

Improved reproducibility of diffusion MRI of the human brain with a four-way blip-up and down phase-encoding acquisition approach

M. Okan Irfanoglu¹  | Neda Sadeghi¹ | Joelle Sarlls² | Carlo Pierpaoli¹

¹Quantitative Medical Imaging Section, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD, USA

²NIH MRI Research Facility, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA

Correspondence

M. Okan Irfanoglu, Quantitative Medical Imaging Section, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD, USA.

Email: irfanoglumo@nih.gov

Funding information

NIH; NIBIB Intramural Program; NINDS Intramural Program

Purpose: To assess the effects of blip-up and -down echo planar imaging (EPI) acquisition designs, with different choices of phase-encoding directions (PEDs) on the reproducibility of diffusion MRI (dMRI)-derived metrics in the human brain.

Methods: Diffusion MRI data in seven subjects were acquired five times, each with five different protocols. The base design included 64 diffusion directions acquired with anterior-posterior (AP) PED, the first and second protocols added reverse phase-encoded $b=0$ s/mm² posterior-anterior (PA) PED images. The third one included 32 directions all with PED acquisitions with opposite polarity (AP and PA). The fourth protocol, also with 32 unique directions used four PEDs (AP, PA, right-left (RL), and left-right (LR)). The scan time was virtually identical for all protocols. The variability of diffusion MRI metrics for each subject and each protocol was computed across the different sessions.

Results: The highest reproducibility for all dMRI metrics was obtained with protocol four (AP/PA-RL/LR, ie, four-way PED). Protocols that used only $b=0$ s/mm² for distortion correction, which are the most widely used designs, had the lowest reproducibility.

Conclusions: An acquisition design with four PEDs, including all DWIs in addition to $b=0$ s/mm² images should be used to achieve high reproducibility in diffusion MRI studies.

KEYWORDS

artifacts, diffusion MRI, distortion correction, EPI, reproducibility

1 | INTRODUCTION

Quantitative diffusion MRI (dMRI) acquisitions, including diffusion tensor imaging (DTI)^{1,2} and high angular resolution diffusion imaging (HARDI)³⁻⁹ are used extensively to

investigate architectural, microstructural, and compositional features of the human brain. However, obtaining reproducible and accurate dMRI results is challenging given that diffusion-weighted images (DWIs), which are collected using echo planar imaging (EPI), are susceptible to various

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Published 2020. This article is a U.S. Government work and is in the public domain in the USA. Magnetic Resonance in Medicine published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.

artifacts (see Ref. [10] for a review). DWI artifacts originate from ghosting, Gibbs ringing, signal drop outs, misalignment due to motion, eddy-currents distortions, as well as abnormal morphology due to distortions produced by static B_0 field inhomogeneities such as magnetic susceptibility variations,¹¹ imperfect shimming, and concomitant fields.¹²

EPI distortions have been shown to have a significant impact on the accuracy of tensor-derived scalar maps^{13,14} and fiber tractography,¹⁵ and can negatively impact consistency of results in multicenter studies.¹⁶ Therefore, proper EPI distortion correction in preprocessing is likely to be relevant to improve dMRI reproducibility both for multi- and single-site studies.^{17,18} Correction of EPI distortions generally requires the acquisition of additional data, such as field-mapping¹¹ or structural images for elastic registration.^{14,19,20} Earlier studies acquired dMRI using a single phase-encoding (PE) direction, most commonly anterior-posterior (AP) or posterior-anterior (PA) directions, and EPI distortion correction was either not performed (ADNI1) or performed using field mapping (ADNI3). More recently, reversed PE, or blip-up blip-down PE EPI distortion correction methods^{17,21-24} have been adopted in diffusion MRI demonstrating generally superior performance to field-mapping. Although it has been shown that it is advantageous to use blip-up blip-down PE for the entire set of DWIs,¹⁷ the most commonly used approach is to acquire blip-up and -down data only for the non-diffusion-weighted, $b = 0$ s/mm² volumes.^{22,24} Large imaging studies including the ABCD study²⁵ and the UK Biobank²⁶ performs blip-up and blip-down acquisitions only on the $b = 0$ s/mm² volumes. The Human Connectome Project (HCP) protocol²⁷ adopted blip-up and blip-down acquisitions for all DWIs (*LR/RL* directions). The Developing Human Connectome Project²⁸ implemented a strategy with diffusion MRI acquisitions spread over all four PE directions (*AP*, *PA*, *RL*, and *LR*).

In this work, we evaluate the effects of using different PE schemes for EPI distortion correction and artifact reduction on the reproducibility of diffusion MRI results. We analyze the reproducibility of dMRI-derived metrics across several scanning sessions in the human brain of healthy subjects. In each session, we acquired a base protocol with all volumes collected with a single direction of PE (eg, *AP*, blip-up only) in which distortion correction was not performed. In the remainder of this manuscript, we refer to this protocol as *Protocol 0*. Then we added commonly used PE schemes for collecting blip-up blip-down datasets, namely: (a) all DWIs collected with a single PE direction adding a single $b = 0$ s/mm² volume collected with reversed PE (eg, *PA*, blip-down), (b) same as (a) but adding six $b \approx 0$ s/mm² volumes, instead of a single one, collected with reversed PE, (c) all volumes (diffusion gradients and b-values) collected as matching pairs between the blip-up and -down data, and (d) a “four-way” acquisition scheme that consists of volumes acquired with the same diffusion gradients and b-values of (c)

but with the volumes partitioned equally in two perpendicular PE orientations with again matching pairs of both blip-up and -down volumes (ie, *AP*, *PA*, left-right (*LR*), and right-left (*RL*)). All protocols had virtually identical scan time.

Our working hypothesis was that the reproducibility of diffusion MRI results for the various *AP-PA* protocols would increase following the order in which we listed them above. In particular, we expect *Protocol 3* to outperform the other *AP* protocols given its previously demonstrated distortion correction robustness.¹⁷ The four-way PE protocol (*AP*, *PA*, *RL*, and *LR*) is also a promising candidate to manifest superior reproducibility, because it would combine the good distortion correction performance of using reverse phase-encoded data for all DWIs¹⁷ with potentially less imaging artifacts. Given that ghosting artifacts manifest themselves at different spatial positions on images acquired with different PE directions, the four-way PE protocol data would suffer from such artifacts to a lesser extent. For instance, images acquired with *RL* PE would be devoid of ghosting artifact originating from the eyes and other anatomical structures ventral to the brain. These differences in spatial positions of ghosting artifacts in the four-way PE protocol would also open the intriguing possibility (which we did not explore in this work) of performing a selective editing of ghosting artifacts when combining DWIs after distortion correction.

2 | METHODS

2.1 | Details of acquisition protocols

Protocol 0: The first acquisition protocol is possibly still the most commonly used, where the entire dataset is acquired with a single PE direction: *AP*. With such a dataset, motion and eddy-currents distortions can still be corrected; however, other EPI distortions can only be corrected through either image registration^{14,19,20} or fieldmapping¹¹ if appropriate data are present, but not with blip-up blip-down approaches. In our experiments, no distortion correction was performed for this protocol, which constitutes our baseline.

Protocol 1: A single $b = 0$ s/mm² is added to *Prot0* with *PA* PE direction to enable blip-up blip-down distortion correction. *Protocol 2*: In addition to the single $b = 0$ s/mm² of *Prot1*, five $b = 50$ s/mm² images with *PA* PE direction are included in the blip-down dataset. This Protocol is conceptually identical to *Prot1*; however, EPI distortion correction should be more robust because it is less susceptible to the imperfections of a single image.

Protocol 3: The same diffusion gradients and b-values are used to acquire both the *AP* and *PA* datasets. After EPI distortion correction, the two datasets are merged into one by pairwise geometric averaging of corresponding volumes.^{17,29} To keep scan times identical to the previous protocols, the number of acquired volumes is halved for each PE direction compared to *Prot0*.

