

RESEARCH ARTICLE OPEN ACCESS

Reproducibility and Reliability of Free-Water-Corrected Diffusion Tensor Imaging of the Brain: Revisited

Tomasz Pieciak  | Guillem París  | Santiago Aja-Fernández  | Antonio Tristán-Vega 

Laboratorio de Procesado de Imagen (LPI), ETSI Telecomunicación, Universidad de Valladolid, Valladolid, Spain

Correspondence: Tomasz Pieciak (tpieciak@tel.uva.es)

Received: 6 August 2025 | **Revised:** 21 November 2025 | **Accepted:** 28 November 2025

Keywords: brain | diffusion tensor imaging | free-water | reliability | reproducibility | separability | white matter

ABSTRACT

Diffusion tensor imaging (DTI) corrected for the free-water (FW) enables the separation of a hindered Gaussian-like profile from an isotropic component, which represents diffusion found in cerebrospinal and interstitial fluids within the extracellular space of grey and white matter. The assessment of the reproducibility and reliability properties of FW-corrected DTI is a crucial factor in demonstrating the potential clinical utility of this refinement, particularly considering the examinations across multiple medical centres. This paper explores the variability, reliability, and separability properties of free-water volume fraction (FWVF) and FW-corrected DTI-based measures in healthy human brain white matter using publicly available test-retest databases acquired in (1) intra-scanner, (2) intra-scanner longitudinal and (3) inter-scanner settings under varying acquisition schemes. Three different estimation techniques to retrieve the FW-corrected DTI parameters tailored to single- or multiple-shell diffusion-sensitising magnetic resonance (MR) acquisitions are investigated: (i) a direct optimization of bi-tensor signal representation in the variational framework, (ii) the region contraction-based approach and (iii) the spherical means technique combined with a correction of diffusion-weighted MR signal prior to DTI estimation. We found the previous suggestion that the FW correction to DTI in a single-shell diffusion-weighted MR acquisition improves the repeatability of DTI-based measures may be data- and methodology-dependent, and does not generalise to multiple-shell scenarios. The study also confirms that the single-shell variational FW-correction method fails to retrieve meaningful information from the mean diffusivity (MD) parameter. In contrast, the combined FW-correction scheme reduces the biological variability of MD, regardless of whether DTI is estimated from single- or multiple-shell data, given that the FWVF used for the correction in both cases is derived from multiple-shell acquisitions. Our experiments have shown that the most reliable and repeatable/reproducible measures, while preserving a moderate separability property, are fractional anisotropy and axial diffusivity estimated in a multiple-shell variant under a combined FW-correction scheme. On the contrary, our results show evidence that the least reliable measures are the mean diffusivity estimated using any FW-correction procedure, as well as the FWVF parameter itself. These results can be used to establish the direction for selecting the most attractive FW-correction DTI scheme for clinical applications in terms of the variability-reliability-separability criterion.

1 | Introduction

Diffusion-weighted magnetic resonance imaging (MRI) is a well-established medical modality that enables the non-invasive probing of random motion of water molecules

in vivo, particularly in brain tissue (Le Bihan and Johansen-Berg 2012). A common approach employed to represent the diffusion-weighted MR signal is single-compartment diffusion tensor imaging (DTI) by providing a set of unique quantitative measures summarising the directional water diffusion

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Human Brain Mapping* published by Wiley Periodicals LLC.

process (Basser et al. 1994; Westin et al. 2002). Extending the standard DTI to the so-called bi-tensor representation has facilitated the separation of hindered diffusion depicted with a tensor-based Gaussian-like profile and the isotropic component, which illustrates the diffusion found in cerebrospinal and interstitial fluids within the extracellular space of grey and white matter (Pierpaoli and Jones 2004; Pasternak et al. 2009). This separation is possible due to suitably selected numerical optimization schemes that enable one to compute the directional DTI profile and the free-water volume fraction (FWVF), a scalar parameter illustrating the fitted isotropic fraction of the bi-tensor representation to the diffusion-sensitised MR signal. Such computations are possible both for single- (Pasternak et al. 2009) and multiple-shell acquisitions (Pasternak et al. 2012; Hoy et al. 2014; Bergmann et al. 2020; Tristán-Vega et al. 2022).

The FW-corrected DTI has been an essential tool in clinical applications, particularly the single-shell FW-corrected DTI, which is primarily employed in cognitive performance evaluation (Maillard et al. 2019), modelling neurodegenerative disorders such as Parkinson's (Ofori et al. 2015) or Alzheimer's (Bergamino et al. 2021; Nakaya et al. 2022), brain ageing (Metzler-Baddeley et al. 2012; Chad et al. 2018), schizophrenia (Carreira Figueiredo et al. 2022) or detecting first episodes of psychosis (Lyall et al. 2018), as the evidence indicates the method improves the specificity of DTI measures (Bergamino et al. 2021; Chad et al. 2023) and the reliability and accuracy of tractometry in brain ageing (Chang et al. 2025). However, the single-shell FW-corrected DTI by Pasternak et al. (2009) has been lately put into stake because it is unclear whether it accurately captures actual anatomy-related changes in the diffusivity profile of brain white matter. For instance, Golub et al. (2021) have shown that the single-shell FW-correction scheme can yield plausible results, but the optimization procedure heavily depends on the initialization strategy, potentially affecting the method's specificity. Correia et al. (2024) discovered a flattening effect of FW-corrected MD profiles with age and FW-corrected FA strong positive correlations with age in some regions, which seem not to be present in a multiple-shell scenario. The authors of these works advocate for considering the multiple-shell rather than single-shell tailored schemes once correcting the DTI for the FW component. As an example, the multiple-shell variant has explicitly demonstrated benefits over the single-shell in brain age predictions (Nemmi et al. 2022).

Although numerous FW-corrected DTI applications have already been demonstrated in a wide range of clinical scenarios, much less attention has been paid to the reproducibility and reliability of the measures obtained from this refinement. Until now, the intra-scanner repeatability (or more generally, the inter-scanner reproducibility) and reliability studies have concentrated primarily on the standard DTI-based metrics under specific acquisition conditions, such as variable magnetic strength fields (Grech-Sollars et al. 2015; Venkatraman et al. 2015; Jakab et al. 2017), multi-band acquisition scheme (Duan et al. 2015), and spatial resolution (Shahim et al. 2017; Zhong et al. 2024), considering particular cohorts like fetal brains (Jakab et al. 2017), neonates (Merisaari et al. 2019) or in the older population (Laguna et al. 2020), in a longitudinal

scenario (Boudreau et al. 2025), across multiple centres (Grech-Sollars et al. 2015) or even under scanner relocation (Melzer et al. 2020). Apart from providing detailed quantitative results, some other studies proposed vendor-agnostic sequences to reduce inter-scanner variabilities (Liu et al. 2024) or drew conclusions on possible solutions to post hoc improve the reproducibility or reliability of DTI-based parameters. For instance, Jakab et al. (2017) presented a negative impact of fetal head motion on DTI repeatability and suggested that this effect can be partially palliated with motion correction algorithms. Ades-Aron et al. (2025) have shown that proper denoising of diffusion-weighted MR data (either in complex or magnitude space) leads to a significant reduction in variability of DTI-based measures and increases statistical power for low signal-to-noise ratio voxels in intra- and inter-scanner, and inter-protocol studies. In the work by Albi et al. (2017), it has been suggested that suppressing the FW component in DTI using the single-shell approach by Pasternak et al. (2009) leads to reduced repeatability errors of standard metrics such as fractional anisotropy (FA) or mean diffusivity (MD) *on average* approximately at 1%pt.¹ However, taking a closer look at the experimental methodology, the work by Albi et al. (2017) may lead to more inquiries than answers, as different numerical optimization schemes have been arranged to compare the standard DTI-based measures and the FW-corrected ones, i.e., the linear least squares procedure versus the non-linear optimization in a variational framework. Besides, the work utilises a questionable single-shell-based FW correction scheme and uses only a test-retest database acquired in a longitudinal intra-scanner scenario.

In this paper, we revisit the study by Albi et al. (2017) and explore the repeatability and reproducibility, reliability and separability of FW-corrected DTI under three different approaches used to correct the DTI for the FW component, namely (1) a single-shell variational scheme by Pasternak et al. (2009), (2) the multiple-shell region contraction-based approach by Hoy et al. (2014), and (3) a prior correction of diffusion-weighted MR signal for the FW component estimated using the spherical means technique by Tristán-Vega et al. (2022), preceded by a standard DTI estimation from FW-corrected signal. We emphasise that single-shell estimation should not be regarded as equivalent to multiple-shell estimation, as it fails to capture the full spectrum of the biological variability of brain tissue. Our study encompasses three databases acquired in (i) intra-scanner, (ii) intra-scanner longitudinal and (iii) inter-scanner scenarios, as well as under various experimental setups.

2 | Materials and Methods

2.1 | In Silico Data

We generate in silico data by composing the signals originating from cellular and free-water compartments, following the formulation:

$$\frac{A(b, \mathbf{g})}{A(0)} = \underbrace{\left(1-f\right) \sum_{k \in B} \alpha_k A_k(b, \mathbf{g})}_{\text{cellular compartment}} + \underbrace{f \exp(-bD_0)}_{\text{free-water compartment}}, f \in [0, 1] \quad (1)$$

with $D_0 = 3.0 \times 10^{-3} \text{ mm}^2/\text{s}$ being the apparent diffusion coefficient for water at a temperature of approximately 37°C, f is the FWVF parameter, and $A_k(b, \mathbf{g})$ is the cellular compartment integrating intra- and extra-axonal parts spherically convolved with the fibre orientation density function $\Phi(\mathbf{g})$

$$A_k(b, \mathbf{g}) = v_k \Phi(\mathbf{g}) \otimes \exp(-b\mathbf{g}^T \mathbf{D}_{\text{ic},k} \mathbf{g}) + (1 - v_k) \Phi(\mathbf{g}) \otimes \exp(-b\mathbf{g}^T \mathbf{D}_{\text{ec},k} \mathbf{g}), \quad (2)$$

where the tensor $\mathbf{D}_{\text{ic},k}$ represents the intra-cellular part and is characterised with two perpendicular null eigenvalues, $\lambda_{\text{ic},k}^{\text{perp}} = 0$, and the tensor $\mathbf{D}_{\text{ec},k}$ is axis-symmetric with a non-zero perpendicular diffusivity $\lambda_{\text{ec},k}^{\text{perp}}$. The cellular compartment defined in Equation (1) can be represented using a single- ($B = \{1\}$) or two-fibre bundle ($B = \{1, 2\}$) with partial fractions $\alpha_1 + \alpha_2 = 1$. The signal given by Equation (1) is then contaminated with a Rician noise (Aja-Fernández and Vegas-Sánchez-Ferrero 2016; Pieciak et al. 2017), as follows $S(b, \mathbf{g}) = |A(b, \mathbf{g}) + N_{\text{re}} + j \cdot N_{\text{im}}|$ with $N_{\text{re}}, N_{\text{im}} \sim N(0, \sigma^2)$. For more information on synthetic data generation scheme, see the Supporting Informations in Pieciak et al. (2023).

