



## dStripe: slice artefact correction in diffusion MRI via constrained neural network

Maximilian Pietsch<sup>a,b,c,\*</sup>, Daan Christiaens<sup>b,d</sup>, Joseph V Hajnal<sup>a,b</sup>, J-Donald Tournier<sup>a,b</sup>

<sup>a</sup>Centre for Medical Engineering, King's College London, London, UK

<sup>b</sup>Centre for the Developing Brain, King's College London, London, UK

<sup>c</sup>Department of Forensic & Neurodevelopmental Sciences, King's College London, London, UK

<sup>d</sup>Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium

### ARTICLE INFO

#### Article history:

diffusion MRI; image artefact removal;  
venetian blind artefact

### ABSTRACT

MRI scanner and sequence imperfections and advances in reconstruction and imaging techniques to increase motion robustness can lead to inter-slice intensity variations in Echo Planar Imaging. Leveraging deep convolutional neural networks as universal image filters, we present a data-driven method for the correction of acquisition artefacts that manifest as inter-slice inconsistencies, regardless of their origin. This technique can be applied to motion- and dropout-artefacted data by embedding it in a reconstruction pipeline. The network is trained in the absence of ground-truth data on, and finally applied to, the reconstructed multi-shell high angular resolution diffusion imaging signal to produce a corrective slice intensity modulation field. This correction can be performed in either motion-corrected or scattered source-space. We focus on gaining control over the learned filter and the image data consistency via built-in spatial frequency and intensity constraints. The end product is a corrected image reconstructed from the original raw data, modulated by a multiplicative field that can be inspected and verified to match the expected features of the artefact. In-plane, the correction approximately preserves the contrast of the diffusion signal and throughout the image series, it reduces inter-slice inconsistencies within and across subjects without biasing the data. We apply our pipeline to enhance the super-resolution reconstruction of neonatal multi-shell high angular resolution data as acquired in the developing Human Connectome Project.

© 2021 Elsevier B. V. All rights reserved.

### 1. Introduction

Diffusion MRI (dMRI) provides unique information about the microstructural properties of brain tissue by sensitisation to the motion of water molecules on the order of micrometers via strong gradient amplitudes. However, this poses a major challenge for in-vivo imaging where bulk subject motion or flow can cause severe phase errors.

In single-shot echo planar imaging (EPI) (Mansfield, 1977;

Wu and Miller, 2017), the k-space data of a 2D image can be encoded within typically 100 ms after a single excitation which effectively freezes motion. A 3D image can be formed by acquiring a stack of parallel 2D EPI images at different slice positions. In EPI, interactions with previous pulses (spin-history effects) and interference across slices (stimulated echo artefacts (Burstein, 1996; Crawley and Henkelman, 1987)), variations in slice timing, imperfect signal unmixing in simultaneous multi-slice (SMS) imaging, and scanner hardware limitations can all lead to inter-slice inconsistencies.

Sensitivity to subject motion and the push to higher in-plane and slice acceleration exacerbates the potential for inter-slice

\*Corresponding author  
e-mail: [maximilian.pietsch@kcl.ac.uk](mailto:maximilian.pietsch@kcl.ac.uk) (Maximilian Pietsch)

inconsistencies. This can destabilise super-resolution reconstruction algorithms and affect downstream data analyses. In this work, we address the problem of removing inter-slice inconsistencies in neonatal dMRI data of the brain, a cohort that is particularly prone to motion and, due to relatively long  $T_1$  relaxation times, spin history effects.

### 1.1. EPI slice intensity inconsistencies

Techniques to reduce acquisition time, such as the use of EPI, often combined with partial Fourier (McGibney *et al.*, 1993), parallel imaging (Larkman and Nunes, 2007; Pruessmann *et al.*, 1999; Griswold *et al.*, 2002), and simultaneous multi-slice (SMS) (Larkman *et al.*, 2001; Barth *et al.*, 2016; Setsompop *et al.*, 2012; Blaimer *et al.*, 2013) are prone to image degradation, particularly in the presence of motion. For example, inter-slice and intra-slice signal leakage in accelerated slice-Grappa EPI acquisitions from neuronal (Cauley *et al.*, 2014; Todd *et al.*, 2016) and non-neuronal origin, such as eye blinking (McNabb *et al.*, 2018) and motion (Kelly *et al.*, 2013; Muresan *et al.*, 2005), cannot be fully suppressed, which can lead to false positive activations in fMRI analysis (Todd *et al.*, 2016; McNabb *et al.*, 2018). Variations in slice timings and subject motion orthogonal to the slice plane result in a temporary disruption to the steady-state and yield pose-, motion- and tissue-dependent intensity modulations (spin-history artefacts) (Friston *et al.*, 1996); this is exacerbated when using SMS techniques, as they can reduce the repetition time (TR) to the order of typical  $T_1$  times for brain tissue. Spin-history artefacts are typically 3 to 7% of the image intensity and difficult to model (Muresan *et al.*, 2005) as they are non-linearly related to motion trajectories (Yancey *et al.*, 2011). They can also be caused by localised motion related to breathing (Friston *et al.*, 1996) and cardiac pulsation (Dagli *et al.*, 1999).

In this study, we aim to remove inter-slice artefacts in neonatal multi-shell dMRI scans that were acquired as part of the developing Human Connectome Project ([www.developingconnectome.org](http://www.developingconnectome.org)). In this data, we observe inter-slice intensity variations of unknown origin: the shell-average of the raw images in scattered source space shows a clear stripe pattern (see figure 1). On a population level, parts of the artefact pattern seem to be linked with the diffusion gradient strength, acquisition order and multiband boundaries. In the shell-average images of a single subject with the least motion in the cohort as well as in the population average across 700 subjects, we observe that the intensity modulation seems to be relatively smooth in-plane, with sharp transitions in the through-plane direction. The stripe modulation artefact has a similar appearance to motion-related slice-wise intensity artefacts. These are frequently observed in the neonatal cohort and can dominate the slice modulation we aim to correct as demonstrated in the shell-average of the subject with median motion in figure 1. The properties of the artefact are described more comprehensively in the Results and Discussion section, where we assess it using spatial and angular constraints applied to estimated destripe fields.

### 1.2. Prior and related work in MRI artefact removal

In fMRI analysis, motion-induced spin-history artefacts are commonly removed via nuisance regressors (Murphy *et al.*, 2013; Kelly *et al.*, 2013) but there is currently no standard method for removing spin-history artefacts from diffusion-weighted images (Andersson and Sotiropoulos, 2016). Variable slice-timing artefacts are best avoided at the setup of the acquisition sequence but this may not be possible for all applications and scanner software versions (Meyer *et al.*, 2014) and retrospective correction might be required for already acquired data.

A suggested retrospective zeroth-order correction of intensity inconsistencies is to apply a single scale factor to even slices to match average intensities between those and adjacent slices (David *et al.*, 2017). While the computational simplicity and transparency is appealing, this approach does not account for spatially variable or tissue-specific intensity modulation, is potentially biased if the true average image intensities do vary between slices and is not applicable to motion corrupted data. Prospective methods have been proposed to deal with similar artefacts in multi-slab acquisition, but to date these do not eliminate the problem (Engström and Skare, 2013; Frost *et al.*, 2014; Van *et al.*, 2011; Parker *et al.*, 1991; Van *et al.*, 2015; Wu *et al.*, 2016; Setsompop *et al.*, 2018; Liao *et al.*, 2018).

### 1.3. Problem formulation and contribution

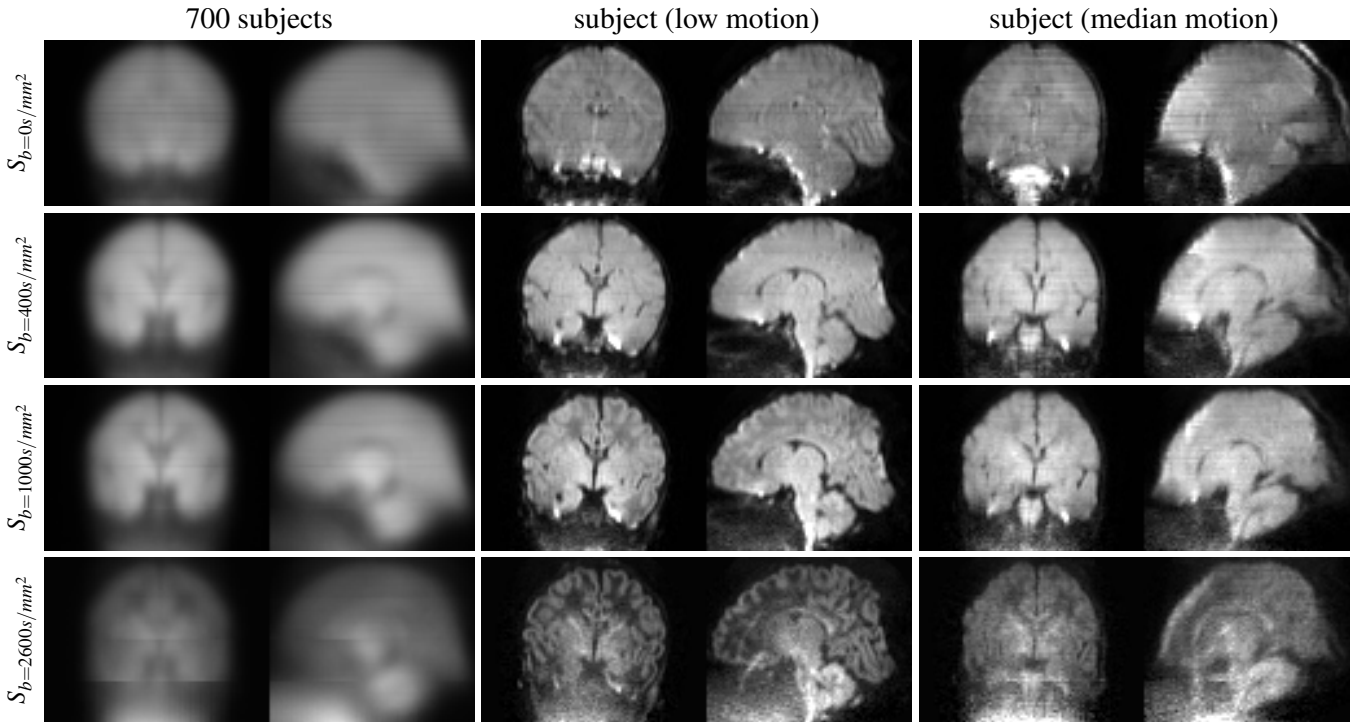
A physical model of the artefact or paired corrupted and uncorrupted data could inform us on how to model the correction; whether the stripe correction is best modelled as a linear or non-linear, local or non-local function of the data, what constraints to use and whether the model is tied to acquisition-space or to the subject anatomy or to a mixture of both. In the absence of this prior knowledge, we aim to make informed and data integrity-preserving decisions.

The effect of the artefact correction should be directly inspectable and, if possible, the model should be guaranteed to produce corrections that match the observed patterns of the artefact. To maintain the trustworthiness of the images, any degree of freedom beyond that should ideally be restricted.

We assume the artefact to be a smooth multiplicative voxel-wise correction field per acquired slice. In the absence of constraints, this choice does not limit the space of possible solutions. However, this choice matters when enforcing inplane smoothness constraints. A multiplicative smooth in-plane field – similar to a B1 inhomogeneity correction field – ensures that the correction approximately preserves local image contrast within each acquired slice.

We build on (Pietsch *et al.*, 2019b) to create a method that facilitates the data-driven and model-free removal of inter-slice inconsistencies of potentially motion-artefacted diffusion-sensitised images in scattered source space or in motion-corrected space. Our framework:

- can be applied retrospectively
- works in tandem with motion correction techniques to remove stripes in the presence of subject motion
- is not tied to a particular  $q$ -space sampling scheme or motion correction technique



**Fig. 1.** Raw (before motion and distortion correction) signal averaged within shells using data from the AP phase-encode direction. The single-subject data is from the scan with the highest (middle) and with the median motion-correction derived quality assessment score (right) in the cohort. The population-averaged raw signal shows clear slice-wise intensity modulation patterns that exhibit a b-value dependency. The population-average pattern is to some degree discernible in the low-motion subject but, in particular for subjects with more motion, the shell-average raw signal contains stripe patterns from motion-related artefacts.

- requires no ground truth data for training
- produces directly human-inspectable correction fields, in the same space as the dMRI data
- uses explicit constraints that locally preserve in-plane image contrast

The source code will be made publicly available at <https://github.com/maxpietsch/dStripe>.

## 2. Methods

### 2.1. Data

The dMRI data were acquired as part of the developing Human Connectome Project. Each dataset consists of 300 volumes at  $1.5 \times 1.5 \times 3$  mm voxel-resolution, acquired in 64 slices with 1.5 mm slice-overlap; interleave 3, shift 2; grid:  $99 \times 99 \times 64$  voxels; multiband factor 4; TR/TE=3800/90 ms; diffusion weightings  $b = 0, 400, 1000,$  and  $2600$  s/mm<sup>2</sup> with 20, 64, 88, and 128 directions, respectively (Hutter *et al.*, 2018; Hughes *et al.*, 2017; Tournier *et al.*, 2020).

