Author(s): Hao, Yuxing; Xu, Huashuai; Xia, Mingrui; Yan, Chenwei; Zhang, Yunge; Zhou, Dongyue; Kärkkäinen, Tommi; Nickerson, Lisa D.; Li, Huanjie; Cong, Fengyu

Title: Removal of site effects and enhancement of signal using dual projection independent component analysis for pooling multi-site MRI data

Year: 2023

Version: Accepted version (Final draft)

Please cite the original version:

Hao, Y., Xu, H., Xia, M., Yan, C., Zhang, Y., Zhou, D., Kärkkäinen, T., Nickerson, L. D., Li, H., & Cong, F. (2023). Removal of site effects and enhancement of signal using dual projection independent component analysis for pooling multi-site MRI data. European Journal of Neuroscience, 58(6), 3466-3487. https://doi.org/10.1111/ejn.16120

Removal of site effects and enhancement of signal using dual projection independent component analysis for pooling multi-site MRI data

Yuxing Hao[a,1], Huashuai Xu[b,1], Mingrui Xia[c,d,e], Chenwei Yan[a], Yunge Zhang[a], Dongyue Zhou[a], Tommi Kärkkäinen[b], Lisa D. Nickerson [f,g,*], Huanjie Li[a,*], Fengyu Cong[a,b,h,i]

a School of Biomedical Engineering, Dalian University of Technology, Dalian, China

b Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

c State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

d Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, China

e IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China

f McLean Imaging Center, McLean Hospital, Belmont, MA, United States

g Department of Psychiatry, Harvard Medical School, Boston, MA, United States

h School of Artificial Intelligence, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China

i Key Laboratory of Integrated Circuit and Biomedical Electronic System, Liaoning Province. Dalian University of Technology, Dalian, China

1 Yuxing Hao and Huashuai Xu contributed equally to this study.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by STI 2030 - Major Projects 2022ZD0211500, Science and Technology Planning Project of Liaoning Province (no. 2022JH2/10700002 and 2021JH1/10400049), National Natural Science Foundation of China [grant numbers 81601484], National Foundation in China [grant number JCKY 2019110B009], National Institutes of Health [NIA RF1 AG078304].

Data availability statement

The simulated MRI data are from https://github.com/ThomasYeoLab/CBIG/tree/master /stable_projects/brain_ parcellation/Schaefer2018_LocalGlobal/Parcellations/.

All real MRI data used in this study are publicly available, including Autism Brain Imaging Data Exchange II dataset (ABIDE II) (http://fcon_1000.projects.nitrc.org/indi/abide/ abide_II.html), travelling subject dataset from DecNef Project Brain Data Repository website (https://bicr-resource.atr.jp/srpbsts/).

* Corresponding authors:
E-mail addresses: hj_li@dlut.edu.cn (Huanjie Li), lisa_nickerson@hms.harvard.edu (Lisa D. Nickerson)

Removal of site effects and enhancement of signal using dual projection independent component analysis for pooling multi-site MRI data

Yuxing Hao[a,1], Huashuai Xu[b,1], Mingrui Xia[c,d,e], Chenwei Yan[a], Yunge Zhang[a], Dongyue Zhou[a], Tommi Kärkkäinen[b], Lisa D. Nickerson [f,g,*], Huanjie Li[a,*], Fengyu Cong[a,b,h,i]

a School of Biomedical Engineering, Dalian University of Technology, Dalian, China

b Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

c State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

d Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, China

e IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China

f McLean Imaging Center, McLean Hospital, Belmont, MA, United States

g Department of Psychiatry, Harvard Medical School, Boston, MA, United States

h School of Artificial Intelligence, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China

i Key Laboratory of Integrated Circuit and Biomedical Electronic System, Liaoning Province. Dalian University of Technology, Dalian, China

1 Yuxing Hao and Huashuai Xu contributed equally to this study.

Abstract

Combining magnetic resonance imaging (MRI) data from multi-site studies is a popular approach for constructing larger datasets to greatly enhance the reliability and reproducibility of neuroscience research. However, the scanner/site variability is a significant confound that complicates the interpretation of the results, so effective and complete removal of the scanner/site variability is necessary to realize the full advantages of pooling multi-site datasets. Independent component analysis (ICA) and general linear model (GLM) based harmonization methods are the two primary methods used to eliminate scanner/site effects. Unfortunately, there are challenges with both ICA-based and GLM-based harmonization methods to remove site effects completely when the signals of interest and scanner/site effects-related variables are correlated, which may occur in neuroscience studies. In this study, we propose an effective and powerful harmonization strategy that implements dual projection (DP) theory based on ICA to remove the scanner/site effects more completely. This method can separate the signal effects correlated with site variables from the identified site effects for removal without losing signals of interest. Both simulations and vivo structural MRI datasets, including a dataset from Autism Brain Imaging Data Exchange II and a traveling subject dataset from the Strategic Research Program for Brain Sciences, were used to test the performance of a DP-based ICA

harmonization method. Results show that DP-based ICA harmonization has superior performance for removing side effects and enhancing the sensitivity to detect signals of interest as compared with GLM-based and conventional ICA harmonization methods.

Keywords: dual projection, harmonization, independent component analysis, magnetic resonance imaging, multi-site, site effects

1 Introduction

It is now common practice to pool multi-site magnetic resonance imaging (MRI) datasets to study brain biomarkers of neuroscience, neuropsychiatry, and neurology to promote rigor and reproducibility of results (Button et al., 2013; Eickhoff et al., 2016; Van Horn & Toga, 2009). However, combining multiple datasets does introduce site-related effects, which confounds effects of interest, and complicates the interpretation of the final results (Casey et al., 1998; Focke et al., 2011; Friedman et al., 2008; Pohl et al., 2016; Takao et al., 2011; Venkatraman et al., 2015; Vollmar et al., 2010; Wegner et al., 2008; Zivadinov & Cox, 2008). Site-related effects arises from differences in scanners manufacturers, field strengths, hardware, software, pulse sequences, quality control, and data quality across sites (Jovicich et al., 2009). It has been shown that differences in acquisition parameters and software and hardware upgrades during data collection using the same scanner have non-negligible effects on almost all image-derived phenotypes from structural images (such as cortical surface and gray matter volume), diffusion-weighted images (such as diffusion tensor image (DTI) measures), and functional MRI (fMRI) data (Groves et al., 2011; Li et al., 2020). Hence, effective removal or deconfounding of site-related variability from the MRI data is a critical step to ensure the accuracy and reproducibility of findings generated from combined datasets.

Several approaches have been proposed for the harmonization of multi-site MRI data, including methods based on the general linear model (GLM) (Fennema-Notestine et al., 2007; Glover et al., 2012; Venkatraman et al., 2015), and data-driven unsupervised learning methods such as independent component analysis (ICA) (Chen et al., 2014; Li et al., 2020) and recently proposed deep learning methods (C Monte-Rubio et al., 2022; Dinsdale et al., 2021; Tian et al., 2022). Two of the most popular methods to eliminate or minimize the site-related effects are based on GLM and ICA (Chen et al., 2014). GLM-based harmonization method is easy to implement and often used in multi-site MRI data studies to minimize the site-related effects, in this case, it utilizes site/study variables as covariates of no interest in group-level GLM analysis to control for site-related effects. Fortin et al. (Fortin et al., 2017) have adapted a GLM-based technique called ComBat (Johnson et al., 2007), an empirical Bayesian method for data harmonization that is popular in the field of genetics, to remove unwanted variation induced by sites while preserving the signal-related variation in neuroimaging studies. ComBat has been applied to harmonize DTI measures (Fortin et al., 2017), cortical thicknesses and functional connectivity measures (Yu et al., 2018), magnetic resonance spectroscopy measures (T. K. Bell

et al., 2022), and positron emission tomography (PET) outcomes (Orlhac et al., 2018) showing good performance for removing site effects. ICA is an unsupervised data-driven statistical method that factorizes or decomposes the image data into a set of statistically independent non-Gaussian components reflecting different sources that generate the measured imaging data. And the site-related components can be removed by regressing them from the original data to generate a harmonized clean dataset for further analysis (Chen et al. 2014). In this case, the ICA has been used to do a data-driven estimation of the site/scanner-related covariates of no interest that are regressed out of the data rather than creating covariates to model the site-scanner effects based on strong assumptions (e.g., regressors are used that assume a constant effect for each site/scanner, which ignores within-site/data to day variations in these effects). ICA is typically applied to harmonize individual MRI modalities, however, our previous work (Li et al., 2020) proposed a harmonization method for multi-modal imaging measures that implemented linked ICA (LICA) (Groves et al., 2011) as a novel approach to eliminate scanner effects from multi-study data. LICA simultaneously decomposes the multi-modal imaging data (for example, structural plus diffusion MRI-derived measures) into a set of multi-modal components and a set of subject loadings quantifying the strength of each multi-modal component in each individual, with components reflecting true signals of interest as multi-modal covariance patterns, as well as artifacts and variability related to uninteresting effects like scanner and site differences. We found that several of the resulting LICA components from an analysis of multi-study data with scanner effects were associated with scanner variations and that these patterns could be effectively regressed from the data to obtain harmonized data relatively free from scanner effects. We showed that multi-modal ICA-based harmonization was more effective at removing scanner-related effects compared with the conventional GLM and single-modality ICA harmonization methods. The reason for its superior performance is that even though all three approaches involve regression to remove scanner effects from the data, the data-driven estimates of the scanner effects from LICA of multi-modal MRI data provided more accurate model of site effects than assuming a constant effect or estimating effects based on single modality ICA to use as nuisance covariates for harmonization.

In the present study, we aim to address another limitation of current methods for harmonization scanner/site effects, namely existing methods for harmonization site/scanner effects ignore the possibility of correlations between these effects and the effects of interest. For the conventional GLM approach, site-related variables are included as covariates of no interest or may be regressed out of data prior to higher-level statistical modeling, which may lead to the removal of interesting signals that are correlated with scanner/site variables and to weaker specificity of harmonization. While ComBat tries to preserve the signal-related variation when harmonization scanner/site effects, similar to the conventional GLM approach, it also assumes a constant effect for all datasets collected from the same site or the same scanner state, thus also ignoring the day-to-day variations in scanner performance. While ICA and LICA can identify scanner/site effects that capture day to day variations in scanner performance (Li

et al., 2020), both approaches are vulnerable to identifying components that reflect a mixture of signal and scanner/site effects, rather than separating the effects into two separate components. In our previous work, to retain signals of interest, only components that were associated with site effects and not signals of interest (e.g., had subject loadings that correlated only with site variables and not variables of interest) were removed from the data while mixed components were retained as a conservative approach to harmonization (Chen et al., 2014; Li et al., 2020). One possibility to address this limitation for ICA-based techniques is to run the ICA with several different model orders to identify a decomposition with stable pure site effects related components not mixed with signals of interest. In practice, it is challenging to do this as different mixtures may arise at different model orders such that it may not be possible to have full separation at any model order.

To solve this problem for ICA-based methods, we propose a new ICA with dual-projection (ICA-DP) technique for harmonization scanner/site effects. In this study, we focus on single modality ICA, with extension to multi-modal ICA to be done in future work. For ICA-DP, mixed components from single-modality ICA are separated into a part related to signal only and a part related only to site effects by applying a projection procedure. The site effects extracted from the mixed components via the projection step are combined with the other ICA components that reflected only site/scanner variance and are then removed from the data using a second projection procedure. Our new method is tested using simulated MRI data and in vivo multi-site datasets to assess the performance of ICA-DP compared to conventional ICA and ComBat harmonization methods.