We generate two sets of synthetic data representing a single fibre bundle ($\alpha_2 = 0$) and a two-fibre bundle case ($\alpha_1 = \alpha_2 = 0.5$). The experimental setup for each dataset comprises 18 replicas of baseline samples and diffusion-weighted samples generated at $b = \{600, 1200\} \text{ s/mm}^2$ and 90 gradient directions per shell distributed according to the sampling defined by the HCP WuMinn project (Van Essen et al. 2013). The signal-to-noise of the data ratio has been defined in terms of baseline signal as $\text{SNR} = A(0)/\sigma$ and fixed to 300. For each set, we also generate noiseless reference data with no FW component, i.e., $f = 0$ in Equation (1).

2.2 | In Vivo Data

We use three publicly available diffusion-weighted MR databases compatible with single- and multiple-shell DTI and FW-corrected DTI. Two databases, namely MICRA (Koller et al. 2021) and Magdeburg (Lehmann et al. 2021), cover repeated scans for each subject acquired using a single scanner (i.e., intra-scanner acquisitions). The third database, the ZJU (Tong et al. 2020), covers multiple scans across identical scanners located in different centres (i.e., inter-scanner scenario).

The acquisition setup for the MICRA database was as follows. The database covers inter-session repeated scans from a single centre. Six healthy volunteers (3F/3M) aged 24–30 were scanned, five times each using a 3T Connectom MRI research scanner (Siemens Healthcare, Erlangen, Germany) equipped with an ultra-strong gradient system at 300 mT/m. Acquisition protocol: single-shot spin-echo echo planar imaging sequence, anterior-posterior (AP) phase-encoding direction, repetition time (TR): 3000 ms, echo time (TE): 59 ms, pulse separation/pulse duration Δ/δ : 24/7 ms, field of view (FOV): 220 × 220 mm², matrix size: 110 × 110, voxel size: 2 × 2 × 2 mm³, b -values: (200, 500, 1200, 2400, 4000, 6000) s/mm² with (20, 20, 30, 61, 61, 61) gradient directions, respectively, 11 non-diffusion-weighted scans in AP direction repeated every twentieth volume and

two non-diffusion-weighted scans in PA direction. For this study, we use only the diffusion-weighted MR samples acquired at $b = \{500, 1200\} \text{ s/mm}^2$.

The acquisition details of two other databases are provided in the Supporting Information S1.

2.3 | Data Preprocessing

The MICRA dataset was preprocessed using the following pipeline: (1) noise removal using the Marčenko-Pastur Principal Component Analysis technique over a window of size 5 × 5 × 5 voxels (MP-PCA; Veraart, Novikov, et al. 2016; Veraart, Fieremans, and Novikov 2016), (2) Gibbs ringing artefacts correction (Kellner et al. 2016), (3) Rician bias correction applied voxel-wise with the formula (Pieciak et al. 2018)

$$M_c(\mathbf{x}) = \frac{1}{2} \left[\langle M(\mathbf{x}) \rangle + \left| \langle M(\mathbf{x}) \rangle^2 - 2\sigma^2(\mathbf{x}) \right|^{1/2} \right]. \quad (3)$$

The correction given by Equation (3) was applied under the assumption that the MP-PCA algorithm generates a proxy for the expectation of signal magnitude, $\langle M(\mathbf{x}) \rangle$, $\sigma^2(\mathbf{x})$ is the spatially-dependent noise standard deviation, (4) susceptibility-induced distortions estimation using the FSL FMRIB Software Library v6 topup tool (Analysis Group, FMRIB, Oxford, UK; Andersson et al. 2003; Smith et al. 2004), (5) head movements and eddy current distortions correction using the FSL eddy (Andersson et al. 2016), and (6) B1 field inhomogeneity correction using the N4 algorithm (Tustison et al. 2010).

2.4 | Free-Water-Corrected Diffusion Tensor Imaging

The FW-corrected DTI is modeled using a two-component representation characterising hindered diffusion using a diffusion tensor, and free diffusion represented by a mono-exponential decay (Pierpaoli and Jones 2004; Pasternak et al. 2009):

$$\frac{S(b, \mathbf{g})}{S(0)} = (1 - f) \exp(-b\mathbf{g}^T \mathbf{D} \mathbf{g}) + f \exp(-bD_0), f \in [0, 1] \quad (4)$$

with $S(b, \mathbf{g})$ being the diffusion-weighted MR signal acquired in direction \mathbf{g} at b -value b , $S(0)$ is a non-diffusion-weighted MR signal, and \mathbf{D} is a symmetric semi-positive matrix of size 3 × 3. The two-component representation given by Equation (4) reduces to the standard DTI for $f = 0$.

2.5 | Estimation Methods: DTI and FW-Corrected DTI

The following methods are considered in this study to estimate FWVF, DTI and FW-corrected DTI:

1. Bi-tensor-S: joint estimation of FWVF and FW-corrected DTI in the variational framework from single-shell acquisitions (Pasternak et al. 2009); the learning rate has

been fixed to 0.005, the optimization uses 200 iterations and the initialization employs the tissue's MD prior set to $0.6 \times 10^{-3} \text{ mm}^2/\text{s}$ (Golub et al. 2021), unless otherwise stated,

2. Bi-tensor-M: joint estimation of FWVF and FW-corrected DTI via the non-linear least squares procedure from multiple-shell acquisitions (Hoy et al. 2014; Henriques et al. 2017),
3. SM: FWVF estimation using the spherical means technique from multiple-shell acquisitions (Tristán-Vega et al. 2022); the spherical harmonics decomposition at the order of $L = 6$ is computed using the inverse linear problem with the regularisation based on the Laplace-Beltrami operator and the regularisation weight set to $\lambda = 0.001$, the parallel diffusivity parameter is fixed to $\lambda_{\text{par}} = 2.0 \times 10^{-3} \text{ mm}^2/\text{s}$ and penalty term ν used to promote prolate convolution kernels was optimised to $\nu = 0.015$ for the MICRA database.
4. DTI: standard DTI estimated via the non-linear least squares (NLLS; Koay et al. 2006),
5. FW-DTI: customised FW-corrected DTI scheme using the NLLS based on a pre-estimated FWVF with the SM technique. This scenario assumes the diffusion-weighted MR signal is corrected for the isotropic component prior to the estimation procedure:

$$\tilde{S}(b, \mathbf{g}) = \left(1 - \hat{f}\right)^{-1} \left(\frac{S(b, \mathbf{g})}{\bar{S}(0)} - \exp(-bD_0) \right) + \exp(-bD_0), \quad (5)$$

where $\tilde{S}(b, \mathbf{g})$ is the normalised diffusion-weighted MR signal after removing the FW component, $\bar{S}(0)$ is the averaged signal across all non-diffusion-weighted MR acquisitions, \hat{f} is the FWVF estimated using the SM technique from multiple-shell data.

We calculate FA, MD, RD (radial diffusivity) and AD (axial diffusivity) measures for each DTI-based technique considered in the study, i.e., Bi-tensor-S, Bi-tensor-M, DTI and FW-DTI. For DTI and FW-DTI, we handle two variants: single- and multiple-shell data. In the former FW-DTI variant,

the FWVF is estimated using the SM approach from multiple-shell data, while the DTI-based parameters are derived from single-shell data. This configuration may reflect a realistic clinical scenario in which the acquisition protocol includes a standard DTI scheme approximately at $b = 1000 \text{ s/mm}^2$ and only a few directions at a lower b -value, around 500 s/mm^2 . Alternatively, if the study includes data acquired at a higher outer shell, for instance, $b = 2000 \text{ s/mm}^2$ (as in the ZJU or Magdeburg databases), it may be advantageous to estimate the FWVF from two shells and then obtain DTI-based measures from the shell at $b = 1000 \text{ s/mm}^2$. In the multiple-shell variant, both the FWVF and DTI parameters are computed from multiple-shell data. Table 1 summarises all methods used in the study.

We used the DIPY library v. 1.9.0 (<https://dipy.org>) for non-linear DTI and DIPY with in-house implementation for FW-DTI fitting procedures delivered in Python 3.11.5 (<https://www.python.org>), NumPy 1.26.4 (<https://numpy.org>) and SciPy 1.11.1 (<https://scipy.org>). The implementation of Bi-tensor-S followed that of (Golub et al. 2021; <https://github.com/mvgolub/FW-DTI-Beltrami>), while Bi-tensor-M employed the one provided with the DIPY library. The SM was estimated using the dMRI-Lab toolbox (<https://www.lpi.tel.uva.es/dmrlab>) run under MATLAB R2023b (The MathWorks Inc., Natick, MA).

2.6 | Data Registration

The FA volumes estimated using a standard DTI at $b = 1200 \text{ s/mm}^2$ for MICRA database were used to register the data to the Montreal Neurological Institute (MNI) space, so-called standard space. Specifically, we linearly registered FA volumes computed with the standard DTI from each scan to the FSL template FMRIB58_FA using the FSL flirt tool under seven degrees of freedom, normalised correlation cost function and spline interpolation. The linearly transformed FA volumes were then non-linearly deformed via the FSL fnirt. All measures from the subjects' native spaces were mapped to the MNI space using a trilinear interpolation. We then retrieved the white matter label from the Johns Hopkins University

TABLE 1 | Summary of FWVF, DTI and FW-corrected DTI estimation procedures used in the study.