The k-space data were reconstructed to 1.5 mm<sup>3</sup> isotropic voxel size and denoised in the complex domain (Veraart *et al.*, 2016). Field maps and brain masks were estimated with FSL `topup` (Andersson *et al.*, 2003) and `bet` (Smith, 2002).

For visualisation of single-subject data, the data ‘sub-CC00083XX10/ses-30900’ was used as it has clearly visible stripe patterns. This baby’s postmenstrual age (PMA) at scan

is 42 weeks and the image lies in the 59th percentile quality score estimated from the motion and outlier weights (Christiaens *et al.*, 2020). To create a population-average template from representative data, we randomly selected 32 subjects from 38 to 42 weeks PMA that have quality assessment scores above the 20th percentile (see figure A.10). Summary data contains the single subject and population template data and additional 11 subjects ranging from 34.3 to 43.2 weeks PMA and quality scores ranging from 3rd to 100th percentile. For network training, we randomly selected 20 scans from the dHCP, none of which are part of the data used for evaluation. The PMA of subjects in the training dataset spans 31 to 43 weeks.

To test generalisation performance to a different cohort and sequence, we use two scans of babies acquired at 29 and at 38 weeks PMA using a Philips Achieva 3T product sequence which exhibits severe slice intensity artefacts due to inhomogeneous slice-timing. The data were acquired with one  $b = 0$  volume and 32 volumes with  $b = 750$  s/mm<sup>2</sup>, TR/TE = 7856/49 ms, voxel size: 1.75 mm inplane, 2 mm through-plane, grid:  $128 \times 128 \times 49$  voxels, using a single phase-encode direction and a regular single-band slice interleaving (interleave 2, shift=1).

### 2.2. Slice-to-volume motion correction and reconstruction framework

The dMRI data were processed using a motion correction and reconstruction algorithm (MCR) with integrated slice-to-volume reconstruction (SVR) (Christiaens *et al.*, 2020). In brief, the reconstruction is based on an iterative estimation of

a data-driven multi-shell low-rank (model-free) data representation (SHARD), slice outlier estimation, and rigid registration algorithm (Christiaens *et al.*, 2019). The reconstruction utilises information from overlapping slices and the native slice profiles for super-resolution deconvolution and is formulated as an inverse problem that iteratively estimates the reconstruction coefficients  $\mathbf{C}$ , defined in the motion-corrected “anatomical” space (the moving subject-aligned reference frame), and the rigid motion parameters  $\boldsymbol{\mu}$  that map between “source” space (the scattered slices in scanner coordinates) and anatomical space by minimising the difference between the acquired signal of a slice in the source space  $\mathbf{R}_s$  and its signal prediction

$$\min_{\mathbf{C}, \boldsymbol{\mu}} \sum_s w_s \|\mathbf{R}_s - \mathbf{B}_s \mathbf{I}(\boldsymbol{\mu}_s) \mathbf{Q}_s(\boldsymbol{\mu}_s) \mathbf{C}\|_2^2 + r(\mathbf{C}).$$

The model consists of the q-space (SHARD) basis  $\mathbf{Q}_s(\boldsymbol{\mu}_s)$ , the linear motion and interpolation operator  $\mathbf{I}(\boldsymbol{\mu}_s)$ , and the blurring and slice selection matrix  $\mathbf{B}_s$  that also incorporates the slice sensitivity profile. Slice weights  $w_s$  are used to reduce the effect of outliers, and a regularisation term  $r(\mathbf{C})$  is used to stabilise the inverse problem. EPI distortions are corrected by unwarping the input dMRI data before each reconstruction step using a field map and the subject motion parameters.

To isolate the effects of dStripe from those of the motion correction, motion, slice-weight and SHARD basis parameters were kept constant after their estimation on the original data for subsequent reconstruction when using source-space destriping (see below).

### 2.3. dStripe approaches: source or anatomical space

MCR provides a mapping between the “source” space of the scanner and the “anatomical” space of the moving subject and separates dropout and other artefacts from anatomical features. The stripe artefact can be affected by the scanner geometry and the subject tissue properties and motion. This raises a question about the best space to operate in.

#### 2.3.1. Anatomical-space destriping

Destriping can be applied in anatomical space to remove residual stripe patterns in the MCR output. This can be performed with any motion correction and reconstruction framework as it operates solely on its output and produces correction fields directly in the space of interest. However, this approach cannot be used to interactively refine the MCR, and will only work well if the destripe mechanism can cope with stripe patterns that are themselves scattered due to motion.

#### 2.3.2. Source-space destriping

Taking advantage of the mapping MCR provides, we can use information from the motion-corrected anatomical space to correct the raw data in source space. For each excitation (single slice or multiband shot), we generate the corresponding *corrected* signal prediction ( $\mathbf{B}_s \mathbf{I}(\boldsymbol{\mu}_s) \mathbf{Q}_s(\boldsymbol{\mu}_s) \mathbf{C}$ ). This allows modulating the corresponding slice in its “native” orientation, potentially increasing the effectiveness of the dStripe algorithm.

This approach relies on the signal representation in anatomical space capturing the contrast of interest together with the

stripe patterns. Similarly to anatomical-space destriping, this approach does not necessarily capture the full source-space stripe pattern as intensity modulations in target space can be attenuated due to interpolation or if they are discarded by the outlier removal or rank-reduced model fit. Moreover, destriping each excitation in “slice-native” space requires destriping a slice-native volume with corresponding motion parameters; for our data, this increases the computational cost by a factor of 16 compared to anatomical-space destriping (64 slices, multiband 4). Finally, to ensure convergence of the output, we need to destripe and perform the subsequent MCR again multiple times (3 times is sufficient in our data), further increasing the computational cost of the source-space destriping approach.

Note that because the dHCP employed overlapping slices with a super resolution reconstruction, stripe patterns are amplified by the slice profile deconvolution in the reconstruction. When projecting the signal to source-space, we therefore modify the slice selection matrix  $\mathbf{B}_s$  to preserve the super-resolved contrast by avoiding slice profile blurring. Instead, we subsequently downscale the dStripe-estimated slice modulation fields to account for the difference in slice thickness.

### 2.4. Metric of stripiness

dMRI images, in particular highly diffusion weighted volumes, have significant anatomical contrast. While stripe patterns can be detected visually with the trained eye, it is not trivial to mathematically separate expected from artefactual through-plane intensity variations.

Inter-slice signal variation can be measured for instance using the standard deviation (SD) of the dMRI signal (or its spherical harmonic representation) over a kernel of neighbouring voxels along the through-plane direction. We expect signal-preserving destriping to reduce the through-plane SD while preserving the equivalent in-plane SD. However, directly optimising a SD-derived measure is not viable because the expected anatomical through-plane SD dominates the artefactual SD by one to two orders of magnitude.

Therefore, we decided to use a neural network trained to remove simulated stripes artificially introduced into the training data. To assess its performance on the original data, we opt to visually rate images using scalar or rotation invariant representations of the diffusion weighted data, as well as derived images, as outlined below.

The angular information of the signal can be expressed in the basis of real, symmetric spherical harmonics (SH) (matrix.readthedocs.io, 2020), which allows to investigate its angular frequency spectrum (Kazhdan *et al.*, 2003). We use the  $l_2$ -norm of the coefficients in a particular harmonic band  $\ell = [0, 2, 4, \dots]$  defined as  $\sqrt{\sum_{m=-\ell}^{\ell} (c_{\ell}^m)^2}$ , where  $c_{\ell}^m$  are the SH coefficients of the dMRI signal in the given voxel and shell. These angular measures, denoted as  $\|\mathbf{S}\|_{\ell}$ , are proportional to the square root of the power spectral density corresponding to that frequency band.

Finally, we investigate data consistency across shells and in the angular domain using (i) a diffusion tensor representation (Basser *et al.*, 1994) and (ii) the brain tissue and free water

maps obtained using multi-shell multi-tissue constrained spherical deconvolution (MT CSD) fits (Jeurissen *et al.*, 2014) (using population-averaged basis functions measured in WM and in CSF voxels (Dhollander *et al.*, 2019; Pietsch *et al.*, 2019a)). Signal changes due to dStripe are assessed quantitatively and visually using the fit residuals and fit-derived measures.

## 2.5. Destriping method

### 2.5.1. The CNN architecture

#	layer	tensor shape	param.
	input volume $\mathbf{S}$	$1, n_x, n_y, n_z$	
1	<i>SeparableConv</i> (3,3,3)	$16, n_x, n_y, n_z$	60
2	<i>BatchNorm</i>		32
3	<i>ConvBlock</i> (3,3,7)		2,128
4	<i>ConvBlock</i> (3,3,7) & <i>ReLU</i>		2,128
5	<i>ConcatPool</i> (2,2,1)	$32, \frac{n_x}{2}, \frac{n_y}{2}, n_z$	
6	<i>BatchNorm</i>		64
7	<i>ConvBlock</i> (3,3,7)		5,792
8	<i>ConvBlock</i> (3,3,7) & <i>ReLU</i>		5,792
9	<i>ConcatPool</i> (.,.,1)	$64, 16, 16, n_z$	
10	<i>SeparableConv</i>	$32, 16, 16, n_z$	2,080
11	<i>SeparableConv</i> & <i>ReLU</i>	$1, 16, 16, n_z$	33
12	<i>DRC</i> : $\mathbf{T} \mapsto \frac{2.0}{1+\exp(-\mathbf{T})} + 10^{-4}$		
13	log transform		
14	x,y lowpass filter (9,9,1)		(81)
15	z highpass filter (1,1,9)		(9)
16	exp transform		
17	x,y upsample	$1, n_x, n_y, n_z$	

**Table 1.** The dStripe network architecture as used during training. Tensor shape denotes number of channels, and inplane extent. Layers with operations aggregating information across multiple spatial locations are highlighted with their (maximum) spatial extent in brackets, those without are pointwise operations or  $1 \times 1 \times 1$  convolutions. For clarity, tensor shapes are omitted if unchanged. Parameter counts for fixed layers are denoted in brackets. Layers and dynamic range constraint (*DRC*) and frequency filters are described in sections 2.5.1 and 2.5.2. For inference, layer #15 is deactivated and its function replaced by FFT-based high-pass-filtering of the network output (see section 2.5.6).

Input to the network is a single 3D volume  $\mathbf{S}$  of dimensions  $n_x, n_y, n_z$  (one channel). The output of the network is the multiplicative field defined in the space of  $\mathbf{S}$  and when element-wise multiplied with the input image ( $\mathbf{S} \odot \mathbf{F}$ ), ideally reverses the stripe-producing mechanism.

To allow potentially taking large parts of the spatial context into account, the network operates on the full field of view of  $\mathbf{S}$  ( $99 \times 99 \times 64$  voxels). To accommodate this data in GPU memory and for performance reasons the network was designed to be relatively small (18,109 trainable parameters). To stabilise training and minimise domain shift, the network is trained with intensity and spatial frequency filter constraints as part of the architecture. For faster training, we use batch normalisation (*BatchNorm*) (Ioffe and Szegedy, 2015; Santurkar *et al.*, 2018). An overview of the network architecture, implemented in pytorch (Paszke *et al.*, 2019), is given in table 1.

For approximate invariance to the image dimensions, all layers use voxel-wise or convolutional operations. Spatially, we aim to represent slice-specific information at native resolution through-plane while limiting the in-plane resolution to yield a

smooth field after upsampling. Hence, the resolution in the slice dimension is preserved throughout the network but is reduced in-plane via pooling layers (#5,#9) to half its original extent, and then adaptively to  $16 \times 16$  voxels. In all convolution layers, the image extent is preserved via zero padding.

Throughout the network, depthwise separable convolutions (Chollet, 2017) are used to limit the number of parameters while allowing the network to learn global relations between feature maps and their respective spatial extents (Hu *et al.*, 2018). The module denoted as *SeparableConv* consist of a set of input-channel-specific  $3 \times 3 \times 3$  convolution filters whose output channels are linearly combined using pointwise  $1 \times 1 \times 1$  convolutions (if the number of output channels exceeds 1).

*ConvBlock* consist of 3 concatenated depthwise  $3 \times 3 \times 3$  convolutions with dilation factors (Yu and Koltun, 2015) of 1, 2, and 3 in the z direction, respectively, followed by a  $1 \times 1 \times 1$  convolution and optionally a *ReLU* activation ( $\mathbf{T} \mapsto \max(0, \mathbf{T})$ ). *ConcatPool* layers concatenate spatial average and maximum pooling layers.

The inplane lowpass filter is implemented as a fixed 2D Gaussian blur filter and the through-plane high-pass filter subtracts the low-frequency filtered tensor from its input using a fixed 1D Gaussian filter ( $\mathbf{T} \mapsto \mathbf{T} - g(\mathbf{T})$ ).

### 2.5.2. Modulation field constraints

To restrict the intensity modulation range, the output of the last convolution and *ReLU* layer (#11) is mapped to the range  $[1.0001, 2.0001]$  via the dynamic range constraint (*DRC*, layer #12)  $\mathbf{T} \mapsto \frac{2.0}{1+\exp(-\mathbf{T})} + 10^{-4}$ . This scaling also facilitates spatial frequency filtering of the multiplicative destripe field in the log-domain (#13) using numerically stable additive operations (#14,15). The in-plane and subsequent through-plane filters are implemented as immutable convolution filters. To suppress high frequency information in-plane, a 2D Gaussian blur filter ( $\sigma = 1.5$  voxels, kernel size  $9 \times 9$ ) is used. The through-plane high-pass filter is similarly implemented by subtracting a 1D Gaussian-filtered version of the image ( $\sigma = 1$  voxels), which allows factoring out low-frequency background modulations (and global offset) in the through-plane direction.

