Multi-contrast anatomical subcortical structures parcellation


DOI
10.7554/eLife.59430

Publication date
2020

Document Version
Final published version

Published in
eLife

License
CC BY

Citation for published version (APA):
Multi-contrast anatomical subcortical structures parcellation

Pierre-Louis Bazin1,2*, Anneke Alkemade1, Martijn J Mulder1,3, Amanda G Henry4, Birte U Forstmann1

1Integrative Model-based Cognitive Neuroscience research unit, University of Amsterdam, Amsterdam, Netherlands; 2Max-Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; 3Psychology Department, Utrecht University, Utrecht, Netherlands; 4Faculty of Archaeology, Leiden University, Leiden, Netherlands

Abstract The human subcortex is comprised of more than 450 individual nuclei which lie deep in the brain. Due to their small size and close proximity, up until now only 7% have been depicted in standard MRI atlases. Thus, the human subcortex can largely be considered as terra incognita. Here, we present a new open-source parcellation algorithm to automatically map the subcortex. The new algorithm has been tested on 17 prominent subcortical structures based on a large quantitative MRI dataset at 7 Tesla. It has been carefully validated against expert human raters and previous methods, and can easily be extended to other subcortical structures and applied to any quantitative MRI dataset. In sum, we hope this novel parcellation algorithm will facilitate functional and structural neuroimaging research into small subcortical nuclei and help to chart terra incognita.

Introduction

Subcortical brain structures are often neglected in neuroimaging studies due to their small size, limited inter-regional contrast, and weak signal-to-noise ratio in functional imaging (Forstmann et al., 2016; Johansen-Berg, 2013). Yet, these small and diverse structures are prominent nodes in functional networks (Marquand et al., 2017; Ji et al., 2019), and they undergo pathological alterations already at early stages of neurodegenerative diseases (Andersen et al., 2014; Koshiyama et al., 2018). Deep brain stimulation surgery, originally performed to reduce motor symptoms in essential tremors, is now a promising therapeutic option in later stages of Parkinson’s disease and movement disorders, as well as refractory psychiatric illnesses in obsessive-compulsive disorder, anorexia, or depression (Forstmann et al., 2017; Mosley et al., 2018). Evolutionary genetics even uncovered that in modern humans, Neanderthal-inherited alleles were preferentially down-regulated in subcortical and cerebellar regions compared to other brain regions (McCoy et al., 2017), suggesting these structures to be essential in making us specifically human.

Despite their importance, these areas are particularly difficult to image. Furthermore, the size, shape, and location of these brain regions changes with development and aging (Fjell et al., 2013; Keuken et al., 2013; Yeatman et al., 2014; Herting et al., 2018). Experience-based plasticity continuously remodels myelin (Tardif et al., 2016; Hill et al., 2018; Turner, 2019), iron and other magnetic substances accumulate with age or pathology (Andersen et al., 2014; Zhang et al., 2018), both bringing changes in the MRI appearance of subcortical regions with diverse tissue characteristics (Draganski et al., 2011; Keuken et al., 2017).

Thus, mapping the structure and function of the subcortex is a major endeavor as well as a major challenge for human neuroscience. Extensive work available from animal brain models unfortunately does not translate in a straightforward way to human subcortical anatomy nor does it shed much light on its involvement in human cognition (Steiner and Tseng, 2017). Besides serious difficulties in
obtaining adequate measures of subcortical neural activity in functional MRI (de Hollander et al., 2017; Miletić et al., 2020), atlases and techniques for labeling accurately and reliably individual subcortical structures have also been scarce (Frazier et al., 2005; Chakravarty et al., 2006; Ahsan et al., 2007; Yelnik et al., 2007; Qiu et al., 2010; Patenaude et al., 2011), typically labeling the thalamus, striatum (or its subdivision into caudate and putamen), and globus pallidus (internal and external segments combined), sometimes the amygdala. However, recent advances in anatomical MRI, combining multiple contrasts and/or quantitative MRI mapping and utilizing the higher resolution achievable with 7 Tesla (7T) and above have started to reduce the gap, each mapping a few additional structures or sub-structures, primarily the iron-rich substantia nigra, red nucleus and sub-thalamic nucleus (Keuken et al., 2013; Xiao et al., 2015; Visser et al., 2016a; Visser et al., 2016b; Wang et al., 2016; Makowski et al., 2018; Ewert et al., 2018; Iglesias et al., 2018; Pauli et al., 2018; Sitek et al., 2019). While these efforts generated valuable atlases, they do not yet enable to identify many subcortical structures in individual subjects. Manual delineation, on the other hand, requires extensive manual labor from highly trained experts which cannot be easily applied to large cohorts or clinical settings.

Here, we propose a new automated parcellation technique to identify and label 17 individual subcortical structures of varying size and composition in individual subjects, based on a large quantitative 7T MRI database (Alkemade et al., 2020), using quantitative maps of relaxation rates R1 and R2* (1/T1 and 1/T2*, respectively) and quantitative susceptibility maps (QSM) as anatomical contrasts. The algorithm, named Multi-contrast Anatomical Subcortical Structure Parcellation (MASSP), follows a Bayesian multi-object approach similar in essence to previous efforts (Fischl et al., 2002; Eugenio Iglesias et al., 2013; Visser et al., 2016a; Garzón et al., 2018), combining shape priors, intensity distribution models, spatial relationships, and global constraints. The main innovation of our approach is to explicitly estimate interfaces between subcortical structures based on a joint model derived from signed distance functions. Modeling interfaces in addition to the structure itself provides a rich basis to encode relationships and anatomical knowledge in shape and intensity priors. A voxel-wise Markovian diffusion regularizes the combined priors for each defined interface, lowering the imaging noise. Finally, the voxel-wise posteriors for the different structures and interfaces are further combined into global anatomical parcels by topology correction and region growing taking into account volumetric priors, which regularizes parcellation results further in smaller nuclei with low or heterogeneous contrast. To validate the results from this new method, in a thorough comparison with expert manual labeling, we show that the proposed method provides results very close from manual raters in many structures and exhibit reasonable bias across the adult lifespan. The method can easily be extended to new structures, can be applied to any quantitative MRI dataset and is available in Open Source as part of Nighres (Huntenburg et al., 2018), a neuroimage analysis package aimed at high-resolution neuroimaging.

