ORIGINAL ARTICLE

Journal Section

Evaluation of six phase encoding based susceptibility distortion correction methods for diffusion MRI

Xuan Gu[1,2] | Anders Eklund[1,2,3]

[1]Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, Linköping, Sweden

[2]Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden

[3]Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, Linköping, Sweden

Correspondence
Xuan Gu, Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, Linköping, Sweden
Email: xuan.gu@liu.se

Funding information
Swedish Research Council, Grant Number: 2017-04889; Linköping University, Grant Name: Linköping University Center for Industrial Information Technology (CENIIT); ITEA3 / VINNOVA funded project "Intelligence based iMprovement of Personalized treatment And Clinical workflow supporT" (IMPACT)

Purpose: Susceptibility distortion impacts diffusion MRI data analysis and is typically corrected during preprocessing. Correction strategies involve three classes of methods: registration to a structural image, the use of a fieldmap or the use of images acquired with opposing phase encoding directions. It has been demonstrated that phase encoding based methods outperform the other two classes but unfortunately, the choice of which phase encoding based method to use is still an open question due to the absence of any systematic comparisons.

Methods: In this paper we quantitatively evaluated six popular phase encoding based methods for correcting susceptibility distortions in diffusion MRI data. We employed a framework that allows for the simulation of realistic diffusion MRI data with susceptibility distortions. We evaluated the ability for methods to correct distortions by comparing the corrected data with the ground truth. We also validated two popular indirect metrics using both simulated data and real data. The two indirect metrics are the difference between the corrected LR and AP data, and the FA standard deviation over the corrected LR, RL, AP and PA data.

Results: We found that *EPIC*, *HySCO* and *TOPUP* offered better correction than the other three methods. The difference between the corrected LR and AP data provide a good in-

dication of the correction performance. We demonstrated that the FA standard deviation over the corrected LR, RL, AP and PA data gave a different ordering of correction quality than the direct metric.

Conclusion: We suggest researchers to use *EPIC*, *HySCO* and *TOPUP* for susceptibility distortion correction. We also suggest that indirect metrics must be interpreted cautiously when evaluating methods for correcting susceptibility distortions in diffusion MRI data.

KEYWORDS

Susceptibility distortion, diffusion MRI

1 | INTRODUCTION

Analysis of diffusion MRI data is confounded by the presence of susceptibility distortions, caused by an off-resonance field induced by differences in magnetic susceptibility at the air-tissue interface. There are a number of techniques available for correcting susceptibility distortions. Broadly, these techniques can be divided into three types: registration based (RB) methods, fieldmap based (FB) methods and phase encoding based (PB) methods. The first approach involves registration of the distorted image to a structural image without distortions. The second approach involves estimating a map of the b_0 inhomogeneities, and using this along with information about the diffusion acquisition protocol to correct for the distortions. The third approach is based on estimating the underlying distortions using additional data acquired with a different phase encoding direction. For example, it is common to collect LR (left right) and RL (right left) images, or AP (anterior posterior) and PA (posterior anterior) images. Phase encoding based techniques have been demonstrated to outperform the other two approaches (Graham et al., 2017; Esteban et al., 2014), at the cost of doubling the scan time.

There are many software packages providing phase encoding based tools for correcting susceptibility distortions, e.g., *animaDistortionCorrection* (aDC) (Voss et al., 2006), *animaBMDistortionCorrection* (aBMDC) (Hedouin et al., 2017), *DR-BUDDI* (Irfanoglu et al., 2015), *EPIC* (Holland et al., 2010), *HySCO* (Ruthotto et al., 2013) and *TOPUP* (Andersson et al., 2003), summarized in Table 1. To date, there is no systematic comparison of existing phase encoding based methods for susceptibility distortion correction. See Table 2 for an overview of previous comparisons of different distortion correction tools.

Tool	Software package	Webpage	Reference
animaDistortionCorrection	ANIMA	https://github.com/Inria-Visages/Anima-Public/wiki/Registration-tools#epi-distortion-correction	Voss et al. (2006)
animaBMDistortionCorrection	ANIMA	https://github.com/Inria-Visages/Anima-Public/wiki/Registration-tools#epi-distortion-correction	Hedouin et al. (2017)
DR-BUDDI	TORTOISE	https://tortoise.nibib.nih.gov/tortoise/v313/10-step-31-after-diffprepdr-buddi	Irfanoglu et al. (2015)
EPIC		https://github.com/dominicholland/EPIC	Holland et al. (2010)
HySCO	SPM	https://bitbucket.org/siawoosh/acid-artefact-correction-in-diffusion-mri/wiki/ACIDSC_wiki_hysco	Ruthotto et al. (2013)
TOPUP	FSL	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/topup/TopupUsersGuide	Andersson et al. (2003)

TABLE 1 Phase encoding based susceptibility distortion correction tools evaluated in this paper.

The lack of ground truth means that evaluations are typically indirect or qualitative (Jezzard and Balaban, 1995;

Wu et al., 2008; Bhushan et al., 2012; Ruthotto et al., 2013; Fritz et al., 2014; Irfanoglu et al., 2015; Taylor et al., 2016; Hedouin et al., 2017; Wang et al., 2017; Irfanoglu et al., 2018). Only a few investigations have been carried out with the presence of a ground truth for evaluation of susceptibility distortion correction (Andersson et al., 2003; Esteban et al., 2014; Graham et al., 2017). Hedouin et al. (2017) compared *aDC*, *aBMDC* and *TOPUP* using phantom data and human data. For the phantom data, both the *aBMDC* and *TOPUP* corrected images appear visually similar for correcting the distortion, while *aDC* gives visually poorer results. *aBMDC* outperformed *aDC* and *TOPUP* by obtaining smaller landmark position errors. For human data, images corrected using *aDC* contain a mismatch around the lateral ventricles compared with respect to a structural (T1-weighted) image . *aBMDC* and *TOPUP* both obtain a corrected image very close to the structural T1-weighted image. *aBMDC* and *TOPUP* show a very high similarity between the two corrected images C_{AP+PA} and C_{LR+RL}, outperforming *aDC*. Irfanoglu et al. (2015) compared *EPIC*, *DR-BUDDI* and *TOPUP* using human data. *DR-BUDDI* produced sharper images than *EPIC* and *TOPUP*, showing clearly visible tissue interfaces. Areas such as the inferior temporal lobes and the olfactory bulbs were more accurately reconstructed by *DR-BUDDI* than *EPIC* and *TOPUP*. *DR-BUDDI* resulted in the lowest variability between the two corrected images C_{AP+PA} and C_{LR+RL}, followed by *TOPUP* and then *EPIC*. Overall, *DR-BUDDI* corrected images showed a higher correlation with the undistorted T2-weighted image than did *EPIC* and *TOPUP*.

In this work, we undertake a comparison of six phase encoding based methods for susceptibility distortion correction using both simulated diffusion data and real diffusion data, see Table 2 for differences between our study and previous comparisons. We used the POSSUM (Drobnjak et al., 2006, 2010) based diffusion MRI simulator (Graham et al., 2016, 2017), in order to produce realistic diffusion data with susceptibility distortions typically seen in real data. Simulated data can provide ground truth that enables direct and quantitative evaluation. Our analysis directly measures the ability to correctly recover distortion-free data by comparing the corrected b_0 image with its ground truth. We also investigate the suitability of two commonly used indirect metrics, i.e. the difference of the corrected data from the LRRL and APPA pairs (Ruthotto et al., 2013; Graham et al., 2017), and the FA standard deviation over the corrected LR, RL, AP and PA data (Wu et al., 2008; Irfanoglu et al., 2015). We hope that this work will enable researchers to make more carefully informed choices when designing their processing pipelines.