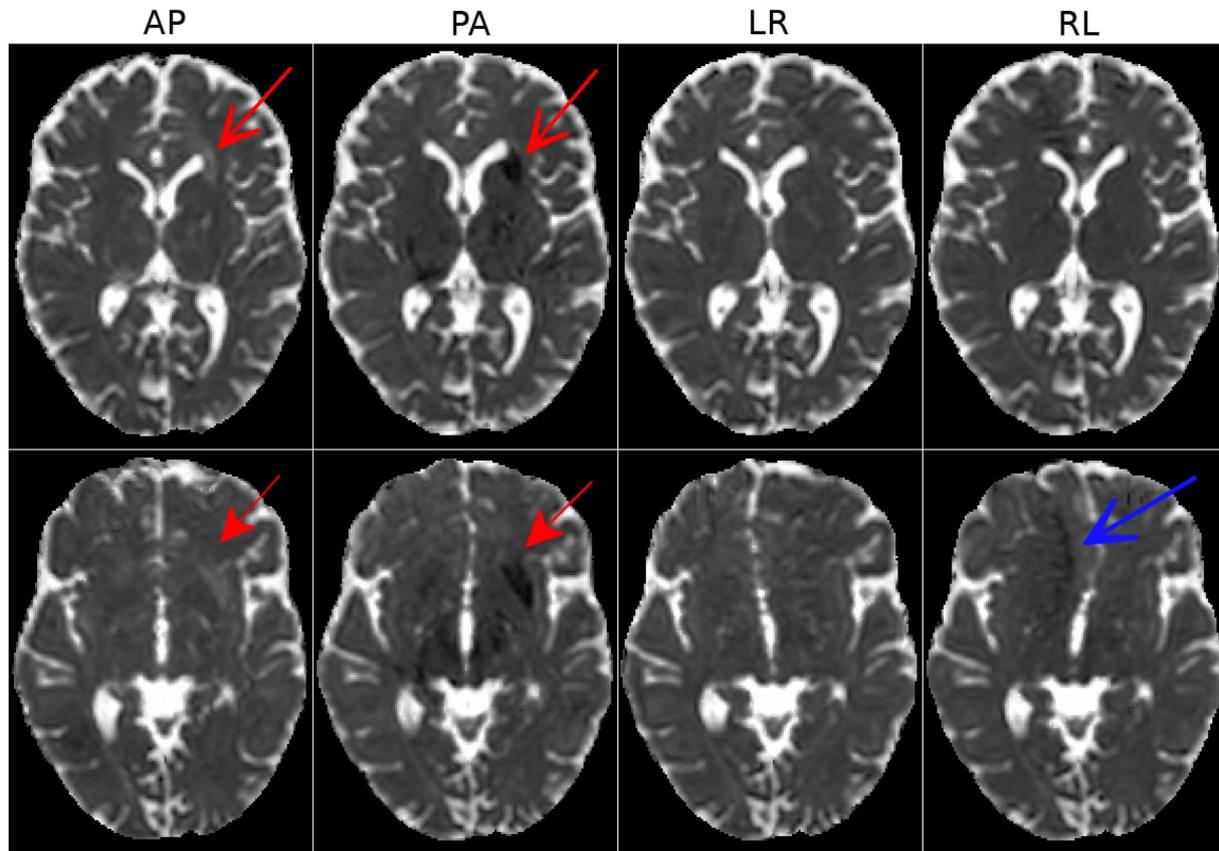


FIGURE 1 Mean diffusivity maps for a single subject for all four phase-encoding directions. All data were corrected for motion and eddy-currents but not other EPI distortions. The region indicated by red arrows display artifactual MD values for *AP* and *PA* phase-encoded data. For the same regions, *LR* and *RL* phase-encoded data seem artifact-free. Similarly, the blue arrow displays the location of a ghost artifact that affect the MD values in *RL*-encoded data. For this region, the other three datasets do not exhibit this ghost artifact

Protocol 4: This proposed design uses four PE directions: *AP*, *PA*, *RL*, and *LR*. *AP* and *PA* pairs have identical experimental designs to yield the *APPA* corrected dataset, as does *RL* and *LR*. However, the *RL/LR* pair is optimized to have different but complementary diffusion gradients to *AP/PA*. Both the *AP/PA* and *RL/LR* diffusion gradients are near-optimal in terms of electrostatic repulsion force³⁰ by themselves; therefore, they can be used independently if desired, but when combined together, they yield the actual optimum set. Even though the *AP* dataset alone has half the number of volumes compared to *Prot3*, the effective number of gradient directions in the processed and combined *APRL* dataset is identical to the *APPA* data of *Prot3*.

2.2 | Dataset

Diffusion MRI data were collected from seven healthy subjects (four males and three females; mean age = 28 years, std age = 3.6), five or six times, over a period of 6 months using a Philips Achieva 3T MRI System. The study was carried out under Institutional Review Board (IRB) approved protocols and all volunteers provided informed consent prior to examination.

Acquisitions were performed using a 32-channel head coil and a SENSE factor of 2 with no simultaneous multislice. For each design, half of the diffusion gradients sampled the northern diffusion hemisphere and half the southern hemisphere to minimize the effects of imaging gradients and eddy-currents on b-matrices. For each protocol, two datasets were acquired: one with a maximum b-value of 1100 s/mm² and one with a maximum b-value of 2700 s/mm² to enable HARDI analysis. DWIs for the DTI regime were acquired with 2 mm isotropic resolution. To achieve identical echo and repetition times among all scans (TE/TR: 92/12875 ms), the HARDI data were acquired at 2.6 mm isotropic resolution. A square field of view was used for all acquisitions to have identical echo train length for both *AP* and *RL* PE.

For *Prot0*, the experimental design included 12 low-*b* s/mm², 8 *b* = 300 s/mm², and 64 maximum b-value images (1100 and 2700) in the *AP* PE direction. For *Prot3*, which included full *AP* and *PA* acquisitions, the number of volumes were halved for each shell to achieve identical scan times. The low-*b* s/mm² images of this *PA* dataset were used to realize the experiments of *Prot1* and 2. For *Protocol4*, the same gradients and b-values used for *Protocol3* were near-optimally split between *AP* and *RL* sets. In addition to dMRI,

fat-suppressed T2W TSE and T1W MPRAGE images were also acquired as structural images. The total scan session for all protocols and the anatomical images took one hour and forty minutes.

To illustrate the level of distortions and artifacts in the data, Figure 1 displays the mean diffusivity (MD) images of one subject at two slice levels for all four PE directions. These datasets were corrected only for motion and eddy-currents distortions but not for other EPI distortions. In this figure, regions indicated by red arrows have artifactual MD values for *AP* and *PA* phase-encoded data, whereas *LR* and *RL* data do not exhibit this behavior. On the contrary, the blue arrow displays the location of a ghost artifact on the *RL* data and for this region, the other three phase-encoded data are not affected.

2.3 | DWI preprocessing

DWI preprocessing was performed separately using two pipelines, *TORTOISE*³¹ and *FSL*,³² to verify that conclusions about the reproducibility of each design were not highly pipeline dependent. Both pipelines were used with their latest available feature sets at the time of writing. With the *TORTOISE* pipeline, the voxelwise B-matrices due to gradient nonlinearities were initially generated,³³ then DWIs were corrected for Gibbs ringing artifacts³⁴ and subsequently for motion and eddy-currents distortions³⁵ and for other EPI distortions¹⁷ while reorienting the voxelwise B-matrices accordingly. For the *FSL* pipeline, susceptibility distortion correction was performed initially²¹ followed by motion and eddy-currents distortion correction³⁶ including outlier rejection,³⁷ slice-to-volume registration,³⁸ and per-volume re-estimation of susceptibility.³⁹

For both pipelines, for protocols with unmatched blip-up and down-data (*Prot1* and *Prot2*), the Jacobian-modulated signals were output in the corrected images. For protocols with matching blip-up and blip-down datasets (*Prot3* and *Prot4*), these two datasets were combined into a single one by pairwise averaging of corresponding corrected volumes. For *Prot4*, the *APPA* and *RLLR* corrected data were simply concatenated.

To provide a common space for longitudinal analysis, the T2W structural image of the first scan session for each subject was manually reoriented to ACPC orientation. The DWIs for all scan sessions for a subject were rigidly aligned to this ACPC-reoriented structural image, with the same registration method for both pipelines.

