

# Internally-consistent and fully-unbiased multimodal MRI brain template construction from UK Biobank: Oxford-MM

Christoph Arthofer<sup>a</sup>, Stephen M. Smith<sup>a</sup>, Gwenaëlle Douaud<sup>a</sup>, Andreas Bartsch<sup>b</sup>, Fidel Alfaro-Almagro<sup>a</sup>, Jesper Andersson<sup>a</sup>, Frederik J. Lange<sup>a</sup>

<sup>a</sup>*Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, United Kingdom*

<sup>b</sup>*Department of Neuroradiology, University of Heidelberg, Germany*

---

## Abstract

Anatomical MRI templates of the brain are essential to group-level analyses and image processing pipelines, as they provide a reference space for spatial normalisation. While it has become common for studies to acquire multimodal MRI data, many templates are still limited to one type of modality, usually either scalar or tensor-based. Aligning each modality in isolation does not take full advantage of the available complementary information, such as strong contrast between tissue types in structural images, or axonal organisation in the white matter in diffusion tensor images. Most existing strategies for multimodal template construction either do not use all modalities of interest to inform the template construction process, or do not use them in a unified framework.

Here, we present multimodal, cross-sectional templates constructed from UK Biobank data: the OMM-1 template, and age-dependent templates for each year of life between 45 to 81. All templates are fully unbiased to represent the average shape of the populations they were constructed from, and internally consistent through jointly informing the template construction process with T1, T2-FLAIR and DTI data. The OMM-1 template was constructed with a multi-resolution, iterative approach using 240 individuals in the 50-55 year age range. The age-dependent templates were estimated using a Gaussian Process, which describes the change in average brain shape with age in 37,330 individuals.

All templates show excellent contrast and alignment within and between modalities. The global brain shape and size is not preconditioned on existing

templates, although maximal possible compatibility with MNI-152 space was maintained through rigid alignment. We showed benefits in registration accuracy across two datasets (UK Biobank and HCP), when using the OMM-1 as the template compared with FSL’s MNI-152 template, and found that the use of age-dependent templates further improved accuracy to a small but detectable extent. All templates are publicly available and can be used as a new reference space for uni- or multimodal spatial alignment.

*Keywords:* Multimodal, Template, Age-dependent, UK Biobank, Registration

---

## 1. Introduction

Anatomical magnetic resonance imaging (MRI) templates of the brain aim to provide representative models of average shape and voxel signal intensity of the populations from which they were constructed. They are essential for many different kind of neuroimaging analyses as they provide a common reference space for the spatial normalization of individual subjects using image registration methods. The resulting transformations and derived measures, such as Jacobian determinant maps, between each individual and a template, and the transformed images in template space, enable the study of intra- and inter-group variability or agreement, unbiased group comparisons of within-subject longitudinal changes and atlas-based segmentation of regions of interest (ROIs) at subject level.

Template construction methods aim to find an average intensity and average shape template, i.e., the shape and intensity of the template should, on average, not be more like any one individual than any other (see section 2.3.2 for mathematical description). This is typically achieved through a series of steps to avoid bias in appearance or shape towards any single individual. The most commonly used method is based on an iterative framework (Guimond et al., 1998, 2000), which was later extended into a multi-resolution approach with a hierarchical processing structure (Grabner et al., 2006; Fonov et al., 2011). First, individual images are corrected for global (affine) misalignment using translation, rotation, scale and shear, which allows for the construction of an initial average affine template. Each individual is then iteratively non-linearly registered to the current template (starting with the affine template in the first iteration), followed by spatial unbiasing of the warps, and resampling of the subject images. Finally, the average across the resampled images



27 becomes the new template and serves as the reference space for the next it-  
28 eration. These steps are repeated until convergence, while warp resolution  
29 and image blurring are adjusted from coarse to fine.

30 Existing templates are often described as uni- or multimodal based on  
31 the number of modalities they comprise. An overview of some of the most  
32 commonly used and some more recent templates can be found in Table 1. In  
33 contrast to one modality in unimodal templates, multimodal templates aim  
34 to provide volumes of different, but anatomically-corresponding, modalities.  
35 This notion of multimodality in most existing templates stems from the post  
36 hoc availability of multiple modalities in template space, but generally does  
37 not refer to the modalities used during the template construction process.  
38 Driving this process with complementary information from different modal-  
39 ities of interest is highly desirable since it can improve registration quality.  
40 For example, the axonal organisation derived from diffusion imaging data can  
41 add valuable information about the white matter, which would not be avail-  
42 able from T1-weighted (T1) images only. Some existing templating methods  
43 use one modality to drive the construction, e.g., T1, and then apply the same  
44 deformation fields to all modalities of interest, e.g., T2-weighted (T2) or dif-  
45 fusion tensor images (DTI) (Rohlfing et al., 2010; Fonov et al., 2011; Gupta  
46 et al., 2016). Others use modalities *derived from* the modality of interest,  
47 e.g., fractional anisotropy (FA) maps from DTI, to drive the construction  
48 and then transform the modality of interest (DTI) with the same transfor-  
49 mations (Zhang et al., 2011; Lv et al., 2022). Estimating deformation fields  
50 based on a subset of modalities or surrogates, and applying the same defor-  
51 mation fields to all other modalities is not optimal. This strategy can lead to  
52 unwanted biasing effects in the template, since not all modalities contribute  
53 to the estimation of the deformation fields that are used for resampling and  
54 spatial unbiasing. This might not have a large impact when using modalities  
55 with similar information content, for example, when estimating a warp based  
56 on T1 images and applying the same warp to T1 and T2 images. However,  
57 for modalities with different information content it could introduce a spa-  
58 tial bias. For example, estimating deformation fields based on structural or  
59 diffusion-derived scalar modalities, and applying them to diffusion tensors  
60 could lead to a bias in the location or orientation of the diffusion data.

61 One fully-unbiased multimodal (FUMM) template was constructed from  
62 individuals in the adolescent brain and cognitive development (ABCD) study.  
63 For this template, eleven scalar modalities, including three structural modal-  
64 ities and eight dMRI-derived modalities but no DTI data were used as in-

Template	Template modalities and maps	#Subjects	Mean age $\pm$ sd (min-max)	Ref.
ICBM MNI305	<b>T1</b>	305 (66f/239m)	23.4 $\pm$ 4.1 (NA)	(Evans et al., 1993)
ICBM 152 linear	<b>T1</b> , T2, PD	152 (66f/86m)	25.02 $\pm$ 4.9 (18-44)	(Mazziotta et al., 1995)
ICBM 152 non-linear 6th gen.	<b>T1</b>	152 (66f/86m)	25.02 $\pm$ 4.9 (18-44)	(Grabner et al., 2006)
ICBM 2009a	<b>T1</b> , T2, PD, T2 relaxometry PVMs (GM, WM, CSF)	152 (66f/86m)	25.02 $\pm$ 4.9 (18-44)	(Fonov et al., 2011)
ICBM 2009b	<b>T1</b> , T2, PD	152 (66f/86m)	25.02 $\pm$ 4.9 (18-44)	(Fonov et al., 2011)
ICBM 2009c	<b>T1</b> , T2, PD PVMs (GM, WM, CSF)	152 (NA)	NA	(Fonov et al., 2011)
ICBM 152 extended nonlinear	<b>T1</b> , T2, PD	152 (66f/86m)	25.02 $\pm$ 4.9 (18-44)	(Fonov et al., 2011)
SRI24	<b>T1</b> , T2, PD, FA, MD, LD, mean DWI, PVMs (GM, WM, CSF), tissue labels, 2 CPMs	12 young (6f/6m) 12 elderly (6f/6m)	25.5 $\pm$ 4.34 (19-33) 77.7 $\pm$ 4.9 (67-84)	(Rohlfing et al., 2010)
Enhanced ICBM DT template	DTI, <b>PVMs (GM,WM,CSF) based on FA/trace map</b>	67 (40f/27m)	f: 27.2 $\pm$ 5.4 (20-39) m: 31.7 $\pm$ 5.6 (22-44)	(Zhang et al., 2011)
Clinical DTI	<b>T1</b> , DTI	48 (NA)	NA	(Gupta et al., 2016)
FOD template	FOD, <b>T1, T2, MD, FA, AFD, CX</b>	50 (25f/25m)	NA (22-35)	(Lv et al., 2022)
MINT ABCD atlas	<b>T1</b> , <b>PVMs (GM, WM)</b> , <b>0<sup>th</sup>, 2<sup>nd</sup> order SHCs of restricted FOD</b> , <b>0<sup>th</sup> order SHC of hindered &amp; free water FODs</b>	BL 11140 (5353f/5787m) FU 7578 (3503f/4075m)	median: 9.9 (8.9-11) median: 11.9 (10.6-13.8)	(Pechewa et al., 2022)
TBI template	<b>T1</b> , <b>DTI</b>	TBI 12 (5f/7m) HC 9 (3f/6m)	35 $\pm$ 12.1 (21-59) 36.2 $\pm$ 8.8 (23-46)	(Avants et al., 2008)
HCP atlas	<b>T1</b> , <b>T2</b> , <b>DTI</b>	971 (520f/451m)	NA (22-35)	(Irfanoglu et al., 2020)
MIITRA atlas	<b>T1</b> , <b>DTI</b>	202 (101f/101m)	80.56 $\pm$ 8.14 (65.2-94.9)	(Wu et al., 2022)

Table 1: Overview of existing unimodal and multimodal templates. **Modalities in bold** are used in the construction. PD...proton-density weighted, LD...longitudinal diffusivity, DWI...diffusion-weighted imaging, PVMs...partial volume maps, CPMs...cortical parcellation maps, SHC...spherical harmonics coefficient, AFD...apparent fibre density, CX...fibre complexity, FOD...fibre orientation distribution, BL...baseline, FU...follow up

put to the Multimodal Image Normalisation Tool (MINT) (Pecheva et al., 2022). Another FUMM templating strategy for scalar and tensor modalities was applied in the construction of the MIITRA template (Wu et al., 2022). The method alternates between registrations within each of the T1 and DTI modalities. In each iteration, deformation fields are estimated within one of the two modalities with a modality-specific registration method. The same transformations are applied to data from both modalities in all iterations except the last, where the DTI data undergoes one more transformation that is not applied to the T1 images. A similar iterative approach, involving multiple repeated registrations with the two methods, is required when spatially normalising individuals to the MIITRA template. Since both modalities drive the template construction the resulting templates are fully unbiased. However, the use of two methods does not provide a unified and internally-consistent framework. To the best of our knowledge, the only two methods that can accommodate both scalar and tensor modalities, and, consequently, allow fully-unbiased and internally-consistent template construction, are Symmetric Normalization for Multivariate Neuroanatomy (SyNMN) (Avants et al., 2008) and DR-TAMAS (Irfanoglu et al., 2016). SyNMN was applied in the construction of a combined T1 and DTI template to investigate traumatic brain injury (TBI) and later in the construction of a template from arterial spin labelling, T2-weighted-Fluid-attenuated inversion recovery (T2-FLAIR), DTI, functional MRI (fMRI), T1 and T2 data (Tustison et al., 2015). The SyNMN tool and templates are not publicly available at the time of writing. DR-TAMAS has been used for the construction of a DTI atlas (Irfanoglu et al., 2020) from the Human Connectome Project Young Adult (Van Essen et al., 2012) dataset (22 - 35 year age range). This DTI template also comprises T1 and T2 volumes, and all modalities were used to drive the registrations during the template construction process. The atlas was constructed from 971 individuals and has good levels of detail and contrast (although not quite as good as might be hoped for, given the quality of the data and the number of subjects).

Most existing multimodal templates provide a single, cross-sectional average of brain shape and intensity from the subjects in a cohort. However, arguably, a template should also be similar to a given population under investigation to reduce the amount of deformation required when aligning individuals to it. The main factor contributing to morphological variability in large datasets is the subjects' age range. As datasets become larger in size and the subjects' age range within datasets increases, it becomes more

difficult to capture the age-related increase in brain shape variability in a single template. Spatiotemporal, or age-dependent templates (ADTs), for sub-populations with smaller age ranges can provide more similar reference spaces. Several construction methods based on discrete bins (Fillmore et al., 2015), kernel regression (Davis et al., 2007; Serag et al., 2012) and neural network-based architectures (Dalca et al., 2019; Xia et al., 2019; Zhao et al., 2019; Wilms et al., 2020) have been described in the literature. However, these have been mainly used for unimodal ADT construction and, to the best of our knowledge, do not publicly provide multimodal templates with scalar and tensor modalities for general use.

### 1.1. Summary of our work

The main contributions of our work include the construction of a cross-sectional, internally consistent and fully-unbiased multimodal, whole-head template, the Oxford-MultiModal-1 (OMM-1), and the development of a modelling and prediction-based approach, which was applied in the construction of multimodal, average-shape ADTs.

The former was obtained from 240 UK Biobank (UKB) individuals (50-55 years, 50% females) with the iterative approach described in (Fonov et al., 2011). First, we constructed an unbiased affine template, which was refined from coarse to fine by iterating through nonlinear registrations, and unbiasing, warping and averaging steps. The ADTs were obtained by nonlinearly registering 37,330 UKB individuals (45-82 years) to OMM-1 and using the acquired deformation fields and corresponding individuals' ages to model the change in average brain shape with age using a Gaussian Process (GP). Finally, the trained model allowed us to predict and apply a mean deformation field for each year of age to derive age-dependent templates from the initial 240 UKB individuals.