Method class	Ground truth availability	Software	Dataset	Conclusion	Reference
RB, FB	No	RB: In-house C++ code using ITK FB: *FSL* function *PRELUDE* and *FUGUE*	5 human subjects	RB showed an overall better performance than FB RB and FB showed different performance in different brain regions	Wu et al. (2008)
FB, PB	No	FB: *FSL* function *PRELUDE* and *FUGUE* PB: *EPIC*	Human subjects	PB provided superior accuracy than FB	Holland et al. (2010)
FB, PB	No	FB: *SPM* fieldmap toolbox PB: *SPM* function *HySCO*	1 human subject	FB gave a better performance than RB	Ruthotto et al. (2013)
RB, FB, PB	Yes	RB: *ANTs* FB: In-house code PB: *FSL* function *TOPUP*	A simulated phantom dataset	PB is the most accurate method	Esteban et al. (2014)
FB, PB	No	FB: *SPM* fieldmap toolbox PB: *FSL* function *TOPUP* and *SPM* function *HySCO*	4 human subjects	PB outperformed FB *TOPUP* outperformed *HySCO*	Fritz et al. (2014)
PB	No	*TORTOISE* function *DR-BUDDI* *EPIC* *FSL* function *TOPUP*	12 human subjects 1 mouse dataset	*DR-BUDDI* performed the best	Irfanoglu et al. (2015)
RB, FB, PB	Yes	RB: *NiftyReg* function *reg_f3d* FB: *FSL* function *PRELUDE* and *FUGUE* PB: *FSL* function *TOPUP*	A simulated dataset 10 human subjects	FB and PB outperformed RB FB was sensitive to partial volume with air	Graham et al. (2017)
PB	No	*ANIMA* function *animaDistortionCorrection* *ANIMA* function *animaBMDistortionCorrection* *FSL* function *TOPUP*	A phantom dataset 5 human subjects	*animaBMDistortionCorrection* performed the best	Hedouin et al. (2017)
RB, FB	No	RB: *ANTs* function *SyN* FB: *FSL* function *FUGUE*	71 human subjects	RB resulted in higher reliability	Wang et al. (2017)
PB	Yes	*ANIMA* function *animaDistortionCorrection* *ANIMA* function *animaBMDistortionCorrection* *TORTOISE* function *DR-BUDDI* *EPIC* *SPM* function *HySCO* *FSL* function *TOPUP*	5 simulated datasets 40 human subjects	*HySCO* and *TOPUP* outperformed the other four methods	This paper

TABLE 2 A list of susceptibility distortion correction evaluation papers.
RB: registration based methods, FB: fieldmap based methods, PB: phase encoding based methods.

2 | DATA

2.1 | Simulated data

The diffusion data was simulated with 11 volumes of $b = 700$ s/mm^2, 12 volumes of $b = 2000$ s/mm^2 and 1 volume of $b = 0$. The input to the POSSUM diffusion MRI simulator is a collection of three 3D anatomical volumes: grey matter, white matter and cerebro-spinal fluid (CSF). The voxel values in these segmentations reflect the proportion of tissue present in each voxel, in the range [0,1]. The input was generated from the T1-weighted structural image from HCP using the *FSL* function *FAST* (Zhang et al., 2001). The representation of diffusion weighting was achieved by a spherical harmonic fit of order $n = 8$ to the $b = 1000$ s/mm^2 shell of the diffusion data, using the *Dipy* (Garyfallidis et al., 2014) module *reconst.shm*. The MR parameters used are listed in Table 3. We used a matrix size of 72×86, 55 slices and a voxel size of 2.5 mm isotropic. The TE was 109 ms, the TR was 700 ms and the flip-angle was 90°. The fieldmap was generated from one phase difference volume and two magnitude volumes (one for each echo time) from HCP using the *FSL* function *fsl_prepare_fieldmap* (Jenkinson, 2003). To generate a tight brain extration for *fsl_prepare_fieldmap*, the brain mask created by the *FSL* function *BET* (Smith, 2002) was further eroded using a 5 mm box kernel. The generated fieldmap was linearly registered to the T1-weighted structural image using the *FSL* function *FLIRT* (Jenkinson and Smith, 2001; Jenkinson et al., 2002). Diffusion data was simulated with four PE directions, i.e., left-right (LR), right-left (RL), anterior-posterior (AP) and posterior-anterior (PA). No other distortions (e.g. eddy-currents and head motion) were included in the simulations. We also simulated a ground truth set, acquired with the same acquisition parameters but no input susceptibility fieldmap. We simulated diffusion data for five subjects (100206, 100307, 100408, 100610, 101006) from the Human Connectome Project (HCP)[1] (Van Essen et al., 2013; Glasser et al., 2013).

	T1 (ms)	T2* (ms)	Spin density	Chemical shift	T2 (ms)
GM	1331	51	0.86	0	75
WM	832	44	0.77	0	70
CSF	3700	500	1	0	500

TABLE 3 A list of MR parameters (relaxation times T1, T2*, spin density, and chemical shift value) used in our simulations within POSSUM.

2.2 | Real data

We used 40 subjects from the developing HCP project (Hughes et al., 2017; Bastiani et al., 2019). It provides diffusion data acquired with four PE directions: AP, PA, LR and RL, enabling evaluation using the indirect metric (i.e. comparing APPA corrected to LRRL corrected). The data was acquired on a 3T Philips Achieva scanner and consists of 4 shells: 20 volumes of $b = 0$, 64 volumes of $b = 400$ s/mm^2, 88 volumes of $b = 1000$ s/mm^2 and 128 volumes of $b = 2600$ s/mm^2. The data was acquired using TR=3800 ms and TE=90 ms. The matrix size is 128×128, the number of slices is 64 and the acquired voxel size is $1.17 \times 1.17 \times 1.5$ mm.

[1] Data collection and sharing for this project was provided by the Human Connectome Project (U01-MH93765) (HCP; Principal Investigators: Bruce Rosen, M.D., Ph.D., Arthur W. Toga, Ph.D., Van J.Weeden, MD). HCP funding was provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS). HCP data are disseminated by the Laboratory of Neuro Imaging at the University of Southern California.

3 | METHODS

For the simulated data, we used the *FSL* function *BET* (Smith, 2002) to create the brain mask from the distortion-free b_0 image. Diffusion tensor fitting and FA calculation were performed using the *FSL* function *dtifit*. For simulated data, we evaluated the ability of each method to recover the correct intensity at each voxel, by computing error maps between the distortion corrected data and ground truth. We also investigated two indirect metrics. One is the difference of the corrected data from the LRRL and APPA pairs (Ruthotto et al., 2013; Graham et al., 2017), and the other is the FA standard deviation over the corrected LR, RL, AP and PA data (Wu et al., 2008; Irfanoglu et al., 2015). Comparing the uncertainty of data from different preprocessing pipelines is a way to determine if one pipeline is better than the other (Sjölund et al., 2018; Gu et al., 2019). For the real data, we used the brain mask provided with the dataset. We corrected for head motion between LR, RL, AP and PA scans by registering all volumes to the first volume using the *FSL* function *FLIRT* (Jenkinson and Smith, 2001; Jenkinson et al., 2002). For real data, we used indirect evaluation. We compared the corrected data from the LRRL pair with the result from the APPA pair, since ideally the corrected results from the two pairs should be the same. For each susceptibility distortion correction tool we used default settings and steps provided by the software's basic help documentation. There is no method referred to in any of the *EPIC* documentation for choosing an alternate phase direction than the y-direction (or the LRRL direction with regards to this study). We manually rotated the LR and RL data 90 degrees in the $x - y$ plane before feeding them to *EPIC*. The data was finally rotated 90 degrees in the other direction after correction. We share our processing scripts on Github [2], such that other researchers can reproduce and extend our findings (Eklund et al., 2017).

