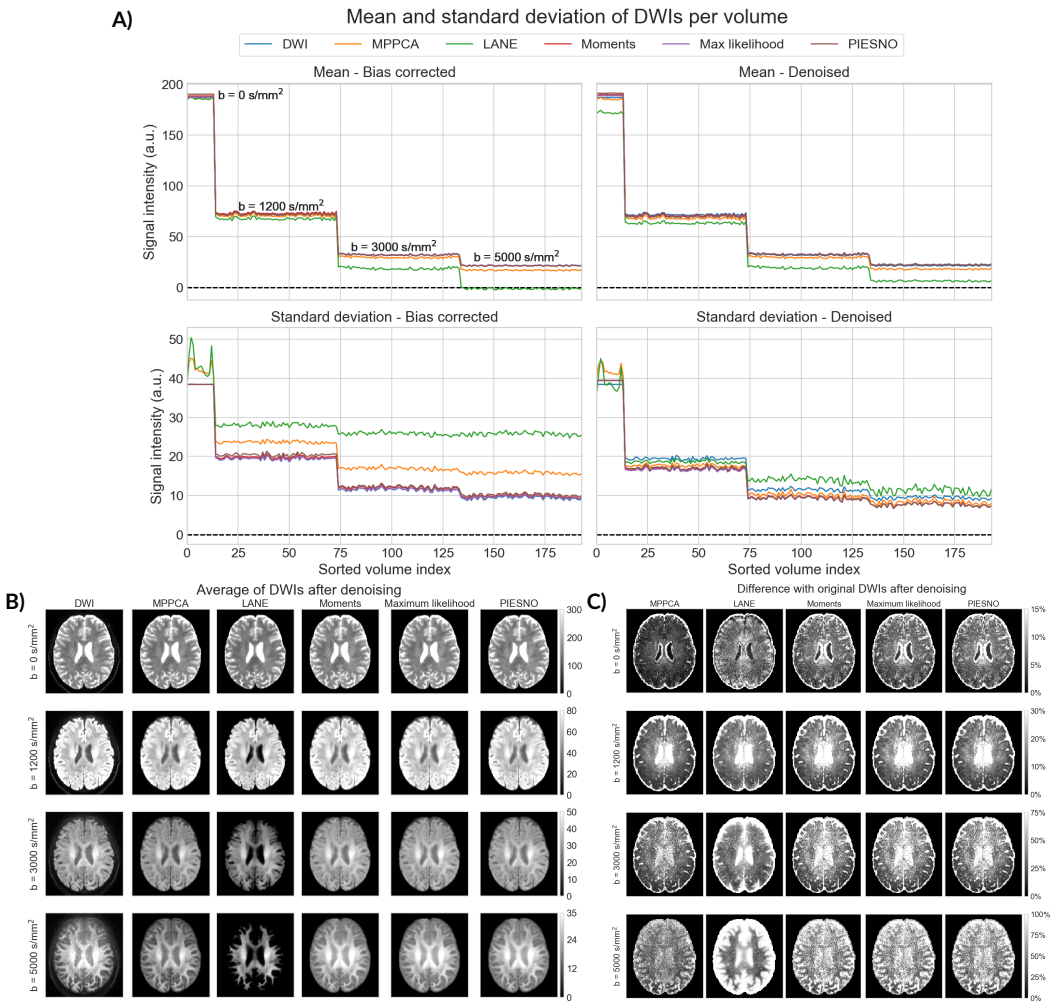

FIGURE 9 Bias correction and denoising of the in vivo dataset from the noise distributions estimated by each method. In **A)**, the left column (resp. right column) shows the result of noncentral chi bias correction (resp. denoising) on the signal value. The top row (resp. bottom row) shows the mean (resp. standard deviation) of the signal inside a white and gray matter mask for each volume. Note how the bias corrected value of LANE goes below 0 (dashed line) due to its high estimation of σ_g. After denoising, the standard deviation of the signal decreases as the b-value increases, an effect which is less noticeable for the bias corrected signal only. However, this effect is less pronounced for the bias corrected signal only in the case of LANE and MPPCA. In **B)**, spatial maps of the original data and after denoising (in each column) from averaged datasets at b = 0 s/mm^2, b = 1200 s/mm^2, b = 3000 s/mm^2 and b = 5000 s/mm^2 (in each row) for each method. Note how each b-value use a different scale to enhance visualization even though the signal intensity is lower for increasing b-values. Panel **C)** shows the difference in percentage between the original data and after denoising using parameters as estimated by each algorithm. Results are similar for all methods for the b = 0 s/mm^2 datasets, but the overestimation of σ_g by LANE produces missing values in the gray matter for b = 3000 s/mm^2 and b = 5000 s/mm^2. In general, denoising reduces the speckled appearance of the data at b = 3000 s/mm^2 and b = 5000 s/mm^2. At b = 5000 s/mm^2, the MPPCA denoised volume is of lower intensity than when obtained by the moments, maximum likelihood equations or PIESNO. This is presumably due to LANE and MPPCA estimating higher values of σ_g than the three other methods.

compared to the three other methods. This is presumably because the signal diverges from a Gaussian distribution at low SNR (26) and especially in noise maps, leading to a misspecification of its parameters when the assumed noise distribution is incorrect. Phantom experiments carried with GRAPPA show similar trends except for PIESNO, which overestimates σ_g as shown in Fig. 2. When erroneously fixing $N = 1$, low intensity voxels where $\eta > 0$ (e.g., gray matter) may be mistakenly included in the distribution after the change of variables, leading to overestimation of σ_g.