Results

The MASSP parcellation method presented here has been trained to parcellate the following 17 structures: striatum (Str), thalamus (Tha), lateral, 3rd and 4th ventricles (LV, 3V, 4V), amygdala (Amg), globus pallidus internal segment (GPI) and external segment (GPe), SN, STN, red nucleus (RN), ventral tegmental area (VTA), fornix (fx), internal capsule (ic), periaqueductal gray (PAG), pedunculopontine nucleus (PPN), and claustrum (Cl), see Figure 1. These structures include the most commonly defined subcortical regions (Str, Tha, Amg, LV), the main iron-rich nuclei (GPI, GPe, RN, SN, STN), as well as smaller, less studied areas (VTA, PAG, PPN, Cl), white matter structures (ic, fx), and the central ventricles (3V, 4V).

MASSP uses a data set of ten expert delineations as a basis for its modeling. From the delineations, an atlas of interfaces between structures, shape skeletons, and interface intensity histograms are generated, and used as prior in a multiple-step non-iterative Bayesian algorithm, see Figure 2 and Materials and methods.

Validation against manual delineations

In a leave-one-out validation study comparing performance with the manual delineations, MASSP performed above 95% of the level of quality of the raters for Str, Tha, 4V, GPe, SN, RN, VTA, ic in terms of Dice overlap, the most stringent of the quality measures (see Figures 3 and 4 and Table 1).
Several of the smaller structures have lower overlap ratios likely due to their smaller size (GPi, STN, PAG, PPN). Structures with an elongated shape (fx, CI) remain challenging, due to the fact that small differences in location can substantially reduce overlap (Bazin et al., 2016). Despite these challenges, when comparing the dilated Dice scores, all structures were above 75% of overlap, with most reaching over 90% of the manual raters ability. Note that the Dice coefficient is very sensitive to size, as smaller structures will have lower overlap ratios for the same number of misclassified voxels. The dilated Dice coefficient is more representative of the variability regardless of size, as the smaller structures can reach high levels of overlap, both in manual and automated parcellations (see Table 1). The average surface distance confirms these results, showing values generally between one and two voxels of distance at a resolution of 0.7 mm, except in the cases of Amg, LV, fx, PPN, and CI. These structures are generally more variable (LV), elongated (fx, CI), or have a particularly low contrast with neighboring regions (Amg, PPN).

Comparison to other automated methods

To provide a basis for comparison, we applied other freely available methods for subcortical structure parcellation to the same 10 subjects. MASSP performs similarly to or better than Freesurfer, FSL FIRST and a multi-atlas registration using ANTs (see Table 2). Multi-atlas registration provides high accuracy in most structures as well, but is biased toward under-estimating the size of smaller and elongated structures where overlap is systematically reduced across the individual atlas subjects. Multi-atlas registration is also quite computationally intensive when using multiple contrasts at high

Figure 1. The 17 subcortical structures currently included in the parcellation algorithm in axial (A), sagittal (B), and coronal (C) views.

Figure 2. The MASSP parcellation pipeline. Atlas priors for interfaces between structures are combined to the MRI data, regularized via probability diffusion and topology correction, and the final structure posteriors are jointly estimated by region growing.
resolution. Finally, MASSP provides many more structures than Freesurfer and FSL FIRST, and can be easily applied to new structures based on additional manual delineations.

**Application to new MRI contrasts**

Quantitative MRI has only become recently applicable in larger studies, thanks in part to the development of integrated multi-parameter sequences (Weiskopf et al., 2013; Caan et al., 2019). Many data sets, including large-scale open databases, use more common T1- and T2-weighted MRI.

---

**Figure 3.** Leave-one-out validation of the structures parcellated by MASSP, compared to the human rater with most neuroanatomical expertise. Scores for the left and right side are computed separately and then combined into box-and-whisker plots.
order to test the applicability of MASSP to such contrasts, we obtained the test-retest subset of the Human Connectome Project (HCP, Van Essen et al., 2013) and applied MASSP to the 45 pre-processed and skull-stripped T1- and T2-weighted images from each of the two test and retest sessions. While performing manual delineations on the new contrasts would be preferable, the model is already rich enough to provide stable parcellations. Test-retest reproducibility is similarly high for MASSP and Freesurfer, and are generally in agreement, see Figure 5 and Table 3.

**Figure 4.** Inter-rater variability for the human expert raters. Scores for the left and right side are computed separately and then combined into box-and-whisker plots.
Biases due to atlas size

A common concern of brain parcellation methods is the risk of biases, as they are typically built from a small number of manual delineations. Our data set is part of a large scale study of the subcortex, for which we obtained manual delineations of the STN, SN, RN, GPe, and GPi on 105 subjects over the adult lifespan (18–80 year old, see Alkemade et al., 2020 for details). First, we investigated the impact of atlas size. We randomly assigned half of the subjects from each decade to two groups, and built atlas priors from subsets of 3, 5, 8, 10, 12, 15, and 18 subjects from the first group. The subjects used in the atlas were taken randomly from each decade (18-30, 31-40, 41-50, 51-60, 61-70, 71-80), so as to maximize the age range represented in each atlas. Atlases of increasing size were