Even though all protocols that were analyzed used equal amount of data and identical scan times, as described in Section 2.1, the directional resolutions of *Protocol3* and *4* were halved compared to the others, which might be undesirable for a HARDI analysis. To test the hypothesis with

a HARDI model, we opted to use MAPMRI.⁴⁰ DTI was computed with nonlinear regression and MAPMRI with constrained quadratic programming with no regularization. The scalar maps of choice for reproducibility analysis included: fractional anisotropy (FA) and trace (TR) from the tensor model (TR = 3* mean diffusivity (MD)), as well as, return-to-origin-probability (RTOP), propagator anisotropy (PA), and non-gaussianity (NG) from the MAPMRI model. Principal eigenvector orientation dispersion (PEOD) was used as the metric to assess directional consistency of the diffusion tensors as described in Ref. [41], where *PEOD* = 0 indicates that all primary eigenvectors are identical and a *PEOD* = 1 indicates that eigenvectors span a uniform half-sphere.

2.4 | Variability analysis

As described in Section 2.3, each scan of a subject was rigidly aligned to the structural image of the first session; therefore, all the DWIs were in correspondence. For each subject, reproducibility between the five visits was assessed via voxelwise standard deviations (SD) and median absolute differences (MAD) for FA, TR, RTOP, PA, and NG maps. PEOD values were computed only within a WM mask. Statistics were computed separately for each design in the native space of the DWIs for both processing pipelines.

To produce a population summary map of reproducibility, a DTI atlas from the data of all seven subjects was created using the *DRTAMAS* software package⁴² and SD and MAD maps for each subject were warped onto the atlas space using the respective deformation fields, and then averaged.

3 | RESULTS

Figure 2 displays the standard deviation of TR maps computed over six scans of a representative subject for each protocol at three slice levels. Bright regions correspond to low TR reproducibility while dark regions correspond to high TR reproducibility. The protocol employing only one PE direction (*AP*), *Prot0*, results in the largest variability. *Prot1* and *Prot2*, which used differing number of $b = 0$ images for the *PA* direction, had slightly improved reproducibility, especially near tissue interfaces between WM/CSF and cortical gyri/sulci. Significant differences between these two protocols were not noticeable, indicating that the $b = 0$ image chosen for *Prot1* was of overall good quality. All these first three protocols suffered from the effects of artifacts that affected the *AP* acquisitions. TR values exhibited low reproducibility in regions indicated by red arrows, including the pons, a frontal white matter region and the centrum semiovale. A detectable improvement in overall reproducibility was achieved by *Prot3*;

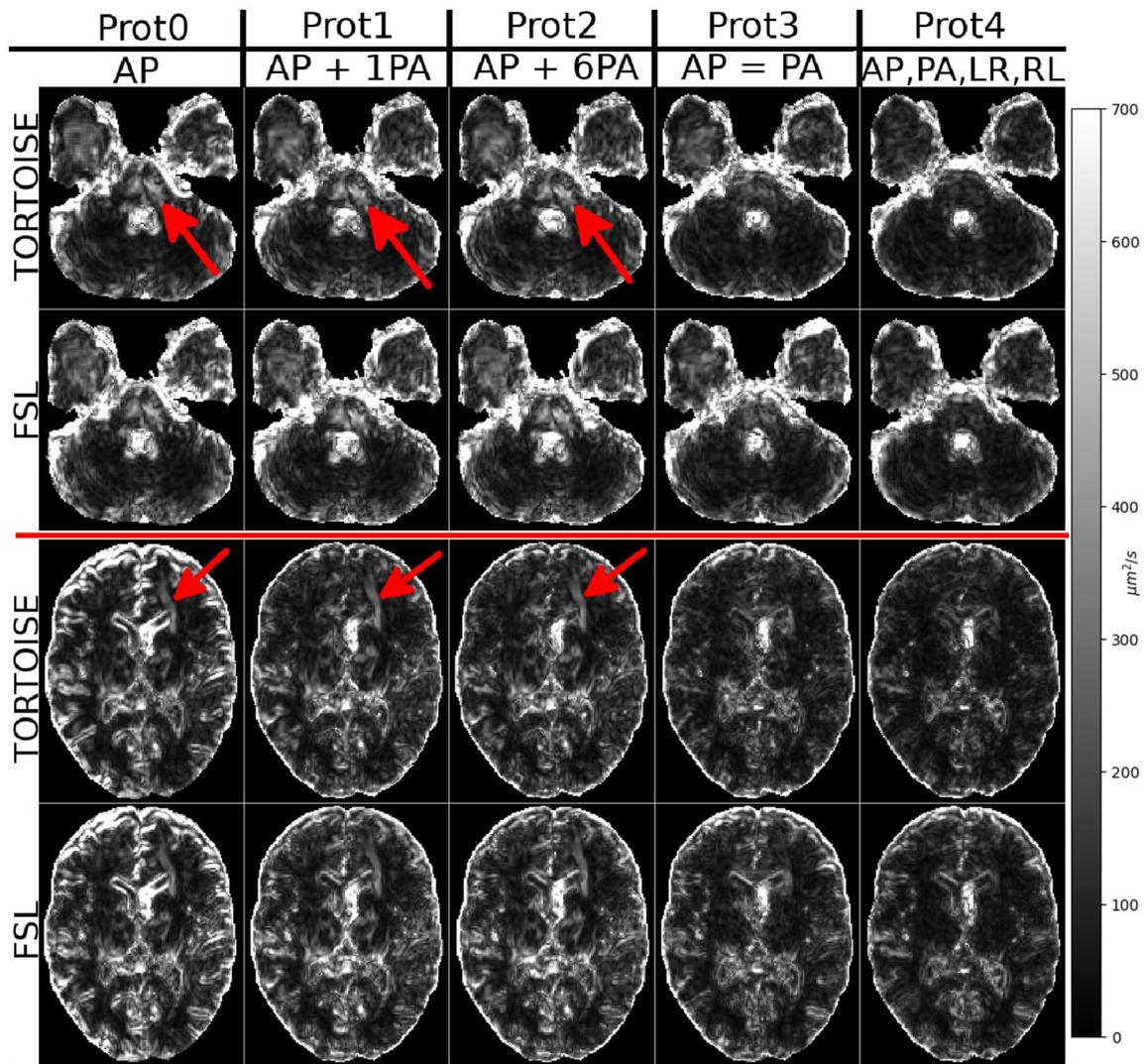


FIGURE 2 Standard deviation of Trace ($3 \times MD$) for a single subject over six scans for each protocol using *TORTOISE* and *FSL*. Brightness indicates increased variability among scans after processing; therefore darker regions are more reproducible. Pure white corresponds to $700 \mu\text{m}^2/\text{s}$. Red arrows point to the locations of high variability due to imaging artifacts such as ghosts. These regions include but are not limited to the pons and regions lateral to the ventricles at mid-brain level. Reproducibility improves with each protocol with Protocol4, that is, the four-way encoded design, yielding the best quality with both processing pipelines

however, the four-way encoding protocol, *Prot4*, virtually eliminated the poor reproducibility in these regions and clearly improved overall reproducibility throughout the brain. Overall, the reproducibility of TR maps computed using *TORTOISE* was slightly better than that of TR maps obtained with *FSL*; however, the trend of improvement in reproducibility across protocols was similar for the two pipelines.

Figure 3 displays the standard deviation of FA for the same subject of Figure 2. Although the areas of high variability in FA may differ from those of TR, the improvement in reproducibility with respect to the various protocols follows the behavior observed for TR: with increasing protocol number, the large variability in the brainstem and WM/CSF tissue interface regions is reduced as indicated by the red arrows. In the pons and in the corpus callosum, the variability of FA

images processed by *TORTOISE* appears lower than those processed by *FSL* for both *Prot3* and *Prot4*.