Method	Single-shell			Multiple-shell		
	FWVF	DTI measures	FW-corrected DTI measures	FWVF	DTI measures	FW-corrected DTI measures
Bi-tensor-S (Pasternak et al. 2009)	✓		✓			
Bi-tensor-M (Hoy et al. 2014)				✓		✓
SM (Tristán-Vega et al. 2022)					✓	
DTI (Koay et al. 2006)		✓			✓	
FW-DTI			✓ 			✓

Note: The methods are classified into single- and multiple-shell data handled in the estimation process.  The DTI-based parameters are estimated from single-shell data, using the FWVF previously estimated from multiple-shell data.

(JHU) WM atlas (Mori et al. 2005) in the standard space and shrunk it using a morphological binary erosion operator with a cross-shaped kernel of size $3 \times 3 \times 3$ to eliminate potential misregistration outliers due to a partial volume effect. All scripting was carried out using the Arturo programming language 0.9.83 (<https://arturo-lang.io>).

2.7 | Variability, Separability and Reliability Assessment

Three characteristics of the FWVF and DTI-related measures are computed, namely reproducibility, reliability, and separability, all three in the standard space. First, we warp the measures from the subjects' native spaces to the standard space using the warping fields obtained from the data registration procedure. Then, we compute the spatially dependent characteristics mentioned above across the repeated scans. Repeatability and reproducibility are explained in terms of the variability index, which defines the general ability to replicate the measure across scans or scanners of the same subject, and they are expressed by the coefficient of variation (CoV). By definition, the smaller the variability, the higher the repeatability and reproducibility. In our study, we distinguish three scenarios: (1) intra-scanner repeatability—the inter-session scans are replicated using the same scanner with a short time interval between the scans, (2) inter-session longitudinal reproducibility—the scans are repeated using the same scanner with a longer time interval between the scans (e.g., several weeks) and possibly affected by confounding factors (e.g., changes in magnetic field drifts) and (3) inter-scanner reproducibility—the scans are repeated using the same scanner type, but installed in different locations. The reliability index reflects the measure's consistency, and it is expected to be high for measures that are stable in value across repeated scans. Decomposing the variance into within-subject and between-subjects variances, as defined by Zuo et al. (2019), enables the computation of a ratio that reflects how much of the total variance is explainable by actual inter-individual differences rather than noise or session variability. The last index, separability, illustrates the metric's ability to capture inter-subject discrepancies.

2.7.1 | Variability

The variability of a metric has been defined as the median value from the CoVs computed across all subjects $s = 1, \dots, S$

$$\text{Variability}(\mathbf{x}) = \text{median}_{s=1, \dots, S} \{ \text{CoV}_s(\mathbf{x}) \} \times 100[\%], \quad (6)$$

where the CoV for subject s , $\text{CoV}_s(\mathbf{x})$, is given by

$$\text{CoV}_s(\mathbf{x}) = \frac{\text{Std. dev}_s(\mathbf{x})}{\text{Mean}_s(\mathbf{x})}. \quad (7)$$

The sample mean $\text{Mean}_s(\mathbf{x})$ and sample standard deviation $\text{Std. dev}_s(\mathbf{x})$ are defined over the stack of M_s scans available for s – th subject (see Supporting Information S1 for details).

2.7.2 | Reliability

The reliability index has been defined using the formulation by Zuo et al. (2019)

$$\text{Reliability}(\mathbf{x}) = \frac{\text{var}_{s=1, \dots, S} \{ \text{Mean}_s(\mathbf{x}) \}}{\text{var}_{s=1, \dots, S} \{ \text{Mean}_s(\mathbf{x}) \} + \text{mean}_{s=1, \dots, S} \{ \text{Var}_s(\mathbf{x}) \}}, \quad (8)$$

where $\text{Var}_s(\mathbf{x})$ is the sample variance of the measure and $\text{var}_{s=1, \dots, S} \{ \cdot \}$ is the population variance calculated across sample means $\text{Mean}_s(\mathbf{x})$, i.e., the normalisation factor in the variance formula equals $1/S$.

2.7.3 | Separability

We define the separability index as follows:

$$\text{Separability}(\mathbf{x}) = \frac{\text{std. dev}_{s=1, \dots, S} \{ \text{Mean}_s(\mathbf{x}) \}}{\text{mean}_{s=1, \dots, S} \{ \text{Mean}_s(\mathbf{x}) \}} \times 100[\%], \quad (9)$$

where $\text{std. dev}_{s=1, \dots, S} \{ \cdot \}$ is the population standard deviation with the normalisation factor also given by $1/S$.

2.7.4 | Diffusion Kernel Density Estimation (DiffKDE)²

The KDE method estimates the probability density function from data samples in a non-parametric way. Given a random sample x_1, \dots, x_n from an unknown probability density function $f(x)$, a non-negative kernel function K and a bandwidth $h > 0$, the formula for the KDE at a point x_0 is given by (Hastie et al. 2009)

$$\hat{f}(x_0; h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_0 - x_i}{h}\right). \quad (10)$$

The KDE-based procedure must be applied with great care, as Silverman's rule-of-thumb bandwidth selection method (Silverman 1986) may yield suboptimal results for non-Gaussian histogram shapes. Thereupon, we employ an advanced DiffKDE approach, which uses the improved Sheather-Jones method to automatically determine the optimal smoothing parameter (Sheather and Jones 1991; Botev et al. 2010). The DiffKDE uses a linear diffusion process (i.e., heat equation) to model the density and minimises the asymptotic mean integrated squared error criterion to select an optimal smoothing parameter. Compared to KDE, which models the density as a sum of kernel functions centred at the data points, the DiffKDE models the density as a solution to a partial differential equation with initial conditions derived from the empirical density of the data.

We compute the DiffKDE over all data points from the white matter area in the standard space of the measure. Representing variability and reliability as density plots diminishes the

influence of potential misregistered voxels or those affected by partial-volume effects on the results.

3 | Experimental Results

This section wraps the experimental results generated for in silico data and MICRA database.

3.1 | In Silico Experiments

In the in silico experiment depicted in Figure 1, we examine different FW-correction DTI schemes and relate the FW-corrected MD and FA measures under $f = 0.2$ to the standard DTI-based equivalents, but obtained from the cellular model only (i.e., $f = 0$ in Equation 1). Among the multiple-shell-based

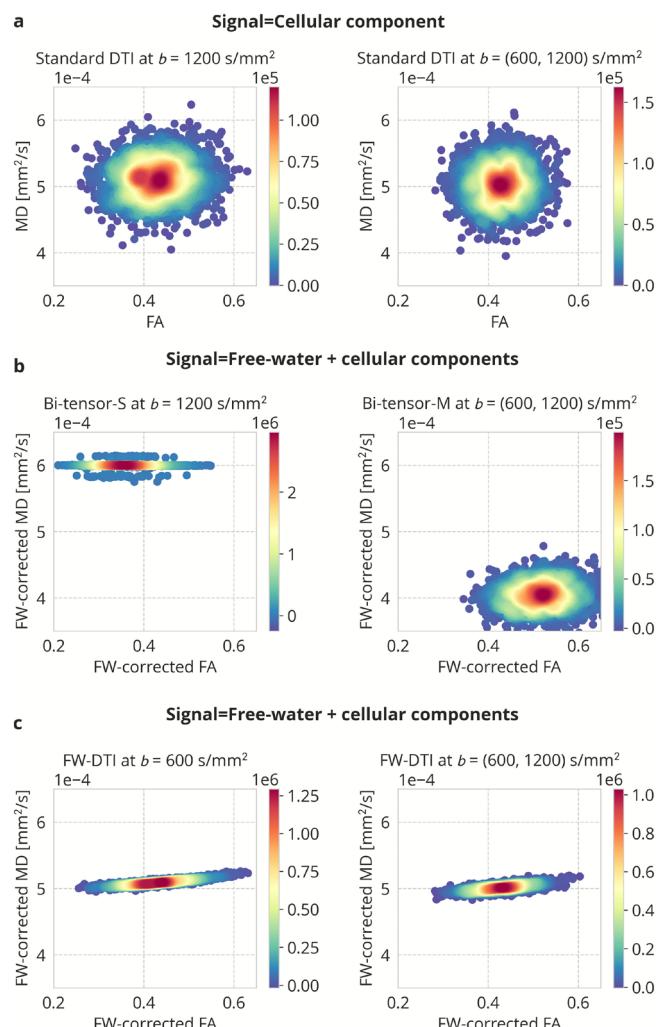


FIGURE 1 | The 2D density plots illustrate the experimental results from an *in silico* model under a two-fibre bundle configuration: (a) the MD and FA parameters estimated using the standard DTI from noiseless reference data covering the cellular component only. The FW-corrected MD and FA estimated from the signal covering FW ($f = 0.2$) and cellular components: (b) Bi-tensor-S and Bi-tensor-M approaches, and (c) FW-DTI customised scheme according to Equation (5). In total, 15^3 samples have been used to obtain a single 2D density plot.

techniques, the FW-DTI customised scheme yields smaller discrepancies of MD and FA from the standard DTI-based MD and FA than the B-tensor-M technique. However, the FW-corrected MD using the FW-DTI scheme is more flattened compared to the FW-corrected MD obtained via the Bi-tensor-M. Regarding the single-shell techniques evaluated at $b = 1200\text{s/mm}^2$, the flattening effect is more evident, albeit the FW-corrected MD using the FW-DTI scheme also follows the general trend of standard DTI-based MD. The Bi-tensor-S method converges to the tissue's prior, i.e., the FW-corrected parameter oscillates around the value of $0.6 \times 10^{-3}\text{mm}^2/\text{s}$.