<table>
<thead>
<tr>
<th>Str</th>
<th>STN</th>
<th>SN</th>
<th>RN</th>
<th>GPe</th>
<th>GPi</th>
<th>Tha</th>
<th>LV</th>
<th>3V</th>
<th>4V</th>
<th>Amg</th>
<th>ic</th>
<th>VTA</th>
<th>fx</th>
<th>PAG</th>
<th>PPN</th>
<th>Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dice overlap</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP</td>
<td>0.893</td>
<td>0.648</td>
<td>0.805</td>
<td>0.870</td>
<td>0.702</td>
<td>0.800</td>
<td>0.867</td>
<td>0.849</td>
<td>0.741</td>
<td>0.869</td>
<td>0.723</td>
<td>0.745</td>
<td>0.570</td>
<td>0.527</td>
<td>0.641</td>
<td>0.496</td>
</tr>
<tr>
<td>Manual</td>
<td>0.897</td>
<td>0.800</td>
<td>0.841</td>
<td>0.875</td>
<td>0.762</td>
<td>0.813</td>
<td>0.877</td>
<td>0.907</td>
<td>0.797</td>
<td>0.882</td>
<td>0.866</td>
<td>0.732</td>
<td>0.574</td>
<td>0.823</td>
<td>0.791</td>
<td>0.665</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.995</td>
<td>0.811</td>
<td>0.957</td>
<td>0.996</td>
<td>0.925</td>
<td>0.987</td>
<td>0.989</td>
<td>0.936</td>
<td>0.936</td>
<td>0.988</td>
<td>0.836</td>
<td>1.020</td>
<td>0.994</td>
<td>0.641</td>
<td>0.814</td>
<td>0.754</td>
</tr>
<tr>
<td>Dilated overlap</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP</td>
<td>0.982</td>
<td>0.919</td>
<td>0.977</td>
<td>0.991</td>
<td>0.909</td>
<td>0.956</td>
<td>0.970</td>
<td>0.929</td>
<td>0.890</td>
<td>0.951</td>
<td>0.891</td>
<td>0.915</td>
<td>0.863</td>
<td>0.756</td>
<td>0.897</td>
<td>0.795</td>
</tr>
<tr>
<td>Manual</td>
<td>0.987</td>
<td>0.988</td>
<td>0.985</td>
<td>0.995</td>
<td>0.953</td>
<td>0.972</td>
<td>0.970</td>
<td>0.967</td>
<td>0.944</td>
<td>0.961</td>
<td>0.978</td>
<td>0.924</td>
<td>0.818</td>
<td>0.957</td>
<td>0.960</td>
<td>0.910</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.995</td>
<td>0.930</td>
<td>0.992</td>
<td>0.995</td>
<td>0.955</td>
<td>0.984</td>
<td>1.000</td>
<td>0.961</td>
<td>0.946</td>
<td>0.991</td>
<td>0.911</td>
<td>0.992</td>
<td>1.059</td>
<td>0.790</td>
<td>0.935</td>
<td>0.879</td>
</tr>
<tr>
<td>Average surface distance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP</td>
<td>0.750</td>
<td>0.911</td>
<td>0.676</td>
<td>0.491</td>
<td>1.140</td>
<td>0.863</td>
<td>1.058</td>
<td>2.690</td>
<td>0.994</td>
<td>0.817</td>
<td>1.476</td>
<td>1.275</td>
<td>1.074</td>
<td>2.950</td>
<td>0.955</td>
<td>1.484</td>
</tr>
<tr>
<td>Manual</td>
<td>0.723</td>
<td>0.508</td>
<td>0.571</td>
<td>0.482</td>
<td>0.902</td>
<td>0.804</td>
<td>0.971</td>
<td>0.615</td>
<td>0.637</td>
<td>0.671</td>
<td>0.779</td>
<td>1.045</td>
<td>1.204</td>
<td>0.703</td>
<td>0.555</td>
<td>0.801</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.971</td>
<td>0.590</td>
<td>0.852</td>
<td>0.996</td>
<td>0.861</td>
<td>0.943</td>
<td>0.916</td>
<td>0.277</td>
<td>0.662</td>
<td>1.020</td>
<td>0.553</td>
<td>0.834</td>
<td>1.161</td>
<td>0.287</td>
<td>0.610</td>
<td>0.619</td>
</tr>
</tbody>
</table>

Table 2. Comparison with multi-atlas registration, Freesurfer, and FSL FIRST.

<table>
<thead>
<tr>
<th>Str</th>
<th>STN</th>
<th>SN</th>
<th>RN</th>
<th>GPe</th>
<th>GPi</th>
<th>Tha</th>
<th>LV</th>
<th>3V</th>
<th>4V</th>
<th>Amg</th>
<th>ic</th>
<th>VTA</th>
<th>fx</th>
<th>PAG</th>
<th>PPN</th>
<th>Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dice overlap</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP</td>
<td>0.893</td>
<td>0.648</td>
<td>0.805</td>
<td>0.870</td>
<td>0.702</td>
<td>0.800</td>
<td>0.867</td>
<td>0.849</td>
<td>0.741</td>
<td>0.869</td>
<td>0.723</td>
<td>0.745</td>
<td>0.570</td>
<td>0.527</td>
<td>0.641</td>
<td>0.496</td>
</tr>
<tr>
<td>Multi-atlas</td>
<td>0.855</td>
<td>0.662</td>
<td>0.760</td>
<td>0.820</td>
<td>0.742</td>
<td>0.796</td>
<td>0.859</td>
<td>0.734</td>
<td>0.660</td>
<td>0.691</td>
<td>0.761</td>
<td>0.718</td>
<td>0.626</td>
<td>0.478</td>
<td>0.674</td>
<td>0.539</td>
</tr>
<tr>
<td>Freesurfer</td>
<td>0.876</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSL FIRST</td>
<td>0.875</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dilated overlap</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP</td>
<td>0.982</td>
<td>0.919</td>
<td>0.977</td>
<td>0.991</td>
<td>0.909</td>
<td>0.956</td>
<td>0.970</td>
<td>0.929</td>
<td>0.890</td>
<td>0.951</td>
<td>0.891</td>
<td>0.915</td>
<td>0.863</td>
<td>0.756</td>
<td>0.897</td>
<td>0.795</td>
</tr>
<tr>
<td>Multi-atlas</td>
<td>0.976</td>
<td>0.938</td>
<td>0.968</td>
<td>0.989</td>
<td>0.947</td>
<td>0.968</td>
<td>0.970</td>
<td>0.920</td>
<td>0.920</td>
<td>0.908</td>
<td>0.921</td>
<td>0.939</td>
<td>0.924</td>
<td>0.798</td>
<td>0.943</td>
<td>0.871</td>
</tr>
<tr>
<td>Freesurfer</td>
<td>0.975</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSL FIRST</td>
<td>0.976</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average surface distance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP</td>
<td>0.750</td>
<td>0.911</td>
<td>0.676</td>
<td>0.491</td>
<td>1.140</td>
<td>0.863</td>
<td>1.058</td>
<td>2.690</td>
<td>0.994</td>
<td>0.817</td>
<td>1.476</td>
<td>1.275</td>
<td>1.074</td>
<td>2.950</td>
<td>0.955</td>
<td>1.484</td>
</tr>
<tr>
<td>Multi-atlas</td>
<td>0.961</td>
<td>0.891</td>
<td>0.858</td>
<td>0.675</td>
<td>0.992</td>
<td>0.882</td>
<td>1.083</td>
<td>1.417</td>
<td>0.932</td>
<td>1.249</td>
<td>1.359</td>
<td>1.129</td>
<td>0.813</td>
<td>1.362</td>
<td>0.794</td>
<td>1.055</td>
</tr>
<tr>
<td>Freesurfer</td>
<td>0.770</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSL FIRST</td>
<td>0.867</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP</td>
<td>0.041</td>
<td>0.017</td>
<td>-0.038</td>
<td>0.007</td>
<td>0.066</td>
<td>0.089</td>
<td>0.040</td>
<td>0.047</td>
<td>0.121</td>
<td>0.047</td>
<td>0.078</td>
<td>0.183</td>
<td>0.032</td>
<td>-0.016</td>
<td>0.026</td>
<td>0.009</td>
</tr>
<tr>
<td>Multi-atlas</td>
<td>0.020</td>
<td>-0.087</td>
<td>-0.009</td>
<td>0.031</td>
<td>0.009</td>
<td>0.014</td>
<td>0.020</td>
<td>-0.003</td>
<td>-0.007</td>
<td>-0.092</td>
<td>-0.038</td>
<td>0.055</td>
<td>-0.067</td>
<td>-0.264</td>
<td>-0.090</td>
<td>-0.269</td>
</tr>
<tr>
<td>Freesurfer</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSL FIRST</td>
<td>-0.100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
constructed by adding subjects to previous atlases, so that atlases of increasing complexity include all subjects from simpler atlases. Results applying these atlases to parcellate the second group are given in Figure 6. As in previous studies (Eugenio Iglesias et al., 2013; Bazin and Pham, 2008), performance quickly stabilized with atlases of more than five subjects (no significant difference in

