



The adverse effect of gradient nonlinearities on diffusion MRI: From voxels to group studies

Hamed Y. Mesri^{*}, Szabolcs David, Max A. Viergever, Alexander Leemans

Image Sciences Institute, University Medical Center Utrecht and Utrecht University, Utrecht, the Netherlands

ARTICLE INFO

Keywords:

Diffusion MRI
Diffusion tensor imaging
Diffusional kurtosis imaging
Gradient nonlinearities
Gradient nonlinearity correction
Fiber tractography
Fractional anisotropy
Mean diffusivity
Group-wise study

ABSTRACT

Nonlinearities of gradient magnetic fields in diffusion MRI (dMRI) can introduce systematic errors in estimates of diffusion measures. While there are correction methods that can compensate for these errors, as presented in the Human Connectome Project, such nonlinear effects are often assumed to be negligible for typical applications, and as a result, gradient nonlinearities are mostly left uncorrected. In this work, we perform a systematic analysis to investigate the effect of gradient nonlinearities on dMRI studies, from voxel-wise estimates to group study outcomes. We present a novel framework to quantify and visualize these effects by decomposing them into their magnitude and angle components. Mean magnitude deviation and fractional gradient anisotropy are introduced to quantify the distortions in the size and shape of gradient vector distributions. By means of Monte-Carlo simulations and real data from the Human Connectome Project, the errors on dMRI measures derived from the diffusion tensor imaging and diffusional kurtosis imaging are highlighted. We perform a group study to showcase the alteration in the significance and effect size due to ignoring gradient nonlinearity correction. Our results indicate that the effect of gradient field nonlinearities on dMRI studies can be significant and may complicate the interpretation of the results and conclusions.

1. Introduction

Newest developments in acquisition and modeling techniques aiming at studying tissue microstructure at sub-voxel scales by diffusion magnetic resonance imaging (dMRI) demand imaging with higher diffusion-weighting (DW) gradient amplitudes (Van Essen et al., 2012; Setsompop et al., 2013; McNab et al., 2013; Ferizi et al., 2015). As a result, the state-of-the-art scanners designed for dMRI studies such as the Connectome scanner utilize DW gradient amplitudes as large as 300 mT/m to increase the q-space resolution and provide better capability of probing tissue microstructure at scales much smaller than the voxel size (Setsompop et al., 2013; McNab et al., 2013; Sotiropoulos et al., 2013; Dyrby et al., 2013).

Utilizing higher gradient amplitudes results in higher gradient field nonlinearities (Setsompop et al., 2013; Bammer et al., 2003). As a result of these spatial nonuniformities, the further a voxel is from the isocenter of the gradient coil system, the more the deviation from the nominal gradient strength it will experience. These effects increase not only by increasing the gradient coil diameter but also by decreasing the gradient coil length.

The nonuniformities of the magnetic field gradients affect both the

imaging and the DW gradients since the same gradient system is used in both steps. The effect of gradient nonlinearities on the imaging gradients results in pixel shifts and the consequent spatial image warping especially at the edges of the image (Michiels et al., 1994; Sumanaweera et al., 1994; Langlois et al., 1999). These effects were known to the manufacturers for a long time and correction techniques based on spherical harmonics (SH) are already in use to compensate for the resulting errors and improve the spatial fidelity (Glover and Pelc, 1983; Janke et al., 2004; Doran et al., 2005; Tao et al., 2015). However, since these correction techniques are already implemented in the scanners, the SH coefficients are not usually provided to the regular users.

Nonuniformities of the DW gradient fields also result in errors in the dMRI measures due to the false assumption of linear gradients over the whole field-of-view (FOV) and ignoring the variations in the gradient vectors at different voxels during the parameter estimation (Sotiropoulos et al., 2013; Bammer et al., 2003; Conturo et al., 1996; Malyarenko et al., 2014; Borkowski et al., 2017; Mohammadi et al., 2012; Nagy et al., 2009). These effects have been previously addressed by Bammer et al. (2003), where the effect of gradient nonlinearity on the measures derived from diffusion tensor imaging (DTI) (Basser et al., 1994) was investigated by using the so-called gradient coil tensor to model the relation between

^{*} Corresponding author. PROVIDI Lab, UMC Utrecht, Q.02.4.45, P.O. Box 85500, 3508 GA, Utrecht, the Netherlands.

E-mail address: hamed@isi.uu.nl (H.Y. Mesri).

the desired and actual gradient fields at a specific location in the FOV. Consequently, a correction technique was proposed by using the gradient coil tensor to calculate the actual gradient field at every voxel of the image.

Bammer's method (Bammer et al., 2003) was used in the Human Connectome Project (HCP) and the gradient coil tensor data were provided for every subject in the cohort. While the correction for gradient nonlinearity was recommended, it was not considered necessary. As such, all the HCP data, which are currently available, still contain the detrimental effects of gradient nonlinearity of the DW gradients.

In this work, we present a novel framework that systematically investigates the importance of correcting gradient field nonlinearity in the HCP cohort before these data were to be used for further neuroscientific investigation. With additional simulations, we generalize the importance of this correction step for all diffusion imaging data. While the absolute magnitude of the errors caused by ignoring the gradient nonlinearity correction (GNC) during preprocessing step may not be solely of specific interest to the research and medical community, understanding their relative contribution and impact on tractography or group-wise studies is highly relevant to interpret the validity of subsequent findings. As such, we also investigate the effect of neglecting GNC on tractography and a group-wise study. As an example for the latter, potential gender differences in terms of diffusion measures are investigated in line with previous reports (Ritchie et al., 2018; Wierenga et al., 2017; Núñez et al., 2018; Tian et al., 2017; Ingallhalikar et al., 2014; Ruigrok et al., 2014; Kanaan et al., 2012; Herting et al., 2012; Menzler et al., 2011; Hsu et al., 2008; Westerhausen et al., 2003). The resulting significance and effect size in this gender group-wise study is then evaluated with and without applying GNC to the dMRI data. Preliminary results of this work were presented at the 2017 and 2018 ISMRM meetings (Mesri et al., 2017a, 2018a).

2. Theory

Due to imperfections of the gradient coil systems, a gradient field applied in one direction also generates gradient fields in the other two orthogonal directions. The resulting nonuniformities in the magnetic field gradients can be modeled as (Bammer et al., 2003):

$$\mathbf{g}_a(\mathbf{r}) = \mathbf{L}(\mathbf{r})\mathbf{g}_n(\mathbf{r}), \quad (1)$$

where $\mathbf{g}_a(\mathbf{r})$ and $\mathbf{g}_n(\mathbf{r})$ represent the actual and nominal (desired) gradients at position \mathbf{r} , respectively. The nominal gradients are predefined assuming that the gradient vector distribution is uniform over the whole FOV. However, in reality, due to the gradient nonlinearity effects, the actual distribution of gradient vectors is not uniform and depends on the spatial position. The actual gradients are the ones which sensitize the MR signal to the underlying diffusion in reality. $\mathbf{L}(\mathbf{r})$. Denotes the gradient coil tensor,¹ which indicates the deviations in the gradient field for each principal axis, i.e.,

$$\mathbf{L}(\mathbf{r}) = \begin{bmatrix} L_{xx}(\mathbf{r}) & L_{xy}(\mathbf{r}) & L_{xz}(\mathbf{r}) \\ L_{yx}(\mathbf{r}) & L_{yy}(\mathbf{r}) & L_{yz}(\mathbf{r}) \\ L_{zx}(\mathbf{r}) & L_{zy}(\mathbf{r}) & L_{zz}(\mathbf{r}) \end{bmatrix}. \quad (2)$$

The element L_{ij} (with $i, j = x, y, z$ of $\mathbf{L}(\mathbf{r})$) represents the i -component of the gradient field when a unit gradient in direction j is intended and can be calculated as:

$$L_{ij} = \left(\frac{\partial B_z^j(\mathbf{r})}{\partial i} \right) / g_j, \quad (3)$$

¹ In order to distinguish scalars, vectors and tensors from each other, we write scalars with normal letters, vectors with lower-case bold letters and tensors with upper-case bold letters throughout this article.

where B_z^j denotes the z -component of the magnetic field as a result of applying the gradient field in direction j and g_j denotes the j -component of the applied gradient vector \mathbf{g}_n . By this information about the gradient coil tensor elements, the actual gradient vector at every location can be calculated using Eq. (1).

Due to gradient field nonlinearities, the actual gradient field produced by the coil deviates from the desired gradient field in both magnitude and direction (Bammer et al., 2003). This deviation can be decomposed into its angular (α) and magnitude (β) components, i.e.,

$$\alpha(\mathbf{r}) = \arccos \left(\frac{\mathbf{g}_a^T(\mathbf{r})\mathbf{g}_n(\mathbf{r})}{|\mathbf{g}_a(\mathbf{r})||\mathbf{g}_n(\mathbf{r})|} \right) \quad (4)$$

and

$$\beta(\mathbf{r}) = \frac{|\mathbf{g}_a(\mathbf{r})|}{|\mathbf{g}_n(\mathbf{r})|}, \quad (5)$$

where $|\mathbf{g}|$ denotes the gradient field magnitude and \mathbf{g}^T denotes the vector transpose of \mathbf{g} . Here, $\alpha(\mathbf{r})$ denotes the angular difference between the nominal and actual gradient vectors, and $\beta(\mathbf{r})$ denotes the ratio of the actual gradient amplitudes to the nominal gradient amplitudes at a given location \mathbf{r} . Fig. Supp.1 illustrates a schematic visualization of how actual gradient vectors could deviate from the nominal vectors at an arbitrary point within the imaging coil.

The change in the orientation of gradient vectors can be decomposed into its roll, pitch and yaw components to specify the rotations around main coordinate axes. However, to reduce the degrees of freedom, in this work, we consider the deviation in the vector's orientation in terms of angular difference between the actual and nominal vectors rather than separate rotations around the main coordinate axes. In the next section, we present a generalized approach based on singular value decomposition (SVD) to analyze the effect of gradient nonlinearities.

The most common approach for acquiring diffusion-weighted images (DWI) is to organize gradient vectors uniformly over a (half-)sphere, for which the radius of the sphere indicates the gradient amplitude. However, due to the nonlinearities of the gradient field, the gradient coil tensor \mathbf{L} may deviate from the identity matrix and as a result the arrangement of gradient vectors would deviate from the uniform distribution over a sphere. By performing a SVD of \mathbf{L} , the effect of gradient nonlinearity on the DW gradients can be investigated in a more structured way. Using the SVD it is possible to decompose the matrix \mathbf{L} as:

$$\mathbf{L} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \quad (6)$$

where $\mathbf{U} = [\mathbf{u}_1 \mathbf{u}_2 \mathbf{u}_3]$ and $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \mathbf{v}_3]$ are unitary matrices corresponding to the left and right singular vectors, respectively. \mathbf{S} indicates the diagonal matrix containing the singular values. The singular values indicate the scaling factors along the orientations identified by the corresponding singular vectors. We introduce the mean magnitude deviation (MMD) of the gradient vectors at a given voxel as the mean of singular values, i.e.

$$\text{MMD} = \frac{s_1 + s_2 + s_3}{3}. \quad (7)$$

Here, s_1, s_2 and s_3 are the singular values in descending order, i.e., $s_1 > s_2 > s_3$. Eq. (6) indicates that under transformation by \mathbf{L} , a unit vector \mathbf{v}_1 is mapped to $s_1\mathbf{u}_1$ (Lay et al., 2016). Therefore, the distribution of gradient vectors at every voxel is transformed from the ideal case of uniform distribution over a sphere into a nonuniform distribution over an ellipsoid, whose main axis is aligned with the first left singular vector \mathbf{u}_1 and is scaled by the first singular value s_1 . Hence, analogous to the FA for the diffusion tensor (DT), we define fractional gradient anisotropy (FGA) as:

$$\text{FGA} = \sqrt{\frac{1}{2} \frac{\sqrt{(s_1 - s_2)^2 + (s_2 - s_3)^2 + (s_3 - s_1)^2}}{\sqrt{s_1^2 + s_2^2 + s_3^2}}}, \quad (8)$$

which can now be used to quantify the anisotropy of the gradient coil tensor. Similar to the directionally encoded color (DEC) maps of the first eigenvector (FE) of the DT (Pierpaoli, 1997; Jones et al., 1997), it is also possible to visualize DEC maps of the first left singular vector of the gradient coil tensor weighted by MMD or FGA.

By our proposed measures, MMD and FGA, we have established a method to analyze the effect of gradient nonlinearities on estimates of diffusion measures. MMD indicates the average deviation in the magnitude of the gradient vectors or simply the change in the volume of the ellipsoid, whereas FGA denotes the anisotropy of the resulting gradient vector distribution or simply the shape of the ellipsoid. $MMD > 1$ results in an expansion of the gradient vector distribution, i.e., acquisitions with a larger b-value than the desired one, and $MMD < 1$ results in a shrinkage of the gradient vector distribution, i.e., acquisitions with a smaller b-value than the desired one. A larger FGA, on the other hand, means larger anisotropy of the gradient vector distribution, which can result in non-equal b-values along different gradient directions.