The OMM-1 and its associated ADTs provide anatomically-corresponding scalar (T1 and T2-FLAIR) and tensor (DTI) volumes. These same modalities were used to drive the construction process by simultaneously informing the nonlinear registrations. These registrations were performed with FSL's MultiModal Registration Framework (MMORF) (Lange et al., 2020a,b), which estimates a single warp by optimizing over an arbitrary number of scalar and tensor input modalities. This ensures internal consistency by avoiding the need to use different registration methods for different modalities, and full-unbiasing of all volumes with respect to all modalities of interest. Our template construction pipeline provides a unified framework that can easily

140 be extended or adjusted to other scalar or tensor modalities of interest. Since  
 141 the OMM-1 is unbiased with respect to the UKB subjects from which it was  
 142 created, its size and shape differs from the most commonly used template, the  
 143 MNI 152 (in its various revisions), as MNI 152 templates are not very close to  
 144 representing the size of the average adult brain. However, the OMM-1 was  
 145 rigidly (6 degrees of freedom) aligned to MNI space, and transformations  
 146 between the two templates are provided, to aid compatibility when switch-  
 147 ing between them. Finally, we have investigated the benefits of using our  
 148 multimodal templating framework for spatial normalization in age-diverse  
 149 populations of two datasets.

## 150 2. Methods

### 151 2.1. Data

152 In this work we used scalar- and tensor-valued, non-defaced brain MRI  
 153 data from UKB (Miller et al., 2016), one of the largest prospective epidemi-  
 154 ological studies to date, which aims to acquire multimodal MR imaging data  
 155 from 100,000 participants.

156 Imaging data from three MRI modalities including T1, T2-FLAIR and  
 157 DTI were used for template construction. T1 provides information about the  
 158 basic anatomical structure of the brain and shows strong contrast between  
 159 the main tissue classes (gray and white matter, and cerebrospinal fluid). It is  
 160 acquired as part of most imaging studies and has become the core modality  
 161 of choice for existing adult human templates. T2-FLAIR was included as  
 162 a second structural modality due to its enhanced contrast of subcortical  
 163 gray matter regions, such as the striatum, pallidum, substantia nigra, red  
 164 nucleus and dentate nucleus, of the olfactory bulbs, but also between normal  
 165 appearing white matter and white matter hyperintensities. Diffusion MRI  
 166 provides information about the properties of the local tissue microstructure  
 167 and white matter tract structure. It makes it possible to estimate a diffusion  
 168 tensor for each voxel (Basser et al., 1994) that adds information about the  
 169 axonal organisation and the preferred directions of diffusion. We decided  
 170 to use non-defaced T1 and T2-FLAIR data to construct a template that is  
 171 sharp and clear in both intra- and extracranial regions and, hence, may be  
 172 useful for a variety of applications.

173 All imaging data were collected at one of three UKB sites using identical  
 174 3T Siemens Skyra scanners running VD13 and a standard Siemens 32-channel

Modality	Voxel size, Matrix	Key parameters
T1	$1.0 \times 1.0 \times 1.0$ mm, 208 $\times$ 256 $\times$ 256	3D MPRAGE, sagittal, R=2, TI/TR=880/2000 ms
T2-FLAIR	$1.05 \times 1.0 \times 1.0$ mm, 192 $\times$ 256 $\times$ 256	FLAIR, 3D SPACE, sagittal, R=2, PF 7/8, fat sat, TI/TR=1800/5000 ms, elliptical
dMRI	$2.0 \times 2.0 \times 2.0$ mm, 104 $\times$ 104 $\times$ 72	MB=3, R=1, TE/TR=92/3600 ms, PF 6/8, fat sat, b=0 s/mm <sup>2</sup> (5 $\times$ + 3 $\times$ phase-encoding-reversed), b=1000 s/mm <sup>2</sup> (50 $\times$ ) b=2000 s/mm <sup>2</sup> (50 $\times$ )

Table 2: UK Biobank brain MRI acquisition parameters for T1, T2-FLAIR and dMRI from (Miller et al., 2016). R=in-plane acceleration factor, MB=multiband factor, PF=partial Fourier

175 receive head coil. A brief overview of the parameters used to acquire T1, T2-  
176 FLAIR and dMRI can be found in Table 2. For a detailed description of the  
177 acquisition protocol in the UKB brain imaging study we refer the reader to  
178 (Miller et al., 2016).

179 The OMM-1 was constructed from 240 individuals uniformly and ran-  
180 domly sampled from the 50-55 year age range (40 individuals per year, 50%  
181 female). The size of the sample was informed by previous investigations on  
182 the Human Connectome Project dataset, where Yang et al. (Yang et al.,  
183 2020) have shown that sample sizes larger than 200 individuals are associ-  
184 ated with only small changes to the final templates. We selected individuals  
185 from the younger end of the UKB age range that provided sufficient data for  
186 uniform sampling. This minimises the appearance of ageing-related features  
187 and, therefore, maximises the utility of the template when applied to studies  
188 involving younger subjects.

189 All UKB images went through the manual and automated quality control  
190 (QC) pipeline described in (Alfaro-Almagro et al., 2018). Although badly  
191 corrupted images are excluded by this pipeline, several additional criteria  
192 for subjects to be considered in our random sample were defined. These re-  
193 quirements included the availability of all three modalities, less than 0.5% of  
194 the total brain volume containing white matter hyperintensities, and small  
195 alignment discrepancies. Alignment discrepancy measurements had been cal-  
196 culated as the correlation ratio between registered within-subject modalities  
197 by the QC pipeline and are available as QC imaging-derived phenotypes  
198 (IDPs) for all three modalities. Extreme scores are potential indicators for

199 poor alignment, or the presence of artefacts or outliers. In particular, we used  
200 IDPs that describe the discrepancies between an individual’s T1 structural  
201 image and the MNI 152 6th gen. (Grabner et al., 2006) after nonlinear align-  
202 ment, and between the T2-FLAIR and the corresponding T1 image, and the  
203 dMRI and the corresponding T1 image after linear alignment. Thresholds  
204 of 0.5 for the T1 and T2-FLAIR and 0.6 for the dMRI discrepancies were  
205 applied, to allow for a large enough sample of subjects from the selected age  
206 groups.

207 For the construction of the age-dependent templates, images from 37,330  
208 (age 45-82 years) individuals, which had T1, T2-FLAIR and dMRI data, were  
209 used. Given this large sample, the image quality at the individual level is  
210 expected to have less impact on the final average templates for age modelling  
211 compared to the smaller sample used for the OMM-1. Therefore, no further  
212 selection criteria were applied.

## 213 2.2. Data preprocessing

214 We used both minimally processed and preprocessed UKB imaging data.  
215 The T1, T2-FLAIR and dMRI volumes of the former are gradient-distortion  
216 corrected, and the T1 and T2-FLAIR volumes are not defaced, i.e., they  
217 include parts of the neck, nose and mouth. The latter had been prepro-  
218 cessed with the standard pipeline described in (Alfaro-Almagro et al., 2018),  
219 which, in addition to gradient-distortion correction, includes defacing, crop-  
220 ping, brain extraction through atlas-based mask propagation, and intensity  
221 inhomogeneity correction of T1 and T2-FLAIR images. Brain-extracted T2-  
222 FLAIR and dMRI images are rigidly co-registered to the corresponding indi-  
223 vidual’s T1 reference space using the B0s as the moving image and boundary-  
224 based registration (Greve and Fischl, 2009) as the cost function in FSL’s  
225 FLIRT (Jenkinson and Smith, 2001). dMRI data are corrected for suscepti-  
226 bility and eddy current distortion, as well as head motion, using FSL’s topup  
227 (Andersson et al., 2003) and eddy (Andersson and Sotiropoulos, 2016; An-  
228 dersson et al., 2016) before fitting the diffusion tensors (Basser et al., 1994)  
229 on the b=1000 images (50 directions) with FSL’s DTIFIT<sup>1</sup>. This standard  
230 preprocessing pipeline was extended for the template construction pipeline.  
231 The binary brain masks in individual dMRI spaces were slightly reduced in

---

<sup>1</sup>Only the b=1000 shell is used for tensor fitting due to the violation of the Gaussian diffusion assumption underlying the diffusion tensor model at higher b-values (or when combining b-values), which would require Kurtosis for correct modelling.



size by smoothing with an un-normalised mean filter (3 x 3 x 3 kernel size, to create a smooth transition between brain and background), thresholding at 0.9, and eroding by one voxel. These masks were used to reduce the impact of noisy DTI voxels at the border of, and outside the brain during nonlinear registrations. Bias fields created with FSL’s FAST were transformed from each individual’s reference spaces to their non-defaced T1 and T2-FLAIR native spaces and used to correct for intensity inhomogeneity in the brain. High intensity values of the scalp in T1 images were smoothly-clamped with a custom function (see Appendix A) to avoid negative effects on the nonlinear registrations during template construction. We did not perform any resampling with the transformations estimated between modalities to avoid the accumulation of interpolation errors.

In the rest of this manuscript, we will use the following notation: the set of  $N$  subjects, where each individual  $n$  has data from three modalities  $M$  is defined as

$$R = \left\{ R_n = \{ R_n^m \}_{m \in \{T1, T2-FLAIR, DTI\}} \right\}_{n=1 \dots N} . \quad (1)$$

### 2.3. Template construction

Our multimodal template and age-dependent template construction pipeline consists of three main parts.

First, an unbiased affine template was constructed by correcting for global (affine) misalignment between individuals (Fig. 1A, Section 2.3.1), which was then rigidly aligned to MNI space (Grabner et al., 2006).

Second, this affine template was used to initialise a nonlinear, hierarchical, multi-resolution templating approach (Fonov et al., 2011), which iterated through registration, unbiasing, transformation, and averaging steps (Fig. 1B, Section 2.3.2). The final nonlinear template, the OMM-1, represents the average shape and intensity of the 240 individuals on which it is based.

Third, the OMM-1 was used as a template to spatially normalise 37,330 individuals from the UKB imaging cohort, resulting in one deformation field for each subject (Fig. 1C, Section 2.3.3). A Gaussian process (GP) was used to model the morphological differences captured by these deformation fields as a function of age. After fitting the model, a mean deformation field was predicted for every year of age between 45-81, and used to generate the corresponding age-dependent template (ADT). In the following sections we will discuss each of these steps in detail.

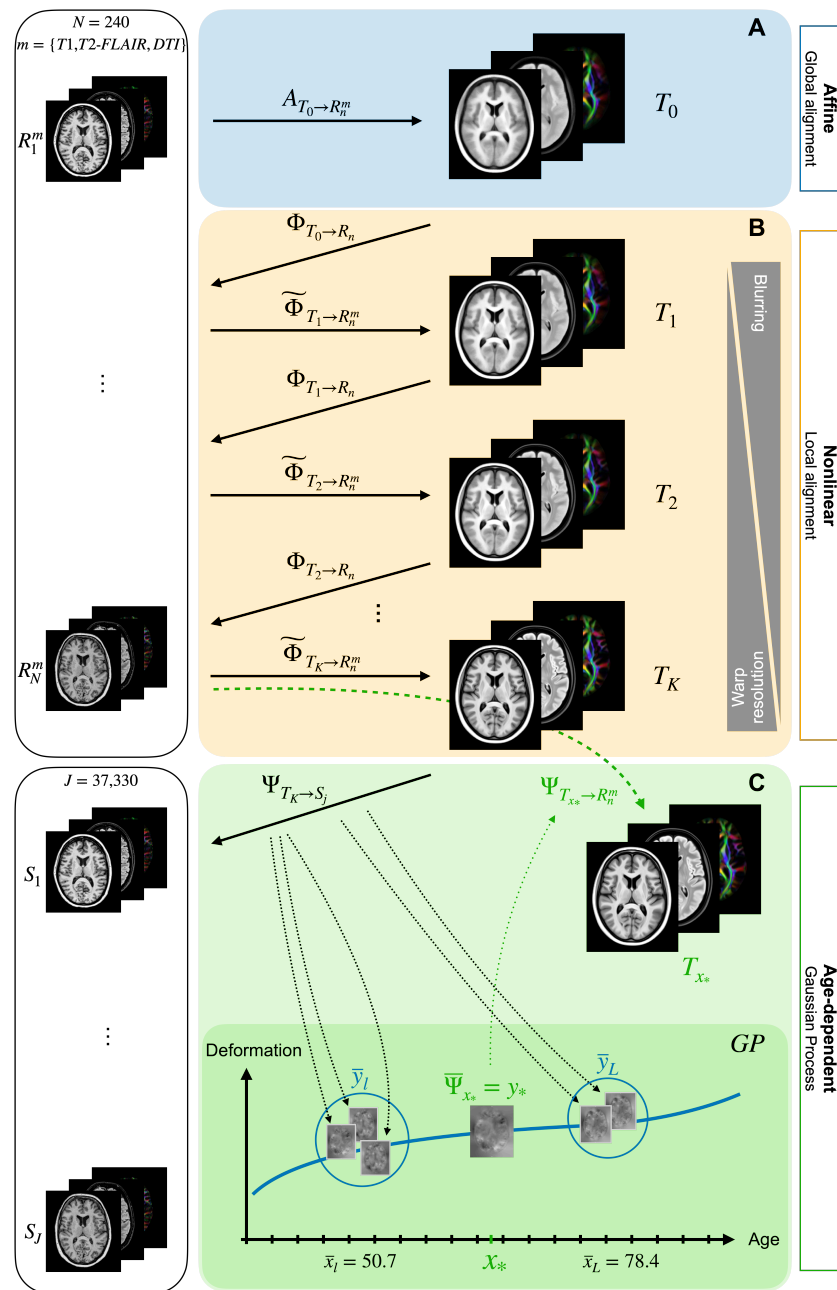


Figure 1: Scalar and tensor-based modalities from 240 UKB individuals were used to construct (A) the unbiased affine template by correcting for global misalignment and (B) the final nonlinear OMM-1 template by iterating through the hierarchical optimization approach. (C) Age-dependent templates were derived from the predictions of a Gaussian process model trained on ages, and warps to OMM-1 space from 37,330 UKB individuals.