**Table 3.** Test-retest comparison with Freesurfer on Human Connectome Project data.

<table>
<thead>
<tr>
<th></th>
<th>Str</th>
<th>STN</th>
<th>SN</th>
<th>RN</th>
<th>GPi</th>
<th>GPe</th>
<th>Tha</th>
<th>LV</th>
<th>3V</th>
<th>4V</th>
<th>Amg</th>
<th>ic</th>
<th>VTA</th>
<th>fx</th>
<th>PAG</th>
<th>PPN</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dice overlap</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP test-retest</td>
<td>0.914</td>
<td>0.701</td>
<td>0.818</td>
<td>0.829</td>
<td>0.791</td>
<td>0.859</td>
<td>0.928</td>
<td>0.881</td>
<td>0.837</td>
<td>0.870</td>
<td>0.866</td>
<td>0.860</td>
<td>0.738</td>
<td>0.774</td>
<td>0.714</td>
<td>0.713</td>
<td>0.785</td>
</tr>
<tr>
<td>Freesurfer test-retest</td>
<td>0.898</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP – Freesurfer</td>
<td>0.876</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dilated overlap</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP test-retest</td>
<td>0.987</td>
<td>0.939</td>
<td>0.977</td>
<td>0.978</td>
<td>0.963</td>
<td>0.977</td>
<td>0.990</td>
<td>0.980</td>
<td>0.979</td>
<td>0.986</td>
<td>0.981</td>
<td>0.973</td>
<td>0.969</td>
<td>0.961</td>
<td>0.965</td>
<td>0.972</td>
<td>0.966</td>
</tr>
<tr>
<td>Freesurfer test-retest</td>
<td>0.986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP – Freesurfer</td>
<td>0.954</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Average surface distance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP test-retest</td>
<td>0.513</td>
<td>0.528</td>
<td>0.467</td>
<td>0.461</td>
<td>0.532</td>
<td>0.508</td>
<td>0.488</td>
<td>0.509</td>
<td>0.391</td>
<td>0.419</td>
<td>0.533</td>
<td>0.536</td>
<td>0.431</td>
<td>0.464</td>
<td>0.428</td>
<td>0.402</td>
<td>0.433</td>
</tr>
<tr>
<td>Freesurfer test-retest</td>
<td>0.876</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASSP – Freesurfer</td>
<td>0.976</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Welch's t-tests between using 18 subjects or any subset of 8 or more for all structures and measures).

**Biases due to age differences**
To more specifically test the influence of age on parcellation accuracy, we defined again six age groups by decade and randomly selected 10 subjects from each group. Each set of subjects was used as priors for the five structures above, and applied to the other age groups. Results are...
summarized in Figure 7. Examining this age bias, we can see a decrease in performance when parcellating subjects in the range of 60 to 80 years of age. The choice of priors seem to have a limited impact, which varies across structures. In particular, using priors from a similar age group is not always beneficial.

### Bias on individual measures

Finally, we investigated the impact of this decrease in performance in the estimation of anatomical quantities, see Figure 8. The bias did affect the morphometric measures of structure volume and thickness, but the effects on the local measure of thickness was reduced compared to the global measure of volume. Quantitative MRI averages were very stable even when age biases are present in the parcellations.

For reference, we report structure volumes, thickness, R1, R2* and QSM values estimated from the entire AHEAD cohort for different age groups, extending our previous work based on manual delineations on a different data set (Keuken et al., 2017; Forstmann et al., 2014). Results are given in Table 4, describing average volumes, thickness, and quantitative MRI parameters for young, middle-aged, and older subjects for the 17 subcortical structures.

### Discussion

Our goal with the MASSP algorithm was to provide a fully automated method to delineate as many subcortical structures as possible on high-resolution structural MRI now available on 7T scanners. We
modeled 17 distinct structures, taking into account location, shape, volume, and quantitative MRI contrasts to provide individual subject parcellations. Based on our results, we can be confident that the automated parcellation technique performs comparably to human experts, providing delineations within one or two voxels of the structure boundaries (dilated Dice overlap over 75% for all structures, including in aging groups). Results were nearly indistinguishable from expert delineations for eight major structures (Str, Tha, 4V, GPe, SN, RN, VTA, ic), and smaller structures retain high levels of overlap, comparable to trained human raters. This parcellation includes the most commonly defined structures (Str, Tha, SN, RN, STN) with overlap scores comparable to those previously

Figure 8. Regression of volume (log scale), structure thickness, R1, R2*, and QSM MRI parameters estimated using manual delineations versus MASSP automated parcellations. Circles show individual data points, linear regression is indicated by a straight line, and 95% confidence interval is given as the shaded area. Pearson correlation coefficients are indicated when significant (p-value<0.01).
reported (Garzón et al., 2018; Visser et al., 2016a; Eugenio Iglesias et al., 2013; Chakravarty et al., 2013; Patenaude et al., 2011). More importantly, it also includes structures seldom or never before considered in MRI atlases and parcellation methods, such as GPe, GPi, VTA, 3V, 4V, ic, fx, PAG, PPN, Cl. The technique handles structures of varying sizes well, as indicated by dilated overlap and boundary distance. Additional structures can be added, if they can be reliably delineated by expert raters on single-subject MRI at achievable resolutions. Some enhancement techniques such as building a multi-subject template (Pauli et al., 2018) or adding a denoising step (Bazin et al., 2019) may be beneficial. Co-registration to a high-precision atlas as in Ewert et al., 2018 may also improve the initial alignment over the MASSP group average template.