Figure 4 provides a summary of the reproducibility of each protocol for the entire population (see Figure caption for the definition of the plotted quantities). For the tensor-derived maps (top row), these summary statistics showed little variation in reproducibility using the first three protocols. However, for both the *FSL* and *TORTOISE* pipelines, *Prot3* and *Prot4* significantly improved reproducibility compared to the first three protocols, with *Prot4* producing the best reproducibility for all metrics. For the MPMRI-derived metrics, *TORTOISE* processing produced results that followed a trend similar to that of the tensor-derived quantities. Curiously, for PA and NG, the *FSL* processing resulted in increased variability with *Prot1* and *Prot2* compared to the no distortion correction protocol, *Prot0*. Regardless of the analyzed metric

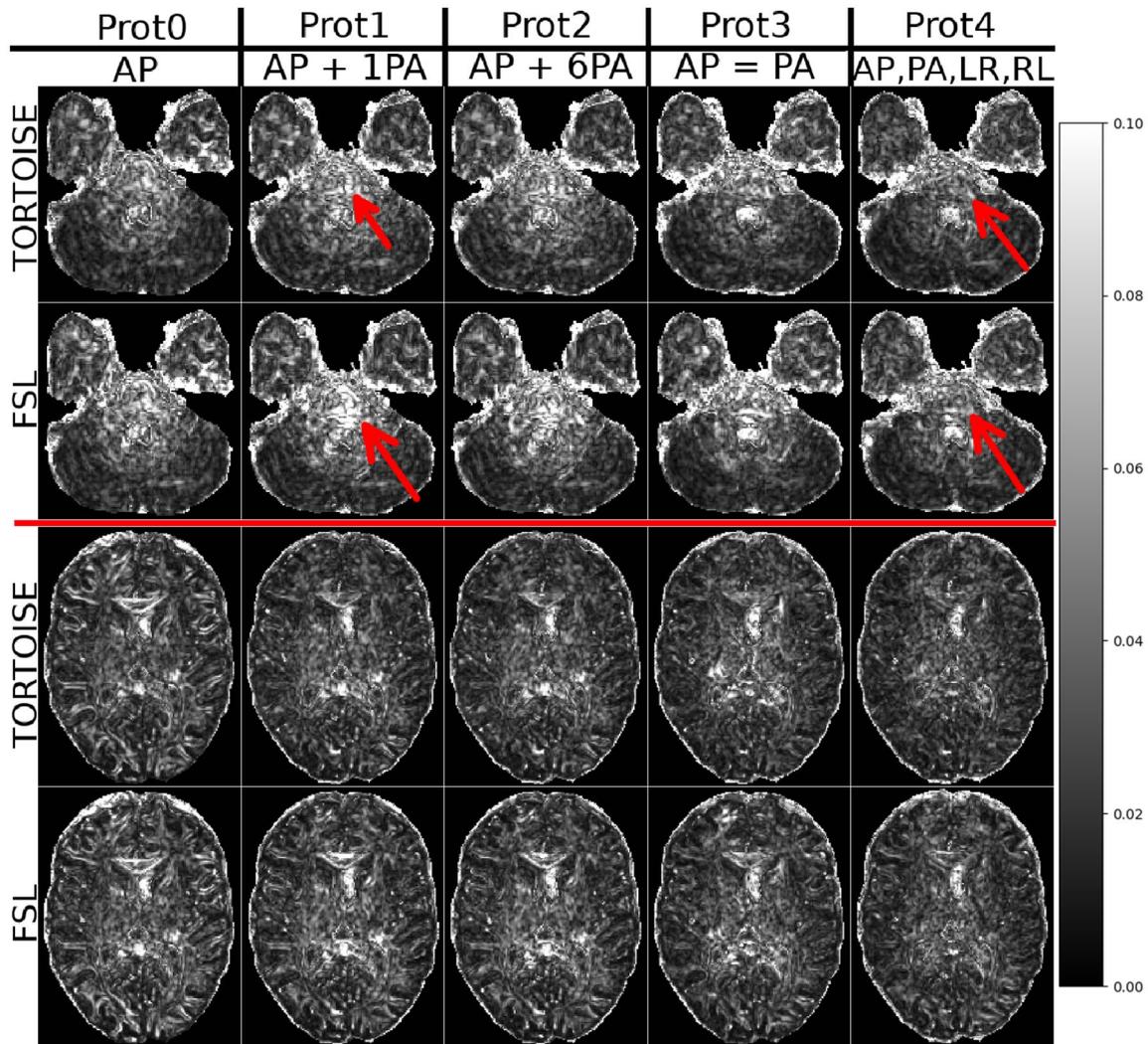


FIGURE 3 Standard deviation of FA for a single subject over six scans for each protocol using *TORTOISE* and *FSL*. Brightness indicates increased variability among scans after processing; therefore, darker regions are more reproducible. Pure white corresponds to 0.1 in FA. Red arrows point to the high variability in the pons for *Prot1* and the improvements achieved with *Prot4*. Reproducibility improves with each protocol with *Protocol4*, that is, the four-way encoded design, yielding the best quality with both processing pipelines

or the processing pipeline, the four-way PE approach provided the best reproducibility even in this population-wide analysis.

To allow an evaluation of the topological patterns of reproducibility at the population level, we report in Figure 5, the population level SD maps of tensor-derived metrics generated with the *TORTOISE* pipeline as described in Section 2.4, for two slice levels. The slice including the brainstem and the cerebellum (top panel) shows that the variability of FA in the pons (green arrows) is high for all protocols that did not include *RL* and *LR* data. For TR and PEOD, some improvement in variability can be appreciated in *Prot3*. However, the protocol that had the lowest variability for all metrics was again *Prot4*. It is also interesting to notice that some high variability in the temporal lobes (red arrows) can be noticed in all *AP* and *PA* protocols, for TR in particular. This high variability, which is due to ghosting from the eye signal into

the temporal lobes, is barely detectable in *Prot4*. The slice including the internal capsule (bottom panel) shows a similar variability improvement with *Prot4*; however, for all metrics, the protocol with no EPI distortion correction performed (*Prot0*) shows clearly higher variability compared to the other protocols that included EPI distortion correction. The light purple arrows highlight the variability at the boundaries of the corpus callosum.

Figure 6 displays the population level variability maps for PA, NG, and RTOP computed from the MAPMRI model. MAPMRI-derived variability maps showed a similar overall pattern to their tensor-derived counterparts with some exceptions. For instance, reproducibility of RTOP seemed to be worse with *Prot1* and *Prot2* compared to the no-correction case (*Prot0*) in the pons and the internal capsule regions. Additionally, at the level of the putamen and thalamus, *Prot1* and *Prot2* seemed to perform worse for all MAPMRI-derived