4 | RESULTS

4.1 | Simulated data

Figure 1 shows the simulated diffusion data and the fieldmap for HCP Subject 100206, the simulated distortions look very realistic. We investigated how different levels of distortion affect the correction performance by simulating three levels of susceptibility distortion, as shown in Figure 2. The distortion level was controlled by dividing the fieldmap by a factor 1, 2 or 4.

4.1.1 | Direct metric

Figure 3 shows the corrected b_0 images using six different methods, along with error maps obtained by calculating the difference compared to the ground truth images. Correction was carried out for LRRL and APPA pairs, respectively, and we used the corrected LR and AP images as the results for the two pairs. For the corrected LR image, *aDC* and *aBMDC* show larger errors for edge voxels, compared to their performance in the LR case. *DR-BUDDI* shows even large errors, not only for edge voxels but also inside the brain. It is clear that *DR-BUDDI* is unable to accurately recover the b_0 image for both LR and AP cases. *EPIC*, *HySCO* and *TOPUP* produced very small errors for both LR and AP cases, mainly along edges.

Figure 4 shows the error maps using the six methods for the three levels of distortion. Correction was carried out for LRRL and APPA pairs, respectively. In addition to visual inspection, we computed the mean absolute error (MAE) (red squares) and the mean squared error (MSE) (red cross) within the brain and their standard deviations (blue bars) for five simulated HCP subjects using the six methods, as shown in Figure 5. The results confirmed what we observed in

[2] https://github.com/xuagu37/SusceptibilityDistortionCorrection

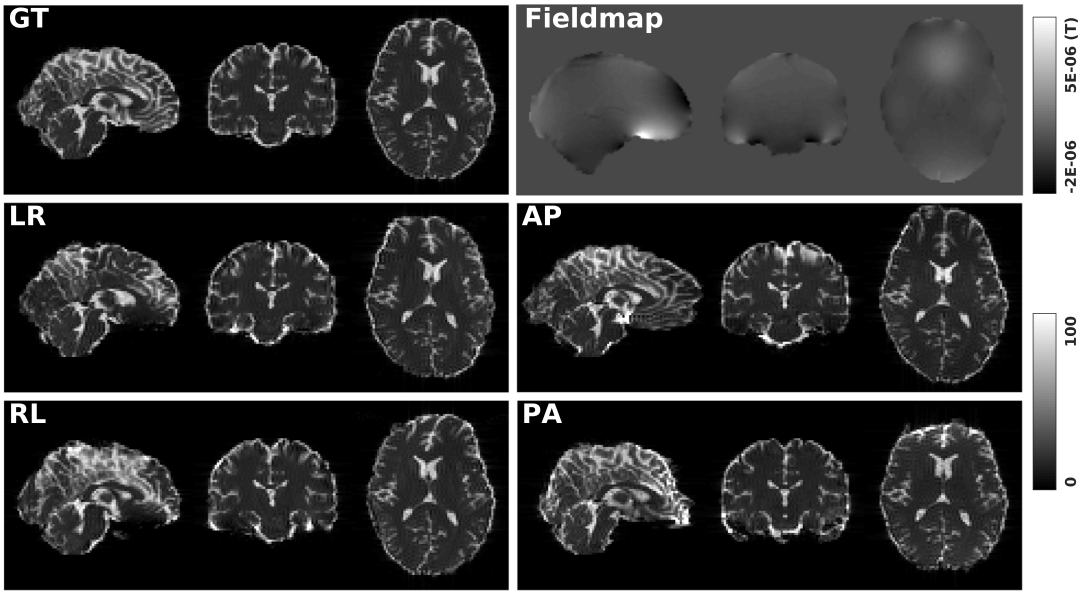

FIGURE 1 GT: simulated diffusion data (without distortions) for HCP subject 100206. Fieldmap: the real fieldmap used to simulate the distortions. LR: simulated diffusion data with LR distortion. RL: simulated diffusion data with RL distortion. AP: simulated diffusion data with AP distortion. PA: simulated diffusion data with PA distortion.

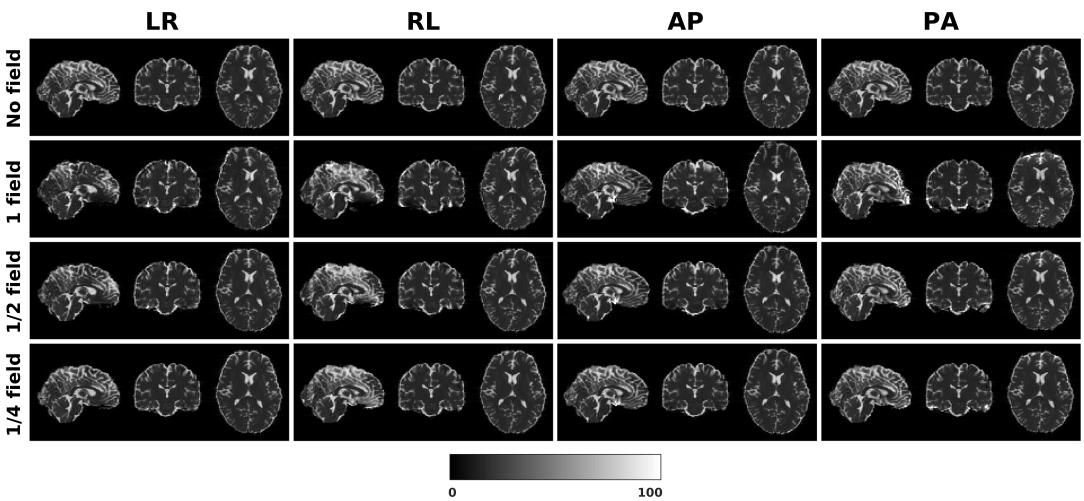

FIGURE 2 Three levels of susceptibility distortion were simulated. 1 field is the original fieldmap. 1/2 field is the original fieldmap divided by 2. 1/4 field is the original fieldmap divided by 4.

Figure 4 and quantitatively demonstrates the accuracy and robustness of the six methods. The MAE and MSE decreased with decreasing field strength, aligned with our predictions.