3. Materials and methods

3.1. Simulations

To investigate how the deviations in the DW gradients propagate into errors in the diffusion estimates, we perform Monte-Carlo simulations based on the diffusion tensor model (Basser et al., 1994). Diffusion tensors were simulated from a given combination of FA and MD values and by assuming axial symmetry of diffusion tensor (**D**) in a specified direction given the fact that for unifibre populations with high FA values, symmetric diffusion profile perpendicular to the main diffusion orientation is expected. Axial symmetry of diffusion tensors has also been considered in the previous works (Jones and Basser, 2004; Leemans and Jones, 2009; Jones et al., 1999; Skare et al., 1997). For a single-voxel, we used the following parameter settings, which are in line with the typical tissue characteristics observed in the brain white matter (WM) and configurations used for dMRI acquisitions: $MD = 0.7 \times 10^{-3} \text{ mm}^2/\text{s}$; $FA = 0.7$; $SNR = 10, 20, 30, \infty$ (noiseless) based on the non-DW images; $b = 1000 \text{ s/mm}^2$; number of gradient directions = 60 (distributed uniformly on the unit half-sphere (Jones et al., 1999; Leemans and Jones, 2009)) and the number of non-DW ($b = 0 \text{ s/mm}^2$) images = 6.

The deviation in the DW gradients for a given combination of angular (α) and magnitude (β) components was simulated by rotating and scaling the nominal gradient vector \mathbf{g}_{nk} , $k = 1, 2, \dots, 60$, with respect to a specified direction **n** to calculate the actual gradient vector, \mathbf{g}_{ak} . α is changed between 0° and 5° , and β is changed between 0.9 and 1.1 reflecting a 10% deviation. Diffusion signal was then calculated by using a predefined non-DW signal intensity (S_0) and the actual (correct) gradient vectors (\mathbf{g}_{ak}) to construct the ground-truth data. Rician noise with standard deviation $\sigma = S_0/SNR$, in which SNR denotes the signal-to-noise ratio of the S_0 , was then added to the simulated signal to create the variable SNR data. Diffusion tensor elements were then estimated using the weighted linear least squares (WLLS) approach (Veraart et al., 2013) once by incorporating the nominal (incorrect) DW gradient vectors, and once by incorporating the actual (correct) DW gradient vectors. The error in the estimated diffusion tensor due to gradient nonlinearity can now be investigated by comparing the results from the use of two above-mentioned gradient configurations.

3.2. Synthetic data

3.2.1. DTI phantom

A numerical phantom was constructed from a representative HCP dataset by incorporating the DT elements calculated from all the non-DW and $b = 1000 \text{ s/mm}^2$ acquisitions using the DTI model and the WLLS approach (Veraart et al., 2013). In the next step, 60 gradient orientations were generated with a uniform distribution over a hemi-sphere based on

the electrostatic repulsion model (Jones et al., 1999). The gradient coil tensor from the same HCP dataset was then applied to the generated gradient vectors to calculate the actual gradient directions for each voxel as in Eq. (1). Single-shell diffusion MRI data were then simulated back using the numerical phantom with 6 non-DW ($b = 0 \text{ s/mm}^2$) images and 60 $b = 1000 \text{ s/mm}^2$ volumes along the actual gradient directions for each voxel.

The effect of noise was considered by adding Rician noise with standard deviation $\sigma = S_0/SNR$ and $SNR \in \{10, 20, 30, 50\}$ to the data. For each SNR level, 50 noise realizations were simulated. Estimation of DTI measures, namely MD, FA, axial diffusivity (AD), radial diffusivity (RD) and the direction of the FE of the DT, were performed in ExploreDTI (Leemans et al., 2009) using the WLLS estimation approach (Veraart et al., 2013) from DTI model (Basser et al., 1994), once by applying the GNC, i.e., using the actual voxel-wise DW gradients, and once by leaving out the GNC, i.e., using the nominal DW gradients.

3.2.2. DKI phantom

To investigate the effect of gradient nonlinearities on DKI measures, the simulation framework and the phantom explained in section 3.2.1 was used to generate synthetic multi-shell diffusion data ($b = 1000, 2000$ and 3000 s/mm^2) with 60 gradient directions and 6 non-DW ($b = 0 \text{ s/mm}^2$) images per shell (198 vol in total). The gradient coil tensor (**L**) was used to calculate the actual voxel-wise gradient vectors for each voxel of the simulations using Eq. (1). The variable SNR data were created by adding the Rician noise with standard deviation $\sigma = S_0/SNR$ and $SNR \in \{10, 20, 30, 50\}$ to the simulated data.

Estimation of DKI measures, namely mean kurtosis (MK), kurtosis anisotropy (KA), radial kurtosis (RK) and axial kurtosis (AK), along with the DTI measures (MD, FA, RD, AD) and the FE of the DT, was performed in ExploreDTI (Leemans et al., 2009) using the DKI model (Jensen et al., 2005) and the WLLS estimation method (Veraart et al., 2013) with constraints on the kurtosis tensor as proposed by Tabesh et al. (2011). These constraints ensure that the estimated kurtosis values are physically plausible and acceptable for biological tissues.

To reveal the errors due to deviations in the magnitude and direction of gradient vectors separately, diffusion parameters were estimated for each noise realization once by applying GNC on the direction of gradient vectors (correction for α only), once by applying GNC on the magnitude of gradient vectors (correction for β only), once by applying GNC on both magnitude and direction (correction for both α and β) and once without GNC. Datasets with GNC for both angular and magnitude deviations were considered as the reference data to calculate the errors.

3.2.3. Tractography from synthetic data

In order to investigate the effect of gradient nonlinearities on tractography results, deterministic fiber tractography, based on DTI, from a single seed point was performed in ExploreDTI (Leemans et al., 2009) on the noise-free and variable SNR data both with and without GNC. The only tracking constraints were maximum length = 100 mm, $FA_{\min} = 0.2$ and step size = 1 mm. The reference tract was calculated from the noise-free dataset ($SNR = \infty$) with GNC and was used to calculate the tract error.

3.3. Real data

3.3.1. Data acquisition

The subjects for this study were collected from the HCP dataset (500 subject release) (Van Essen et al., 2012; Glasser et al., 2013; Sotiropoulos et al., 2013), out of which 410 healthy subjects (244 females and 166 males) had their full diffusion data available. Age of the subjects at the time of data acquisition was between 22 and 36 years (mean age for all subjects: 29.1 years, standard deviation: 3.5 years; mean age for females: 29.2 years, standard deviation: 3.4 years; mean age for males: 28.9 years, standard deviation: 3.6 years).

The dMRI data were acquired as part of HCP (Van Essen et al., 2012)

using a 3T Connectome Scanner adapted from a Siemens Skyra (Siemens AG, Erlangen, Germany) with an echo planar imaging (EPI) sequence, isotropic resolution of 1.25 mm and TR/TE = 5520 ms/89.50 ms. The multi-shell dMRI data ($b = 1000, 2000$ and 3000 s/mm^2) were collected with 90 DW gradient directions and 6 non-DW acquisitions per shell (288 volumes in total).

3.3.2. Data analysis and parameter estimation

Correction for subject motion, eddy current induced distortions, B0 inhomogeneity, and imaging gradient nonlinearity distortions was already performed by the HCP team as explained in (Glasser et al., 2013).

Estimation of DKI and DTI parameters from the full multi-shell dataset (3 shells with $b = 1000, 2000$ and 3000 s/mm^2) using the DKI model was performed in ExploreDTI (Leemans et al., 2009) using the WLLS approach (Veraart et al., 2013) with constraints on kurtosis tensor as proposed by Tabesh et al. (2011). DTI parameters were also estimated from the single-shell subset of the data ($90 b = 1000 \text{ s/mm}^2$ and 9 non-DW images) using the DTI model and WLLS (Veraart et al., 2013) approach in ExploreDTI (Leemans et al., 2009).

To investigate the effect of GNC on magnitude and direction deviations separately, the parameter estimations for the multi-shell data was performed once with GNC on angular deviations (correction for α effect), once with GNC on magnitude deviations (correction for β effect), once with GNC on both angular and magnitude deviations (correction for both α and β effects) and once without GNC (by simply using the nominal b-matrix throughout the whole volume). The calculated diffusion metrics were then transformed into MNI152 space using the transform functions provided by the HCP data. The dataset with GNC correction for both angular and magnitude deviations was considered as the reference data to calculate errors.

All the computations were performed using a customized MATLAB (The MathWorks, Natick, MA) script on a Dell (Dell Inc., Round Rock, TX) multi-core parallel processing system with 72 Intel(R) Xeon(R) E7-8870 v3 @2.10Ghz dual cores and access to 1 TB of RAM.

3.3.3. Tractography from real data

DTI-based deterministic tractography was performed on both non-GNC and GNC dMRI data in ExploreDTI (Leemans et al., 2009). A single seed point was used with maximum tract length = 100 mm, $FA_{\min} = 0.2$ and step size = 1 mm during tracking.

The seed point for each subject was determined by transforming the mask of a single voxel seed point in the MNI152 space to each subject's native space using the transformation functions provided by the HCP data. Then, the voxel with the maximum intensity was considered as the seed point for tractography. The reference tract was calculated from the dataset with GNC for both angular and magnitude deviations and was then used to calculate the tract error.

3.3.4. Regional values

The regional values of the DKI and DTI parameters were calculated by taking the mean of the parameter values at the parcellated regions in the brain for each subject as explained in (Mesri et al., 2017b).

Parcellation of brain regions for each subject was performed using the "wmparc" atlas in FreeSurfer (Fischl, 2012) by the HCP team. 179 brain regions common among all the subjects were identified, out of which 165 regions (after excluding CSF, ventricles, etc.) were considered for this study (see supplementary Table Supp.1).

3.4. Evaluation metrics

3.4.1. Percent error

The errors in the diffusion scalars were evaluated by calculating the percent error (η) between the corrected measures (using the actual gradients) and uncorrected (using the nominal gradients), i.e.

$$\eta_X(\mathbf{r}) = \frac{|X_{\text{Uncorr}}(\mathbf{r}) - X_{\text{Corr}}(\mathbf{r})|}{|X_{\text{Corr}}(\mathbf{r})|} \times 100, \quad (9)$$

where X_{Corr} and X_{Uncorr} indicate diffusion measure X , e.g., FA or MD, etc., at location \mathbf{r} with and without GNC, respectively.

3.4.2. Angular error

The error in the direction of the FE was defined as the angle (θ_{FE}) between the FE of DT estimated by using actual (correct) gradients and the FE from the DT estimated by using the nominal (incorrect) gradients, i.e.,

$$\theta_{\text{FE}}(\mathbf{r}) = \arccos\left(\frac{\mathbf{e}_{\text{Uncorr}}^T(\mathbf{r}) \cdot \mathbf{e}_{\text{Corr}}(\mathbf{r})}{|\mathbf{e}_{\text{Uncorr}}(\mathbf{r})||\mathbf{e}_{\text{Corr}}(\mathbf{r})|}\right), \quad (10)$$

where $\mathbf{e}_{\text{Corr}}(\mathbf{r})$ and $\mathbf{e}_{\text{Uncorr}}(\mathbf{r})$ denote the FE of the DT at location \mathbf{r} with and without GNC, respectively.

3.4.3. Tract propagation error

The error made in tract pathway propagation is defined as the point-by-point Euclidean distance between the tract calculated without GNC and the reference tract calculated with GNC, i.e.

$$E_j = |\mathbf{r}_j - \mathbf{R}_j|, \quad (11)$$

where $\mathbf{r}_j = (x_j, y_j, z_j)$ are the coordinates of the j -th point on the non-GNC tract, $\mathbf{R}_j = (X_j, Y_j, Z_j)$ are the coordinates of the j -th point on the GNC tract and $|\mathbf{R}|$ denotes the Euclidean norm of \mathbf{R} . A schematic visualization of the calculation of error metric is depicted in Fig. Supp.2.

3.5. Statistical analyses

All statistical analyses were performed using non-parametric permutation tests in Permutation Analysis of Linear Models (PALM) (Winkler et al., 2014; Nichols and Holmes, 2002; Holmes et al., 1996) version alpha104 with 10^4 permutations. We used nonparametric permutations as it was proved to be more efficient in producing fewer false positives than parametric methods (Eklund et al., 2016). For better precision and visualization, the p-values were computed in logarithmic scale ($-\log_{10}(p)$). Calculation speed was accelerated using tail approximation (Winkler et al., 2016a) in PALM. Correction for multiple comparisons with family wise error-rate (FWER) adjustment was applied and the significance of a test was determined at $p < 0.05$. The effect size measure Cohen's d was also calculated for each test.