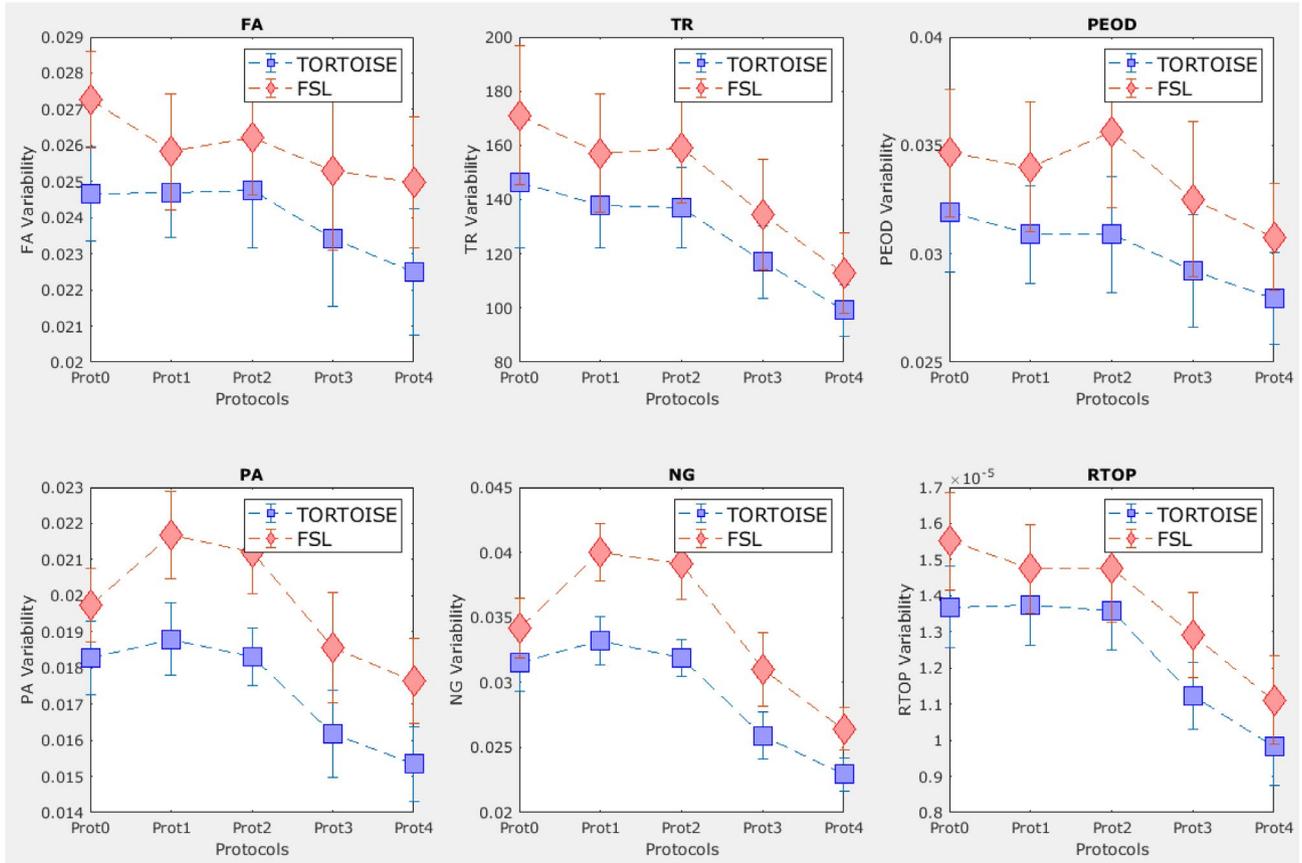


FIGURE 4 For each protocol, these graphs plot the average values across the population of seven subjects of the median from all voxels in the brain of the standard deviation of the metric of interest computed across all longitudinal scans acquired in each subject. The error bars represent the standard deviation of the average values across the population. For tensor-derived metrics, addition of data from different phase-encoding directions improves reproducibility for both pipelines. For MPMRI, adding a single $b = 0$ image to *PA* phase-encoding (*Prot1*) affects the sensitivity of the fitting in a negative way, yielding worse reproducibility. Four-way phase encoding provides the best results for both modalities and both pipelines

metrics. Despite these exceptions with *Prot1* and *Prot2*, *Prot3* again significantly improved reproducibility and *Prot4* was confirmed to produce the lowest variability for all metrics among all the approaches.

Table 1 displays the relative variability of each protocol for all metrics compared to the case where no distortion correction was performed (*Prot0*), which was set as reference with a value of 100%. For nearly all DTI-derived metrics, the variability was reduced with all successive protocol number, with *Prot1* and *Prot2* generally displaying a similar behavior. MPMRI-derived measures also exhibited similar patterns, with *Prot1-2* performing slightly worse than the baseline, but *Protocol3* and *Protocol4* performing significantly better. The four-way PE scheme, that is, *Protocol4* was the least variable, that is, most reproducible with all metrics.

4 | DISCUSSION

Low reproducibility of diffusion-derived metrics has been a major obstacle to the adoption of quantitative diffusion MRI in the clinical setting. The main goal of this work was to

examine the effects on diffusion MRI reproducibility of different strategies for the acquisitions of DWIs that are suitable for reversed PE (or blip-up blip-down) based EPI distortion correction techniques.^{17,21,22,24} Traditionally, in a clinical setting, DWIs are acquired with a single PE direction, either AP or PA and no EPI distortion correction is performed. Our first hypothesis was that correcting for EPI distortions in the DWIs would have resulted in better reproducibility of results in comparison to the clinical default of no correction (*Prot0*). Somewhat surprisingly, the protocols employing the popular EPI distortion correction strategies used in most large diffusion MRI multicenter studies: *Prot1* and *Prot2* (ie, using only $b = 0$ images in the *PA* direction for the correction) did not result in an appreciable improvement of reproducibility compared to the no correction case. For the *TORTOISE* pipeline, the reproducibility for the first three protocols were near identical for both the tensor-derived and MPMRI-derived metrics, whereas for the *FSL* pipeline, in some cases, reproducibility of *Prot1* and *Prot2* was even slightly worse compared to performing no distortion correction, *Prot0*. This surprising outcome might be attributed to the following: as observed in Ref. [17], an EPI distortion correction strategy

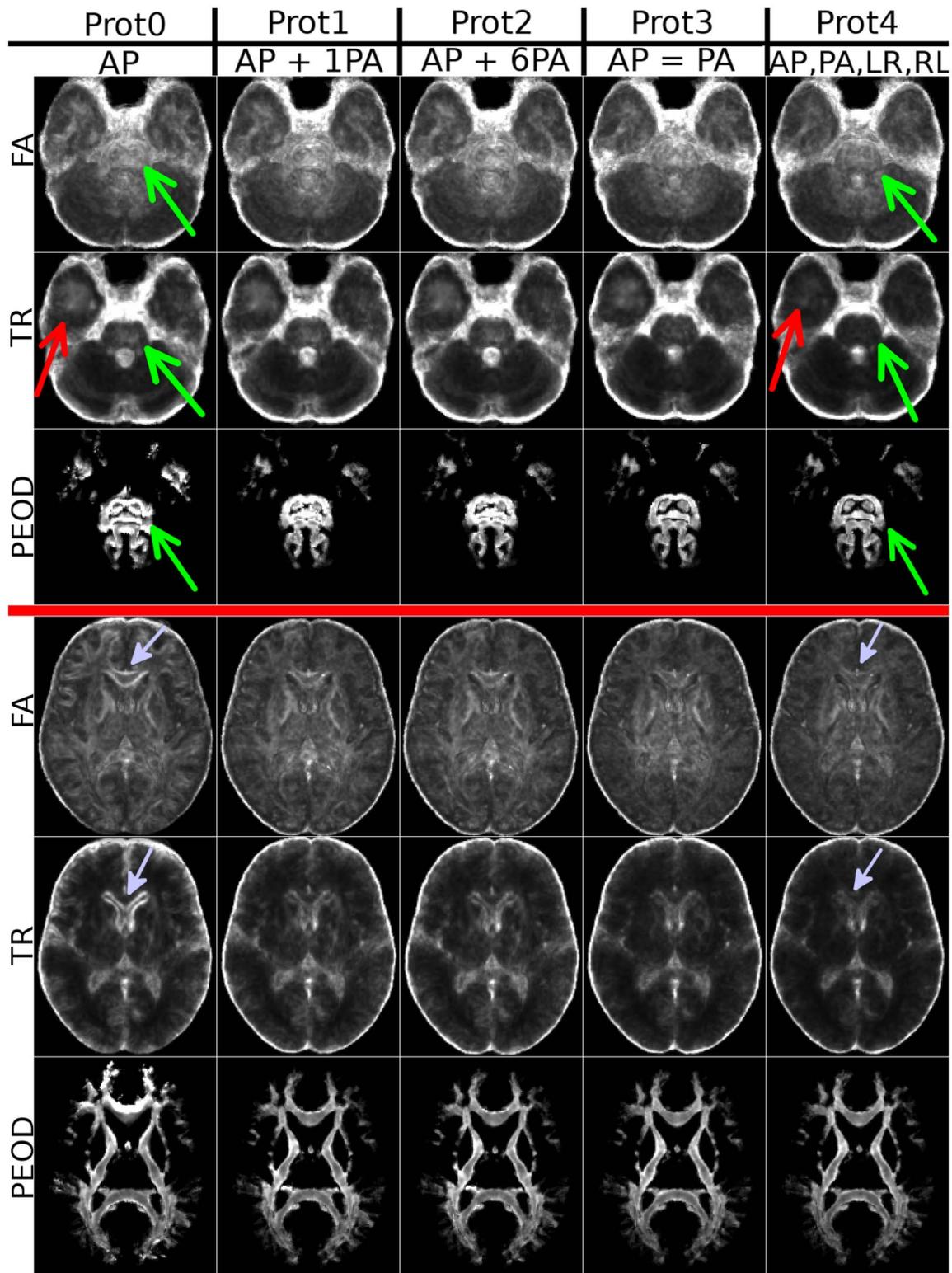


FIGURE 5 Population level tensor-derived average standard deviation maps from the *TORTOISE* pipeline at two slice levels. Columns represent the different protocols and the rows contain different tensor-derived modalities: FA, TR, and PEOD, respectively. The separation of data along more phase encoding directions improves the reproducibility for all metrics. The green arrows indicate the reproducibility improvements achieved in the pons region with Prot4 data compared to Prot0. The red arrows point to the temporal lobes which suffer from high variability in all protocols except Prot4 due to the ghosts of the eyes manifesting in this region. Light purple arrows point to the genu of corpus callosum which suffers from high variability in the non-distortion corrected data due to misalignments caused by EPI distortions