The presence of tissue in voxels used for noise estimation might compromise the accuracy of the estimated distributions as shown in Section 4.1. This can be explained by the lower number of noise only voxels available to the proposed methods and PIESNO and to difficulty in separating the signal from the noise for MPPCA and LANE at low SNR. Using measured noise maps is not a foolproof solution as by definition they set $\eta = 0$, while the (unknown) noiseless signal from tissues is $\eta > 0$. As the noise distribution may depend on η (28), this means that its parameters (e.g., from a GRAPPA reconstruction) will be inherently different than the one estimated from noise maps. This effect can be seen in Fig. 6, where the estimated values of N from noise maps and DWIs are close to 1 for SENSE as expected in theory. For GRAPPA, they are either overestimated and underestimated in regions of the phantom and overestimated in background regions as N locally depends on η. Accurate estimation of σ_g and N over signal regions still remains an open challenge. Nevertheless, the median of the estimated distribution of σ_g is closer to the true distribution when using noise maps than when using DWIs for the proposed methods. Such noise map measurements could therefore provide improved signal distribution estimation for, e.g., body or cardiac imaging, where no intrinsic background measurements are available.

Effects of parallel imaging and multiband in a phantom

Section 4.2 presented results from a scanned phantom using SENSE coupled with multiband imaging. While no ground truth is available, a SENSE acceleration should provide a Rician signal distribution ($N = 1$) and σ_g should increase with $\sqrt{R}$ (28). Fig. 7 shows that for a common SENSE factor, all values of σ_g estimated with $MB = 3$ are larger than at lower factors. The use of multiband imaging should not influence the estimation of σ_g as it only reduces the measured signal, and not the noise component unlike SENSE does. Indeed, estimated values of σ_g are stable until $MB = 3$ or $R = 2$ and $MB = 2$ is used; this is possibly due to signal leakage from adjacent slices with higher factors (40, 41). Noise maps are less affected by this artifact, which is already present when $R = 2$ and $MB = 2$, as adjacent voxels have low values, similarly to unaffected voxels. However, leaking signal in DWIs might impact parameters estimation as it can be interpreted as an increase in SNR and therefore a lower noise contribution than expected. Estimation of σ_g is also increasing approximately with $\sqrt{R}$ for all methods as expected (28). While we can not quantify these results, this follows the synthetic experiments with SENSE shown in Fig. 4, where PIESNO and the proposed methods were more precise in estimating σ_g from noise maps.