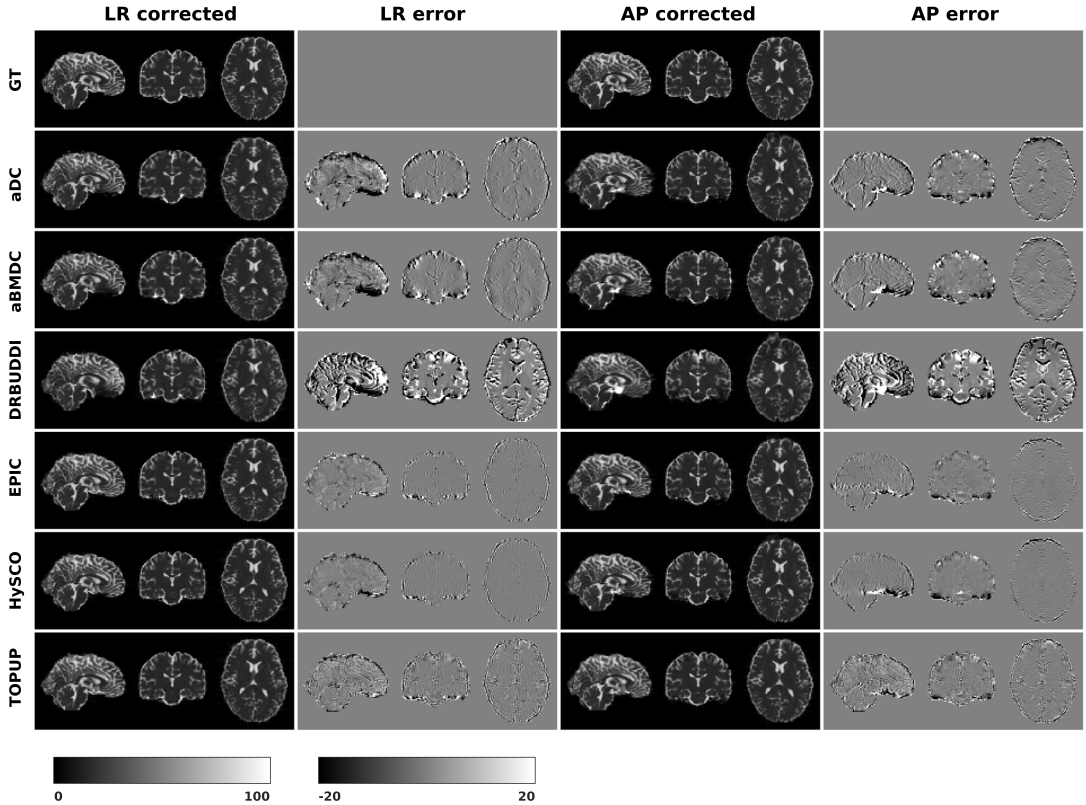

FIGURE 3 Corrected data for simulated HCP Subject 100206 using the six methods. Error maps were obtained by calculating the difference compared to the ground truth. Correction was carried out for LRRL and APPA pairs, respectively. Note that aDC and aBMDC work better for APPA correction, compared to LRRL correction.

4.1.2 | Indirect metric

To investigate the suitability of LRRL-APPA differences as an indirect metric we plot the corrected LR and AP data, and their differences, as shown in Figure 6. Ideally, the corrected LR and the corrected AP would be identical. The whole-brain mean absolute difference (MAD) and mean squared difference (MSD) were computed for every correction method, as shown in Figure 7. The results show large MAD and MSE for *aDC*, *aBMDC* and *DR-BUDDI*. *EPIC*, *HySCO* and *TOPUP* performed almost equally well and *TOPUP* produced a slightly smaller standard deviation over subjects. The results are not completely consistent with previous results in Figure 3 and 5. These results demonstrate that the indirect metric (difference maps) shows a different ordering of correction performance compared to the direct metric (error maps).

To investigate the suitability of FA standard deviation as an indirect metric we plot the FA standard deviation over the corrected LR, RL, AP and PA data, as shown in Figure 8. Ideally, the corrected LR, RL, AP and PA would be identical, which would make a zero FA standard deviation. The whole-brain mean of the standard deviation was computed for every correction method, as shown in Figure 9. The results show large standard deviations for *aDC* and *aBMDC*. *TOPUP* produces the smallest FA standard deviation and it is also the most robust over different subjects.

4.2 | Real data

Figure 10 shows the distorted diffusion data from a dHCP subject scanned with four different PE directions. The distortions are more pronounced in regions close to tissue-air interfaces, such as the frontal poles and the temporal lobes near the petrous bone. To evaluate the six methods, we computed the difference between the corrected LR and AP for the dHCP data, as shown in Figure 11. Higher difference values are visible in regions most affected by magnetic susceptibility variations such as the boundary regions of the brain. Figure 12 reports the whole-brain mean of the difference, for the 40 processed subjects. *EPIC*, *HySCO* and *TOPUP* performed better than the other three methods, which resembled the results for simulated data in Figure 7.

The processing time of one simulated HCP dataset for the six softwares are 2.8 (*aDC*), 35.5 (*aBMDC*), 1370.5 (*DR-BUDDI*), 38.0 (*EPIC*), 8.1 (*HySCO*) and 195.7 (*TOPUP*) seconds, respectively. Please note that *aBMDC*, *DR-BUDDI*, *EPIC* and *HySCO* use several CPU threads to speedup the processing. We used a computer with 32 GB RAM and an Intel(R) Xeon(R) Silver 4114 2.20 GHz CPU (containing 10 cores, which can run 20 threads in parallel).