2 Methods

2.1 Dual-projection harmonization improvement

2.1.1 Traditional ICA-based harmonization

ICA decomposition model for group structural MRI data can be expressed as:

$$\mathbf{Y} = \mathbf{LS}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{M \times N}$ denotes the original data to be decomposed, $\mathbf{S} \in \mathbb{R}^{R \times N}$ contains independent spatial maps and $\mathbf{L} \in \mathbb{R}^{M \times R}$ contains their corresponding loadings. And M, N and R denote the number of subjects, voxels and components, respectively.

When removing site effects of multi-site structural MRI data, the original ICA-based harmonization methods (Chen et al., 2014; Li et al., 2020) only eliminate the pure site-related components (only related to site effects) to avoid discarding useful information such as diagnoses or symptom measures. For comparison, we designate this ICA-based harmonization method as ICA-SP (single projection) as it only uses one step of projecting:

$$\mathbf{Y}_{clean}^{ICA-SP} = \mathbf{Y}_{orig} - \mathbf{L}_N \cdot pinv(\mathbf{L}_N) \cdot \mathbf{Y}_{orig}, \tag{2}$$

where $\mathbf{L}_N$ is the loadings of pure site effects components (components related to site effects)

and $\mathbf{Y}_{clean}^{ICA-SP}$ is the harmonized data derived from ICA-SP harmonization methods.

Though it may preserve signal-related information well, it is too soft to remove the site effects as it does nothing with the mixed components and is more than likely to find site effects in its harmonized data.

2.1.2 Proposed ICA-DP harmonization

The ICA-DP harmonization procedure is summarized in Fig. 1. ICA-DP is inspired by the dual-regression approach for projecting a participant's fMRI data onto a set of spatial maps derived from ICA of multi-subject fMRI data to identify subject-level spatial maps corresponding to each group level component (Beckmann et al., 2009; Filippini et al., 2009; Nickerson et al., 2017).

First, the subject series is decomposed by ICA, and the resulting subject loadings **L**, of each component that reflects the strength of the corresponding variables represented in the IC map (could be interesting signal, scanner/site effects, or a mixture) are labeled as loadings for pure site effects components $\mathbf{L}_N$, pure signal components $\mathbf{L}_S$, or mixed components $\mathbf{L}_M$ by calculating the correlations between the loadings and all the signal and site effects variables (i.e., components exhibiting significant associations p<0.05 are identified as signal- and/or site effects-related ones).

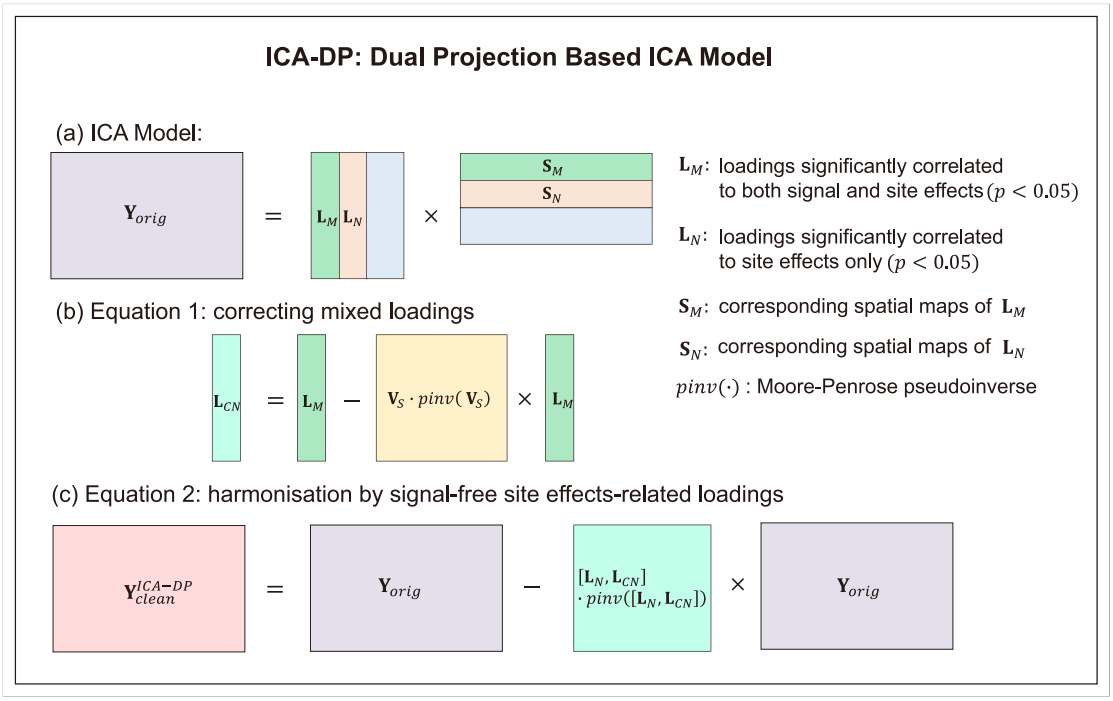

Fig. 1. The procedures of ICA-DP harmonization method. (a) Identifying the loadings extracted by ICA that related to site effects variables (including mixed ones that significantly correlated to both site effects and signal, and the ones only significantly correlated to site effects). (b) Correcting the mixed loadings to only site effects-related ones ($\mathbf{L}_{CN}$) by projecting out signal-related information. (c) Obtaining cleaned data by removing the integral site effects-related components ($\mathbf{L}_N, \mathbf{L}_{CN}$).

The first projection procedure is used to separate the signal effects out from $\mathbf{L}_M$ as below:

$$\mathbf{L}_{CN} = \mathbf{L}_M - \mathbf{V}_S \cdot pinv(\mathbf{V}_S) \cdot \mathbf{L}_M, \quad (3)$$

[1]https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal/Parcellations/

where $\mathbf{V}_S$ denotes the variables of interest (signals to be preserved, e.g., age, gender or health condition), $\mathbf{L}_{CN}$ denotes the corrected site effects contributions to the mixed components, and $pinv(\cdot)$ denotes the Moore-Penrose inverse (pseudoinverse) of a non-square matrix. Thus, the signal information is projected out from $\mathbf{L}_M$ and we can identify the site effects for the mixed components.

Then $[\mathbf{L}_N, \mathbf{L}_{CN}]$ represent the total site-related effects present in the data, which are then cleaned from the subject series via a second projection procedure:

$$\mathbf{Y}_{clean}^{ICA-DP} = \mathbf{Y}_{orig} - [\mathbf{L}_N, \mathbf{L}_{CN}] \cdot pinv([\mathbf{L}_N, \mathbf{L}_{CN}]) \cdot \mathbf{Y}_{orig}, \quad (4)$$

where $\mathbf{Y}_{orig}$ denotes the subject series of spatial maps and $\mathbf{Y}_{clean}^{ICA-DP}$ denotes the harmonized MRI data free from site/scanner effects that can be used for further analysis.

2.2 Study Data

2.2.1 Simulated Data

The simulated structural MRI data, including 100 subjects, were generated in this study. For each subject, the data was generated by computing 10 spatial maps and one set of ground truth subject loadings (Eq. 1). Each component map was multiplied by the corresponding subject loading, and then they were added together to obtain the simulated MRI data for each subject. The spatial maps were gotten by combining different areas of the standard brain template as shown in Fig. 2[1]. To make our simulated data much closer to real MRI data, two kinds of spatial maps were simulated, one is all the spatial maps of 10 components were spatially independent and the other is two components' spatial maps were overlapped (Fig. 2). For each condition, 100 subjects were generated, and the subject-specific data shared the same spatial maps, and the difference was the weights in its loadings corresponding to the spatial maps. Three different types of relationships between subject loadings and signal/site effects variables were simulated in this study: (1) signal variable was not significantly correlated to site effects variables; subject loadings were linearly correlated to signal and (or) site effects variables (Table 1). Among the 10 components, the first four components were significantly related to signal and (or) site effects variables, and the other components were not related to the variables we are interested in. Components 1 and 2 are mixed components, which are related with both signal and site effects variables. The difference is that component 1 is much more related to signal, and component 2 is more correlated to site effects. Component 3 is pure site effects components, which only significantly correlated with site effects variable. Component 4 is a pure signal component that only significantly correlated with signal variable; (2) Signal variable is significantly correlated to site effects variable, subject loadings are linearly correlated to signal and (or) site effects variables (Table 2). Since the signal variable was significantly correlated to the site effects variable, there were no pure signal or site effects components under this condition, so we selected the first 2 components as mixed components, component 1 is much more related to the signal, and component 2 is much more correlated to site effects. Three

[1] https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal/Parcellations/

different correlation levels (from low to high) between signal and site effects variables were simulated in this study to show the harmonization power of ICA-DP.

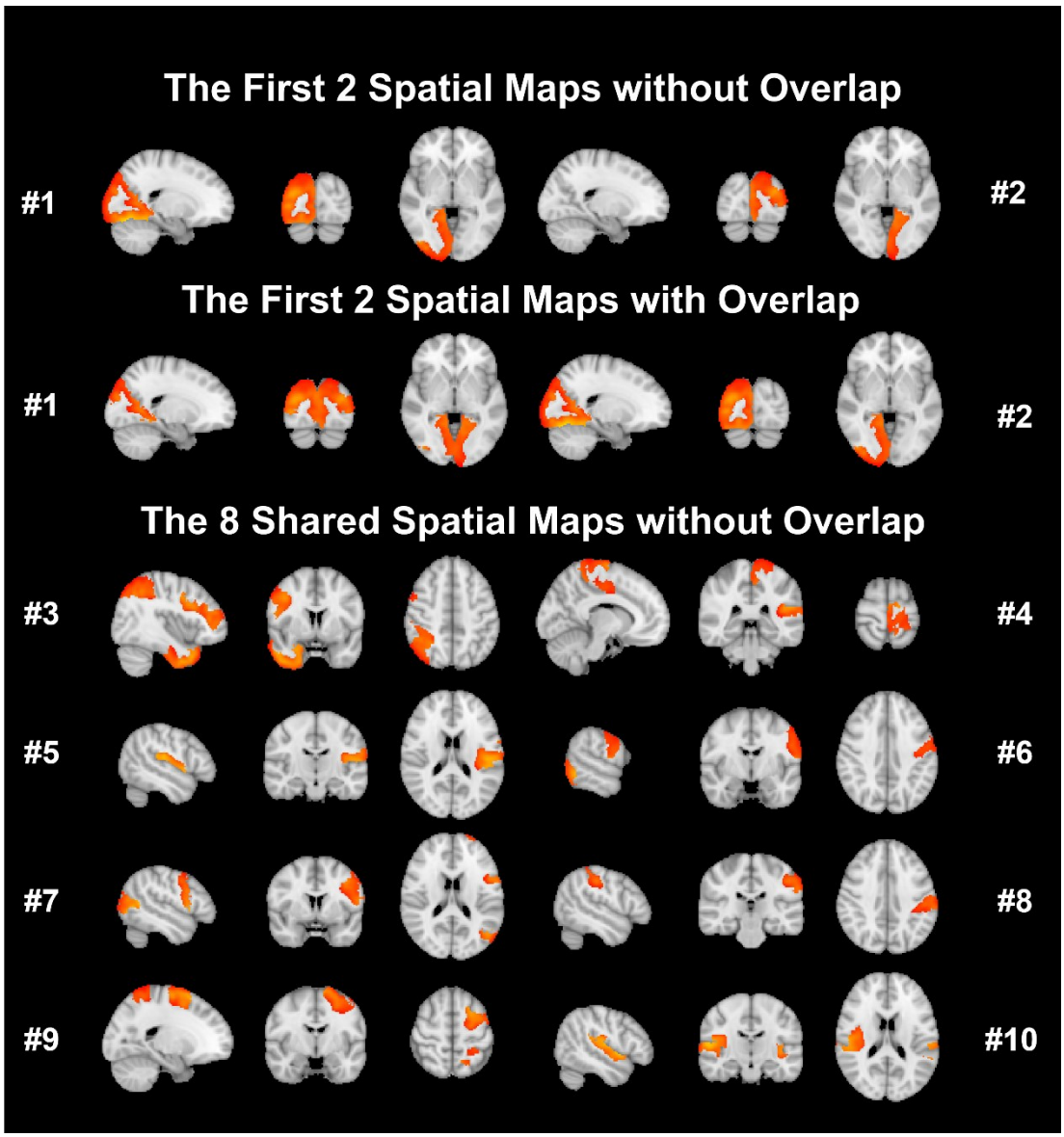

Fig. 2. Ten independent brain spatial maps used to simulate MRI data. Situation1: there is no overlap among all the spatial maps; Situation 2: The first two components were spatially overlapped, and the other 8 components share the same maps with situation 1.