In the case of estimation using DWIs as input, this expected increase in σ_g for increasing SENSE factor is less obvious e.g., LANE estimates of σ_g decrease from $R = 2$ to $R = 1$ for the no multiband case. As MPPCA and LANE also estimate η, it could explain the larger variance of σ_g as η fundamentally depends on the microstructural content of each voxel, which is complex and subject to large spatial variations, e.g., notably across DWIs. In the noise maps, we have observed that MPPCA and LANE estimated $\eta > \sigma_g$ in all cases (results not shown) even though, by definition, we have $\eta = 0$. This overestimation of η in turn leads to lower estimates of σ_g. The use of multiband and the inherent signal leakage at high factors could explain this overestimation of η and underestimation of σ_g for all tested cases. In the case of SENSE, the proposed methods estimated approximately $N = 1$ in all cases, suggesting robustness to multiband artifacts.

Estimation of noise distributions in an in vivo dataset

To complement earlier sections, an in vivo dataset combining parallel and multiband imaging was analyzed in Section 4.3. Fig. 8 shows a large range of estimates for σ_g across methods. In particular, the moments, maximum likelihood equations and PIESNO estimate small values of σ_g (between 0 and 1) and large values of N. However, they also successfully discard voxels contaminated by multiband artifacts. As seen in panel **B)** of Fig. 5, N is overestimated when the proposed methods use DWIs as input, which could explain these results from a 32 channels coil. In panel **A)**, MPPCA and LANE estimates of σ_g with DWIs are likely affected by multiband artifacts as the median is larger than the signal level at b = 5000 s/mm^2. This indicates a possible overestimation as σ_g should be lower than the measured signal at the highest b-value. For PIESNO and the proposed methods, the median σ_g is lower than the median of the reference b = 5000 s/mm^2 data. An overestimation of N could explain the low values of σ_g estimated by the proposed methods.

Fig. 9 shows the result of each method on a bias correction and denoising task. In panel **A)**, the standard deviation of the signal (bottom left panel) is increased after bias correction for MPPCA (orange line) and LANE (green line) and decreased (around the same level) for the three other methods when compared to the uncorrected data (blue line). The situation is similar after denoising, but to a lesser extent, while the moments, maximum likelihood equations and PIESNO follow the same signal level as the unprocessed data on average. Regarding the mean of the signal itself, LANE is on average lower or close to 0 after bias correction, indicating potential degeneracies due to overestimation of σ_g which would put the signal artificially below the noise floor (21). We also note that PIESNO either estimated $\sigma_g = 0$ as it failed to identify noise voxels for some slices or estimated σ_g as the same value for every other slice, possibly due to a mismatch of the Rician distribution assumption. When PIESNO estimated $\sigma_g = 0$, we instead set σ_g to the value of adjacent slices to avoid degenerated cases during bias correction. From panels **B)** and **C)**, the results of all methods are visually similar except for LANE (especially at b = 3000 s/mm^2 and b = 5000 s/mm^2), indicating that the NLSAM denoising algorithm treated different values of σ_g in the same way. This is because the optimal regularized solution (which depends on σ_g) is piecewise constant (22, 42) and can tolerate small deviations in σ_g. Finally, MPPCA, the moments and likelihood equations and PIESNO perform similarly, even if they estimated different values of σ_g and N, with MPPCA showing slightly lower signal intensity at b = 5000 s/mm^2. This could be due to the bias correction having a larger effect when σ_g is larger, increasing the standard deviation of the resulting signal. As shown in panel **C)**, the difference with the original dataset for MPPCA is lower than the proposed methods or PIESNO, even though the estimated value of σ_g was larger. This could be due to residual noise still present at low b-values for MPPCA due to a misspecification of N despite the larger values of σ_g.