3.5.1. Effects on scalar diffusion measures

The effect of GNC on DKI (MK, KA, RK and AK) and DTI (MD, FA, RD and AD) scalars from the DKI model was evaluated with voxel-wise paired t-tests, where every metric in every subject was compared with and without GNC. The test was also applied on the DTI scalars estimated from the DTI model. Threshold-free cluster enhancement (Smith and Nichols, 2009) was applied and the effect size for each metric from the DTI and DKI models was calculated, i.e., $d_{X,\text{DTI}}$ and $d_{X,\text{DKI}}$, where X indicates the diffusion metric of interest, e.g., FA. This setup tested the paired or within-subject differences to check if on average GNC leads to a larger or smaller value.

3.5.2. Angular deviation of the first eigenvector

The angular error between the FEs of the corrected and uncorrected datasets was tested. The voxelwise error maps which were calculated in the native space for each subject were transformed to the MNI space. We used a one-sample t -test to check whether on average the angular error was zero. Independent and symmetric errors were assumed for this test to use sign flipping because the residuals are symmetric around zero. The

tests tend to be conservative if symmetry is violated. Therefore, this assumption is realistic even if the difference is skewed after the mean has been discounted (i.e., residuals). Threshold-free cluster enhancement (Smith and Nichols, 2009) was applied for these tests as well.

3.5.3. Model dependency of GNC

To check if the effect of gradient nonlinearities is model-dependent, we evaluated the effect of neglecting GNC on DTI scalars computed from both DTI and DKI. To this end, first, the difference between uncorrected and corrected measures for every scalar and every subject was calculated for DTI:

$$\Delta X_{DTI} = X_{DTI,corr} - X_{DTI,uncorr} \quad (12)$$

and for DKI:

$$\Delta X_{DKI} = X_{DKI,corr} - X_{DKI,uncorr}, \quad (13)$$

where X denotes a given DTI metric, e.g. FA.

In the next step, voxel-wise paired t -tests of the differences (ΔX_{DTI} and ΔX_{DKI}) across all subjects were performed to compare the model dependency of the errors. Threshold-free cluster enhancement (Smith and Nichols, 2009) was also applied for these tests.

In case of significance, the difference in the effect sizes from the original tests (section 3.5.1), i.e.,

$$\delta_X = d_{X,DKI} - d_{X,DTI} \quad (14)$$

was reported to indicate the model dependency of GNC. Here, $d_{X,DTI}$ and $d_{X,DKI}$ denote the effect sizes of the original tests for the DTI and DKI models, respectively, as described in section 3.5.1.

3.5.4. Effects of GNC on group-wise studies

To check how neglecting GNC can alter the results and conclusions derived from group-wise studies, we used previously reported potential gender differences in terms of diffusion measures (Ritchie et al., 2018; Wierenga et al., 2017; Núñez et al., 2018; Tian et al., 2017; Ingalhalikar et al., 2014; Ruigrok et al., 2014; Kanaan et al., 2012; Herting et al., 2012; Menzler et al., 2011; Hsu et al., 2008; Westerhausen et al., 2003) as an example to evaluate the differences in the significance and effect size with and without applying GNC to the dMRI data.

ROI-based differences between female and male subjects were calculated using unpaired two-sample t -tests of regional values as explained in our previous work (Mesri et al., 2018b). We implemented the same tests for the uncorrected datasets (without GNC) as well, separately. Since volume had been demonstrated to affect diffusion estimates (Vos et al., 2011), the volume of each region per subject was considered as a nuisance co-variate in the design matrix. p -values were corrected for multiple comparisons across ROIs, also by considering multiple contrasts and modalities (Winkler et al., 2016b).

The dependency of gender differences on GNC was determined by subtracting the non-GNC values from the GNC ones in all DTI and DKI scalars for every ROI in each subject, i.e.,

$$\Delta Y = Y_{Corr} - Y_{Uncorr} \quad (15)$$

where Y denotes the regional value of the diffusion parameter of interest for each ROI and subject, separately.

In the next step, an unpaired two-sample t -test on gender was performed on the aforementioned regional differences ΔY . This is equivalent to the interaction part of a repeated or 2-way ANOVA with 2 levels per subject, where the levels represent the cases with and without GNC. A significant difference in a region means that the difference is driven by the GNC for the given scalar. Exchangeable errors along with independent and symmetric errors were assumed, which allow sign-flipping shuffles to increase the statistical power. The errors are considered exchangeable because their joint distribution is invariant with respect to

permutations. Correction for multiple contrasts and modalities was applied.

4. Results

4.1. Magnitude and angular deviations from a HCP subject

Fig. 1 illustrates the spatial pattern and prevalence of the deviations in the magnitude and direction of DW gradients. Note that deviations increase with the distance from the isocenter of the gradient coil system. As a result, voxels towards the edge of the brain typically experience larger deviations compared to the voxels in the center of the brain.

Fig. 2 shows the joint-histogram of the mean angular (α) and magnitude deviations (β) from all the HCP subjects considered in this study. The figure indicates that some combinations of α and β are more probable, which are mainly distributed over arcs with different radii. It can be seen in Fig. 2 that the realistic range for α is between 0° and 5° and for β is between 0.9 and 1.1 ($\pm 10\%$ deviations).

Due to gradient nonlinearities, the distribution of gradient vectors deviates from the desired uniform distribution over a sphere with a given radius (b-value). As an example, Fig. 3 demonstrates how the distribution of the gradient vectors for a typical voxel in Fig. 1 was transformed from the ideally spherical case in Fig. 3(a) to form an ellipsoid in Fig. 3(b).

Fig. 4 shows the MMD-DEC and FGA-DEC maps of the representative subject (see Fig. 1), demonstrating that the principal orientation, MMD and FGA depend on the position of the voxel within the FOV. The deviations are higher in voxels at the outer part of the brain, whereas at the inner part and around the isocenter of the gradient coil system, the deviations are minimal. The first left singular vector (u_1) indicates the orientation along which the gradient vector field is deviated the most. Note that in the frontal areas, the main direction of deviations is along the left-right orientation meaning that the gradient vectors are tilted towards this orientation as illustrated by the red color. Since such a bias depends on the interaction between the deviations and the underlying tissue characteristics, it is difficult to predict estimation errors in diffusion measures and how these will impact the outcome of further analyses.

Fig. Supp.3 shows the mean MMD and FGA maps across all subjects included in this study transformed into MN152 space. The center of the coil represents a smooth region of MMD = 1 and FGA = 0. The smoothness in the FGA and MMD maps arise from the fact that the location of the isocenter of the coil and the spread of the effects can be slightly different for each subject and depend on the head size and the positioning of the subject inside the scanner.

4.2. Simulations

Fig. 5 presents the single-voxel Monte-Carlo simulations. The errors in FA, MD and FE with respect to different values of magnitude and angle deviations are shown for three different SNR levels. Fig. 5(a–c) shows that as the angular deviation (α) increases, the errors in the diffusion metrics increase. However, the error due to α is less pronounced at lower SNR levels. Especially, the errors due to smaller angular deviations are almost swamped by noise and cannot be detected at lower SNR values. For instance, the mean FA percent error due to $\alpha = 1^\circ$ and $\beta = 1$ (no magnitude deviation) was increased by 0.3% from the $\alpha = 0^\circ$ case at SNR = 30, while it was increased by 1.45% in the noise-free case.

Fig. 5(d–f) shows the errors in the diffusion metrics as a function of magnitude deviation (β), with no angular deviation ($\alpha = 0$). As it is shown in Fig. 5(d,f), the effect of magnitude deviation on FA and FE in the noise-free case is zero and only the bias and uncertainty due to noise increase as the SNR decreases. On the other hand, Fig. 5(e) shows that the effect of magnitude deviation on MD is scaled linearly with β for the noise-free case. The bias for the variable SNR simulations without angular and magnitude deviations ($\alpha = 0$ and $\beta = 1$) is only due to noise. The effect of magnitude deviations is less pronounced at lower SNR levels.

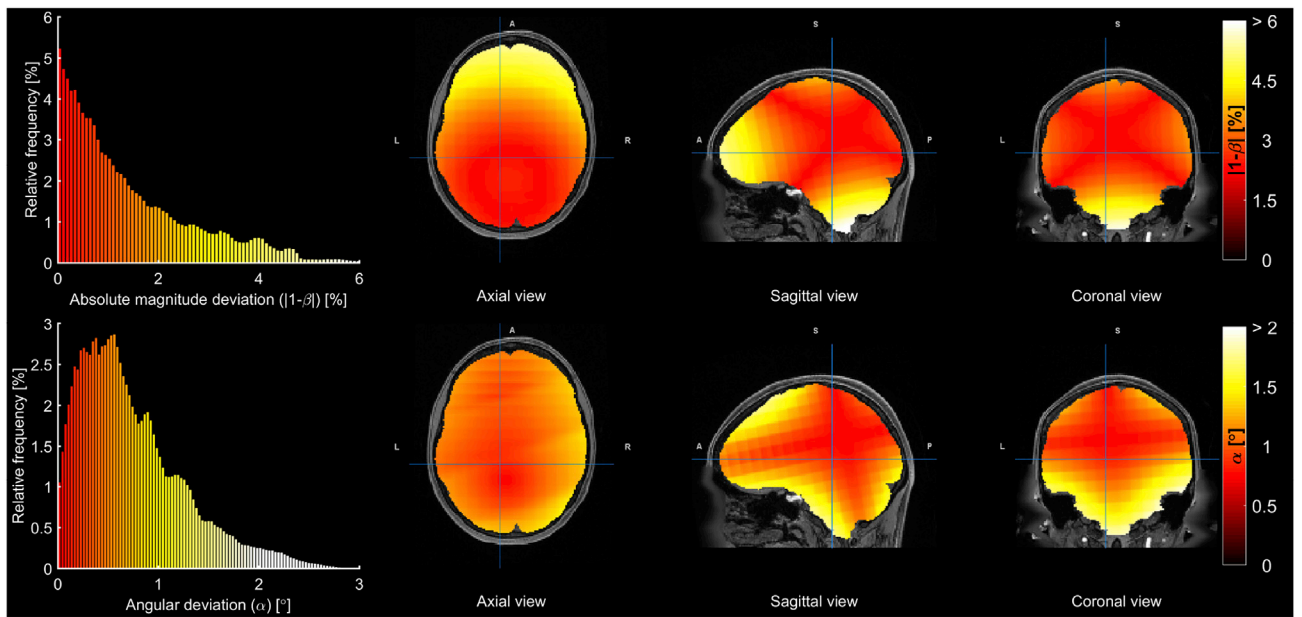


Fig. 1. Histograms and spatial distribution of deviations in the magnetic field gradients for a representative subject from the HCP. Upper row: Percentage magnitude deviation ($|1 - \beta|$); Lower row: Angular deviation (α).

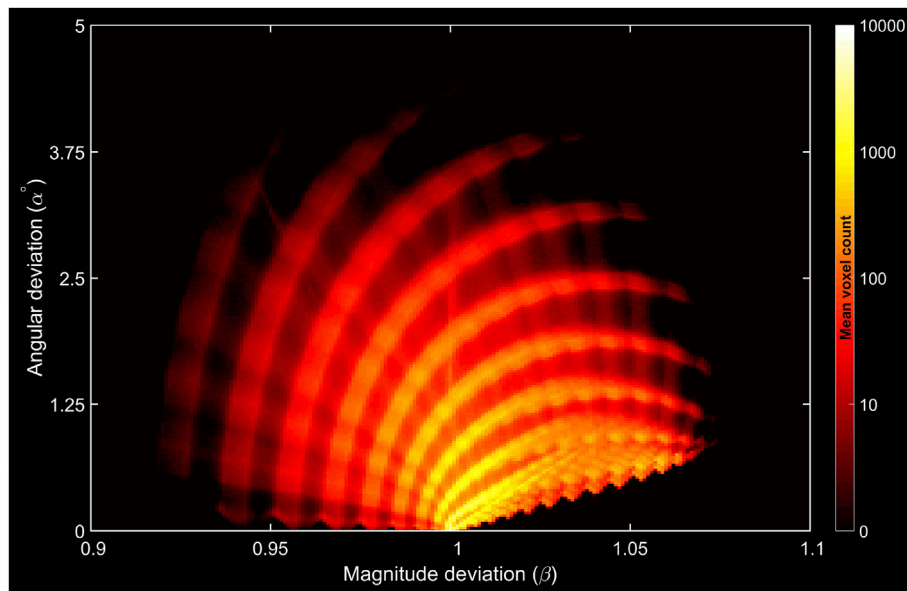


Fig. 2. Joint histogram of the mean angular (α) and magnitude (β) deviations calculated from all HCP subjects considered in this study.

4.3. Synthetic data

Fig. 6 shows the mean percent errors for the DKI and DTI metrics from phantom simulations with DKI at SNR = 30 when no GNC was applied. It can be inferred from the results that (i) the spatial pattern of the gradient nonuniformities translates into errors in the diffusivity metrics, (ii) the errors in the other metrics (FA and DKI metrics) show a stronger dependency to the underlying tissue parameters, and (iii) the average percent errors for DTI metrics are larger in comparison to the DKI metrics. The maximum percent errors for the DKI metrics are in the order of 3%, while the maximum percent errors for the DTI metrics are in the order of 10%.

