RESEARCH ARTICLE

NMR IN BIOMEDICINE **WILEY**

Mean kurtosis-Curve (MK-Curve) correction improves the test–retest reproducibility of diffusion kurtosis imaging at 3 T

Ernst Christiaanse[1,2] | Patrik O. Wyss[1] | Anke Scheel-Sailer[3,4] | Angela Frotzler[5] |
Dirk Lehnick[6] | Rajeev K. Verma[1] | Markus F. Berger[1] |
Alexander Leemans[2] | Alberto De Luca[2,7]

[1]Department of Radiology, Swiss Paraplegic Centre, Nottwil, Switzerland

[2]Image Sciences Institute, Division Imaging & Oncology, University Medical Center Utrecht, Utrecht, the Netherlands

[3]Rehabilitation and Quality Management, Swiss Paraplegic Centre, Nottwil, Switzerland

[4]Department of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland

[5]Clinical Trial Unit, Swiss Paraplegic Centre, Nottwil, Switzerland

[6]Department of Health Sciences and Medicine, Biostatistics and Methodology, University Lucerne, Lucerne, Switzerland

[7]Neurology Department, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands

Correspondence
Ernst Christiaanse, Department of Radiology, Swiss Paraplegic Centre, Guido-A.Zaech-Strasse 1, CH-6207 Nottwil, Switzerland.
Email: ernst.christiaanse@paraplegie.ch

Funding information
This study was funded by the Swiss Paraplegic Foundation (research program Radiology).

Abstract

Diffusion kurtosis imaging (DKI) is applied to gain insights into the microstructural organization of brain tissues. However, the reproducibility of DKI outside brain white matter, particularly in combination with advanced estimation to remedy its noise sensitivity, remains poorly characterized. Therefore, in this study, we investigated the variability and reliability of DKI metrics while correcting implausible values with a fit method called mean kurtosis (MK)-Curve. A total of 10 volunteers (four women; age: 41.4 ± 9.6 years) were included and underwent two MRI examinations of the brain. The images were acquired on a clinical 3-T scanner and included a T1-weighted image and a diffusion sequence with multiple diffusion weightings suitable for DKI. Region of interest analysis of common kurtosis and tensor metrics derived with the MK-Curve DKI fit was performed, including intraclass correlation (ICC) and Bland–Altman (BA) plot statistics. A p value of less than 0.05 was considered statistically significant. The analyses showed good to excellent agreement of both kurtosis tensor- and diffusion tensor-derived MK-Curve–corrected metrics (ICC values: 0.77–0.98 and 0.87–0.98, respectively), with the exception of two DKI-derived metrics (axial kurtosis in the cortex: ICC = 0.68, and radial kurtosis in deep gray matter: ICC = 0.544). Non-MK-Curve–corrected kurtosis tensor-derived metrics ranged from 0.01 to 0.52 and diffusion tensor-derived metrics from 0.06 to 0.66, indicating poor to moderate reliability. No structural bias was observed in the BA plots for any of the diffusion metrics. In conclusion, MK-Curve–corrected DKI metrics of the human brain can be reliably acquired in white and gray matter at 3 T and DKI metrics have good to excellent agreement in a test–retest setting.

1 | INTRODUCTION

Diffusion-weighted imaging (DWI) is a noninvasive technique based on the Brownian movement of water.[1] Biological tissues are complex structures composed of various cells and membranes, and the diffusion of water molecules may deviate from Gaussian behavior,[2] which is a core assumption of the diffusion tensor model. This non-Gaussian behavior becomes more apparent when collecting DWI with higher b-values or longer diffusion times. For this reason, diffusion tensor imaging (DTI) cannot take advantage of the information content of data acquired in the brain with b-values above a value of about 1200 s/mm^2.[3] The deviation from the normal distribution of water molecules in the experiment time—termed excess kurtosis—is related to membrane restriction effects[4] or changes in diffusional heterogeneity,[5,6] because of, for example, intravoxel incoherent motion. Diffusion kurtosis imaging (DKI) quantifies the excess kurtosis as an extension of DTI taking into account the non-Gaussian distribution of water molecules in tissues.[3,5] To fit the DKI model, a minimal DWI experiment setup should include 21 or more diffusion-weighted directions distributed over at least two nonzero diffusion weightings (in addition to the nondiffusion-weighted images).

Multiple studies have assessed the test–retest reliability of DTI,[7–9] showing good reproducibility and within-subject coefficients of variation (wsCVs) of 2% to 3% for fractional anisotropy (FA) and mean diffusivity (MD). However, only a few studies have investigated the test–retest reliability of DKI in the brain in healthy individuals[10–12]: of these, one study was conducted in three patients with traumatic brain injury and four healthy controls (HCs),[13] and one study in small vessel disease,[14] reporting good to excellent test–retest reproducibility. Of all these studies, only two included gray matter (GM) regions in their test–retest assessment.[10,11]

In addition, compared with DTI, DKI is inherently more noise sensitive because of a higher degree of freedom (i.e., the larger number of fitted parameters of 22 vs. seven). This also propagates to scalar maps commonly quantified from DKI, such as the mean kurtosis (MK), which requires the square of the mean diffusivity to be determined and is thus prone to noise amplification. Therefore, MK maps often exhibit implausible values, introducing undesired intrasubject and intersubject variability. Various strategies have been proposed to stabilize the DKI fit, with the most common being spatial smoothing, that is, the replacement of implausible values with the average value of an adjacent voxel.

Recently, a DKI correction method named MK-Curve has been proposed to recover voxel-wise implausible DKI estimates without averaging or smoothing.[15] In this work, we evaluate whether the MK-Curve method can be applied to a clinically feasible DKI protocol to improve repeatability and reproducibility compared with naive DKI estimates. We estimate the repeatability and reproducibility of a DKI fit with and without MK-Curve correction over two scans in both white matter (WM) and GM regions by evaluating various interscan reproducibility and test–retest reliability metrics in the healthy brain.

2 | MATERIALS AND METHODS

2.1 | DKI fit with the MK-Curve method

The general equation of the DKI signal representation can be written as

$$S = S_0 e^{-b\boldsymbol{D} + \frac{1}{6}b^2\overline{D}^2 K,}$$

where S is the signal measured at b-value (diffusion weighting) along a specific direction, S_0 is the nondiffusion-weighted signal, D is the diffusion tensor, $\overline{D}$ the mean diffusivity, and K the fourth-order kurtosis tensor. Fitting the DKI representation implies determining 22 unique parameters (S_0, six parameters of the symmetric tensor D, and 15 parameters of the symmetric tensor K). Because of the relatively large number of free parameters, and because of the presence of a quadratic term linking the kurtosis and diffusion tensor results, DKI estimates are particularly sensitive to noise amplification. Recently, Zhang and colleagues[15] introduced the MK-Curve method approach to remove spurious kurtosis estimates. This framework is based on the idea that physiologically plausible estimates of MK should fall within the range $0 \leq MK \leq 3$, and that erroneous MK values mostly originate from anomalous values of S_0 because of, for example, Gibbs ringing artefacts, among others.[6] Accordingly, in the MK-Curve method, the value of S_0 is iteratively adapted until $0 \leq MK \leq 3$, without altering or discarding any diffusion-weighting measurement. The corrected S_0 lies in the range between 0.5 and two times the original measured S_0 and is in line with the original work of Zhang et al. These corrected S_0 values should not be regarded as ground truth or true measurements but rather as a mathematical construct to facilitate plausible MK values.

2.2 | Study population

All participants were informed about the aim of the study and provided written informed consent. The local ethics committee approved this study (PB 2019–00333). The study was conducted according to the declaration of Helsinki and guidelines of good clinical practice.

We included 10 healthy volunteers (four women; mean age: 41.4 ± 9.6 years, range: 25.1–57.9 years) with no history of cervical trauma, traumatic brain injury, cervical surgery, signs of neurological impairment, or neurological disease.

2.3 | MRI acquisition

We acquired the data on a dedicated clinical 3-T Philips Achieva scanner (Release 5.4.1, Philips Healthcare, Best, The Netherlands) with a 32-channel head coil. The acquisition protocol consisted of (i) a high-resolution sagittal 3D T1-weighted anatomical sequence (TR = 8 ms, TE = 3.7 ms, voxel size = 1 x 1 x 1 mm^3, acquisition time = 150 s) and (ii) a diffusion MRI sequence (TR = 13,000 ms, TE = 80 ms, FOV = 240 x 240 x 152 mm^3, voxel size = 2.5 x 2.5 x 2.5 mm^3, matrix size = 96 x 96, 61 slices, SENSE acceleration factor = 2, acquisition time = 900 s), resulting in a total acquisition time of 20 min. The diffusion MRI sequence is based on a standard Human Connectome Project b-values scheme with 63 diffusion-weighted volumes (seven at b = 0 s/mm^2, six at b = 500 s/mm^2, 20 at b = 1000s/mm^2, and 30 at b = 2000s/mm^2).

2.4 | Diffusion MRI processing

We processed and analyzed the acquired diffusion data with MATLAB R2020a (Mathworks Inc., version 9.8) using ExploreDTI[16] (version 4.8.6) and MRIToolkit (www.mritoolkit.org).[17] The data processing pipeline included signal drift correction,[18] de-noising with Marchenko–Pastur Principal Component Analysis (MPPCA),[19] Gibbs ringing correction,[20] correction for subject motion,[21] eddy current distortions, and EPI deformations. We set the sliding window size used for the MPPCA denoising to 5 x 5 x 5, comprising a total of 125 voxels. We estimated the diffusion and kurtosis tensors with linear least square estimator and corrected for unfeasible values with the MK-Curve method.[15] We derived the kurtosis metrics axial kurtosis (AK), radial kurtosis (RK), mean kurtosis (MK), and kurtosis anisotropy (KA), as previously described by Poot et al.[22] using 1024 sampling directions. We calculated the DTI metrics from the tensor quantified during the DKI fit, and not with a separate DTI fit.