Table 1 Pearson correlation coefficients and corresponding *p* values by correlating variables and loadings.

#Component	Signal Variable (*r*/*p*)	Site Effects Variable (*r*/*p*)
1	0.9425(<0.0001)	0.3999(<0.0001)
2	0.2999(0.0024)	0.9728(<0.0001)
3	--	0.5999(<0.0001)
4	0.5999(<0.0001)	--

Note: Component loadings are linearly correlated with signal and site effects variables, while the signal variable is not significantly correlated to the site effects variable. Components 1 (more related to signal) and 2 (more related to site effects) are mixed components. Component 3 is only related to the site effects variable, and component 4 is only related to the signal variable. The relationship of loadings and variables is expressed by *r*-value and *p*-value. -- denotes not significantly correlated.

Table 2 Pearson correlation coefficients and corresponding *p* values by correlating variables and loadings.

Correlation between Signal and Site Effects	#Component	Signal	Site effects
0.2999 (2.4e-3)	1	0.7946(<0.0001)	0.2412(<0.0001)
	2	0.2590(0.0093)	0.7959(<0.0001)
0.4999 (<0.0001)	1	0.7962(<0.0001)	0.4279(<0.0001)
	2	0.3447(0.0005)	0.7859(<0.0001)
0.6999 (<0.0001)	1	0.7957(<0.0001)	0.5932(<0.0001)
	2	0.4993(<0.0001)	0.7761(<0.0001)

Note: Component loadings are linearly correlated with signal and site effects variables, while signal variable is significantly correlated to site effects variable. Components 1 (more related to signal) and 2 (more related to site effects) are mixed components. The relationship of loadings and variables is expressed by *r*-value and *p*-value. Three different correlation levels between signal and site effects variables are simulated in this study with *r*-values of 0.2999 (p=0.0024), 0.4999 (p=1.2e-7), and 0.6999 (p=5.6e-16), corresponding to low, medium and high correlation levels.

2.2.2 Multi-site MRI data from ABIDE II

High spatial resolution structural MRI data of 606 subjects (including Autism Spectrum Disorder (ASD) patients: 225, Healthy Controls (HC): 381) were obtained from Autism Brain Imaging Data Exchange II dataset (ABIDE II) (http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html). The data were collected from 13 different sites, all the data were acquired from 3T scanner with different manufacturers (Siemens, Philips, and GE) (Di Martino et al., 2017). The acquisition parameters: scanner/site and imaging-related details, including repetition time (TR), echo time (TE), flip angle (FA), and voxel size in Table 3. The demographic information: ASD/HC, gender, and age are summarized in Table 4. Subjects with ASD could be divided into two categories: ASD only and ASD with comorbidity (Attention-Deficit/Hyperactivity Disorder, anxiety or others) (Di Martino et al., 2017).

Table 3 Scanning parameters and demographic information of the multi-site ABIDE II data.

Sites	Scanners	TR/TE (ms)	FA (degree)	Voxel Size
EMC	GE MR750	1664/4.24	16	0.9×0.9×0.9
ETH	PhilipsAchieva	3000/3.9	8	0.9×0.9×0.9
GU	Siemens TriTim	2530/3.5	7	1×1×1
IU	Siemens TriTim	2400/2.3	8	0.7×0.7×0.7
KKI	PhilipsAchieva	3500/3.7	8	1×1×1
KUL	PhilipsAchieva	2000/4.6	8	1×1×1.2
OHSU	Siemens TriTim	2300/3.58	10	1×1×1.1
ONRC	Siemens Skyra	2200/2.88	13	0.8×0.8×0.8
SU	GE SIGNA	5.9/1.8	11	1×1×1
UCD	Siemens TriTim	2000/3.16	8	1×1×1
UCLA	Siemens TriTim	2300/2.86	9	1×1×1.2
UM	GE Healthcare	-/-	12	1×1×1
USM	Siemens TriTim	2300/2.91	9	1×1×1.2

Note: The data were collected from 13 different sites: Erasmus University Medical Center (EMC), ETH Zürich (ETH,), Georgetown University (GU), Indiana University (IU), Kennedy Krieger Institute (KKI), Katholieke Universiteit Leuven (KUL), Oregon Health and Science University (OHSU), Olin Neuropsychiatry Research Center (ONRC), Stanford University (SU), University of California Davis (UCD), University of California Los Angeles (UCLA), University of Miami (UM), University of Utah School of Medicine (USM).

In this study, the site differences are defined as nuisance variables to eliminate, while group differences (ASD/HC), age and gender are regarded as signal variables. The correlation coefficients among these variables are summarized in Table 5. Since site differences are categorical variables and calculating the correlation coefficients between categorical variables and numeric variables directly is not achievable, we used ANOVA to calculate the significant levels of signal variables and site effects variables to identify the independent components related to site effects significantly. For ICA analysis, the components that only significantly correlated to site variables were regarded as pure site effects components, and the components that significantly correlated to both site and signal variables were regarded as mixed components.

Table 4 Demographic information of the multi-site ABIDE II data collected from 13 sites.

Sites	ASD/HC	ASD with comorbidity	Male/Female	Age Range (Mean)	Full IQ Standard Score (Mean)
EMC	18/20	11	31/7	6.40~10.66 (8.28)	--
ETH	8/17	--	25/0	13.83~29.42 (22.43)	82~133 (112.84)
GU	33/43	--	51/25	8.06~13.88 (10.74)	92~149 (119.17)
IU	18/18	--	28/8	17~54 (24.61)	80~135 (116.36)
KKI	32/133	29	103/62	8.02~12.99 (10.36)	83~143 (113.26)
KUL	7/0	1	7/0	18~25 (21.71)	73~146 (103.86)
OHSU	33/51	23	52/32	7~15 (10.94)	72~140 (113.11)
ONRC	16/29	5	32/13	18~31 (23.24)	86~146 (111.76)
SU	15/17	2	29/3	8.42~12.99 (10.99)	93~151 (115.31)
UCD	13/13	4	19/7	12~17.83 (15.03)	83~128 (107.96)
UCLA	12/12	--	19/5	7.75~15.03 (11.04)	78~141 (107.96)
UM	7/12	--	14/5	7.3~14.3 (10.32)	98~144 (115.72)
USM	13/16	--	24/5	9.12~38.86 (21.21)	73~144 (108.5)

Table 5 The relationship of signal and site effects variables for real MRI data (*p* values).

Correlation	Site	ASD vs HC	Age	Gender
Site	<0.0001	<0.0001	<0.0001	0.0004
ASD vs HC	<0.0001	<0.0001	-	<0.0001
Age	<0.0001	-	<0.0001	-
Gender	0.0004	<0.0001	-	<0.0001

2.2.3 Traveling subjects

To further validate the site effects removing efficiency of ICA-DP, the high spatial resolution structural MRI data from a traveling-subject dataset including 9 healthy participants (all males, age: 27±2.6) scanned at 12 different sites from the DecNef Project Brain Data Repository (https://bicr-resource.atr.jp/srpbsts/) were used in this study. For T1-weighted MRI data of the 12 different sites, there were two phase-encoding directions (PA and AP), three MRI manufacturers (Siemens, GE, and Philips) with seven scanner types (TimTrio, Verio, Skyra, Spectra, MR750W, SignaHDxt, and Achieva) and four channels per coil (8, 12, 24, and 32) (Maikusa et al., 2021; Tanaka et al., 2021). Scanning parameters, including repetition time (TR), echo time (TE), flip angle (FA), and voxel size, are summarized in Table 6. Three sites (i.e., ATT, UTO, and YC2) were excluded for harmonization analysis because they contain duplicate data (there are 7 duplicate images in ATT and ATV, 2 same images within both YC2 and UTO).

As the images from this dataset are the same groups under different sites, site effects variables and subject variables (including subject labels, age, etc.) are uncorrelated.

Table 6 Scanning parameters of the traveling-subject dataset.

Sites	Scanners	TR/TE (ms)	FA (degree)	Voxel Size
ATT	SiemensTimTrio	2300/2.98	9	1×1×1
ATV	Siemens Verio	2300/2.98	9	1×1×1
COI	Siemens Verio	2300/2.98	9	1×1×1
HKH	Siemens Spectra	1900/2.38	10	0.8×0.75×0.75
HUH	GE Signa HDxt	6788/1.928	20	1×1×1
KPM	Philips Achieva	7.1/3.31	10	1×1×1
KUS	Siemens Skyra	2300/2.98	9	1×1×1
KUT	SiemensTimTrio	2000/3.4	8	0.9375×0.9375×1
SWA	Siemens Verio	2300/2.98	9	1×1×1
UTO	GE MR750W	7.7/3.1	11	1×1.0156×1.0156
YC1	Philips Achieva	6.99/3.176	9	1×1×1

Note: The datasets include 9 healthy subjects undergoing T1-weighted MRI scans at 12 different sites, and all of them used 3T scanners and the same acquisition parameters but with different manufacturers and hardware versions (Siemens, GE, and Philips).

2.3 Data analysis

For both the ABIDE II dataset and the traveling dataset, the modulated gray matter (GM) images were analyzed with FSL-VBM (Douaud et al., 2007) (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/ FSLVBM), an optimized VBM protocol (Good et al., 2001) carried out with FSL tools (Smith et al., 2004). First, structural images were brain-extracted and grey matter-segmented before being registered to the MNI 152 standard space using non-linear registration (Andersson et al., 2007). The resulting images were averaged and flipped along the x-axis to create a left-right symmetric, study-specific grey matter template. Second, all native grey matter images were non-linearly registered to this study-specific template and "modulated" to correct for local expansion (or contraction) due to the non-linear component of the spatial transformation. The modulated GM images were then smoothed with an isotropic Gaussian kernel with a sigma of 3 mm.

2.4 Comparison of ICA-DP with other harmonization methods

General linear model (GLM) harmonization(Maikusa et al., 2021; Venkatraman et al., 2015; Yamashita et al., 2019) and ComBat harmonization(Fortin et al., 2017) are two main model-baesd methods for removing site effect differences (see the Supplementary Information for detailed descriptions of these methods). To examine the advantages of our proposed methods, we compared ICA-DP with ICA-SP and these model-based harmonization methods in terms of site effect removal and biological variability preservation.

2.4.1 Site effects removing with ICA-SP/DP methods

Firstly, ICA was applied to the non-harmonized data to identify pure site effects components, pure signal components and mixed components. For simulated data, the Pearson correlation

coefficient between subject loadings and signal (or site effects) variable was used to measure the properties of components. Those components only related with the signal variable ($p < 0.05$) were identified as pure signal components; those components only related with the site effects variable ($p < 0.05$) were identified as pure site effects components; those components related with the signal and site effects variables ($p < 0.05$) were identified as mixed components.