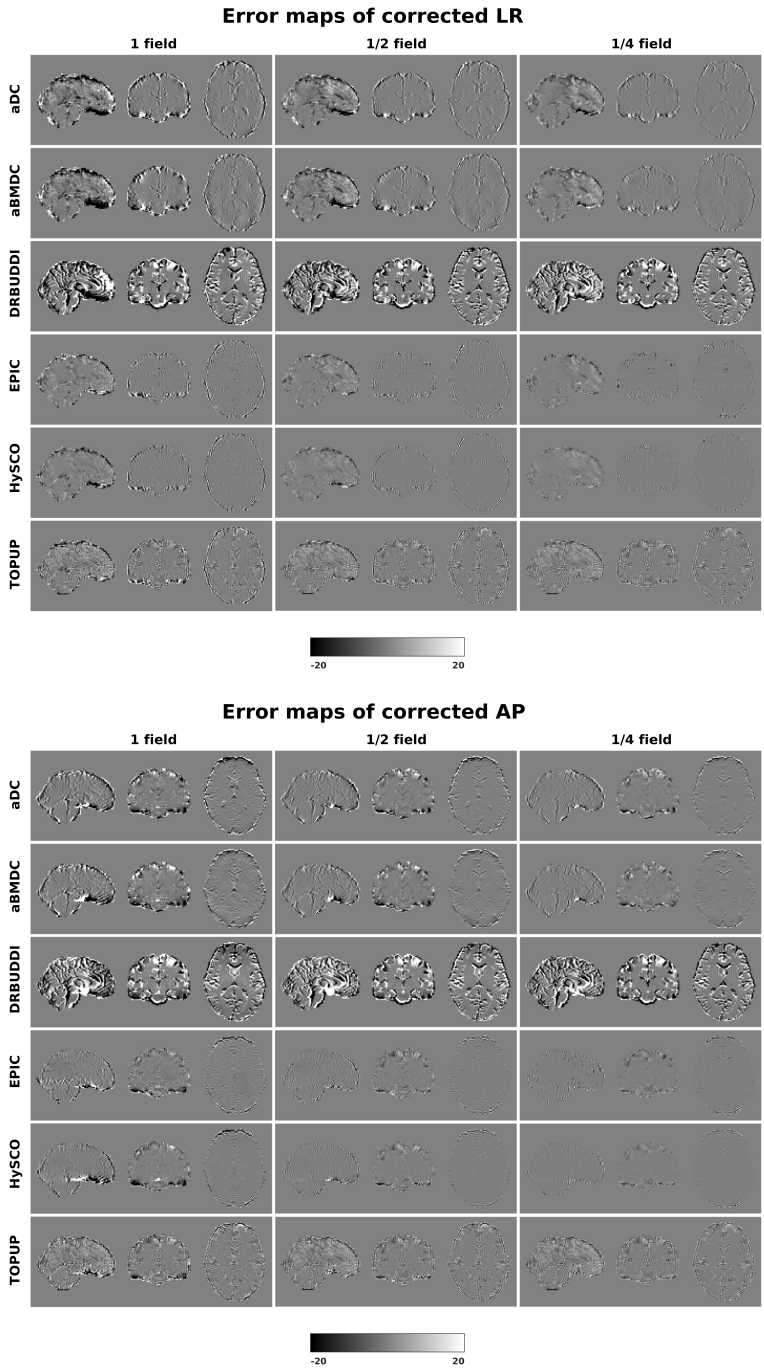

FIGURE 4 Error maps for simulated HCP Subject 100206 using the six methods. Three levels of susceptibility distortion were simulated. Correction was carried out for LRRL (left) and APPA (right) pairs, respectively.

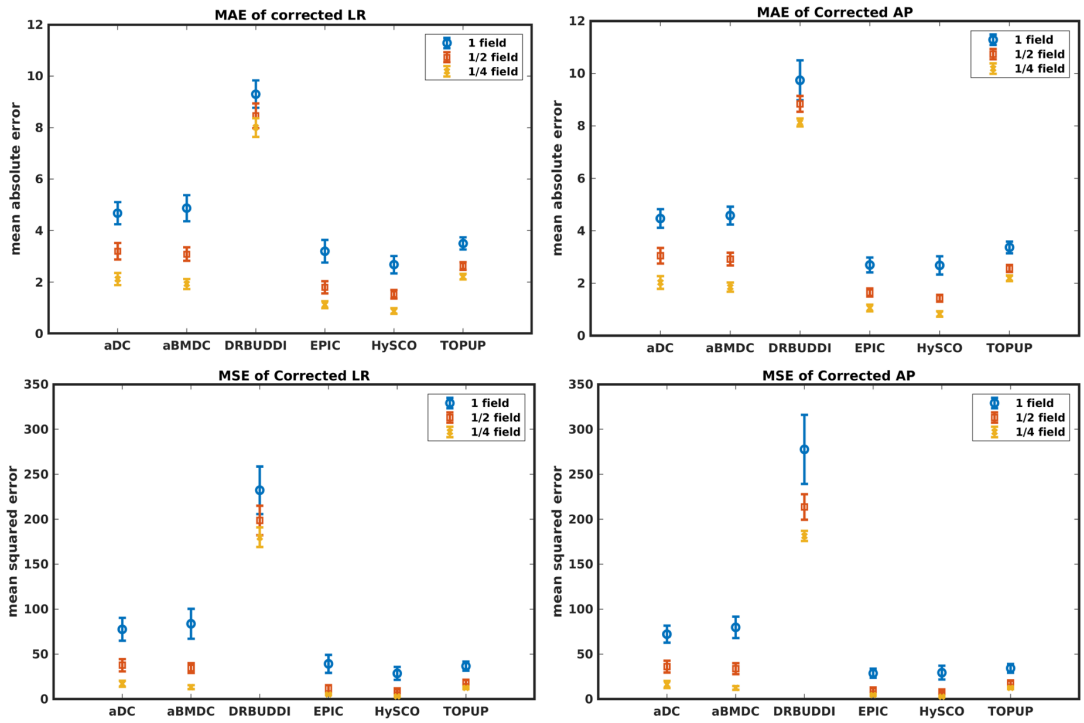

FIGURE 5 MAE (top) and MSE (bottom) for five simulated HCP subjects using the six methods. The error bars represent standard deviation over subjects. Correction was carried out for LRRL (left) and APPA (right) pairs, respectively. Three levels of susceptibility distortion were simulated.

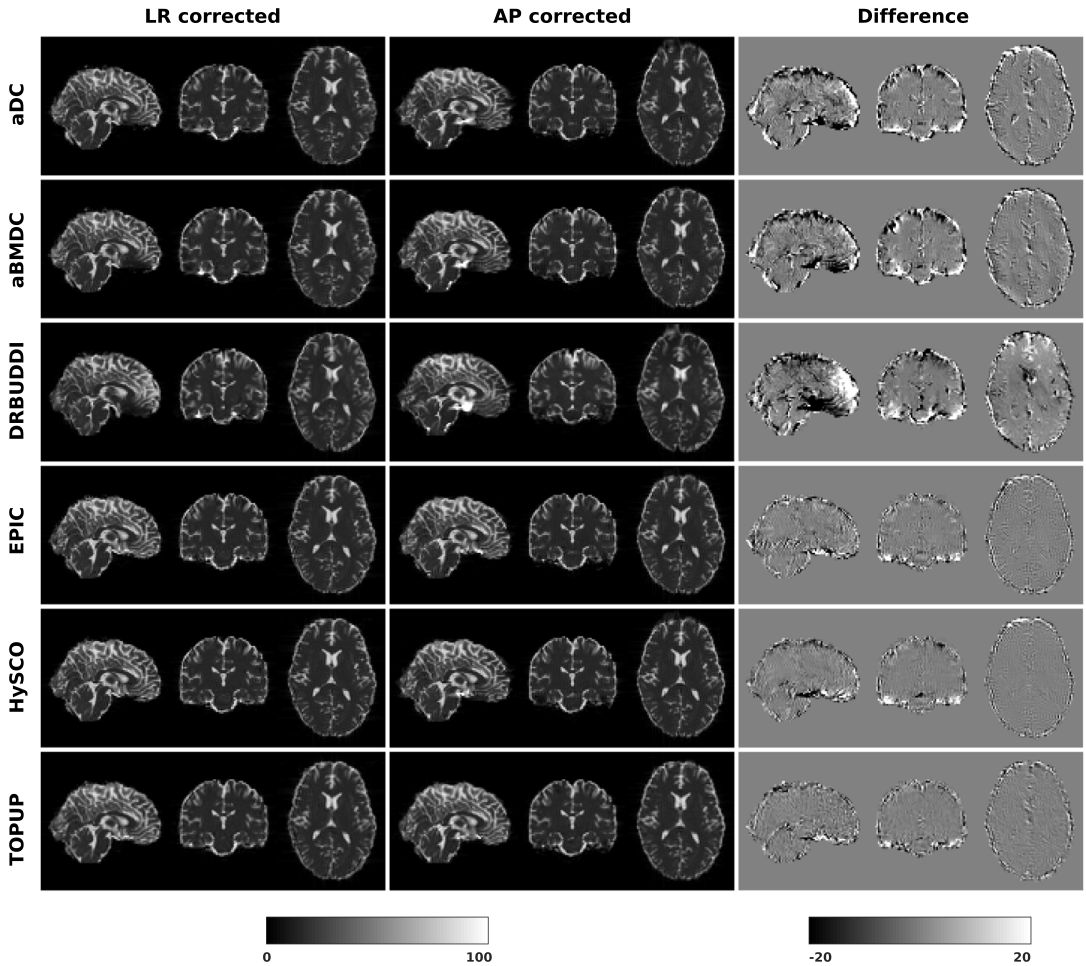

FIGURE 6 Difference between corrected LR and AP for simulated HCP Subject 100206 using the six methods. *TOPUP* provides the smallest difference between the two corrections.