3.2 | Visual Inspection of the Measures

We now move to in vivo experiments, first by visually inspecting the measures in the subject's native coordinate system. We have chosen a single subject from the MICRA database (sub-01, ses-01) and displayed the measures in Figure 2. In Figure 2a, we illustrate the FWVF estimated under multiple-shell data at $b = \{500, 1200\}\text{s/mm}^2$ using the SM and Bi-tensor-M approaches, and single-shell data at $b = 1200\text{s/mm}^2$ with the Bi-tensor-S. The results obtained with Bi-tensor-M and Bi-tensor-S differ despite both techniques are conceptually based on a direct optimization of Equation (4), albeit using distinct numerical schemes. Next, in Figure 2b,c, the DTI measures estimated from multiple- and single-shell acquisitions have been illustrated in the following rows: (i) standard DTI (no free-water correction), (ii) FW-corrected DTI according to Equation (5) and (iii) Bi-tensor-M or Bi-tensor-S approach. We observe increased FA values and decreased MD, AD and RD over the brain using all three FW-correction schemes examined, i.e., the FW-DTI, Bi-tensor-M and Bi-tensor-S. Besides, the FW-corrected MD parameter estimated using the Bi-tensor-S approach exhibits flattened spatial characteristics over the brain (see Figure 2c). The median (interquartile range) value of FW-corrected MD computed with the Bi-tensor-S approach over the WM area is 0.6006 (0.6 – 0.6015).

3.3 | Variability Maps

In the experiment shown in Figure 3, we visually inspect spatially dependent variability maps of the measures in the standard space computed according to Equation (6) for the MICRA dataset.

In Figure 3a, we illustrate the variability of the FWVF estimated from: (i) multiple-shell data using the SM, (ii) multiple-shell data using the Bi-tensor-M procedure and (iii) single-shell data via the Bi-tensor-S. Among the three techniques mentioned above, the highest variability is observed with the Bi-tensor-M approach, with the median CoV over the white matter area at 19.9 %, compared to SM (CoV = 11.78 %) and Bi-tensor-S (CoV = 11.57 %).

Next, in Figure 3b,c, we present the variability of DTI metrics computed with the standard DTI and under a FW-correction from multiple- and single-shell data. The first observation is that the variability of the FA parameter follows different spatial characteristics compared to MD, AD and RD. Specifically, the variability of FA is remarkably lower in the white matter compared

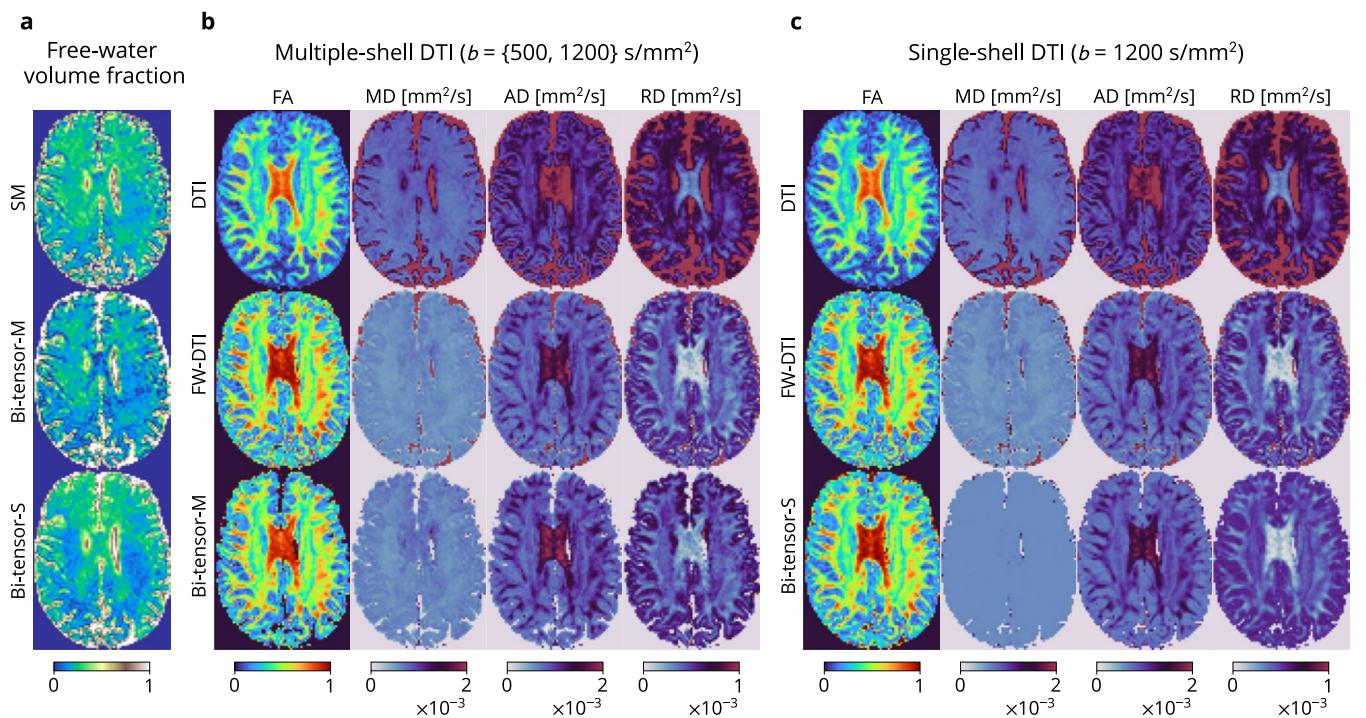


FIGURE 2 | Estimated microstructural measures for a selected MICRA acquisition (sub-01, ses-01, slice 36): (a) FWVF estimated from multiple-shell data (SM and Bi-tensor-M approaches) and single-shell data (Bi-tensor-S), (b) DTI-based measures estimated from multiple-shell data using a standard DTI, and two free water-correction methodologies: FW-DTI according to Equation (5) and Bi-tensor-M, and (c) DTI-based measures estimated from single-shell data using the standard DTI, FW-DTI and Bi-tensor-S. The FWVF parameter for FW-DTI was pre-estimated using the SM approach from multiple-shell data in both scenarios presented in panels (b) and (c).

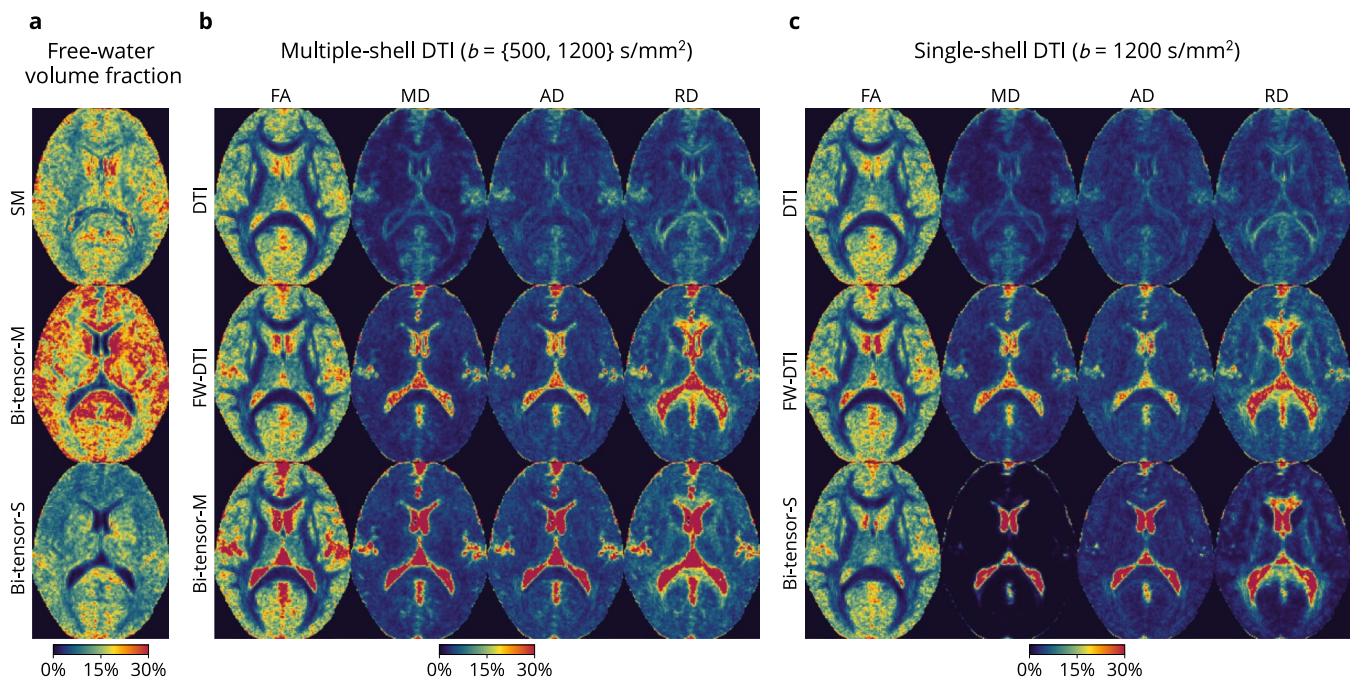


FIGURE 3 | Inter-session variability maps of the measures defined in the standard space for MICRA database according to the coefficient of variation defined by Equation (6): (a) FWVF estimated from multiple-shell data (SM and Bi-tensor-M approaches) and single-shell data (Bi-tensor-S), (b) DTI-based measures estimated from multiple-shell data using a standard DTI, and two FW-correction methodologies: FW-DTI according to Equation (5) and Bi-tensor-M, and (c) DTI-based measures estimated from single-shell data with the standard DTI, FW-DTI and Bi-tensor-S. The FWVF parameter for FW-DTI was pre-estimated using the SM approach from multiple-shell data in both scenarios presented in panels (b) and (c).

to the grey matter, while the variability of MD, AD, and RD is more homogeneous across the white and grey matter areas. The median CoV of multiple-shell-based FA parameter varies only

slightly across the methods considered: 4.35% (DTI), 4.27% (FW-DTI) and 4.56% (Bi-tensor-M). For the multiple-shell MD, the results are more pronounced: 2.76% (DTI), 3.22% (FW-DTI)