Age biases are present both in expert manual delineations and automated parcellation techniques. Age trajectories in volume and quantitative MR parameters indicate systematic shifts in contrast intensities and an increasing variability with age, associated with changing myelination, iron deposition, and brain atrophy (Draganski et al., 2011; Daugherty and Raz, 2013; Fjell et al., 2013; Keuken et al., 2017). These changes seem only to impact the parcellation accuracy for age groups beyond 60 years and age-matched priors did not provide specific improvements, thus indicating that an explicit modeling of age effects may be required to further improve parcellation quality in elderly populations. These results also point to exercising caution when applying automated parcellation methods to study morphometry in elderly or diseased populations, where measured differences may include biases. They also point out that while global volume and local thickness are indeed affected by such biases, quantitative MRI measures are much more robust. Note that this bias is likely present in many automated methods, although they have not been systematically investigated due to the extensive manual labor required. Interestingly, biases also exist in expert delineations: when the size or shape of a structure is refined in neuroanatomical studies, experts may become more or less conservative in their delineations. Automated methods provide a more objective measure in such case, as the source of their bias is explicitly encoded in the atlas prior delineations and computational model. Important applications of subcortical parcellation also include deep-brain stimulation surgery (Ewert et al., 2018), where the number of structures parcellated by MASSP can

### Table 4. Mean volume and quantitative MRI values for each age group.

<table>
<thead>
<tr>
<th>Age</th>
<th>Str</th>
<th>STN</th>
<th>SN</th>
<th>RN</th>
<th>GPe</th>
<th>GPe</th>
<th>Tha</th>
<th>LV</th>
<th>3V</th>
<th>4V</th>
<th>Amg</th>
<th>ic</th>
<th>VTA</th>
<th>fx</th>
<th>PAG</th>
<th>PPN</th>
<th>Cl</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-40</td>
<td>10656</td>
<td>118</td>
<td>566</td>
<td>253</td>
<td>567</td>
<td>1366</td>
<td>7112</td>
<td>7524</td>
<td>1895</td>
<td>1391</td>
<td>1315</td>
<td>4335</td>
<td>254</td>
<td>1632</td>
<td>250</td>
<td>193</td>
<td>843</td>
</tr>
<tr>
<td>41-60</td>
<td>10572</td>
<td>124</td>
<td>583</td>
<td>256</td>
<td>586</td>
<td>1403</td>
<td>7492</td>
<td>8850</td>
<td>2024</td>
<td>1408</td>
<td>1363</td>
<td>4495</td>
<td>264</td>
<td>1808</td>
<td>255</td>
<td>195</td>
<td>830</td>
</tr>
<tr>
<td>61-80</td>
<td>10734</td>
<td>130</td>
<td>584</td>
<td>260</td>
<td>586</td>
<td>1397</td>
<td>7463</td>
<td>9142</td>
<td>2023</td>
<td>1407</td>
<td>1321</td>
<td>4407</td>
<td>272</td>
<td>1910</td>
<td>259</td>
<td>192</td>
<td>829</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Thickness (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-40</td>
</tr>
<tr>
<td>41-60</td>
</tr>
<tr>
<td>61-80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>qR1 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-40</td>
</tr>
<tr>
<td>41-60</td>
</tr>
<tr>
<td>61-80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>qR2* (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-40</td>
</tr>
<tr>
<td>41-60</td>
</tr>
<tr>
<td>61-80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QSM (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-40</td>
</tr>
<tr>
<td>41-60</td>
</tr>
<tr>
<td>61-80</td>
</tr>
</tbody>
</table>
help neurosurgeons orient themselves more easily, although precise targeting will still require manual refinements, especially in neurodegenerative diseases.

We observed that dilated overlap, that is, the overlap of structures up to one voxel, provided a measure of accuracy largely independent of size, for automated or manual delineations. Imprecision in the range of one voxel in the boundary is to be expected due to partial voluming which impacts Dice overlap. The dilated overlap measure is a better representative of performance and indicates that conservative or inclusive versions of the subcortical regions can be obtained by eroding or dilating the estimated boundary by a single voxel. Such masks may be useful when separating functional MRI signals between neighboring nuclei or when locating smaller features inside a structure. Additionally, the Bayesian estimation framework provides voxel-wise probability values, which can also be used to further weight the contribution of each voxel within a region in subsequent analyses.

In summary, our method provides fast and accurate parcellation for subcortical structures of varying size, taking advantage of the high resolution offered by 7T and the specificity of quantitative MRI. The algorithm is based on an explicit model of structures given in a Bayesian framework and is free of tuning parameters. Given a different set of regions of interest or different populations, new priors can be automatically generated and used as the basis for the algorithm. If more MRI contrasts are available, the method can also be augmented to take them into account. The main requirement for the technique is a set of manual delineations of all the structures of interest in a small group of representative subjects. Performance may further improve with the number of included structures, as the number of distinct interfaces increases, refining in particular the intensity priors. In future works, we plan to include more structures or sub-structures and model the effects of age on the priors. We hope that the method, available in open source, will help neuroscience researchers to include more subcortical regions in their structural and functional imaging studies.

Materials and methods

Data acquisition

Our parcellation method has been developed for the MP2RAGEME sequence (Caan et al., 2019). Briefly, the MP2RAGEME consists of two interleaved MPRAGEs with different inversions and four echoes in the second inversion. Based on these images, one can estimate quantitative MR parameters of R1, R2* and QSM. In this work, we used the following sequence parameters: inversion times TI1,2 = 670 ms, 3675.4 ms; echo times TE1 = 3 ms, TE2,1–4 = 3, 11.5, 19, 28.5 ms; flip angles FA1,2 = 4˚, 4˚; TRGRE1,2 = 6.2 ms, 31 ms; bandwidth = 404.9 MHz; TRMP2RAGE = 6778 ms; SENSE acceleration factor = 2; FOV = 205 x 205 x 164 mm; acquired voxel size = 0.7 x 0.7 x 0.7 mm; acquisition matrix was 292 x 290; reconstructed voxel size = 0.64 x 0.64 x 0.7 mm; turbo factor (TFE) = 150 resulting in 176 shots; total acquisition time = 19.53 min.

T1-maps were computed using a look-up table (Marques et al., 2010). T2*-maps were computed by least-squares fitting of the exponential signal decay over the multi-echo images of the second inversion. R1 and R2* maps were obtained as the inverse of T1 and T2*. For QSM, phase maps were pre-processed using iHARPERELLA (integrated phase unwrapping and background phase removal using the Laplacian) of which the QSM images were computed using LSQR (Li et al., 2014). Skull information was removed through creation of a binary mask using FSL’s brain extraction tool on the reconstructed uniform T1-weighted image and then applied to the quantitative contrasts (Smith, 2002). As all images were acquired as part of a single sequence, no co-registration of the quantitative maps was required (see Figure 9).