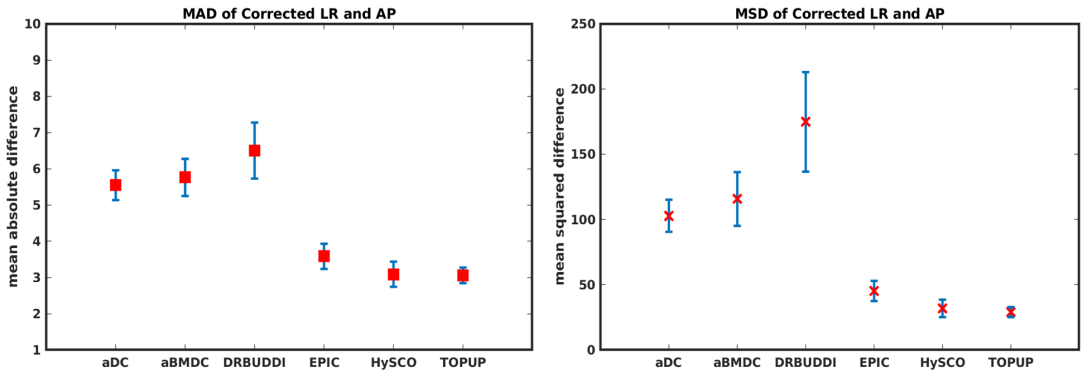

FIGURE 7 MAD and MSD between corrected LR and AP for five simulated HCP subjects using the six methods. The error bars represent the standard deviation over subjects.

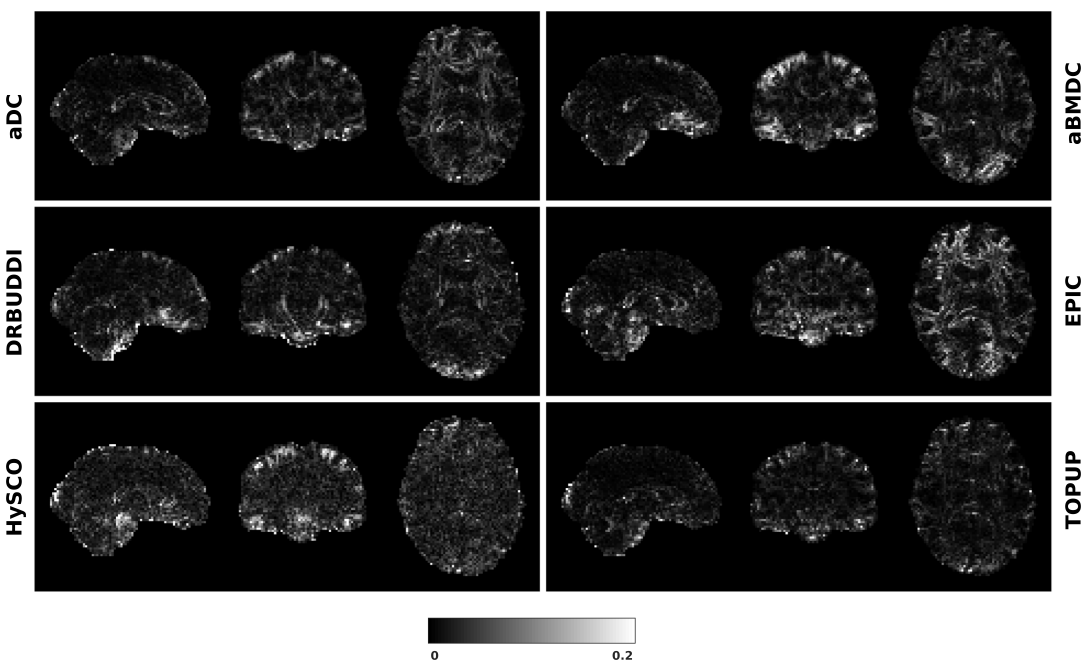

FIGURE 8 FA standard deviation over the corrected LR, RL, AP and PA data for simulated HCP Subject 100206 using the six methods. *Topup* produces the smallest standard deviation.

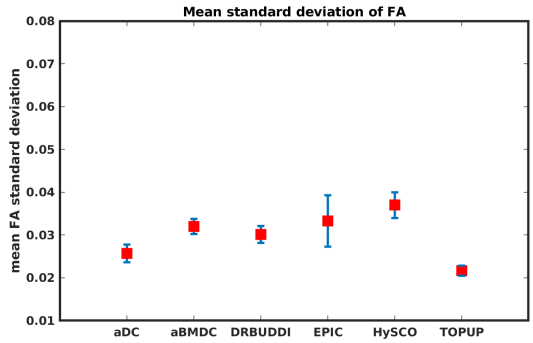

FIGURE 9 Mean FA standard deviation for five simulated HCP subjects using the six methods. The error bars represent standard deviation over subjects.

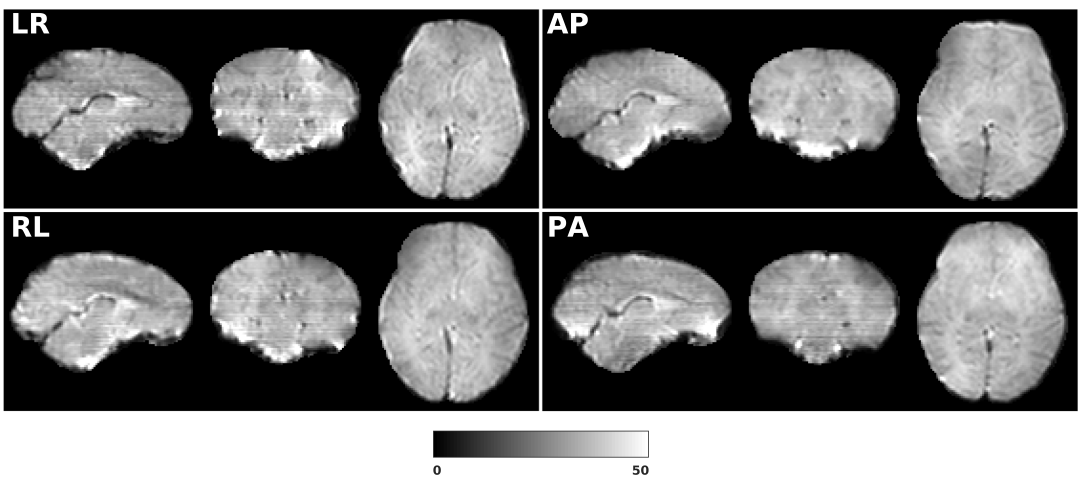

FIGURE 10 LRRL and APPA pairs for dHCP Subject CC00069XX12.