### 2.3.1. Affine template construction

An initial affine template was constructed from the preprocessed, brain-extracted  $R^{T1}$  images. One subject was randomly selected as a reference space and the remaining subjects were affinely registered to this reference with 12 degrees of freedom (DOF). To avoid the introduction of a bias towards the brain geometry of the reference individual, the transformation from each individual's space to the mid-space of all subjects was calculated using FSL's *midtrans* function. We performed preliminary tests with different subjects as an initial reference, and confirmed that we did not find any difference in the final results. Brain-extracted  $R^{T1}$  images were resampled into this unbiased space by applying the corresponding transformations to them. The first affine template was created by calculating voxel-wise the median over the resampled images, which provides a sharper group average compared with taking the mean at this early stage (Fig. 1A) and was found to improve registration performance in the subsequent iterations.

This initial template was rigidly (6 DOF) aligned to the space of the asymmetric version of the nonlinear 6th gen. ICBM 152 template (MNI 152) (Grabner et al., 2006) included in FSL, to maximise similarity between the spaces while avoiding shearing and scaling effects. The final set of linear transformations  $A_n = \{A_{T_0 \rightarrow R_n^{T1}}, A_{T_0 \rightarrow R_n^{T2-FLAIR}}, A_{T_0 \rightarrow R_n^{DTI}}\}$  for each individual were created by concatenating the corresponding rigid transformation from each modality's native space to T1 reference space, the affine transformation from the T1 reference space to the unbiased template space, and the rigid transformation to the space of the new template  $T_0$ . Non-defaced images were transformed from their native spaces to  $T_0$  by applying the corresponding concatenated transformations using spline interpolation. Additionally, T1 and DTI brain masks were resampled with the same transformations using trilinear interpolation.

The voxel-wise median of the resampled images for each of the modalities provided the final affine template with three volumes  $T_0 = \{T_0^{T1}, T_0^{T2-FLAIR}, T_0^{DTI}\}$ . Similarly, mean T1 and DTI brain masks were created from the transformed masks in template space.

### 2.3.2. Nonlinear template construction

As we have stated previously, it is desirable that a template not be biased towards any particular individual (or subset of individuals) in the population from which it is constructed. By biased, we mean that the template should not, on average, be more like any one subject than any other. There are

two ways in which a template might appear more similar to an individual: in its shape, and in its appearance - where appearance refers to the voxel intensities. Consequently, a template may exhibit either a shape bias, an appearance bias, or both, unless care is taken to avoid this.

Shape (spatial) bias can be avoided by ensuring that, following registration to the template, the average displacement from the template to each individual is minimised across the population. Appearance (intensity) bias can be avoided by ensuring that, following registration to the template, the average image dissimilarity metric used to drive the registration is minimised across the population. Dissimilarity metrics commonly used by registrations methods include mean squared difference, cross-correlation and mutual information.

Fonov et al. (2011) formalised this concept as finding the template  $T$  that simultaneously minimises Equations 2 and 3, which address spatial and intensity bias respectively. The former (Eq. 2) minimises the magnitude of the nonlinear deformations  $\Phi_{T \rightarrow R_n}$  required to warp the template  $T$  to each subject  $R_n$ , and the latter (Eq. 3) minimises the mean squared intensity difference between the template  $T$  and each warped subject  $\Phi_{T \rightarrow R_n}(T)$ . Note that Equation 3 is specific to our case where MMORF optimises an image dissimilarity metric that is a version of the sum of squared-differences, and would differ if, for example, cross-correlation was used instead.

$$\arg \min_T \left[ \sum_{n=1}^N |\Phi_{T \rightarrow R_n}|^2 \right] \quad (2)$$

$$\arg \min_T \left[ \sum_{n=1}^N (T - \Phi_{T \rightarrow R_n}(T))^2 \right] \quad (3)$$

In practice, these two steps are interleaved at each of multiple iterations. In iteration  $k$ , Equation 2 is minimised by “undoing” (inverting and applying) the average across all nonlinear deformations  $\Phi_{T_k \rightarrow R_n}$  required to warp the template  $T_k$  to each subject  $R_n$ , and Equation 3 is minimised by simple voxelwise averaging of the warped intensities  $\Phi_{T_k \rightarrow R_n}(T_k)$  across all subjects.

Given this understanding of unbiasing at each stage/iteration of the template construction pipeline, the optimal, unbiased, nonlinear OMM-1 template  $T$  was constructed by iterating over the following three steps (Fig. 1B).

1. Deformation fields  $\Phi_{T_k \rightarrow R_n}$  are estimated by nonlinearly registering each individual to the template from the previous iteration  $T_{k-1}$ , with

the affine template  $T_0$  being used as reference space for the first iteration. Registrations were performed with MMORF (Lange et al., 2020a) and were informed with both scalar modalities (T1, T2-FLAIR) and the tensor-valued modality (DTI) from individual and reference space. MMORF optimizes the following total cost function

$$\mathcal{C}_{TOT} = \lambda_{T1}\mathcal{C}_{T1} + \lambda_{T2-FLAIR}\mathcal{C}_{T2-FLAIR} + \lambda_{DTI}\mathcal{C}_{DTI} + \lambda_{REG}\mathcal{C}_{REG} \quad (4)$$

with all modalities contributing equally in the optimisation procedure (*i.e.*,  $\lambda_{T1} = \lambda_{T2-FLAIR} = \lambda_{DTI} = 1$ ). The mean squared error is calculated between scalar images and the mean squared Frobenius norm is used as a cost function for tensors.

Registrations were initialised with the corresponding linear transformations  $A_n$  estimated during the construction of the affine template. Note that non-brain-extracted individual images and templates were used for the scalar channels, which poses additional challenges. Inclusion of the skull can negatively affect registration quality in nearby cortical regions, and the face and neck have larger anatomical and positional variability compared to the brain. To reduce the potential impact of extracranial tissue on the deformations close to the brain, and improve registration quality in the face and neck, different levels of relative regularisation were imposed on intra- and extracranial regions. This was achieved through modulation of the T1 brain mask in template space. Larger weights were given to regions inside the brain, *i.e.*, reducing the relative level of regularisation to allow for more aggressive deformations, and smaller weights were given to regions outside the brain, *i.e.*, increasing the relative level of regularisation to constrain the deformations. The intra-to-extracranial weight ratio was approximately 8-to-1. Similarly, a weighted average DTI mask with smoothly decreasing weights at the edge of brain tissue in reference space and eroded DTI brain masks in individuals' native spaces were used to reduce the potential negative effect of poor/noisy tensor fitting around brain boundaries that are often seen in DTI. No masks were required for T2-FLAIR since tissue outside the brain already appears dark and does not strongly drive the registration relative to brain tissue.

2. The average deformation field  $\bar{\Phi}_{T_k}$  was calculated with

$$\bar{\Phi}_{T_k} = \frac{1}{N} \sum_{i=1}^N \Phi_{T_k \rightarrow R_n} \quad (5)$$

and used to spatially unbias the template. This unbiasing step was performed by composing its inverse  $\bar{\Phi}_{T_k}^{-1}$  with each individual deformation field and the corresponding rigid and affine transformations:

$$\tilde{\Phi}_{T_k \rightarrow R_n^m} = \bar{\Phi}_{T_k}^{-1} \circ \Phi_{T_k \rightarrow R_n^m} \circ A_{T_0 \rightarrow R_n^m} \quad (6)$$

Individuals' modalities in their respective native spaces were resampled to the new unbiased template space in one step by applying  $\tilde{\Phi}_{T_k \rightarrow R_n^m}$  with spline interpolation. Tensors were reoriented with FSL's *vecreg* tool, which uses the preservation of principal directions algorithm (Alexander et al., 2001). T1 and DTI brain masks were resampled with trilinear interpolation using the same transformations.

3. New mean masks and a new template  $T_k$  were created in unbiased space by taking the average over the resampled images for each modality. This new unbiased template with its three volumes served as a reference space in the next iteration  $k + 1$ .

We performed a total of  $K = 18$  iterations, allowing for coarse to fine improvements, with three iterations at each of six hierarchical levels. A large grid spacing of 32 mm and a blurring kernel of 8 mm FWHM was used for the MMORF registrations (step 1) in the three iterations at the first hierarchical level. These parameter values were halved for each level, down to 1 mm and 0.25 mm (respectively) at the last hierarchical level. An overview of the MMORF registration parameters can be found in Appendix B.

### 2.3.3. Age-dependent template construction

UKB individuals  $S = \{S_j\}_{j=1 \dots J}$  ( $J = 37,330$ ) from the 45-82 year age range were first affinely and then nonlinearly registered to OMM-1. Similar to the previous nonlinear registrations, we used MMORF with all three MRI modalities (Fig. 1C) and the registration parameters in Appendix B. The estimated set of deformation fields  $\Psi = \{\Psi_{T_K \rightarrow S_j}\}_{j=1 \dots J}$  in OMM-1 space was used to model the average change in morphology with age using Gaussian process (GP) regression (Rasmussen and Williams, 2005), i.e. the objective can be stated as finding the function that best models the change in brain morphology as captured by the deformation fields given the subjects' ages  $\mathbf{x}$ . The trained GP allowed the prediction of a mean output deformation field  $\tilde{\Psi}$  in OMM-1 space for any (observed or unobserved) age  $x_*$ . Note that here we did not model differences in overall brain size and, consequently, the

397 nonlinear deformation fields without their affine transformation components  
398 were used.

399 Gaussian processes generalize the concept of Gaussian probability distri-  
400 butions from stochastic variables to stochastic functions, and can be written  
401 as  $\mathbf{f}(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$  or  $y = f(x) + \epsilon$  with additive independent  
402 Gaussian noise  $\epsilon$ . The GP is specified by its prior mean function  $m(\mathbf{x})$ ,  
403 which is usually set to zero, and covariance function  $k(\mathbf{x}, \mathbf{x}')$ , whose form  
404 has to be manually chosen. Conceptually these can be seen as continuous  
405 generalisations of the mean vector and covariance matrix used to describe  
406 multivariate normal distributions of random variables. The joint distribu-  
407 tion of the observed training input and output pair  $(\mathbf{x}, \mathbf{y})$  and unobserved  
408 pair  $(x_*, y_*)$  can be written as

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \mathbf{K}(\mathbf{x}, \mathbf{x}) + \sigma_n^2 \mathbf{I} & \mathbf{k}(\mathbf{x}, x_*) \\ \mathbf{k}(x_*, \mathbf{x}) & k(x_*, x_*) \end{bmatrix}\right) \quad (7)$$

409 where  $\mathbf{I}$  is the identity matrix and  $\sigma_n$  describes the variance of the noise,  
410 with larger  $\sigma_n$  resulting in a smoother function.  $\mathbf{K}(\mathbf{x}, \mathbf{x})$  is a  $J \times J$  matrix of  
411 covariances between all training inputs,  $\mathbf{k}(\mathbf{x}, x_*)$  and  $\mathbf{k}(x_*, \mathbf{x})$  are vectors of  
412 covariances between training and query inputs, and  $k(x_*, x_*)$  is the variance  
413 of the query input.

414 As will become more apparent from Equations 10-11, the calculation of  
415  $\mathbf{K}(\mathbf{x}, \mathbf{x}) + \sigma_n^2 \mathbf{I}$  becomes increasingly computationally challenging with larger  
416  $J$ . To reduce the computational burden, the training input and output  
417 data were stratified along the age axis into half-yearly bins. Additionally,  
418 each bin was split into two sub-bins, where each individual within an age  
419 bin was randomly assigned to one of the corresponding sub-bins. This in-  
420 troduction of some variability within each bin was done to better condi-  
421 tion the estimation of the hyperparameters by making the estimates less  
422 correlated. This aggregation considerably reduced the size of the training  
423 dataset from the initial  $J = 37,330$  data points to  $L = 148$  (2 sub-bins  $\times$   
424 74 half-yearly mean age bins  $\bar{\mathbf{x}}$ ) and corresponding mean deformation fields  
425  $\bar{\mathbf{y}}$  as respective input and output for ages 45-81 years.

426 The noise term in Equation 7 assumes that the noise is constant for every  
427 data point. This assumption would hold if every sub-bin was assigned the  
428 same number of individuals. However, when stratifying over age, the sub-bin  
429 averages were taken over different numbers of individuals because of the non-  
430 uniform age distribution in the initial dataset—with fewer individuals for the



youngest and oldest age groups. Assuming noise to be constant for all sub-bins would introduce a bias. To account for this non-uniformity, the identity matrix  $\mathbf{I}$  in Equation 7 was replaced with a weight matrix  $\mathbf{W}$  containing  $\frac{1}{p_l}$  in the diagonal, where  $p_l$  is the number of individuals assigned to age bin  $\ell$ . This down-weights the noise variance for, and increases the confidence in, bins pooled from a larger number of individuals, and vice versa. The joint distribution from Equation 7 becomes

$$\begin{bmatrix} \bar{\mathbf{y}} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \mathbf{K}(\bar{\mathbf{x}}, \bar{\mathbf{x}}) + \sigma_n^2 \mathbf{W} & \mathbf{k}(\bar{\mathbf{x}}, x_*) \\ \mathbf{k}(x_*, \bar{\mathbf{x}}) & k(x_*, x_*) \end{bmatrix}\right) \quad (8)$$

The choice of covariance function and its associated hyperparameters is important since it defines the properties of the functions generated during inference. Here, a squared exponential kernel was used, which has strong smoothness assumptions, and is therefore in line with the expected smooth changes in the brain with age. The kernel for calculating the covariance between two ages  $x$  and  $x'$  can be written as

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{(x - x')^2}{2\ell^2}\right) \quad (9)$$

where  $\sigma_f$  is a scaling factor, and  $\ell$  is the length scale. Intuitively, a larger  $\sigma_f$  increases both the magnitude and the variability of the fitted function, and a larger length scale  $\ell$  increases the dispersion and covariance between more distant ages, leading to a smoother function, which is less influenced by noise and overfitting.