6 | CONCLUSIONS

We presented a new, fully automated framework for characterizing the noise distribution from a diffusion MRI dataset using the moments or maximum likelihood equations of the Gamma distribution. This requires only magnitude data, without the use of dedicated maps or parameters intrinsic to the reconstruction process, which may be challenging to obtain in practice. The proposed method is fast and robust to artifacts as voxels not adhering to the noise distribution can be automatically discarded using an outlier rejection step. Experiments using parallel MRI and multiband imaging on simulations, an acquired phantom and an in vivo dataset have shown how modern acquisition techniques complicate estimation of the signal distribution due to artifacts at high acceleration factor. This issue can be alleviated with the use of noise only measurements or by limiting the acceleration factor to prevent signal leakage. We also have shown how signal bias correction and denoising can tolerate some misestimation of the noise distribution using an in

vivo dataset. Noteworthy is that the theory we presented also applies to any other MRI weighting using large samples of magnitude data (e.g., functional MRI, dynamic constrast enhanced MRI). This could help multicenter studies or data sharing initiatives to include knowledge of the noise distribution in their analysis in a fully automated way to better account for inter-scanner effects.

A | SUPPLEMENTARY MATERIALS

A.1 | Estimating parameters of the Gamma distribution

Estimation using the method of moments

For any given distribution, we can estimate its parameters by relating the samples and the theoretical expression of its moments. The Gamma distribution is parametrized as $Gamma(\alpha, \beta)$ and has a probability distribution function of

$$pdf(t|\alpha, \beta) = \frac{t^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} \exp(-t/\beta)dt \quad [13]$$

with $t, \alpha, \beta > 0$ and $\Gamma(x)$ the gamma function. The first moments are analytically given by (Chap. 5 43), (29).

$$\mu_{gamma} = \alpha\beta,\ \sigma^2_{gamma} = \alpha\beta^2, \quad [14]$$

In this paper, the Gamma distribution parameters are $Gamma(\alpha = N, \beta = 1)$ after the change of variable $t = m^2/(2\sigma_g^2)$ for our particular case. Since we have $\beta = 1$, this leads to a special case where the mean and variance are *equal* with a value of $\alpha = N$ and can be expressed only in terms of the magnitude signal m. For simplicity, we will only use the mean μ_{gamma} and variance σ^2_{gamma} to estimate the required parameters N and σ_g^2, but higher order moments could also be used. However, in practice, they might accumulate numerical errors due to the higher powers involved and are not used here since two equations are enough to estimate the two parameters. Starting from the analytical expression given by Eq. (14), we have for the special case $Gamma(N, 1)$

$$\mu_{gamma} = \alpha,\ \sigma^2_{gamma} = \alpha \quad [15]$$

Which we can compute using the sample mean and sample variance formulas such that

$$\alpha = \frac{1}{K}\sum_{k=1}^{K} t_k = \frac{1}{K}\sum_{k=1}^{K} t_k^2 - \left(\frac{1}{K}\sum_{k=1}^{K} t_k\right)^2 \quad [16]$$

Substituting the equation for the moments in terms of $t = m^2/2\sigma_g^2$, we obtain

$$\frac{1}{K}\sum_{k=1}^{K}\frac{m_k^2}{2\sigma_g^2} = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{m_k^2}{2\sigma_g^2}\right)^2 - \left(\frac{1}{K}\sum_{k=1}^{K}\frac{m_k^2}{2\sigma_g^2}\right)^2 \quad [17]$$

$$\Rightarrow \quad \frac{1}{2K\sigma_g^2}\sum_{k=1}^{K} m_k^2 = \frac{1}{4K\sigma_g^4}\sum_{k=1}^{K} m_k^4 - \frac{1}{4K^2\sigma_g^4}\left(\sum_{k=1}^{K} m_k^2\right)^2 \quad [18]$$

$$\Rightarrow \quad \sum_{k=1}^{K} m_k^2 = \frac{1}{2K\sigma_g^2}\left(K\sum_{k=1}^{K} m_k^4 - \left(\sum_{k=1}^{K} m_k^2\right)^2\right) \quad [19]$$

$$\Rightarrow \quad 2K\sigma_g^2 = \frac{K\sum_{k=1}^{K} m_k^4 - \left(\sum_{k=1}^{K} m_k^2\right)^2}{\sum_{k=1}^{K} m_k^2} \quad [20]$$

$$\Rightarrow \quad \sigma_g = \frac{1}{\sqrt{2K}}\sqrt{\frac{K\sum_{k=1}^{K} m_k^4 - \left(\sum_{k=1}^{K} m_k^2\right)^2}{\sum_{k=1}^{K} m_k^2}} \quad [21]$$

$$\Rightarrow \quad \sigma_g = \frac{1}{\sqrt{2}}\sqrt{\frac{\sum_{k=1}^{K} m_k^4}{\sum_{k=1}^{K} m_k^2} - \frac{1}{K}\sum_{k=1}^{K} m_k^2} \quad [22]$$