Figs. Supp.4(a) and Supp.4(b) show the residual errors in the cases where the GNC is applied only on the direction (α -correction) or on the magnitude (β -correction) of the gradient vectors, respectively. It can be

inferred from the figures that α -correction has a very limited impact on the residual errors as the errors are almost identical to the case in which no GNC was applied. On the other hand, β -correction results in diminished residual errors compared to the case with no GNC correction in Fig. 6.

Fig. Supp.5 shows the mean percent errors for DTI metrics calculated from synthetic data with DTI at SNR = 30, which indicates that the spatial pattern of errors follows the spatial pattern of the gradient nonuniformities. This is especially visible for the diffusivity metrics (MD, RD and AD), where the pattern of errors is similar to the pattern of magnitude deviation in Fig. 1 (upper row). This is expected as it was shown earlier that the errors in diffusivity metrics scale linearly with respect to magnitude deviations. The errors in FA indicate a stronger dependency on the tissue parameters compared to the diffusivity metrics.

Comparing the errors of the DTI metrics from the DTI model in Fig. Supp.5 to the errors of the DTI metrics from the DKI model in the

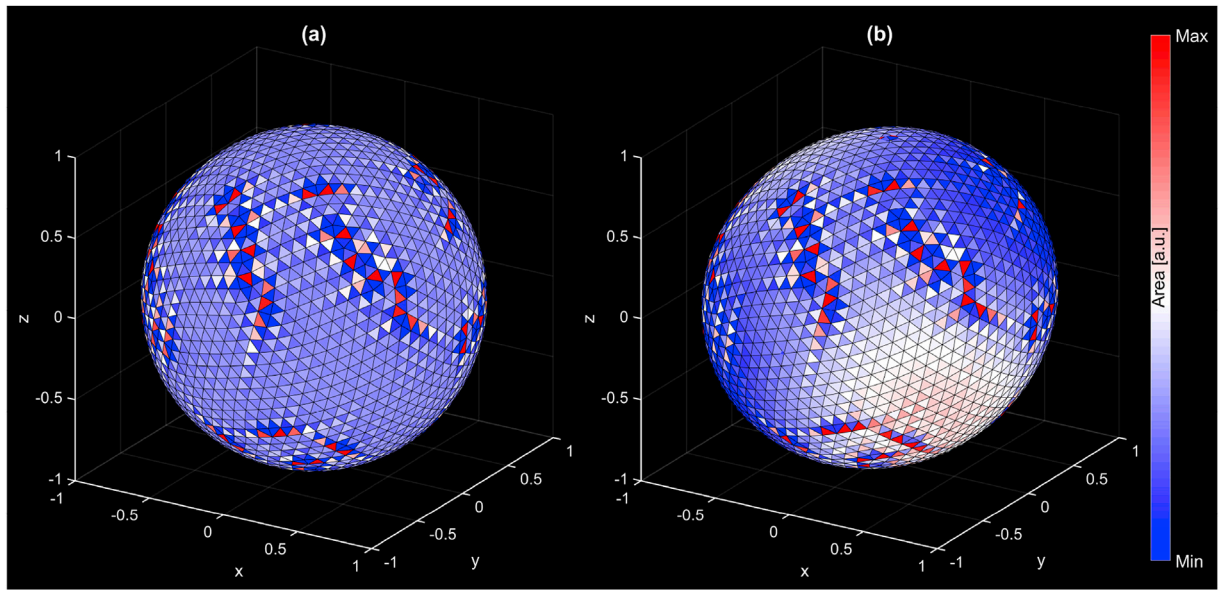


Fig. 3. (a) Nominal and (b) actual distribution of diffusion gradient vectors at an arbitrary voxel from a representative HCP dataset. Due to gradient field nonlinearity, the actual distribution of gradient vectors deviates from the ideal case of uniform distribution over a sphere. This can be identified by the change in the shape of the distribution from a sphere in (a) to an ellipsoid in (b) and the consequent change in the area of surface triangles.

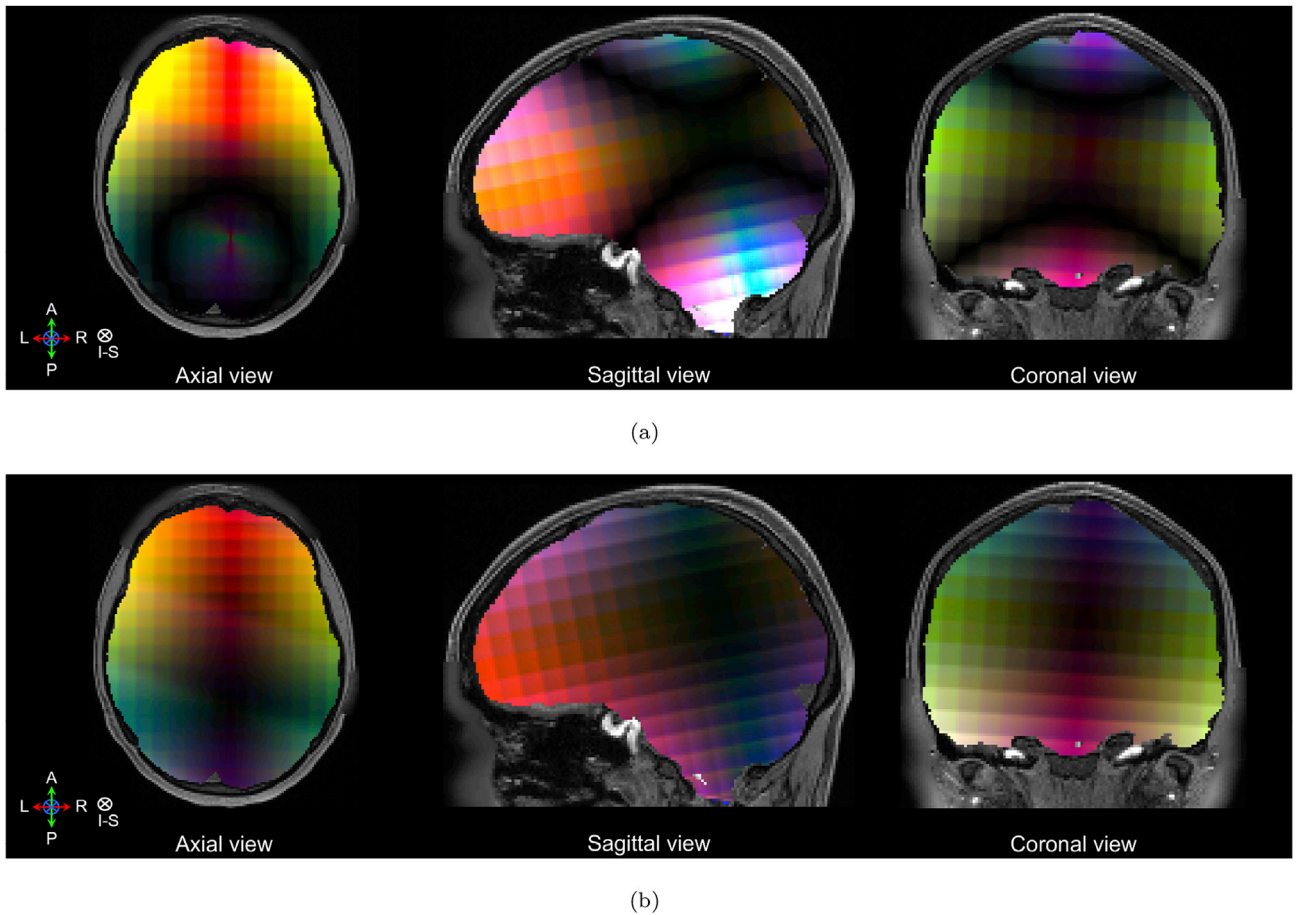


Fig. 4. (a) MMD-DEC and (b) FGA-DEC maps from a representative HCP subject. MMD values are between 0.923 and 1.049 and FGA Values are between 0 and 0.047. The color encoding indicates the conventional DEC map color encoding, i.e. red: left-right, green: anterior-posterior, and blue: up-down.

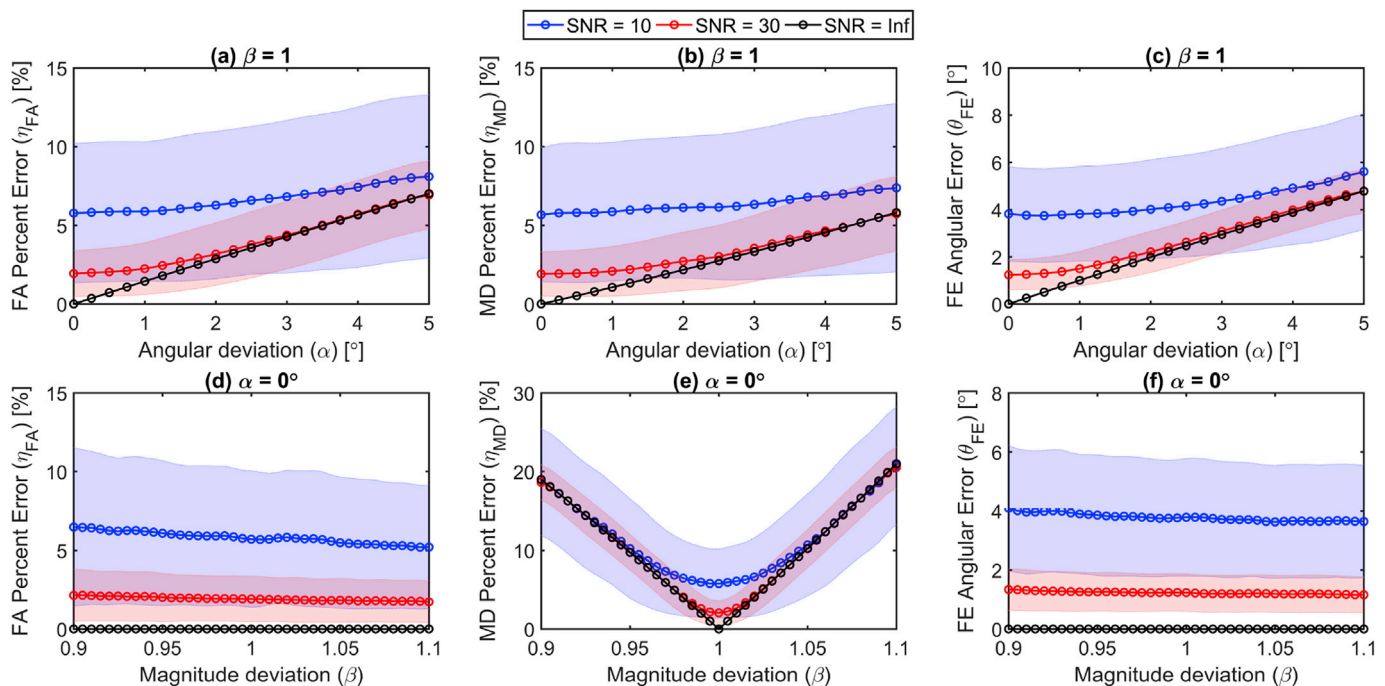


Fig. 5. Results of the single-voxel simulations at three different SNR levels. (a–c) FA percent error (η_{FA}), MD percent error (η_{MD}) and FE angular error (θ_{FE}) as a function of α at $\beta = 1$. The errors increase as the angular deviation increases. The effect of α is swamped by noise at lower SNR levels. (d–f) η_{FA} , η_{MD} and θ_{FE} as a function of β at $\alpha = 0^\circ$. The error in MD linearly scales with β for the noise-free case. At lower SNR levels, the effect of small β is swamped by noise and only the noise bias itself is visible. The effect of β on FA and FE is zero.

