Structural MRI Based Parcellation of The Human Brain Using Spatial Hierarchical Clustering

Tobias Tietz (✉ tobias.tietz@uni-duesseldorf.de)  
Heinrich-Heine University  https://orcid.org/0000-0001-7009-0708

Simon Eickhoff  
Institute of Neuroscience and Medicine (INM-1, INM-7)

Holger Schwender  
Heinrich-Heine University

Research Article

Keywords: structural MRI, spatial hierarchical clustering, ensemble clustering, Voxel-based morphometry, brain parcellation, clustering stability

Posted Date: January 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1254932/v1

License: ☑️ Ⓦ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Structural MRI based parcellation of the human brain using spatial hierarchical clustering

Tobias Tietz\textsuperscript{1*}, Simon B. Eickhoff\textsuperscript{2,3,4} and Holger Schwender\textsuperscript{1}

\textsuperscript{1*}Mathematical Institute, Heinrich-Heine University, Düsseldorf, Germany.
\textsuperscript{2}Institute of Neuroscience and Medicine (INM-1, INM-7), Research Centre Juelich, Germany.
\textsuperscript{3}Institute of Systems Neuroscience, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany.
\textsuperscript{4}Institute of Clinical Neuroscience and Medical Psychology, Heinrich-Heine University, Düsseldorf, Germany.

*Corresponding author(s). E-mail(s):
tobias.tietz@uni-duesseldorf.de;
Contributing authors: simon.eickhoff@uni-duesseldorf.de; holger.schwender@uni-duesseldorf.de;

Abstract
Parcellating the human brain into neurobiologically meaningful regions serves two (key) purposes, i.e. to gain neurobiological insight on the modules of the brain and to achieve dimensionality reduction for large-scale studies. Data-driven parcellation approaches based on magnetic resonance imaging (MRI) mainly differ from each other by the modalities that they are derived from and the parcellation algorithm. Here we extensively evaluate parcellation approaches which are based on one specific modality, i.e. grey matter volume deduced from T1-weighted structural MRI scans. The newly proposed spatial hierarchical variable clustering method SPARTACUS is employed as parcellation algorithm and its performance is compared with other spatial hierarchical agglomerative clustering
(SHAC) algorithms as well as with spatial spectral clustering on both simulated data and a single-site structural MRI data set of older adults using two metrics, i.e. clustering quality and clustering stability across subsamples. Our analysis reveals that SPARTACUS and Ward’s SHAC outperform the other competing methods. Using spatial ensemble clustering methods we could further improve the quality of the parcellations. Performing parameter tuning of the granularity (considering 2-1000 brain regions) we identify multiple interesting numbers of brain regions, where the corresponding parcellations may reflect different levels of brain organization. Finally, we observe that our parcellations converge well above chance with other structural MRI based parcellations but hardly better or even worse than chance with parcellations from another modality.

**Keywords:** structural MRI, spatial hierarchical clustering, ensemble clustering, Voxel-based morphometry, brain parcellation, clustering stability

---

**Acknowledgements**

Computational infrastructure and support were provided by the Centre for Information and Media Technology at Heinrich-Heine University Düsseldorf.

---

**1 Introduction**

The human brain is clearly the most complex organ of the human body. In order to get a better understanding on how the brain works, a crucial and widely accepted concept is that the brain is organized into spatially contiguous, specialized brain regions (cortical areas and subcortical nuclei), which are inter-connected by large-scale networks (Eickhoff et al, 2018a). These brain regions should be of large within homogeneity and between heterogeneity with respect to different neurobiological modalities, where the boundaries should be consistent among different modalities (Eickhoff et al, 2018a). While the first modalities are histological-based, e.g., investigating cyto- and myeloarchitecture in postmortem brains, the development of high-quality magnetic...
resonance imaging (MRI) techniques gives rise to a variety of modalities measured in vivo, e.g., functional specialization, functional/structural connectivity or grey matter volume.