Finally, exponential scaling (#16) and in-plane upsampling (#17) yield the multiplicative field at the target resolution while preserving in-plane smoothness and through-plane resolution.

### 2.5.3. Training and augmentation

Training is performed on 20 dMRI datasets in 500 epochs with a batch size of one, with the Adam optimiser (Kingma and Ba, 2014) and triangular learning rate scheduler  $[1, 5] \times 10^{-4}$  (80 training iterations per half-cycle) (Smith, 2017). From each dMRI dataset, 10 randomly selected volumes are selected for each *b*-value and all images divided by its 99th percentile intensity.

Data augmentation was performed by random reorientation via dihedral transformations (90 degree rotations and axis-aligned reflections) such that all three axes (AP, LR, IS) are used as the slice direction with equal probability. Rotated versions of the data can be assumed stripe-free, and are included in the augmentation to minimise the influence of any stripes that might have been present prior to image augmentation.

Slice-wise intensity stripe patterns were simulated to allow random multiplicative factors per slice that are correlated in time according to the slice interleave pattern. For all slices  $i$  of a temporally continuous excitation sequence block  $\tau \in [0, 1, 2]$  ( $i \bmod 3 = \tau$ ), multiplicative slice-wise intensity modulations  $m_i$  were drawn from the positive-clipped squared normal distribution  $\max\left(\left(\mathcal{N}(\mu_\tau, \sigma_\tau^2)\right)^2, 10^{-10}\right)$ , with block-specific variance  $\sigma_\tau^2 = 0.05$  and random but fixed centre drawn from a uniform distribution  $\mu_\tau \sim \mathcal{U}(0.9, 1.1)$ . This was repeated for all 3 excitation blocks. Finally the modulation vector  $\mathbf{m}$  was normalised to unit geometric mean to approximately preserve global scaling effects. We denote  $\mathcal{M}$ , the slice-wise intensity modulation distribution that is defined as the observable consequence of this slice artefact generating process.

#### 2.5.4. Stein's Unbiased Risk Estimator-based image recovery

Consider the image reconstruction problem to infer an unknown image  $\mathbf{u} \in \mathbb{R}^n$  from corrupted measurements  $\mathbf{v} = \mathbf{u} + \delta$ , with corruptions  $\delta$ . This problem is commonly solved by using prior knowledge about the properties of the image. Convolutional neural networks are powerful models to encode image properties but typically require large amounts of paired training data.

In this work, we have only one corrupted measurement and no ground truth data. Therefore, instead of supervised learning with paired data, we use a training technique that is based on Stein's Unbiased Risk Estimator (SURE) (Stein, 1981), that allows reconstructing an unobserved image  $\mathbf{u}$  from individual additive Gaussian noise corrupted measurements as for instance demonstrated in (Soltanayev and Chun, 2018; Metzler et al., 2018). SURE is a measure of the expected generalisation loss ("risk") of an estimator of the mean of a data-generating process  $\mathbf{v} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbb{1}\sigma^2)$  with unknown mean  $\boldsymbol{\mu}$ . Assuming the measurements are corrupted by standard additive homoscedastic noise and that the estimator  $f_\Theta$  with parameters  $\Theta$  is weakly differentiable, we can express the expected reconstruction error as (Stein, 1981)

$$\mathbb{E} \left[ \frac{1}{n} \|\mathbf{u} - f_\Theta(\mathbf{v})\|_2^2 \right] = \mathbb{E} \left[ \frac{1}{n} \|\mathbf{v} - f_\Theta(\mathbf{v})\|_2^2 \right] - \sigma^2 + \frac{2\sigma^2}{n} \nabla_{\mathbf{v}} \cdot f_\Theta(\mathbf{v}), \quad (1)$$

where the expectation term measures the bias, and the divergence term  $\nabla_{\mathbf{v}} \cdot f_\Theta(\mathbf{v}) = \sum_{i=1}^n \frac{\partial f_{\Theta,i}(\mathbf{v})}{\partial v_i}$  expresses the response of the model to input perturbations (model variance). Remarkably, SURE does not require access to  $\boldsymbol{\mu}$  as  $\sigma$  can be estimated from the data (Efron, 2004) and  $\|\mathbf{v} - f_\Theta(\mathbf{v})\|_2^2$  is directly accessible for a given model (Stein, 1981).  $\nabla_{\mathbf{v}} \cdot f_\Theta(\mathbf{v})$  can be hard or impossible to derive analytically for complex  $f_\Theta$ . However, for bounded functions  $f_\Theta$  with intractable derivatives or prohibitively high-dimensional parameter spaces, such as deep convolutional neural networks, Monte-Carlo techniques can be used to estimate the divergence

$$\nabla_{\mathbf{v}} \cdot f_\Theta(\mathbf{v}) = \lim_{\epsilon \rightarrow 0} \mathbb{E} \left[ \mathbf{p}^T \frac{f_\Theta(\mathbf{v} + \epsilon \mathbf{p}) - f_\Theta(\mathbf{v})}{\epsilon} \right],$$

with the normally distributed noise vector  $\mathbf{p} \sim \mathcal{N}(\mathbf{0}, \mathbb{1})$ . As shown by (Ramani et al., 2008), due to the high dimensionality of the data, this can be approximated with a single or few samples and a small non-zero  $\epsilon$  via

$$\nabla_{\mathbf{v}} \cdot f_\Theta(\mathbf{v}) \approx \mathbf{p}^T \frac{f_\Theta(\mathbf{v} + \epsilon \mathbf{p}) - f_\Theta(\mathbf{v})}{\epsilon}. \quad (2)$$

#### 2.5.5. Loss functions

The training loss is evaluated inside the brain mask  $\mathbf{M}$ , which contains  $n_{\mathbf{M}}$  true elements, and depends on the slice direction after the dihedral transformation  $d$ . For simplicity of notation, let  $\mathbf{S} \odot \mathbf{m}$  denote the broadcast-multiplication of each slice  $\mathbf{S}_{:,s}$  with the modulation  $m_s$  and  $d(\mathbf{S}) \odot \mathbf{m}$  the transformed and modulated image. Along the AP and LR axes, the training loss consists of the sum of  $J_{\text{aug}}$ , the mean squared loss between original signal  $\mathbf{S}_o$  and augmented  $\mathbf{S}_{d,m} = d(\mathbf{S}) \odot \mathbf{m}$  corrected by the network  $f_\Theta$

$$J_{\text{aug}} = \frac{1}{n_{\mathbf{M}}} \|d(\mathbf{S}_o) - f_\Theta(\mathbf{S}_{d,m})\|_{2,\mathbf{M}}^2$$

and  $J_{\text{const}}$ , the mean squared loss that penalises altering the original image (free of stripe artefacts in that direction):

$$J_{\text{const}} = \frac{1}{n_{\mathbf{M}}} \|d(\mathbf{S}_o) - f_\Theta(d(\mathbf{S}_o))\|_{2,\mathbf{M}}^2.$$

For slice directions along IS, we construct the MC SURE-based loss using an augmented image as input. Following equations 1 and 2 and limiting the divergence to the non-negative domain to bound the loss, we use

$$J_{\text{SURE}} = \frac{1}{n_{\mathbf{M}}} \|\mathbf{S}_{d,m} - f_\Theta(\mathbf{S}_{d,m})\|_{2,\mathbf{M}}^2 - \sigma_{\text{SURE}}^2 + 2 \frac{\sigma_{\text{SURE}}^2}{n_{\mathbf{M}}} |\rho_{\text{SURE}}|$$

with  $\sigma_{\text{SURE}} = \text{SD}(\mathbf{S}_{d,m} - d(\mathbf{S}_o))_{\mathbf{M}}$  the standard deviation of the signal change due to image augmentation inside the brain mask, and  $\rho_{\text{SURE}}$  the model variance penalty to input perturbations which is estimated using

$$\rho_{\text{SURE}} = \left\langle \mathbf{P} \odot \frac{f_\Theta(\mathbf{S}_{d,m} + \epsilon \mathbf{P}) - f_\Theta(\mathbf{S}_{d,m})}{\epsilon}, \mathbf{M} \right\rangle_F,$$

with  $\epsilon = 10^{-3}$  and the calibrated perturbation image  $\mathbf{P} = (\mathbf{S}_{d,s} \odot \hat{\mathbf{m}} - \mathbf{S}_{d,s}) / \text{SD}(\mathbf{S}_{d,m} \odot \hat{\mathbf{m}} - \mathbf{S}_{d,m})_{\mathbf{M}}$  generated using an additional simulated slice modulation vector  $\hat{\mathbf{m}}$  drawn from  $\mathcal{M}$ , and  $\langle \cdot, \cdot \rangle_F$  the Frobenius inner product which denotes an element-wise multiplication and summation over all elements.

#### 2.5.6. Modifications for inference: attention and iterative field updates

Training is performed as described, however during inference we can improve performance and avoid issues we have observed. First, destripe performance can be improved by incrementally estimating the destriped image by applying the neural network to the previous best estimate for up to three iterations after which the resulting images are approximately converged.

However, naïve multiplication of field updates would increasingly relax the field's frequency constraints. Second, in-plane, the field estimates require additional low-pass filtering due to upsampling artefacts of the bilinear interpolation layer (#17). Third, the field estimate can be unreliable in areas dominated by noise or ill-defined due to the lack of signal of interest and close to the edge of the field of view. This can, through the field frequency constraints, negatively impact the field in adjacent brain areas.

The approach to address these issues consists of the following modifications to the final stages of the network, applied during inference only:

- 1) We **remove the built-in high-pass filter** (layer #15) and replace its function with a FFT-based high-pass filter ( $h_z$ ) that is applied to the (upsampled) field.
- 2) To **prevent frequency spectrum drift** over iterations  $i$ , we do not apply the through-plane filter  $h_z$  (described above) to the output of the network, but instead to the composition of the current best estimate of the field  $\mathbf{F}_i$  and its update:

$$\mathbf{F}_{i+1} \leftarrow h_z(\mathbf{F}_i \odot f_{\Theta}(\mathbf{S}_o \odot \mathbf{F}_i))$$

For this filter, we use a parametric FFT-based frequency filter which allows high-quality frequency constraints that can be easily adapted retrospectively. We use a 4th order Butterworth filter ( $\mathbf{b}_1$ : normalised frequency cutoff: 21/32, padding: 17) to block low-frequency components.

3) To **address edge effects** we implement an attention mechanism based on the brain mask that attenuates high-frequency contributions to the field from outside trusted areas. This modulation has to be applied before high-pass frequency filtering to reduce leakage of the field into areas of interest. Hence, the attention mechanism is incorporated into the through-plane frequency filter  $h_z$ . It uses a through-plane frequency separation into low- and mid- to high-frequency components and a spatial attention map  $\mathbf{A}$  with voxel-values in  $[0, 1]$  to smoothly blend these frequency bands in the image domain.

Specifically,  $h_z(\mathbf{F})$  is defined as

$$\begin{aligned} \text{step 1: } \mathbf{F} &\leftarrow \exp(\log(\mathbf{F}/\mathbf{F}_{\text{lp,3D}}) \odot \mathbf{A}) \odot \mathbf{F}_{\text{lp,3D}} \\ \text{step 2: } \mathbf{F} &\leftarrow \mathbf{F}/\text{fft}_z^{-1}(\text{fft}_z(\mathbf{F}) \odot \mathbf{b}_1), \end{aligned}$$

where  $\mathbf{F}_{\text{lp,3D}}$  is a lowpass-filtered version of the field (3D Gaussian-blur,  $\sigma = 5 \times 5 \times 11$  voxels). In step 1, the field's medium- to high-frequency information ( $\mathbf{F}/\mathbf{F}_{\text{lp,3D}}$ ) is exponentiated voxel-wise with  $\mathbf{A}$ , which down-weights this frequency band in untrusted areas (where  $\mathbf{A} < 1$ ), while leaving low-frequency components unaltered. In step 2, this locally filtered field is subsequently high-pass filtered through-plane with the Butterworth filter  $\mathbf{b}_1$ . The key is that the Gaussian filter allows estimating the low-frequency background without causing distortion to the high-frequency components of interest and  $\mathbf{A}$  is spatially smooth, allowing a weighted blending of low- with medium- to high-frequency components. In areas with low trust, the low-frequency field dominates in step 1, hence high-frequencies from these areas contribute little to the filtering in step 2.

We use a brain mask  $\mathbf{M}$  to create the attention map  $\mathbf{A}$ . In plane, to gradually reduce contributions from outside the brain

mask while preserving attention inside the brain,  $\mathbf{M}$  is dilated by the full width at one-tenth maximum of a Gaussian filter and subsequently blurred (13 dilations,  $\sigma = 9 \times 9$  voxels). Through-plane, this image is smoothed further ( $\sigma = 3$  voxels), down-weighting contributions of the most inferior and superior parts of the masked area.

- 4) As a final step, the field is low-pass filtered in-plane with a cutoff frequency chosen to suppress **in-plane upsampling artefacts** (3rd order Butterworth filter, normalised frequency cutoff: 2/32, padding: 24).