For the real MRI data from ABIDE II and the traveling-subject data, we used the Pearson correlation coefficient and ANOVA to identify signal, site effects and mixed components (age, gender, and group difference(ASD/HC) were signal variables of interest). The subject loadings from one site were divided into one variable, then 13 levels- for ABIDE II and 9 levels- for the traveling-subject data. ANOVA was used to calculate the significant levels of subject loadings and site differences and identify the independent components related to site effects significantly. Finally, those components whose *p*-values of ANOVA were significant (using Bonferroni correction to adjust for multiple comparisons, adjusted $p < 0.05$) were identified as site effects components. The intersection of signal components and site-related components were mixed components.

For the ICA-SP method, only pure site effects components were used to regress the site effects. For the ICADP method, all the site effects components, including the mixed ones, were used for site effects removal. The site effects extracted from the mixed and pure site effects components will be regarded as the integral site-related effects for removal by ICA-DP. All the ICA analyses were based on MATLAB and FSL MELODIC.

2.4.2 Site effects removing with GLM and ComBat methods

For the GLM-based harmonization method, the site difference is regarded as the covariates to be regressed out. For ComBat harmonization method, firstly, ComBat normalizes the data by removing the effect of the overall mean and signal variables. Then, using an empirical Bayes framework, ComBat estimates additive and multiplicative site effects. The final harmonized data could be obtained by removing these site effects and adding the signal-related information back. In our study, the performance of GLM was only evaluated with simulation data. The main reason is that ComBat is a GLM-derived model and is more powerful than the original GLM model. Thus, for real MRI data, only the performance of ICA model and ComBat model were compared.

2.4.3 Evaluating the harmonization results

For simulated MRI data, ICA was utilized to the non-harmonized and harmonized data to extract and identify the signal- and site effects-related components to compare the harmonization effects of all the methods.

For the real MRI data, a set of analyses were applied to show the performance of site effects elimination and biological variability preservation (including age effects and group difference (ASD/HC)) for all the harmonization methods. For site effects removal evaluation, T-distributed stochastic neighbor embedding (t-SNE) was used to visualize the heterogeneity

related to sites of non-harmonized data and harmonized data by projecting their dominant features into a 2D space. We could assess the efficiency of the harmonization methods by checking whether there are site-clustered distributions or noticeable inter-site heterogeneities after site effects removal. Group F-test was also applied to both the non-harmonized data and the harmonized data derived from different methods to test the significant difference regions caused by the site difference. For age effects evaluation, the Pearson correlation coefficient between median GM and age of all the subjects from ABIDE II data was calculated to show their relationship. The median GM per subject was obtained by calculating the median of 100 regions of interest (using a standard brain template, FSLMNI152_1mm, in FSL parcelled by Schaefer et al., 2018). One-way ANOVA analysis was also applied to both the non-harmonized data and the harmonized data derived from different methods to test the significant difference regions caused by age. For group difference (ASD/HC) evaluation, group t-test was applied to both the non-harmonized data and the harmonized data derived from different methods to test the significant difference regions caused by the group difference. FWE-corrected $p < 0.05$ using non-parametric permutation testing with threshold-free cluster enhancement (TFCE) (Smith & Nichols, 2009) in FSL's Randomise (Winkler et al., 2014), with 5,000 permutations was used to find the significant regions.

3 Results

3.1 Simulation Harmonization Results

For simulated data, the data were decomposed into 10 components based on the simulation. Fig. 3 shows the signal- and site-effects-related components of simulated data extracted by ICA, before and after harmonization, when the signal variable is not significantly correlated to the site-effects variable. The results are shown in Fig. 3(a) when the spatial maps of all 10 components are spatially independent. Fig. 3(b) shows the results when the first two components are spatially overlapped. When the signal variable is not correlated to the site effects variable, the results for spatially independent and spatial dependent data are similar. All the harmonization methods could remove pure site effects component #3 and preserve pure signal component #4. However, the site effects cannot be removed from the mixed components #1 (more related to signal for the original data) and #2 (more related to site effects for the original data) by ICA-SP method. The performance of ICA-DP, GLM and ComBat were comparable under this situation, the site effects in the mixed components #1 and #2 were effectively removed and the signal effect was enhanced by increasing its correlation levels with the signal variable after ICA-DP, GLM and ComBat harmonization. The two mixed components were combined into one component that is only significantly related with signal variables. The site effects-related regions were also removed after harmonization with ICA-DP, GLM, and ComBat methods.

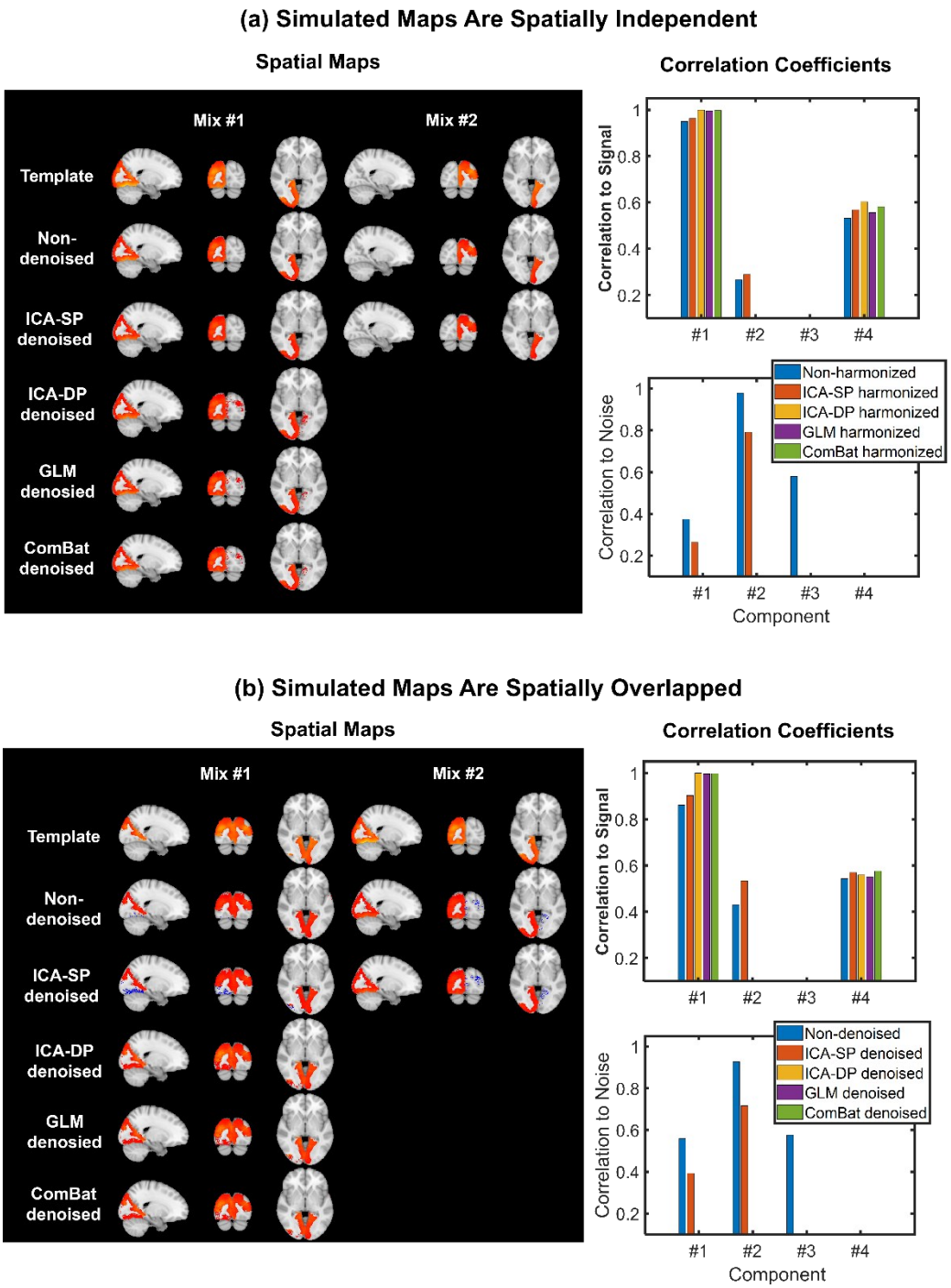

Fig. 3. Harmonization effects on the signal- and site-effects-related components when the signal variable is not significantly correlated to the site-effects variable. (a) Results when the spatial maps of all 10 components were spatially independent. (b) Results when the spatial maps of the first two components were spatially overlapped. For the non-harmonized data, components #1 and #2 are two mixed components, named Mix #1 (more related to the signal variable) and Mix #2 (more related to the site effects variable), component #3 is pure site effects component and component #4 is a pure signal component. The pure site effects component was removed by all the harmonization methods. The two mixed components were combined into one component that is only significantly related with signal variable. The site effects-related regions were also removed after harmonization with ICA-DP, GLM, and ComBat methods. ICA-SP cannot harmonize the site effects from the two mixed components.

Fig. 4 shows the harmonization effects on the two mixed components when the signal variable significantly correlates to the site effects variable. Three different correlation levels between the signal and site effects variables were simulated. Fig. 4(a) shows the results when the spatial maps of all 10 components were spatially independent. Fig. 4(b) shows the results when the spatial maps of the first two components were spatially overlapped. Among all the harmonization methods, only ICA-DP could effectively weaken the site effects while strengthening the signal effects when signal and site effects variables are correlated for two types of simulated data.

ICA-SP harmonization could not remove the site effects from the mixed components, as both the extracted spatial maps and correlation coefficients between subject loadings and site effects variable of the mixed components were not changed after being harmonized by ICA-SP, compared to that of non-harmonized data.

GLM harmonization showed the most aggressive harmonization performance while eliminating the site effects-related information at the expense of destroying the signal-related information. After being harmonized by GLM, both components #1 and #2 were not correlated to site effects variable. Besides, component #2 was not correlated to the signal variable any longer and the correlation between component #1 and signal variable also became lower, which became more severe with the increasing correlation levels between signal and site effects variables.

ComBat harmonization could not remove the site effects when the mixed component is more related with signal effects (component #1) and showed aggressive harmonization performance that also removes signal effects when the mixed component was more correlated with site effects variable (component #2). When all the 10 components are spatially independent, after ComBat harmonization, both spatial maps and subject loadings of the mixed component #1 were not changed, in contrast, the mixed component #2 was not related with both signal and site effects variables any longer. When the two mixed components are spatially overlapped, the site effects could not be effectively removed by ComBat when the mixed component was more related to signal variable (component #1) and showed aggressive harmonization performance that removed some signal effects when the mixed component was more related with site effects (component #2). Both spatial maps and subject loadings of the mixed component #1 (more related to signal) did not change significantly, in contrast the mixed component #2 (more related to site effects) was less correlated to both signal and site effects variables.

After being harmonized by ICA-DP, the original mixed components #1 and #2 were merged into a single component that was more related to signal variable and the site-related effects were effectively decreased. Some spatial areas related to site effects were also removed (highlighted with white circles), especially for lower correlation between signal and site effects variables. Fig. 4(b) shows that the removed spatial parts only cover the unique parts of component #2 and do not involve the overlapping parts. Though the merged component after being harmonized by ICA-DP was still mixed component, the correlation coefficient between its loading and

signal variable was strengthened for all the signal to site effects correlation levels, and the correlation levels to site effects variable were contributed by the inherent relationship between signal and site effects variables. Thus, there is no site-specific effect in the mixed component after being harmonized by ICA-DP. ICA-DP showed the most powerful harmonization performance, which could remove all the site-specific effects and enhance the signal effects.