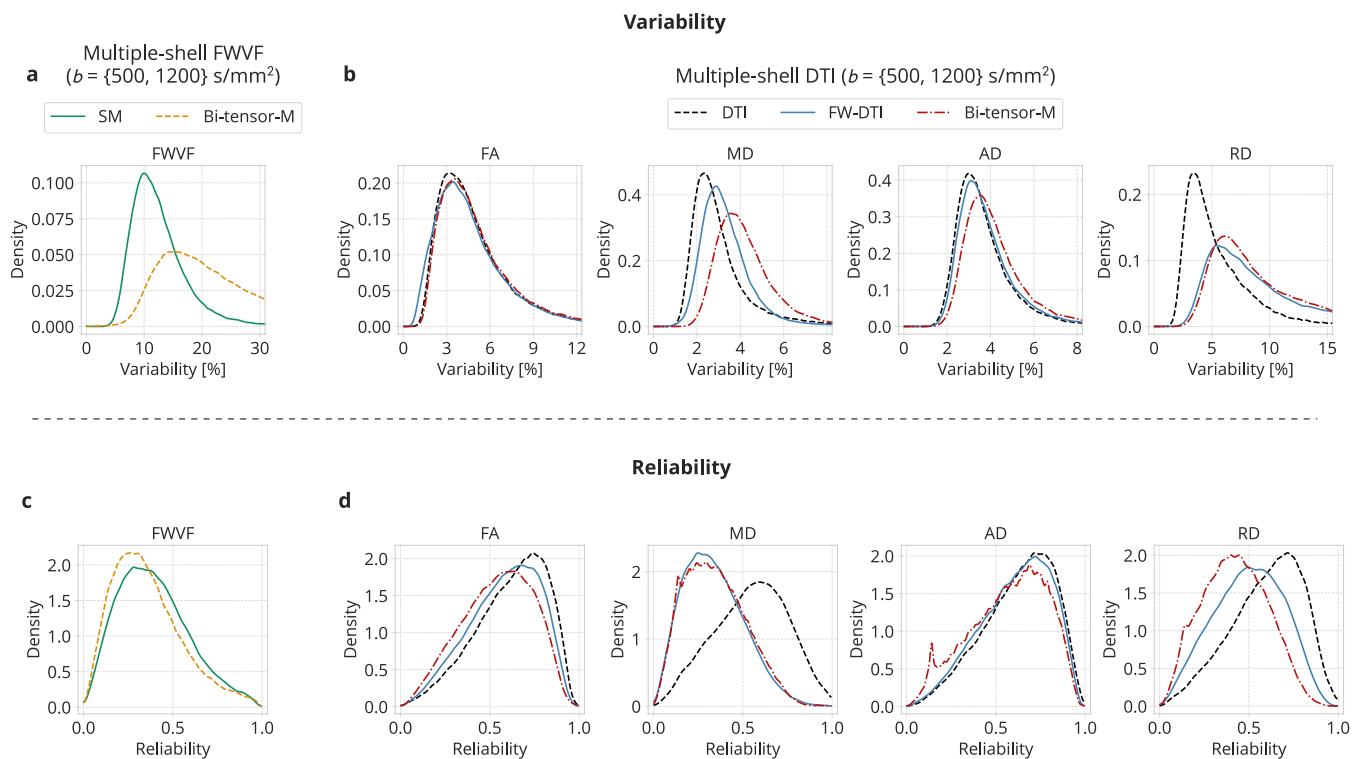


FIGURE 4 | Inter-session kernel density-based variability (top) and reliability (bottom) indices computed for MICRA database over the white matter area: (a, c) FWVF estimated from multiple-shell data (SM and Bi-tensor-M approaches), (b, d) DTI-based measures estimated from multiple-shell data using standard DTI, FW-DTI according to Equation (5) and Bi-tensor-M. The FWVF parameter for FW-DTI was pre-estimated from multiple-shell data using the SM approach.

and 4.11% (Bi-tensor-M). However, in the case of MD estimated with Bi-tensor-S, the results are contentious—the experiment has revealed the smallest variability across all the measures considered in the study.

3.4 | Density-Based Variability and Reliability Indices

From this point onward, only the results for the multiple-shell data are presented in the main paper. For the extended results, including single-shell data, we refer the reader to the Supporting Information S1.

The following two experiments assess the variability and reliability of the measures across the white matter in the form of density plots. The plots have been estimated using the DiffKDE approach (Sheather and Jones 1991; Botev et al. 2010) for multiple-shell measures previously illustrated in Figure 2 and are depicted in Figure 4.

The FWVF parameter, no matter how it is estimated, is poorly reproducible over the white matter, to a much lesser extent than the DTI and FW-corrected DTI measures. Density plots representing the variability (i.e., CoV parameter) of the FWVF parameter reach their peaks at $\text{CoV} = 10.08\%$ (SM), and $\text{CoV} = 15\%$ (Bi-tensor-M). In general, the plots representing the variability for the measures considered in the study exhibit unimodal, predominantly positively skewed densities. Our results demonstrate the FW-correction primarily does not

improve the repeatability/reproducibility of DTI-related measures compared to the standard DTI case, and even if it does, the changes are minute. We found the contrary behaviour of multiple-shell Bi-tensor-M scheme—the FW-correction has led to an increase in the variability of MD/AD/RD measures (see Figure 4b).

Our experiments also reveal that the measures characterised with the lowest reliability are multiple-shell FWVF and FW-corrected MD (see Figure 4c,d). Notably, the FW-correction to MD has led to a significant decrease in the reliability parameter, regardless of the numerical method used to estimate the measure, i.e., the reliability peak is far below 0.5. At large, the plots representing the reliability index exhibit unimodal and negatively skewed densities for the evaluated measures (cf. to the densities of the variability parameter). In general, the reliability of standard DTI-based measures is roughly equal to or better than the reliability of equivalent measures corrected for the FW component.

3.5 | Variability, Reliability and Separability

We now proceed to quantitative experiments that demonstrate two relationships: (i) reliability versus variability and (ii) separability versus variability.

In the first experiment depicted in Figure 5, we put together the population's first (25th percentile), second (median) and third (75th percentile) quartiles computed for the variability

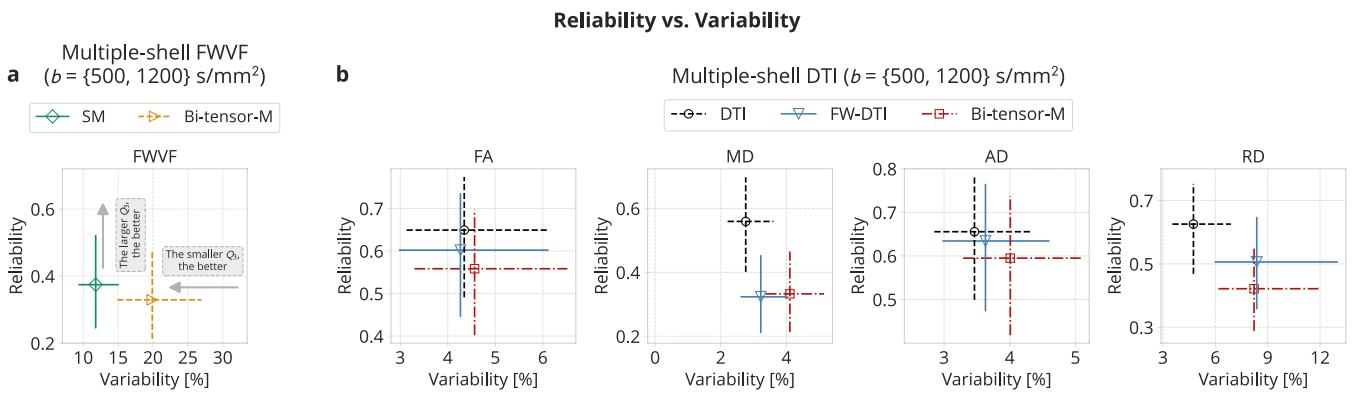


FIGURE 5 | Inter-session reliability versus variability plots computed for MICRA database over the white matter area: (a) FWVF estimated from multiple-shell data (SM and Bi-tensor-M approaches), (b) DTI-based measures estimated from multiple-shell data using standard DTI, FW-DTI according to Equation (5) and Bi-tensor-M. The FWVF parameter for FW-DTI was pre-estimated from multiple-shell data using the SM approach. The markers present median values calculated over the white matter area, while the horizontal and vertical lines represent distances between the first $Q_1 = 0.25$ (25th percentile) and third $Q_3 = 0.75$ (75th percentile) quartiles. The horizontal and vertical ranges for plots representing multiple- and single-shell equivalent cases have been fixed.

and reliability indices. The ‘population’ is understood here as all voxels taken from the white matter region defined in the standard space. Note that the smaller the third quartile for the variability index, the better, while the larger the third quartile for the reliability, the better. As mentioned in the previous section, the FWVF is characterised by the highest variability among all measures, typically several times higher than DTI-based measures, and a relatively low reliability. However, the variability and reliability indices vary between the methods used to estimate the FWVF. Nevertheless, the SM technique has shown superior behaviour over the Bi-tensor-M in terms of reliability, though the variability index does not give a clear answer to which method varies more. The multiple-shell DTI-based measures with no FW correction are more reproducible and more reliable than FW-corrected equivalents using the Bi-tensor-M approach. The experiments show evidence that correcting DTI measures for the FW component via Equation (5) would be a more appropriate solution in terms of variability-reliability criterion, given that it potentially avoids such declines in reproducibility and reliability indexes, as observed in the Bi-tensor-M approach.

In the final experiment, we strive to establish optimal measures in terms of reliability, variability and separability indices. In the bar charts presented in Figure 6a, we relate the median reliability to median variability. Both indices are ordered according to the median reliability. The indices were computed over the white matter area in the standard space. In general, the results presented in charts-based plots illustrate that the highest reliability measures are generally reproducibly approving, with the variability being less than 5%. In this class, one can identify AD, FA, and RD computed from standard DTI, with the last two measures characterised by higher variability and separability than the AD. Contrarily, the measures characterised by low reliability, such as the FWVF, are typically highly variable.