Therefore, it is possible to estimate the Gaussian noise standard deviation using Eq. (22) and the values of magnitude data m_k, assuming that the voxels considered here do not contain any object signal. With the value of the noise variance σ_g^2 now known, going back to the original Gamma distribution $Gamma(\alpha = N, \beta = 1)$ yields the number of coils N as previously shown by Eq. (7)

$$N = \alpha = \mu_{gamma} = \frac{1}{2K\sigma_g^2}\sum_{k=1}^{K} m_k^2 \quad [23]$$

Estimation using maximum likelihood equations

An alternative to the method of moments to estimate parameters from a given distribution is to solve the equations derived from its likelihood function for each unknown parameter. Given a set of observed data, maximizing the likelihood function from a known distribution (or equivalently, the log of the likelihood function) yields a set of equations to estimate its parameters. For the $Gamma(\alpha, \beta)$ distribution, maximizing the log likelihood by equating the partial derivative to 0 for each parameter yields (30)

$$\frac{1}{K\beta}\sum_{k=1}^{K} t_k - \alpha = 0 \quad [24]$$

$$\log(\beta) + \frac{d}{d\alpha}\log(\Gamma(\alpha)) - \frac{1}{K}\sum_{k=1}^{K}\log(t_k) = 0 \quad [25]$$

Since we have $\alpha = N$ and $\beta = 1$, in this special case Eq. (24) is the same as Eq. (23).

Combining Eqs. (24) and (25) yields an implicit equation to estimate σ_g, which can be written as

$$f(\sigma_g) = \psi\left(\frac{1}{2K\sigma_g^2}\sum_{k=1}^{K} m_k^2\right) - \frac{1}{K}\sum_{k=1}^{K}\log(m_k^2) + \log(2\sigma_g^2) = 0 \quad [26]$$

$$f'(\sigma_g) = \frac{1}{K\sigma_g^3}\left[\psi'\left(\frac{1}{2K\sigma_g^2}\sum_{k=1}^{K} m_k^2\right)\sum_{k=1}^{K} m_k^2 - 2K\sigma_g^2\right] = 0 \quad [27]$$

and Eq. (25) can be rewritten as an implicit equation of N

$$f(N) = \psi(N) - \frac{1}{K}\sum_{k=1}^{K}\log(m_k^2/2\sigma_g^2) = 0 \quad [28]$$

$$f'(N) = \psi'(N) = 0 \quad [29]$$

where $\psi(x) = \frac{d}{dx}\log(\Gamma(x))$ is the digamma function and ψ' is the derivative of ψ, called the polygamma function. Eqs. (27) and (29) can be solved numerically using Newton's method provided we have a starting estimate x_0. The update rule for Newton's method at iteration n is therefore

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad [30]$$

For the first iteration, a starting estimate x_0 to approximate the solution is needed. For Eq. (27), we use $x_0 = \sigma_{m_N}$ while a starting estimate for Eq. (29) is given by (44) considering $y = \frac{1}{K}\sum_{k=1}^{K}\log(m_k^2/2\sigma_g^2)$.

$$x_0 = \psi^{-1}(y) \approx \begin{cases} \exp(y) + 1/2 & \text{if } y \geq -2.22 \\ -1/(y + \psi(1)) & \text{if } y < -2.22 \end{cases} \quad [31]$$

In practice, we have observed that 5 iterations of Eq. (30) were sufficient to reach $|x_n - x_{n-1}| < 10^{-13}$.

A.2 | Automated identification of noise only voxels

This appendix outlines the proposed algorithm and details for a practical implementation. Our implementation is also freely available at `https://github.com/samuelstjean/autodmri` and will be a part of ExploreDTI (45). The synthetic and acquired datasets used in this manuscript are also available (35).