An accurate parcellation into neurobiologically meaningful regions does not only help to get a better understanding of the topology and function of the brain. It can also be used to communicate neurobiological results, e.g., task-based activation patterns, or to perform data compression for large-scale studies (Glasser et al, 2016; Eickhoff et al, 2018a). However, the human brain is a highly complex structure that evidently exhibits deviating patterns among different neurobiological modalities. This makes the creation of a brain atlas challenging and it is still unclear, whether a universal brain atlas exists at all (Eickhoff et al, 2018b). Even though there probably will not be a final brain atlas in the nearby future, the concept of a brain atlas is one of the most important concepts in the field of neuroimaging for describing and analyzing brain organization (Eickhoff et al, 2018a).

A large number of different brain atlases have been proposed in the literature, which mainly differ from each other by the modalities that they are derived from and by the parcellation approach (see, e.g., Moghimi et al (2021) for a detailed review). Parcellation approaches can be classified into manual, partly automated and fully automated approaches (Glasser et al, 2016). While the former two involve manual labelling by expert neuroanatomists, e.g., the cortical areas of the Desikan-Killany atlas (Desikan et al, 2006) are manually identified or the areas of a multimodal atlas by Glasser et al (2016) are delineated using an objective semi-automated neuroanatomical approach, the latter rely exclusively on computer algorithms and can be further divided into local boundary mapping and global clustering approaches (Eickhoff et al, 2018b). Local boundary mapping is, e.g., applied by (Gordon et al, 2016) (resting-state...
Schaefer et al (2018) employ a gradient-weighted Markov Random Field (gwMRF) method (resting-state fMRI), which is a hybrid method integrating both boundary mapping and clustering. Examples of clustering algorithms applied to MRI data are contiguity constrained spectral clustering (Craddock et al, 2012) (resting-state fMRI), a combination of region growing and spatially constrained hierarchical Ward clustering (Blumensath et al, 2013) (resting-state fMRI), a combination of principal component analysis (PCA) for feature reduction and K-means clustering (Thirion et al, 2014) (task-based fMRI) or orthonormal projective non-negative matrix factorization based clustering (Varikuti et al, 2018; Sotiras et al, 2015) (structural MRI).

Clustering based parcellation approaches are extensively evaluated, e.g., in application to task-based fMRI data (Thirion et al, 2014) or resting-state fMRI data (Arslan et al, 2018). While the analyses by Thirion et al (2014) reveal that spatially constrained hierarchical Ward clustering (Ward Jr, 1963; Carvalho et al, 2009) outperforms the other methods under consideration with respect to reproducability and accuracy, the experiments by Arslan et al (2018) could not identify a favored clustering method with respect to the considered evaluation measures. However, Arslan et al (2018) characterize the performance of spatially constrained agglomerative hierarchical clustering (SHAC) to reside in-between the performance of K-means (low reproducability and high accuracy) (Lloyd, 1982; MacQueen, 1967) and spatially constrained spectral clustering (high reproducability and low accuracy) (Shi and Malik, 2000; Ng and Han, 2002; Yuan et al, 2015), i.e. it generates spatially contiguous parcels, entailing an improved reproducability of the resulting parcellations, while simultaneously achieving a fairly high accuracy. However, only a few parcellations (Varikuti et al, 2018; Sotiras et al, 2015) have been derived by
applying clustering methods to structural MRI data, where an extensive investigation of the performance of SHAC methods in application to structural MRI data is missing in the literature up till now.

Therefore, in this article, our goal is to derive brain parcellations using SHAC methods based on one specific modality, namely T1-weighted structural MRI data. For this, since the objects to be clustered are voxels, i.e. variables, we propose a new spatially constrained hierarchical agglomerative variable clustering procedure, referred to as SPARTACUS, introducing contiguity constraints into a hierarchical variable clustering method by Vigneau and Qannari (2003). SPARTACUS’s performance (with respect to clustering quality and clustering stability) is extensively evaluated and compared with other SHAC methods as well as with spatially constrained spectral clustering (SSPEC) in application to simulated data and the 1000BRAINS data set (Caspers et al., 2014) including structural scans of older adults. The results of this investigation may provide an additional view on human brain organization and improve the understanding of the mechanisms of the human brain.