## 2.6. Degrees of freedom

As outlined in the introduction (1), in this work we assumed that the field needs to be modelled as a function of position and  $b$ -value (see figure 1), and direction of the diffusion weighting. We can control the effective spatial degree of freedom of dStripe by modifying the minimum inplane grid size of the network architecture (layer #9), by modifying the frequency filter parameters, or by applying filters to the final upsampled field.

In anatomical-space destriping, angular degrees of freedom of the final upsampled full rank field ( $\mathbf{F}_{b,\ell,m,x,y,z}$ ) can be reduced for instance via a projection to a spherical harmonics representation with reduced  $\ell_{\text{max}}$  or by shell-averaging, or it can be derived from shell-averaged dMRI signal projections.

For the purpose of investigating the effect of a reduction of the spatial and angular degrees of freedom of the field (see section 3.1), we calculate the geometric mean of the field inside the brain mask across the angular domain (denoted  $\mathbf{F}_{b,x,y,z}$ ) and also averaged within slices ( $\mathbf{F}_{b,z}$ ) and apply these rank-reduced fields to the dMRI data via broadcasting operations.

## 3. Results and Discussion

### 3.1. Degrees of freedom of slice modulation field

Figure 2 shows that dStripe needs to operate in the spatial, angular and  $b$ -value domains: restricting the field  $\mathbf{F}_{b,\ell,m,x,y,z}$  that has the same rank as the data, in the angular ( $\mathbf{F}_{b,x,y,z}$ ) and spatial domain ( $\mathbf{F}_{b,x,y,z}$ ,  $\mathbf{F}_{b,z}$ ) leads to increased residual stripes, justifying the design decisions of the method.

In figure 2, the  $l_2$ -norm of the resulting data as well as the difference to the data pre-dStripe are shown for representative  $b$  and  $\ell$  value combinations for a representative subject. In the shell-average images, shell-specific but slice-wise constant multiplicative scale factors (figure 2b) reduce the stripe pattern but a smoothly varying field (figure 2c) reduces the stripe pattern further. However, stripe patterns in the angular domain remain nearly unchanged. The full dStripe field attenuates stripe patterns in the angular domain and reduces the power in higher angular frequency components. By design, an inplane reduction in high-frequency SH power can be attributed to smoothly varying intensity modulations of individual dMRI volumes and is therefore likely caused by a reduction of angular variance due to the removal of stripe artefacts.

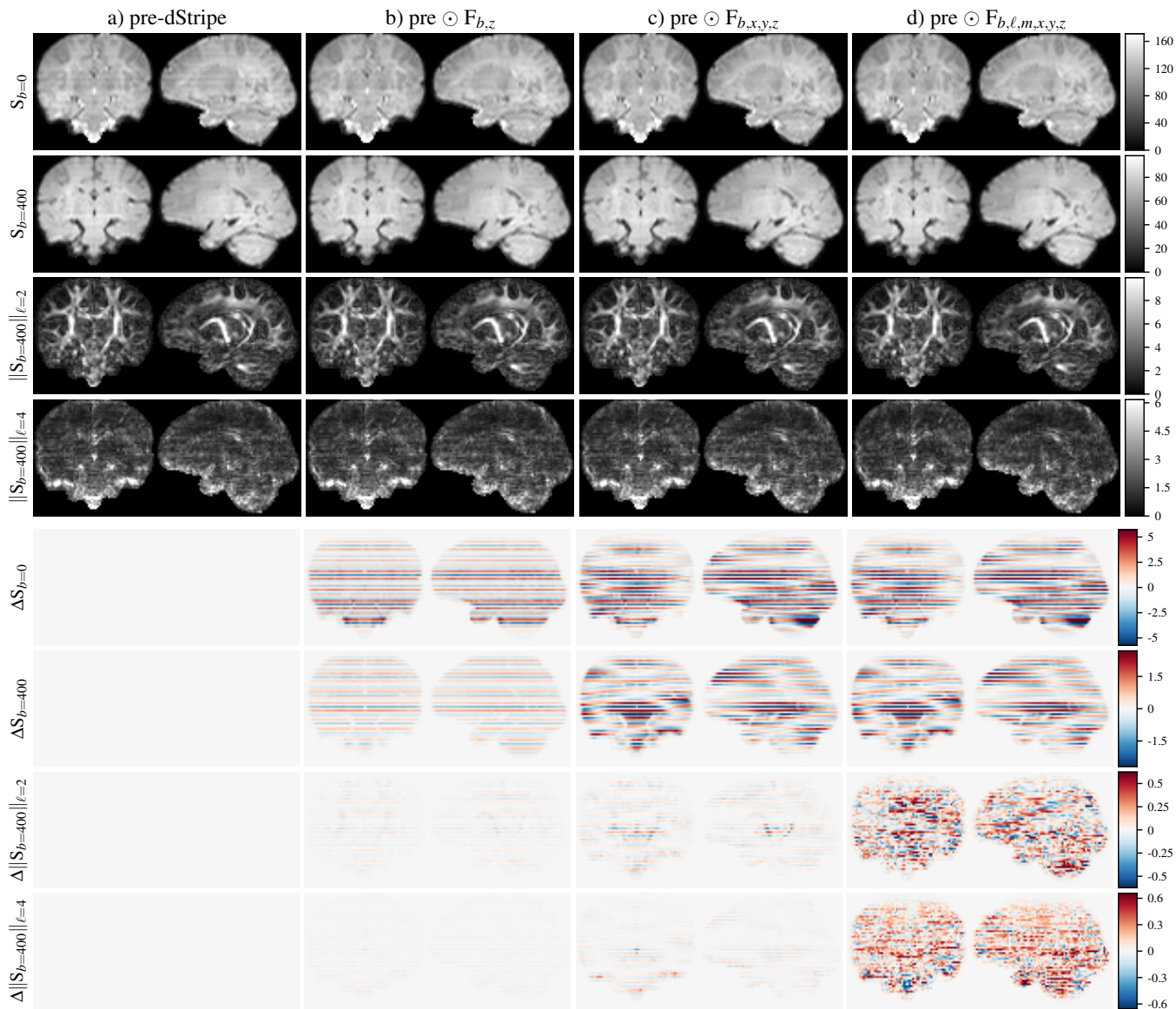


Fig. 2. Exemplary components of the  $l_2$ -norm spectrum with increasing degrees of freedom of the correction field (top) and the difference in the  $l_2$ -norm spectrum due to dStripe (pre-dStripe - post-dStripe) shown below. a): pre-dStripe: original data after motion correction, b) 1D b-value specific correction applied, c) 3D scalar b-value specific field applied, d) full correction without rank constraints. b-values in units of  $s/mm^2$ .

### 3.2. Space-dependency and subject motion

If the artefact is tied to scanner space, it is desirable to destripe in source space; similarly if it is fixed to the subject tissue, dStripe should be performed in anatomical space. Given the observation that the field has consistent features in source space of a cohort with relatively prevalent motion, it is reasonable to at least partially model the field in source space. However, it is unclear if it is best to correct stripe patterns in scattered source or in the motion-corrected anatomical space as this depends not just on the nature of the artefact but also on how well it can be corrected given our framework, where we rely on a motion-corrected model of the data.

The properties of the estimated slice modulation artefact depend on the space in which it is estimated. In particular, the super-resolution reconstruction increases inter-slice intensity variations compared to the patterns observed in the source

data acquired with overlapping slices. The presence of subject motion also affects the spatial frequency content of the artefact: assuming a smooth field in source space, subject rotation can cause higher frequency patterns in motion-corrected space. We rely on a motion-corrected representation of the data, where artefacts originating from source space might be smeared out or might not be fixable given the frequency constraints of the field. Similarly, when projected to source space, stripes observed in the motion corrected signal could be distributed over multiple scattered (oblique) slices.

A possible source of destripe field inconsistencies for source-space methods is a potential field of view or pose-dependency of the method. To estimate the slice-specific field, dStripe requires spatially contiguous adjacent slices, in our case a complete “slice-native” volume for context. For source-space destriping, intra-volume motion can affect the pose of the slice-



native volume for each slice of interest, introducing a possible source of variance into the destripe field estimation, which can potentially be amplified or introduce local artefacts in the subsequently required motion correction, especially if super-resolution algorithms are used. In anatomical-space destriping, the pose is fixed eliminating this source of inter-slice and intra-volume variance.

Finally, the choice of the space in which dStripe is performed has implications for the nature of the change to the motion corrected data. A smooth modulation applied to scattered source space data can improve subsequent (super-resolution) motion correction but also introduces local changes in its output due to the interpolation and aggregation of data from multiple slices; dStripe performed in anatomical space guarantees smoothly varying fields in the space where the data is used for analysis.

In the following sections, we attempt to remove stripes in both spaces independently and compare the data in anatomical space. This does not prove the origin of the artefact but assesses the stripe artefact removal potential for each space, given the framework.

Figure 3 shows the shell-averaged data in motion-corrected space without dStripe, with dStripe in anatomical space and in source space. The shell-average signal changes most in the lower  $b$ -values, source-space destriping changes the signal slightly more, but the difference between source- and anatomical-space destriping are much smaller than between data pre- and post-dStripe. The difference between anatomical- and source-space dStripe at the top of the brain is related to the implementation of the attention filter that can be gradually circumvented with further dStripe iterations. In the presence of motion, potential field of view edge effects of the network can appear in more central slices, in particular if motion caused the brain to be cropped during the acquisition.

Figure 4 shows raw  $b = 0$  data and corresponding fields (estimated in source space) for a single volume and averaged across the shell. By design, the fields are smooth in-plane and relatively sharp through-plane. In source space, using source-space destriping, the average field shares similarity with that in individual volumes but is attenuated overall. This attenuation can originate from a non-stationary of the modulation artefact, for instance caused by blurring due to subject motion if the field is tissue dependent, or assuming a fixed field in source-space, it could indicate variance in the estimated dStripe field due to varying subject pose.

### 3.3. Performance

To evaluate performance we use the following objectives: post-dStripe, dMRI data has to appear visually less stripy and have reduced variance in the through-plane direction. Furthermore, we use DTI and MT CSD fits to evaluate the consistency of the data across shells and in the angular domain.

#### 3.3.1. spatial dMRI signal variance

The local standard deviation of the (motion corrected) dMRI signal evaluated in 1D patches of 7 voxels inside the brain mask aligned in-plane ( $x,y$ ) and through-plane ( $z$ ) of each spherical harmonic volume are shown in figure 5. The analysis is split by

$b$ -value and harmonic band. Note that in-plane and through-plane values of (anatomical) variance are not equal (see left column) and anatomical and inter-subject variance dominate changes due to dStripe. Performing a subject-specific comparison, dStripe reduces through-plane variance while approximately preserving in-plane signal variance. Anatomical-space destriping tends to reduce in-plane variance more, source-space destriping reduces through-plane variance in the outer shell more which is partially driven by the difference in attention filters but also due to the amplitude of changes (see figure 3).

#### 3.3.2. Diffusion signal representation fits

Shell-average fit residuals and fit-derived maps are shown for a subject in figure 6. Stripy residual maps for both data representations indicates that the fit residuals are sensitive to stripe patterns; MD and tissue and fluid volume fraction maps particularly show clear patterns of stripe artefacts prior to dStripe. After (anatomical-space) dStripe, both residual maps and MD and volume fraction maps show clearly reduced stripe patterns while preserving anatomical contrast.