Fig. 4. Harmonization effects on the two mixed components when the signal variable is significantly correlated to the site effects variable. Mix #1 is more related to the signal variable, and Mix #2 is more related to the site effects variable for the non-harmonized data. Three different correlation levels between signal and site effects variables

were simulated in this study to test the harmonization performance of all the harmonization methods. (a) When the spatial maps of all 10 components were spatially independent. (b) Results when the spatial maps of the first two components were spatially overlapped. After being harmonized by ICA-DP, Mix #1 and #2 were merged into a single component more related to signal variable, and the site-related effects were effectively decreased. Some spatial areas related to site effects were also removed (highlighted with a white circle), especially for lower correlation between signal and site effects variables. Among all the harmonization methods, only ICA-DP could effectively weaken the site effects while strengthening the signal effects when signal and site effects variables are correlated.

3.2 Real Datasets Harmonization Results

After harmonization, we performed a set of analyses to show the elimination of site effects and the preservation of biological variability, i.e., HC/ASD and age for the ABIDE II dataset, and subject heterogeneity for the traveling subject dataset. For the data from ABIDE II, they were decomposed into 50, 100 and 150 independent components, by calculating the correlation levels of subjects' loadings and variables with the analysis of variance (ANOVA) for each component, we identified the numbers of pure site effects components were 27, 56 and 96, respectively, and the numbers of mixed components were 23, 42 and 49, respectively. The traveling-subject data were decomposed into 50 independent components, and 10 pure site effects components and 16 mixed components were identified.

Fig. 5 shows the tSNE-2D projection of ADIDE II and traveling subject datasets before and after harmonization. The t-SNE was utilized to project the data into two dimensions by using the two dominant features of the non-harmonized data and harmonized data, to visualize the distribution of site effects and indicate whether it could be eliminated after harmonization. For the ABIDE II dataset, the data points of the non-harmonized data showed site-clustered distribution as most of the centers had their own specific cluster area, except for some intersections among centers UCLA, OHSU, and ETH. And the site-clustered distribution disappeared after being harmonized by any of the harmonization methods. For the traveling subject dataset, the projected data points of the non-harmonized data from the same subject tend to be clustered into one cluster, i.e., the first two projected features were dominated by the subject heterogeneity rather than site effects. Though significant difference was not found before and after harmonization, the subject heterogeneity was well preserved after harmonization.

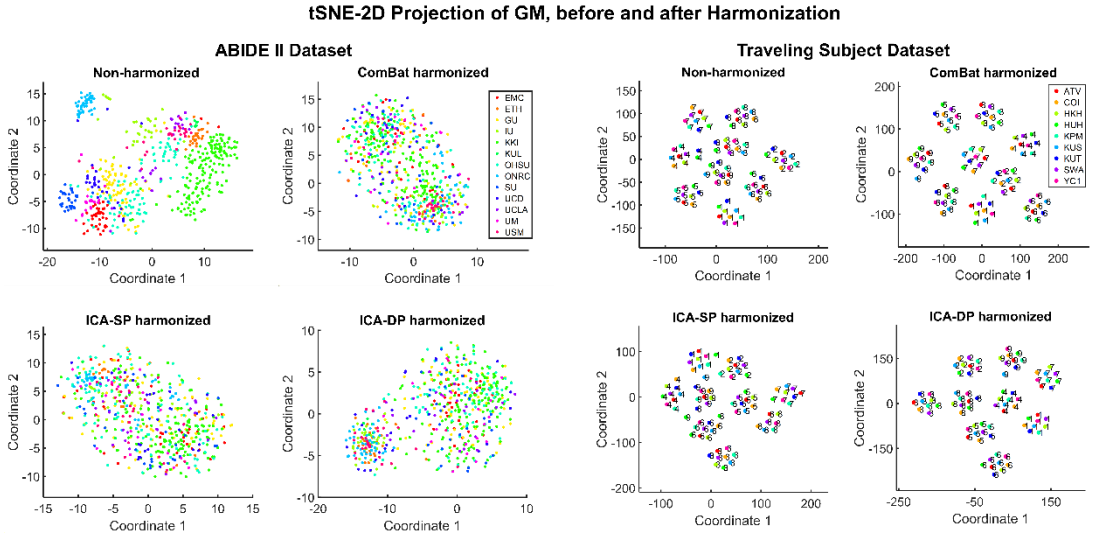

Fig. 5. Dimension reduction visualization by t-SNE before and after harmonization for ABIDE II and traveling subject datasets. The site-cluster distribution of the ABIDE II dataset before harmonization indicated the site effects, and it decreased when the data points were randomly distributed after harmonization. For the traveling subject dataset, the subject-cluster distribution indicated the dominance of subject heterogeneity, as the subjects from this dataset are the same ones scanned at different centers (the data points were labeled by subject numbers). There was no significant difference before and after harmonization, and subject heterogeneity was well preserved after harmonization.

In Fig. 6(a), diagnostic plots were presented for all the subjects from the two datasets, and the different colors represent different sites. For each subject, the GM measurements were summarized into a boxplot. The different range of GM values among sites was reduced after harmonization, and the ICA-based harmonization showed efficient reduction. Fig. 6(b) shows the median GM values distribution of the subjects from different sites. The standard deviation values for the medians of Median GM were calculated across subjects. After harmonization, the site effects decreased noticeably for all the harmonization methods.

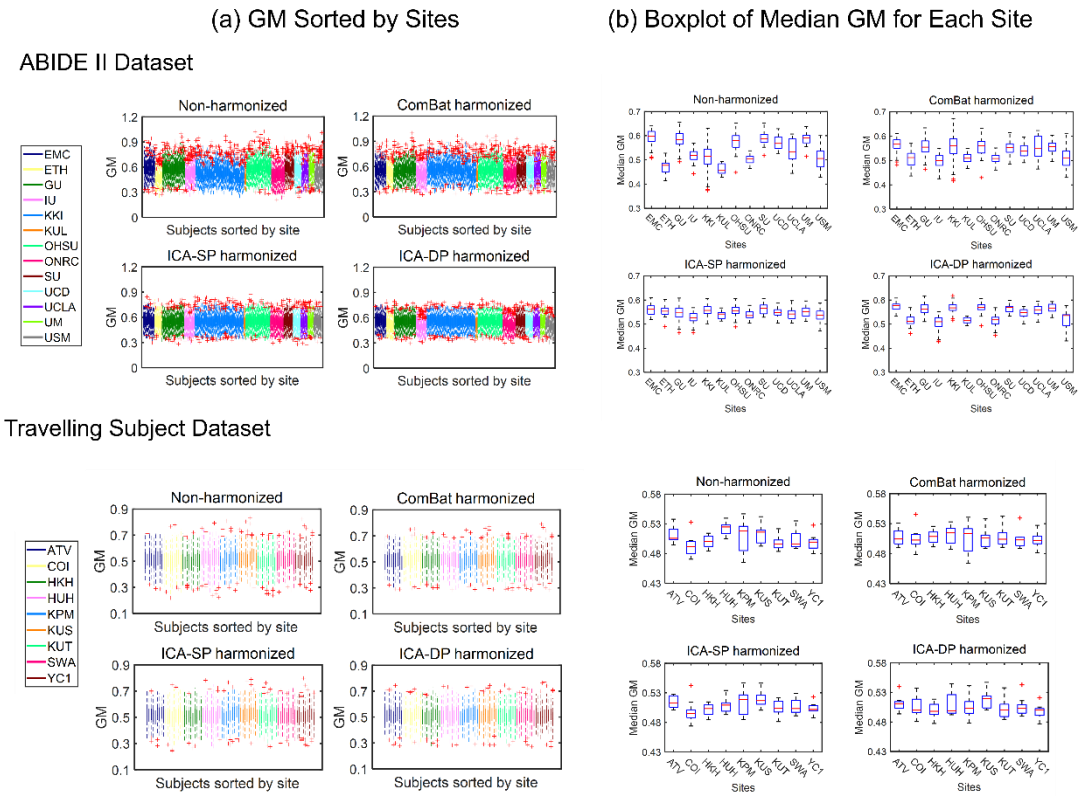

Fig. 6. (a) Site-sorted boxplots of GM. Each boxplot represents the GM values distribution of 100 regions of interest (ROI) for every subject. The ranges indicated differences among sites and among subjects. (b) Site-sorted boxplots of median GM. Each boxplot represents the distribution of median GM values for all subjects from the same site. The fluctuates indicated the inter-site difference. The standard deviation values for the medians of Median GM across subjects, before and after harmonization, are 1)ABIDE II Dataset: 0.0472 (Non-harmonized), 0.0250 (ComBat harmonized), 0.0209 (ICA-SP harmonized), 0.0251 (ICA-DP harmonized); 2)Traveling Subject Dataset: 0.0115 (Non-harmonized), 0.0046 (ComBat harmonized), 0.0078 (ICA-SP harmonized), 0.0071 (ICA-DP harmonized).

The boxplots presented in Fig. 7, for non-harmonization data and harmonization data, summarized the distribution of the median GM for each subject, revealing heterogeneity among different subjects. The subject heterogeneity was not destroyed by all the harmonization methods and preserved well after being harmonized by ICA-DP, and the trends of median GM for each subject, before and after harmonization, are shown in Fig. 7(b). Besides, the intra-subject difference (represented by the height of each box in Fig. 7 (a) and the standard variation of median GM for each subject in Fig. 7 (c)), as a representation of site effects, had been most significantly reduced after ComBat and ICA-DP harmonization (Fig. 7(c)).

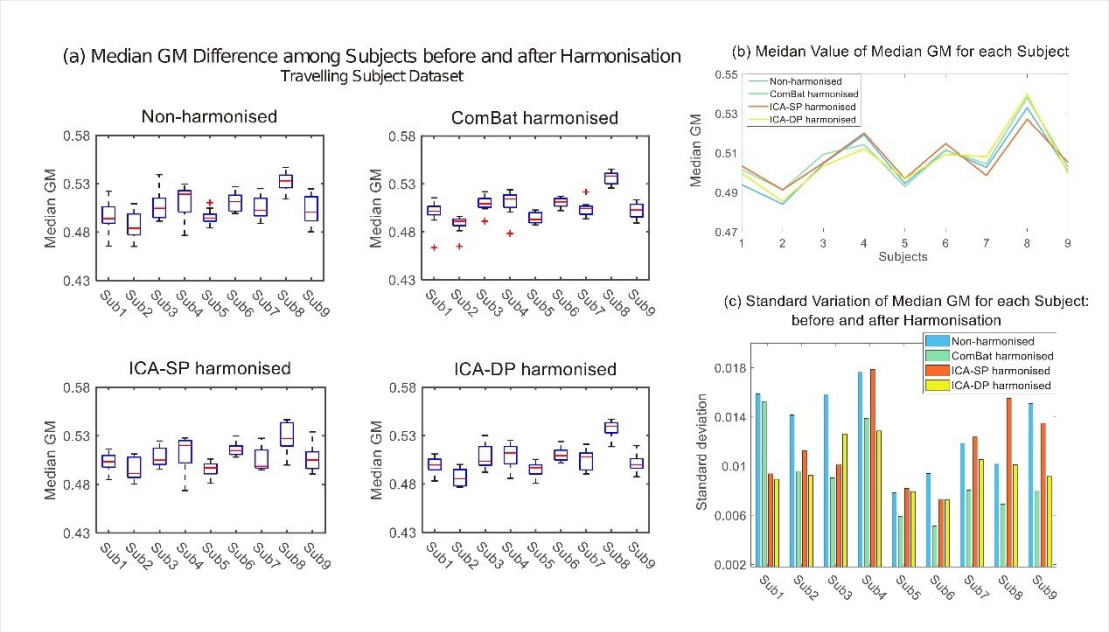

Fig. 7. (a) Subject-sorted boxplots of median GM. The fluctuates indicated the inter-subject difference and the height of each box indicated the intra-subject (inter-site) difference. The standard deviation values for the medians of Median GM across subjects, before and after harmonization, are 0.0147 (Non-harmonized), 0.0139 (ComBat harmonized), 0.0116 (ICA-SP harmonized), and 0.0149 (ICA-DP harmonized). (b) The median values of Meidan GM for each subject, before and after harmonization. The trends show the difference among subjects. (c) The standard deviation value of median GM for each subject, before and after harmonized.