Algorithm 1: Main algorithm to identify voxels belonging to the Gamma distribution

Data: 4D DWIs data, probability level $p = 0.05$, length of the search interval $I = 50$, $N_{min} = 1$, $N_{max} = 12$
Result: σ_g, N, mask of background only voxels
Compute the *median* of the whole dataset;
foreach *2D Slice of the 4D dataset* **do**
  Compute the upper bound $\sigma_{g_{max}} = median/\sqrt{2\,icdf(N_{max}, 1/2)}$;
  Compute the search interval $\Phi = [1\sigma_{g_{max}}/I, 2\sigma_{g_{max}}/I, \ldots, I\sigma_{g_{max}}/I]$;
  while σ_g, N *not converged* **do**
    Compute $\lambda_- = icdf(\alpha, p/2)$ and $\lambda_+ = icdf(\alpha, 1 - p/2)$;
    foreach $\sigma_{candidate} \in \Phi$ **do**
      Apply change of variable $t = \frac{data^2}{2\sigma^2_{candidate}}$;
      Find voxels from the gamma distribution mask_current $= \left(\lambda_- \leq \sum_{k=1}^{K} t_k\right) \bigcap \left(\sum_{k=1}^{K} t_k \leq \lambda_+\right)$;
      if *number of voxels in mask_current > mask* **then**
        mask = mask_current;
      end
    end
    Compute σ_g with the voxels inside the mask using Eq. (6) or Eq. (10);
    Compute N with the voxels inside the mask using Eq. (7) or Eq. (9);
    Set $N_{min} = N_{max} = N$;
    Set $\Phi = [0.95\sigma_g, 0.96\sigma_g, \ldots, 1.05\sigma_g]$;
  end
end

Conflict of interest

The authors have declared no conflict of interest.

References

[1] Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine* 1999; 42(5): 952–962. doi: 10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S

[2] Griswold Ma, Jakob PM, Heidemann RM, et al. Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). *Magnetic Resonance in Medicine* 2002; 47(6): 1202–1210. doi: 10.1002/mrm.10171

[3] Storey P, Frigo FJ, Hinks RS, et al. Partial k-space reconstruction in single-shot diffusion-weighted echo-planar imaging. *Magnetic Resonance in Medicine* 2007; 57(3): 614–619. doi: 10.1002/mrm.21132

[4] Nunes RG, Hajnal JV, Golay X, Larkman DJ. Simultaneous slice excitation and reconstruction for single shot EPI. *Proc. Intl. Soc. Mag. Reson. Med* 2006; 13(2): 293.

[5] Moeller S, Yacoub E, Olman CA, et al. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain FMRI. *Magnetic Resonance in Medicine* 2010; 63(5): 1144–1153. doi: 10.1002/mrm.22361

[6] Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine* 2007; 58(6): 1182–1195. doi: 10.1002/mrm.21391

[7] Paquette M, Merlet S, Gilbert G, Deriche R, Descoteaux M. Comparison of sampling strategies and sparsifying transforms to improve compressed sensing diffusion spectrum imaging. *Magnetic Resonance in Medicine* 2015; 73(1): 401–416. doi: 10.1002/mrm.25093

[8] Dietrich O, Raya JG, Reeder SB, Ingrisch M, Reiser MF, Schoenberg SO. Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics.. *Magnetic resonance imaging* 2008; 26(6): 754–62. doi: 10.1016/j.mri.2008.02.001

[9] Sotiropoulos SN, Moeller S, Jbabdi S, et al. Effects of image reconstruction on fiber orientation mapping from multichannel diffusion MRI: Reducing the noise floor using SENSE. *Magnetic Resonance in Medicine* 2013; 70(6): 1682–1689. doi: 10.1002/mrm.24623

[10] Holdsworth SJ, O'Halloran R, Setsompop K. The quest for high spatial resolution diffusion-weighted imaging of the human brain in vivo. *NMR in Biomedicine* 2019; 32(4): e4056. doi: 10.1002/nbm.4056