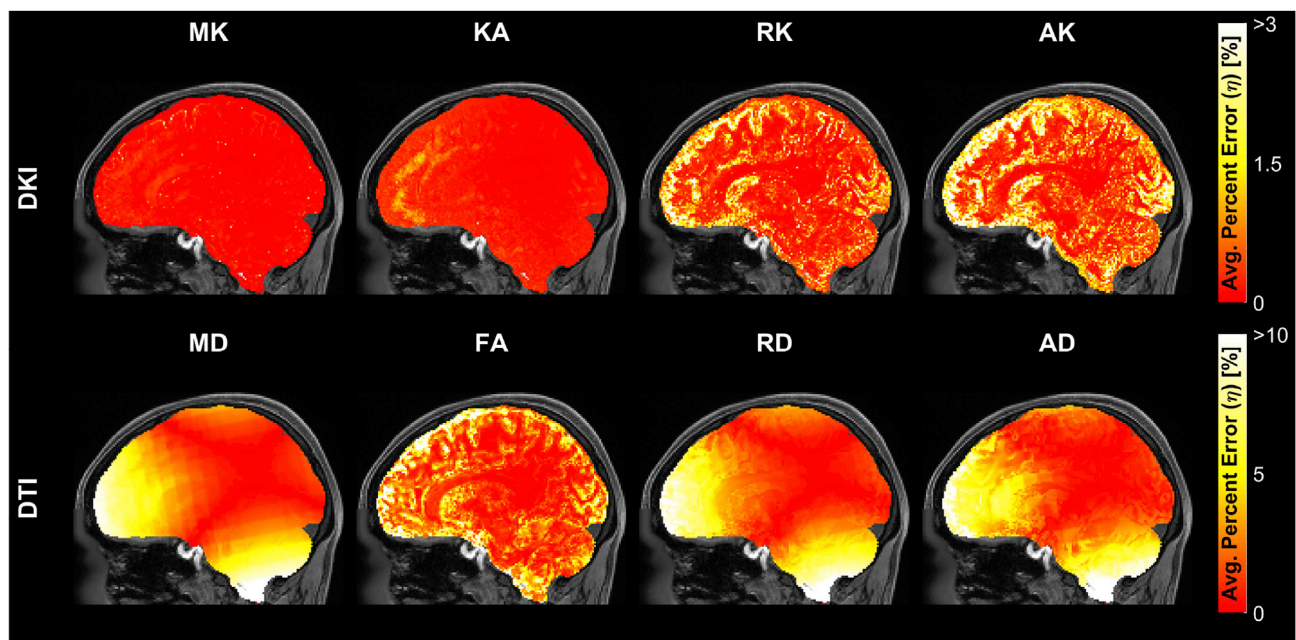


Fig. 6. Mean percent errors of DKI (upper row) and DTI (lower row) metrics from a simulation phantom consistent with the HCP dataset using the DKI model at SNR = 30, without applying GNC. The errors follow the pattern of the gradient deviations and are larger for voxels further from the isocenter of the coil. The mean percent error is larger for DTI metrics (MD, FA, RD, and AD) compared to DKI metrics (MK, KA, RK and AK). Note the different color scales for the upper and lower rows.

bottom row of Fig. 6, it can be inferred that the errors are in the same range and follow a similar pattern regardless of the model (DTI or DKI) used for their calculation.

4.4. Real data

Fig. 7 shows the average percent errors for the DTI and DKI metrics

calculated using the DKI model across all the subjects in MNI space. Fig. Supp.6 presents the equivalent results for DTI metrics computed with the DTI model. By comparing both results, one can see that the errors in the DTI metrics calculated using the DTI model are in the same order as the errors in the DTI metrics calculated using the DKI model. The results in Fig. 7 indicate the same trends as for the synthetic data, where the errors in diffusivity measures follow the nonlinear gradient field

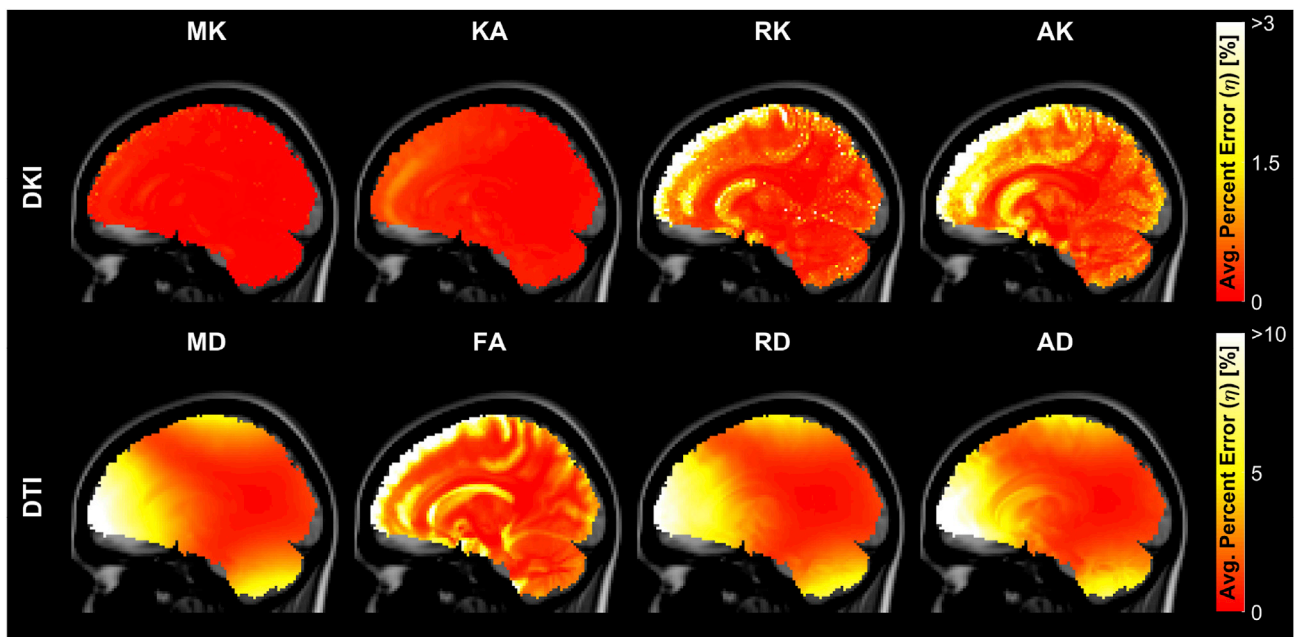


Fig. 7. Mean percent errors of diffusion metrics derived from the DKI model for real datasets transformed to the MNI152 space without applying GNC. The average percent errors for DTI metrics (MD, FA, RD, and AD) calculated from the DKI model are in the same level as for the errors in DTI metrics calculated from the DTI model. The average percent errors are larger for DTI metrics compared to DKI metrics (MK, KA, RK and AK). Note the different color scales for the upper and lower rows.

pattern.

Figs. Supp.7(a) and Supp.7(b) depict the residual errors when the gradient vectors are corrected only for angular (α) and magnitude (β) deviations, respectively. Comparing the results to the case with no GNC in Fig. 7, it can be inferred that the residual errors when only α -correction is applied are in the same range as the errors when no GNC was applied. On the other hand, the residual errors in the case of β -correction are drastically decreased. Especially, for the DTI measures, the residual errors are nearly eliminated with β -correction.

4.5. Statistical analyses with HCP datasets

4.5.1. Scalar diffusion measures

Fig. 8 shows the effect sizes (Cohen's d) for the voxels with a significant difference in diffusion scalars computed with the DKI model between the HCP datasets processed with and without GNC. The positive and negative numbers (red and blue colors, respectively) indicate the increase and decrease in the metrics as a result of applying GNC. Fig. 8 indicates that the effect sizes for the DTI metrics (lower row) are larger

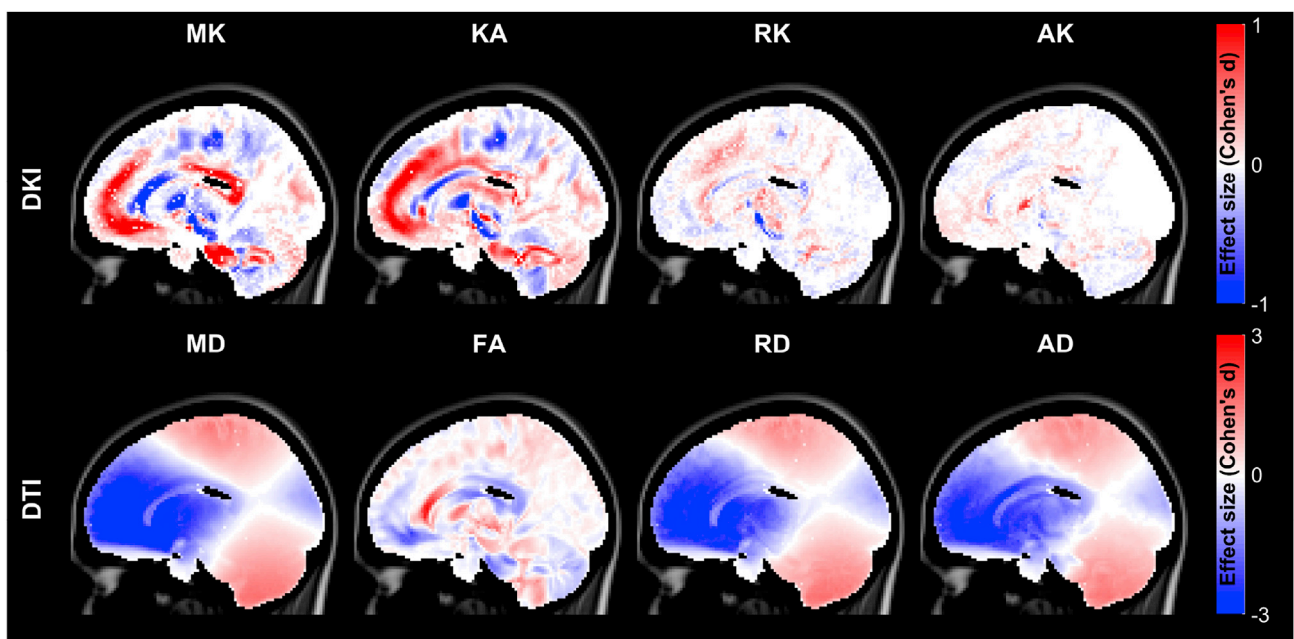


Fig. 8. Results of the paired t-tests for the diffusion metrics calculated from the DKI model for real datasets transformed to the MNI152 space. The effect sizes for the DTI metrics are larger compared to the DKI metrics. Note the different color scales for the upper and lower rows. The patterns of the significant voxels for diffusivity metrics (MD, RD, AD) follow the pattern of gradient nonuniformities. Red and blue colors indicate increase and decrease in the metrics, respectively, due to gradient nonlinearity correction.

than the effect sizes for the DKI metrics (upper row). The effect sizes for the DKI metrics as well as FA show more tissue dependency compared to the effect sizes for the diffusivity metrics, which mainly follow the pattern of gradient field nonuniformities. Larger effect sizes (in the order of 3) are observed for the diffusivity measures in voxels further from the isocenter of the gradient coil system.

Fig. Supp.8 shows the effect sizes for the voxels with a significant difference in DTI scalars computed from the DTI model with and without GNC, indicating similar patterns to those of the DTI scalars computed from the DKI model in Fig. 8.

4.5.2. Diffusion orientation

The results of the single one-sample *t*-test of the FE angular error is presented in Fig. 9. The FE angular errors in WM regions with high anisotropic diffusion such as the corpus callosum, corticospinal tract and optic radiata have larger mean and smaller variance, which results in relatively larger effect sizes compared to the regions with more isotropic diffusion such as the GM.

4.5.3. Model dependency

The change in the effect sizes for the voxels with a significant difference between DKI and DTI models is presented in Fig. 10(a). Four cases can be expected here as shown in Fig. 10(b). The GNC can result in an increase in the estimated parameter (cases 2 and 4) or a decrease (cases 1 and 3). On the other hand, the effect size with the DKI model can be larger (cases 1 and 2) or smaller (cases 3 and 4) than the ones with DTI. The other possible combinations were not observed in the results. In Fig. 10(a), a larger effect size with DKI and DTI are shown by red and blue color scales, respectively. The results demonstrate differences in the effect sizes in the order of 0.3 which can be considered relatively small compared to the effect sizes observed in the original tests. It can also be understood from the results that the trend of change in the effect sizes for FA is opposite to the trend of change for the diffusivity measures.

4.5.4. ROI based group-wise analysis

The change in the effect sizes of group-wise *t*-tests for gender difference as a result of applying GNC is presented in Fig. 11. Only regions with a significant dependency of gender difference on GNC are highlighted. Six scenarios can be considered here:

1. Significance with and without GNC with an increase in the effect size after applying GNC (highlighted by red);
2. Significance with and without GNC, but a decrease in the effect size after applying GNC (highlighted by blue);
3. Gaining significance after applying GNC with an increase in the effect size (highlighted by green);

4. Losing significance after applying GNC with a decrease in the effect size (highlighted by magenta);
5. No significance with and without GNC but an increase in the effect size after applying GNC (highlighted by yellow); and
6. No significance with and without GNC but a decrease in the effect size after applying GNC (highlighted by cyan).

4.6. Tractography based analysis

Tractography results from the synthetic data at two different SNR levels, SNR = 30 (with 50 noise realizations) and SNR = ∞ , are shown in Fig. 12(a) and 12(b), respectively. The tracts from datasets with and without GNC are depicted by green and red streamlines, respectively. Deviations within green and/or red tracts are only due to noise, while deviations between red and green tracts are due to the GNC.

Fig. 13 shows the tract error ranges (25 percentile, median and 75 percentile) from the synthetic data with and without GNC at different SNR levels. The tract from the noise-free dataset with GNC is considered as the ground-truth tract for both cases. The results indicate that (i) the tract error increases as the distance from the seed point increases, which is also known as accumulation of error; (ii) the uncertainty of tracts increases as the SNR level decreases, this is more visible for the datasets with GNC; and (iii) as the SNR increases, the tracts converge to the ground-truth tract for the GNC datasets, while for the datasets without GNC, the tracts converge to a false tract at SNR = ∞ , which indicates a systematic bias in the tractography results when the GNC is ignored.