Anatomical structure delineations

Manual delineations of subcortical structures were performed by two raters trained by an expert anatomist, according to protocols optimized to use the better contrast or combination of contrasts for each structure and to ensure a consistent approach across raters. The following 17 structures were defined on a group of 10 subjects (average age 24.4, eight female): striatum (Str), thalamus (Tha), lateral, 3rd and 4th ventricles (LV, 3V, 4V), amygdala (Amg), globus pallidus internal segment (GPI) and external segment (GPe), SN, STN, red nucleus (RN), ventral tegmental area (VTA), fornix (fx), internal capsule (ic), periaqueductal gray (PAG), pedunculopontine nucleus (PPN), and claustrum (Cl). Separate masks for left and right hemisphere were delineated except for 3V, 4V, and fx. In the
following the algorithm treats each side separately, resulting in a total of 31 distinct structures (see Figure 1).

**Anatomical interface priors**

In order to inform the algorithm, we built a series of priors derived from the manual delineations. Each subject was first co-registered to a MP2RAGEME anatomical template built from 105 subjects co-aligned with the MNI2009b atlas (Fonov et al., 2011) with the SyN algorithm of ANTs (Avants et al., 2008) using successively rigid, affine, and non-linear transformations, high levels of regularization as recommended for the subcortex (Ewert et al., 2019) and mutual information as cost function.

The first computed prior is a prior of anatomical interfaces, recording the most likely location of boundaries between the different structures, defined as follows. Given two delineated structures $i,j$, let $\phi_i, \phi_j$ be the signed distance functions to their respective boundary, that is, $\phi_i(x)$ is the Euclidean distance of any given voxel to the boundary of $i$, with a negative sign inside the structure. Then we define the interface $B_{ij}$ with the distance function $d_{ij}$:

$$d_{ij}(x) = \min(\phi_i(x), \phi_j(x) - \delta, 0)$$ (1)
where $\delta$ is a scale parameter for the thickness of the interface. These interfaces functions are not symmetrical, as the intensity inside $i$ next to $j$ is generally different from the intensity inside $j$ next to $i$. Based on this definition, the prior for a given interface based on $N$ manual delineations is given by:

$$P(x \in B_{ij}) \sim \frac{1}{\sqrt{2\pi \sigma_{ij}^2(x)}} \exp - \frac{1}{2} \frac{(x - \mu_{ij}(x))^2}{\sigma_{ij}^2(x)}$$

$$\mu_{ij}(x) = \frac{1}{N} \sum_{n \in N} d_{ij,n}(x), \quad \sigma_{ij}(x) = \sqrt{\frac{1}{N} \sum_{n \in N} (d_{ij,n}(x) - \mu_{ij}(x))^2 + \delta}$$

(2)

These probability functions are calculated for all possible configurations including $ij$, which represent the inside of each structure. We thus have a total of $N^2$ functions, but only a few are non-zero at a given voxel $x$, and we may keep only the 16 largest values to account for any number of interfaces in 3D (Bazin et al., 2007). Finally, we need to scale the prior to be globally consistent with the priors below by assuming that the 95th percentile of the highest kept $P(x \in B_{ij})$ values have a probability of 0.95. The scale parameter $\delta$ is set to one voxel, representing the expected amount of partial voluming. The resulting interface prior is shown in Figure 10A.

**Anatomical skeleton priors**

Next, we defined priors for the skeleton of each structure, representing their essential shape regardless of exact boundaries (Blum, 1973). As we are mostly interested in the most likely components of the skeleton or medial axis $S_i$, we follow a simple method to estimate its location:
\[ S_i = \left\{ x \mid |\nabla \phi_i(x)| < \frac{1}{2} \right\} \]  

(3)

We define as \( s_i(x) \) the signed distance function of this discrete skeleton, and define prior probabilities as above:

\[ P(x \in S_i) = \frac{1}{\sqrt{2\pi\sigma^2(x)}} \exp \left( -\frac{1}{2} \frac{(x - \mu(x))^2}{\sigma^2(x)} \right) \]

\[ \mu_i(x) = \frac{1}{N} \sum_{n \in N} s_{in}(x) \quad \sigma_i(x) = \sqrt{\frac{1}{N} \sum_{n \in N} (s_{in}(x) - \mu_i(x))^2 + \delta} \]

(4)

The skeletons are defined inside each structure, which implies \( P(x \in S_i) \leq P(x \in B_{ij}) \). To respect this relationship, we scale \( P(x \in S_i) \) with the same factor as \( P(x \in B_{ij}) \) but use \( \sqrt{P(x \in S_i)} \) when combining probabilities during the estimation stage. The obtained anatomical skeleton priors are given on Figure 10B.

**Interface intensity priors**

While anatomical priors already provide rich information, they are largely independent of the underlying MRI. From the co-aligned quantitative MRI maps and manual delineations, we defined intensity priors for every interface \( ij \), in the form of intensity histograms to ensure a flexible representation of intensity distributions. Given a quantitative contrast \( R_n(x) \), we built a histogram \( H_{ij,n} \) for each subject \( n \) and interface \( ij \). Histograms have 200 bins covering the entire intensity range within a radius of 10 mm from any of the delineated structures. To obtain an average histogram, we combine each histogram with a weighting function \( w_{ij}(x) \) giving the likelihood of the subject’s intensity measurement compared to the group:

\[ w_{ij,n}(x) = P(x \in B_{ij}) \frac{1}{\sqrt{2\pi\sigma^2(x)}} \exp \left( -\frac{1}{2} \frac{(x - \mu_R(x))^2}{\sigma_R(x)^2} \right) \]

(5)

where \( \mu_R(x) \) is the median of the \( R_n(x) \) values at \( x \), and \( \sigma_R(x) \) is 1.349 times the inter-quartile range of \( R_n(x) \). These are robust estimators of the mean and standard deviation, used here to avoid biases by intensity outliers. To further combine the R1, R2*, and QSM contrasts we take the geometric mean of the histogram probabilities:

\[ H_{ij}(x) = \prod_n H_{ij,n}(R(x))^{1/3} \]

**Global volume priors**

The last type of priors extracted from manual delineations are volume priors for each of the structure. Here, we assume a log-normal distribution for the volumes \( V_i \) and simply estimate the mean \( \mu_{V,i} \) and standard deviation \( \sigma_{V,i} \) of \( \log V_{i,n} \) over the subjects.

