Muti-shell diffusion MRI harmonisation and enhancement challenge (MUSHAC): progress and results

Muti-shell diffusion MRI harmonisation and enhancement challenge (MUSHAC): progress and results

Lipeng Ning[1,2], Elisenda Bonet-Carne[3], Francesco Grussu[3], Farshid Sepehrband[4], Enrico Kaden[3], Jelle Veraart[5], Stefano B. Blumberg[3], Can Son Khoo[3], Marco Palombo[3] Jaume Coll-Font[6,2], Benoit Scherrer[6,2], Simon K. Warfield[6,2], Suheyla Cetin Karayumak[1,2], Yogesh Rathi[1,2], Simon Koppers[7], Leon Weninger[7], Julia Ebert[7], Dorit Merhof[7], Daniel Moyer[4], Maximilian Pietsch[8], Daan Christiaens[8], Rui Teixeira[8], Jacques-Donald Tournier[8], Andrey Zhylka[9], Josien Pluim[9], Greg Parker[10], Umesh Rudrapatna[10], John Evans[10], Cyril Charron[10], Derek K. Jones[10,11], and Chantal W.M. Tax[10]

[1] Brigham and Women's Hospital, Boston, United States
[2] Harvard Medical School, Boston, United States
[3] University College London, London, United Kingdom
[4] University of Southern California, Los Angeles, United States
[5] New York University, New York, NY, United States
[6] Boston Children's Hospital, Boston, United States
[7] RWTH Aachen University, Aachen, Germany
[8] King's College London, London, United Kingdom
[9] Eindhoven University of Technology, Eindhoven, Netherlands
[10] CUBRIC, School of Psychology, Cardiff University, Cardiff, United Kingdom
[11] School of Psychology, Australian Catholic University, Melbourne, Australia

Abstract. We present a summary of competition results in the multi-shell diffusion MRI harmonisation and enhancement challenge (MUSHAC). MUSHAC is an open competition intended to stimulate the development of computational methods that reduce scanner- and protocol-related variabilities in multi-shell diffusion MRI data across multi-site studies. Twelve different methods from seven research groups have been tested in this challenge. The results show that cross-vendor harmonization and enhancement can be performed by using suitable computational algorithms such as deep convolutional neural networks. Moreover, parametric models for multi-shell diffusion MRI signals also provide reliable performances.

Keywords: Diffusion MRI · Harmonisation · Spherical harmonics · Deep learning · Parametric model.

1 Introduction

Multi-center clinical studies on mental disorders usually need to combine diffusion magnetic resonance imaging (dMRI) data acquired from different scanners to increase the statistical power and sensitivity of studies. But inter-scanner

variability and different acquisition protocols could introduce inherent variability in dMRI data which is a major limitation in multi-center dMRI studies [1–3]. Therefore, there is a pressing need for reliable approaches to harmonise scanner- and/or protocol-related variability in dMRI data [4–8].

As a continuation of the single-shell diffusion MRI harmonisation challenge [9, 10], the multi-shell diffusion MRI harmonisation and enhancement challenge (MUSHAC) is an open competition that provides datasets from different scanners to enable the development new harmonisation algorithms for multi-shell dMRI data. It evaluates the performances of different algorithms using a test dataset from the same scanners based on several dMRI measures that are relevant in clinical studies. This work reports the preliminary evaluation results from MUSHAC.

Data: The datasets are a subset of the benchmark database in [9, 10]. MUSHAC uses the data of 15 healthy volunteers who were scanned on a 3T Siemens Prisma and 3T Siemens Connectom Scanner. On both scanners, dMRI images were acquired with a 'standard' (ST) protocol and a 'state-of-the-art' (SA) protocol. The ST data from both scanners have an isotropic resolution of 2.4 mm, TE = 89 ms and TR=7.2 s. Both ST dMRI data have 30 directions at b=1200, 3000 s/mm^2. On the other hand, the Prisma-SA data has a higher isotropic resolution of 1.5mm, TE=80 ms, TR = 7.1s and 60 directions at the same b-values while the Connectom-SA data has an even higher resolution of 1.2 mm with TE=68 ms, TR=5.4 s and 60 directions. Additional b=0 s/mm^2 images were also acquired with matched TE and TR across scanners. Moreover, structural MPRAGEs (e.g. T1-weighted anatomical MRI scans) were acquired on both scanners. The data from 10 randomly selected subjects were used as training data and the remaining 5 subjects were used for testing.