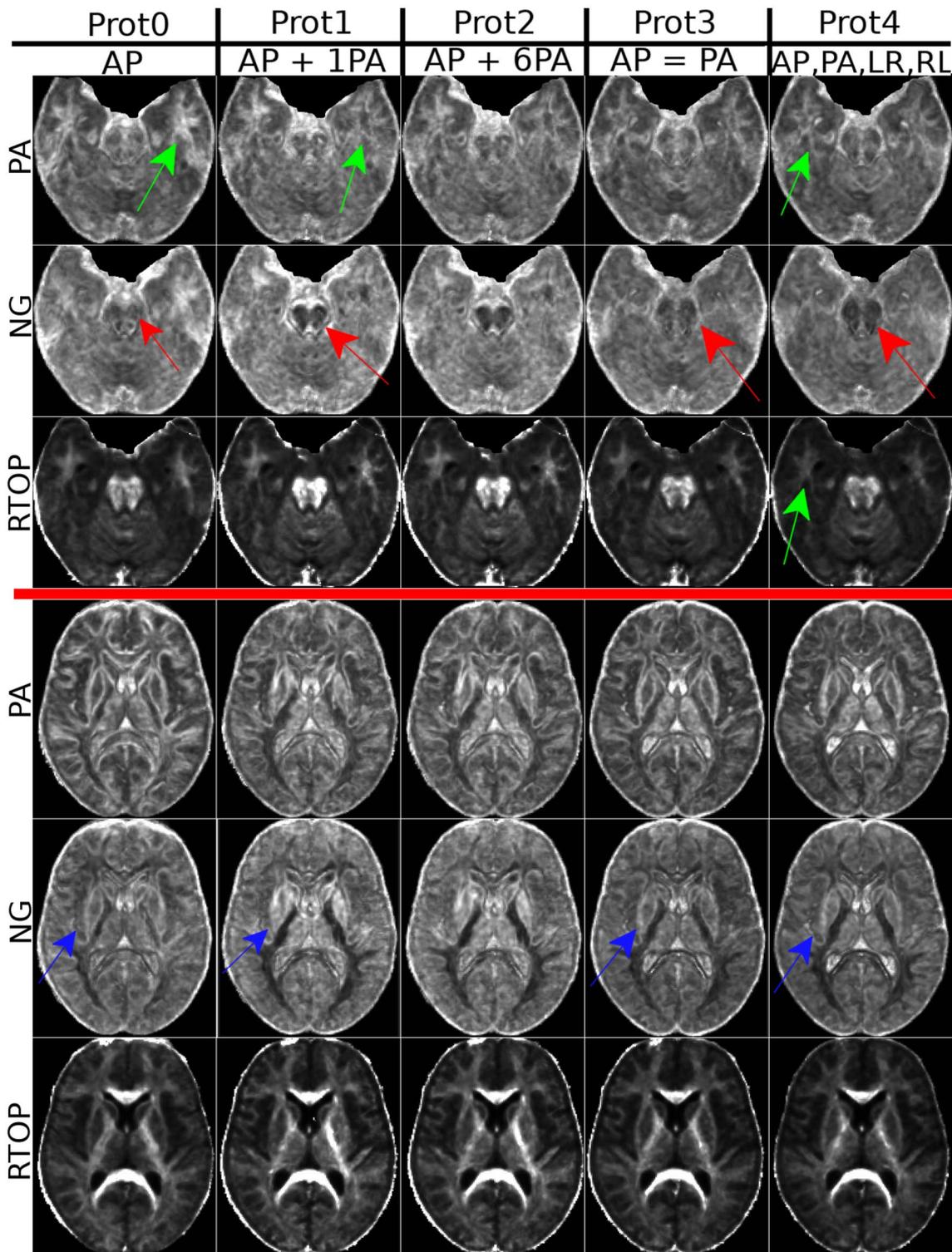


FIGURE 6 Population level MPMRI-derived average standard deviation maps from the *TORTOISE* pipeline at two slice levels. Columns represent the different protocols and the rows contain different MPMRI modalities: propagator anisotropy (PA), non-gaussianity (NG), and return-to-origin-probability (RTOP), respectively. Green arrows point to the high variability in the temporal lobes. Red arrows indicate the improvements achieved in reproducibility in the pons region with all protocols that included an EPI distortion correction step compared to the data which did not (Prot0). Blue arrows indicate the reproducibility improvements in the internal capsule region again with distortion correction. The reader should note that even though the reproducibility significantly improved for the internal capsule with all protocols compared to Prot0, with Protocols 1 and 2 the reproducibility worsened for the caudate region

TABLE 1 Absolute (left part of cells) and relative (right part of cells) variability of each protocol

	<i>Protocol0</i>	<i>Protocol1</i>	<i>Protocol2</i>	<i>Protocol3</i>	<i>Protocol4</i>
	Absolute Median σ /Relative Median σ				
TR	143/100%	134/94%	136/95%	116/81%	97/68%
FA	0.025/100%	0.025/100%	0.026/102%	0.024/96%	0.023/92%
PEOD	0.031/100%	0.030/97%	0.030/97%	0.029/92%	0.028/88%
PA	0.018/100%	0.019/102%	0.018/100%	0.016/88%	0.015/83%
NG	0.031/100%	0.033/105%	0.032/101%	0.026/82%	0.023/72%
RTOP	$1.4E^{-5}/100\%$	$1.4E^{-5}/100\%$	$1.4E^{-5}/99\%$	$1.1E^{-5}/82\%$	$1.E^{-5}/71\%$

Notes: Relative variability was computed with respect to *Prot0*, that is, the protocol where no distortion correction was performed ($\sigma_{Prot_i}/\sigma_{Prot0}$). The unit for TR is $\mu\text{m}^2/\text{s}$ and the other metrics are unitless. The statistics were computed at the population level. *Prot4* is the best performing protocol with the most reduction in variability for all metrics.

that only uses $b=0\text{ s/mm}^2$ images can only correct the distortion of the contour of regions that are homogeneous in the $b=0\text{ s/mm}^2$ images, as no information is present to guide the correction within the structure. This may lead to spurious deformations within the structures that in turn may lead to poor alignment of anatomical features and consequently lead to higher variability than performing no distortion correction at all.

The dataset which had blip-up and blip-down images for all the DWIs in the *AP* and *PA* directions, that is, *Prot3*, showed a significant improvement in reproducibility compared to the previously mentioned protocols. Reproducibility of all imaging metrics was further improved with the four-way protocol (*Prot4*), which also used the *LR* and *RL* PE data. Overall, a small systematic difference was observed between *TORTOISE* and *FSL* pipelines with *TORTOISE* data variability being generally lower than that of *FSL*; however, the overall reproducibility trend across different protocols is present in data processed with either pipeline.

The magnitude of reduction in variability with the four-way protocol compared to the baseline was spatially varying, with some regions exhibiting more significant improvements than others but nearly all brain voxels showed improvements. The most remarkable reproducibility improvements were found both in regions susceptible to severe EPI distortions and in regions that suffered from ghosting artifacts. For the former category, the TR in the genu of the corpus callosum showed a reduction of 30% in variability ($1-\sigma_{Prot4}/\sigma_{Prot0}$). For the latter category, the TR variability in the temporal lobes, which contained a ghost of the eyes in *AP* encoded data, exhibited a reduction in variability of 50%. Also, the pons, which suffered from ghosting of the surrounding CSF regions, showed a decrease in variability of 30% for TR and 18% for FA. In addition to the improvements in regions where obvious artifacts were present, the four-way PE protocol also provided the lowest overall whole-brain variability (Table 1), which has the obvious benefit of improving the statistical power, when, for instance, exploring the differences between

healthy and patient populations. Moreover, *Prot4* showed the most homogeneous variability across brain regions (Figure 5), which is a desirable feature because it achieves the goal of having the same statistical power across brain regions for a given number of recruited subjects.