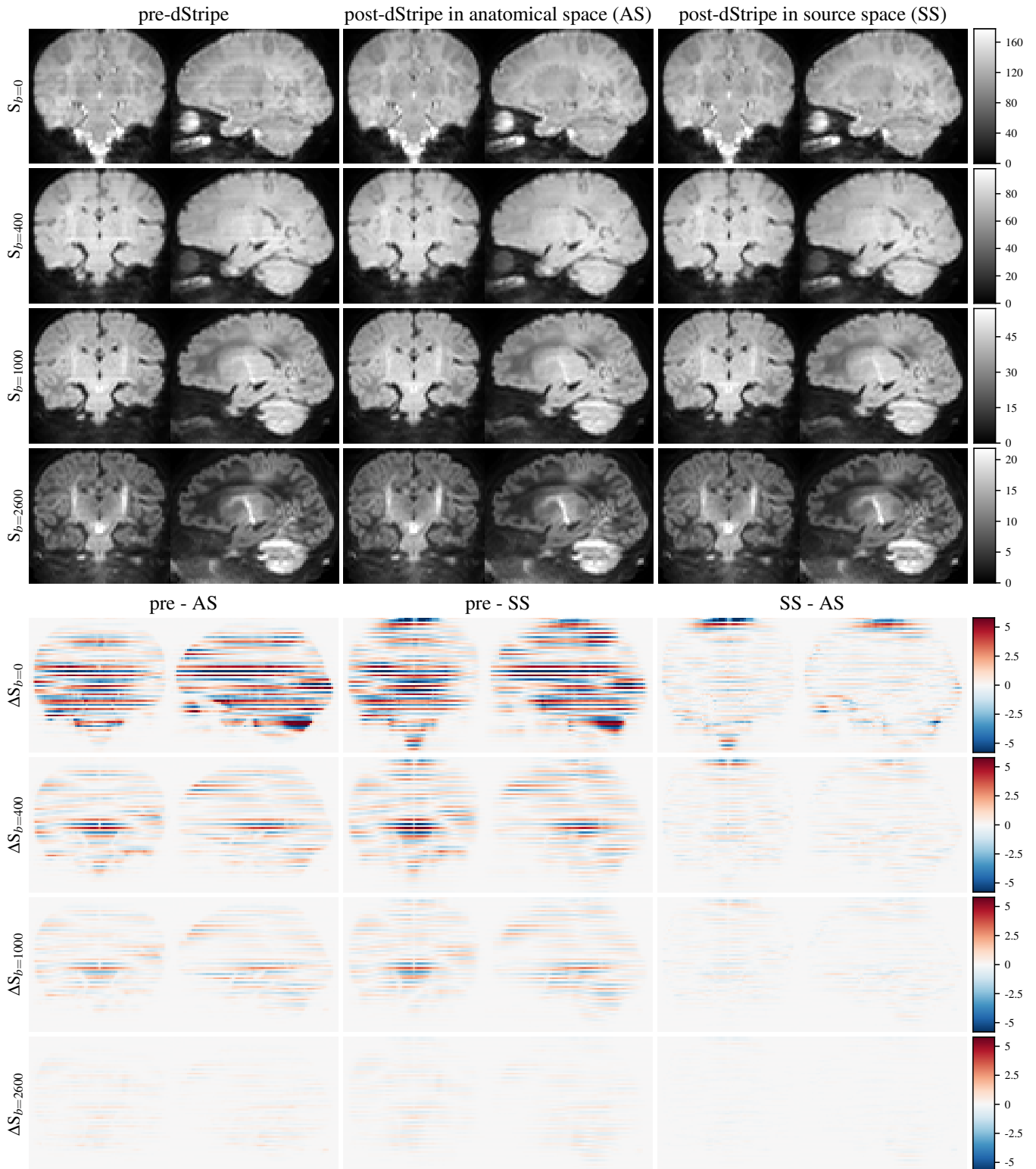
Using diffusion tensor and MT CSD root mean squared fit residuals as a proxy for data consistency across shells and directions, figure 7 shows that dStripe increases data consistency across all shells for both signal representations. Fit residuals are reduced mostly in the  $b = 400s/mm^2$  and  $b = 1000s/mm^2$  shells. Source-space dStripe compared to anatomical-space dStripe tends to yield slightly lower residuals with the exception of the  $b = 0s/mm^2$  and  $b = 2600s/mm^2$  shell for the DT fit. For both fits and spaces a larger relative reduction of residuals can be observed for data with higher residuals in the  $b = 400s/mm^2$  and  $b = 1000s/mm^2$  shells; dStripe improves residuals the most for high-residual data, which tend to coincide with more subject motion, and source-space dStripe slightly outperforms anatomical-space dStripe in this regime.

#### 3.4. Population-level effects and anatomical bias

As demonstrated, dStripe reduces stripe patterns in the dMRI data and overall increases its consistency within subjects. Here, we use data from 32 age-matched subjects to assess the effect of dStripe on data consistency across subjects and whether the dStripe method introduces systematic changes (bias) across subjects related to specific anatomical locations. Specifically, to assess the spatial distribution of the signal changes within each subject in a common reference frame and to measure bias possibly tied to subject anatomy, we jointly coregister each subjects' pre- and post-dStripe data to create a common population-average template for pre- and post-dStripe data. Below we analyse changes to signal properties between groups, first across the whole brain then spatially-resolved.

##### 3.4.1. Whole-brain analysis

Figure 8 shows the relative change of CSD-derived quantities inside the brain that can be attributed to applying dStripe. We display the bias using histograms of the voxel-wise relative change in the population-average fluid and tissue component volume fractions and the tissue angular power ( $A$ ). After

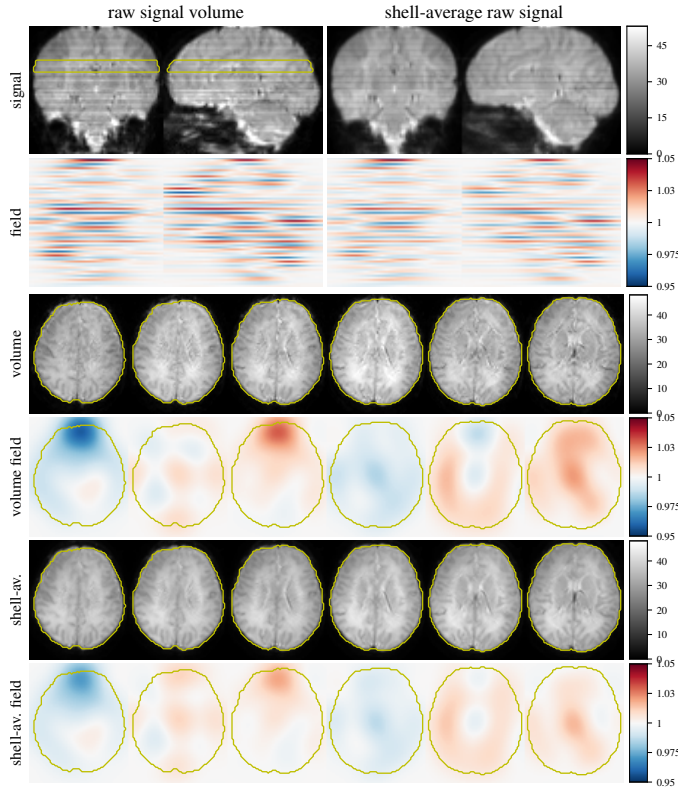


**Fig. 3.** Exemplary shell-average signal pre- and post-dStripe and shell-average signal difference images displayed in motion-corrected anatomical-space.

dStripe, on average, the volume fractions in the template components are close to constant (fluid: +0.12%, tissue: -0.07%), the tissue angular energy is slightly reduced in the  $\ell = 2$  (-0.30%) and  $\ell = 4$  (-0.44%) band and slightly increased in the  $\ell = 6$  band (+1.00%). This is in line with an observed overall

reduction in angular power in single-subject dMRI signal (see figure 2d).

In the coregistered data, we use the intra-template cross-subject standard deviation as a measure of data consistency. Figure 8 (B) shows the histograms of the relative reduction in



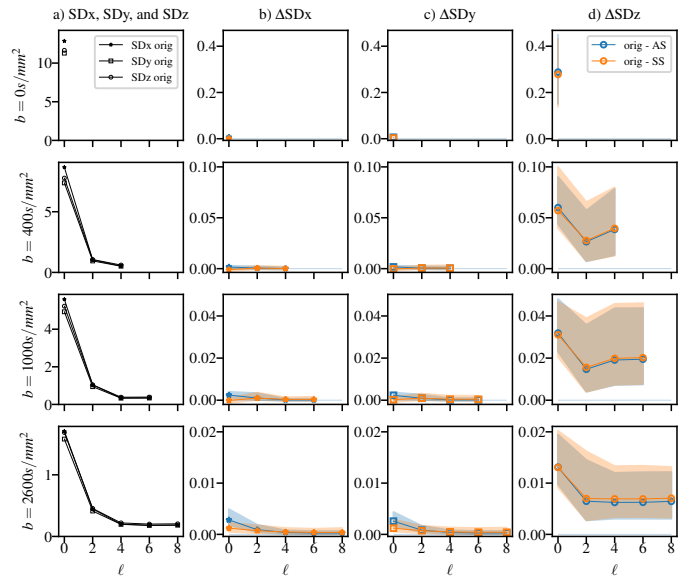
**Fig. 4.** Signal and slide modulation field in scattered source space. Shown is an individual  $b=0$  volume with relatively little motion and the  $b=0$  source-space average signal as well as dStripe fields estimated in source space through-plane and inplane. Note that the averaged signal contains signal dropout and other artefacts that are suppressed by the motion correction algorithm and are therefore not accounted for in the dStripe field. Through-plane frequency constraints apply in “slice-native” space, high-frequency variations are not enforced in source space shown here.

voxel-wise cross-subject standard deviation due to dStripe. The relative inter-subject variation is most decreased by dStripe in the fluid volume fraction and in the  $\ell = 4$  band of the tissue signal. On average, dStripe decreases inter-subject standard deviation nearly across the entire power spectrum of the fluid and tissue components (figure 8f: -1.61%, g: -0.64%, h: -0.97%, i: -2.1%, j: -0.24%) indicating higher data consistency across subjects after dStripe.

Performing dStripe in anatomical space or in source space yields close to identical results but source-space dStripe exhibits a higher dispersion of the voxel-wise population-average tissue volume fractions and exhibits slightly wider inter-subject standard deviation (compare lines in figure 8).

### 3.4.2. Spatially-resolved analysis

Here we use data from the relatively homogeneous group of 32 subjects selected for template creation to investigate whether the application of dStripe causes systematic changes (bias) or increased variance of the dMRI signal in certain anatomical locations. After alignment of the data to the population-average template space, the heterogeneity of subject position in source-space should dampen the influence of any subject-specific stripe patterns. Any systematic difference between coregistered and averaged pre-dStripe data and post-dStripe data beyond random



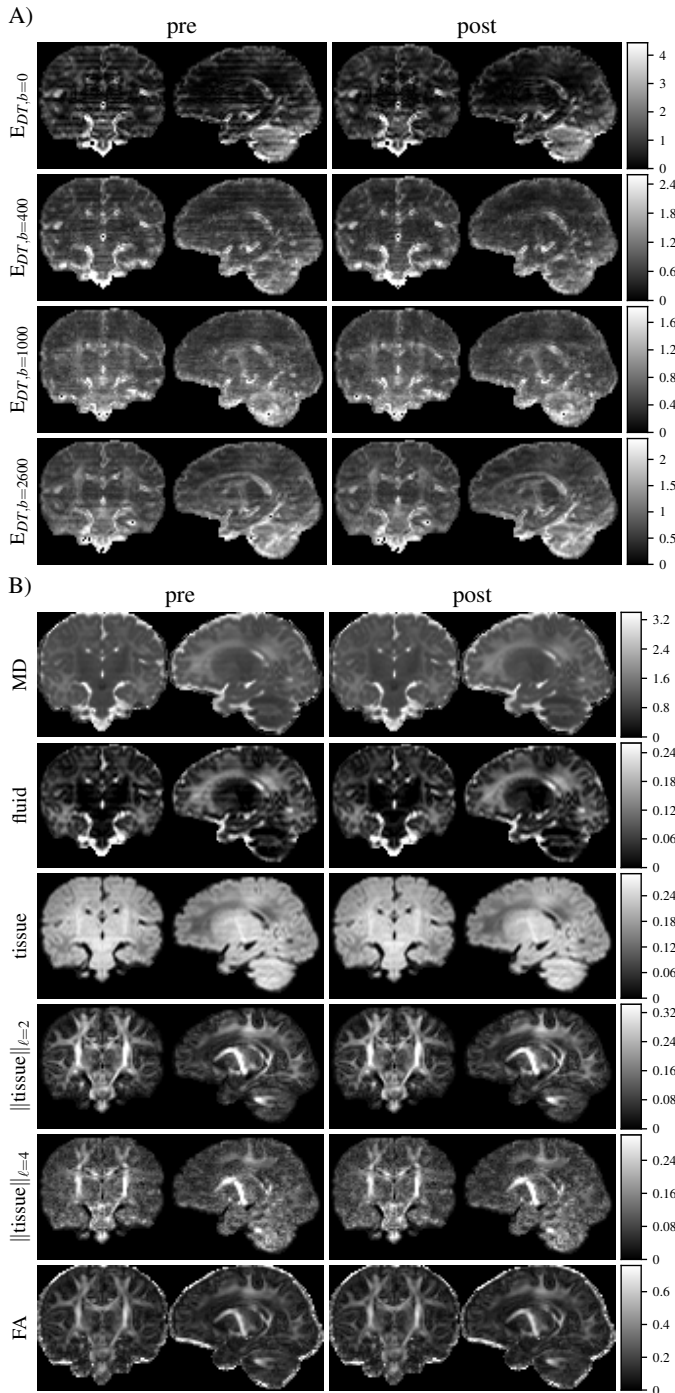
**Fig. 5.** The local standard deviation of the DWI signal in the SH-basis in plane (SDx, SDy) and through-plane (SDz) by harmonic order of the original data after reconstruction (a) and the subject-wise change after destriping (pre-post) in anatomical space (AS) and source space (SS). The through-plane SD is reduced throughout the spectrum, the in-plane SD is approximately preserved. Both dStripe methods perform similarly, SS dStripe is slightly superior for data with larger residuals.  $n=44$ .

variations due to the finite number of images and subject poses in source-space indicates either systematic anatomical bias of the method or a dependence of the stripe artefact on subject anatomy.

Figure 9 shows the corresponding images of the population-average (a), its absolute change (b) and the change in within-template standard deviation (d) due to dStripe as well as a map of the within-subject standard deviation across subjects in template space (c). We use these to spatially assess variance and bias associated with dStripe.