The hyperparameters  $\sigma_f$ ,  $\sigma_n$  and  $\ell$  can be estimated by maximizing the marginal likelihood given by

$$\log p(\bar{\mathbf{y}}|\sigma_n, \sigma_f, \ell) = -\frac{1}{2}\bar{\mathbf{y}}^T(\mathbf{K} + \sigma_n^2 \mathbf{W})^{-1}\bar{\mathbf{y}} - \frac{1}{2}\log|\mathbf{K} + \sigma_n^2 \mathbf{W}| \quad (10)$$

where  $\bar{\mathbf{y}}$  is a matrix containing the vectorized deformation fields in the rows, and  $\mathbf{K}$  is the covariance matrix where element  $K_{ij}$  is the covariance between two ages  $\bar{x}_i$  and  $\bar{x}_j$ . The Nelder-Mead simplex method has shown robust estimates when minimizing the negated function, which was optimized over all voxels to estimate the set of hyperparameters.

Using the estimated hyperparameters, the predictive mean can be calculated with

$$y_* = k(x_*, \bar{\mathbf{x}})(\mathbf{K}(\bar{\mathbf{x}}, \bar{\mathbf{x}}) + \sigma_n^2 \mathbf{W})^{-1}\bar{\mathbf{y}}. \quad (11)$$

where the output  $y_*$  is the mean deformation field  $\bar{\Psi}_{x_*}$  for the corresponding age  $x_*$ . One deformation field was predicted for each year in the age range 45-81. The inverse of this deformation was concatenated with the initial 240 subjects' linear and nonlinear transformations such that

$$\Psi_{T_{x_*} \rightarrow R_n^m} = \bar{\Psi}_{x_*}^{-1} \circ \tilde{\Phi}_{T_k \rightarrow R_n^m} \quad (12)$$

before resampling the corresponding modalities to age-specific template space in one step. Averaging over each of the resampled modalities provided the corresponding ADT  $T_{x_*}$ .

#### 2.4. Validation and applications

Convergence of the OMM-1 template construction process was assessed with three metrics including the root mean squared (RMS) difference, the root mean squared percentage (RMSP) difference, and Pearson's correlation (PC) between consecutive iterations of average warps and T1, T2-FLAIR and DTI volumes. The Frobenius norm (FN) between consecutive iterations was additionally calculated for DTI volumes. Metrics based on the average warps show the improvement with respect to the first objective function of finding the average shape template, while metrics based on the T1, T2-FLAIR and DTI volumes show changes with respect to the second objective function of finding the average intensity template.

Our ADTs were visually assessed and the prediction-derived 81-year ADT was compared to two directly-estimated templates with the same age. The first of these two templates was constructed by registering all 101 UKB individuals in the 80-81 year range (mean age of 80.44 years) to OMM-1 space. The estimated deformation fields were spatially unbiased and applied to the corresponding images, which allowed the construction of a directly-estimated ADT-81. As a second template for comparison we used the existing older adult MIITRA template (mean age of 80.56 years) (Wu et al., 2022).

Age-related differences were assessed with the distortion given by average Jacobian determinant maps. One map was derived for each corresponding GP-predicted deformation field. Volume differences in the subcortical structures were quantified by summing over corresponding ROIs in the Jacobian determinant maps. The ROI masks were created by warping FSL FIRST (Patenaude et al., 2011) segmentation masks from individuals to OMM-1 space before averaging and binarizing them with a threshold of 0.5.

Finally, we investigated whether there is an advantage in registering individuals to the standard OMM-1 template via an age-matched ADT over

493 registering them directly to the standard OMM-1. These results were also  
 494 compared to registering directly to the MNI 152 template. 148 held-out UKB  
 495 test subjects (50% female) were uniformly sampled from the 45-81 age range  
 496 and registered to (1) the OMM-1 directly, (2) the GP-derived ADT corre-  
 497 sponding to the individual’s age, and (3) the MNI 152 directly. FSL FLIRT  
 498 (Jenkinson and Smith, 2001) was used for all affine registrations followed  
 499 by MMORF (Lange et al., 2020a) for all nonlinear registrations. Registra-  
 500 tion parameters were identical for (1) and (2) with T1, T2-FLAIR and DTI  
 501 driving the nonlinear registrations, and T1-only driving the nonlinear reg-  
 502 istration for (3). DKT atlas ROIs (Desikan et al., 2006) had been created  
 503 for each individual with FreeSurfer (Fischl et al., 2004) as part of the UKB  
 504 preprocessing pipeline. These ROIs were transformed from individuals’ na-  
 505 tive spaces to generic template space using the direct warp to OMM-1 as  
 506 estimated in (1), the composed warp from individual to ADT and from ADT  
 507 to OMM-1 for (2), and the direct warp to MNI 152 in (3) using trilinear  
 508 interpolation. The Dice similarity coefficient for each transformed binarized  
 509 ROI was calculated for every possible pairing of subjects for each of the three  
 510 approaches. We repeated the same tests with 100 out-of-sample individuals  
 511 and their corresponding Destrieux atlas Freesurfer ROIs (Destrieux et al.,  
 512 2010) from the Human Connectome Project (HCP) (Van Essen et al., 2013)  
 513 using T1-only, T1-only with brain extraction, and T1 and DTI modalities  
 514 for registrations in (1) and (2). Due to the younger age range in the HCP,  
 515 we used the youngest 45-year ADT as a reference space for all registrations  
 516 in (2).

## 517 3. Results

### 518 3.1. OMM-1

519 The template after the last iteration of each hierarchical level is shown  
 520 in Figure 2A. Qualitatively, a gradual increase in contrast and sharpness is  
 521 noticeable in all three modalities with more rapid change over the early it-  
 522 erations and less change in the later iterations. Quantitative measurements  
 523 of convergence towards the average intensity and average shape are shown  
 524 in Figure 2B. The change in intensities and the change in average warps  
 525 between consecutive iterations show a similar pattern with all metrics. A  
 526 sharp increase in difference occurs after switching to a finer registration level  
 527 before stabilising in the following iterations at the same level. These can be  
 528 seen as sharp spikes in the differences measured with RMS, RMSP and FN,

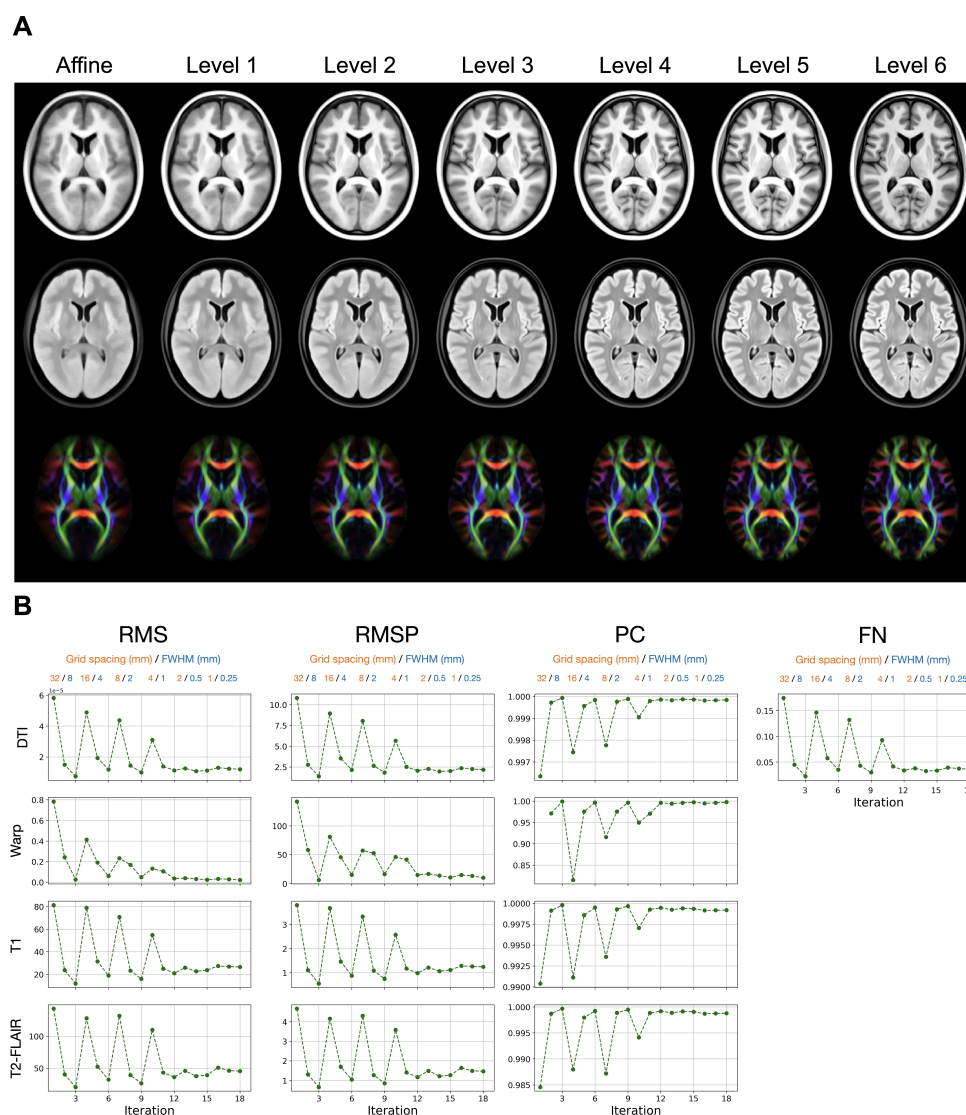


Figure 2: (A) Rows show the three template modalities: T1, T2-FLAIR and DTI (visualised using a principal diffusion direction colour-coded FA map where AP=Green, LR=Red, IS=Blue). The improvement in contrast and alignment after the final iteration of each hierarchical level is noticeable for each of the three modalities as the size of the blurring kernel and the grid spacing are reduced from coarse to fine. (B) Convergence measured with several metrics — root mean squared (RMS), root mean squared percentage (RMSP), Pearson correlation (PC) and Frobenius norm (FN) — shows the difference between template intensities for each modality and between average warps of consecutive iterations. A sharp difference can be observed with all metrics after switching to a new hierarchical level, which is followed by an improvement in the following iterations at the same level. These fluctuations stabilise towards the last iterations indicating convergence.

529 and sharp drops with PC. These fluctuations become smaller in the later  
530 iterations, which indicates convergence.

531

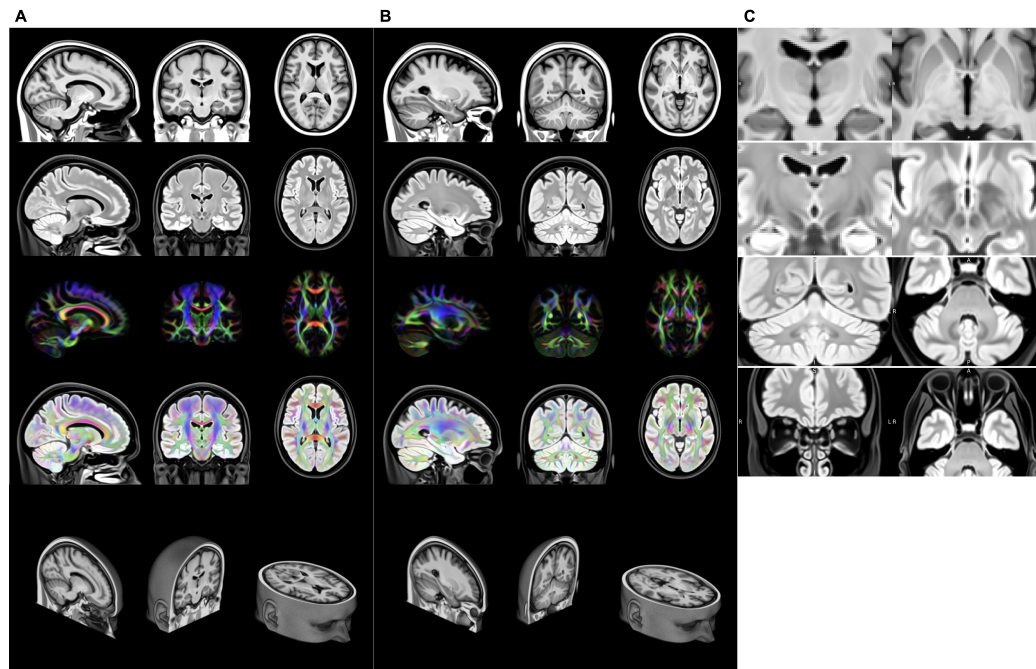


Figure 3: Two different sets of slices (A) and (B) through the modalities of the final OMM-1 template: T1, T2-FLAIR, DTI (visualised using a principal diffusion direction colour-coded FA map where AP=Green, LR=Red, IS=Blue), DTI overlaid on T2-FLAIR, and the corresponding slice in the 3D T1 volume. Scalar T1 and T2-FLAIR volumes exhibit very good contrast. The DTI volume shows excellent sharpness and orientational consistency. Alignment between scalar and tensor modalities is exceptional as can be seen in the overlay of DTI and T2-FLAIR. Facial features such as the ears, nose and eyes show a high level of detail. Coronal and axial slices show the right hemisphere on the left and vice versa. Zoomed-in views of four ROIs in (C) highlight the excellent alignment across all 240 participants. The medial medullary lamina - separating internal from external globus pallidus - is clearly visible in the axial view of the T1 volume. On the T2-FLAIR volume a clear separation of the subthalamic nucleus and substantia nigra can be seen in the coronal view, as well as the dentate nucleus in the cerebellum and the olfactory bulbs are clearly visible.