**Voxel-wise posterior probabilities**

When parcellating a new subject, we first co-register its R1, R2*, and QSM maps jointly to the template and use the inverse transformation to deform the anatomical priors into subject space. Then we derive voxel-wise posteriors as follows:

\[ P(x \in B_{ij} | R(x)) \sim P(x \in B_{ij}) H_{ij}(x) \text{ if } i \neq j \]

and

\[ P(x \in B_{ij} | S_i(x), R(x)) \sim \max(P(x \in B_{ij}), P(x \in S_i)^{1/2}) H_{ij}(x) \]

(6)

Once again we should compute all possible combinations, but due to the multiplication of the priors we can restrict ourselves to the 16 highest probabilities previously estimated. To balance the contribution of the anatomical priors and the intensity histograms, we also need to normalize the intensity priors sampled on the subject’s intensities. We use the same approach, namely assuming that the 95th percentile of the highest kept \( H_{ij}(x) \) values have a probability of 0.95, separately for each contrast. The voxel-wise parcellation and posteriors obtained are shown in Figure 11A.
Markovian diffusion

The voxel-wise posteriors are independent from each other and do not reflect the continuous nature of the structures. The next step is to combine information from neighboring voxels. We define a sparse Markov Random Field model for the posteriors:

Figure 11. Successive parcellation results: (A) voxel-wise posteriors and parcellation, (B) diffused posteriors and parcellation, (C) topology-corrected posteriors and final region-growing parcellation.
Neuroscience

Bazin et al.

Research article

\[ P(x \in B_{ij}|R,S,C) = \sum_{y \in C(x)} P(y \sim x|R)P(y \in B_{ij}|R,S,C) \]  

with \( P(y \sim x|R) = \prod_k \exp\left(-(R(y) - R(x))^2 / 2\sigma_k^2\right) \), where \( \sigma_k \) is the median of the standard deviations \( \sigma_{ij,k} \) of the contrast histograms \( H_{ij}(R(x)) \). The neighborhood \( C(x) \) is defined as \( x \) itself and the four 26-connected neighboring voxels with highest probability \( P(y \sim x|R) \), thus representing the neighbors most likely to be connected to \( x \). The model is similar to a diffusion process and can be estimated with an iterated conditional modes (ICM) approach, updating sequentially the probabilities (Bazin and Pham, 2007):

\[ P(x \in B_{ij}|R,S,C) \leftarrow \sum_{y \in C(x)} P(y \sim x|R)P(y \in B_{ij}|R,S,C) \]

from the initial voxel-wise posteriors until the ratio of changed parcellation labels decreases below 0.1, typically within 50–80 iterations. The diffused probabilities and parcellation are shown in Figure 11B.

Topology correction

The final step of the parcellation algorithm takes a global view of the individual structures, growing from the highest posterior values inside toward the boundaries. This region growing approach makes the implicit assumption that posterior maps should be monotonically decreasing from inside to outside, which is not necessarily the case. Therefore, we perform first a topology correction step on the individual structure posteriors \( P(x \in i|R,B,S,C) = \max_{ij} P(x \in B_{ij}|R,S,C) \) with a fast marching algorithm (Bazin and Pham, 2007). While the corrected posterior is very similar to the original one (see Figure 11C), it ensures that all regions obtained by growing to a threshold have spherical object topology.

Anatomical region growing

Last, we turn the posteriors into optimized parcellations, by growing them concurrently (to avoid overlaps) until the target volume for each structure is reached. Given the volume \( V_i(R,B,S,C) \) of the parcellation of the diffused and topology-corrected posteriors, we define the following target volume:

\[ \tilde{V}_i = P(V_i|\mu_{V,i},\sigma_{V,i})V_i + (1 - P(V_i|\mu_{V,i},\sigma_{V,i}))\exp\mu_{V,i} \]

taking a weighted average of the volume estimated from the data and the prior volume. This approach ensures that even in extreme cases where some structures have low posteriors, they are still able to grow to a plausible size. The region growing algorithm is driven from the most likely voxels, defined as \( P(x \in i|R,B,S,C) = \max_{j \in i} P(x \in j|R,B,S,C) \), and further modulated to follow isocontours of the skeleton prior:

\[ P(x \sim y) \sim P(y \in i|R,B,S,C) - \max_{j \in i} P(y \in j|R,B,S,C) \]

Directionality of internal structures is a useful tool for understanding mechanical function in bones (Maquer et al., 2015). Here, we adapt this concept by using the skeleton isocontours as a representation of internal directionality, maintaining the intrinsic shape of structures. Thus, voxels with highest probability compared to the other structures and with similar distance to the internal skeleton are preferentially selected. The final parcellation is given in Figure 11C.

Validation metrics

To validate the method against manual expert delineations, we compared the MASSP results and the expert delineations with the following three measures:

1. The Dice overlap coefficient (Dice, 1945) \( D(A,B) = \frac{2A \cap B}{A + B} \), which measures the strict overlap between voxels in both delineation;
2. The dilated overlap coefficient $dD(A, B) = \frac{A[d(B)] + B[d(A)]}{A + B}$, where $d(\cdot)$ is a dilation of the delineation by one voxel, which measures the overlap between delineations allowing for one voxel of uncertainty;

3. The average surface distance $asd(A, B)$, measuring the average distance between voxels on the surface boundary of the first delineation to the other one and reciprocally, which measures the distance between both delineations.

We computed all three measures for the manual delineations from the two independent raters, as well as the ratio of overlaps (automated over manual) and distances (manual over automated) to compare both performances, as detailed in the Results section.

Comparisons with other automated methods
To assess the performance of MASSP compared to existing parcellation tools, we ran Freesurfer (Fischl et al., 2002), FSL FIRST (Patenaude et al., 2011) and a multi-atlas registration approach (co-registering 9 of the 10 manually delineated subjects on the remaining one with ANTs [Avants et al., 2008] and labeling each structure by majority voting, similarly to the MAGeT Brain approach of Chakravarty et al., 2013). Freesurfer and FIRST were run on the skull-stripped R1 map, while the multi-atlas approach used all three R1, R2*, and QSM contrasts. All methods were compared in terms of Dice overlap, dilated overlap and average surface distance. We also assessed the presence of a systematic volume bias, defined as the average of the signed difference of the estimated structure volume to the manually delineated volume, normalized by the manually delineated volume.