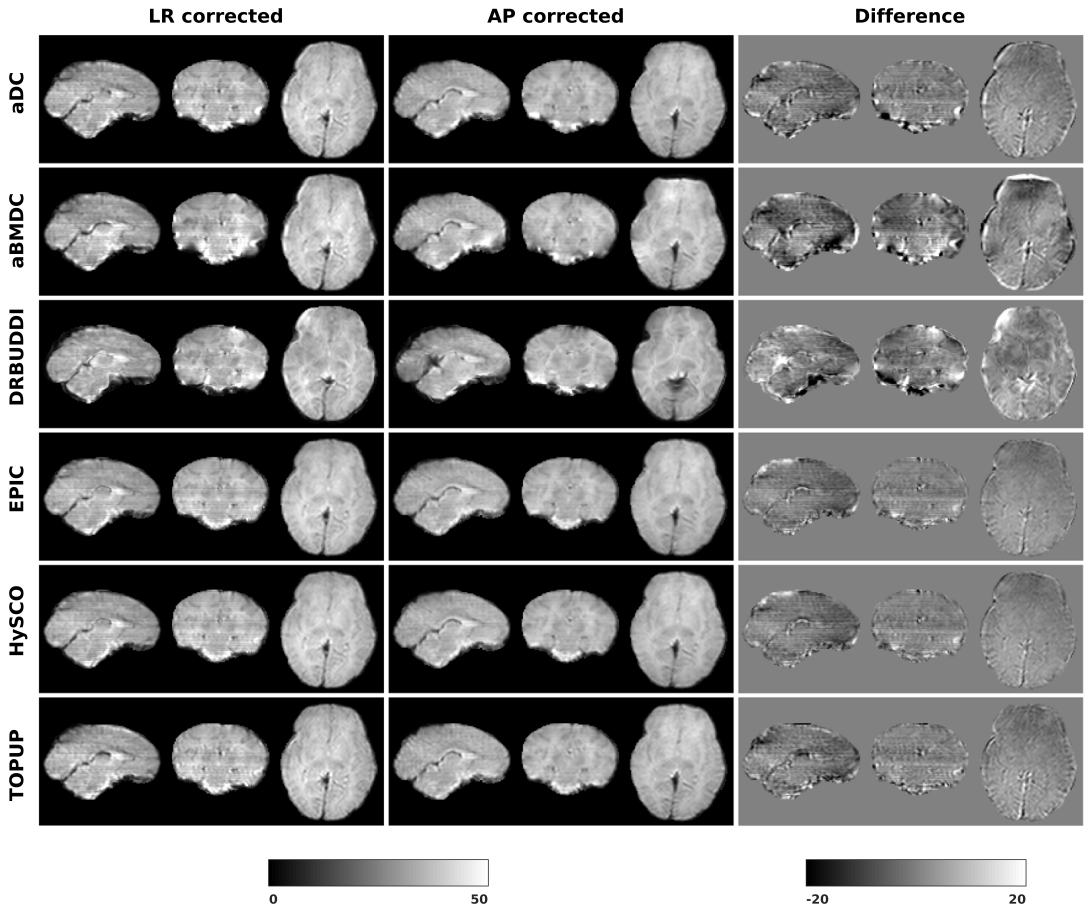

FIGURE 11 Difference between corrected data from LRRL and APPA pairs for dHCP Subject CC00069XX12 using the six methods.

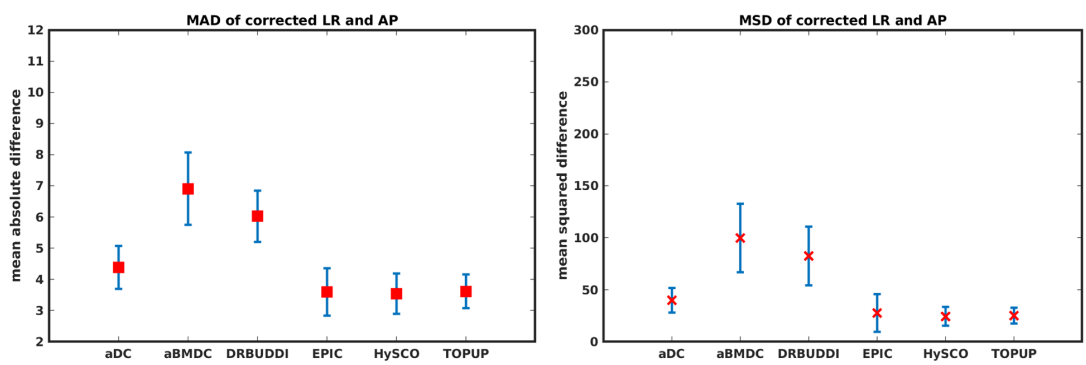

FIGURE 12 MAD and MSD between corrected data from LRRL and APPA pairs for 40 dHCP subjects using the six methods. The error bars represent the standard deviation over subjects. In all cases *TOPUP* produces the smallest difference.

5 | DISCUSSION

In this paper, we used both simulated and real data to evaluate six phase encoding based methods for correcting susceptibility distortions. This work is important given that phase encoding based methods have been demonstrated to outperform the other two classes of approaches, and are very frequently used in diffusion data analysis pipelines. It is thus essential to carefully evaluate phase encoding based correction techniques and their limitations. By this work we aim to answer the following two questions. Which method of the six provides the best distortion correction? Are the indirect metrics suitable for measuring the distortion correction performance?

Before comparing the quality of the corrections provided by each of these tools, it is important to note that each tool has its customizable parameters. The default parameter settings were used in this work, as it was reasoned that this would be representative of the way the tools were most often used. It was also reasoned that changing default parameters or influencing the inputs prior to correction would lead to a less fair comparison. It is therefore possible that the corrections in this work may not represent the best corrections attainable by use of these tools.

The error map directly measures the ability to correctly recover distortion-free data for different methods. Based on our experiments, we found that *EPIC*, *HySCO* and *TOPUP* were generally superior to the other three methods in achieving better performance for the direct metric, see Figure 5, and *HySCO* is more computationally efficient. *EPIC*, *HySCO* and *TOPUP* demonstrated almost equal results but the order of performance is slightly different for different PE pairs. However, there is no method referred to in any of the *EPIC* documentation for choosing an alternate phase direction than the y-direction (or the LRRL direction with regards to this study). It is indeed often the case that the phase encoding direction is chosen to be the y-direction, however there are many images where this is not the case. In this respect, *TOPUP* and *HySCO* offer more flexibility without the need to manually rotate the LR and RL data 90 degrees in the $x - y$ plane. *TOPUP* was found to work only for data volumes with even size on x, y and z directions since the default configuration file (*b02b0.cnf*) performs a factor of 2 subsampling. There are two solutions suggested to eliminate this problem; either cropping or adding dummy data to the 3D volume. *HySCO* was found to be the easiest correction tool to use, and in Figure 5 it even slightly outperformed *EPIC* and *TOPUP*.

Indirect metrics are often used to evaluate distortion correction for real data due to the absence of ground truth. We investigated the ability of two indirect metrics to measure the correction quality. We used the simulated data to validate two of the most promising indirect metrics for correction quality, i.e., the difference between corrected LR and AP data, and the FA standard deviation over the corrected LR, RL, AP and PA data. The first indirect metric, as shown in Figure 6 and 11, roughly confirmed what we observed for the direct metric in Figure 5. However, we also found that the indirect metrics can sometimes provide different conclusions compared to direct metrics. For example, the FA standard deviation over the corrected LR, RL, AP and PA data can appear small despite the data having been poorly corrected, see Figure 9 and 4. The reason might be that the FA values are not affected greatly by the error which mostly occurs at the edge voxels (and FA is typically low close to the edge of the brain). Therefore, the FA standard deviation should not be interpreted as a measure of distortion correction on its own. We therefore suggest that indirect metrics must be interpreted cautiously. We have here focused on susceptibility distortions, but real diffusion data are more complicated as head motion and eddy currents will introduce more artifacts. These additional artifacts may explain why direct and indirect metrics provide different conclusions regarding the performance of the six methods.