532 Two different sets of slices of the final OMM-1 template are shown in  
533 Figure 3. The T1 and T2-FLAIR volumes are visually sharp with excel-  
534 lent contrast. In the DTI volume, white matter fibre tracts appear clear

535 and bright, and both isotropic and anisotropic regions are very well aligned.  
 536 This is especially noticeable in the cerebellum and the brainstem tracts.  
 537 All volumes, scalar, and DTI, show very good alignment, as judged by eye.  
 538 This indicates that there are no systematic, rigid registration errors between  
 539 modalities. Our strategy estimates a single warp from and applies it to all  
 540 modalities in individual spaces, which ensures correct cross-modality align-  
 541 ment in template space (assuming that modalities in individual spaces are  
 542 correctly aligned). Extracerebral structures, such as the sinuses are clearly  
 543 visible in the T2-FLAIR volume, as well as the olfactory tract and the ol-  
 544 factory bulbs. Subcortical GM structures are sharply defined, with comple-  
 545 mentary contrast provided by the T1 and T2-FLAIR modalities. (Fig. 3B).  
 546

### 547 *3.2. Age-dependent templates*

548 The prediction-based ADTs for selected ages in steps of five years are  
 549 illustrated in Figure 4A. The scalar and tensor volumes show consistently  
 550 high quality of both contrast and crispness, and good alignment between  
 551 modalities. Expected age-related differences, such as increases in ventricle  
 552 sizes and sulcal widening, are visible in the templates, while the overall shape  
 553 of anatomical brain structures and the folding pattern remains stable across  
 554 all ages. This is confirmed when looking at the Jacobian determinant maps,  
 555 where the largest differences occur in the ventricles. Cortical GM thinning  
 556 is noticeable in the insular cortex and the inferior frontal gyrus.

557 Age-related differences in volume can also be directly derived from the  
 558 templates as exemplified in subcortical structures in Figure 4B, where an age-  
 559 related loss in volume is noticeable for all structures. A small unexpected  
 560 increase in the hippocampus and amygdala volumes was found before age 52.  
 561 We further investigated whether this increase is caused by an artefact of the  
 562 modelling or the data by comparing measurements from the GP-estimated  
 563 48 year Jacobian determinant map with those derived from the average Ja-  
 564 cobian determinant map of all 45-50 year old individuals. We found that our  
 565 model underestimated the relative volume by approximately 1.5% for the  
 566 hippocampus and 1.1 % for the amygdala in this age range, which will be  
 567 partly caused by the smaller number of individuals in this age range, leading  
 568 to larger uncertainty in the GP estimates. A similar pattern in the hip-  
 569 pocampus was observed outside the main data range in (Janahi et al., 2022),  
 570 where GPs were directly applied to volume measurements. However, for the  
 571 hippocampus, similar trajectories in volume difference due to the data have



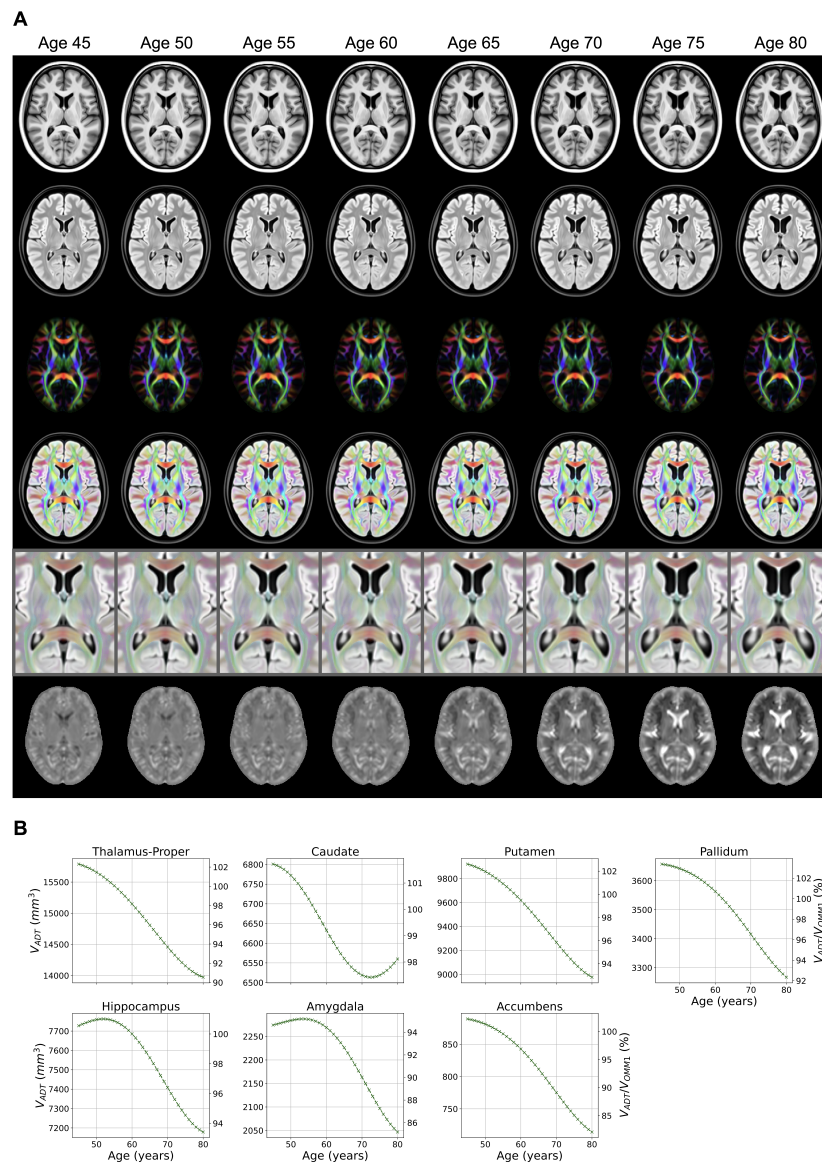


Figure 4: (A) The increase in ventricle size and sulcal widening can be observed across all modalities in this subset of age-dependent templates (right hemisphere on the left). The expansion seen in ventricles and sulci and the thinning of GM are visible in the corresponding log-Jacobian maps with dark values indicating contraction and bright values expansion of the main OMM-1 template. The  $\log|J|$  intensity range is set to  $\log(0.5) - \log(2.0)$ . (B) Volume measurements directly derived from the templates show characteristic age-related differences in subcortical structures. Absolute volumes are given on the left y-axis and percent volume in comparison to OMM-1 on the right y-axis.



572 been reported in the majority subgroup of the study population in (Fraser  
573 et al., 2021).

574 The increase in volume seen in the caudate from 72 years is an artefact  
575 caused by the segmentation masks, where the increasingly large ventricles  
576 start bleeding into the caudate ROI at older ages.

577  
578 Figure 5 visually illustrates the age-related differences between the OMM-  
579 1 and the 81 year ADTs. It also highlights the high similarity between the  
580 directly-estimated ADT, the GP-estimated ADT, and the MIITRA template  
581 (Wu et al., 2022). Note that the GP-estimated ADT was derived by trans-  
582 forming the original 240 OMM-1 individuals from the 50-55 year age range  
583 through a GP-predicted warp, while the directly-estimated ADT was derived  
584 by transforming and averaging individuals with a mean age of 80.44 years.  
585 MIITRA was created through an iterative process from a cohort with a mean  
586 age of 80.56 years. While the general shape of all three templates is very  
587 similar, there are some differences. The directly-estimated ADT has slightly  
588 larger ventricles than both the GP-estimated ADT and MIITRA. It is also  
589 slightly blurrier in cortical areas compared to the GP-estimated ADT. Our  
590 GP-estimated ADT shows improved sharpness in subcortical areas, while  
591 MIITRA shows improved sharpness in cortical areas such as the occipital  
592 lobe. The striped appearance of the striatum is visible in the MIITRA, but  
593 not in the OMM-1 template.

594 The MNI 152 is provided as a commonly-used reference template and  
595 highlights the large difference in scale compared to the OMM-1, ADT and  
596 MIITRA templates. The cerebral volume of the OMM-1 is approximately  
597 1,419,081 mm<sup>3</sup>, which is more than 465 ml less than the 1,884,594 mm<sup>3</sup> of  
598 the MNI 152 and much more similar to the median and mean volumes of  
599 1,433,335 mm<sup>3</sup> and 1,440,417 mm<sup>3</sup>, respectively, in the entire UKB.

600

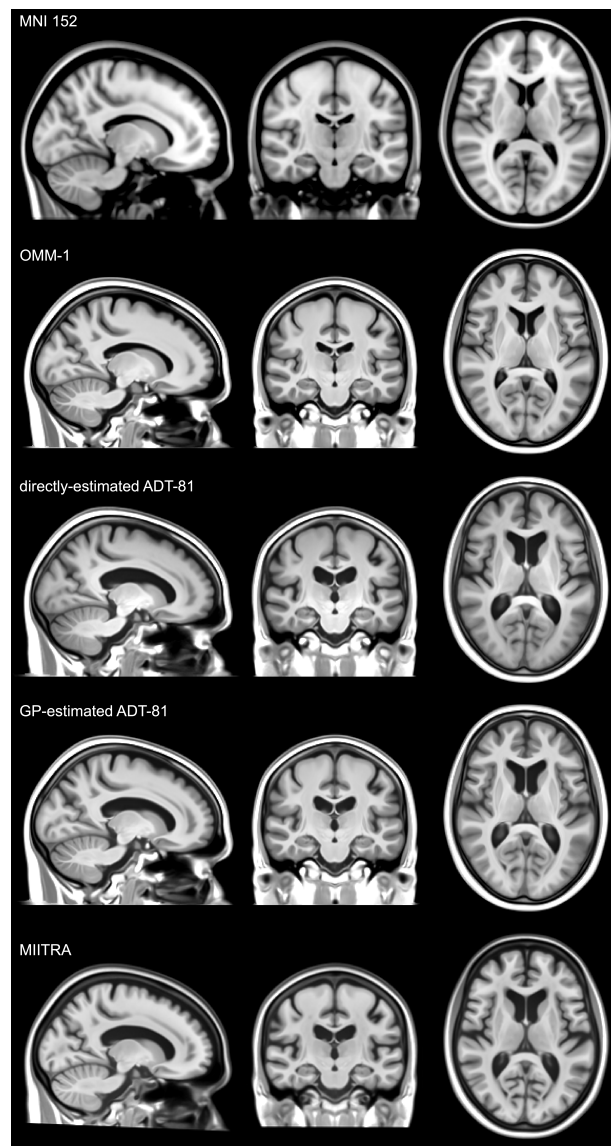


Figure 5: Visual comparison of the T1 volumes of the MNI 152, OMM-1, 81-year ADT directly estimated from 81 year old UKB individuals in OMM-1 space, 81-year ADT estimated with the GP, and MIITRA template (rigidly aligned) (Wu et al., 2022). The MNI 152 is provided as a reference and highlights the large global scale differences in MNI templates. Age-related difference between the OMM-1 and the two versions of the ADT-81 are well noticeable. The GP-estimated ADT-81 shows high similarity in overall shape and appearance, and age-related features such as ventricle size and sulcal widening, with both the directly-estimated ADT-81 and MIITRA. The GP-estimated ADT was derived from the GP-transformed, original 240 OMM-1 individuals (50-55 year age range). The directly-estimated ADT and MIITRA were derived from older adults with mean ages of 80.44 years and 80.56 years respectively.