Preprocessing: The b0 volumes were corrected for EPI distortions by applying FSL TOPUP on reversed phase-encoding pairs [11]. The data was corrected for eddy current induced distortions, subject motion, EPI distortions [12], and gradient-nonlinearity distortions [13] with FSL TOPUP/eddy and in-house software kindly provided by Massachusetts General Hospital (MGH). Spatiotemporally varying b-vectors and b-values due to gradient nonlinearities of the Connectom scanner were made available [14–16]. All data were affinely registered to Prisma-ST using the corresponding fractional anisotropy (FA) maps with appropriate b-matrix rotation.

Tasks: This competition included two tasks on multi-shell dMRI data harmonisation . Specifically, Task 1 is to predict Connectom-ST data using Prisma-ST data where both datasets have matching TE, TR and angular and spatial resolution. The second task includes the prediction of Prisma-SA data (Task 2a) and Connectom-SA data (Task 2b) given Prisma-ST data.

Evaluation: For the test datasets, fractional anisotropy (FA) and mean diffusivity (MD) were obtained by fitting the diffusion tensor model to the low b-shell dMRI signals using the weighted least squares algorithm. The zeroth and second order rotationally invariant spherical harmonic (RISH) features [6], i.e. L0 and L2, were computed for both b-shells. Moreover, two multi-shell dMRI measures, including mean kurtosis (MK) [17] and the return-to-origin probability (RTOP) measure [18], were also computed using the DIPY [19] toolbox. All measures were computed from the predicted data of each subject and compared to the ground-truth measures derived from the acquired data, which were not released to the participants. The performances were evaluated in brain regions specified by brain masks excluding the cerebellum, obtained with the Geodesic Information Flow (GIF) algorithm [20]. Then, the percentage error (PE) and the absolute-value of PE (APE) were computed globally in a brain mask excluding cerebellum and regionally based on FreeSurfer regions excluding the cerebellum. The reference values for PE and APE in Task 1 were computed by comparing the dMRI measures from Connectom-ST and Prisma-ST data.

Algorithms: Twelve different algorithms from seven research groups were evaluated for Task 1. Eight of these algorithms were also evaluated for Task 2a and Task 2b. Based on the similarities of the underlying methods, these algorithms can be grouped into three categories. The first category includes three interpolation-based algorithms, where Algorithm 1 uses spherical harmonic (SH) functions to interpolate signals along angular directions and Algorithms 2 and 3 interpolate the multi-shell dMRI signals using a parametric model for multi-shell dMRI signals. The second type of algorithms is based on regressions. In particular, Algorithm 4 uses linear regression based on RISH features, and Algorithms 5 and 6 are based on regression forest. The third type of algorithms use either deep convolutional neural networks (CNN), including Algorithms 7, 8, 10, 11, 12, or a fully connected feed-forward network, including Algorithm 9, to learn mappings between dMRI signals using SH representations. The twelve algorithms include but are not limited to the methods proposed in [21, 8, 22].

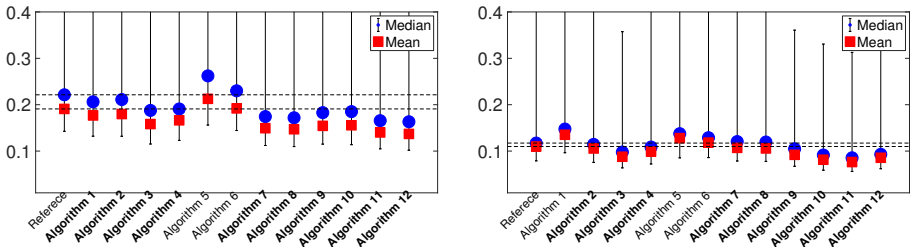
2 Results

2.1 Scanner-to-scanner harmonisation (Task 1)

Most algorithms could significantly reduce inter-scanner variabilities in the dMRI measures. All deep-CNN based approaches had similar performances in this task and are in general better than the other two types of algorithms. In particular, Figure 1a and Figure 1b illustrate the APE for the zero-order (L0) RISH features at the high b-shell and MK, respectively. The blue and red markers represent the median and mean values in global evaluations and the vertical bars represent the 20-80 percentiles of the distributions. The 80-percentile bars go beyond the displayed range of the plots. The two dashed horizontal lines show the reference

median and mean values obtained by comparing the ground-truth measures from acquired data.