ACKNOWLEDGEMENTS

This research was supported by the Swedish Research Council (grant 2017-04889), Linköping University Center for Industrial Information Technology (CENIIT) and the ITEA3 / VINNOVA funded project ”Intelligence based iMprovement

of Personalized treatment And Clinical workflow supporT" (IMPACT). We used the data provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. Part of these results were obtained using data made available from the Developing Human Connectome Project funded by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. [319456]. The authors would like to thank Mark Graham for providing help with the POSSUM diffusion MRI simulator.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Andersson, J. L., Skare, S. and Ashburner, J. (2003) How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*, **20**, 870–888.

Bastiani, M., Andersson, J. L., Cordero-Grande, L., Murgasova, M., Hutter, J., Price, A. N., Makropoulos, A., Fitzgibbon, S. P., Hughes, E., Rueckert, D. et al. (2019) Automated processing pipeline for neonatal diffusion MRI in the developing Human Connectome Project. *NeuroImage*, **185**, 750–763.

Bhushan, C., Haldar, J. P., Joshi, A. A. and Leahy, R. M. (2012) Correcting susceptibility-induced distortion in diffusion-weighted MRI using constrained nonrigid registration. In *Proceedings of the Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 1–9. IEEE.

Drobnjak, I., Gavaghan, D., Süli, E., Pitt-Francis, J. and Jenkinson, M. (2006) Development of a functional magnetic resonance imaging simulator for modeling realistic rigid-body motion artifacts. *Magnetic Resonance in Medicine*, **56**, 364–380.

Drobnjak, I., Pell, G. S. and Jenkinson, M. (2010) Simulating the effects of time-varying magnetic fields with a realistic simulated scanner. *Magnetic Resonance Imaging*, **28**, 1014–1021.

Eklund, A., Nichols, T. E. and Knutsson, H. (2017) Reply to brown and behrmann, cox, et al., and kessler et al.: Data and code sharing is the way forward for fMRI. *Proceedings of the National Academy of Sciences*, **114**, E3374–E3375.

Esteban, O., Daducci, A., Caruyer, E., O'Brien, K., Ledesma-Carbayo, M. J., Bach-Cuadra, M. and Santos, A. (2014) Simulation-based evaluation of susceptibility distortion correction methods in diffusion MRI for connectivity analysis. In *International Symposium on Biomedical Imaging*, 738–741.

Fritz, L., Mulders, J., Breman, H., Peters, J., Bastiani, M., Roebroeck, A., Andersson, J., Ashburner, J., Weiskopf, N. and Goebel, R. (2014) Comparison of EPI distortion correction methods at 3T and 7T. In *Annual Meeting of the Organization for Human Brain Mapping*.

Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M. and Nimmo-Smith, I. (2014) Dipy, a library for the analysis of diffusion MRI data. *Frontiers in Neuroinformatics*, **8**, 8.

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. et al. (2013) The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, **80**, 105–124.

Graham, M. S., Drobnjak, I., Jenkinson, M. and Zhang, H. (2017) Quantitative assessment of the susceptibility artefact and its interaction with motion in diffusion MRI. *PLOS One*, **12**, e0185647.

Graham, M. S., Drobnjak, I. and Zhang, H. (2016) Realistic simulation of artefacts in diffusion MRI for validating post-processing correction techniques. *NeuroImage*, **125**, 1079–1094.

Gu, X., Eklund, A., Özarslan, E. and Knutsson, H. (2019) Using the wild bootstrap to quantify uncertainty in mean apparent propagator MRI. *Frontiers in Neuroinformatics*, **13**, 43.

Hedouin, R., Commowick, O., Bannier, E., Scherrer, B., Taquet, M., Warfield, S. K. and Barillot, C. (2017) Block-matching distortion correction of echo-planar images with opposite phase encoding directions. *IEEE Transactions on Medical Imaging*, **36**, 1106–1115.

Holland, D., Kuperman, J. M. and Dale, A. M. (2010) Efficient correction of inhomogeneous static magnetic field-induced distortion in echo planar imaging. *NeuroImage*, **50**, 175–183.

Hughes, E., Cordero-Grande, L., Murgasova, M., Hutter, J., Price, A., Gomes, A. D. S., Allsop, J., Steinweg, J., Tusor, N., Wurie, J. et al. (2017) The Developing Human Connectome: announcing the first release of open access neonatal brain imaging. In *Annual Meeting of the Organization for Human Brain Mapping*.

Irfanoglu, M. O., Modi, P., Nayak, A., Hutchinson, E. B., Sarlls, J. and Pierpaoli, C. (2015) DR-BUDDI (diffeomorphic registration for blip-up blip-down diffusion imaging) method for correcting echo planar imaging distortions. *NeuroImage*, **106**, 284–299.

Irfanoglu, M. O., Sarlls, J., Nayak, A. and Pierpaoli, C. (2018) Evaluating corrections for eddy-currents and other EPI distortions in diffusion MRI: methodology and a dataset for benchmarking. *Magnetic Resonance in Medicine*, **81**, 2774–2787.

Jenkinson, M. (2003) Fast, automated, n-dimensional phase-unwrapping algorithm. *Magnetic Resonance in Medicine*, **49**, 193–197.

Jenkinson, M., Bannister, P., Brady, M. and Smith, S. (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, **17**, 825–841.

Jenkinson, M. and Smith, S. (2001) A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, **5**, 143–156.

Jezzard, P. and Balaban, R. S. (1995) Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine*, **34**, 65–73.

Ruthotto, L., Mohammadi, S., Heck, C., Modersitzki, J. and Weiskopf, N. (2013) Hyperelastic susceptibility artifact correction of DTI in SPM. In *Bildverarbeitung für die Medizin*, 344–349. Springer.

Sjölund, J., Eklund, A., Özarslan, E., Herberthson, M., Bånkestad, M. and Knutsson, H. (2018) Bayesian uncertainty quantification in linear models for diffusion MRI. *NeuroImage*, **175**, 272–285.

Smith, S. M. (2002) Fast robust automated brain extraction. *Human Brain Mapping*, **17**, 143–155.

Taylor, P. A., Alhamud, A., van der Kouwe, A., Saleh, M. G., Laughton, B. and Meintjes, E. (2016) Assessing the performance of different DTI motion correction strategies in the presence of EPI distortion correction. *Human Brain Mapping*, **37**, 4405–4424.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H. et al. (2013) The WU-Minn human connectome project: an overview. *NeuroImage*, **80**, 62–79.

Voss, H. U., Watts, R., Uluğ, A. M. and Ballon, D. (2006) Fiber tracking in the cervical spine and inferior brain regions with reversed gradient diffusion tensor imaging. *Magnetic Resonance Imaging*, **24**, 231–239.

Wang, S., Peterson, D. J., Gatenby, J. C., Li, W., Grabowski, T. J. and Madhyastha, T. M. (2017) Evaluation of field map and nonlinear registration methods for correction of susceptibility artifacts in diffusion MRI. *Frontiers in Neuroinformatics*, **11**, 17.

Wu, M., Chang, L.-C., Walker, L., Lemaitre, H., Barnett, A. S., Marenco, S. and Pierpaoli, C. (2008) Comparison of EPI distortion correction methods in diffusion tensor MRI using a novel framework. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 321–329. Springer.

Zhang, Y., Brady, M. and Smith, S. (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, **20**, 45–57.