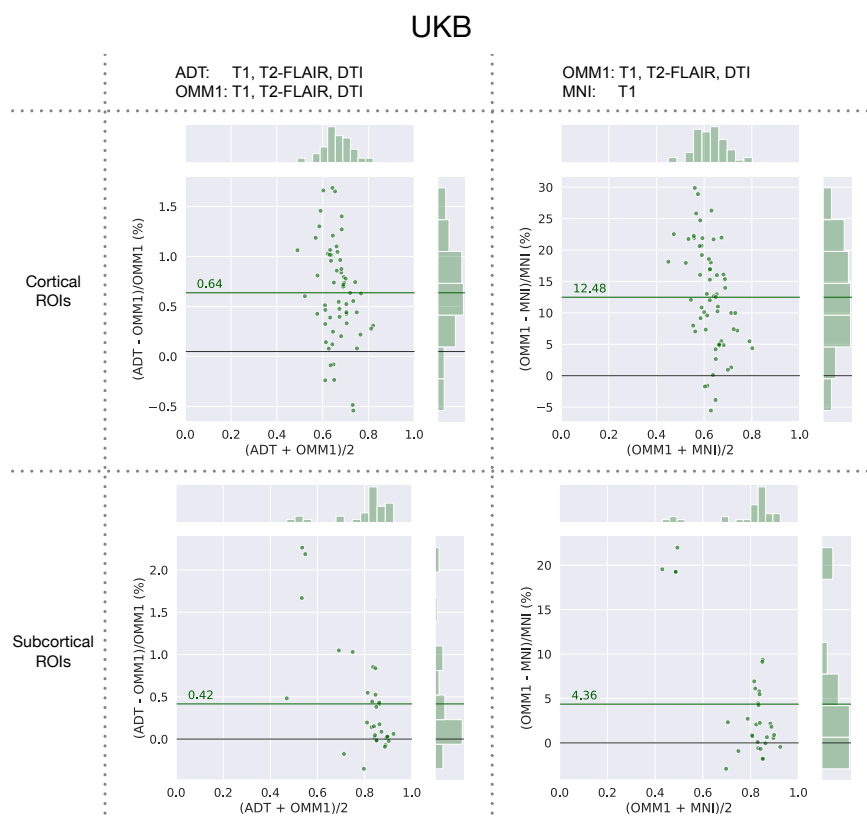


Figure 6: UK Biobank: relative differences between Dice coefficients obtained by warping the same masks with transformations estimated when directly registering to OMM-1 (OMM1), indirectly registering to OMM-1 via the individual’s corresponding age-dependent template (ADT), and registering to MNI 152. Each dot indicates one ROI and the green line shows the average percentage difference over all ROIs. On average ADTs slightly outperform OMM-1 and OMM-1 outperforms MNI by a larger margin.

### 3.3. Application to spatial normalisation

Relative differences in aggregated pairwise dice coefficients for each ROI on UKB data are shown in Figure 6. On average over all ROIs spatial alignment to the corresponding ADT slightly outperforms alignment to OMM-1 by 0.64% in cortical and 0.42% in subcortical ROIs. Alignment to OMM-1 outperforms MNI 152 by a larger margin of 12.48% in cortical and 4.36% in subcortical ROIs. ADTs compared to OMM-1, and OMM-1 compared to MNI 152 achieve significantly larger Dice overlaps in 49 and 51 out of 63 cortical ROIs respectively. ADTs compared to OMM-1, and OMM-1 compared

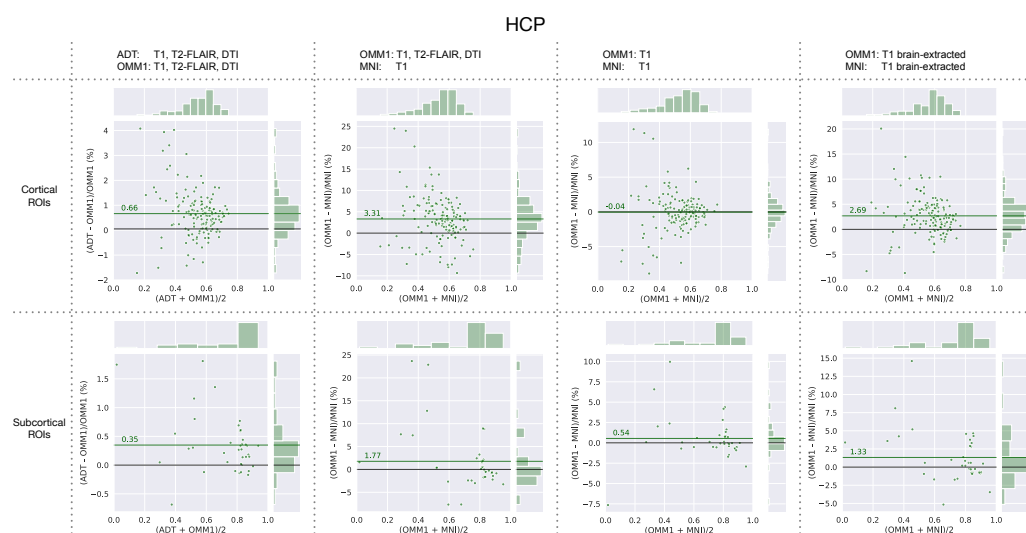


Figure 7: HCP: relative differences between Dice coefficients obtained by warping the same masks with transformations estimated when directly registering to OMM-1 (T1, DTI), indirectly registering to OMM-1 via the individual’s corresponding ADT, and registering to MNI 152. Additionally, registrations based on T1-only and brain-extraction T1-only were compared for OMM-1 and MNI 152. Each dot indicates one ROI and the green line shows the average percentage difference over all ROIs.

to MNI 152 achieve significantly larger Dice overlaps in 22 and 23 out of 32 subcortical ROIs respectively.

Similar results can be replicated in an out-of-sample (non-UKB) cohort from the HCP (Fig. 7) where the 45 year ADT outperforms OMM-1 on average by 0.66% in cortical ROIs and 0.35% in subcortical ROIs. Multimodal registration to OMM-1 outperforms single-modal T1 registration to MNI by 3.31% and 1.77% respectively. Using only T1 for the registration to OMM-1 performs equally well as registration to MNI for cortical ROIs and 0.54% better for subcortical ROIs. Using preprocessed, brain-extracted images and templates for T1-based registrations showed a large improvement of OMM-1 over MNI 152 for both cortical and subcortical ROIs.

## 4. Discussion

We have presented the construction of the OMM-1, a fully-unbiased, internally-consistent, multimodal template of the brain including parts of

the neck and face, averaging across 240 UKB individuals in the 50-55 year age range. GP regression was used to model deformation fields with age after registering 37,330 UKB individuals to OMM-1, and allowed the prediction of an average warp for each year of age in the corresponding 45-81 age range. These warp predictions were used to resample the original 240 OMM-1 subjects and create one multimodal ADT for each year of age. Test subjects from the UKB and HCP were registered to the OMM-1 directly, the ADT corresponding to the age of the individual, and MNI 152 to compare their performance in spatial normalisation tasks.

OMM-1 and ADTs consist of T1, T2-FLAIR and DTI volumes, and the same modalities were used to jointly inform the template construction process through the use of MMORF for all nonlinear registrations. The scalar volumes of all templates provide excellent contrast and exceptional anatomical detail, and the DTI volume appears sharp in isotropic and anisotropic regions. Our strategy inherently provides optimal cross-modality alignment between template volumes since all modalities are resampled through the same warps.

The OMM-1 is rigidly (6 DOF) aligned to MNI 152 space to provide a basic level of comparability, while avoiding scaling effects, to preserve the average brain size and shape of the UKB population. The scaling factor (Jacobian determinant of the affine transformation) between the OMM-1 and MNI 152 indicates an approximately 1.33 times larger cerebral volume of the MNI 152, which is also visually noticeable in Figure 5. Although this difference in scale might not have a large observable impact when used for spatial normalisation, it is not optimal as it will require additional unnecessary distortion for the majority of subjects. However, we recognise that considerable effort has been put into the development of atlases and analysis of studies in MNI space, and we therefore provide a deformation field that maps between MNI and OMM-1 space to allow these to be used with, or adapted for, our new template.

The GP model was used to create ADTs in steps of one year between ages 45 and 81, but allows for the construction of ADTs on a continuous scale within this age range. We did not test the prediction of templates outside this age range, since the GP model fit becomes increasingly uncertain when extrapolating.

Visual inspection (Figure 5) shows that our GP-estimated ADT-81 is highly similar in appearance to an ADT directly estimated from 81-year UKB individuals. It also has highly similar shape- and age-related features with

663 respect to ventricle size, cortical folding and global shape and size compared  
664 to the MIITRA template, which was directly constructed from an older adult  
665 cohort. The slightly larger ventricles of the directly-estimated ADT will likely  
666 be caused by the smaller age range (80.0-81.0 years) of the subjects used in  
667 its construction. This is in stark contrast to 65.2-94.9 years for MIITRA  
668 and the weighted contributions according to the GP hyperparameters for the  
669 GP-estimated ADT.

670 The GP-estimated ADT shows improved sharpness in subcortical areas,  
671 while MIITRA shows improved sharpness in cortical areas, especially in the  
672 occipital lobe. Improvements in the MIITRA template in these areas are  
673 likely due to their weighted averaging approach, where intensities more sim-  
674 ilar to the median intensity across subjects at a voxel location receive higher  
675 weights. Notably, the striped appearance of the striatum is visible in the  
676 MIITRA, but not in the OMM-1. We have identified two reasonable causes  
677 for this. The first is that the regularisation metric used by MMORF to  
678 generate OMM-1, compared to that used by ANTs/DR-TAMAS to gener-  
679 ate MIITRA, will more strongly penalise the deformations required to align  
680 the stripes across individuals. Since MMORF optimises the structural and  
681 DTI alignment simultaneously, the deformations required to align the stripes  
682 would negatively impact the alignment of the tensors in that region (partic-  
683 ularly in terms of orientation alignment). This is supported by experiments  
684 (not shown) where we generated templates using only the T1 channel, and  
685 in which the stripes were partly visible (but still not to the extent seen in  
686 MIITRA). The second is that there does not appear to be a clear biological in-  
687 dication on the consistency of the stripes (= pontes grisei caudatolentiformes  
688 alternating with white matter forming the internal capsule) in number or ex-  
689 act location across individuals. Consequently, we do not believe that this  
690 negatively affects the use of OMM-1 as a registration target, even for older  
691 subjects.

692 We found that using the GP approach over simpler methods (such as ker-  
693 nel regression or the direct estimation of a template for each year) produced  
694 ADTs where morphology not affected by ageing (*e.g.*, the folding pattern of  
695 the cortex) remained far more stable as a function of time. Our GP-based  
696 approach is also much more time-efficient compared to the repeated, direct  
697 construction of templates for specific ages. The most time-intensive tasks  
698 have to be performed only once, *i.e.*, the iterative construction of a template  
699 (OMM-1) and the training of the GP. The estimated hyperparameters can  
700 then be re-used for the prediction of new templates thereafter.

701 The loss of subcortical volume in Figure 4 is mostly in line with results  
 702 previously reported in the literature (Walhovd et al., 2005; Vinke et al., 2018;  
 703 Nobis et al., 2019; Wang et al., 2019). The increase in hippocampus and  
 704 amygdala volumes for the younger ages was unexpected and appears to be  
 705 related to the smaller number of individuals available in this age range. This  
 706 is similar to the results obtained by Janahi et al. (Janahi et al., 2022), where  
 707 a GP was applied to extracted volume measurements. In comparison to  
 708 direct measurements from the data, our model-derived ADTs underestimate  
 709 the volume by approximately 1.1 % to 1.5 % for this age range. However, it  
 710 should be noted that a similar increase in hippocampus volume for this age  
 711 range has previously been reported (Fraser et al., 2021).

712 The increase in caudate volume seen from age 72 was caused by the  
 713 increasing size of the ventricles at older ages bleeding into the caudate ROI. In  
 714 addition to absolute estimated volumes, we show relative volumes normalized  
 715 by the volumes estimated in generic OMM-1 space. These ratios change at  
 716 different rates for different ROIs as would be expected. The volume change  
 717 of the amygdala is a slight outlier in that it does not reach 100%. The  
 718 reason for both of these deviations is the higher warp resolution of 1 mm  
 719 used for the registrations in the construction of the OMM-1, compared to  
 720 2 mm for the GP-estimated ADTs. Increasing the warp resolution to 1  
 721 mm would produce more fine-grained average warps that might mitigate this  
 722 effect. However, given the increasingly large size of the dataset and the  
 723 corresponding number of required nonlinear registrations, 2 mm was found  
 724 to be a reasonable compromise, and is comparable to, or better than, the  
 725 standard settings of other registration methods.

726 The presence of pathologies such as white matter hyperintensities or mi-  
 727 crobleeds at older ages have an impact on registration and subsequently on  
 728 segmentation accuracy in ADTs. Images of individuals in the 50-55 year age  
 729 range provide enhanced tissue contrast, more detail in anatomical structures,  
 730 and less pathologies than images from older individuals. This enhanced qual-  
 731 ity was the main motivation behind choosing 240 younger subjects for the  
 732 construction of the OMM-1, and is an advantage when using the templates  
 733 for spatial normalization, which is generally the main use case for population-  
 734 based templates. Common age-related pathologies not present in these 240  
 735 individuals will also not be present in ADTs, whose image intensities are  
 736 all derived from the the same 240 individuals. Similarly, DTI volumes of  
 737 the ADTs will show expected differences in shape but not in the derived FA  
 738 and MD maps. However, despite the use of these 240 younger individuals



739 for the GP-estimated ADTs, appearance did not show substantial differences  
740 compared to the directly-estimated ADT-81 (from 80-81 year old UKB indi-  
741 viduals) and the older-adult MIITRA template.

742 On average, alignment to OMM-1 via registration to an individual’s age-  
743 corresponding ADT shows slightly better spatial normalization performance  
744 than registering directly to OMM-1 in both UKB and HCP test subjects.  
745 Although the MNI 152 template is outside the age range of the UKB, it is  
746 commonly used for adult studies of all ages and, as such, it was included in our  
747 comparison. The use of OMM-1 and ADTs as template spaces outperformed  
748 MNI 152 on both UKB and HCP test subjects for the majority of ROIs. The  
749 performance of MNI 152 was similar to T1-only registration to OMM-1 when  
750 the skull was included. This can be explained by the large scalp signal present  
751 in the UKB T1 images and, consequently, also in OMM-1 and ADTs that  
752 does not match the characteristics of HCP data. The sharp improvement of  
753 T1-only registrations when used with a brain mask can be observed in Figure  
754 7. Hence, we recommend the use of a brain mask for the T1 template volume  
755 when used as registration reference in uni- and multimodal registrations. It  
756 is also recommended to use a mask for DTI to avoid any potential impact of  
757 noisy tensors outside the brain. We supply such masks in template space as  
758 part of our OMM-1 release.