With *Prot1* and *Prot2*, MAPMRI-derived metric reproducibility exhibited a different behavior than their DTI counterparts, even worsening the reproducibility in some regions compared to the protocol where no distortion correction was performed. This can be attributed to the following: MAPMRI is a more complex and flexible model than DTI. Therefore, it is more sensitive to the imperfections of the data during the fitting process. When performing EPI distortion correction with a full *AP* dataset using only the $b=0$ images for the *PA* PE (*Prot1-2*), any imaging artifacts such as ghosts in these *PA* $b=0$ images would cause the estimated deformation fields to be inaccurate. These variations in the deformation fields would in turn cause the signals to be different for the same anatomical location in the longitudinal scans. Therefore, the highly flexible MAPMRI fitting then would assign these spurious signals as features of the apparent diffusion propagator causing the low reproducibility of the derived metrics. Another interesting observation regarding MAPMRI is that the reductions in reproducibility for MAPMRI with *Prot1-2* compared to *Prot0* were not systematic but they were regionally dependent. As can be observed in Figure 6, the reproducibility actually improved with *Prot1* in the internal capsule but at the same time worsened in the basal ganglia region, in particular in the globus pallidum, which typically has lower SNR in the $b=0$ images, resulting in unstable deformation fields.

Combining *AP* phase-encoded data and *RL* phase-encoded data would not be feasible in case the distortions are not fully corrected and the PE bandwidth, TE, and TR are not the same for the two datasets. For our experiments, we ensured that *AP* and *RL* acquisition parameters were identical and that we verified that EPI distortion correction was adequate for combining the two datasets. Several

strategies to combine the *AP/PA* and *RL/LR* data could be considered. The simplest approach would be the simple concatenation of the two datasets. Given that the locations of ghosting artifacts are different in the *AP* and *RL* datasets, more sophisticated combination approaches would be aimed at identifying these artifacts and favoring the dataset that is artifact-free. For artifact identification, several strategies can be considered including: (a) *Iterative re-weighting-based fitting* approaches similar to *RESTORE*⁴³ or *REKINDLE*,⁴⁴ which can be retrofitted to handle *AP* and *RL* data separately for voxelwise outlier identification, (b) *Postprocessing registration* approaches which handle the *APPA* and *RLLR* data separately without any combination but which perform an additional successive diffusion MRI based registration,^{42,45,46} and (c) *Computer vision-based artifact identification* approaches which detect artifacts using image processing or machine learning techniques and pass this information as weights to a tailored fitting process. In this work, we chose the simplest approach that is concatenating the two datasets. We reasoned that if reproducibility is improved with simple concatenation, any future, smarter way of combining the two datasets could only improve the results.

4.1 | Limitations of the current study and future directions

In this work, the duration of a single scanning session was close to 2 hours, which was prohibitively long to acquire data for any additional experiments that we would like to have conducted. Such experiments and other limitations of the current study are discussed below.

Although it has been clearly shown that dual PE both $b = 0$ images and DWIs is advantageous,¹⁷ this approach we have evaluated in *Prot3* has not gained wide acceptance. The main concern in adapting such a strategy is that by reverse PE all DWIs, the number of unique diffusion gradient directions in the dataset is cut in half. In other words, the directional resolution of the diffusion sampling is compromised. In this work, we followed this design for *Prot3*; therefore the directional resolution of our gradient scheme was half of that of the protocols that reversed the PE of the $b = 0$ images only (*Prot1 Prot2*). However, prospective studies do not need to follow this sampling scheme. For applications that are sensitive to gradient direction resolution, such as fiber tractography, the gradient directions of the baseline protocol can simply be split among the four PE directions, without any penalty in directional sampling resolution. A similar approach has been adopted in the UK dHCP project.²⁸

In this study, we analyzed the diffusion MRI reproducibility using only one scanner. Ideally, this study should be

extended to other vendors and scanner models to generalize the superiority of the four-way PE and to encourage its clinical adaptation. It should be noted that the implementation of the four-way PE protocol might not be straightforward in some scanners. Additional research licenses might be required for such acquisitions in clinical settings. Even with these licenses, special attention has to be paid to keeping all four acquisitions identical in parameters, such as FoV, TE, TR, and diffusion times. Additionally, the minimum achievable TE might be slightly penalized for *RL* and *LR* PE directions due to peripheral nerve stimulation related restrictions. For instance, in this work, the minimum TEs for *RL* and *LR* was about 7 ms longer than those of *AP* and *PA*. At the expense of this small reduction in SNR for *AP* and *PA*, all four TEs were manually set to the longer of these two values. For clinical settings, the cooperation from scanner manufacturers for an easy implementation of such a four-way protocol with a user friendly parameter optimization would be very beneficial.

Given scan time limitations, one protocol that is missing from the set of our tested protocols is the unique *RL* and *LR* protocol acquired with the same directional sampling resolution of *AP* and *PA* protocols. Therefore, we could not assess whether a purely *RL* and *LR* protocol would have achieved even higher reproducibility than the hybrid four-way protocol that was tested.

4.2 | Conclusions and practical recommendations

In summary, the four-way PE protocol that we have proposed was shown to provide a very relevant improvement in reproducibility, for both diffusion tensor and higher order diffusion metrics, compared to performing no distortion correction or using current blip-up blip-down acquisition and distortion correction approaches used in large quantitative dMRI studies. It is important to remind readers that these comparisons were performed using datasets collected using identical scan times.

In order to improve the reproducibility and provide a robust acquisition of dMRI data, we suggest to adopt some of the strategies that we have used in this study that are not commonly employed in clinical acquisitions, namely:

1. *Gradient distribution over PEs*: In case maintaining high directional resolution of diffusion sampling is desired, the gradient sampling scheme should be partitioned across the four PE directions. Each of these subsets should still maintain a near optimal sampling distribution in three-dimensional gradient vector space³⁰ so that each PE direction dataset would be intrinsically balanced and potentially analyzed independently.

2. *Imaging gradients*: The diffusion gradients for each PE direction should span the entire sphere instead of a half-sphere to balance the potential effects of imaging gradients.^{47,48}
3. *Parameter harmonization*: Once the diffusion gradient sampling over PEs is finalized, the other acquisition parameters such as TE/TR have to be harmonized over all acquisitions. Given that typically *RL* and *LR* acquisitions have more stringent TE limitations, we advise to first investigate the minimum achievable TE and TR for *RL* and *LR* acquisitions and then set the same values for *AP* and *PA* acquisitions.
4. *Scanner considerations*: As stated above, implementations of the four-way protocol is not straightforward in some scanners. For implementations on different scanner platforms, the website https://tortoise.nibib.nih.gov/tortoise/data_acquisition contains specific instructions and pointers.
5. *Preprocessing pipeline considerations*: The preprocessing pipeline that is intended to be used might have specific acquisition requirements. For instance, the *TORTOISE* pipeline used in this work is designed, in a user-friendly way, to handle four-way PE datasets; however, it requires a distortion-free T2W fat-suppressed anatomical image at several steps of the processing. Other pipelines might have different requirements.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the National Institute of Biomedical Imaging and Bioengineering and National Institute of Neurological Disorders and Stroke in the National Institutes of Health. The contents of this work do not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred. The authors would also like to thank Susan Fulton and Steven Newman for their efforts in subject recruitment and Dr. Alan Koretsky, the principal investigator of the clinical protocol under which this study was conducted.