Application to new MRI contrasts
Before applying MASSP to unseen contrasts, we need to convert its intensity prior histograms $H_{ij,R}$ to the new intensities. In order to perform this mapping, we first created a groupwise median of the HCP subjects, by co-registering every subject to the MASSP template using ANTs with non-linear registration and both T1w, T2w contrasts matched to the template’s R1 and R2* maps. The histogram bins are then updated as follows:

$$H_{\text{bin},ij,R} = \sum_{x \in B_{ij}} P(x \in B_{ij})H_{ij,R1}H_{ij,R2}H_{ij,QSM}$$

adding the joint probability of the quantitative contrasts weighted by their importance for each interface to define the new intensity histograms. This model is essentially projecting the joint likelihood of the MASSP contrasts onto the new contrasts, assuming that the co-registration between the two is accurate enough. With these new histograms, we compared the test-retest reliability and overall agreement of MASSP with Freesurfer parcellations included in the HCP pre-processed data set.

Measurement of structure thickness
Finally, when comparing derived measures obtained over the lifespan with MASSP compared to manual delineations, we explored the utility of a shape thickness metric, based on the medial representation. Given the signed distance function $\varphi_i$ of the structure boundary and $s_i$ of the structure skeleton, the thickness is given by:

$$th_i(x) = 2(s_i(x) - \varphi_i(x))$$

Like in cortical morphometry, thickness is a local measure, defined everywhere inside the structure, and expected to provide additional information about anatomical variations. Indeed, a similar measure of shape thickness has recently been able to highlight subtle anatomical changes in depression (Ho et al., 2020).

Software implementation
The proposed method, Multi-contrast Anatomical Subcortical Structure Parcellation (MASSP), has been implemented as part of the Nighres toolbox (Huntenburg et al., 2018), using Python and Java for optimized processing. The software is available in open source from (release-1.3.0) and . A
complete parcellation pipeline is included with the Nighres examples. Computations take under 30 min per subject on a modern workstation.

Acknowledgements
We thank Josephine Groot, Nikita Berendonk, Nicky Lute for their help collecting the AHEAD database, and Wietse van der Zwaag and Matthijs Caan for their help in setting up the MP2RAGEME sequence. We also thank Steven Miletic and Dagmar Timmann for stimulating discussions around this topic, and all undergraduate students who contributed to the manual delineations. This work was supported by a NWO Vici grant (BF) and a NWO STW grant (AA, BF). HCP data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Additional information

**Funding**

<table>
<thead>
<tr>
<th>Funder</th>
<th>Grant reference number</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nederlandse Organisatie voor Wetenschappelijk Onderzoek</td>
<td>VICI</td>
<td>Birte U Forstmann</td>
</tr>
<tr>
<td>Nederlandse Organisatie voor Wetenschappelijk Onderzoek</td>
<td>STW</td>
<td>Anneke Alkemade Birte U Forstmann</td>
</tr>
</tbody>
</table>

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

**Author contributions**
Pierre-Louis Bazin, Conceptualization, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - original draft; Anneke Alkemade, Data curation, Formal analysis, Validation, Investigation, Methodology, Writing - review and editing; Martijn J Mulder, Resources, Data curation, Software, Methodology, Writing - review and editing; Amanda G Henry, Conceptualization, Methodology, Writing - review and editing; Birte U Forstmann, Conceptualization, Resources, Formal analysis, Supervision, Funding acquisition, Project administration, Writing - review and editing

**Author ORCIDs**
Pierre-Louis Bazin [https://orcid.org/0000-0002-0141-5510](https://orcid.org/0000-0002-0141-5510)
Anneke Alkemade [https://orcid.org/0000-0002-3234-353X](https://orcid.org/0000-0002-3234-353X)
Amanda G Henry [https://orcid.org/0000-0002-2923-4199](https://orcid.org/0000-0002-2923-4199)
Birte U Forstmann [https://orcid.org/0000-0002-1005-1675](https://orcid.org/0000-0002-1005-1675)

**Ethics**

Human subjects: Informed consent and consent to publish, including consent to publish anonymized imaging data, was obtained for all subjects. Ethical approval was obtained with the University of Amsterdam Faculty of Social and Behavioral Sciences LAB Ethics Review Board, with ERB number 2016-DP-6897.

**Decision letter and Author response**

Decision letter [https://doi.org/10.7554/eLife.59430.sa1](https://doi.org/10.7554/eLife.59430.sa1)
Author response [https://doi.org/10.7554/eLife.59430.sa2](https://doi.org/10.7554/eLife.59430.sa2)
Data availability

The tool presented in this article is available in open source on Github (https://github.com/nighres/nighres). The atlases necessary to run the algorithm have been deposited on the University of Amsterdam FigShare (https://doi.org/10.21942/uva.12074175.v1 and https://doi.org/10.21942/uva.12301106.v2). A single sample subject data set has been deposited on the University of Amsterdam FigShare (https://doi.org/10.21942/uva.12280316.v2). All the measurements used to generate the figures included in the article have been deposited on the University of Amsterdam FigShare (https://doi.org/10.21942/uva.12452444.v1).

The following previously published datasets were used:

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Year</th>
<th>Dataset title</th>
<th>Dataset URL</th>
<th>Database and Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alkemade A, Mulder MJ, Groot JM, Isaacs BR, van Berendonk N, Lute N, Isherwood SJ, Bazin P-L, Forstmann BU</td>
<td>2020</td>
<td>The Amsterdam Ultra-high field adult lifespan database (AHEAD): A freely available multimodal 7 Tesla submillimeter magnetic resonance imaging database</td>
<td><a href="https://doi.org/10.21942/uva.10007840.v1">https://doi.org/10.21942/uva.10007840.v1</a></td>
<td>FigShare / University of Amsterdam / Amsterdam University of Applied Sciences, 10.21942/uva.10007840.v1</td>
</tr>
<tr>
<td>Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K</td>
<td>2013</td>
<td>The WU-Minn Human Connectome Project</td>
<td><a href="https://www.humanconnectome.org/study/hcp-young-adult">https://www.humanconnectome.org/study/hcp-young-adult</a></td>
<td>WU-Minn HCP Retest Data, hcp-young-adult</td>
</tr>
</tbody>
</table>

References


da de Hollander G, Keuken MC, van der Zwaag W, Forstmann BU, Trampe R. 2017. Comparing functional MRI protocols for small, iron-rich basal ganglia nuclei such as the subthalamic nucleus at 7 T and 3 T. *Human Brain Mapping* **38**:3226–3248. DOI: https://doi.org/10.1002/hbm.23586, PMID: 28345164


Yeatman JD, Wandell BA, Mezer AA. 2014. Lifespan maturation and degeneration of human brain white matter. *Nature Communications* **5**:4932. DOI: https://doi.org/10.1038/ncomms5932, PMID: 25230200