759 Whether or not the benefits of slightly improved registration outweigh  
760 the added complexity of using ADTs will likely depend on the specifics of a  
761 particular study. In many cases it may be sufficient to simply use the OMM-1  
762 directly. However, should the use of an ADT be preferred (e.g., the 80 year  
763 ADT for an older population study), then our template provides a natural  
764 way to compare results with those from other studies using the generic OMM-  
765 1, as well as making atlases defined in generic OMM-1 space available in any  
766 ADT. In addition to age, a new GP model could also be conditioned on other  
767 attributes such as sex. To get high confidence predictions, this approach  
768 requires datasets with sufficiently well-represented sub-populations with the  
769 attribute of interest.

770 Our multimodal templating strategy provides a framework for integrat-  
771 ing complementary information from scalar- and tensor-valued modalities with  
772 MMORF in a fully-unbiased and internally-consistent way. The use of our  
773 templates in combination with MMORF can largely improve accuracy in  
774 spatial normalisation tasks and the availability of spatially corresponding in-  
775 formation from anatomical images such as T1 and T2-FLAIR and diffusion  
776 tensors from dMRI will greatly benefit the interpretation of results in tem-

plate space. OMM-1 and all pre-constructed ADTs, the code for template construction, and MMORF and the MMORF config files used for the non-linear registrations can be publicly accessed via (Arthofer et al., 2023). Our template construction pipeline is not limited to these specific modalities and can be readily applied to other modalities and datasets of interest. In the future we hope to further extend the field of view of the OMM-1 to include the whole neck and face, which could add further benefits for MEG and EEG studies.

## 5. Acknowledgements

This research was funded by a Wellcome Trust Collaborative Award (215573/Z/19/Z). The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z). Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This research has been conducted using the UK Biobank Resource under Application Number 8107. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

## Appendix A. Dealing with high intensity values

Clamping of high scalp intensities was achieved with a custom sigmoid clamping function

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } 0 < x \leq k \\ k/(2 * (1 + \exp(-8 * (x - k)/k))) + 0.75 * k & \text{if } x > k \end{cases} \quad (\text{A.1})$$

where  $x$  is a voxel's intensity and  $k$  the mean WM intensity.

## 805 **Appendix B. MMORF registration parameters**

806 Table Appendix B contains the MMORF registration parameters used  
 807 for the construction of the OMM-1 and age-dependent templates. The main  
 808 difference is in the maximum warp resolution which is 1 mm for the former  
 809 and 2 mm for the latter. The choice of reducing the warp resolution from  
 810 1 mm to 2 mm was mainly motivated by savings in runtime, while still  
 811 maintaining a high level of detail.

	OMM-1	Age-dependent templates
; general		
warp_res_init	32	32
warp_scaling	1 1 2 2 2 2 2	1 1 2 2 2 2
lambda_reg	4.0e5 3.7e-1 3.1e-1 2.6e-1 2.2e-1 1.8e-1 1.5e-1	4.0e5 3.7e-1 3.1e-1 2.6e-1 2.2e-1 1.8e-1
hires	3.9	3.9
optimiser_max_it_lowres	10	5
optimiser_max_it_hires	10	5
; whole head T1		
use_implicit_mask	0	0
use_mask_ref_scalar	1 1 1 1 1 1 1	1 1 1 1 1 1
use_mask_mov_scalar	0 0 0 0 0 0 0	0 0 0 0 0 0
fwhm_ref_scalar	8.0 8.0 4.0 2.0 1.0 0.5 0.25	8.0 8.0 4.0 2.0 1.0 0.5
fwhm_mov_scalar	8.0 8.0 4.0 2.0 1.0 0.5 0.25	8.0 8.0 4.0 2.0 1.0 0.5
lambda_scalar	1 1 1 1 1 1 1	1 1 1 1 1 1
estimate_bias	1	1
bias_res_init	32	32
lambda_bias_reg	1e9 1e9 1e9 1e9 1e9 1e9 1e9	1e9 1e9 1e9 1e9 1e9 1e9
; whole head FLAIR		
use_implicit_mask	0	0
use_mask_ref_scalar	0 0 0 0 0 0 0	0 0 0 0 0 0
use_mask_mov_scalar	0 0 0 0 0 0 0	0 0 0 0 0 0
fwhm_ref_scalar	8.0 8.0 4.0 2.0 1.0 0.5 0.25	8.0 8.0 4.0 2.0 1.0 0.5
fwhm_mov_scalar	8.0 8.0 4.0 2.0 1.0 0.5 0.25	8.0 8.0 4.0 2.0 1.0 0.5
lambda_scalar	1 1 1 1 1 1 1	1 1 1 1 1 1
estimate_bias	1	1
bias_res_init	32	32
lambda_bias_reg	1e9 1e9 1e9 1e9 1e9 1e9 1e9	1e9 1e9 1e9 1e9 1e9 1e9
; DTI		
use_mask_ref_tensor	1 1 1 1 1 1 1	1 1 1 1 1 1
use_mask_mov_tensor	1 1 1 1 1 1 1	1 1 1 1 1 1
fwhm_ref_tensor	8.0 8.0 4.0 2.0 1.0 0.5 0.25	8.0 8.0 4.0 2.0 1.0 0.5
fwhm_mov_tensor	8.0 8.0 4.0 2.0 1.0 0.5 0.25	8.0 8.0 4.0 2.0 1.0 0.5
lambda_tensor	1 1 1 1 1 1 1	1 1 1 1 1 1

Table B.3: MMORF parameters used for the registrations when constructing the OMM-1 and age-dependent templates. Given the large number of individuals used in the age-dependent template modelling and the large associated computational requirements the highest warp resolution was set to 2 mm, in contrast to 1 mm used for the OMM-1.

## 812 References

- 813 Alexander, D., Pierpaoli, C., Basser, P., Gee, J., 2001. Spatial transforma-  
814 tions of diffusion tensor magnetic resonance images. *IEEE Transactions*  
815 *on Medical Imaging* 20, 1131–1139. doi:10.1109/42.963816.
- 816 Alfaro-Almagro, F., Jenkinson, M., Bangerter, N.K., Andersson, J.L.,  
817 Griffanti, L., Douaud, G., Sotiropoulos, S.N., Jbabdi, S., Hernandez-  
818 Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy,  
819 P., Rorden, C., Daducci, A., Alexander, D.C., Zhang, H., Drag-  
820 onu, I., Matthews, P.M., Miller, K.L., Smith, S.M., 2018. Image  
821 processing and Quality Control for the first 10,000 brain imag-  
822 ing datasets from UK Biobank. *NeuroImage* 166, 400–424. URL:  
823 <https://linkinghub.elsevier.com/retrieve/pii/S1053811917308613>,  
824 doi:10.1016/j.neuroimage.2017.10.034.
- 825 Andersson, J.L., Skare, S., Ashburner, J., 2003. How to correct  
826 susceptibility distortions in spin-echo echo-planar images: applica-  
827 tion to diffusion tensor imaging. *NeuroImage* 20, 870–888. URL:  
828 <https://www.sciencedirect.com/science/article/pii/S1053811903003367>,  
829 doi:https://doi.org/10.1016/S1053-8119(03)00336-7.
- 830 Andersson, J.L., Sotiropoulos, S.N., 2016. An integrated ap-  
831 proach to correction for off-resonance effects and subject move-  
832 ment in diffusion mr imaging. *NeuroImage* 125, 1063–1078. URL:  
833 <https://www.sciencedirect.com/science/article/pii/S1053811915009209>,  
834 doi:https://doi.org/10.1016/j.neuroimage.2015.10.019.
- 835 Andersson, J.L.R., Graham, M.S., Zsoldos, E., Sotiropoulos, S.N.,  
836 2016. Incorporating outlier detection and replacement into a  
837 non-parametric framework for movement and distortion correc-  
838 tion of diffusion mr images. *NeuroImage* 141, 556–572. URL:  
839 <https://www.sciencedirect.com/science/article/pii/S1053811916303068>,  
840 doi:https://doi.org/10.1016/j.neuroimage.2016.06.058.
- 841 Arthofer, C., Smith, S., Douaud, G., Bartsch, A., Andersson, J., Lange, F.,  
842 2023. Multimodal MRI brain templates from UK Biobank: Oxford-MM.  
843 URL: [osf.io/s9ge4](https://osf.io/s9ge4), doi:10.17605/OSF.IO/S9GE4.

- 844 Avants, B., Duda, J.T., Kim, J., Zhang, H., Pluta, J., Gee, J.C., Whyte,  
845 J., 2008. Multivariate analysis of structural and diffusion imaging  
846 in traumatic brain injury. *Academic Radiology* 15, 1360–1375. URL:  
847 <https://www.sciencedirect.com/science/article/pii/S1076633208003954>,  
848 doi:<https://doi.org/10.1016/j.acra.2008.07.007>.
- 849 Bassar, P., Mattiello, J., Lebihan, D., 1994. Estimation of the  
850 effective self-diffusion tensor from the nmr spin echo. *Jour-*  
851 *nal of Magnetic Resonance, Series B* 103, 247–254. URL:  
852 <https://www.sciencedirect.com/science/article/pii/S1064186684710375>,  
853 doi:<https://doi.org/10.1006/jmrb.1994.1037>.
- 854 Dalca, A., Rakic, M., Gutttag, J., Sabuncu, M., 2019. Learning conditional  
855 deformable templates with convolutional networks. *Advances in neural*  
856 *information processing systems* 32.
- 857 Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S., 2007. Population shape  
858 regression from random design data, in: 2007 IEEE 11th International  
859 Conference on Computer Vision, pp. 1–7. doi:10.1109/ICCV.2007.4408977.
- 860 Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C.,  
861 Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman,  
862 B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling  
863 system for subdividing the human cerebral cortex on mri scans into  
864 gyral based regions of interest. *NeuroImage* 31, 968–980. URL:  
865 <https://www.sciencedirect.com/science/article/pii/S1053811906000437>,  
866 doi:<https://doi.org/10.1016/j.neuroimage.2006.01.021>.
- 867 Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Auto-  
868 matic parcellation of human cortical gyri and sulci using stan-  
869 dard anatomical nomenclature. *NeuroImage* 53, 1–15. URL:  
870 <https://www.sciencedirect.com/science/article/pii/S1053811910008542>,  
871 doi:<https://doi.org/10.1016/j.neuroimage.2010.06.010>.
- 872 Evans, A., Collins, D., Mills, S., Brown, E., Kelly, R., Peters, T., 1993. 3d  
873 statistical neuroanatomical models from 305 mri volumes, in: 1993 IEEE  
874 Conference Record Nuclear Science Symposium and Medical Imaging Con-  
875 ference, pp. 1813–1817 vol.3. doi:10.1109/NSSMIC.1993.373602.

- 876 Fillmore, P.T., Phillips-Meek, M.C., Richards, J.E., 2015. Age-  
877 specific mri brain and head templates for healthy adults from 20  
878 through 89 years of age. *Frontiers in Aging Neuroscience* 7. URL:  
879 <https://www.frontiersin.org/articles/10.3389/fnagi.2015.00044>,  
880 doi:10.3389/fnagi.2015.00044.
- 881 Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat,  
882 D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness,  
883 V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically Parcel-  
884 lating the Human Cerebral Cortex. *Cerebral Cortex* 14, 11–22. URL:  
885 <https://doi.org/10.1093/cercor/bhg087>, doi:10.1093/cercor/bhg087,  
886 arXiv:<https://academic.oup.com/cercor/article-pdf/14/1/11/1193353/bhg087.pdf>.
- 887 Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry,  
888 R.C., Collins, D.L., 2011. Unbiased average age-appropriate at-  
889 lases for pediatric studies. *NeuroImage* 54, 313–327. URL:  
890 <https://www.sciencedirect.com/science/article/pii/S1053811910010062>,  
891 doi:<https://doi.org/10.1016/j.neuroimage.2010.07.033>.
- 892 Fraser, M.A., Walsh, E.I., Shaw, M.E., Abhayaratna, W.P., Anstey,  
893 K.J., Sachdev, P.S., Cherbuin, N., 2021. Longitudinal trajecto-  
894 ries of hippocampal volume in middle to older age community  
895 dwelling individuals. *Neurobiology of Aging* 97, 97–105. URL:  
896 <https://www.sciencedirect.com/science/article/pii/S0197458020303213>,  
897 doi:<https://doi.org/10.1016/j.neurobiolaging.2020.10.011>.
- 898 Grabner, G., Janke, A.L., Budge, M.M., Smith, D., Pruessner, J., Collins,  
899 D.L., 2006. Symmetric atlasing and model based segmentation: An ap-  
900 plication to the hippocampus in older adults, in: Larsen, R., Nielsen, M.,  
901 Sparring, J. (Eds.), *Medical Image Computing and Computer-Assisted In-*  
902 *tervention – MICCAI 2006*, Springer Berlin Heidelberg, Berlin, Heidelberg.  
903 pp. 58–66.
- 904 Greve, D.N., Fischl, B., 2009. Accurate and robust brain image align-  
905 ment using boundary-based registration. *NeuroImage* 48, 63–72. URL:  
906 <https://www.sciencedirect.com/science/article/pii/S1053811909006752>,  
907 doi:<https://doi.org/10.1016/j.neuroimage.2009.06.060>.
- 908 Guimond, A., Meunier, J., Thirion, J.P., 1998. Automatic computation of  
909 average brain models, in: Wells, W.M., Colchester, A., Delp, S. (Eds.),