We applied the SPM-based Cat12 toolbox[23] to conduct brain segmentation using both a WM atlas[24] and a GM atlas,[25] resulting in 224 regions of interest (ROIs). We selected three subsets consisting of 206 ROIs in both the left (103) and right (103) hemispheres, a priori omitting 18 ROIs corresponding to ventricle areas. In total, we used 140 ROIs in cortical regions, 12 ROIs in deep GM, and 54 ROIs in deep WM. We calculated the signal-to-noise ratio (SNR) of every dataset by voxel-wise division of the mean of the b = 0 s/mm^2 images divided by their standard deviation (SD).

To facilitate the presentation and interpretation of the results, we combined the ROIs defined by the atlases in higher level groups representative of the (i) cortex, (ii) deep GM, (iii) deep WM (Table S1), and jointly as (iv) whole brain (i.e., (i) + (ii) + (iii) = sum of cortex + deep GM + deep WM).

2.5 | Statistical analysis

We performed all statistical analyses using R Statistics[26] (version 4.0.2) and considering the aforementioned four groups of ROIs (cortex, deep GM, deep WM, and whole-brain ROI). We plotted the figures using the packages ggplot2,[27] cowplot,[28] blandr,[29] and ggExtra.[30] The boxplots show data including medians and quartiles. We quantitatively analyzed the data calculating the mean value (MV), the SD, and the wsCV (100*sqrt[mean {var^2/mean^2}]) by combining both measurement time points. For test–retest analysis, we calculated the intraclass correlation coefficient (ICC),[31] the square of the coefficient of determination (R^2), the mean and the intrasubject SD of the differences between the first and the second measurement time points (MV_{Diff}, SD_{Diff}), the coefficient of repeatability (CR = 1.96*SD_{Diff}), and the repeatability index (RI = 100% CR/MV_{Diff}). Based on the rating according to Koo and Li,[32] the ICC values indicated: poor (ICC < 0.5), moderate (0.5 < ICC < 0.75), good (0.75 < ICC < 0.9), and excellent (ICC > 0.9). To characterize the Bland–Altman (BA) plots, we calculated the bias and the upper and lower limits of agreement.

To assess the influence of the time interval between the measurements, the data of the participants were split into two groups (< 3 vs. ≥ 3 days) and were compared showing density plot. A difference of the distribution between the two groups was assessed by a Kolmogorov–Smirnov test.
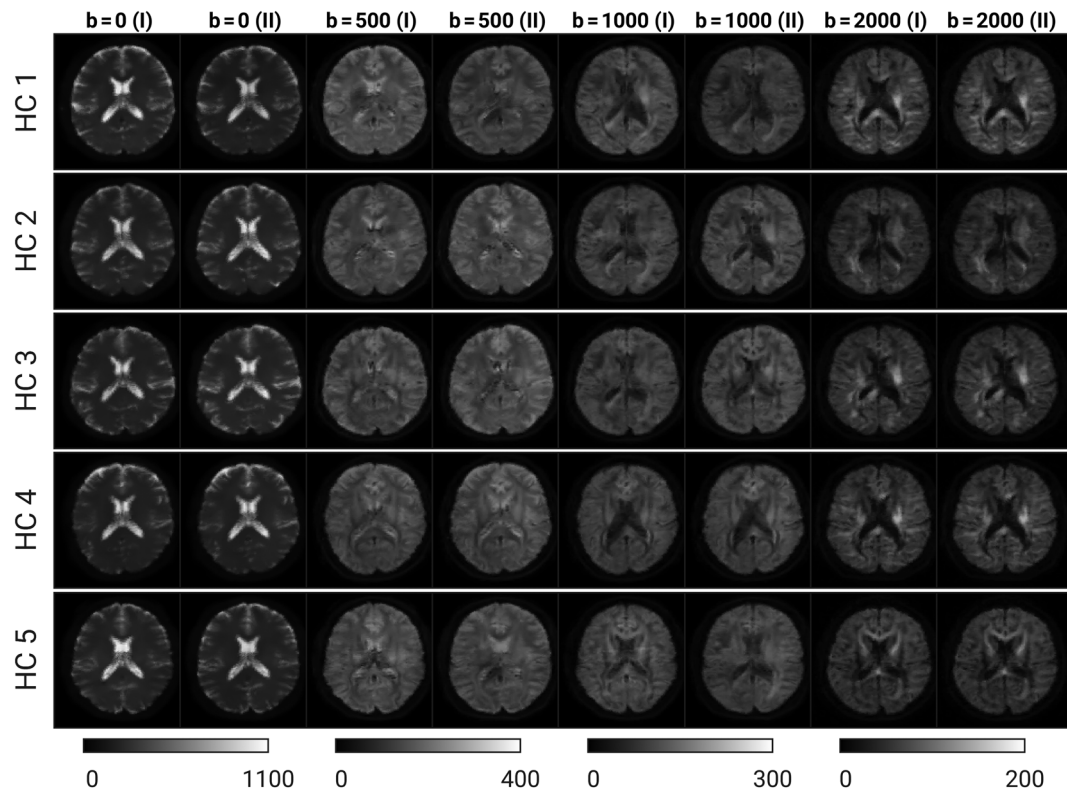
FIGURE 1 Single-slice averaged b-values of five participants at time point 1 (TP1) and time point 2 (TP2). Shown is a single axial slice of the averaged b = 0 (number of signal averages = 7), b = 500 (6), b = 1000 (20), and b = 2000 (30) values of five healthy controls (HCs) at TP 1 (I) and TP 2 (II)
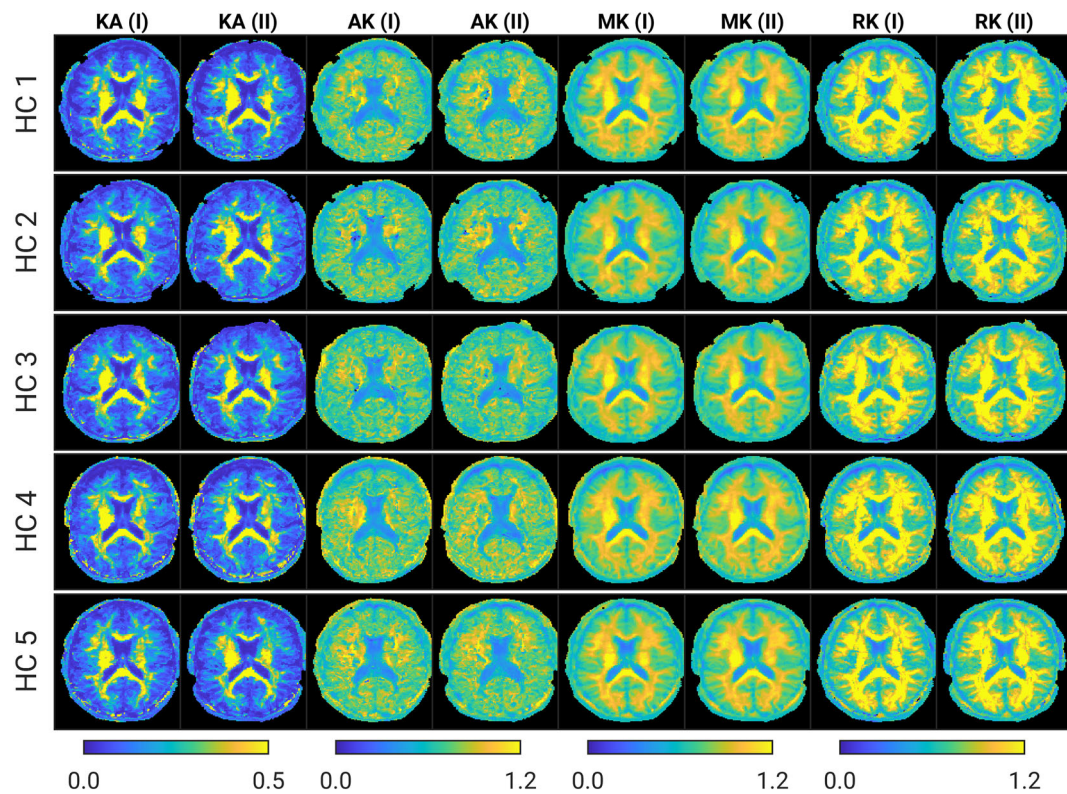
FIGURE 2 Single-slice averaged DKI metrics of five participants at time point 1 (TP1) and time point 2 (TP2). Shown is a single axial slice of the average kurtosis anisotropy (KA), axial kurtosis (AK), mean kurtosis (MK), and radial kurtosis (RK) values of five healthy controls (HCs) at TP1 (I) and TP2 (II). DKI, diffusion kurtosis imaging

TABLE 1 MV, SD, intersubject CV, and WSCV ($100 * \sqrt{\text{mean}[\text{var}^2/\text{mean}^2]}$) of the difference between time point 1 (TP1) and time point 2 (TP2) for (A) DKI and (B) DTI metrics with MK-Curve correction

(A) DKI

Region Metric	Whole brain				Cortex				Deep GM				Deep WM			
	MV_{diff}	SD_{diff}	CV (%)	WSCV (%)	MV_{diff}	SD_{diff}	CV (%)	WSCV (%)	MV_{diff}	SD_{diff}	CV (%)	WSCV (%)	MV_{diff}	SD_{diff}	CV (%)	WSCV (%)
KA	2.5e-02	4.4e-02	67	10.0	1.7e-02	4.2e-02	51.3	9.3	2.7e-02	2.6e-02	38.7	13.4	4.5e-02	4.6e-02	37	10.9
AK	2.9e-02	3.5e-02	12	4.3	2.5e-02	3.4e-02	7.5	3.6	4.1e-02	3.8e-02	10.1	4.6	3.6e-02	3.8e-02	19	5.6
MK	1.5e-02	2.4e-02	15	2.1	1.3e-02	2.5e-02	11.8	2.0	1.6e-02	1.6e-02	6.0	2.0	2.2e-02	2.2e-02	11	2.2
RK	4.3e-02	6.0e-02	29	4.8	2.7e-02	3.3e-02	21.8	3.6	4.1e-02	3.7e-02	7.7	5.3	8.4e-02	9.1e-02	17	7.0