Figure 2a illustrates the reference distribution of the PE in different brain regions between the acquired dMRI data. The frontal lobe and occipital lobe consistently have higher errors than other brain regions, which may be caused by stronger susceptibility distortions in these regions. On the other hand, Figure 2b illustrates the region-wise PE corresponding to the output of Algorithm 12 which significantly reduced the errors in these ROI.



(a) Task 1: APE of high-b-value L0

(b) Task 1: APE of MK

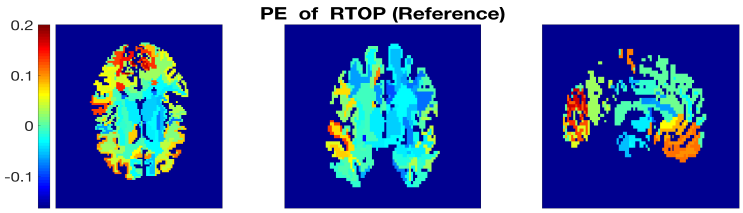
Fig. 1: Scanner-to-scanner harmonisation results. The red square, blue circle and solid black lines show the mean, median and 20-80 percentile of the the absolute percentage error (APE) of a) the high-b-value L0 RISH feature, and b) the MK measure. The algorithms with mean APE lower than the reference are highlighted using bold font.

2.2 Spatial- and angular resolution enhancement (Tasks 2a and 2b)

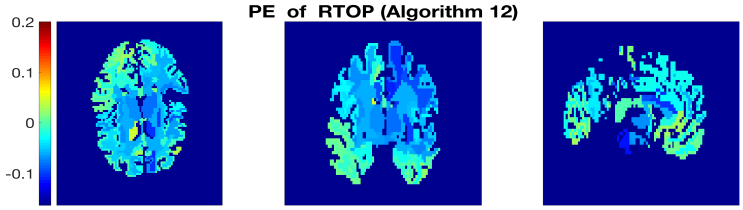
Figures 3a and 3b show the APE of the L0 RISH features for the b=3000 s/mm^2 signals and the MK measure, respectively, for Task 2a. Figures 3c and 3d show the corresponding results for Task 2b. In this case, no reference lines are plotted since there is no ground-truth due to different spatial resolutions. It is interesting to note that the interpolation-based Algorithm 3 has consistently good performances in these tasks.

3 Discussion and Conclusion

We present preliminary results on a comparison of the performance of several harmonisation algorithms that estimate mappings between multi-shell dMRI data acquired using different scanners and/or protocols. These algorithms include interpolation-based approaches, regression-based approaches and deep-learning based methods. Our results show that harmonisation algorithms could significantly reduce scanner- or protocol-related differences in dMRI data. These algorithms generally have lower accuracy for the spatial- and angular-resolution

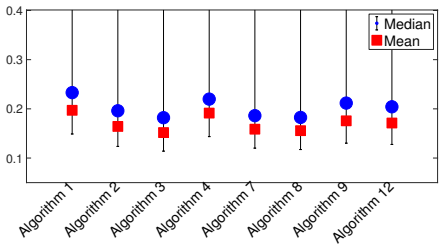


(a) Region-wise PE of RTOP (Reference)

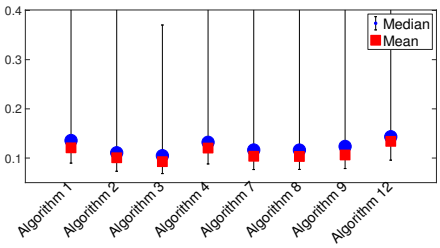


(b) Region-wise PE of RTOP (Algorithm 12)

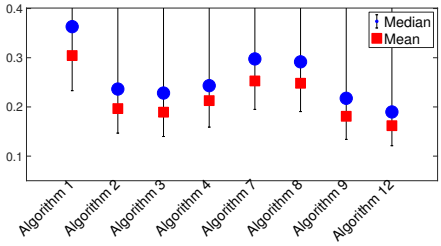
Fig. 2: Scanner-to-scanner harmonisation results. a) Region-wise percentage error (PE) of RTOP between the acquired data. b) Region-wise PE of RTOP from the output of Algorithm 12.