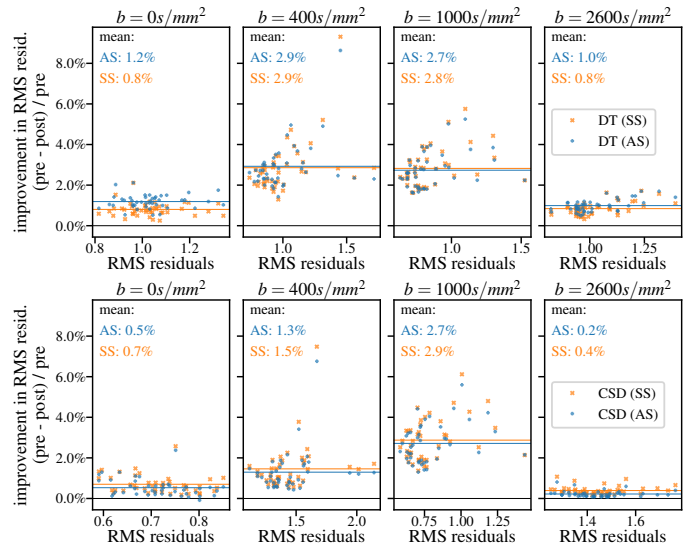
In the majority of locations, the inter-subject variance is reduced by dStripe, in particular in the tissue  $\ell = 4$  band (figure 9d), indicating that dStripe increases consistency throughout the brain and, despite the spatial inhomogeneity of the across-subject standard deviation of signal changes (figure 9c), does not introduce inconsistencies at specific locations. Within individual subjects, the fluid component is more affected by dStripe than the tissue component (figure 9c) but no such systematic effect is observable in the population average, indicating that the variance due to the stripe modulation in the fluid component is larger than in the tissue component.

In the template, systematic effects across subjects due to dStripe (bias) are small (figure 9b) compared to the inter-subject standard deviation of within-subject signal changes (figure 9c). This shows that across subjects, the changes due to dStripe tend to be independent of anatomical location. The most prominent local effect on the average compartment volume fractions occur superior to the body of the corpus callosum and in the vicinity of the pons, medulla and cerebellum and the superior frontal lobe, areas prone to pulsation and distortion artefacts. A consistent local reduction of angular power can be observed superior



**Fig. 6.** Shell-specific root mean squared DT fit residuals **E** (A) and DT and MT CSD fit-derived quantities (B). For both fits, dStripe reduces the interslice variation within the residuals and in the DT-derived mean diffusivity (MD), as well as CSD-derived fluid and tissue fraction maps. Fractional anisotropy (FA) and tissue anisotropy appear visually unchanged.

to the body of the corpus callosum (cingulum and parts of the fornix), most prominent in the template in the tissue  $\ell = 2$  and  $\ell = 4$  and FA images (b). This systematic reduction in angular power in the population-average template resembles a feature in appearance and amplitude that can be observed in the angular spectrum of the unregistered cohort-average ( $N=700$ ) raw data signal. This creates some ambiguity about whether and how



**Fig. 7.** Relative improvement of the shell-average root mean square residuals for diffusion tensor (top row) and MT CSD (bottom row) fits after destriping compared to fitting non-destriped data for  $n=44$  subjects.

much of this change should be attributed to anatomical bias of the network rather than a genuine need for consistent change as our reference point for zero expected change in the population-average space might not be neutral in this area.

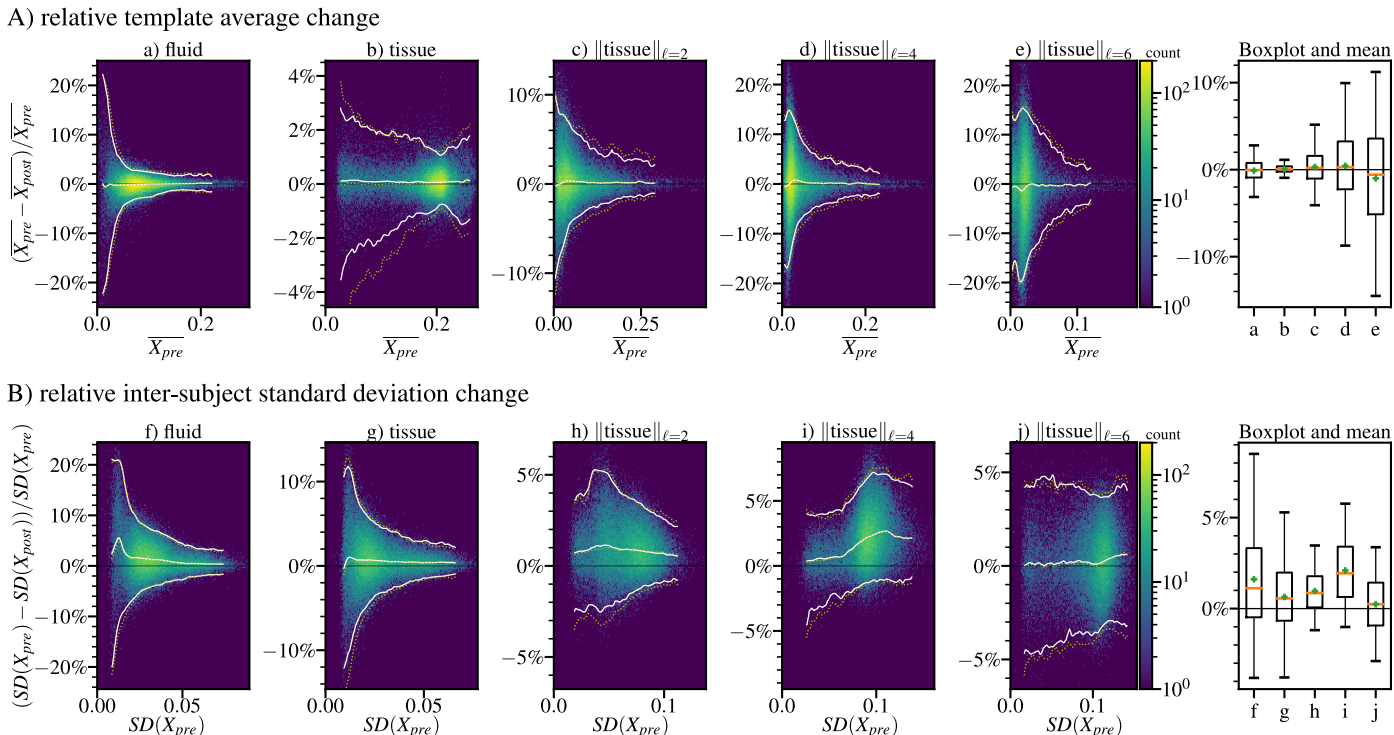
### 3.5. Generalisation

To generalise to unseen data, dStripe needs to remove interslice intensity artefacts while preserving the integrity of the data. As with other deep learning techniques (Antun *et al.*, 2020), the transferability to other cohorts and acquisition parameters needs to be investigated on a case by case basis as generalisation performance is in general dependent on the quantity and quality of the training data and the expected domain shift.

On dHCP data, we observe that dStripe is relatively data-efficient. When trained on a single scan ( $N=1$ ) and applied to that training data, this network achieves a similar reduction in stripe patterns compared to the network trained on 20 different scans (see figure B.11 b-e). This data-efficiency suggests that for target domains similar to the dHCP data, the network training and inference can be readily transferred to other and much smaller cohorts.

When the network is trained on two scans (29 and 38 weeks PMA) acquired with the Philips Achieva 3T product sequence (1  $b=0$  and 32  $b=750$   $s/mm^2$  volumes) that exhibit inhomogeneous slice timing artefacts, resulting network performance varies between training runs when assessed on these scans and on dHCP data (figure B.11 f-h). On this training data, the network requires more than the 66 unique volumes to reliably converge to a good solution.

The performances of networks trained on 20 or 1 dHCP scans when applied to the product sequence data are on a par with that of the best instance trained on the native data (figure B.11 B, C, columns b-e vs h) suggesting that domain adaptation is feasible between both datasets. For data-limited applications for which self-supervised network training is not stable, more ex-



**Fig. 8.** 2D histograms of the fluid and tissue components densities and angular energy spectra (band-specific  $l_2$  norm of the spherical harmonic coefficients) using non-destriped data (pre-dStripe) and data destriped in anatomical-space (post-dStripe). **A** shows the population-average change due to dStripe (comparing the template images), **B** compares the within-template voxel-wise standard deviation across subjects between original and destriped data.  $X$  is a placeholder for the respective quantities denoted in the plot titles. Boxplots summarise the 5, 25, 50, 75, and 95 percentile and mean (+) in each band. White lines demark the (conditional) 5%, 50%, 95% probability density of the 2D histograms, yellow dotted lines show the results for source-space destriping.

tensive data augmentation, domain adaptation, pooling of training data across cohorts, transfer learning, and model selection techniques are possible avenues to improve performance.

### 3.6. Limitations

The focus of this work is to develop and evaluate a framework that allows removing stripe artefacts from motion corrupted data without access to ground truth data. Our algorithm choices were guided by observations within our data and are potentially tied to the choice of SVR algorithm. The dependency on the motion corrected prediction can limit the destriping of the source data when subject motion causes stripe patterns to be absent in the anatomical-space – they can be either spatially smeared out, averaged with other data, or removed by the outlier detection mechanism.

This study is a first step at modelling inter-slice intensity modulations. We do not model interactions across slices, volumes (directions) and we do not explicitly use the motion trajectory in the dStripe model. Inter-slice interactions, for instance caused by inhomogeneous g-factor maps biasing the multiband reconstruction, could require unmixing of the signal, which would break the modulation field assumptions made in this work. This and longer-range bias across multiband pack boundaries can be explored in future work.

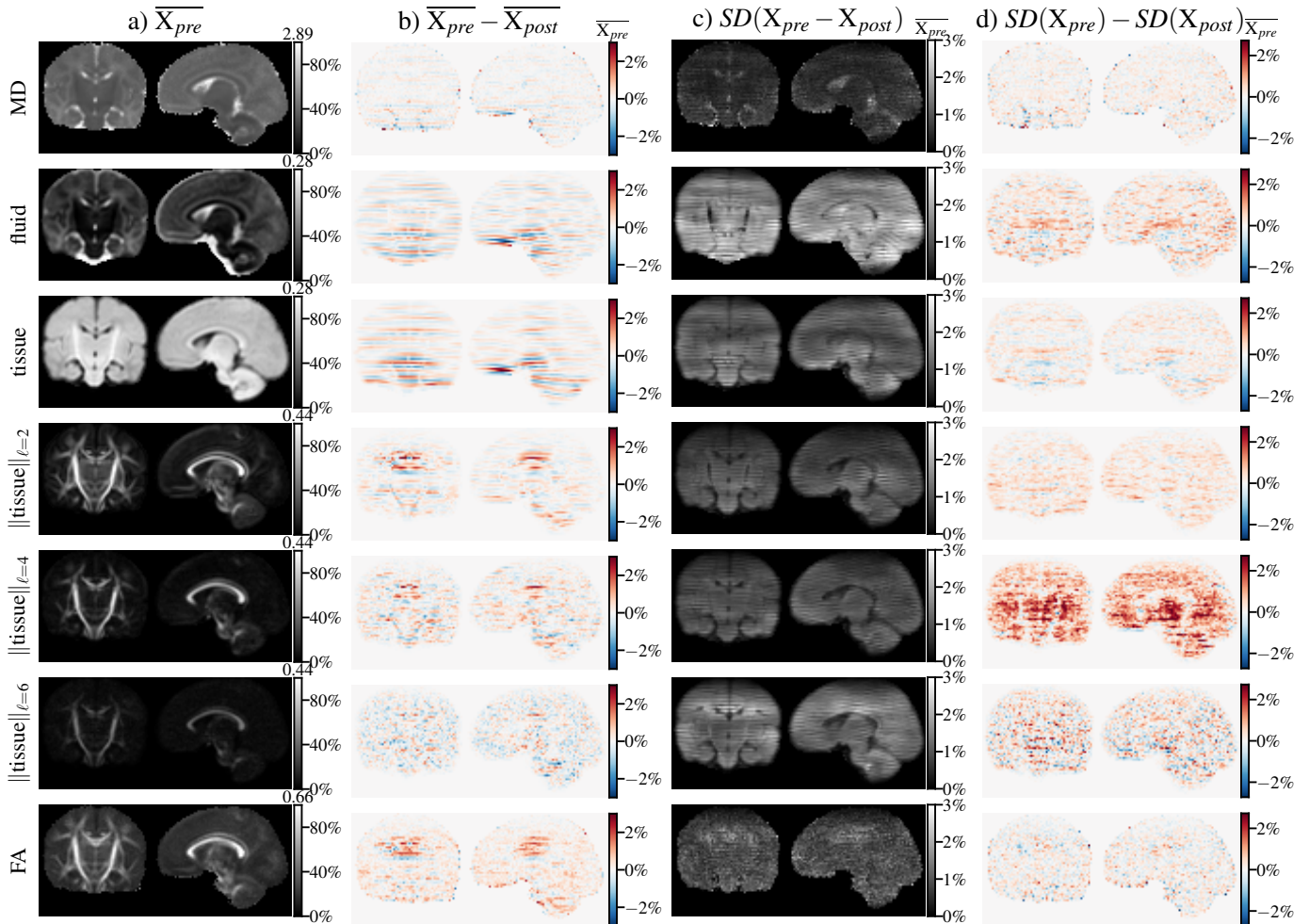
We do not take tissue-dependencies of the stripe artefact into account. However, if needed, the signal could be decomposed into distinct compartments that separate long T1 from short T1

species, each destriped independently using the same approach as described here, and subsequently recombined. This approach requires the ability to robustly decompose the signal into components relevant for the artefact – in the presence of stripe artefacts.