(B) DTI

Region Metric	Whole brain				Cortex				Deep GM				Deep WM			
	MV_{diff}	SD_{diff}	CV (%)	WSCV (%)	MV_{diff}	SD_{diff}	CV (%)	WSCV (%)	MV_{diff}	SD_{diff}	CV (%)	WSCV (%)	MV_{diff}	SD_{diff}	CV (%)	WSCV (%)
FA	1.9e-02	2.3e-02	55	8.6	1.5e-02	1.7e-02	49.5	8.9	3.0e-02	2.6e-02	33.4	10.9	2.9e-02	3.0e-02	24	6.7
AD*	4.8e-05	6.0e-05	17	4.8	3.9e-05	4.9e-05	10.5	4.9	5.6e-05	6.0e-05	14.8	4.7	6.7e-05	7.9e-05	20	4.4
MD*	3.7e-05	4.7e-05	15	4.9	3.4e-05	4.0e-05	11.1	5.0	4.6e-05	5.0e-05	16.3	4.8	4.5e-05	6.1e-05	20	4.8
RD*	3.8e-05	4.6e-05	20	5.7	3.5e-05	3.9e-05	14.8	5.4	4.8e-05	5.0e-05	19.3	5.9	4.5e-05	5.9e-05	25	6.5

Abbreviations: AD, axial diffusivity; AK, axial kurtosis; CV, coefficient of variation; DKI, diffusion kurtosis imaging; DTI, diffusion tensor imaging; FA, fractional anisotropy; GM, gray matter; KA, kurtosis anisotropy; MD, mean diffusivity; MK, mean kurtosis; MV, mean value; RD, radial diffusivity; RK, radial kurtosis; SD, standard deviation; WM, white matter; WSCV, within-subject coefficient of variation.
*(mm^2/s).

TABLE 2 ICC, R^2, RI (100%*CR/MV), and bias of the data at time point 1 (TP1) and time point 2 (TP2) of 10 healthy volunteers for (A) DKI and (B) DTI metrics with MK-Curve correction. In addition, the total number of possible datapoints, the number of excluded values (implausible values), and the percentage (numbers excluded) of the sum of valid datapoints $+/-$ 2*SD of the mean is shown

(A) DKI

Metric	ROI	ICC	R^2	RI	Bias	Total number of possible values (i.e., number of measurements 20 * number of ROIs)	Numbers excluded (implausible values)	Numbers excluded (%)	Beyond 2 sigma	Beyond 2 sigma (%)
KA	Whole brain	0.947	0.898	13.1	-3.5e-03	2060	0	0	78/2060	3.8
KA	Cortex	0.873	0.781	11.8	-3.9e-03	1,400	0	0	17/1400	1.2
KA	Deep GM	0.855	0.774	8.8	$-1.1e-02$	120	0	0	8/120	6.7
KA	Deep WM	0.911	0.830	16.8	$-1.0e-03$	540	0	0	32/540	5.9
AK	Whole brain	0.862	0.751	39.6	$-6.2e-03$	2060	1	0.05	86/2059	4.2
AK	Cortex	0.677	0.484	51.5	$-6.8e-03$	1,400	1	0.07	40/1399	2.9
AK	Deep GM	0.770	0.609	61.4	$-5.3e-03$	120	0	0	8/120	6.7
AK	Deep WM	0.930	0.867	25.1	$-4.7e-03$	540	0	0	32/540	5.9
MK	Whole brain	0.975	0.951	6.6	1.7e-04	2060	0	0	52/2060	2.5
MK	Cortex	0.955	0.914	6.9	$-1.1e-04$	1,400	0	0	11/1400	0.8
MK	Deep GM	0.892	0.800	5.5	$-3.7e-03$	120	0	0	6/120	5.0
MK	Deep WM	0.957	0.917	6.1	1.8e-03	540	0	0	32/540	5.9
RK	Whole brain	0.964	0.932	14.8	1.1e-02	2060	5	0.2	115/2055	5.6
RK	Cortex	0.973	0.950	9.6	9.1e-03	1,400	0	0	74/1400	5.3
RK	Deep GM	0.544	0.323	14.2	6.1e-03	120	0	0	6/120	5.0
RK	Deep WM	0.843	0.716	18.5	1.9e-02	540	5	0.009	35/535	6.5

(B) DTI

Metric	ROI	ICC	R^2	RI	Bias	Total number of possible values (i.e., number of measurements 20 * number of ROIs)	Numbers excluded (implausible values)	Numbers excluded (%)	Beyond 2 sigma	Beyond 2 sigma (%)
FA	Whole brain	0.978	0.957	22.5	1.2e-03	2060	0	0	129/2060	6.3
FA	Cortex	0.972	0.945	22.9	$-6.5e-04$	1,400	0	0	62/1400	4.4
FA	Deep GM	0.871	0.762	33.3	6.1e-03	120	0	0	8/120	6.7
FA	Deep WM	0.919	0.847	18.6	5.0e-03	540	0	0	34/540	6.3
AD	Whole brain	0.938	0.883	11.4	$-9.4e-06$	2060	3	0.1	110/2057	5.3

TABLE 2 (Continued)

(B) DTI

Metric	ROI	ICC	R^2	RI	Bias	Total number of possible values (i.e., number of measurements 20 * number of ROIs)	Numbers excluded (implausible values)	Numbers excluded (%)	Beyond 2 sigma	Beyond 2 sigma (%)
AD	Cortex	0.887	0.793	9.8	−8.7e-06	1,400	0	0	74/1400	5.3
AD	Deep GM	0.881	0.806	14.0	−1.2e-05	120	0	0	7/120	5.8
AD	Deep WM	0.935	0.875	13.4	−1.1e-05	540	3	0.6	24/537	4.5
MD	Whole brain	0.916	0.844	11.5	−7.5e-06	2060	0	0	93/2060	4.5
MD	Cortex	0.896	0.809	9.8	−5.8e-06	1,400	0	0	68/1400	4.9
MD	Deep GM	0.894	0.827	14.4	−1.3e-05	120	0	0	4/120	3.3
MD	Deep WM	0.927	0.863	15.0	−1.1e-05	540	0	0	20/540	3.7
RD	Whole brain	0.940	0.886	13.5	−6.2e-06	2060	0	0	90/2060	4.4
RD	Cortex	0.925	0.859	11.1	−4.2e-06	1,400	0	0	65/1400	4.6
RD	Deep GM	0.900	0.826	16.7	−1.3e-05	120	0	0	6/120	5.0
RD	Deep WM	0.918	0.845	19.7	−9.9e-06	540	0	0	23/540	4.3

Abbreviations: AD, axial diffusivity; AK, axial kurtosis; CR, coefficient of repeatability; DKI, diffusion kurtosis imaging; DTI, diffusion tensor imaging; FA, fractional anisotropy; GM, gray matter; ICC, intraclass-correlation coefficient; KA, kurtosis anisotropy; MD, mean diffusivity; MK, mean kurtosis; MV, mean value; R2, square of the coefficient of determination; RD, radial diffusivity; RI, repeatability index; RK, radial kurtosis; ROI, region of interest; SD, standard deviation; WM, white matter.

3 | RESULTS

3.1 | Study population and data quality

All participants completed both measurement sessions and all the collected datasets had sufficient quality (Figure 1) for the subsequent analysis. The time between the sessions ranged from 2 h to 15 days (Table S2). The brain segmentation using both the WM MORI atlas[24] and the GM neuromorphometrics atlas[25] provided 206 ROIs (see the METHODS AND MATERIALS section). Taking all 10 subjects together and combining left and right hemispheres, the numbers of subject-specific datapoints at ROI level were 1400 (cortex), 120 (deep GM), 540 (deep WM), and 2060 (whole-brain ROI). Figure S1 shows the segmentation results and visualizes the ROI alignment with anatomical structure. A list of acronyms and a list of all the ROIs included in these subsets is available in Table S3.

Figure 1 shows a single axial slice of the averaged b = 0 s/mm², b = 500 s/mm², b = 1000 s/mm², and b = 2000 s/mm² images of the first five HCs at time point 1 (I) (TP1) and time point 2 (II) (TP2). Besides some folding artefacts (because of accelerated parallel imaging), no major artefacts or misalignments were observed across the subjects' averaged maps. Figure S2 shows the MV of the SNR based on the seven nonweighted images, which are realigned in MNI space before averaging over all 20 measurements. The MV ± SD of the SNR is 17.9 ± 6.2 in deep GM and 22.9 ± 10.1 in deep WM and, overall, it varies from 8.6 to 42.2 in the 95% range (2.5%–97.5%) (Table S4). The ROI-wise calculated SNR is listed in Table S1.