[11] Duchesne S, Chouinard I, Potvin O, et al. The Canadian Dementia Imaging Protocol: Harmonizing National Cohorts. *Journal of Magnetic Resonance Imaging* 2019; 49(2): 456–465. doi: 10.1002/jmri.26197

[12] Emaus MJ, Bakker MF, Peeters PHM, et al. MR Imaging as an Additional Screening Modality for the Detection of Breast Cancer in Women Aged 50–75 Years with Extremely Dense Breasts: The DENSE Trial Study Design. *Radiology* 2015; 277(2): 527–537. doi: 10.1148/radiol.2015141827

[13] Sakaie K, Zhou X, Lin J, Debbins J, Lowe M, Fox RJ. Technical Note: Retrospective reduction in systematic differences across scanner changes by accounting for noise floor effects in diffusion tensor imaging. *Medical Physics* 2018; 45(9): 4171–4178. doi: 10.1002/mp.13088

[14] Tax CM, Grussu F, Kaden E, et al. Cross-scanner and cross-protocol diffusion MRI data harmonisation: A benchmark database and evaluation of algorithms. *NeuroImage* 2019; 195: 285–299. doi: 10.1016/j.neuroimage.2019.01.077

[15] Mirzaalian H, Ning L, Savadjiev P, et al. Multi-site harmonization of diffusion MRI data in a registration framework. *Brain Imaging and Behavior* 2018; 12(1): 284–295. doi: 10.1007/s11682-016-9670-y

[16] Collier Q, Veraart J, Jeurissen B, et al. Diffusion kurtosis imaging with free water elimination: A bayesian estimation approach. *Magnetic Resonance in Medicine* 2018; 80(2): 802–813. doi: 10.1002/mrm.27075

[17] Sakaie K, Lowe M. Retrospective correction of bias in diffusion tensor imaging arising from coil combination mode. *Magnetic Resonance Imaging* 2017; 37: 203–208. doi: 10.1016/j.mri.2016.12.004

[18] Veraart J, Fieremans E, Novikov DS. Diffusion MRI noise mapping using random matrix theory. *Magnetic Resonance in Medicine* 2016; 76(5): 1582–1593. doi: 10.1002/mrm.26059

[19] Koay CG, Özarslan E, Pierpaoli C. Probabilistic Identification and Estimation of Noise (PIESNO): A self-consistent approach and its applications in MRI. *Journal of Magnetic Resonance* 2009; 199(1): 94–103. doi: 10.1016/j.jmr.2009.03.005

[20] Tabelow K, Voss HU, Polzehl J. Local estimation of the noise level in MRI using structural adaptation. *Medical Image Analysis* 2015; 20(1): 76–86. doi: 10.1016/j.media.2014.10.008

[21] Koay CG, Özarslan E, Basser PJ. A signal transformational framework for breaking the noise floor and its applications in MRI. *Journal of Magnetic Resonance* 2009; 197(2): 108–119. doi: 10.1016/j.jmr.2008.11.015

[22] St-Jean S, Coupé P, Descoteaux M. Non Local Spatial and Angular Matching: Enabling higher spatial resolution diffusion MRI datasets through adaptive denoising. *Medical Image Analysis* 2016; 32(2016): 115–130. doi: 10.1016/j.media.2016.02.010

[23] Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain.. *NeuroImage* 2012; 61(4): 1000–16. doi: 10.1016/j.neuroimage.2012.03.072

[24] Landman B, Bazin PL, Prince J. Diffusion tensor estimation by maximizing rician likelihood. *Proceedings of the IEEE International Conference on Computer Vision* 2007: 1–8. doi: 10.1109/ICCV.2007.4409140

[25] St-Jean S, De Luca A, Viergever MA, Leemans A. Automatic, Fast and Robust Characterization of Noise Distributions for Diffusion MRI. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G. , eds.)Springer International Publishing 2018: 304–312. doi: 10.1007/978-3-030-00928-1_35

[26] Gudbjartsson H, Patz S. The rician distribution of noisy mri data. *Magnetic Resonance in Medicine* 1995; 34(6): 910–914. doi: 10.1002/mrm.1910340618