Fig. 8 shows the group-level analysis for site effects from the two datasets. The non-harmonized GM data was globally affected by the site effects for both datasets. Although the site effects had been alleviated by the ICA-SP method, it could not remove them sufficiently. After being harmonized by ICA-DP and ComBat, no significant regions were associated with site variable for both datasets (FWE-corrected $p < 0.05$).

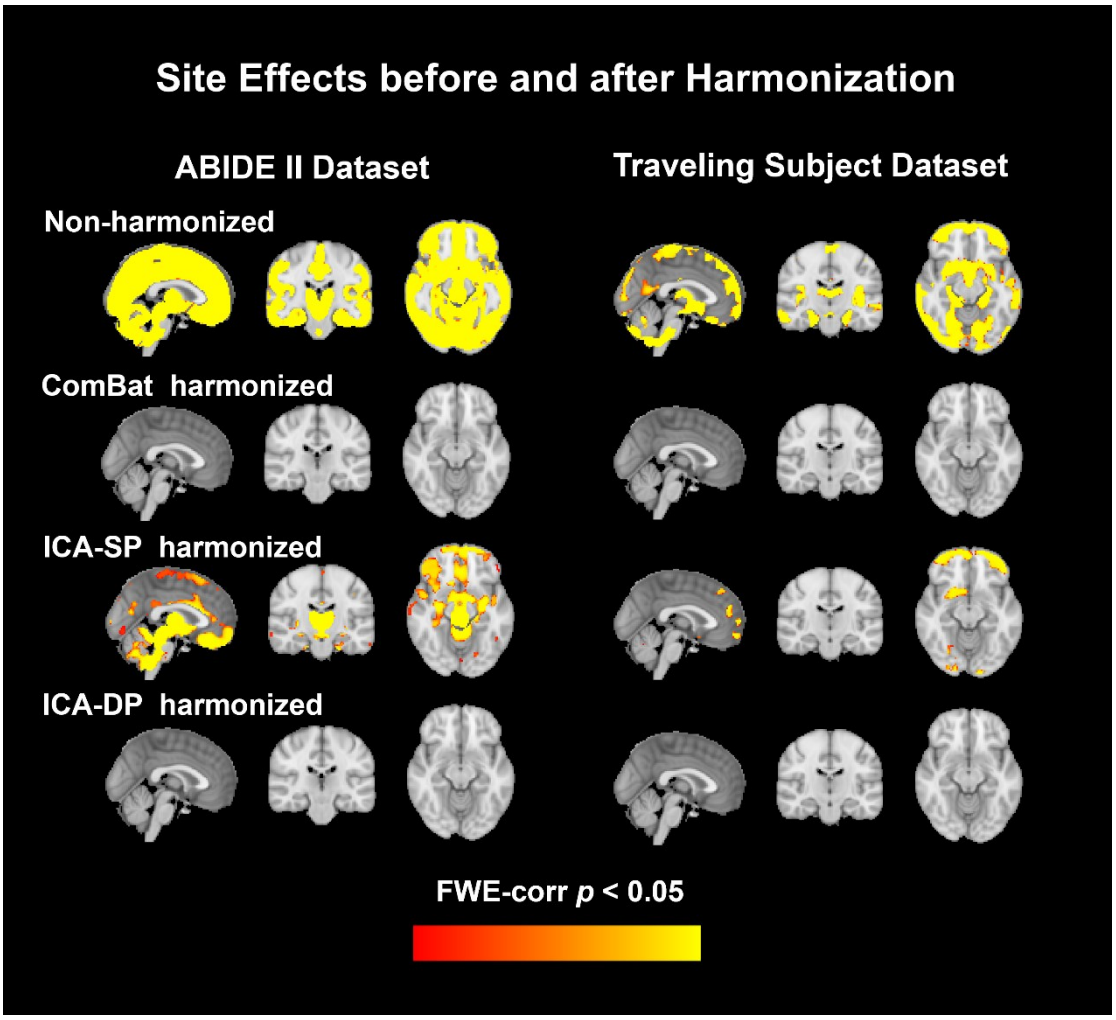

Fig. 8. Group-level analysis for site effects before and after harmonization. Site effects were removed completely by Combat and ICA-DP. Though ICA-SP reduced the site effects, some significant regions still could be found.

As a biological variable of interest, age effects of the ABIDE II dataset, before and after harmonization, were shown in Fig. 9. The correlation between age and median GM before and after harmonization with the ABIDE dataset was shown in Fig. 9(a). The median GM were sorted by age and the data from different scanning centers were in different colors. Pearson correlation coefficients between age and median GM from non-harmonized data and harmonized data were calculated, which were -0.4746 (Non- harmonized), -0.5689 (ComBat harmonized), -0.3617 (ICA-SP harmonized), -0.8493 (ICA-DP harmonized), respectively. The correlation coefficients indicated that the negative correlation between GM and age was strengthened by ICA-DP and ComBat, especially for ICA-DP. Fig. 9(b) shows the group-level analyses for age. Site effects confound us to find the true age effects. The negative age effects were not found in the non-harmonized data because of the existence of site effects, removal of the effects by all the harmonization methods, especially for ICA-DP, could reveal the negative age effects that are not detected from the non-harmonized data.

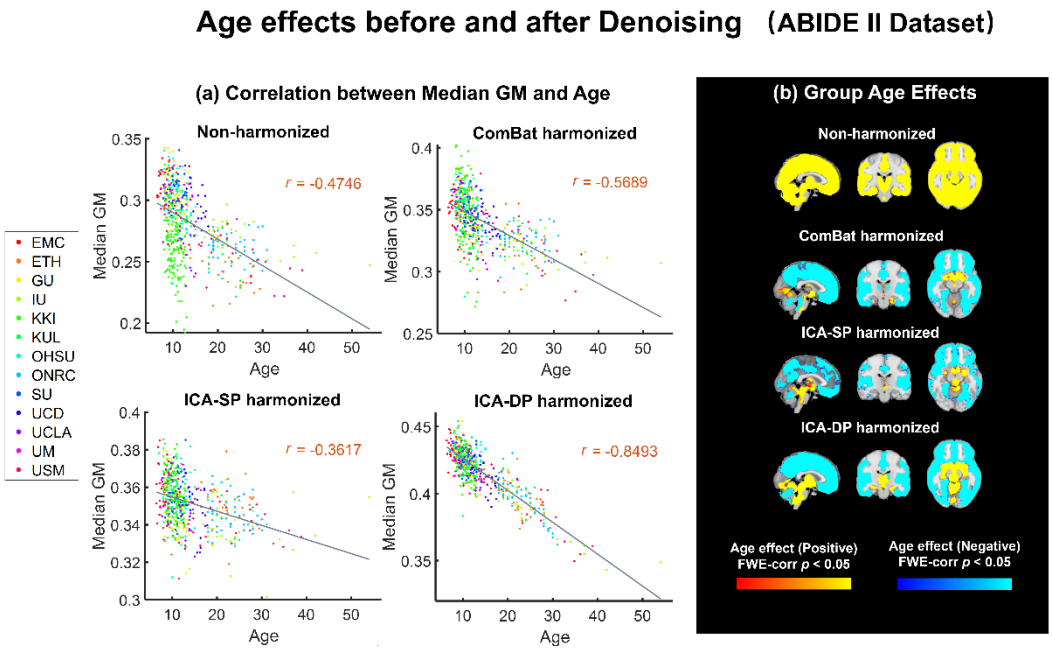

Fig. 9. (a) Relationship between age and median GM before and after site effects harmonization. The Pearson correlation coefficient was -0.4746 (Non-harmonized), -0.5689 (ComBat harmonized), -0.3617 (ICA-SP harmonized), and -0.8493 (ICA-DP harmonized). (b) Group-level analysis of GM maps for age effects before and after data harmonization. The negative age effects (significance level) were enhanced after harmonization, and they could not be detected in the non-harmonized data when testing age group differences.

Fig. 10 shows the group difference (ASD/HC) before and after harmonization. ICA-DP increased the group effects by detecting more significantly different regions related to ASD and HC, while ComBat and ICA-SP decreased the group effects as no significant regions were tested from the data harmonized by them. Compared to the non-thresholded group difference maps from non-harmonized data (first row), the regions associated with group difference (ASD/HC) from ICA-DP-based harmonized data could also be found in the non-harmonized data. In other words, ICA-DP only strengthened the signal that should be there rather than reintroducing artifacts.

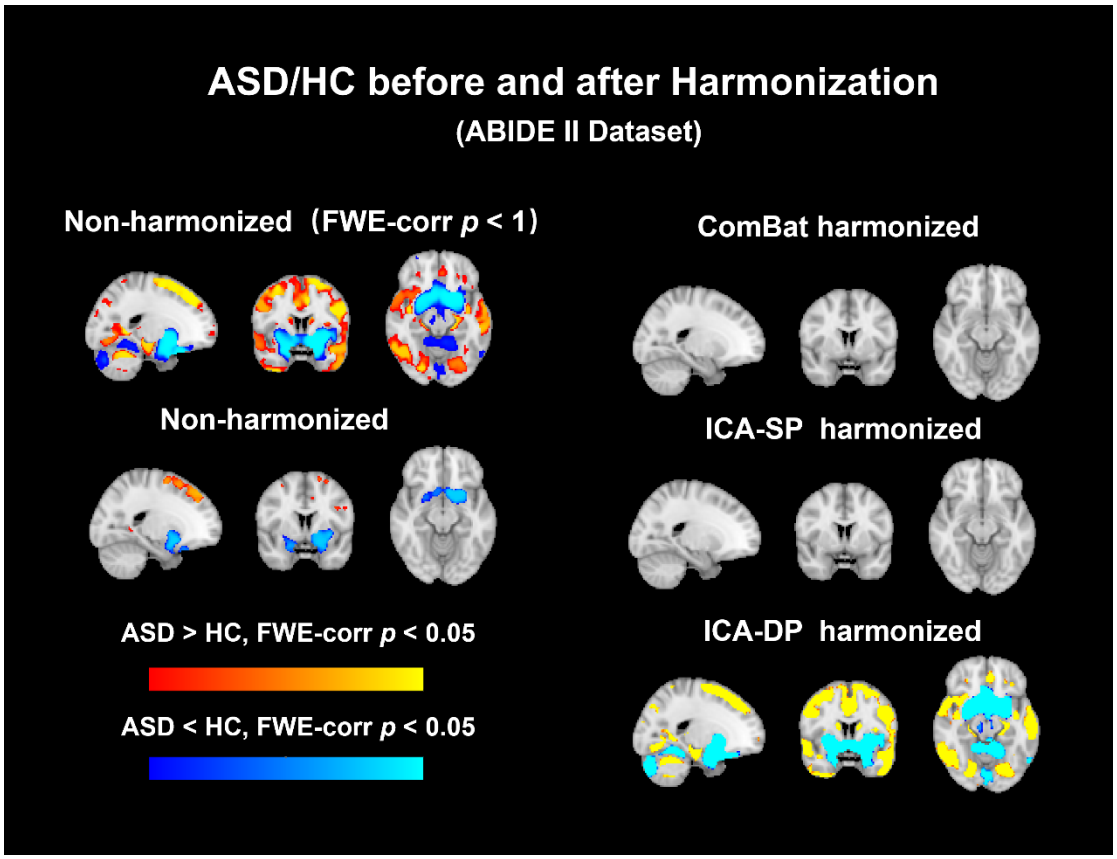

Fig. 10. Group-level analysis of GM maps for group difference (ASD/HC) before and after data harmonization. No significant regions were detected from the data harmonized by ComBat and ICA-SP, while ICA-DP could increase the significance of the regions related to ASD/HC. FWE-corr $p < 1$ was shown for non-harmonized data to indicate that the regions tested from ICA-DP harmonized data were not reintroduced artifacts.