Figure 2 shows a single axial slice of the averaged KA, AK, MK, and RK values of the first five HCs at TP1 (I) and TP2 (II) with MK-Curve correction. As mentioned previously, DKI is noise sensitive. This is shown by Figure S3, which shows numerous artefacts in the DKI data without MK-Curve correction, illustrating the implausibility of the noncorrected DKI fit. Similarly, Figure S4 shows a single axial slice of the averaged DTI metrics (FA, AD, MD, and RD) at two time points in the first five HCs without relevant artefacts or misalignments.
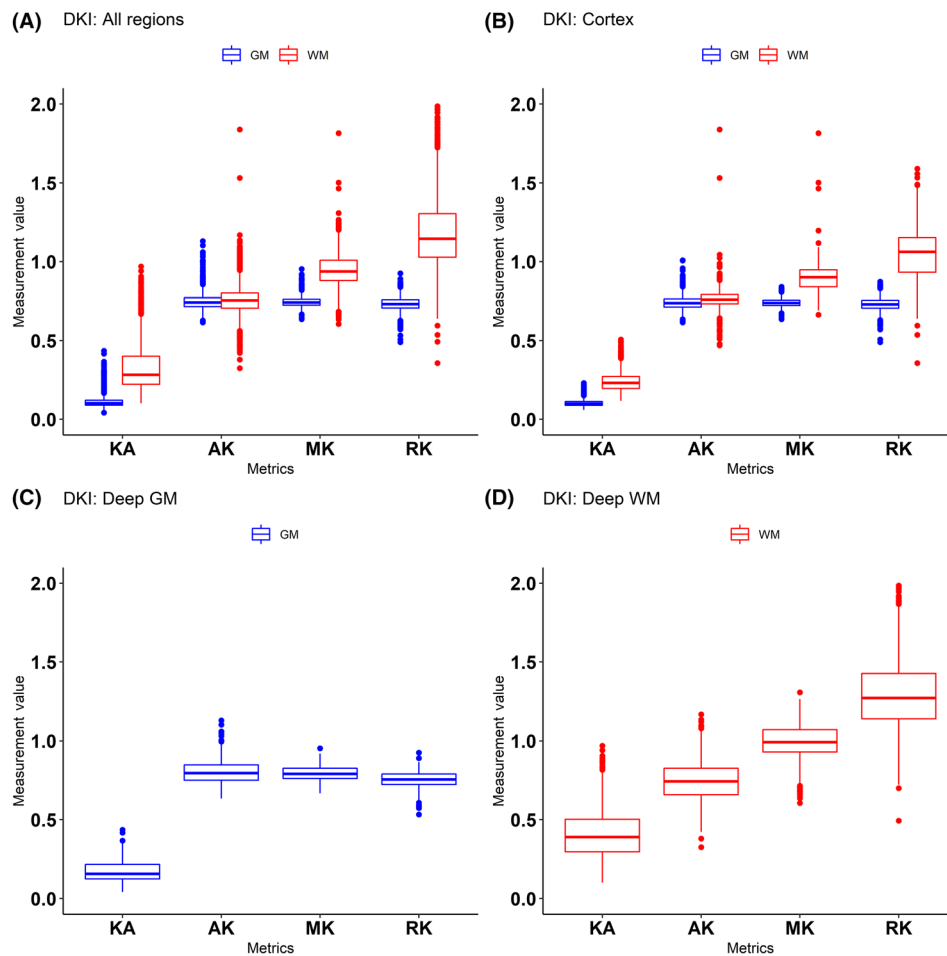
FIGURE 3 Boxplots showing the DKI measurement values with MK-Curve correction of DKI data for the metrics KA, AK, MK, and RK in GM (blue) and WM (red) for (A) All regions (whole brain), (B) The cortex, (C) Deep GM, and (D) Deep WM. Outliers are denoted by a dot. AK, axial kurtosis; DKI, diffusion kurtosis imaging; GM, gray matter; KA, kurtosis anisotropy; MK, mean kurtosis; RK, radial kurtosis; WM, white matter

3.2 | Test–retest assessments

Table 1 shows the MV, SD, and coefficient of variation (CV) for MK-Curve–corrected DKI and DTI metrics of the difference between TP1 and TP2. As expected, the intersubject CV values are much higher when compared with the wsCV, with a large variability across the subjects. All DKI and DTI metrics show a wsCV of less than 13%. The non-MK-Curve–corrected MV, SD, and CV of the DKI and DTI metrics are shown in Table S5, with a wsCV of up to 56%.

Table 2 includes the ICC, R^2, and RI values as correlation statistics and bias. All DKI and DTI metrics show good to excellent interscan agreement, with the exceptions of moderate agreement for AK in cortex (ICC = 0.68) and RK in deep GM (ICC = 0.54). R^2 is above 0.70 for all DKI and DTI metrics, except for AK in cortex ($R^2 = 0.48$), and RK in deep GM ($R^2 = 0.32$).

The RI depends on the metric and selected region and varies between 5.5% and 61.4% for DKI metrics and between 9.8% and 33.3% for DTI metrics. The highest RI values were observed for AK and were lowest for MK. The ICC, R^2, and RI of the non-MK-Curve–corrected DKI and DTI metrics are shown in Table S6. The ICC for DKI metrics varies from 0.006 to 0.52 and diffusion tensor-derived metrics from 0.06 to 0.66.

Figure 3 visualizes the boxplots of each kurtosis metric with MK-Curve correction of the two time points combined in the different regions of the brain. In general, the variance of the values is similar across regions but shows a larger variance in WM areas compared with GM areas. AK, RK, KA, and MK assume higher values in WM compared with GM. Measurement values are lowest for KA and highest for RK. Without MK-Curve correction, the DKI values show a larger spread at baseline (Figure 4) compared with the MK-Curve–corrected DKI data.
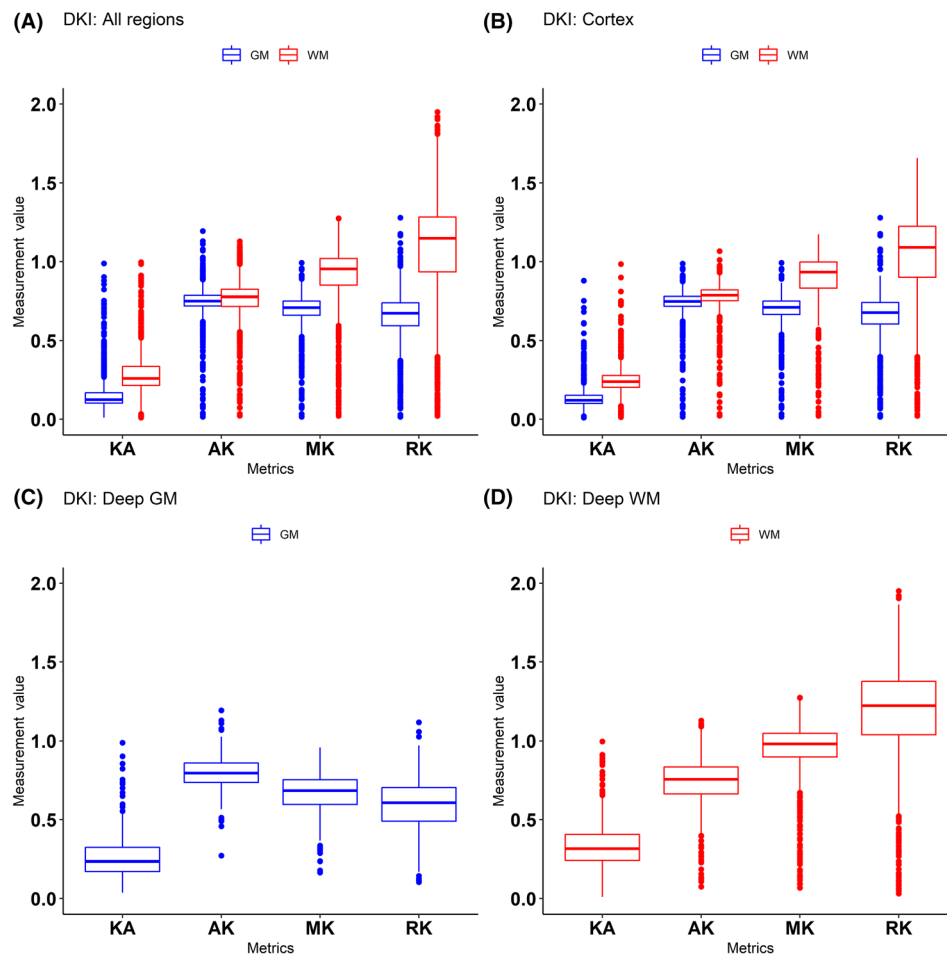
FIGURE 4 Boxplots showing the DKI measurement values without MK-Curve correction of DKI data for the metrics KA, AK, MK, and RK in GM (blue) and WM (red) for (A) All regions (whole brain), (B) The cortex, (C) Deep GM, and (D) Deep WM. Outliers are denoted by a dot. AK, axial kurtosis; DKI, diffusion kurtosis imaging; GM, gray matter; KA, kurtosis anisotropy; MK, mean kurtosis; RK, radial kurtosis; WM, white matter

Figure 5 shows the boxplots of each DTI metric. Compared with the results of the DKI metrics (Figure 3), the DTI metrics show a higher variation. RD shows the lowest values and interquartile range across the regions. The values of AD and MD in the cortex are higher in the GM ROIs, whereas the values in the deep GM ROIs are lower than those in deep WM ROIs.

Figures 6 and 7 show the scatterplots of each DKI metric measurement with MK-Curve correction (Figure 6) and without MK-Curve correction (Figure 7) of the first (x-axis) and second (y-axis) time point as well as the regression line (black). The combined scatterplots of each region and DKI metric with MK-Curve correction and without MK-Curve correction are shown in Figures S5 and S6. Most of the regression lines are well aligned with the 45° reference line (i.e., perfect reproducibility). DKI estimates without MK-curve correction proved less reproducible (Figures 7 and S6). In particular, a higher number of outliers and deviation of the regression lines from the 45° reference lines were observed.

Conversely, the BA plots of each DKI metric (Figure 8) and each metric and region combined (Figure S7) with MK-Curve correction show a good agreement, with values close to zero and no apparent biases (Table 2).

Figure 9 shows the scatterplots of the diffusion tensor metrics, which present a good correlation with the plotted regression lines, especially for FA. The metric FA values differ most in the deep GM from the reference line. The scatterplots of all ROIs combined of each metric are shown in Figure S8. BA plots of the DTI metrics (Figure 10) as well as the BA plots of combined DTI metrics (Figure S9) show good agreement with a normal distribution and lower values compared with the DKI metrics.