[27] Heidemann RM, Anwander A, Feiweier T, Knösche TR, Turner R. k-space and q-space: Combining ultra-high spatial and angular resolution in diffusion imaging using ZOOPPA at 7T. *NeuroImage* 2012; 60(2): 967–978. doi: 10.1016/j.neuroimage.2011.12.081

[28] Aja-Fernández S, Vegas-Sánchez-Ferrero G, Tristán-Vega A. Noise estimation in parallel MRI: GRAPPA and SENSE.. *Magnetic resonance imaging* 2014; 32(3): 281–90. doi: 10.1016/j.mri.2013.12.001

[29] Weisstein EW. Gamma Distribution. From MathWorld—A Wolfram Web Resource http://mathworld.wolfram.com/GammaDistribution.html Last accessed 2017-10-09

[30] Thom HCS. A Note on the Gamma Distribution. *Monthly Weather Review* 1958; 86(4): 117–122.

[31] Caruyer E, Daducci A, Descoteaux M, Houde JC, Thiran JP, Verma R. Phantomas: a flexible software library to simulate diffusion MR phantoms. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM'14)*;17: 6407 2014.

[32] Maier-Hein KH, Neher PF, Houde JC, et al. The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications* 2017; 8(1): 1349. doi: 10.1038/s41467-017-01285-x

[33] Neher PF, Laun FB, Stieltjes B, Maier-Hein KH. Fiberfox: Facilitating the creation of realistic white matter software phantoms. *Magnetic Resonance in Medicine* 2014; 72(5): 1460–1470. doi: 10.1002/mrm.25045

[34] Aja-Fernández S, Tristán-Vega A. A review on statistical noise models for Magnetic Resonance Imaging. *LPI, ETSI Telecomunicacion, Universidad de Valladolid, Spain, Tech. Rep* 2015(TECH-LPI2013-01): 1–30.

[35] St-Jean S, De Luca A, Tax CMW, Viergever MA, Leemans A. Datasets for 'Automated characterization of noise distributions in diffusion MRI data'. *Zenodo* 2018. doi: 10.5281/zenodo.2483105

[36] Descoteaux M, Angelino E, Fitzgibbons S, Deriche R. Regularized, fast, and robust analytical Q-ball imaging. *Magnetic Resonance in Medicine* 2007; 58(3): 497–510. doi: 10.1002/mrm.21277

[37] Pieciak T, Rabanillo-Viloria I, Aja-Fernandez S. Bias correction for non-stationary noise filtering in MRI. In: *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*: 307–310IEEE 2018

[38] Lønning K, Putzky P, Sonke JJ, Reneman L, Caan MW, Welling M. Recurrent inference machines for reconstructing heterogeneous MRI data. *Medical Image Analysis* 2019; 53: 64–78. doi: 10.1016/j.media.2019.01.005

[39] Hutchinson EB, Avram AV, Irfanoglu MO, et al. Analysis of the effects of noise, DWI sampling, and value of assumed parameters in diffusion MRI models.. *Magnetic resonance in medicine* 2017; 78(5): 1767–1780. doi: 10.1002/mrm.26575

[40] Todd N, Moeller S, Auerbach EJ, Yacoub E, Flandin G, Weiskopf N. Evaluation of 2D multiband EPI imaging for high-resolution, whole-brain, task-based fMRI studies at 3T: Sensitivity and slice leakage artifacts. *NeuroImage* 2016; 124(Pt A): 32–42. doi: 10.1016/j.neuroimage.2015.08.056

[41] Barth M, Breuer F, Koopmans PJ, Norris DG, Poser BA. Simultaneous multislice (SMS) imaging techniques.. *Magnetic resonance in medicine* 2016; 75(1): 63–81. doi: 10.1002/mrm.25897

[42] Tibshirani RJ, Taylor J. The solution path of the generalized lasso. *The Annals of Statistics* 2011; 39(3): 1335–1371. doi: 10.1214/11-AOS878

[43] Papoulis A. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education 1991.

[44] Minka TP. Estimating a Dirichlet distribution. *Microsoft research* 2012.

[45] Leemans A, Jeurissen B, Sijbers J, Jones D. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine* 2009; 17(2): 3537.