Considering the FW-corrected DTI, we observe FA and AD rectified with the customised scheme from multiple-shell data are more reliable, reproducible and somewhat separable compared to the Bi-tensor-M approach. The previously discussed

FW-corrected MD with the Bi-tensor-S, although it has revealed extremely low variability, is a non-reliable measure (see the extended results in Figure S14). Our experiments have also demonstrated that the FW-corrected MD under Bi-tensor-M and FW-DTI are also the least reliable among the DTI measures. Overall, the MD measure, computed in any FW-correction variant is the least reliable and separable parameter among all DTI-based parameters considered in this study. The FWVF is also non-reliable but seems to be a highly separable parameter (see Figure 6b). Interestingly, the FW-corrected RD obtained with any FW correction scheme is moderately reliable, but reasonably highly variable and separable measure, second only to the FWVF.

4 | Discussion

This paper re-examines the previous findings made by Albi et al. (2017) that the FW correction to DTI from single-shell diffusion-weighted MR data, as proposed by Pasternak et al. (2009), enhances the repeatability of DTI-based measures, such as FA or MD. Our study extrapolates that study to other FWVF estimation techniques and FW-correction DTI schemes particularly tailored to multiple-shell data. Although the results we report focus on the MICRA database, they are corroborated in the Supporting Information S1 by two additional databases with distinct characteristics, which include intra-scanner longitudinal and inter-scanner variabilities. Our results suggest that the improved repeatability of the FW-corrected DTI compared to a standard DTI in a single-shell scenario observed by Albi et al. (2017) may be data- and methodology-dependent, and does not generalize to multiple-shell FW correction schemes. On the contrary, we have shown that the FW-corrected DTI in a multiple-shell scenario, using a region contraction-based technique by Hoy et al. (2014), leads to systematic declines in repeatability/reproducibility and reliability compared to the standard DTI, as the number of degrees of freedom in the optimization procedure is larger. Our study shows evidence that the most reliable and repeatable (and reproducible) measures are FA, AD and RD estimated from the standard DTI and among

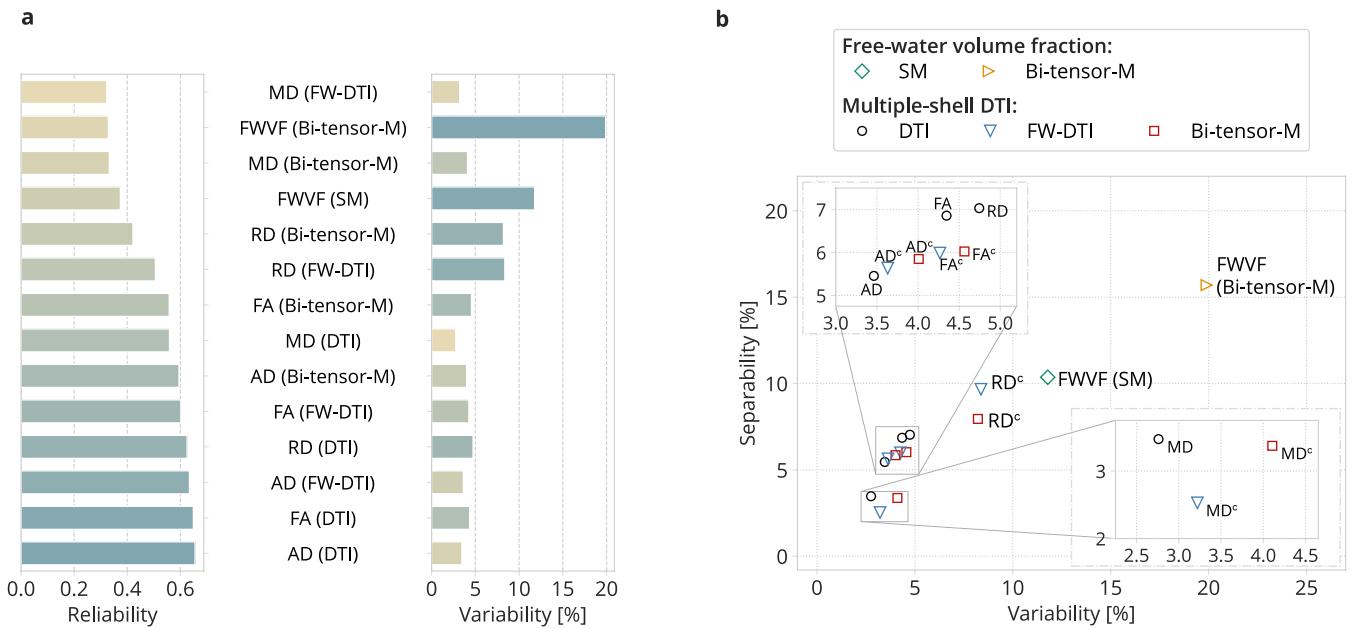


FIGURE 6 | (a) Box-plots presenting inter-session median reliability and median variability of multiple-shell indices computed from the MICRA database over the white matter area. The colours in the variability box-plot reflect the ordering according to the variability index value (i.e., the lower the value, the more beige the colour). (b) A diagram representing the separability index as a function of the variability index. A single marker refers to the median variability and median separability values calculated over the white matter area in the standard space. The FW-corrected DTI measures are annotated with a subscript ‘c’.

the FW-corrected DTI-based measures the FA and AD estimated from a previously corrected diffusion-weighted MR signal under a multiple-shell variant. In contrast, the least reliable and separable measure is the MD obtained from any FW correction approach, as well as the FWVF parameter itself, no matter whether estimated separately via the SM technique or jointly with the DTI.

In general, one can follow several approaches to correct the DTI for the FW, either using single- or multiple-shell diffusion-weighted MR data. The first and most popularised approach by Pasternak et al. (2009), which we refer to here as the Bi-tensor-S, directly optimises the bi-tensor representation given by Equation (4) within a variational framework. This formulation enables estimation of both the FWVF and FW-corrected DTI measures in a joint optimisation procedure using only single-shell diffusion-weighted MR data acquired approximately at $b = 1000\text{s/mm}^2$. However, the method requires the initialization scheme to be carefully selected (Parker et al. 2020; Golub et al. 2021) and yet it can fail to retrieve biological information from the MD-corrected parameter (see Figure 1b). The second group of methods optimise Equation (4) using the numerical schemes tailored for multiple-shell acquisitions (Pasternak et al. 2012; Hoy et al. 2014; Bergmann et al. 2020). As a representative, in this study, we follow the region contraction-based method by Hoy et al. (2014), which we call the Bi-tensor-M. Our experiments demonstrate that, while the Bi-tensor-M technique preserves biological information in the MD parameter, both the FW-corrected MD and FA parameters remain biased (see Figure 1b). Recently, the advantage of FWVF estimated using multiple-shell over single-shell has been demonstrated in the context of brain age estimation (Nemmi et al. 2022) and healthy brain ageing (Correia

et al. 2024). An alternative solution is to estimate the FWVF parameter, correct the diffusion-weighted MR signal for the FW component and then re-estimate the standard DTI from the corrected signal (Pieciak et al. 2023; Chang et al. 2025; Guadilla et al. 2025). Here, the biological variability of the FW-corrected MD parameter appears limited, while the MD and FA parameters are relatively free of the bias seen in the Bi-tensor-M approach (cf. Figure 1c to Figure 1b). The last group consists of deep learning-based approaches that aim to find a non-linear mapping between diffusion-weighted MR signal and FW or FW-corrected DTI parameters (Molina-Romero et al. 2018; Weninger et al. 2020).

We start our discussion by commenting on the sanity checks, displaying the estimated measures using different techniques. Figure 2d depicts a particularly flattened FW-corrected MD characteristic computed with the Bi-tensor-S approach. We note the flattening effect has been previously explored by Golub et al. (2021) in the context of in silico experiments and by Correia et al. (2024) in brain ageing. To put it differently, the FW-corrected MD using the Bi-tensor-S turns out to be the tissue’s prior. Contrary to Bi-tensor-S, the FW-corrected MD computed using the Bi-tensor-M and FW-DTI approaches has enabled us to discriminate between WM and GM areas. As for other FW-corrected measures, increased FA and decreased MD/AD/RD parameters over the white matter are consistent across the datasets considered in our study and with previous reports (Metzler-Baddeley et al. 2012; Hoy et al. 2014; Golub et al. 2021; Pieciak et al. 2023).

It is noteworthy that the FWVF estimated using SM, Bi-tensor-S, and Bi-tensor-M actually presents the effective FW fraction confounded by T_2 relaxation, but it is typically

used as a proxy for the FWVF (Pasternak et al. 2009; Golub et al. 2021). Although none of the above-mentioned methods directly model freely diffusing water, they somewhat aggregate the diffusion found as the cerebrospinal fluid and interstitial fluid in the extracellular space of grey and white matter (Pasternak et al. 2009). Intrinsically, the FWVF may be biased by other pools, such as blood perfusion, which affects the signal at low b -values (Rydhög et al. 2017).

In the study by Albi et al. (2017), it has been suggested that the FW correction to DTI from a single-shell acquisition improves the longitudinal test-retest repeatability of FA and MD metrics by reducing the CoV *on average* approximately at 1%pt. Our study partially corroborates these results, illustrating an improvement in FA/MD repeatability for MICRA database in terms of median CoV over the WM at 0.3%pt./2.5%pt. However, this pattern might not be general, given the results from other databases presented in the Supporting Information S1. The recent study by Correia et al. (2024) discovered the flattening effect of FW-corrected MD parameter with age. Our study has revealed an excellent reproducibility of FW-corrected MD, which directly explains the flattened spatial characteristics of the measure. In other words, the Bi-tensor-S provides the prior for the MD measure (here, assumed to be $0.6 \times 10^{-3} \text{ mm}^2/\text{s}$) rather than the value contemplating the actual FW-corrected MD parameter. Moreover, the FW-corrected MD parameter with the Bi-tensor-S approach is neither reliable nor separable (see the extended results demonstrated in Figure S14). This is a consequential result that may raise questions about the repeatability (and thus, the trustworthiness) of previous findings in the brain studies based on a single-shell Bi-tensor-S-based MD parameter. In general, the FW-corrected MD parameter estimated using any single- or multiple-shell-based method considered in our experiments exhibits low reliability and separability, which clarifies the factiously excellent reproducibility observed in Figure 3. Interestingly, the reliability of FW-corrected MD measure, regardless of the method used, is consistently lower than the standard DTI-based MD (see Figure 5b). The results obtained with the single-shell Bi-tensor-S method do not translate to the multiple-shell scenario, as the FW-corrected MD, AD and RD parameters computed with Bi-tensor-M approach reveal increased variability compared to the standard DTI (see Figure 4b and Figure 5b). A direct reason for the variability growth observed in the Bi-tensor-M approach is the increased number of degrees of freedom in the optimised cost function. The variability of a FW-corrected DTI might decrease if one pre-estimates the FWVF using an external method, such as the SM (Tristán-Vega et al. 2022), corrects the diffusion-weighted MR signal for the FW component and then re-estimates the DTI using a standard procedure. Remarkably, ‘fixing’ the FWVF in the optimization process does not reduce the reliability of the FW-corrected measures compared to the Bi-tensor-M, as indicated by the experiment depicted in Figure 5b.