Fig. 11 shows the harmonization performance of the two ICA-based harmonization methods under different choices of component numbers when running ICA algorithms. The ICA-SP could not remove the site effect completely, though it decreased more site effects as the number of components increased, and the information related to signal variable (ASD/HC) could not be detected from the data harmonized by it under any component number choosing, indicating that this kind of soft harmonization based on ICA could neither remove the site effects completely nor reveal the information related to covariates of interest. In contrast, after being harmonized by ICA-DP, the information that related to site effects could not be tested and there were some regions that significantly correlated to ASD/HC could be revealed from the harmonized data, indicating good performance and importance for eliminating site effects and the ability to unveil the signal related information concurrently and showing no affection of which number of components were chosen.

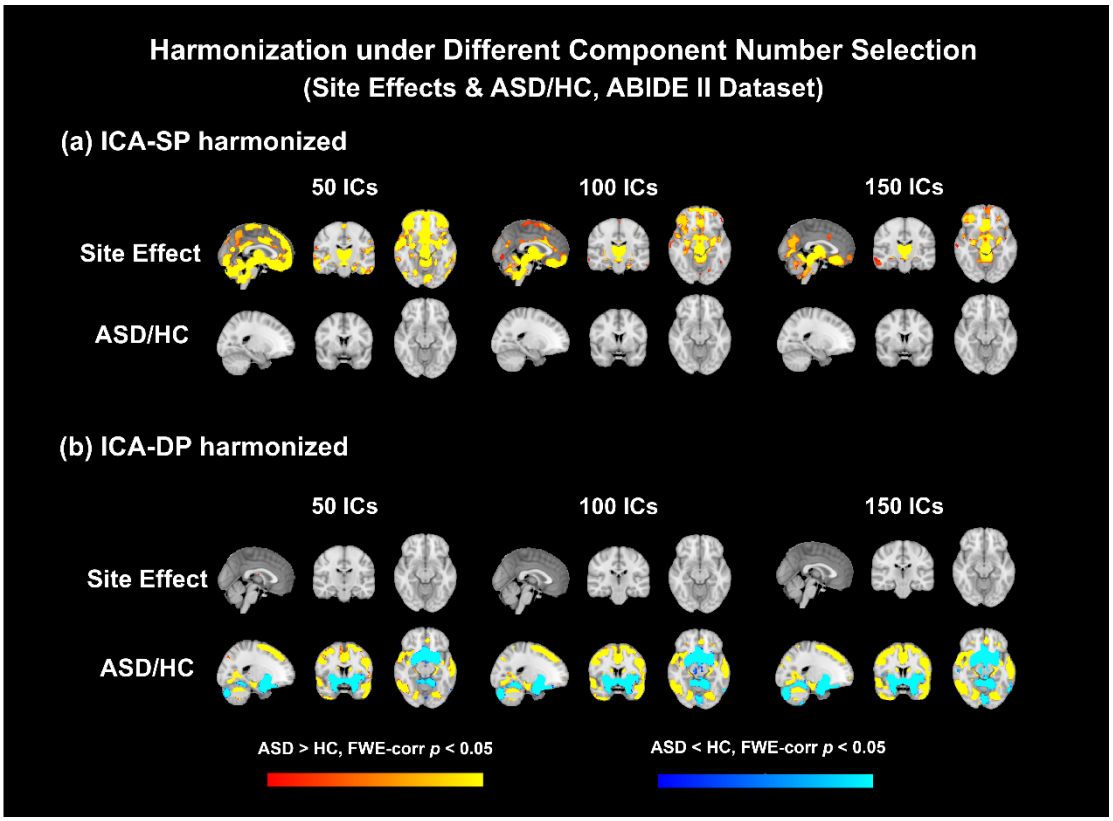

Fig. 11. Harmonization performance of the two ICA-based harmonization methods under different component numbers (50, 100, and 150). After being harmonized by ICA-DP, site effects were removed completely and ASD/HC group differences were significantly enhanced (the regions related to ASD/HC could not be detected before harmonization) without the affection of component number.

4 Discussion

In this paper, we proposed a dual-projection ICA-based harmonization method that can remove the site effects more effectively and completely while enhancing the signal effects. This method shows superior performance when the site effects are also related with signal variables. The Matlab codes of ICA-DP harmonization are available at https://github.com/Yuxing-Hao/ICA-DP_Harmonization.git.

ICA-DP was developed to clean the site effects more completely and effectively when the site and signal effects are correlated. Based on the simulation results (Figs. 3, 4), it was found that ICA-DP can eliminate the site effects effectively while enhancing the signal-related information, whether the site and signal variables are correlated. The benefits of ICA-DP are contributed by two reasons: first, ICA-DP can extract all the site-related effects with the first projection even though some site effects are mixed with the signal effects that could not be extracted from original ICA method; second, ICA-DP is more stable than original ICA method, as it can extract and remove all the site effects without the limitation of model order selection; third, compared to standard GLM and ComBat model whose regressors have the same value for all participants in the same sites, site-related effects are better captured by ICA-DP, which can capture day-to-day variations in scanner performance. Compared with ICA-DP, ICA-SP,

GLM and ComBat showed different defects when harmonization the site effects. For ICA-SP, it encountered a problem that it failed to eliminate the site effects completely when the site-related components also significantly correlated to signal variables (Figs. 3, 4). The GLM-based harmonization method performed well and was comparable to ICA-DP when the signal and site effects variables are not significantly correlated (Fig. 3(a)). However, GLM showed aggressive harmonization performance that will remove or decrease the signal-related information when the signal and site effects variables are significantly correlated. The signal distortion became worse when the correlation coefficients between the signal and site effects variables increased (Fig. 4). Though ComBat-based harmonization method showed better performance than GLM and original ICA-based harmonization performance which can effectively remove site effects while keeping or strengthening some signal effects when signal and site variables are not correlated. However, ComBat could not completely remove the site effects when signal and site variables are correlated when the mixed components are more correlated with the signal (Fig. 4). Meanwhile, we found that ComBat showed aggressive harmonization that also removed signal effects when the mixed components are more related with site effects.

Based on the results of the ABIDE II dataset and the traveling-subject dataset, ICA-DP also shows superior performance on harmonization site effects than ICA-SP and ComBat. The significant regions of GM that related to site effects, which cannot be completely removed by ICA-SP, were not detectable after harmonizing with ICA-DP and ComBat (Fig. 8). The inter-site variation of the traveling data and the intra-subject variation among nine sites for the traveling subject dataset were most significantly reduced after harmonization with ICA-DP (Fig. 6). Moreover, subject heterogeneity of the traveling subject dataset was also well preserved after being harmonized by ICA-DP (Figs. 5, 7).

In addition, ICA-DP also shows superior performance in enhancing biological variability (i.e., age effects and group difference between ASD and HC) compared with ICA-SP and ComBat based on the results of the ABIDE II dataset. Site effects hinder us from finding the true age and group effects. The age effects detected with GM of the non-harmonized ABIDE II are opposite with the recognized results (Gennatas et al., 2017; Groves et al., 2012). After being harmonized by ICA-SP, ICA-DP, and ComBat, the true age effects on GM were discovered. Among all the three methods, ICA-DP finds the most significant regions related to age and the negative correlation between age and GM was most strongly enhanced by ICA-DP. The relationship of median GM and age was enhanced from -0.4746 (non-harmonized) to -0.8493 after being harmonized by ICA-DP (Fig. 9). In addition, compared to ICA-SP and ComBat, only ICA-DP enhanced the group effects (ASD vs. HC) by detecting more significantly group different regions which cannot be effectively detected by the non-harmonized data (Fig. 10), indicating the importance of removing site effects in multi-site data. The notable enhancements of the biological variabilities (i.e., age effects and ASD vs. HC) may be attributed to the larger proportion of site-related components that we selected for ICA-DP harmonization, which could increase the weights of signal we are interested in and make the signal-related information

easier to be detected. On the other hand, this may lead to the other variables that we are not interested in not being well preserved. To protect other variables that we may be interested in, we just need to add these variables to $\mathbf{V}_S$ in the first projection of the ICA-DP harmonization method (Eq. (1)). Thus, the ICA-DP is the most effective method for harmonizing site effects and preserving biological variability among the methods discussed above. Moreover, unlike ICA-SP, the performance of ICA-DP in site effects harmonization and signal enhancement was not affected by the number of components chosen for ICA decomposition. It could clean the site effects and strengthen the signal under any selected component number (Fig. 11).

Finally, as a limitation of the proposed harmonization method, when the site effects variable is strongly related to the signal variable (Fig. 4), ICA-DP could not eliminate the intersection effects that are related with both site and signal variables (neither do other methods except GLM, the most aggressive one destroying signal-related information severely), thus the harmonized data are still correlated to site effects because of the inherent correlation between site effects and signal variables (Nevertheless, the correlation values in our simulation is really high and hardly appear in real data study). Another limitation to consider is the selection of the number of components to be extracted by ICA. Though we validated that the performance of ICA-DP in site effects harmonization and signal enhancement were not affected by the number of components chosen for ICA decomposition under three different selections for the number of components (see Figure. 11), it is not sufficient and verification methods should be developed further. However, to some extent, ICA-DP allows users to choose the number of components according to their own standard.

Overall, the dual-projection harmonization method is more effective and powerful in removing site effects while preserving signal-related information than other methods mentioned above, and can enhance the sensitivity to detect signals of interest and remove all the effects that are only contributed by site difference. Compared to Combat, it is a data-driven method rather than utilizing the manually designed covariates for regressing. ICA-DP harmonization method has great potential for large-scale multi-site studies to produce combined data free from study-site confounds.

5 Conclusion

While combing the multi-site MRI data has great convenience that enhances the statistical results and obviates some of the shortcomings of the single-site study, the site effects come naturally, confounding the MRI data analysis and making the results hard to interpret. The traditional methods designed to eliminate the site effects encounter incomplete or aggressive harmonization, i.e., cannot eliminate the site effects well or may destroy the signal-related information. To tackle these shortcomings, we proposed a dual-projection data-driven method based on ICA, which can better eliminate the site effects and preserve the signals of interest. And we strongly recommend that researchers use the ICA-DP method to harmonize the MRI

data as it can extract subject-specific loadings that correspond to the signal or site effects variable.

Abbreviations: DP, dual-projection; DTI, diffusion tensor image; GLM, general linear model; GM, gray matter; ICA, independent component analysis; ICA-DP, ICA-dual projection; ICA-SP, ICA-single projection; LICA, linked ICA; MRI, magnetic resonance imaging; PET, positron emission tomography; t-SNE, t-distributed stochastic neighbor embedding.

References

Andersson, J. L., Jenkinson, M., & Smith, S. (2007). Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2. *FMRIB Analysis Group of the University of Oxford*, *2*(1), e21.