Fig. 14(a) depicts the tracts from a representative subject with and without GNC. It can be seen in the figure that the tract from the dataset without GNC deviates from the tract with GNC as it propagates from the seed point. Fig. 14(b) demonstrates the range (25 percent, median and 75 percent intervals) of the tract error across all the subjects considered in this study. The tract error indicates that as the distance from the seed point increases, the bias and uncertainty in the tracts increase.

5. Discussion

In this work, we presented a systematic analysis of the effect of DW gradient field nonuniformities on the diffusion metrics, tractography results and group wise studies. Our results from simulations and real data from the HCP demonstrated that nonlinearities of the gradient fields translate into the errors in the diffusion measures. Results from the statistical analyses indicate that the errors in diffusion measures are significant and may result in false interpretations. To showcase this, we investigated the effect of ignoring GNC on the results and conclusions from an example group study. Our results showed that ignoring GNC can lead to alterations in the scientific outcomes and conclusions from group

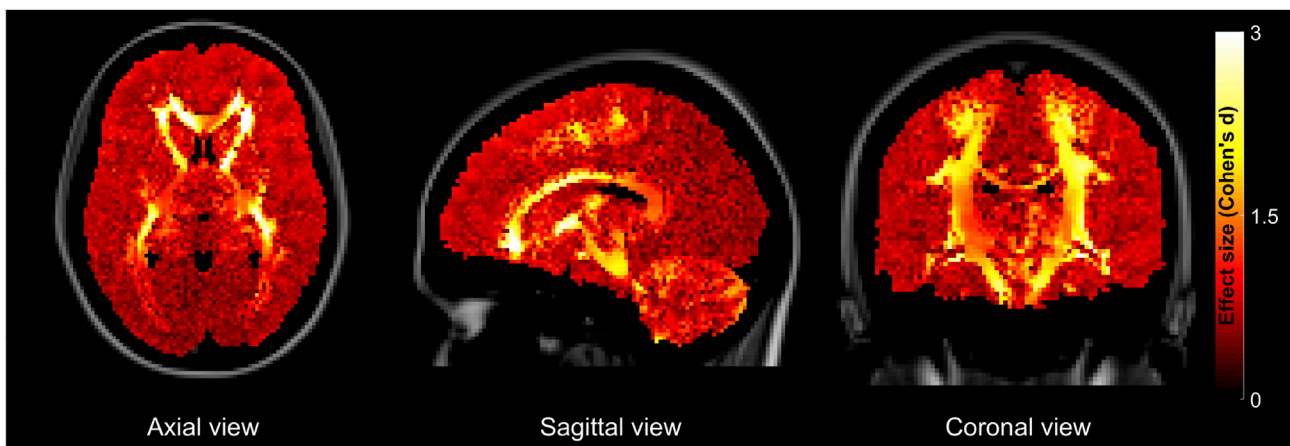
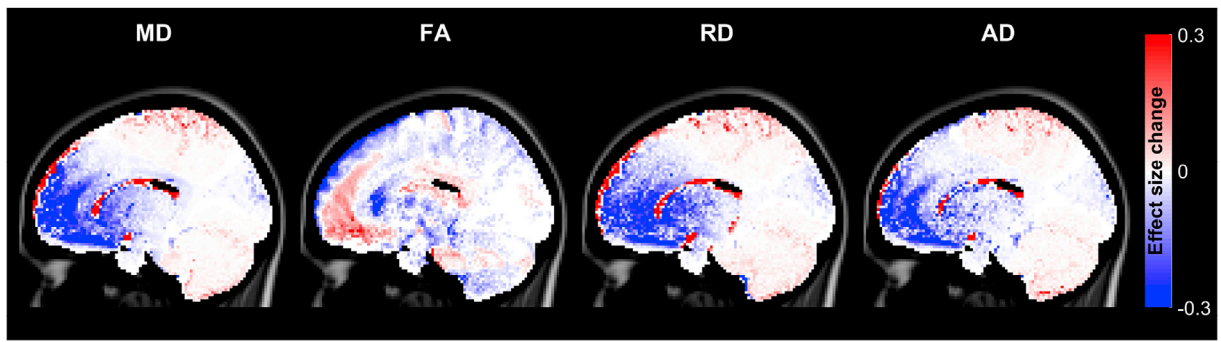
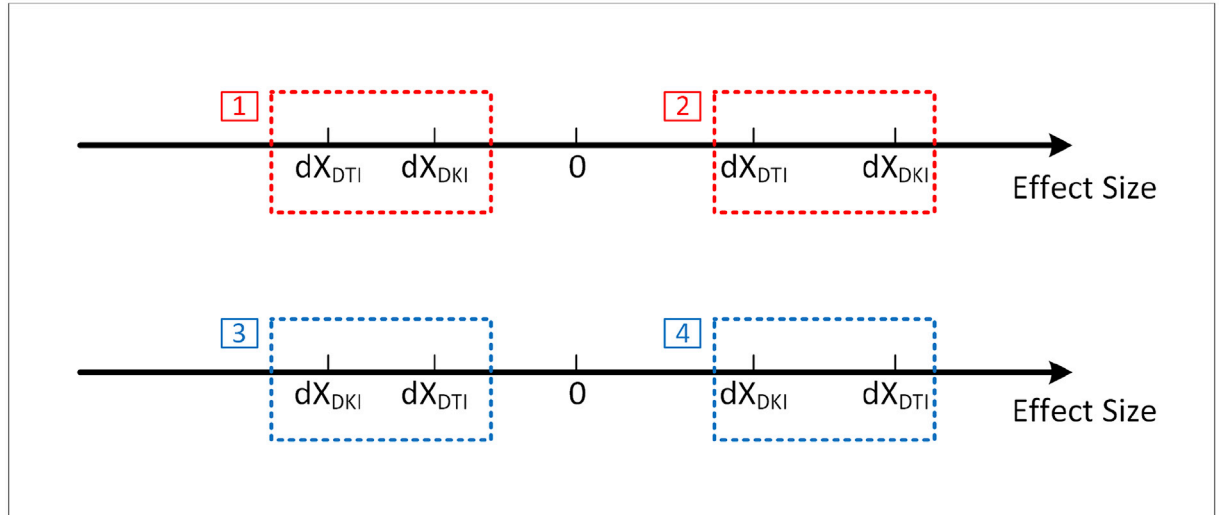


Fig. 9. Effect sizes of the one sample *t*-test of the FE angular error from real datasets transformed to the MNI152 space. The effect sizes in white matter regions are relatively larger due to higher mean and lower variance.



(a)



(b)

Fig. 10. (a) Differences in the effect sizes of the DTI metrics calculated from DKI and DTI. Red and blue color scales indicate larger or smaller effect sizes for parameters estimated from the DKI compared to the ones estimated from the DTI. (b) 4 possible combinations of effect sizes from the original tests with the DKI and DTI models.

based studies. Ignoring GNC was also shown to have a negative impact on the results from tractography based studies.

5.1. Single-voxel simulations

Simulation of gradient nonlinearity effects on the diffusion metrics was performed at the voxel level by assuming different combinations of angular and magnitude deviations at a number of noise levels. The results indicated that the effect of gradient nonlinearities is more pronounced at higher SNR levels, especially, for the smaller values of α and β . The effect of β on the FA and FE was shown to be zero, while its effect on the diffusivity metrics (MD, RD and AD) was shown to linearly scale with β . On the other hand, the effect of α was shown as a relevant factor on the errors for all metrics.

Our results imply that ignoring the effect of gradient imperfections is more problematic for high SNR acquisition systems and high quality data. Moreover, the deviation ranges used in this study for simulations were inline with the ranges observed in the HCP dataset (see Fig. 2), which are in general larger than those for the typical scanners. The gradient nonlinearities may be ignored for typical datasets because the effect of these imperfections are swamped by noise. However, as the acquisition strategies and hardware improve and high quality data become more accessible (e.g. HCP), correcting for these effects becomes more significant to avoid the confounding bias in scientific findings.

5.2. Whole-brain simulations and real data

The results from the whole-brain simulations using a phantom consistent with the HCP data demonstrated that the nonlinearities in the gradient magnetic fields translate into errors in the estimated diffusion measures. It was shown in Fig. 6 that the errors in diffusivity measures follow the pattern of field nonlinearities and increase as the distance of the voxel from the isocenter of the gradient system increases. However, errors in the FA were shown to have a tissue dependency as well, with higher errors in the regions with lower FA values. This can complicate data interpretation when using these metrics in the presence of pathology, since the change in the parameter value cannot be merely identified as a pathology due to the concomitant gradient nonlinearities.

The results of synthetic data with the DKI model in Fig. 6 indicated that, in general, DKI parameters have smaller errors compared to DTI parameters. This is due to the fact that the DKI is a higher order model with more degrees of freedom and is more prone to noise. Therefore, the effect of gradient nonlinearity is swamped by the effect of noise and is not detected. It was also shown that the errors in DKI metrics have a tissue dependency, which can also complicate the interpretation of results when no correction is applied. The results from the synthetic data with DTI model in Fig. Supp.5 indicated that the percent errors for the DTI measures computed using DTI model is similar to the errors for the DTI measures from synthetic data with DKI model in Fig. 6.

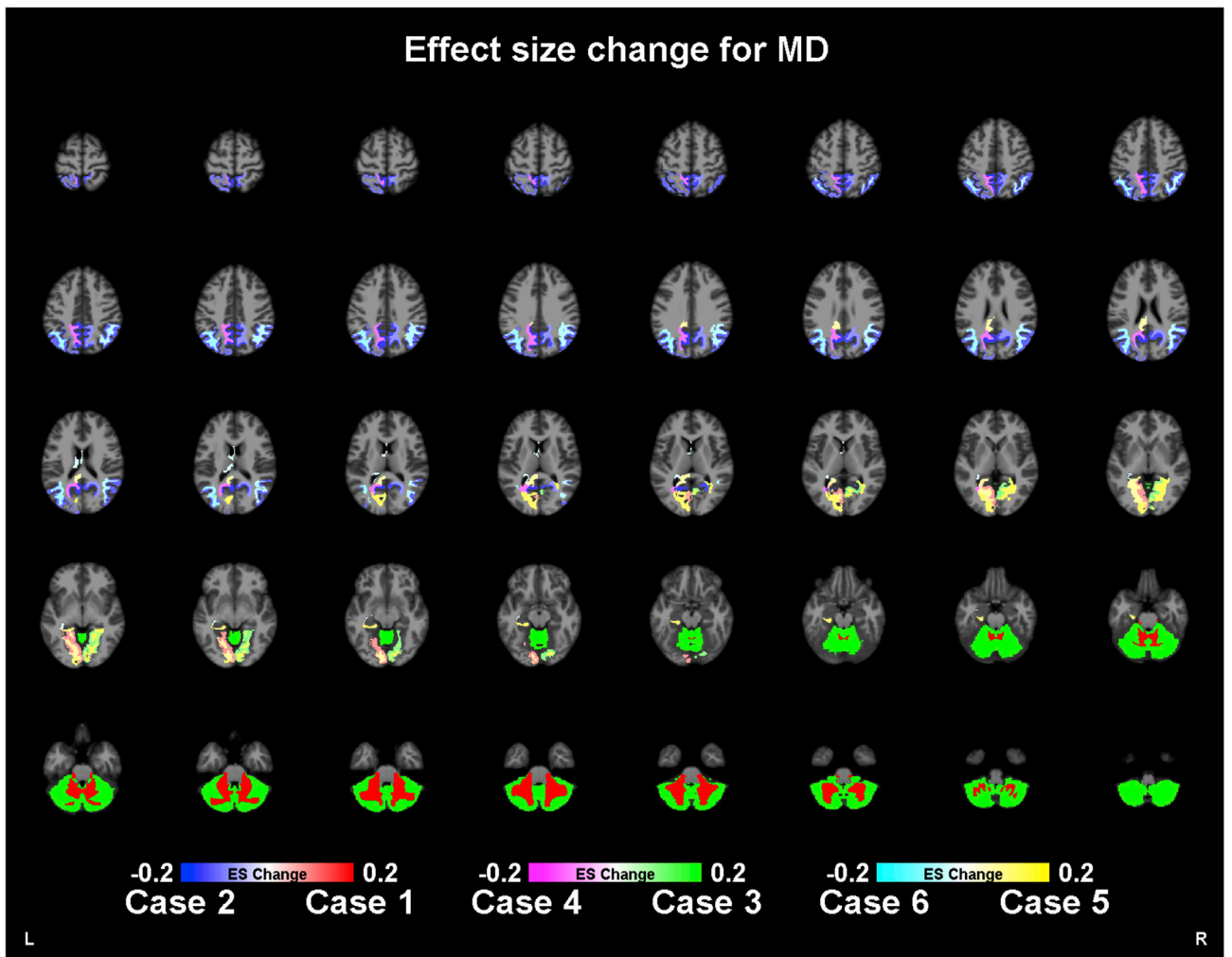


Fig. 11. Changes in the effect sizes of the group-wise gender comparison after applying GNC. Six cases are observed here: 1. Red: significant before and after applying GNC with an increase in the effect size; 2. Blue: significant before and after applying GNC with a decrease in the effect size; 3. Green: gaining significance after applying GNC; 4. Magenta: losing significance after applying GNC; 5. Yellow: no significance before and after applying GNC but an increase in the effect size; and 6. Cyan: no significance before and after applying GNC but a decrease in the effect size.