The comparison of the two groups (< 3 vs. ≥ 3 days) to examine the influence of time between measurements is shown by the density plot (Figure S10A). The distribution of the first versus the second group shows a similar shape with some outliers. Applying a Kolmogorov–Smirnov test did not show significant differences for KA ($p = 0.72$), AK ($p = 0.72$), and RK ($p = 1.00$). However, the test revealed significant differences in the distribution for the MK values ($p = 0.01$).
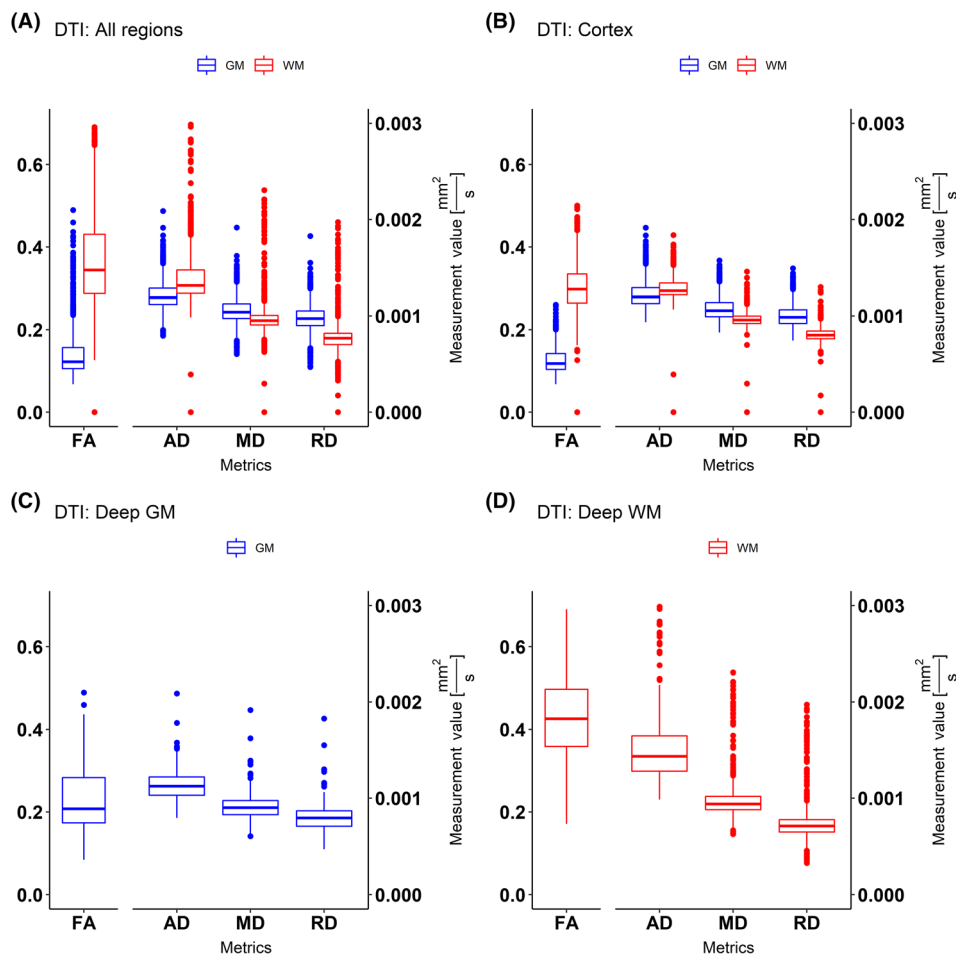
FIGURE 5 Boxplots showing DTI measurements derived from DKI fit for the metrics FA, AD, MD, and RD in GM (blue) and WM (red) for (A) All regions (whole brain), (B) Cortex, (C) Deep GM, and (D) Deep WM. Outliers are denoted by a dot. FA has a different scaling compared with the other metrics. AD, axial diffusivity; DKI, diffusion kurtosis imaging; DTI, diffusion tensor imaging; FA, fractional anisotropy; GM, gray matter; MD, mean diffusivity; RD, radial diffusivity; WM, white matter

4 | DISCUSSION

In the current study, we investigated the interscan reproducibility and test–retest reliability of DKI and DTI metrics derived with a DKI fit in different regions of the brain of 10 healthy subjects at 3 T. We observed good to excellent agreement of both MK-Curve–corrected kurtosis tensor- and diffusion tensor-derived metrics, with ICC values in the range 0.77–0.98 for the former, and 0.87–0.98 for the latter. However, only moderate agreement was observed for the DKI metrics AK (ICC = 0.68) in cortex regions and RK in deep GM (ICC = 0.54). The ICC values of the non-MK-Curve–corrected diffusion kurtosis and diffusion tensor-derived metrics are considerably lower, with values for the former of 0.006–0.52 and 0.06–0.66 for the latter, indicating poor to moderate reliability.

Our results show a good agreement of each DKI metric in the different brain regions with some regional variation, which is comparable with previously published data.[10–14] The metrics show regional variation reflecting the different underlying architecture of the nervous tissue. ICC values show good to excellent agreement, with the exception of AK (cortex) and RK (deep GM).

The study by Kasa et al.[10] reported slightly higher wsCVs with values of less than 20% for all metrics compared with the values acquired in this study. They analyzed preprocessed data from the Human Connectome Project at a slightly lower resolution (1.25 mm^3) in the WM of six commonly affected lobes in neurological disorders and in the GM of five lobes, not covering the whole brain, as was carried out in this study. Shahim et al.[13] report mean wsCVs of DKI (≤ 4.5%) and DTI (< 7%) metrics in three concussive traumatic brain injury patients and four HCs. We report slightly higher wsCVs, indicating higher intrasubject variability of the measurements, representing the fact that subject noise more strongly influences the DKI fit model compared with the standard DTI model. In addition, the reported AD and FA values are higher than previously reported[7–9] and show the model complexity of the DKI fit (and DTI-derived metrics thereof).
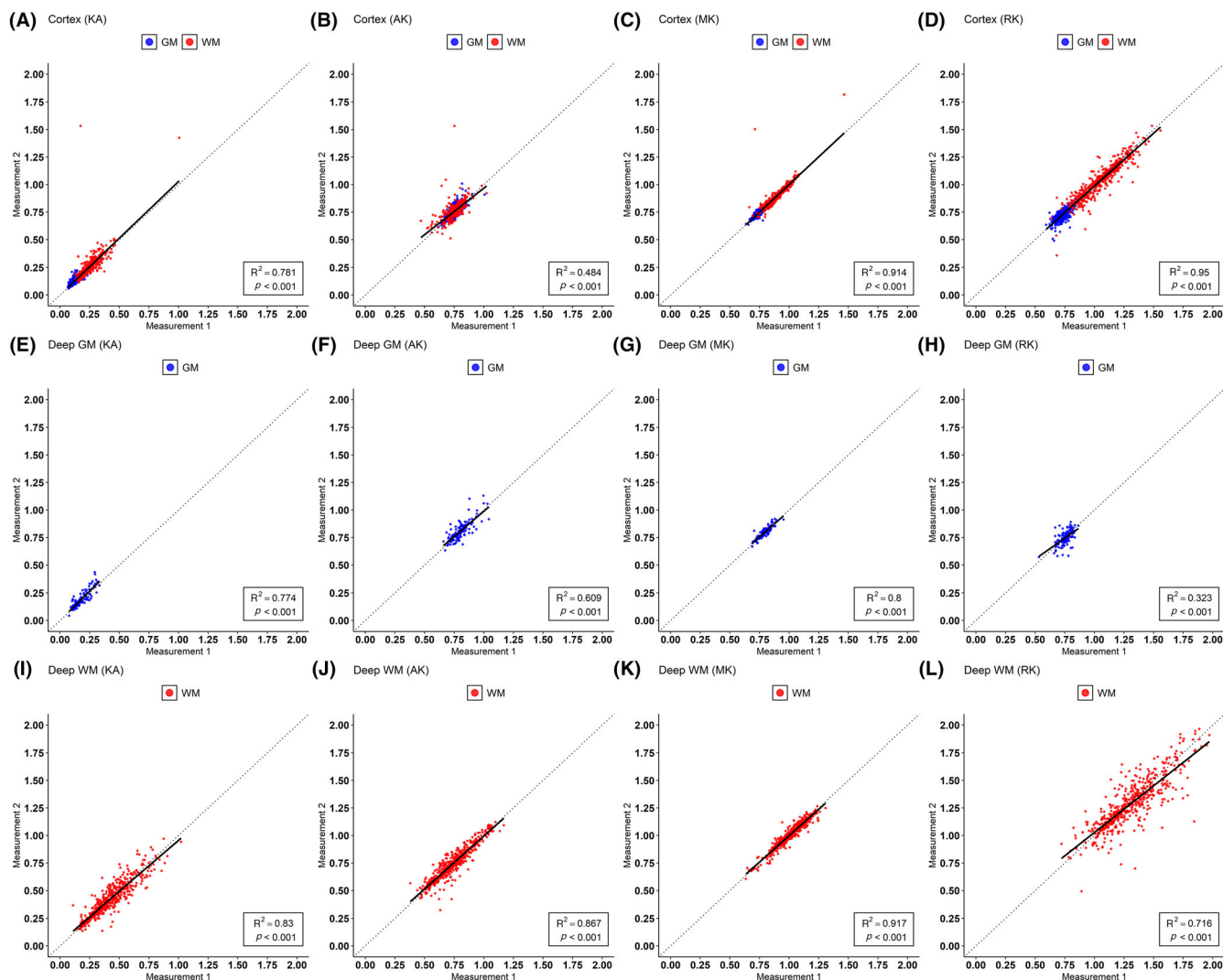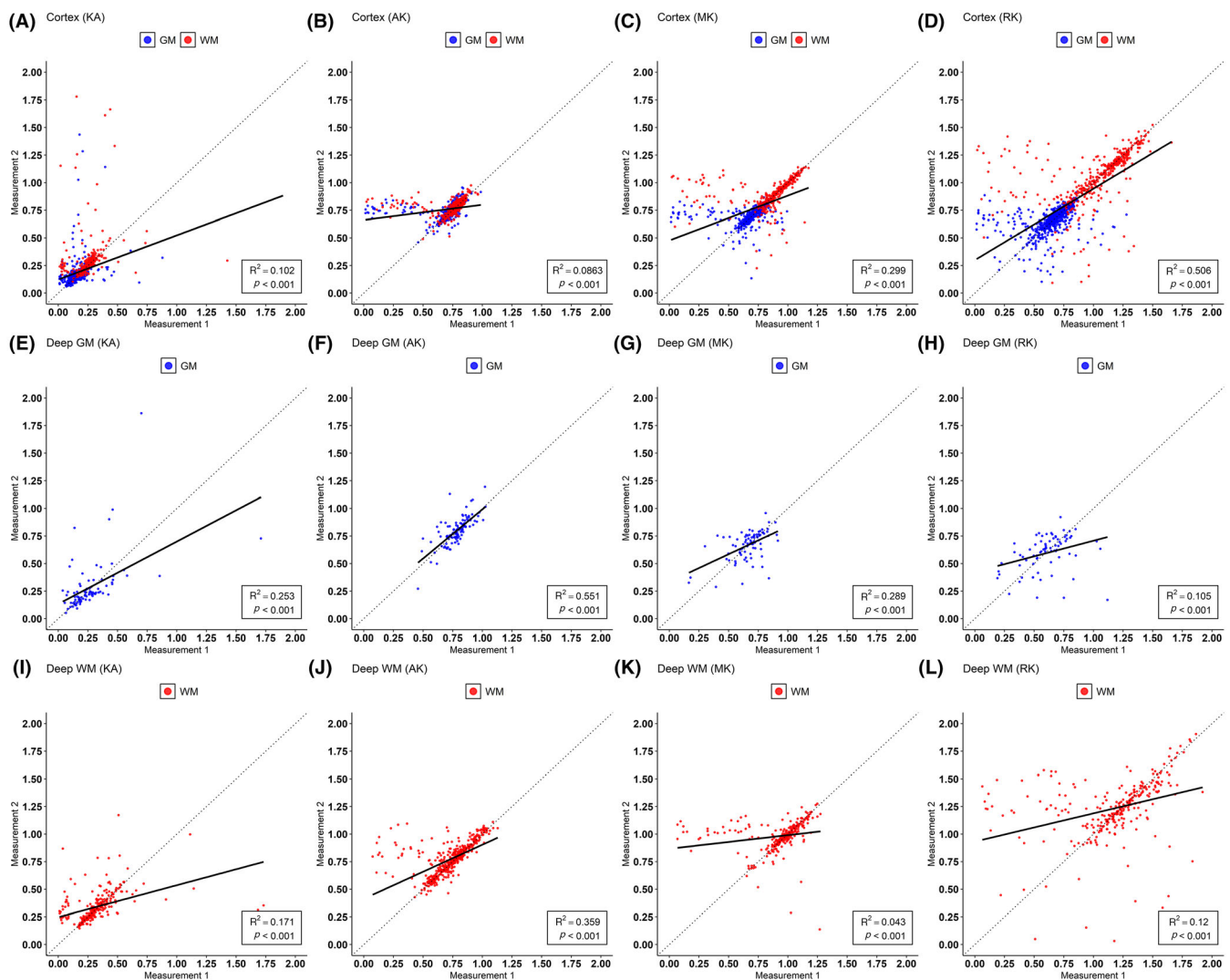
FIGURE 6 Correlation scatterplots showing the DKI data of the first measurement (x-axis) and second measurement (y-axis) with MK-Curve correction of DKI data. The single metrics are shown for in (A–D) The Cortex, (E–H) Deep GM, and (I–L) Deep WM. The regression line is shown in black and R^2 is denoted. DKI, diffusion kurtosis imaging; GM, gray matter; MK, mean kurtosis; R^2, square of the coefficient of determination; WM, white matter