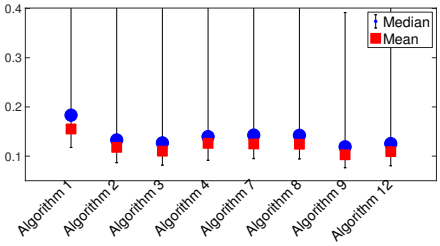
(a) Task 2a: APE of high-b-value L0



(b) Task 2a: APE of MK



(c) Task 2b: APE of high-b-value L0



(d) Task 2b: APE of MK

Fig. 3: Spatial- and angular-resolution enhancement results. a) APE of high-b-value L0 RISH feature in Task 2a, b) APE of MK in Task 2a, c) APE of high-b-value L0 RISH feature in Task 2b, and d) APE of MK in Task 2b.

enhancement tasks than the harmonisation task. But the approach based on parametric modeling of multi-shell signals has consistently good performance over tasks.

Regarding potential limitations, susceptibility distortions may contribute more to voxel or region-wise inter-protocol residuals than protocol-related variabilities, leading to reduced performances for most algorithms in some ROIs. Different preprocessing methods could be explored to reduce systematic errors in these regions. Alternatively, the errors in these regions can also be reduced using suitable harmonisation algorithms as shown in Figure 2b.

Acknowledgement

CMWT is supported by a Rubicon grant (680-50-1527) from the Netherlands Organisation for Scientific Research (NWO) and Wellcome Trust grant (096646/Z/11/Z). LN is supported in part by NIH grants R21MH115280 and R21MH116352. FG is funded by the Horizon2020-EU.3.1 CDS-QuaMRI grant (ref.: 634541) and by the Engineering and Physical Sciences Research Council (EPSRC: ref.: EP/M020533/1). DKJ was supported by MRC grant MR/K004360/1. Scan costs were supported by the National Centre for Mental Health (NCMH) with funds from Health and Care Support Wales and by the Wellcome Trust. JV is a Postdoctoral Fellow of the Research Foundation - Flanders (FWO; grant number 12S1615N). AZ has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765148. SK was supported by the International Research Training Group 2150 of the German Research Foundation (DFG).

References

1. Vollmar, C., O'Muircheartaigh, J., Barker, G.J., Symms, M.R., Thompson, P., Kumari, V., Duncan, J.S., Richardson, M.P., Koepp, M.J.: Identical, but not the same: Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0t scanners. NeuroImage **51**(4) (2010) 1384 – 1394
2. Landman, B.A., Farrell, J.A., Jones, C.K., Smith, S.A., Prince, J.L., Mori, S.: Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. NeuroImage **36**(4) (2007) 1123 – 1138
3. Landman, B.A., Huang, A.J., Gifford, A., Vikram, D.S., Lim, I.A.L., Farrell, J.A., Bogovic, J.A., Hua, J., Chen, M., Jarso, S., Smith, S.A., Joel, S., Mori, S., Pekar, J.J., Barker, P.B., Prince, J.L., van Zijl, P.C.: Multi-parametric neuroimaging reproducibility: A 3-T resource study. NeuroImage **54**(4) (2011) 2854 – 2866
4. Pohl, K.M., Sullivan, E.V., Rohlfing, T., Chu, W., Kwon, D., Nichols, B.N., Zhang, Y., Brown, S.A., Tapert, S.F., Cummins, K., Thompson, W.K., Brumback, T., Colrain, I.M., Baker, F.C., Prouty, D., Bellis, M.D.D., Voyvodic, J.T., Clark, D.B., Schirda, C., Nagel, B.J., Pfefferbaum, A.: Harmonizing DTI measurements across scanners to examine the development of white matter microstructure in 803 adolescents of the NCANDA study. NeuroImage **130** (2016) 194 – 213