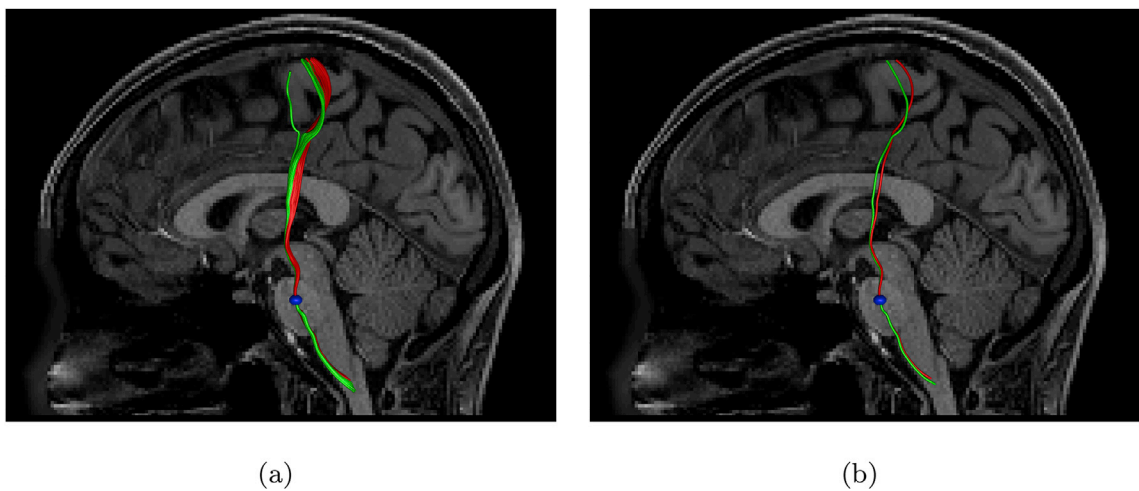


Fig. 12. Tractography results from synthetic data with (a) $\text{SNR} = 30$ or (b) $\text{SNR} = \infty$ (i.e. noise-free) with GNC (green) and without GNC (red).

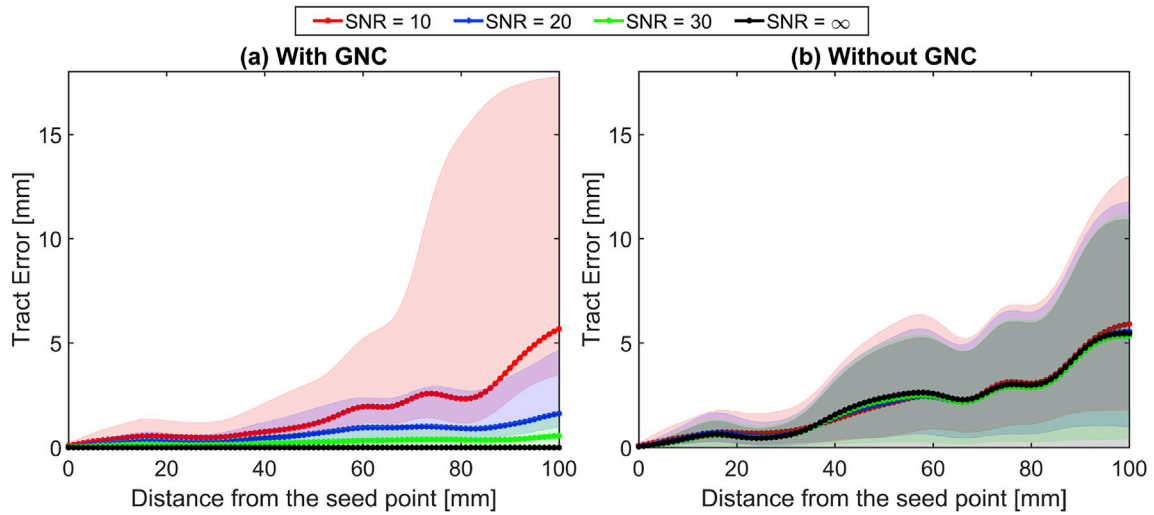


Fig. 13. Tract error ranges from the synthetic datasets, (a) with GNC and (b) without GNC. Tract errors and uncertainties increase as the distance from the seed point and noise level increase. As the SNR increases, the tracts from datasets with GNC converge to the ground-truth tract, while the tracts from datasets without GNC converge to a false tract indicating a systematic bias in the tractography results for datasets without GNC.

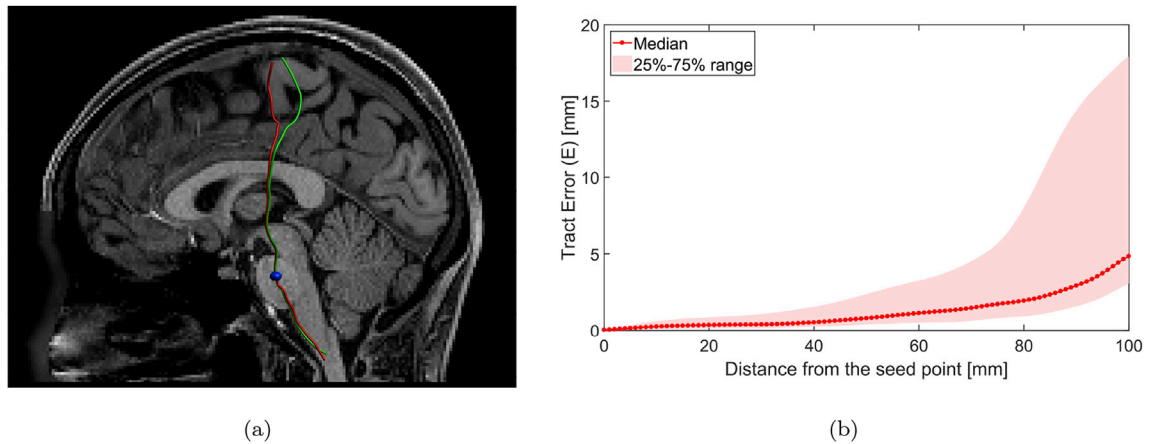


Fig. 14. Results of tractography from real HCP datasets. (a) Tracts from a representative subject with GNC (green) and without GNC (red). (b) Range of the tract error for all subjects in the study. The further the distance from the seed point, the higher the error and uncertainty.

The effect of applying GNC on α and β components separately were shown in Figs. Supp.4 and Supp.7 for synthetic and real data, respectively. It could be concluded from the results that considering the typical range of angular and magnitude deviations in the HCP data, applying GNC on α had a small correction effect while applying GNC on β resulted in a diminished residual errors, especially, for the DTI measures. It can be understood from these results that given the range of α and β components in the HCP datasets, the impact of β -correction is larger than the impact of α -correction.

The percent errors from the real data with DTI and DKI models presented in Figs. Supp.6 and 7 are in line with their counterparts from synthetic data and acknowledge the conclusions drawn previously. In general, the pattern of errors in diffusivity measures follow the pattern of gradient nonlinearities and DTI metrics have larger errors compared to the DKI metrics. However, for the real data, a higher variability in the results is expected due to different head sizes and different positioning of patients in the scanner.

5.3. Statistical tests and group-wise analysis

The results from the voxel-wise paired t-tests of the difference in the diffusion metrics between real data with and without GNC were

presented in Fig. 8 and Supp.8 for the DKI and DTI models, respectively. In general, we observe that (i) similar to the percent errors, the pattern of effect sizes for diffusivity metrics follow the pattern of gradient field nonuniformities; (ii) the effect sizes for DTI measures are larger than those for the DKI metrics; and (iii) the effect sizes for FA and all the DKI metrics show tissue dependency.

The one-sample t-tests of the errors in the FE also indicate effect sizes in the order of 3, which can be considered large. The effect sizes also show tissue dependency, which is a result of the interaction of the underlying tissue characteristics (e.g. MD, FA and the principal orientation of the diffusion tensor) with the gradient field inhomogeneity characteristics (e.g., MMD, FGA, and the principal orientation of the gradient coil tensor) at an arbitrary voxel. Additionally, placement of patients inside the scanner, patient's head size, as well as the size of gradient coil system may also differ for different acquisitions, patients and scanners. This implies that ignoring GNC will result in errors which depend on a large number of parameters under different scenarios. Hence, it would be complicated to disentangle their effects after calculations are complete. It is therefore important to correct for these effects in the early stages of processing and parameter estimation.

The sensitivity of the diffusion models, namely DTI and DKI, to the gradient nonlinearity artifacts was examined by comparing the

significance and effect sizes for the errors in the DTI metrics calculated from both models. The results from the paired t-tests for these models in Fig. 8 and Fig. Supp.8 indicate that the effect sizes for errors in DTI metrics from both models are in a similar range. The difference in the effect sizes as a result of using different models, presented in Fig. 10, indicates that the changes in effect sizes are relatively small with respect to the ranges from the original tests. Hence, we can conclude that DTI and DKI models are prone to the errors caused by gradient nonlinearity effects to approximately the same extent and that the errors can not be minimized by simply choosing a different diffusion model.

To showcase the effect of gradient field nonlinearities on the results from group-wise studies, we considered the already reported group-wise differences in the regional values of diffusion measures between females and males (Ritchie et al., 2018; Wierenga et al., 2017; Núñez et al., 2018; Tyan et al., 2017; Ingallhalikar et al., 2014; Ruigrok et al., 2014; Kanaan et al., 2012; Herting et al., 2012; Menzler et al., 2011; Hsu et al., 2008; Westerhausen et al., 2003) as an example. The motive was to understand how much ignoring GNC can alter the results and conclusions from a group study. To this end, we compared the results of the study with and without GNC to check whether the differences are significant, and if so, to what extent the effect sizes change. It was shown that applying GNC could alter the significance of the test for regions, or change the effect sizes without changing the significance. Six different cases were identified for all possible scenarios (see Fig. 11 and table Supp.1). In general, our results indicate that ignoring GNC could alter the significance and effect size of the statistical test, especially for regions falling in the areas where the gradient nonlinearity effects are stronger. For example, GM regions show more changes in the effect size since they are mostly located further from the isocenter of the gradient coil, where the gradient nonlinearities are larger. However, these changes also depend on the effect sizes from the original tests, and as a result, significant changes occur in the WM regions as well. This implies that the interpretations and conclusions drawn from group-wise studies may be subject to errors due to ignoring GNC.

5.4. Tractography

The effect of ignoring GNC on tractography results was investigated using tractography with synthetic data as well as the real data from the HCP. We utilized the tract error, i.e., the mean geometrical distance between the reconstructed tract and the ground-truth tract, as a tool to evaluate the errors in the tract pathways. The results of tractography from a single seed point with synthetic data presented in Figs. 12 and 13 indicated that the tractography results can be affected by the gradient nonlinearities. It was shown that in the case of non-GNC, as the SNR increases, tract pathway reconstructions converge to a false tract pathway representing a systematic bias. The errors accumulate as the tract propagates further away from the seed point. On the other hand, the uncertainties due to the noise in the tractography results without applying GNC is suppressed. This can be explained by considering that the systematic bias changes the trajectories in such a way that the variance is also affected. For instance, the tracts might be changed to a direction for which less deviations can occur. The error pattern also depends on the areas that the tract passes through, the underlying tissue characteristics, as well as the noise distribution. The results of the tractography from the real data in Fig. 14 were in line with the results from the simulations showcasing the accumulation of the tract error as the fiber trajectories propagate from the seed point. Our results imply that these artifacts can result in a considerable bias in the tractography results, which may affect data interpretation.

6. Conclusions and future work

The results of this study implicate that the effect of gradient field nonlinearities on the diffusion MRI studies can be significant and may result in false outcomes and conclusions. These artifacts are different for

different subjects and different acquisitions due to for example, the differences in the subjects' head size or different placement of the subjects in the scanner. Therefore, GNC must be applied for each subject and each acquisition separately. We conclude from this study that correction for gradient nonlinearity artifacts should not be ignored especially for high quality datasets and where the correction techniques and necessary gradient coil tensor data are provided by the scanner or the cohort as for the HCP.

In this work, the effect of ignoring GNC was investigated in the absence of other artifacts and confounding factors, e.g. motion, intensity inhomogeneity, etc. Hence, a reasonable follow-up study would be to investigate the interplay of the artifacts and the effect of gradient nonlinearities in presence of other confounding factors or pathology.