Previous reports on DTI variability relied on metrics derived from the DTI fit itself, whereas here we are reporting the variability of FA and MD derived from a DKI fit. Considering that DKI fits a much larger number of parameters (22 vs. seven), it is reasonable to expect larger variability for the estimated metrics. Furthermore, the data were acquired on a clinical scanner with a time interval varying between hours and weeks, reflecting clinical practice. The paper by Konieczny et al.[14] reported excellent intersite reproducibility for DKI and DTI metrics with an ICC of more than 0.93 in 10 patients with small vessel disease using tract-based spatial statistics to compensate for misregistration errors biasing the results.[33] Also, a custom mask was used to exclude all areas typically susceptible to cerebrospinal fluid (CSF) partial volume effects, and a different fit model was used to estimate the DKI metrics, which influences the variability.

In addition to previous papers, we investigated the influence of the time interval between measurements. As illustrated by Figure S10, there is an overall good overlap between the two measurements with a few outliers, but a time dependence of the MK metric. In the first group, the difference of the values between the two measurements is greater than that for the second group.

To address the fact that DKI is an unstable method, Henriques et al.[12] focused on a DKI estimator from powder-averaged diffusion-weighted data with comparison with other fit models. Compared with our study, however, this method can only compute MK. We applied a different method (MK-Curve), which corrects voxel-wise only implausible voxels focusing on MK. Correcting all voxels in commonly used regularization methods could influence measurements in voxels that otherwise would not be corrected by using only the MK-Curve method. This also potentially prevents the introduction of partial volume effects when adjacent voxels derive from different brain tissues.[15]

FIGURE 7 Correlation scatterplots showing the DKI data of the first measurement (x-axis) and second measurement (y-axis) without MK-Curve correction of DKI data. The single metrics are shown for in (A–D) The Cortex, (E–H) Deep GM, and (I–L) Deep WM. The regression line is shown in black and R^2 is denoted. DKI, diffusion kurtosis imaging; GM, gray matter; MK, mean kurtosis; R^2, square of the coefficient of determination; WM, white matter

The key concept behind the MK-Curve method is to adjust the value of the nonweighted image to obtain plausible and reliable DKI estimates. Different from methods enforcing constraints in the optimization function or regularizing diffusion-weighted images, manipulating S_0 offers two advantages: (i) nonweighted data are simpler to manipulate because they are independent from the underlying tissue anisotropy; and (ii) there typically are few nonweighted measurements, ensuring the correction only affects a minimal amount of data, and does not alter the relations between diffusion-weighted images at different b-values.

No consensus standard acquisition protocol has been established for DKI, but it is important to obtain a balance between image quality and acquisition time. The scan time is mainly determined by the number of b-values and it should be kept in mind that b-values above 3000 s/mm² result in poorer fit quality.[34] Kasa et al.[10] demonstrated that comparable reproducibility can be obtained by using either a three or a two multishell protocol with lower b-values; the latter is the study protocol we opted for in this study. Technologies such as multiband might enable doubling the number of directions acquired in the same time, which will probably further reduce the CV.

A different way to detect and address possible outliers also depends on the type of model used to estimate the diffusion tensor or kurtosis tensor from the diffusion data. Despite the use of an advanced fit method in our study, we still observed the presence of a few outliers in some regions, which might be given by artifacts. However, different methods exist to estimate the diffusion tensor and kurtosis tensor from the diffusion data, such as maximum likelihood, least squares, Bayesian, and the robust approaches.[35] The linear least squares estimate is widely used to
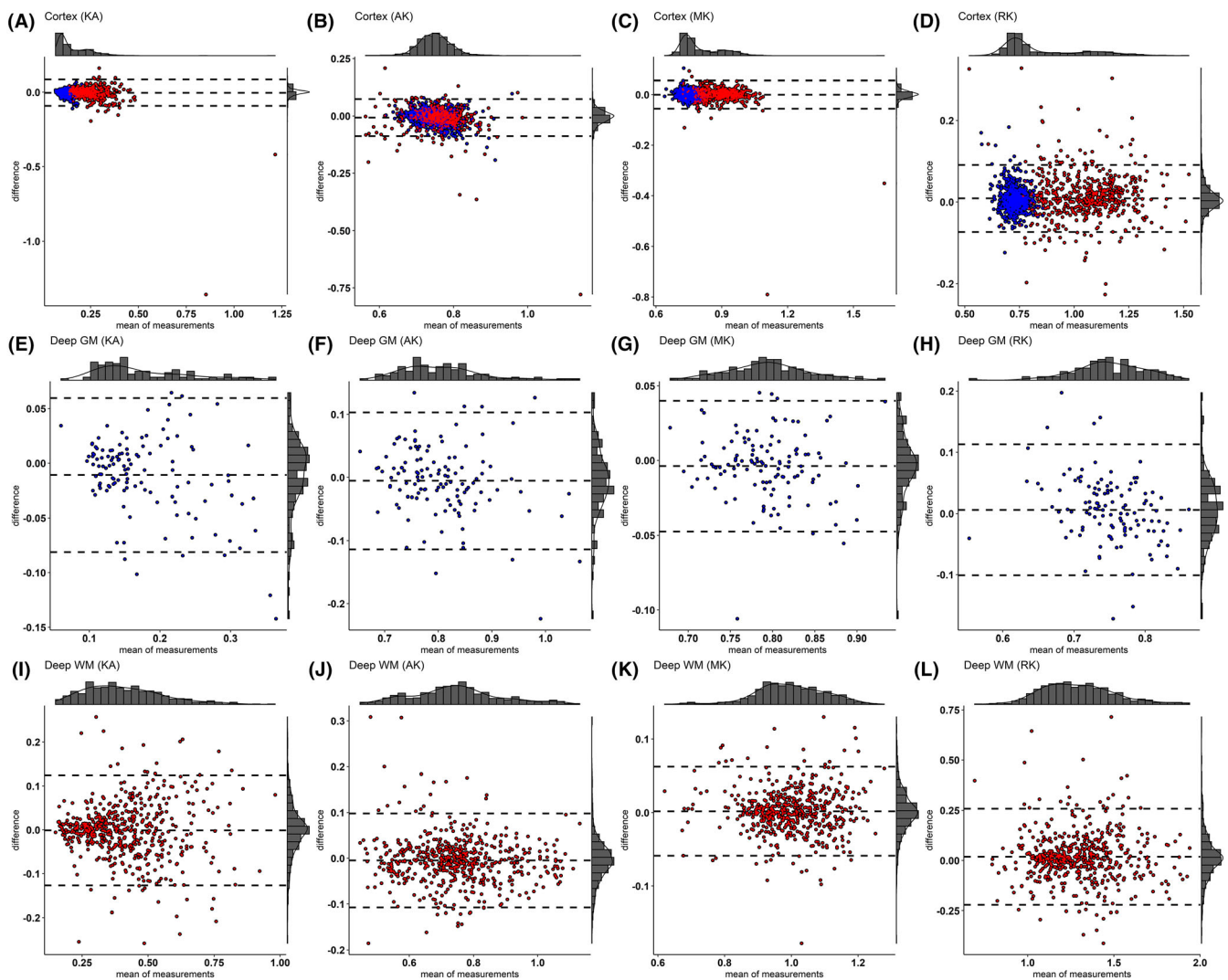
FIGURE 8 Bland–Altman plots showing the DKI data. The single metrics are shown for in (A–D) The cortex, (E–H) Deep GM, and (I–L) Deep WM. The distribution of the datapoints along the axes is shown at the margin of the plots. DKI, diffusion kurtosis imaging; GM, gray matter; WM, white matter

calculate the diffusion parameters but is particular sensitive to outliers and should therefore be avoided. Robust fit models with constraints, such as REKINDLE,[36] could be a valid alternative in the presence of many motion artefacts.