## 4. Conclusions and outlook

We presented a data-driven method for the removal of stripe artefacts from dMRI data. dStripe reduces stripe artefacts from the shell-average and the angular signal components, and thus decreasing DTI and MT CSD fit residuals across shells. Single-subject component and DTI-derived images appear visually less stripy and inter-subject variation is reduced indicating improved data consistency across subjects.

Applying the dStripe approach in source space (which is likely the space of origin of the slice modulation artefact) slightly outperforms anatomical-space destriping in terms of the reduction of inter-slice signal variation and of signal representation fit residuals. However, it is limited in its applicability as it requires control over the SVR framework’s forward-projection and comes with a high computational cost. We showed that anatomical-space dStripe is a suitable substitute as it produces similar results and performs equally in terms of data consistency. For cohorts with less motion we expect the results of anatomical-space dStripe to be even more favourable, in particular for data acquired using lower multiband factors.



**Fig. 9.** Effect of dStripe on diffusion tensor and MT CSD-derived quantities aligned to the population-average template ( $n=32$ ). The template (population average  $\bar{X}$ ) pre-dStripe is shown in column a) and intensity values in all other columns are expressed relative to the maximum intensity in a). The effect of dStripe on the population-average is shown in b). Column c) displays the standard deviation (SD) of the within-subject changes across subjects due to dStripe, indicating the expected effect on a subject-level. The change to the within-group heterogeneity measured as the difference between standard deviation across aligned subjects pre- and post-dStripe data is shown in column d).

While our pipeline is optimised for diffusion MRI, it could be adapted for other modalities and applications such as the detection and removal of stripe artefacts in transcranial ultrasound imaging (Vignon *et al.*, 2010), line and area levelling in atomic force microscopy (Raposo *et al.*, 2007; Wang *et al.*, 2018), or venetian blind artefact removal in multiple overlapping thin slab acquisitions (Blatter *et al.*, 1991).

## 5. Acknowledgements

This work received funding from the European Research Council under the European Union’s Seventh Framework Programme ([FP7/2007-2013/ERC] grant agreement no. [319456] dHCP project), and was supported by the Wellcome/EPSRC Centre for Medical Engineering at King’s College London [WT 203148/Z/16/Z]; the Medical Research Council [MR/K006355/1] and by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London. The views expressed are those of the author(s) and not neces-

sarily those of the NHS, the NIHR or the Department of Health. We would like to thank Prof. Serena Counsell for access to the Philips Achieva 3T product sequence data.

## References

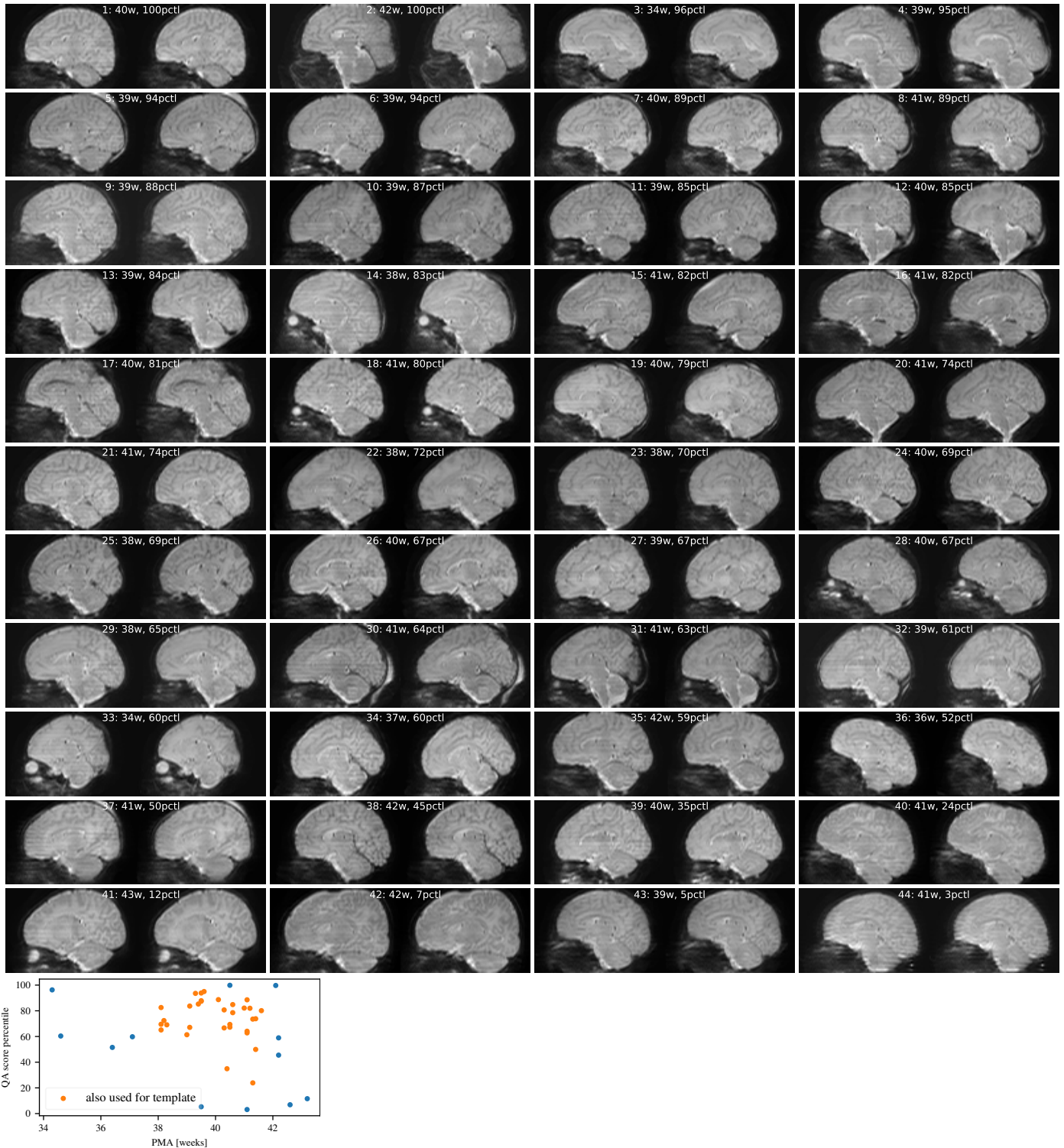
- Andersson, J.L., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* 20, 870–888. doi:10.1016/s1053-8119(03)00336-7.
- Andersson, J.L., Sotiropoulos, S.N., 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion mr imaging. *Neuroimage* 125, 1063–1078. doi:10.1016/j.neuroimage.2015.10.019.
- Antun, V., Renna, F., Poon, C., Adcock, B., Hansen, A.C., 2020. On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proceedings of the National Academy of Sciences* doi:10.1073/pnas.1907377117.
- Barth, M., Breuer, F., Koopmans, P.J., Norris, D.G., Poser, B.A., 2016. Simultaneous multislice (sms) imaging techniques. *Magnetic resonance in medicine* 75, 63–81. doi:10.1002/mrm.25897.
- Basser, P.J., Mattiello, J., LeBihan, D., 1994. Mr diffusion tensor spectroscopy and imaging. *Biophysical journal* 66, 259–267. doi:10.1016/s0006-3495(94)80775-1.

- Blaimer, M., Choli, M., Jakob, P.M., Griswold, M.A., Breuer, F.A., 2013. Multiband phase-constrained parallel mri. *Magnetic resonance in medicine* 69, 974–980. doi:10.1002/mrm.24685.
- Blatter, D., Parker, D., Robison, R., 1991. Cerebral mr angiography with multiple overlapping thin slab acquisition. part i. quantitative analysis of vessel visibility. *Radiology* 179, 805–811. doi:10.1148/radiology.179.3.2027996.
- Burstein, D., 1996. Stimulated echoes: Description, applications, practical hints. *Concepts in Magnetic Resonance* 8, 269–278. doi:10.1002/(sici)1099-0534(1996)8:4<269::aid-cmr3>3.0.co;2-x.
- Cauley, S.F., Polimeni, J.R., Bhat, H., Wald, L.L., Setsompop, K., 2014. Inter-slice leakage artifact reduction technique for simultaneous multislice acquisitions. *Magnetic resonance in medicine* 72, 93–102. doi:10.1002/mrm.24898.
- Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258.
- Christiaens, D., Cordero-Grande, L., Hutter, J., Price, A.N., Deprez, M., Hajnal, J.V., Tournier, J.D., 2019. Learning compact q-space representations for multi-shell diffusion-weighted mri. *IEEE transactions on medical imaging* 38, 834–843. doi:10.1109/TMI.2018.2873736.
- Christiaens, D., Cordero-Grande, L., Pietsch, M., Hutter, J., Price, A.N., Hughes, E.J., Vecchiato, K., Deprez, M., Edwards, A.D., Hajnal, J.V., Tournier, J.D., 2020. Scattered slice SHARD reconstruction for motion correction in multi-shell diffusion MRI. *NeuroImage*, 117437doi:10.1016/j.neuroimage.2020.117437.
- Crawley, A., Henkelman, R., 1987. A stimulated echo artifact from slice interference in magnetic resonance imaging. *Medical physics* 14, 842–848. doi:10.1118/1.596011.
- Dagli, M.S., Ingeholm, J.E., Haxby, J.V., 1999. Localization of cardiac-induced signal change in fmri. *Neuroimage* 9, 407–415. doi:10.1006/nimg.1998.0424.
- David, S., Heemskerk, A., Viergever, M., Leemans, A., 2017. Closing the venetian blinds: A processing strategy for correcting stripe artifacts in diffusion mri data. in: *ESMRMB Annual Scientific Meeting*, p. 215.
- Dhollander, T., Mito, R., Raffelt, D., Connelly, A., 2019. Improved white matter response function estimation for 3-tissue constrained spherical deconvolution. in: *Proc. Intl. Soc. Mag. Reson. Med.*, p. 555.
- Efron, B., 2004. The estimation of prediction error: covariance penalties and cross-validation. *Journal of the American Statistical Association* 99, 619–632. doi:10.1198/016214504000000692.
- Engström, M., Skare, S., 2013. Diffusion-weighted 3d multislabs echo planar imaging for high signal-to-noise ratio efficiency and isotropic image resolution. *Magnetic resonance in medicine* 70, 1507–1514. doi:10.1002/mrm.24594.
- Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., Turner, R., 1996. Movement-related effects in fmri time-series. *Magnetic resonance in medicine* 35, 346–355. doi:10.1002/mrm.1910350312.
- Frost, R., Miller, K.L., Tijssen, R.H., Porter, D.A., Jezzard, P., 2014. 3d multi-slab diffusion-weighted readout-segmented epi with real-time cardiac-reordered k-space acquisition. *Magnetic resonance in medicine* 72, 1565–1579. doi:10.1002/mrm.25062.
- Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A., 2002. Generalized autocalibrating partially parallel acquisitions (grappa). *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 47, 1202–1210. doi:10.1002/mrm.10171.
- Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141. doi:10.1109/cvpr.2018.00745.
- Hughes, E.J., Winchman, T., Padormo, F., Teixeira, R., Wurie, J., Sharma, M., Fox, M., Hutter, J., Cordero-Grande, L., Price, A.N., Allsop, J., Bueno-Conde, J., Tumor, N., Arichi, T., Edwards, A.D., Rutherford, M.A., Counsell, S.J., Hajnal, J.V., 2017. A dedicated neonatal brain imaging system. *Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 78, 794–804. doi:10.1002/mrm.26462.
- Hutter, J., Tournier, J.D., Price, A.N., Cordero-Grande, L., Hughes, E.J., Malik, S., Steinweg, J., Bastiani, M., Sotiropoulos, S.N., Jbabdi, S., et al., 2018. Time-efficient and flexible design of optimized multishell hardi diffusion. *Magnetic resonance in medicine* 79, 1276–1292. doi:10.1002/mrm.26765.
- Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
- Jeurissen, B., Tournier, J.D., Dhollander, T., Connelly, A., Sijbers, J., 2014. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data. *NeuroImage* 103, 411–426. doi:10.1016/j.neuroimage.2014.07.061.
- Kazhdan, M., Funkhouser, T., Rusinkiewicz, S., 2003. Rotation invariant spherical harmonic representation of 3 d shape descriptors. in: *Symposium on geometry processing*, pp. 156–164.
- Kelly, M.E., Duff, E.P., Bijsterbosch, J.D., Voets, N.L., Filippini, N., Moeller, S., Xu, J., Yacoub, E.S., Auerbach, E.J., Ugurbil, K., et al., 2013. An assessment of motion artefacts in multi band epi for high spatial and temporal resolution resting state fmri. in: *Proceedings of the ISMRM*, Abstract.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Larkman, D.J., Hajnal, J.V., Herlihy, A.H., Coutts, G.A., Young, I.R., Ehnholm, G., 2001. Use of multicoil arrays for separation of signal from multiple slices simultaneously excited. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 13, 313–317. doi:10.1002/1522-2586(200102)13:2<313::aid-jmri1045>3.0.co;2-w.
- Larkman, D.J., Nunes, R.G., 2007. Parallel magnetic resonance imaging. *Physics in Medicine & Biology* 52, R15. doi:10.1088/0031-9155/52/7/r01.
- Liao, C., Stockmann, J., Tian, Q., Bilgic, B., Manhard, M.K., Wald, L.L., Setsompop, K., 2018. High-fidelity, high-isotropic resolution diffusion imaging through gslider acquisition with b1+ & t1 corrections and multi-coil b0 shim array. *arXiv preprint arXiv:1811.05473*.
- Mansfield, P., 1977. Multi-planar image formation using nmr spin echoes. *Journal of Physics C: Solid State Physics* 10, L55. doi:10.1088/0022-3719/10/3/004.
- McGibney, G., Smith, M., Nichols, S., Crawley, A., 1993. Quantitative evaluation of several partial fourier reconstruction algorithms used in mri. *Magnetic resonance in medicine* 30, 51–59. doi:10.1002/mrm.1910300109.
- McNabb, C.B., Lindner, M., Shen, S., Murayama, K., Burgess, L.G., Johnstone, T., 2018. Inter-slice leakage and intra-slice aliasing in simultaneous multi-slice echo-planar images. *bioRxiv*, 458042doi:10.1101/458042.
- Metzler, C.A., Mousavi, A., Heckel, R., Baraniuk, R.G., 2018. Unsupervised learning with stein’s unbiased risk estimator. *arXiv preprint arXiv:1805.10531*.
- Meyer, M., Biber, A., Koch, M., 2014. Chasing the zebra. the quest for the origin of a stripe artifact in diffusion-weighted mri. in: *Student Conference Medical Engineering Science 2014: Proceedings*, GRIN Verlag. p. 199.
- mrtrix.readthedocs.io, 2020. Spherical Harmonics documentation. URL: [https://mrtrix.readthedocs.io/en/3.0.0/concepts/spherical\\_harmonics.html](https://mrtrix.readthedocs.io/en/3.0.0/concepts/spherical_harmonics.html).
- Muresan, L., Renken, R., Roerdink, J.B., Duifhuis, H., 2005. Automated correction of spin-history related motion artefacts in fmri: simulated and phantom data. *IEEE transactions on biomedical engineering* 52, 1450–1460. doi:10.1109/tbme.2005.851484.
- Murphy, K., Birn, R.M., Bandettini, P.A., 2013. Resting-state fmri confounds and cleanup. *Neuroimage* 80, 349–359. doi:10.1016/j.neuroimage.2013.04.001.
- Parker, D.L., Yuan, C., Blatter, D.D., 1991. Mr angiography by multiple thin slab 3d acquisition. *Magnetic resonance in medicine* 17, 434–451. doi:10.1002/mrm.1910170215.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. Pytorch: An imperative style, high-performance deep learning library. in: *Wallach, H., Larochelle, H., Beygelzimer, A., d’Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., pp. 8024–8035.
- Pietsch, M., Christiaens, D., Hutter, J., Cordero-Grande, L., Price, A.N., Hughes, E., Edwards, A.D., Hajnal, J.V., Counsell, S.J., Tournier, J.D., 2019a. A framework for multi-component analysis of diffusion mri data over the neonatal period. *NeuroImage* 186, 321–337. doi:10.1016/j.neuroimage.2018.10.060.
- Pietsch, M., Christiaens, D., Tournier, J.D., Hajnal, J.V., 2019b. D-stripe: correction for stripe artefacts in diffusion mri using a combined deep neural network and svr approach. in: *Proceedings of the ISMRM*, Abstract.
- Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., 1999.