Acknowledgments

The research is supported by VIDI Grant 639.072.411 from the Netherlands Organization for Scientific Research (NWO).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.116127>.

References

- Bammer, R., Markl, M., Barnett, A., Acar, B., Alley, M.T., Pelc, N.J., Glover, G.H., Moseley, M.E., 2003. Analysis and generalized correction of the effect of spatial gradient field distortions in diffusion-weighted imaging. *Magn. Reson. Med.* 50, 560–569.
- Basser, P.J., Mattiello, J., LeBihan, D., 1994. MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267.
- Borkowski, K., Kłodowski, K., Figiel, H., Krzyżak, A.T., 2017. A theoretical validation of the B-matrix spatial distribution approach to diffusion tensor imaging. *Magn. Reson. Imag.* 36, 01–06.
- Conturo, T.E., McKinstry, R.C., Aronovitz, J.A., Neil, J.J., 1996. Diffusion MRI: precision, accuracy and flow effects. *NMR Biomed.* 8, 307–332.
- Doran, S.J., Charles-Edwards, L., Reinsberg, S.A., Leach, M.O., 2005. A complete distortion correction for MR images: I. Gradient warp correction. *Phys. Med. Biol.* 50, 1343–1361.
- Dyrby, T.B., Søgaard, L.V., Hall, M.G., Pito, M., Alexander, D.C., 2013. Contrast and stability of the axon diameter index from microstructure imaging with diffusion MRI. *Magn. Reson. Med.* 70, 711–721.
- Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci.* 113, 7900–7905.
- Ferizi, U., Schneider, T., Witzel, T., Wald, L.L., Zhang, H., Wheeler-Kingshott, C.A., Alexander, D.C., 2015. White matter compartment models for in vivo diffusion MRI at 300mT/m. *Neuroimage* 118, 468–483.
- Fischl, B., 2012. *NeuroImage* 62, 774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021>.
- Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., 2013. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80, 105–124.
- Glover, G.H., Pelc, N.J., 1983. In: Method for Correcting Image Distortion Due to Gradient Nonuniformity. URL: <https://patents.google.com/patent/US4591789A/en>.
- Herting, M.M., Maxwell, E.C., Irvine, C., Nagel, B.J., 2012. The impact of sex, puberty, and hormones on white matter microstructure in adolescents. *Cerebr. Cortex* 22, 1979–1992.
- Holmes, A.P., Blair, R.C., Watson, G., Ford, I., 1996. Nonparametric analysis of statistic images from functional mapping experiments. *J. Cereb. Blood Flow Metab.* 16, 7–22.
- Hsu, J.L., Leemans, A., Bai, C.H., Lee, C.H., Tsai, Y.F., Chiu, H.C., Chen, W.H., 2008. Gender differences and age-related white matter changes of the human brain: a diffusion tensor imaging study. *Neuroimage* 39, 566–577.
- Ingallhalikar, M., Smith, A., Parker, D., Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Hakonarson, H., Gur, R.E., Gur, R.C., Verma, R., 2014. Sex differences in the structural connectome of the human brain. *Proc. Natl. Acad. Sci.* 111, 823–828.
- Janke, A., Zhao, H., Cowin, G.J., Galloway, G.J., Doddrell, D.M., 2004. Use of spherical harmonic deconvolution methods to compensate for nonlinear gradient effects on MRI images. *Magn. Reson. Med.* 52, 115–122.
- Jensen, J.H., Helpert, J.A., Ramani, A., Lu, H., Kaczynski, K., 2005. Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. *Magn. Reson. Med.* 53, 1432–1440.
- Jones, D.K., Basser, P.J., 2004. Squashing peanuts and smashing pumpkins: how noise distorts diffusion-weighted MR data. *Magn. Reson. Med. : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 52, 979–993.

- Jones, D., Williams, S., Horsfield, M., 1997. Full representation of white matter fibre direction on one map via diffusion tensor analysis. In: *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 5 Canada, Vancouver, p. 1743.
- Jones, D.K., Horsfield, M.A., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn. Reson. Med.* : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 42, 515–525.
- Kanaan, R.A., Allin, M., Picchioni, M., Barker, G.J., Daly, E., Shergill, S.S., Woolley, J., McGuire, P.K., 2012. Gender differences in white matter microstructure. *PLoS One* 7, e38272.
- Langlois, S., Desvignes, M., Constans, J., Revenu, M., 1999. MRI geometric distortion: a simple approach to correcting the effects of non-linear gradient fields. *J. Magn. Reson. Imaging* 9, 821–831.
- Lay, D.C., Lay, S.R., McDonald, J., 2016. *Linear Algebra and its Applications*, fifth ed. Leemans, A., Jones, D.K., 2009. The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn. Reson. Med.* 61, 1336–1349.
- Leemans, A., Jeurissen, B., Sijbers, J., Jones, D., 2009. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine* 17, 3537.
- Malyarenko, D.I., Ross, B.D., Chenevert, T.L., 2014. Analysis and correction of gradient nonlinearity bias in apparent diffusion coefficient measurements. *Magn. Reson. Med.* 71, 1312–1323.
- McNab, J.A., Edlow, B.L., Witzel, T., Huang, S.Y., Bhat, H., Heberlein, K., Feiweier, T., Liu, K., Keil, B., Cohen-Adad, J., Tisdall, M.D., Folkerth, R.D., Kinney, H.C., Wald, L.L., 2013. The human connectome Project and beyond: Initial applications of 300mT/m gradients. *Neuroimage* 80, 234–245.
- Menzler, K., Belke, M., Wehrmann, E., Krakow, K., Lengler, U., Jansen, A., Hamer, H.M., Oertel, W.H., Rosenow, F., Knake, S., 2011. Men and women are different: diffusion tensor imaging reveals sexual dimorphism in the microstructure of the thalamus, corpus callosum and cingulum. *Neuroimage* 54, 2557–2562.
- Mesri, H.Y., Froeling, M., Viergever, M., Heemskerk, A., Leemans, A., 2017. Investigating the adverse effect of gradient nonuniformities on diffusion MRI measures: Do we need to worry?. In: *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 25, p. 3534. Honolulu, USA. <http://indexsmart.miramart.com/ISMRM2017/PDFfiles/3534.html>.
- Mesri, H.Y., Konijn, M.C., Viergever, M., Leemans, A., 2017. Baseline measures for diffusional kurtosis imaging in the human brain: results from the Human Connectome Project. *Magnetic Resonance Materials in Physics, Biology and Medicine* 30, 412–413.
- Mesri, H.Y., David, S., Viergever, M.A., Leemans, A.A., 2018. Investigating the effect of gradient nonlinearities on diffusional kurtosis imaging parameters: results from the human connectome Project. In: *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 26, p. 1647. Paris.
- Mesri, H.Y., David, S., Viergever, M., Leemans, A., 2018. Investigating the performance of diffusional kurtosis imaging for group-wise analyses: a study from the human connectome Project. In: *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 26, p. 3097. Paris.
- Michiels, J., Bosmans, H., Pelgrims, P., Vandermeulen, D., Gybels, J., Marchal, G., Suetens, P., 1994. On the problem of geometric distortion in magnetic resonance images for stereotactic neurosurgery. *Magn. Reson. Imag.* 12, 749–765.
- Mohammadi, S., Nagy, Z., Möller, H.E., Symms, M.R., Carmichael, D.W., Josephs, O., Weiskopf, N., 2012. The effect of local perturbation fields on human DTI: Characterisation, measurement and correction. *Neuroimage* 60, 562–570.
- Nagy, Z., Alexander, D., Weiskopf, N., 2009. Measuring and correcting errors that occur in diffusion weighted images due to non-ideal gradient linearity. In: *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine Honolulu*, p. 1377.
- Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25.
- Núñez, C., Theofanopoulou, C., Senior, C., Cambra, M.R., Usall, J., Stephan-Otto, C., Brébion, G., 2018. A large-scale study on the effects of sex on gray matter asymmetry. *Brain Struct. Funct.* 223, 183–193.
- Pierpaoli, C., 1997. Oh no! One more method for color mapping of fiber tract direction using diffusion MR imaging data. In: *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 5. Canada, Vancouver, p. 1741.
- Ritchie, S.J., Cox, S.R., Shen, X., Lombardo, M.V., Reus, L.M., Alloza, C., Harris, M.A., Alderson, H.L., Hunter, S., Neilson, E., Liewald, D.C.M., Auyeung, B., Whalley, H.C., Lawrie, S.M., Gale, C.R., Bastin, M.E., McIntosh, A.M., Deary, I.J., 2018. Sex differences in the adult human brain: evidence from 5216 UK biobank participants. *Cerebr. Cortex* 28, 2959–2975.
- Ruigrok, A.N., Salimi-Khorshidi, G., Lai, M.C., Baron-Cohen, S., Lombardo, M.V., Tait, R.J., Suckling, J., 2014. A Meta-Analysis of Sex Differences in Human Brain Structure. <https://doi.org/10.1016/j.neubiorev.2013.12.004>. <https://www.sciencedirect.com/science/article/pii/S0149763413003011>. ?via={%}3Dihub.
- Setsompop, K., Kimmlingen, R., Eberlein, E., Witzel, T., Cohen-Adad, J., McNab, J.A., Keil, B., Tisdall, M.D., Hoecht, P., Dietz, P., Cauley, S.F., Tountcheva, V., Matschl, V., Lenz, V.H., Heberlein, K., Potthast, A., Thein, H., Van Horn, J., Toga, A., Schmitt, F., Lehne, D., Rosen, B.R., Wedeen, V., Wald, L.L., 2013. Pushing the limits of in vivo diffusion MRI for the human connectome Project. *Neuroimage* 80, 220–233.
- Skare, S., Hedehus, M., Moseley, M.E., Li, T.Q., 1997. Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI. *Journal of magnetic resonance (San Diego, Calif)* 147, 340–352, 2000.
- Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44, 83–98.
- Sotiropoulos, S.N., Jbabdi, S., Xu, J., Andersson, J.L., Moeller, S., Auerbach, E.J., Glasser, M.F., Hernandez, M., Sapiro, G., Jenkinson, M., Feinberg, D.A., Yacoub, E., Lenglet, C., Van Essen, D.C., Ugurbil, K., Behrens, T.E.J., 2013. Advances in diffusion MRI acquisition and processing in the human connectome Project. *Neuroimage* 80, 125–143.
- Sumanaweera, T., Glover, G., Song, S., Adler, J., Napel, S., 1994. Quantifying MRI geometric distortion in tissue. *Magn. Reson. Med.* 31, 40–47.
- Tabesh, A., Jensen, J.H., Ardekani, B.A., Helpert, J.A., 2011. Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. *Magn. Reson. Med.* 65, 823–836.
- Tao, S., Trzasko, J.D., Shu, Y., Huston, J., Bernstein, M.A., 2015. Integrated image reconstruction and gradient nonlinearity correction. *Magn. Reson. Med.* 74, 1019–1031.
- Tyan, Y.S., Liao, J.R., Shen, C.Y., Lin, Y.C., Weng, J.C., 2017. Gender differences in the structural connectome of the teenage brain revealed by generalized q-sampling MRI. *Neuroimage: Clinical* 15, 376–382.
- Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., 2012. In: *The Human Connectome Project: A Data Acquisition Perspective*. <https://doi.org/10.1016/j.neuroimage.2012.02.018>. URL: <http://www.sciencedirect.com/science/article/pii/S1053811912001954>. arXiv: NIHMS150003.
- Veraart, J., Sijbers, J., Sunaert, S., Leemans, A., Jeurissen, B., 2013. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. *Neuroimage* 81, 335–346.
- Vos, S.B., Jones, D.K., Viergever, M.A., Leemans, A., 2011. Partial volume effect as a hidden covariate in DTI analyses. *Neuroimage* 55, 1566–1576.
- Westerhausen, R., Walter, C., Kreuder, F., Wittling, R.A., Schweiger, E., Wittling, W., 2003. The influence of handedness and gender on the microstructure of the human corpus callosum: a diffusion-tensor magnetic resonance imaging study. *Neurosci. Lett.* 351, 99–102.
- Wierenga, L.M., Sexton, J.A., Laake, P., Giedd, J.N., Tamnes, C.K., 2017. A key characteristic of sex differences in the developing brain: Greater variability in brain structure of boys than girls. *Cerebr. Cortex* 28, 1–11.
- Winkler, A.M., Ridgway, G.R., Webster, M.A., Smith, S.M., Nichols, T.E., 2014. Permutation inference for the general linear model. *Neuroimage* 92, 381–397.
- Winkler, A.M., Ridgway, G.R., Douaud, G., Nichols, T.E., Smith, S.M., 2016. Faster permutation inference in brain imaging. *Neuroimage* 141, 502–516.
- Winkler, A.M., Webster, M.A., Brooks, J.C., Tracey, I., Smith, S.M., Nichols, T.E., 2016. Non-parametric combination and related permutation tests for neuroimaging. *Hum. Brain Mapp.* 37, 1486–1511.