ORCID

M. Okan Irfanoglu  <https://orcid.org/0000-0001-8351-1516>

REFERENCES

1. Basser PJ, Mattiello J, Le Bihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson.* 1994;103:247-254.
2. Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Chiro GD. Diffusion tensor MR imaging of the human brain. *Radiology.* 1996;201:637-648.
3. Frank LR. Characterization of anisotropy in high angular resolution diffusion-weighted MRI. *Magn Reson Med.* 2002;47:1083-1099.
4. Jansons KM, Alexander DC. Persistent angular structure: new insights from diffusion MRI data. *Inverse Prob.* 2003;19:1031-1046.
5. Tuch DS. *Diffusion MRI of Complex Tissue Structure* [PhD thesis]. Cambridge, MA: Harvard University; 2002.
6. Tuch DS. Q-ball imaging. *Magn Reson Med.* 2004;52:1358-1372.
7. Tuch DS, Weisskoff RM, Belliveau JW, Wedeen VJ. High angular resolution diffusion imaging of the human brain. In: Proceedings of International Society of Magnetic Resonance in Medicine, Philadelphia, PA; 1999:321.
8. Wedeen VJ, Tuch D, Weigel M, Dou J, Weisskoff R, Chessler D. Mapping fiber orientation spectra in cerebral white matter with fourier-transform diffusion MR. In: Proceedings of International Society of Magnetic Resonance in Medicine, Denver, CO; 2000;8:82.
9. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage.* 2012;61:1000-1016.
10. Pierpaoli C. Artifacts in diffusion MRI. In: D.K. Jones (Ed.), *Diffusion MRI: Theory, Methods, and Applications*. Oxford University Press; 2010.
11. Jezzard P, Balaban RS. Correction for geometric distortion in echo planar images from B_0 field variations. *Magn Reson Med.* 1995;34:65-73.
12. Du YP, Zhou XJ, Bernstein MA. Correction of concomitant magnetic field-induced image artifacts in nonaxial echo-planar imaging. *Magn Reson Med.* 2002;48:509-515.
13. Kennis M, Rooij SV, Kahn RS, Geuze E, Leemans A. Choosing the polarity of the phase-encoding direction in diffusion MRI: does it matter for group analysis? *Neuroimage Clin.* 2016;11:539-547.
14. Wu M, Chang LC, Walker L, et al. Comparison of EPI distortion correction methods in diffusion tensor MRI using a novel framework. Proceedings of MICCAI, New York, NY. 2008;11:321-329.
15. Irfanoglu MO, Walker L, Sarlls J, Marengo S, Pierpaoli C. Effects of image distortions originating from susceptibility variations and concomitant fields on diffusion MRI tractography results. *Neuroimage.* 2012;15:275-288.
16. Walker L, Curry M, Nayak A, Lange N, Pierpaoli C. A framework for the analysis of phantom data in multicenter diffusion tensor imaging studies. *Human Brain Mapping.* 2013;34:2439-2454.
17. Irfanoglu MO, Modi P, Nayak A, Hutchinson EB, Sarlls J, Pierpaoli C. DR-BUDDI: (diffeomorphic registration for blip-up blip-down diffusion imaging) method for correcting echo planar imaging distortions. *Neuroimage.* 2015;106:284-289.
18. Nayak A, Wilde E, Taylor B, et al. A living phantom study to evaluate the echo planar imaging (EPI) distortion correction effects in reducing inter-site variability. In: Proceedings of ISMRM, Paris, France; 2018:1635.
19. Kybic J, Thevenaz P, Nirkko A, Unser M. Unwarping of unidirectionally distorted EPI images. *IEEE Trans Med Imaging.* 2000;19:80-93.
20. Tao R, Fletcher PT, Gerber S, Whitaker RT. A variational image-based approach to the correction of susceptibility artifacts in the alignment of diffusion weighted and structural MRI. *Inform Process Med Imaging.* 2009;21:651-663.
21. Andersson JL, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage.* 2003;20:870-888.
22. Bowtell RW, McIntyre DJO, Commandre MJ, Glover PM, Mansfield P. Correction of geometric distortion in echo planar images. In: Proceedings of 2nd Annual Meeting of the SMR San Francisco, San Francisco, CA; 1994:411.

23. Chang H, Fitzpatrick JM. A Technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. *IEEE TMI*. 1992;11:319-329.
24. Holland D, Kuperman JM, Dale AM. Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging. *Neuroimage*. 2010;50:175-183.
25. Hagler DJ, Hatton S, Cornejo MD, et al. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *NeuroImage*. 2019;202:116091.
26. Miller K, Alfaro-Almagro F, Bangerter N, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. 2016;19:1523-1536.
27. Sotiropoulos SN, Jbabdi S, Xu J, et al. Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage*. 2013;80:125-143.
28. Bastiani M, Andersson JLR, Cordero-Grande L, et al. Automated processing pipeline for neonatal diffusion MRI in the developing Human Connectome Project. *NeuroImage*. 2019;185:750-763.
29. Morgan PS, Bowtell RW, McIntyre DJO, Worthington BS. Correction of spatial distortion in EPI due to inhomogeneous static magnetic fields using reversed gradient method. *J Magn Reson Imaging*. 2004;19:499-507.
30. Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn Reson Med*. 1999;42:515-525.
31. Pierpaoli C, Walker L, Irfanoglu MO, et al. TORTOISE: an integrated software package for processing of diffusion MRI data. In: Proceedings of International Society of Magnetic Resonance in Medicine, Stockholm, Sweden; 2010:1597.
32. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. 2004;23:208-219.
33. Barnett AS, Irfanoglu MO, Rodgers B, Landmann B, Pierpaoli C. Correction of systematic errors in DTI imaging caused by gradient nonlinearity using gradient field maps measured by diffusion imaging of an isotropic diffusion phantom. In: Proceedings of International Society for Magnetic Resonance in Medicine, Montreal, Canada; 2019:3396.
34. Kellner E, Dhital B, Kiselev VG, Reisert M. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magn Reson Med*. 2016;76:1574-1581.
35. Rohde GK, Barnett AS, Basser PJ, Marengo S, Pierpaoli C. Comprehensive approach for correction of motion and distortion in diffusion weighted MRI. *Magn Reson Med*. 2004;51:103-114.
36. Andersson JL, Sotiropoulos SN. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*. 2016;125:1063-1078.
37. Andersson JLR, Graham MS, Zsoldos E, Sotiropoulos SN. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage*. 2016;141:556-572.
38. Andersson JLR, Graham MS, Drobjnak I, Zhang H, Filippini N, Bastiani M. Towards a comprehensive framework for movement and distortion correction of diffusion MR images: within volume movement. *NeuroImage*. 2017;152:450-466.
39. Andersson JLR, Graham MS, Drobjnak I, Zhang H, Campbell J. Susceptibility-induced distortion that varies due to motion: correction in diffusion MR without acquiring additional data. *NeuroImage*. 2018;171:277-295.
40. Özarslan E, Koay CG, Shepherd TM, et al. Mean apparent propagator (MAP) MRI: a novel diffusion imaging method for mapping tissue microstructure. *NeuroImage*. 2013;78:16-32.
41. Basser PJ, Pajevic S. Statistical artifacts in diffusion tensor MRI (DT-MRI) caused by background noise. *Magn Reson Med*. 2000;44:41-50.
42. Irfanoglu MO, Nayak A, Jenkins J, et al. DR-TAMAS: diffeomorphic registration for tensor accurate alignment of anatomical structures. *NeuroImage*. 2016;132:439-454.
43. Chang LC, Jones DK, Pierpaoli C. RESTORE: robust estimation of tensors by outlier rejection. *Magn Reson Med*. 2005;53:1088-1095.
44. Tax CMW, Otte WM, Viergever MA, Dijkhuizen RM, Leemans A. REKINDLE: robust extraction of kurtosis INDices with linear estimation. *Magn Reson Med*. 2015;73:794-808.
45. Qiao Y, Sun W, Shi Y. FOD-based registration for susceptibility distortion correction in connectome imaging. *Connectomics in NeuroImaging*. Cham: Springer International Publishing; 2018:11-19.
46. Raffelt D, Tournier JD, Fripp J, Crozier S, Connelly A, Salvado O. Symmetric diffeomorphic registration of fibre orientation distributions. *NeuroImage*. 2011;56:1171-1180.
47. Mattiello J, Basser PJ, LeBihan D. Analytical calculation of the b matrix in diffusion imaging. In: D. LeBihan (Ed.), *Diffusion and Perfusion Magnetic Resonance Imaging*. Raven Press; 1995:77-90.
48. Neeman M, Freyer JP, Sillerud LO. A simple method for obtaining cross-term-free images for diffusion anisotropy studies in NMR microimaging. *Magn Reson Med*. 1991;21:138-143.

How to cite this article: Irfanoglu MO, Sadeghi N, Sarlls J, Pierpaoli C. Improved reproducibility of diffusion MRI of the human brain with a four-way blip-up and down. *Magn Reson Med*. 2021;85:2696–2708. <https://doi.org/10.1002/mrm.28624>