- Sense: sensitivity encoding for fast mri. *Magnetic resonance in medicine* 42, 952–962. doi:10.1002/(sici)1522-2594(199911)42:5<952::aid-mrm16>3.0.co;2-s.
- Ramani, S., Blu, T., Unser, M., 2008. Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms. *IEEE Transactions on image processing* 17, 1540–1554. doi:10.1109/tip.2008.2001404.
- Raposo, M., Ferreira, Q., Ribeiro, P., 2007. A guide for atomic force microscopy analysis of soft-condensed matter. *Modern research and educational topics in microscopy* 1, 758–769.
- Santurkar, S., Tsipras, D., Ilyas, A., Madry, A., 2018. How does batch normalization help optimization?, in: *Advances in Neural Information Processing Systems*, pp. 2483–2493.
- Setsompop, K., Fan, Q., Stockmann, J., Bilgic, B., Huang, S., Cauley, S.F., Nummenmaa, A., Wang, F., Rathi, Y., Witzel, T., et al., 2018. High-resolution in vivo diffusion imaging of the human brain with generalized slice dithered enhanced resolution: Simultaneous multislice (g slider-sms). *Magnetic resonance in medicine* 79, 141–151. doi:10.1002/mrm.26653.
- Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., Wald, L.L., 2012. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic resonance in medicine* 67, 1210–1224. doi:10.1002/mrm.23097.
- Smith, L.N., 2017. Cyclical learning rates for training neural networks, in: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE. pp. 464–472. doi:10.1109/wacv.2017.58.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Human brain mapping* 17, 143–155. doi:10.1002/hbm.10062.
- Soltanayev, S., Chun, S.Y., 2018. Training deep learning based denoisers without ground truth data, in: *Advances in Neural Information Processing Systems*, pp. 3257–3267.
- Stein, C.M., 1981. Estimation of the mean of a multivariate normal distribution. *The annals of Statistics* , 1135–1151 doi:10.1214/aos/1176345632.
- Todd, N., Moeller, S., Auerbach, E.J., Yacoub, E., Flandin, G., Weiskopf, N., 2016. Evaluation of 2d multiband epi imaging for high-resolution, whole-brain, task-based fmri studies at 3t: sensitivity and slice leakage artifacts. *Neuroimage* 124, 32–42. doi:10.1016/j.neuroimage.2015.08.056.
- Tournier, J.D., Christiaens, D., Hutter, J., Price, A.N., Cordero-Grande, L., Hughes, E., Bastiani, M., Sotiropoulos, S.N., Smith, S.M., Rueckert, D., et al., 2020. A data-driven approach to optimising the encoding for multi-shell diffusion mri with application to neonatal imaging. *NMR in Biomedicine* 33, e4348. doi:10.1002/nbm.4348.
- Van, A.T., Aksoy, M., Holdsworth, S.J., Kopeinigg, D., Vos, S.B., Bammer, R., 2015. Slab profile encoding (pen) for minimizing slab boundary artifact in three-dimensional diffusion-weighted multislabs acquisition. *Magnetic resonance in medicine* 73, 605–613. doi:10.1002/mrm.25169.
- Van, A.T., Hernando, D., Sutton, B.P., 2011. Motion-induced phase error estimation and correction in 3d diffusion tensor imaging. *IEEE transactions on medical imaging* 30, 1933–1940. doi:10.1109/tmi.2011.2158654.
- Veraart, J., Novikov, D.S., Christiaens, D., Ades-Aron, B., Sijbers, J., Fieremans, E., 2016. Denoising of diffusion mri using random matrix theory. *NeuroImage* 142, 394–406. doi:10.1016/j.neuroimage.2016.08.016.
- Vignon, F., Shi, W.T., Yin, X., Hoelscher, T., Powers, J.E., 2010. The stripe artifact in transcranial ultrasound imaging. *Journal of Ultrasound in Medicine* 29, 1779–1786. doi:10.7863/jum.2010.29.12.1779.
- Wang, Y., Lu, T., Li, X., Wang, H., 2018. Automated image segmentation-assisted flattening of atomic force microscopy images. *Beilstein journal of nanotechnology* 9, 975–985. doi:10.3762/bjnano.9.91.
- Wu, W., Koopmans, P.J., Frost, R., Miller, K.L., 2016. Reducing slab boundary artifacts in three-dimensional multislabs diffusion mri using nonlinear inversion for slab profile encoding (npen). *Magnetic resonance in medicine* 76, 1183–1195. doi:10.1002/mrm.26027.
- Wu, W., Miller, K.L., 2017. Image formation in diffusion mri: a review of recent technical developments. *Journal of Magnetic Resonance Imaging* 46, 646–662. doi:10.1002/jmri.25664.
- Yancey, S.E., Rotenberg, D.J., Tam, F., Chiew, M., Ranieri, S., Biswas, L., Anderson, K.J., Baker, S.N., Wright, G.A., Graham, S.J., 2011. Spin-history artifact during functional mri: potential for adaptive correction. *Medical physics* 38, 4634–4646. doi:10.1118/1.3583814.
- Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122* .

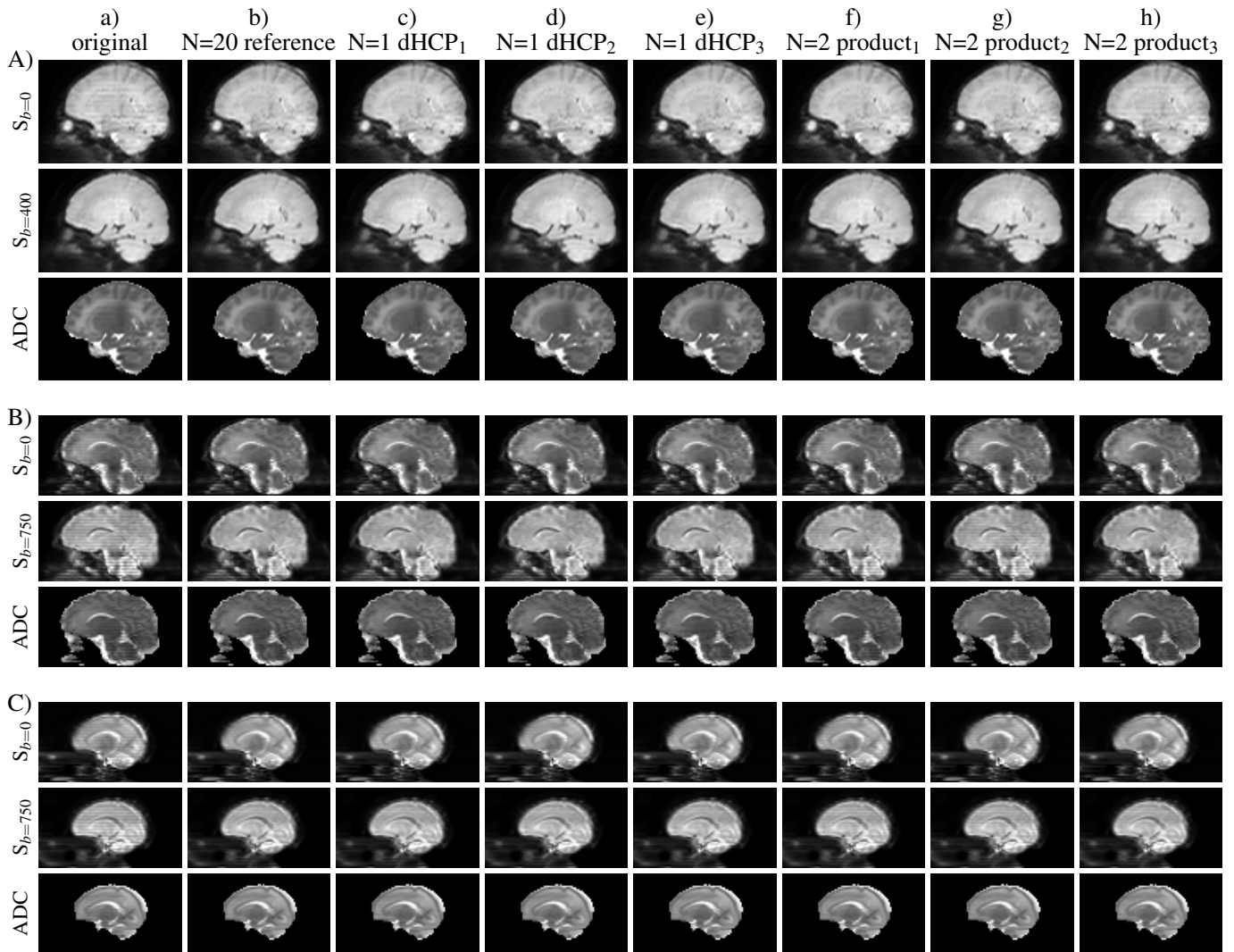


## Appendix A. Data overview



**Fig. A.10.** Top: sagittal cross-sections through  $b=0$  data pre-dStripe (left) and post-dStripe (right) for the 44 diffusion MRIs used for the analysis of the dStripe method. Images are sorted by percentile quality assessment score (higher is better). Bottom: Age at scan versus percentile quality assessment score. Data used for single-subject analysis and in the template is marked as such.

## Appendix B. Generalisation



**Fig. B.11.** Shell-average signal and ADC maps of the original (a) and denoised (b-h) data of a dHCP scan (A) and two scans (B, C) acquired with a Philips Achieva 3T product sequence which exhibit severe slice intensity artefacts due to inhomogeneous slice-timing. The columns correspond to the reference network trained on N=20 dHCP scans or to network instances either trained on the one dHCP scan shown in A) or on the two scans acquired with the product sequence. Instances from repeated training runs are denoted with subscripts 1-3. All scans shown in this figure were part of the training of the respective networks – except for the N=20 reference network.