We selected the most important regions and a priori avoided ROIs with expected strong susceptibility artefacts and/or B0 inhomogeneities, such as the ventricle system or CSF areas, therefore reducing the number of regions from 224 to 206, resulting in the group whole-brain ROI. In addition, we grouped these ROIs according to their anatomy and function into three groups (cortex, deep GM, and deep WM) to examine the behavior of DKI and DTI metrics in these areas. DKI typically uses higher gradient strengths compared with DTI, with b-values of up to 3000 s/mm^2. Because of limited maximal gradient magnitude, this is achieved by increasing the gradient duration, which results in longer acquisition times, making DKI more prone to movement artefacts[37] and produces additional outliers. This is visible in folding artefacts, which originate from the required accelerated parallel imaging. Motion correction and value adjustment control accounts for this in postprocessing. However, in our study, we achieved overall sufficient SNR to quantify the DKI model with an acquisition time for the diffusion sequence of 15 min, achieving SNR values of between 10 and 46. Values may locally reach more than 40, and the majority lie in a suitable range. The SNR was on average 18 (GM) to 23 (WM), which is typical for a clinical study, and in line with commonly available research datasets (e.g., the SNR of the MASSIVE[38]), is around 16.[39] While differences between DKI estimates with and without the MK-Curve method are expected to decrease with increasing SNR, in their original work, Zhang et al. showed that the MK-Curve method also improved DKI estimates with high quality data, such as those from the Human Connectome Project.[15] Accordingly, we recommend the use of the MK-Curve method or other recently proposed alternatives,[10,12,14] when fitting DKI to both clinical and research quality data.
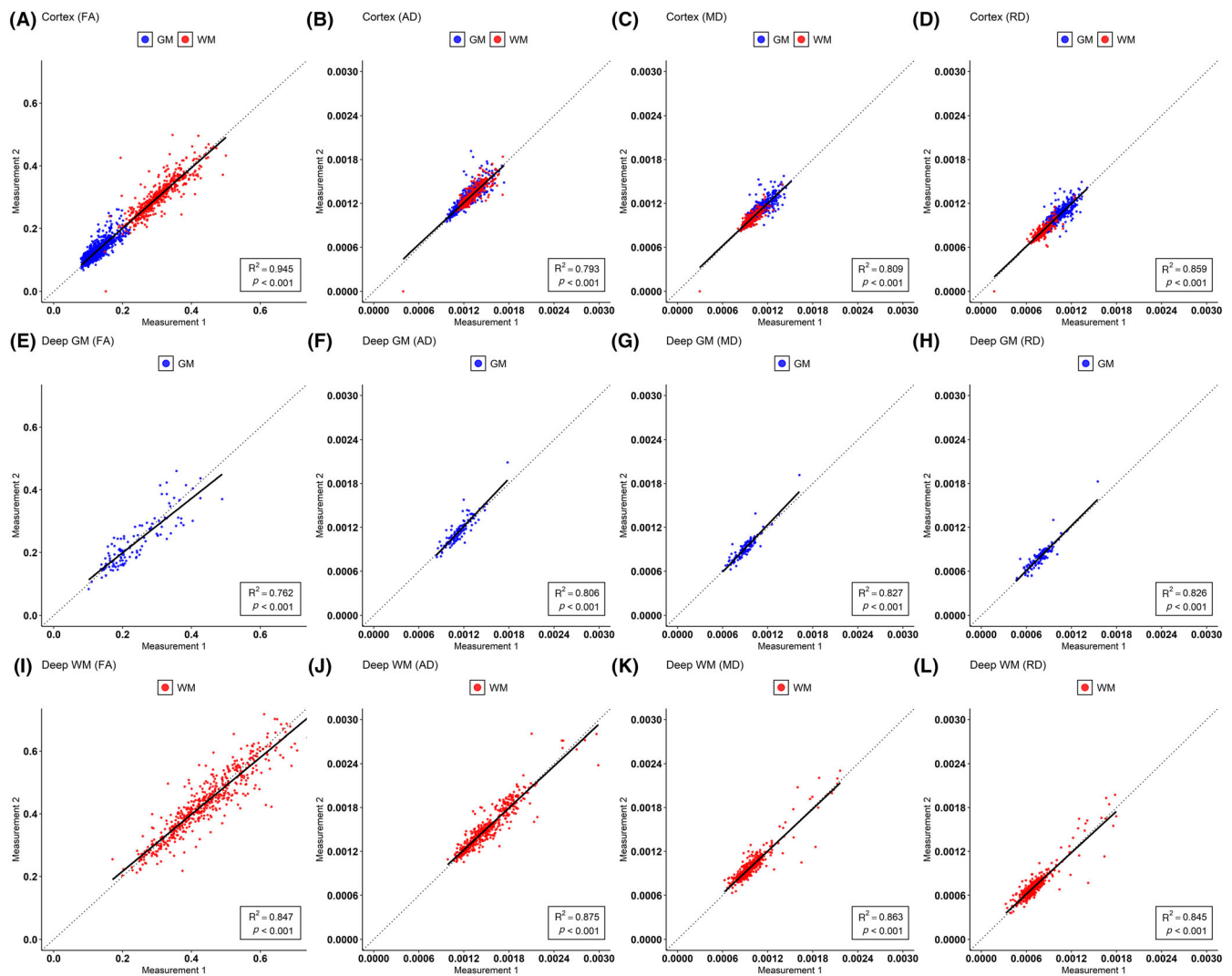
FIGURE 9 Correlation scatterplots showing the DTI data of the first measurement (x-axis) and second measurement (y-axis). The single metrics are shown for in (A–D) The cortex, (E–H) Deep GM, and (I–L) Deep WM. The regression line is shown in black and R^2 is denoted. DTI, diffusion tensor imaging; GM, gray matter; R^2, square of the coefficient of determination; WM, white matter

4.1 | Limitations

The limitations of this study include the application of one regularization method and the number of participants (10 HCs). As previously mentioned, DKI relative to DTI requires a longer acquisition time, which makes this method more sensitive to artefacts because of patient movement.

Focusing on large ROIs would enhance the statistical power of any subsequent analysis, but this requires a priori knowledge on specific areas of interest, which might not be available in explorative studies. For this reason, we provide here an overview of DTI and DKI metrics, their variability and repeatability in multiple ROIs of variable size, which can guide the design of future studies providing information on which regions are more or less reliable to study with a clinically feasible DKI protocol.

Our analysis of a large number of ROIs of different sizes, compared with previous studies, could therefore be of value to clinical studies, especially those with high-SNR data.

As mentioned previously, we calculated the SNR of every dataset by voxel-wise division of the mean of the $b = 0$ s/mm^2 images divided by their SD. If one considers the voxel-wise SD to be a sum of thermal noise and tissue heterogeneity, our approach leads to a conservative underestimation of the final SNR.

There are different strategies to calculate the SNR that should be preferred where possible, for instance, by using the SD of the background signal outside the brain. However, this might not be straightforward to implement on a clinical MRI system like the one employed in this study, because the images are automatically masked in the background as part of the MRI reconstruction.
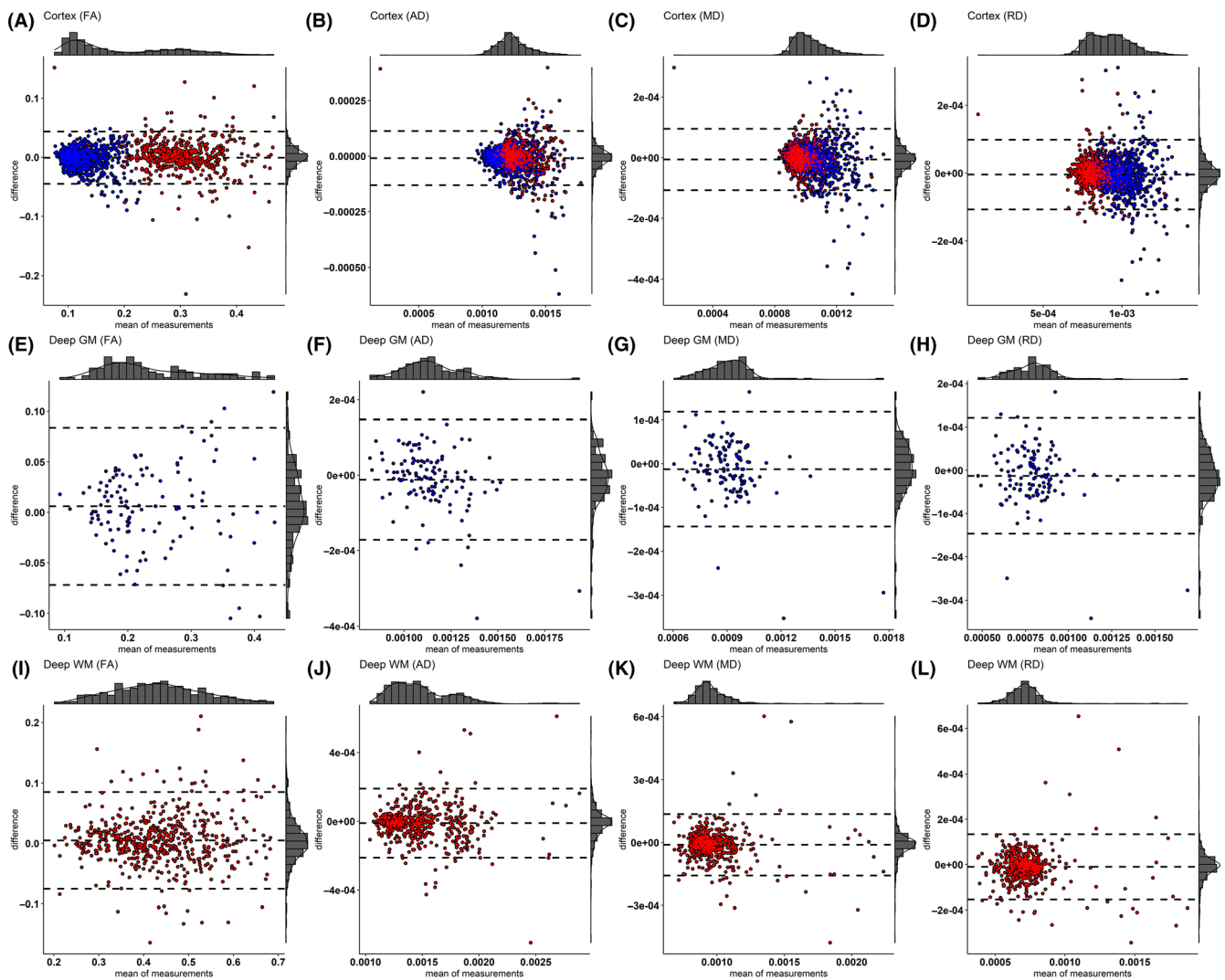
FIGURE 10 Bland–Altman plots showing the DTI data. The single metrics are shown for in (A–D) The cortex, (E–H) Deep GM, and (I–L) Deep WM. The distribution of the datapoints along the axes is shown at the margin of the plots. DTI, diffusion tensor imaging; GM, gray matter; WM, white matter