- 910 Medical Image Computing and Computer-Assisted Intervention — MIC-  
911 CAI'98, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 631–640.
- 912 Guimond, A., Meunier, J., Thirion, J.P., 2000. Aver-  
913 age brain models: A convergence study. *Computer Vi-*  
914 *sion and Image Understanding* 77, 192–210. URL:  
915 <https://www.sciencedirect.com/science/article/pii/S1077314299908159>,  
916 doi:<https://doi.org/10.1006/cviu.1999.0815>.
- 917 Gupta, V., Malandain, G., Ayache, N., Pennec, X., 2016. A framework for  
918 creating population specific multimodal brain atlas using clinical t1 and  
919 diffusion tensor images, in: Fuster, A., Ghosh, A., Kaden, E., Rath, Y.,  
920 Reiser, M. (Eds.), *Computational Diffusion MRI*, Springer International  
921 Publishing, Cham. pp. 99–108.
- 922 Irfanoglu, M., Nayak, A., Pierpaoli, C., 2020. Diffusion mri atlases  
923 from the human connectome project data, in: 2020 Proceedings of  
924 the International Society for Magnetic Resonance Imaging. URL:  
925 <https://cds.ismrm.org/protected/20MProceedings/PDFfiles/0848.html>.
- 926 Irfanoglu, M.O., Nayak, A., Jenkins, J., Hutchinson, E.B.,  
927 Sadeghi, N., Thomas, C.P., Pierpaoli, C., 2016. DR-TAMAS:  
928 Diffeomorphic registration for tensor accurate alignment of  
929 anatomical structures. *NeuroImage* 132, 439–454. URL:  
930 <https://www.sciencedirect.com/science/article/pii/S1053811916001750>,  
931 doi:<https://doi.org/10.1016/j.neuroimage.2016.02.066>.
- 932 Janahi, M., Aksman, L., Schott, J.M., Mokrab, Y., Altmann, A.,  
933 2022. Nomograms of human hippocampal volume shifted by polygenic  
934 scores. *eLife* 11, e78232. URL: <https://doi.org/10.7554/eLife.78232>,  
935 doi:10.7554/eLife.78232.
- 936 Jenkinson, M., Smith, S., 2001. A global optimisa-  
937 tion method for robust affine registration of brain im-  
938 ages. *Medical Image Analysis* 5, 143–156. URL:  
939 <https://www.sciencedirect.com/science/article/pii/S1361841501000366>,  
940 doi:10.1016/S1361-8415(01)00036-6.
- 941 Lange, F.J., Ashburner, J., Smith, S.M., Andersson, J.L.,  
942 2020a. A symmetric prior for the regularisation of elas-

- 943 tic deformations: Improved anatomical plausibility in non-  
944 linear image registration. *NeuroImage* 219, 116962. URL:  
945 <https://www.sciencedirect.com/science/article/pii/S1053811920304481>,  
946 doi:<https://doi.org/10.1016/j.neuroimage.2020.116962>.
- 947 Lange, F.J., Smith, S.M., Bertelsen, M.F., Khrapitchev, A.A., Manger, P.R.,  
948 Mars, R.B., Andersson, J.L.R., 2020b. Multimodal mri template creation  
949 in the ring-tailed lemur and rhesus macaque, in: Špiclin, Ž., McClelland,  
950 J., Kybic, J., Goksel, O. (Eds.), *Biomedical Image Registration*, Springer  
951 International Publishing, Cham. pp. 141–150.
- 952 Lv, J., Zeng, R., Ho, M.P., D'Souza, A., Calamante, F.,  
953 2022. Building a tissue-unbiased brain template of fiber ori-  
954 entation distribution and tractography with multimodal regis-  
955 tration. *Magnetic Resonance in Medicine*, mrm.29496 URL:  
956 <https://onlinelibrary.wiley.com/doi/10.1002/mrm.29496>,  
957 doi:10.1002/mrm.29496.
- 958 Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J.,  
959 1995. A Probabilistic Atlas of the Human Brain: Theory and  
960 Rationale for Its Development: The International Consortium  
961 for Brain Mapping (ICBM). *NeuroImage* 2, 89–101. URL:  
962 <https://www.sciencedirect.com/science/article/pii/S1053811985710129>,  
963 doi:10.1006/nimg.1995.1012.
- 964 Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub,  
965 E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L.R.,  
966 Griffanti, L., Douaud, G., Okell, T.W., Weale, P., Dragonu, I., Garratt,  
967 S., Hudson, S., Collins, R., Jenkinson, M., Matthews, P.M., Smith, S.M.,  
968 2016. Multimodal population brain imaging in the UK Biobank prospec-  
969 tive epidemiological study. *Nature Neuroscience* 19, 1523–1536. URL:  
970 <http://www.nature.com/articles/nn.4393>, doi:10.1038/nn.4393.
- 971 Nobis, L., Manohar, S.G., Smith, S.M., Alfaro-Almagro, F., Jenk-  
972 inson, M., Mackay, C.E., Husain, M., 2019. Hippocampal vol-  
973 ume across age: Nomograms derived from over 19,700 peo-  
974 ple in uk biobank. *NeuroImage: Clinical* 23, 101904. URL:  
975 <https://www.sciencedirect.com/science/article/pii/S2213158219302542>,  
976 doi:<https://doi.org/10.1016/j.nicl.2019.101904>.

- 977 Patenaude, B., Smith, S.M., Kennedy, D.N., Jenkinson, M.,  
978 2011. A bayesian model of shape and appearance for sub-  
979 cortical brain segmentation. *NeuroImage* 56, 907–922. URL:  
980 <https://www.sciencedirect.com/science/article/pii/S1053811911002023>,  
981 doi:<https://doi.org/10.1016/j.neuroimage.2011.02.046>.
- 982 Pecheva, D., Iversen, J.R., Palmer, C.E., Watts, R., Jernigan,  
983 T.L., Hagler, D.J., Dale, A.M., 2022. Multimodal image nor-  
984 malisation tool (mint) for the adolescent brain and cogni-  
985 tive development study: the mint abcd atlas. *bioRxiv* URL:  
986 <https://www.biorxiv.org/content/early/2022/08/16/2022.08.09.503395>,  
987 doi:10.1101/2022.08.09.503395, arXiv:<https://www.biorxiv.org/content/early/2022/08/16/2022.08.09.503395>.
- 988 Rasmussen, C.E., Williams, C.K.I., 2005. Gaussian Pro-  
989 cesses for Machine Learning. The MIT Press. URL:  
990 <https://doi.org/10.7551/mitpress/3206.001.0001>,  
991 doi:10.7551/mitpress/3206.001.0001.
- 992 Rohlfing, T., Zahr, N.M., Sullivan, E.V., Pfefferbaum, A.,  
993 2010. The SRI24 multichannel atlas of normal adult human  
994 brain structure. *Human Brain Mapping* 31, 798–819. URL:  
995 <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.20906>,  
996 doi:10.1002/hbm.20906. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.20906>.
- 997 Serag, A., Aljabar, P., Ball, G., Counsell, S.J., Boardman, J.P., Rutherford,  
998 M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D., 2012. Construction  
999 of a consistent high-definition spatio-temporal atlas of the developing  
1000 brain using adaptive kernel regression. *NeuroImage* 59, 2255–2265. URL:  
1001 <https://linkinghub.elsevier.com/retrieve/pii/S1053811911011360>,  
1002 doi:10.1016/j.neuroimage.2011.09.062.
- 1003 Tustison, N.J., Shrinidhi, K.L., Wintermark, M., Durst, C.R., Kandel, B.M.,  
1004 Gee, J.C., Grossman, M.C., Avants, B.B., 2015. Optimal Symmetric  
1005 Multimodal Templates and Concatenated Random Forests for Supervised  
1006 Brain Tumor Segmentation (Simplified) with ANTsR. *Neuroinformat-*  
1007 *ics* 13, 209–225. URL: <https://doi.org/10.1007/s12021-014-9245-2>,  
1008 doi:10.1007/s12021-014-9245-2.
- 1009 Van Essen, D., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.,  
1010 Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Della

- 1011 Penna, S., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-  
1012 Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld,  
1013 R., Petersen, S., Prior, F., Schlaggar, B., Smith, S., Snyder, A.,  
1014 Xu, J., Yacoub, E., 2012. The human connectome project: A  
1015 data acquisition perspective. *NeuroImage* 62, 2222–2231. URL:  
1016 <https://www.sciencedirect.com/science/article/pii/S1053811912001954>,  
1017 doi:<https://doi.org/10.1016/j.neuroimage.2012.02.018>. connectivity.
- 1018 Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Ya-  
1019 coub, E., Ugurbil, K., 2013. The wu-minn human connec-  
1020 tome project: An overview. *NeuroImage* 80, 62–79. URL:  
1021 <https://www.sciencedirect.com/science/article/pii/S1053811913005351>,  
1022 doi:<https://doi.org/10.1016/j.neuroimage.2013.05.041>. mapping the Con-  
1023 nectome.
- 1024 Vinke, E.J., de Groot, M., Venkatraghavan, V., Klein, S., Niessen, W.J.,  
1025 Ikram, M.A., Vernooij, M.W., 2018. Trajectories of imaging markers in  
1026 brain aging: the rotterdam study. *Neurobiology of Aging* 71, 32–40. URL:  
1027 <https://www.sciencedirect.com/science/article/pii/S0197458018302513>,  
1028 doi:<https://doi.org/10.1016/j.neurobiolaging.2018.07.001>.
- 1029 Walhovd, K.B., Fjell, A.M., Reinvang, I., Lundervold, A., Dale, A.M.,  
1030 Eilertsen, D.E., Quinn, B.T., Salat, D., Makris, N., Fischl, B.,  
1031 2005. Effects of age on volumes of cortex, white matter and sub-  
1032 cortical structures. *Neurobiology of Aging* 26, 1261–1270. URL:  
1033 <https://www.sciencedirect.com/science/article/pii/S0197458005001673>,  
1034 doi:<https://doi.org/10.1016/j.neurobiolaging.2005.05.020>.
- 1035 Wang, Y., Xu, Q., Luo, J., Hu, M., Zuo, C., 2019. Effects of age and  
1036 sex on subcortical volumes. *Frontiers in Aging Neuroscience* 11. URL:  
1037 <https://www.frontiersin.org/articles/10.3389/fnagi.2019.00259>,  
1038 doi:10.3389/fnagi.2019.00259.
- 1039 Wilms, M., Bannister, J.J., Mouches, P., MacDonald, M.E., Rajashekar,  
1040 D., Langner, S., Forkert, N.D., 2020. Bidirectional modeling and analy-  
1041 sis of brain aging with normalizing flows, in: Kia, S.M., Mohy-ud Din,  
1042 H., Abdulkadir, A., Bass, C., Habes, M., Rondina, J.M., Tax, C., Wang,  
1043 H., Wolfers, T., Rathore, S., Ingallhalikar, M. (Eds.), *Machine Learning  
1044 in Clinical Neuroimaging and Radiogenomics in Neuro-oncology*, Springer  
1045 International Publishing, Cham. pp. 23–33.

- 1046 Wu, Y., Ridwan, A.R., Niaz, M.R., Qi, X., Zhang, S., Alzheimer's  
1047 Disease Neuroimaging Initiative, Bennett, D.A., Arfanakis, K., 2022.  
1048 Development of high quality t1-weighted and diffusion tensor templates  
1049 of the older adult brain in a common space. *NeuroImage* 260, 119417. URL:  
1050 <https://www.sciencedirect.com/science/article/pii/S1053811922005341>,  
1051 doi:<https://doi.org/10.1016/j.neuroimage.2022.119417>.
- 1052 Xia, T., Chartsias, A., Tsafaris, S.A., 2019. Consistent brain ageing syn-  
1053 thesis, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou,  
1054 S., Yap, P.T., Khan, A. (Eds.), *Medical Image Computing and Computer*  
1055 *Assisted Intervention – MICCAI 2019*, Springer International Publishing,  
1056 Cham. pp. 750–758.
- 1057 Yang, G., Zhou, S., Bozek, J., Dong, H.M., Han, M., Zuo, X.N.,  
1058 Liu, H., Gao, J.H., 2020. Sample sizes and population differences  
1059 in brain template construction. *NeuroImage* 206, 116318. URL:  
1060 <https://linkinghub.elsevier.com/retrieve/pii/S1053811919309097>,  
1061 doi:10.1016/j.neuroimage.2019.116318.
- 1062 Zhang, S., Peng, H., Dawe, R.J., Arfanakis, K., 2011. Enhanced icbm dif-  
1063 fusion tensor template of the human brain. *NeuroImage* 54, 974–984. URL:  
1064 <https://www.sciencedirect.com/science/article/pii/S1053811910011894>,  
1065 doi:<https://doi.org/10.1016/j.neuroimage.2010.09.008>.
- 1066 Zhao, Q., Adeli, E., Honnorat, N., Leng, T., Pohl, K.M., 2019. Variational  
1067 autoencoder for regression: Application to brain aging analysis, in: Shen,  
1068 D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T.,  
1069 Khan, A. (Eds.), *Medical Image Computing and Computer Assisted In-*  
1070 *tervention – MICCAI 2019*, Springer International Publishing, Cham. pp.  
1071 823–831.