5. Fortin, J.P., Parker, D., Tun, B., Watanabe, T., Elliott, M.A., Ruparel, K., Roalf, D.R., Satterthwaite, T.D., Gur, R.C., Gur, R.E., Schultz, R.T., Verma, R., Shinohara, R.T.: Harmonization of multi-site diffusion tensor imaging data. NeuroImage **161** (2017) 149 – 170
6. Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., Grant, G., Marx, C., Morey, R., Flashman, L., George, M., McAllister, T., Andaluz, N., Shutter, L., Coimbra, R., Zafonte, R., Coleman, M., Kubicki, M., Westin, C., Stein, M., Shenton, M., Rathi, Y.: Inter-site and inter-scanner diffusion MRI data harmonization. NeuroImage **135** (2016) 311 – 323
7. Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., Karmacharya, S., Grant, G., Marx, C.E., Morey, R.A., Flashman, L.A., George, M.S., McAllister, T.W., Andaluz, N., Shutter, L., Coimbra, R., Zafonte, R.D., Coleman, M.J., Kubicki, M., Westin, C.F., Stein, M.B., Shenton, M.E., Rathi, Y.: Multi-site harmonization of diffusion MRI data in a registration framework. Brain Imaging and Behavior **12**(1) (Feb 2018) 284–295
8. Karayumak, S.C., Bouix, S., Ning, L., James, A., Crow, T., Shenton, M., Kubicki, M., Rathi, Y.: Retrospective harmonization of multi-site diffusion MRI data acquired with different acquisition parameters. NeuroImage **184** (2019) 180 – 200
9. Tax, C., et al.: Cross-vendor and cross-protocol harmonisation of diffusion tensor imaging data: a comparative study. In: ISMRM-ESMRMB. Number 0471 (2018)
10. Tax, C., et al.: Cross-scanner and cross-protocol diffusion mri data harmonisation: A benchmark database and evaluation of algorithms. submitted (2018)
11. Andersson, J.L., Skare, S., Ashburner, J.: How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. NeuroImage **20**(2) (2003) 870 – 888
12. Andersson, J.L., Sotiropoulos, S.N.: An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. NeuroImage **125** (2016) 1063 – 1078
13. 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., Essen, D.C.V., Jenkinson, M.: The minimal preprocessing pipelines for the human connectome project. NeuroImage **80** (2013) 105 – 124
14. Bammer, R., Markl, M., Barnett, A., Acar, B., Alley, M., Pelc, N., Glover, G., Moseley, M.: Analysis and generalized correction of the effect of spatial gradient field distortions in diffusion-weighted imaging. Magnetic Resonance in Medicine **50**(3) 560–569
15. 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., Essen, D.C.V., Ugurbil, K., Behrens, T.E.: Advances in diffusion MRI acquisition and processing in the Human Connectome Project. NeuroImage **80** (2013) 125 – 143
16. Rudrapatna, S.U., Parker, G.D., Roberts, J., Jones, D.K.: Can we correct for interactions between subject motion and gradient-nonlinearity in diffusion MRI? In: ISMRM. Number 1206 (2018)
17. Jensen, J.H., Helpern, J.A., Ramani, A., Lu, H., Kaczynski, K.: Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. Magnetic Resonance in Medicine **53**(6) (2005) 1432–1440
18. Özarslan, E., Koay, C.G., Shepherd, T.M., Komlosh, M.E., Irfanoglu, M.O., Pierpaoli, C., Basser, P.J.: Mean apparent propagator (MAP) MRI: A novel diffusion imaging method for mapping tissue microstructure. NeuroImage **78** (2013) 16 – 32

19. Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., Nimmo-Smith, I., Contributors, D.: Dipy, a library for the analysis of diffusion MRI data. Frontiers in Neuroinformatics **8** (2014) 8
20. Cardoso, M.J., Modat, M., Wolz, R., Melbourne, A., Cash, D., Rueckert, D., Ourselin, S.: Geodesic information flows: Spatially-variant graphs and their application to segmentation and fusion. IEEE Transactions on Medical Imaging **34**(9) (2015) 1976–1988
21. Scherrer, B., Schwartzman, A., Taquet, M., Sahin, M., Prabhu, S.P., Warfield, S.K.: Characterizing brain tissue by assessment of the distribution of anisotropic microstructural environments in diffusion-compartment imaging (DIAMOND). Magnetic Resonance in Medicine **76**(3) (2016) 963–977
22. Blumberg, S.B., Tanno, R., Kokkinos, I., Alexander, D.C.: Deeper image quality transfer: Training low-memory neural networks for 3D images. In Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G., eds.: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. (2018) 118–125