The experiments demonstrated consistency in the variability of FWVF against the DTI and FW-corrected DTI measures across all three datasets. Specifically, the FWVF estimated using any method considered in the study demonstrates a higher CoV over the white matter area compared to all DTI-based

measures examined in this study. The experimental results do not provide a clear answer as to which multiple-shell-based technique (i.e., the region contraction-based or the spherical means) is superior in terms of reproducibility. However, the Bi-tensor-M technique is trailing behind the SM considering the reliability index.

Finally, we note the study goes beyond the standard intra-site repeatability or inter-site reproducibility, as it also explores longitudinal reproducibility. Such longitudinal reproducibility evaluation is particularly important in a clinical scenario once the features observed in the images are expected to demonstrate the evolution of the brain between the scans, or confounding factors such as different operators handling the scanner or magnetic field drifts (Lehmann et al. 2021; Boudreau et al. 2025).

5 | Conclusions

This paper studies the variability, reliability and separability properties of FW-corrected DTI in the healthy human brain. We explore different methodologies used to correct the DTI for the FW compartment, depending on whether the diffusion-weighted MR acquisitions are single- or multiple-shell-based, and evaluate them using three publicly available databases acquired in inter-session, intra-scanner longitudinal and inter-scanner scenarios. Our study has shown that one should not only look for the maximal repeatability (or reproducibility) of FW-corrected DTI measures in the brain studies, but also assess the reliability and separability indices, as it has been particularly observed with the FW-corrected MD parameter using the single-shell variational method. Importantly, how to correct the DTI for the FW is of great importance—the behaviour of the single-shell method appears to be data-dependent, with questionable enhancement in variability and reliability, as well as the FW-corrected MD parameter being anatomically non-meaningful. The multiple-shell FW-correction contraction-based technique has shown a reduced reproducibility and reliability of the measures compared to the standard DTI, the results being consistent across the evaluated data. As a conclusive remark, the FW correction to DTI should not be considered in terms of conceivable ‘improvement’ in the repeatability (or reproducibility) and reliability, but rather as a methodology that provides information about the probed tissue, which the standard DTI partially hides. However, the choice of the FW-correction scheme should be made to avoid impacting the reproducibility, reliability, and accuracy of DTI-based parameters. As a solution, we suggest employing a customised FW-correction scheme, i.e., estimating the FWVF externally, correcting the diffusion-weighted MR signal for the FW, and then re-estimating the DTI using a standard procedure. Our experiments have provided evidence that this strategy improves the repeatability and reproducibility while not (significantly) affecting the reliability of the FW-corrected measures compared to the multiple-shell FW-correction of bi-tensor representation, while maintaining a relatively low error in the estimated MD and FA parameters. However, we note that this technique may result in a reduction of the biological information conveyed by the MD-corrected measure, although it still exhibits a lower average bias compared to the Bi-tensor-M approach.

Acknowledgements

This work was funded by Junta de Castilla y León and Fondo Social Europeo Plus (FSE+) under research grant VA156P24; Agencia Estatal de Investigación (Ministerio de Ciencia e Innovación of Spain) with research grants PID2021-124407NB-I00 and PID2024-158963NB-I00. Tomasz Pieciak acknowledges the Polish National Agency for Academic Exchange for grant PPN/BEK/2019/1/00421 under the Bekker programme. Guillem París was funded by the Consejería de Educación de Castilla y León and the European Social Fund through the 'Ayudas para financiar la contratación predoctoral de personal investigador - Orden EDU/1100/201712/12' program. Open Access funding enabled and organized by CRUE/BUCLE 2025 Gold.

The authors acknowledge Dr. Nico Lehmann for sharing the Magdeburg database.

The authors did not use generative AI in the writing of this manuscript.

Funding

This work was supported by Junta de Castilla y León and Fondo Social Europeo Plus (FSE+) (VA156P24), Ministerio de Ciencia e Innovación (PID2021-124407NB-I00, PID2024-158963NB-I00), Narodowa Agencja Wymiany Akademickiej (PPN/BEK/2019/1/00421) and Consejería de Educación de Castilla y León and the European Social Fund (Orden EDU/1100/201712/12).

Ethics Statement

Ethics approval was waived for this study due to the use of external data.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data comes from the papers by Koller et al. (2021), Lehmann et al. (2021) and Tong et al. (2020).

Endnotes

¹ Percent point.

² DiffKDE is a computational method used to estimate the probability density function and should not be confused with diffusion-weighted MR technique.

References

Ades-Aron, B., S. Coelho, G. Lemberskiy, et al. 2025. "Denoising Improves Cross-Scanner and Cross-Protocol Test–Retest Reproducibility of Diffusion Tensor and Kurtosis Imaging." *Human Brain Mapping* 46, no. 4: e70142.

Aja-Fernández, S., and G. Vegas-Sánchez-Ferrero. 2016. *Statistical Analysis of Noise in MRI*. Springer International Publishing.

Albi, A., O. Pasternak, L. Minati, et al. 2017. "Free Water Elimination Improves Test–Retest Reproducibility of Diffusion Tensor Imaging Indices in the Brain: A Longitudinal Multisite Study of Healthy Elderly Subjects." *Human Brain Mapping* 38, no. 1: 12–26.

Andersson, J. L., M. S. Graham, E. Zsoldos, and S. N. Sotropoulos. 2016. "Incorporating Outlier Detection and Replacement Into a Non-Parametric Framework for Movement and Distortion Correction of Diffusion MR Images." *NeuroImage* 141: 556–572.

Andersson, J. L., S. Skare, and J. Ashburner. 2003. "How to Correct Susceptibility Distortions in Spin-Echo Echo-Planar Images: Application to Diffusion Tensor Imaging." *NeuroImage* 20, no. 2: 870–888.

Basser, P. J., J. Mattiello, and D. LeBihan. 1994. "Estimation of the Effective Self-Diffusion Tensor From the NMR Spin Echo." *Journal of Magnetic Resonance. Series B* 103, no. 3: 247–254.

Bergamino, M., R. R. Walsh, and A. M. Stokes. 2021. "Free-Water Diffusion Tensor Imaging Improves the Accuracy and Sensitivity of White Matter Analysis in Alzheimer's Disease." *Scientific Reports* 11, no. 1: 6990.

Bergmann, Ø., R. Henriques, C. F. Westin, and O. Pasternak. 2020. "Fast and Accurate Initialization of the Free-Water Imaging Model Parameters From Multi-Shell Diffusion MRI." *NMR in Biomedicine* 33, no. 3: e4219.

Botev, Z. I., J. F. Grotowski, and D. P. Kroese. 2010. "Kernel Density Estimation via Diffusion (2010)." *Annals of Statistics* 38, no. 5: 2916–2957.

Boudreau, M., A. Karakuzu, A. Boré, et al. 2025. "Longitudinal Reproducibility of Brain and Spinal Cord Quantitative MRI Biomarkers." *Imaging Neuroscience* 3: imag_a_00409.

Carreira Figueiredo, I., F. Borga, O. Pasternak, F. E. Turkheimer, and O. D. Howes. 2022. "White-Matter Free-Water Diffusion MRI in Schizophrenia: A Systematic Review and Meta-Analysis." *Neuropsychopharmacology* 47, no. 7: 1413–1420.

Chad, J. A., O. Pasternak, D. H. Salat, and J. J. Chen. 2018. "Re-Examining Age-Related Differences in White Matter Microstructure With Free-Water Corrected Diffusion Tensor Imaging." *Neurobiology of Aging* 71: 161–170.

Chad, J. A., N. Sochen, J. J. Chen, and O. Pasternak. 2023. "Implications of Fitting a Two-Compartment Model in Single-Shell Diffusion MRI." *Physics in Medicine and Biology* 68, no. 21: 215012.

Chang, K., L. Burke, N. LaPiana, et al. 2025. "Free Water Elimination Tractometry for Aging Brains." *Imaging Neuroscience* 3: IMAG.a.991. <https://doi.org/10.1162/IMAG.a.991>.

Correia, M. M., R. N. Henriques, M. Golub, S. Winzeck, and R. G. Nunes. 2024. "The Trouble With Free-Water Elimination Using Single-Shell Diffusion MRI Data: A Case Study in Ageing." *Imaging Neuroscience* 2: 1–17.

Duan, F., T. Zhao, Y. He, and N. Shu. 2015. "Test–Retest Reliability of Diffusion Measures in Cerebral White Matter: A Multiband Diffusion MRI Study." *Journal of Magnetic Resonance Imaging* 42, no. 4: 1106–1116.

Golub, M., R. N. Henriques, and R. G. Nunes. 2021. "Free-Water DTI Estimates From Single b-Value Data Might Seem Plausible but Must Be Interpreted With Care." *Magnetic Resonance in Medicine* 85, no. 5: 2537–2551.

Grech-Sollars, M., P. W. Hales, K. Miyazaki, et al. 2015. "Multi-Centre Reproducibility of Diffusion MRI Parameters for Clinical Sequences in the Brain." *NMR in Biomedicine* 28, no. 4: 468–485.

Guadilla, I., A. R. Fouto, A. Ruiz-Tagle, I. Esteves, G. Caetano, and N. A. Silva. 2025. "White Matter Alterations in Episodic Migraine Without Aura Patients Assessed With Diffusion MRI: Effect of Free Water Correction." *Journal of Headache and Pain* 26, no. 1: 1–13.

Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Henriques, N. R., A. Rokem, E. Garyfallidis, S. St-Jean, E. T. Peterson, and M. M. Correia. 2017. "Optimization of a Free Water Elimination Two-Compartment Model for Diffusion Tensor Imaging." *ReScience* 3, no. 1: 2.

Hoy, A. R., C. G. Koay, S. R. Kecskemeti, and A. L. Alexander. 2014. "Optimization of a Free Water Elimination Two-Compartment Model for Diffusion Tensor Imaging." *NeuroImage* 103: 323–333.

Jakab, A., R. Tuura, C. Kellenberger, and I. Scheer. 2017. "In Utero Diffusion Tensor Imaging of the Fetal Brain: A Reproducibility Study." *NeuroImage: Clinical* 15: 601–612.

Kellner, E., B. Dhital, V. G. Kiselev, and M. Reisert. 2016. "Gibbs-Ringing Artifact Removal Based on Local Subvoxel-Shifts." *Magnetic Resonance in Medicine* 76, no. 5: 1574–1581.

Koay, C. G., L. C. Chang, J. D. Carew, C. Pierpaoli, and P. J. Basser. 2006. "A Unifying Theoretical and Algorithmic Framework for Least Squares Methods of Estimation in Diffusion Tensor Imaging." *Journal of Magnetic Resonance* 182, no. 1: 115–125.

Koller, K., U. Rudrapatna, M. Chamberland, et al. 2021. "MICRA: Microstructural Image Compilation With Repeated Acquisitions." *NeuroImage* 225: 117406.

Laguna, P. A. L., A. J. Combes, J. Streffer, et al. 2020. "Reproducibility, Reliability and Variability of FA and MD in the Older Healthy Population: A Test-Retest Multiparametric Analysis." *NeuroImage: Clinical* 26: 102168.

Le Bihan, D., and H. Johansen-Berg. 2012. "Diffusion MRI at 25: Exploring Brain Tissue Structure and Function." *NeuroImage* 61, no. 2: 324–341.

Lehmann, N., N. Aye, J. Kaufmann, et al. 2021. "Longitudinal Reproducibility of Neurite Orientation Dispersion and Density Imaging (NODDI) Derived Metrics in the White Matter." *Neuroscience* 457: 165–185.

Liu, Q., L. Ning, I. A. Shaik, et al. 2024. "Reduced Cross-Scanner Variability Using Vendor-Agnostic Sequences for Single-Shell Diffusion MRI." *Magnetic Resonance in Medicine* 92, no. 1: 246–256.

Lyall, A. E., O. Pasternak, D. G. Robinson, et al. 2018. "Greater Extracellular Free-Water in First-Episode Psychosis Predicts Better Neurocognitive Functioning." *Molecular Psychiatry* 23, no. 3: 701–707.

Maillard, P., E. Fletcher, B. Singh, et al. 2019. "Cerebral White Matter Free Water: A Sensitive Biomarker of Cognition and Function." *Neurology* 92, no. 19: e2221–e2231.

Melzer, T. R., R. J. Keenan, G. J. Leeper, et al. 2020. "Test-Retest Reliability and Sample Size Estimates After MRI Scanner Relocation." *NeuroImage* 211: 116608.

Merisaari, H., J. J. Tuulari, L. Karlsson, et al. 2019. "Test-Retest Reliability of Diffusion Tensor Imaging Metrics in Neonates." *NeuroImage* 197: 598–607.

Metzler-Baddeley, C., M. J. O'Sullivan, S. Bells, O. Pasternak, and D. K. Jones. 2012. "How and How Not to Correct for CSF-Contamination in Diffusion MRI." *NeuroImage* 59, no. 2: 1394–1403.

Molina-Romero, M., B. Wiestler, P. A. Gómez, M. I. Menzel, and B. H. Menze. 2018. "Deep Learning With Synthetic Diffusion MRI Data for Free-Water Elimination in Glioblastoma Cases." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 98–106. Springer International Publishing.

Mori, S., S. Wakana, P. C. Van Zijl, and L. M. Nagae-Poetscher. 2005. *MRI Atlas of Human White Matter*. Elsevier.

Nakaya, M., N. Sato, H. Matsuda, et al. 2022. "Free Water Derived by Multi-Shell Diffusion MRI Reflects Tau/Neuroinflammatory Pathology in Alzheimer's Disease." *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 8, no. 1: e12356.

Nemmi, F., M. Levardon, and P. Péran. 2022. "Brain-Age Estimation Accuracy Is Significantly Increased Using Multishell Free-Water Reconstruction." *Human Brain Mapping* 43, no. 7: 2365–2376.

Ofori, E., O. Pasternak, P. J. Planetta, et al. 2015. "Longitudinal Changes in Free-Water Within the Substantia Nigra of Parkinson's Disease." *Brain* 138, no. 8: 2322–2331.

Parker, D., A. A. Ould Ismail, R. Wolf, et al. 2020. "Freewater esti-matoR Using iNtErpolated iniTialization (FERNET): Characterizing Peritumoral Edema Using Clinically Feasible Diffusion MRI Data." *PLoS One* 15, no. 5: e0233645.

Pasternak, O., M. E. Shenton, and C. F. Westin. 2012. "Estimation of Extracellular Volume From Regularized Multi-Shell Diffusion MRI." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 305–312. Springer Berlin Heidelberg.

Pasternak, O., N. Sochen, Y. Gur, N. Intrator, and Y. Assaf. 2009. "Free Water Elimination and Mapping From Diffusion MRI." *Magnetic Resonance in Medicine* 62, no. 3: 717–730.

Pieciak, T., S. Aja-Fernandez, and G. Vegas-Sanchez-Ferrero. 2017. "Non-Stationary Rician Noise Estimation in Parallel MRI Using a Single Image: A Variance-Stabilizing Approach." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, no. 10: 2015–2029.

Pieciak, T., G. París, D. Beck, et al. 2023. "Spherical Means-Based Free-Water Volume Fraction From Diffusion MRI Increases Non-Linearly With Age in the White Matter of the Healthy Human Brain." *NeuroImage* 279: 120324.

Pieciak, T., I. Rabanillo-Viloria, and S. Aja-Fernández. 2018. "Bias Correction for Non-Stationary Noise Filtering in MRI." In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 307–310. IEEE.

Pierpaoli, C., and D. K. Jones. 2004. "Removing CSF Contamination in Brain DT-MRIs by Using a Two-Compartment Tensor Model." In *International Society for Magnetic Resonance in Medicine Meeting*, 1215. International Society for Magnetic Resonance in Medicine Meeting (ISMRM).

Rydhög, A. S., F. Szczepankiewicz, R. Wiestam, et al. 2017. "Separating Blood and Water: Perfusion and Free Water Elimination From Diffusion MRI in the Human Brain." *NeuroImage* 156: 423–434.

Shahim, P., L. Holleran, J. H. Kim, and D. L. Brody. 2017. "Test-Retest Reliability of High Spatial Resolution Diffusion Tensor and Diffusion Kurtosis Imaging." *Scientific Reports* 7, no. 1: 11141.

Sheather, S. J., and M. C. Jones. 1991. "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 53, no. 3: 683–690.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC.

Smith, S. M., M. Jenkinson, M. W. Woolrich, et al. 2004. "Advances in Functional and Structural MR Image Analysis and Implementation as FSL." *NeuroImage* 23: S208–S219.

Tong, Q., H. He, T. Gong, et al. 2020. "Multicenter Dataset of Multi-Shell Diffusion MRI in Healthy Traveling Adults With Identical Settings." *Scientific Data* 7, no. 1: 157.

Tristán-Vega, A., G. París, R. de Luis-García, and S. Aja-Fernández. 2022. "Accurate Free-Water Estimation in White Matter From Fast Diffusion MRI Acquisitions Using the Spherical Means Technique." *Magnetic Resonance in Medicine* 87, no. 2: 1028–1035.

Tustison, N. J., B. B. Avants, P. A. Cook, et al. 2010. "N4ITK: Improved N3 Bias Correction." *IEEE Transactions on Medical Imaging* 29, no. 6: 1310–1320.

Van Essen, D. C., S. M. Smith, D. M. Barch, et al. 2013. "The WU-Minn Human Connectome Project: An Overview." *NeuroImage* 80: 62–79.

Venkatraman, V. K., C. E. Gonzalez, B. Landman, et al. 2015. "Region of Interest Correction Factors Improve Reliability of Diffusion Imaging Measures Within and Across Scanners and Field Strengths." *NeuroImage* 119: 406–416.

Veraart, J., E. Fieremans, and D. S. Novikov. 2016. "Diffusion MRI Noise Mapping Using Random Matrix Theory." *Magnetic Resonance in Medicine* 76, no. 5: 1582–1593.

Veraart, J., D. S. Novikov, D. Christiaens, B. Ades-Aron, J. Sijbers, and E. Fieremans. 2016. "Denoising of Diffusion MRI Using Random Matrix Theory." *NeuroImage* 142: 394–406.

Weninger, L., S. Koppers, C. H. Na, K. Juetten, and D. Merhof. 2020. "Free-Water Correction in Diffusion MRI: A Reliable and Robust Learning Approach." In *Computational Diffusion MRI: MICCAI Workshop, Shenzhen, China, October 2019*, 91–99. Springer International Publishing.

Westin, C. F., S. E. Maier, H. Mamata, A. Nabavi, F. A. Jolesz, and R. Kikinis. 2002. "Processing and Visualization for Diffusion Tensor MRI." *Medical Image Analysis* 6, no. 2: 93–108.

Zhong, J., X. Liu, Y. Hu, et al. 2024. "Robustness of Quantitative Diffusion Metrics From Four Models: A Prospective Study on the Influence of Scan-Rescans, Voxel Size, Coils, and Observers." *Journal of Magnetic Resonance Imaging* 60, no. 4: 1470–1483.

Zuo, X. N., T. Xu, and M. P. Milham. 2019. "Harnessing Reliability for Neuroscience Research." *Nature Human Behaviour* 3, no. 8: 768–771.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information.