5 | CONCLUSIONS

In conclusion, we demonstrated that the MK-Curve–corrected DKI model reliably quantifies kurtosis tensor and diffusion tensor metrics, which can be applied in longitudinal studies for studying and assessing changes in disease with mandatory careful inspection of artefacts.

ORCID

Dirk Lehnick https://orcid.org/0000-0003-1836-2811
Alexander Leemans https://orcid.org/0000-0002-9306-6126
Alberto De Luca https://orcid.org/0000-0002-2553-7299

REFERENCES

1. Stejskal EO, Tanner JE. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J Chem Phys*. 1965;42(1):288-292. doi:10.1063/1.1695690

2. Kärger J. NMR self-diffusion studies in heterogeneous systems. *Adv Colloid Interface Sci*. 1985;23:129-148. doi:10.1016/0001-8686(85)80018-X

3. Jensen JH, Helpern JA. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR Biomed*. 2010;23(7):698-710. doi:10.1002/nbm.1518

4. Fieremans E, Jensen JH, Helpern JA. White matter characterization with diffusional kurtosis imaging. *Neuroimage*. 2011;58(1):177-188. doi:10.1016/j.neuroimage.2011.06.006

5. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. *Magn Reson Med*. 2005;53(6):1432-1440. doi:10.1002/mrm.20508

6. De Luca A, Bertoldo A, Froeling M. Effects of perfusion on DTI and DKI estimates in the skeletal muscle. *Magn Reson Med*. 2017;78(1):233-246. doi:10.1002/mrm.26373

7. Ciccarelli O, Parker GJ, Toosy AT, et al. From diffusion tractography to quantitative white matter tract measures: a reproducibility study. *Neuroimage*. 2003;18(2):348-359. doi:10.1016/s1053-8119(02)00042-3

8. Pfefferbaum A, Adalsteinsson E, Sullivan EV. Replicability of diffusion tensor imaging measurements of fractional anisotropy and trace in brain. *J Magn Reson Imaging*. 2003;18(4):427-433. doi:10.1002/jmri.10377

9. Heiervang E, Behrens TE, Mackay CE, Robson MD, Johansen-Berg H. Between session reproducibility and between subject variability of diffusion MR and tractography measures. *Neuroimage*. 2006;33(3):867-877. doi:10.1016/j.neuroimage.2006.07.037

10. Kasa LW, Haast RAM, Kuehn TK, et al. Evaluating high spatial resolution diffusion kurtosis imaging at 3T: reproducibility and quality of fit. *J Magn Reson Imaging*. 2021;53(4):1175-1187. doi:10.1002/jmri.27408

11. Ades-Aron B, Veraart J, Kochunov P, et al. Evaluation of the accuracy and precision of the diffusion parameter EStImation with Gibbs and NoisE removal pipeline. *Neuroimage*. 2018;183:532-543. doi:10.1016/j.neuroimage.2018.07.066

12. Henriques RN, Jespersen SN, Jones DK, Veraart J. Toward more robust and reproducible diffusion kurtosis imaging. *Magn Reson Med*. 2021;86(3):1600-1613. doi:10.1002/mrm.28730

13. Shahim P, Holleran L, Kim JH, Brody DL. Test-retest reliability of high spatial resolution diffusion tensor and diffusion kurtosis imaging. *Sci Rep*. 2017;7:11141. doi:10.1038/s41598-017-11747-3

14. Konieczny MJ, Dewenter A, Telgte AT, et al. Multi-shell diffusion MRI models for white matter characterization in cerebral small vessel disease. *Neurology*. 2020;96:e698-e708. doi:10.1212/WNL.0000000000011213

15. Zhang F, Ning L, O'Donnell LJ, Pasternak O. MK-curve - characterizing the relation between mean kurtosis and alterations in the diffusion MRI signal. *Neuroimage*. 2019;196:68-80. doi:10.1016/j.neuroimage.2019.04.015

16. Leemans A JB, Sijbers J, Jones DK. ExploreDTI: A Graphical Toolbox for Processing, Analyzing, and Visualizing Diffusion MR Data. 2009:3537.

17. De Luca A. MRIToolkit. http://www.mritoolkit.org

18. Vos SB, Tax CM, Luijten PR, Ourselin S, Leemans A, Froeling M. The importance of correcting for signal drift in diffusion MRI. *Magn Reson Med*. 2017;77(1):285-299. doi:10.1002/mrm.26124

19. Veraart J, Novikov DS, Christiaens D, Ades-Aron B, Sijbers J, Fieremans E. Denoising of diffusion MRI using random matrix theory. *Neuroimage*. 2016;142:394-406. doi:10.1016/j.neuroimage.2016.08.016

20. Perrone D, Aelterman J, Pižurica A, Jeurissen B, Philips W, Leemans A. The effect of Gibbs ringing artifacts on measures derived from diffusion MRI. *Neuroimage*. 2015;120:441-455. doi:10.1016/j.neuroimage.2015.06.068

21. Leemans A, Jones DK. The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn Reson Med*. 2009;61(6):1336-1349. doi:10.1002/mrm.21890

22. Poot DH, den Dekker AJ, Achten E, Verhoye M, Sijbers J. Optimal experimental design for diffusion kurtosis imaging. *IEEE Trans Med Imaging*. 2010;29(3):819-829. doi:10.1109/TMI.2009.2037915

23. Gaser C, Dahnke R. CAT-A Computational Anatomy Toolbox for the Analysis of Structural MRI Data. 2016.

24. Rezende TJR, Campos BM, Hsu J, et al. Test-retest reproducibility of a multi-atlas automated segmentation tool on multimodality brain MRI. *Brain Behav*. 2019;9(10):e01363. doi:10.1002/brb3.1363

25. Bakker R, Tiesinga P, Kotter R. The scalable brain atlas: instant web-based access to public brain atlases and related content. *Neuroinformatics*. 2015;13(3):353-366. doi:10.1007/s12021-014-9258-x

26. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

27. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag; 2016. doi:10.1007/978-3-319-24277-4

28. Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. https://wilkelab.org/cowplot/

29. Datta D. blandr: a Bland-Altman Method Comparison package for R. https://github.com/deepankardatta/blandr

30. Attali D, Baker C. ggExtra: Add Marginal Histograms to 'ggplot2', and More 'ggplot2' Enhancements. https://github.com/daattali/ggExtra

31. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428. doi:10.1037//0033-2909.86.2.420

32. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

33. Bach M, Laun FB, Leemans A, et al. Methodological considerations on tract-based spatial statistics (TBSS). *Neuroimage*. 2014;100:358-369. doi:10.1016/j.neuroimage.2014.06.021

34. Lu H, Jensen JH, Ramani A, Helpern JA. Three-dimensional characterization of non-Gaussian water diffusion in humans using diffusion kurtosis imaging. *NMR Biomed*. 2006;19(2):236-247. doi:10.1002/nbm.1020

35. Koay CG. Least squares approaches to diffusion tensor estimation. In: Jones DK, ed. *Diffusion MRI: theory, methods, and applications*. Oxford University Press; 2010. doi:10.1093/med/9780195369779.003.0016

36. Tax CM, Otte WM, Viergever MA, Dijkhuizen RM, Leemans A. REKINDLE: robust extraction of kurtosis INDices with linear estimation. *Magn Reson Med*. 2015;73(2):794-808. doi:10.1002/mrm.25165

37. Li X, Yang J, Gao J, et al. A robust post-processing workflow for datasets with motion artifacts in diffusion kurtosis imaging. *PLoS ONE*. 2014;9(4):e94592. doi:10.1371/journal.pone.0094592

38. Froeling M, Tax CMW, Vos SB, Luijten PR, Leemans A. "MASSIVE" brain dataset: Multiple acquisitions for standardization of structural imaging validation and evaluation. *Magn Reson Med*. 2017;77(5):1797-1809. doi:10.1002/mrm.26259

39. De Luca A, Ianus A, Leemans A, et al. On the generalizability of diffusion MRI signal representations across acquisition parameters, sequences and tissue types: Chronicles of the MEMENTO challenge. *Neuroimage*. 2021;240:118367. doi:10.1016/j.neuroimage.2021.118367

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.