Beckmann, C. F., Mackay, C. E., Filippini, N., & Smith, S. M. (2009). Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *NeuroImage*, *47*(Supp1 1), S148.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Bell, T. K., Godfrey, K. J., Ware, A. L., Yeates, K. O., & Harris, A. D. (2022). Harmonization of multi-site MRS data with ComBat. *NeuroImage*, *257*, 119330. https://doi.org/10.1016/j.neuroimage.2022.119330

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafo, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

C Monte-Rubio, G., Segura, B., P Strafella, A., van Eimeren, T., Ibarretxe-Bilbao, N., Diez-Cirarda, M., Eggers, C., Lucas-Jimenez, O., Ojeda, N., & Pena, J. (2022). Parameters from site classification to harmonize MRI clinical studies: Application to a multi-site Parkinson's disease dataset. *Human Brain Mapping*, *43*(10), 3130–3142.

Casey, B. J., Cohen, J. D., O'Craven, K., Davidson, R. J., Irwin, W., Nelson, C. A., Noll, D. C., Hu, X., Lowe, M. J., Rosen, B. R., Truwitt, C. L., & Turski, P. A. (1998). Reproducibility of fMRI results across four institutions using a spatial working memory task. *NeuroImage*, *8*(3), 249–261. https://doi.org/10.1006/nimg.1998.0360

Chen, J., Liu, J., Calhoun, V. D., Arias-Vasquez, A., Zwiers, M. P., Gupta, C. N., Franke, B., & Turner, J. A. (2014). Exploration of scanning effects in multi-site structural MRI studies. *Journal of Neuroscience Methods*, *230*, 37–50. https://doi.org/10.1016/j.jneumeth.2014.04.023

Di Martino, A., O'connor, D., Chen, B., Alaerts, K., Anderson, J. S., Assaf, M., Balsters, J. H., Baxter, L., Beggiato, A., & Bernaerts, S. (2017). Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Scientific Data*, *4*(1), 1–15.

Dinsdale, N. K., Jenkinson, M., & Namburete, A. I. L. (2021). Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. *NeuroImage*, *228*(December 2020), 117689. https://doi.org/10.1016/j.neuroimage.2020.117689

Douaud, G., Smith, S., Jenkinson, M., Behrens, T., Johansen-Berg, H., Vickers, J., James, S., Voets, N., Watkins, K., & Matthews, P. M. (2007). Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. *Brain*, *130*(9), 2375–2386.

Eickhoff, S., Nichols, T. E., Van Horn, J. D., & Turner, J. A. (2016). Sharing the wealth: Neuroimaging data repositories. *NeuroImage*, *124*, 1065–1068. https://doi.org/10.1016/j.neuroimage.2015.10.079

Fennema-Notestine, C., Gamst, A. C., Quinn, B. T., Pacheco, J., Jernigan, T. L., Thal, L., Buckner, R., Killiany, R., Blacker, D., Dale, A. M., Fischl, B., Dickerson, B., & Gollub, R. L. (2007). Feasibility of multi-site clinical structural neuroimaging studies of aging using legacy data. *Neuroinformatics*, *5*(4), 235–245. https://doi.org/10.1007/s12021-007-9003-9

Filippini, N., MacIntosh, B. J., Hough, M. G., Goodwin, G. M., Frisoni, G. B., Smith, S. M., Matthews, P. M., Beckmann, C. F., & Mackay, C. E. (2009). Distinct patterns of brain activity in young carriers of the APOE-ε4 allele. *Proceedings of the National Academy of Sciences*, *106*(17), 7209–7214.

Focke, N. K., Helms, G., Kaspar, S., Diederich, C., Tóth, V., Dechent, P., Mohr, A., & Paulus, W. (2011). Multi-site voxel-based morphometry - Not quite there yet. *NeuroImage*, *56*(3), 1164–1170. https://doi.org/10.1016/j.neuroimage.2011.02.029

Fortin, J. P., Parker, D., Tunç, 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. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, *161*(July), 149–170. https://doi.org/10.1016/j.neuroimage.2017.08.047

Friedman, L., Stern, H., Brown, G. G., Mathalon, D. H., Turner, J., Glover, G. H., Gollub, R. L., Lauriello, J., Lim, K. O., Cannon, T., Greve, D. N., Bockholt, H. J., Belger, A., Mueller, B., Doty, M. J., He, J., Wells, W., Smyth, P., Pieper, S., … Potkin, S. G. (2008). Test-retest and between-site reliability in a multicenter fMRI study. *Human Brain Mapping*, *29*(8), 958–972. https://doi.org/10.1002/hbm.20440

Gennatas, E. D., Avants, B. B., Wolf, D. H., Satterthwaite, T. D., Ruparel, K., Ciric, R., Hakonarson, H., Gur, R. E., & Gur, R. C. (2017). Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. *Journal of Neuroscience*, *37*(20), 5065–5073.

Glover, G. H., Mueller, B. A., Turner, J. A., Van Erp, T. G. M., Liu, T. T., Greve, D. N., Voyvodic, J. T., Rasmussen, J., Brown, G. G., Keator, D. B., Calhoun, V. D., Lee, H. J., Ford, J. M., Mathalon, D. H., Diaz, M., O'Leary, D. S., Gadde, S., Preda, A., Lim, K. O., … Potkin, S. G. (2012). Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. *Journal of Magnetic Resonance Imaging*, *36*(1), 39–54. https://doi.org/10.1002/jmri.23572

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N., Friston, K. J., & Frackowiak, R. S. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage*, *14*(1), 21–36.

Groves, A. R., Beckmann, C. F., Smith, S. M., & Woolrich, M. W. (2011). Linked independent component analysis for multi-modal data fusion. *Neuroimage*, *54*(3), 2198–2217.

Groves, A. R., Smith, S. M., Fjell, A. M., Tamnes, C. K., Walhovd, K. B., Douaud, G., Woolrich, M. W., & Westlye, L. T. (2012). Benefits of multi-modal fusion analysis on a large-scale dataset: life-span patterns of inter-subject variability in cortical morphometry and white

matter microstructure. *Neuroimage*, *63*(1), 365–380.

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Jovicich, J., Czanner, S., Han, X., Salat, D., Kouwe, A. Van Der, Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Rosas, D., Makris, N., Gollub, R., & Dale, A. (2009). MRI-derived measurements of human subcortical, ventricular andintracranial brain volumes: Reliability effects of scan sessions,acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *NeuroImage*, *46*(1), 177–192. https://doi.org/10.1016/j.neuroimage.2009.02.010.MRI-derived

Li, H., Smith, S. M., Gruber, S., Lukas, S. E., Silveri, M. M., Hill, K. P., Killgore, W. D., & Nickerson, L. D. (2020). Denoising scanner effects from multi-modal MRI data using linked independent component analysis. *Neuroimage*, *208*, 116388.

Maikusa, N., Zhu, Y., Uematsu, A., Yamashita, A., Saotome, K., Okada, N., Kasai, K., Okanoya, K., Yamashita, O., & Tanaka, S. C. (2021). Comparison of traveling-subject and ComBat harmonization methods for assessing structural brain characteristics. *Human Brain Mapping*, *42*(16), 5278–5287.

Nickerson, L. D., Smith, S. M., Ongur, D., & Beckmann, C. F. (2017). Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Frontiers in Neuroscience*, *11*, 115.

Orlhac, F., Boughdad, S., Philippe, C., Stalla-Bourdillon, H., Nioche, C., Champion, L., Soussan, M., Frouin, F., Frouin, V., & Buvat, I. (2018). A postreconstruction harmonization method for multicenter radiomic studies in PET. *Journal of Nuclear Medicine*, *59*(8), 1321–1328. https://doi.org/10.2967/jnumed.117.199935

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., De Bellis, M. D., Voyvodic, J. T., Clark, D. B., Schirda, C., … Pfefferbaum, A. (2016). Harmonizing DTI measurements across scanners to examine the development of white matter microstructure in 803 adolescents of the NCANDA study. *NeuroImage*, *130*, 194–213. https://doi.org/10.1016/j.neuroimage.2016.01.061

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., & Flitney, D. E. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23*, S208--S219.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, *44*(1), 83–98.

Takao, H., Hayashi, N., & Ohtomo, K. (2011). Effect of scanner in longitudinal studies of brain volume changes. *Journal of Magnetic Resonance Imaging*, *34*(2), 438–444. https://doi.org/10.1002/jmri.22636

Tanaka, S. C., Yamashita, A., Yahata, N., Itahashi, T., Lisi, G., Yamada, T., Ichikawa, N.,

Takamura, M., Yoshihara, Y., & Kunimatsu, A. (2021). A multi-site, multi-disorder resting-state magnetic resonance image database. *Scientific Data*, *8*(1), 1–15.

Tian, D., Zeng, Z., Sun, X., Tong, Q., Li, H., He, H., Gao, J. H., He, Y., & Xia, M. (2022). A deep learning-based multisite neuroimage harmonization framework established with a traveling-subject dataset. *NeuroImage*, *257*(March), 119297. https://doi.org/10.1016/j.neuroimage.2022.119297

Van Horn, J. D., & Toga, A. W. (2009). Multisite neuroimaging trials. *Current Opinion in Neurology*, *22*(4), 370–378. https://doi.org/10.1097/WCO.0b013e32832d92de

Venkatraman, V. K., Gonzalez, C. E., Landman, B., Goh, J., Reiter, D. A., An, Y., & Resnick, S. M. (2015). Region of interest correction factors improve reliability of diffusion imaging measures within and across scanners and field strengths. *Neuroimage*, *119*, 406–416.

Vollmar, C., O'Muircheartaigh, J., Barker, G. J., Symms, M. R., Thompson, P., Kumari, V., Duncan, J. S., Richardson, M. P., & Koepp, M. J. (2010). Identical, but not the same: Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. *NeuroImage*, *51*(4), 1384–1394. https://doi.org/10.1016/j.neuroimage.2010.03.046

Wegner, C., Filippi, M., Korteweg, T., Beckmann, C., Ciccarelli, O., De Stefano, N., Enzinger, C., Fazekas, F., Agosta, F., Gass, A., Hirsch, J., Johansen-Berg, H., Kappos, L., Barkhof, F., Polman, C., Mancini, L., Manfredonia, F., Marino, S., Miller, D. H., … Matthews, P. M. (2008). Relating functional changes during hand movement to clinical parameters in patients with multiple sclerosis in a multi-centre fMRI study. *European Journal of Neurology*, *15*(2), 113–122. https://doi.org/10.1111/j.1468-1331.2007.02027.x

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage*, *92*, 381–397.

Yamashita, A., Yahata, N., Itahashi, T., Lisi, G., Yamada, T., Ichikawa, N., Takamura, M., Yoshihara, Y., Kunimatsu, A., Okada, N., Yamagata, H., Matsuo, K., Hashimoto, R., Okada, G., Sakai, Y., Morimoto, J., Narumoto, J., Shimada, Y., Kasai, K., … Imamizu, H. (2019). Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. In *PLoS Biology*. https://doi.org/10.1101/440875

Yu, M., Linn, K. A., Cook, P. A., Phillips, M. L., McInnis, M., Fava, M., Trivedi, M. H., Weissman, M. M., Shinohara, R. T., & Sheline, Y. I. (2018). Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Human Brain Mapping*, *39*(11), 4213–4227. https://doi.org/10.1002/hbm.24241

Zivadinov, R., & Cox, J. L. (2008). Is functional MRI feasible for multi-center studies on multiple sclerosis? *European Journal of Neurology*, *15*(2), 109–110. https://doi.org/10.1111/j.1468-1331.2007.02030.x