In order to improve clustering quality, we employ spatial ensemble clustering (SEC) methods. Ensemble clustering methods (Monti et al., 2003; Strehl and Ghosh, 2002) combine multiple parcellations in a cluster ensemble and use a consensus function to obtain a final ensemble parcellation. E.g., Bellec et al. (2010) propose an ensemble clustering procedure called bootstrap analysis of stable clusters for resting-state fMRI data.

A critical challenge for any parcellation procedure is to select the number of brain regions. Since the human brain is organized into multiple levels, the correct number of brain regions might not exist. Instead, it is more likely that different numbers of brain regions indicate different levels of brain organization (Eickhoff et al., 2018b). Many studies have identified interesting numbers of
clusters by calculating internal validation scores from the information within
the input data (Kelly et al, 2010; Pravitasari et al, 2020). Clustering sta-
tility (Von Luxburg, 2010) is another procedure employed by many studies
to identify interesting numbers of clusters (Van Oort et al, 2018; Yeo et al,
2011). Clustering stability builds upon the idea that biological truth should be
reflected by parcellations that are stable across different subsamples (Eickhoff
et al, 2018b). We employ both a subsample based internal validation approach
and a subsample based clustering stability approach to search for interesting
numbers of brain regions.

This article is structured as follows. First, the relevant spatial clustering
methods and the methods to identify interesting numbers of brain regions
are presented. Subsequently, we evaluate and compare the performance of
these methods in a simulation study. Moreover, we apply these methods to
the 1000BRAINS data set including structural brain scans of older subjects.
As further quality feature, the convergence of the final brain parcellations
with existing anatomical atlases as well as alternative atlases generated by
(semi-)algorithmic approaches based on MRI data is analyzed.

2 Materials and Methods

2.1 1000BRAINS data set

A structural MRI data set is obtained from the 1000BRAINS study (Caspers
et al, 2014). A range of sequences are considered in the 1000BRAINS study
which are all scanned at a single site on a 3 Tesla MR scanner, using the same
imaging protocol for each subject. Hence, reliable and homogeneous data are
produced for a large number of subjects. Here, an anatomical 3D T1-weighted
MP RAGE (Mugler III and Brookeman, 1990) sequence is considered from this
study with the following scanning parameters (Caspers et al, 2014): Repetition
time = 2.25 s, echo time = 3.03 ms, inversion time = 900 ms, field of view = 
256 × 256 mm², flip angle = 9° and voxel resolution = 1.5 × 1.5 × 1.5 mm³. This 
sequence includes structural scans of 693 older subjects (age: 55-75 years; 47% 
females) (Varikuti et al, 2018). The VBM8 toolbox (http://www.neuro.uni-
jena.de/vbm8) is employed to preprocess the structural MRI data as described 
in Varikuti et al (2018). The resulting grey matter images consist of 344,383 
voxels and are saved in a 693 × 344383 data matrix. However, since two subjects 
have exactly the same voxel intensity values, these two subjects are removed 
from the data set. Moreover, since five voxels are not spatially contiguous 
to the other voxels, also these voxels are removed from the data set. Hence, 
further analysis is based on a 691 × 344378 data matrix. In accordance with 
Varikuti et al (2018) this data set is referred to as the 1000BRAINS data set.

2.2 Simulation study

A simulation study is conducted in which 3D images with known cluster labels 
are simulated. In order to consider the multilevel organization of the human 
brain, 3D images are simulated with nested clusters, i.e. each larger cluster 
is further subdivided into two smaller clusters. Hereby, voxels from the same 
smaller cluster are simulated to have a larger pairwise correlation than voxels 
from the same larger cluster (but from different smaller clusters) which, again, 
have a larger correlation than voxels from different larger clusters. Since the 
performance of the spatial clustering methods under investigation might be 
influenced by whether the true parcellation is balanced or unbalanced, sim-
ulations with balanced parcellations, i.e. all clusters are of equal size, and 
unbalanced parcellations, i.e. the clusters are of three different sizes, are con-
sidered. In the following, we briefly describe the setup of the simulation study. 
We present a more detailed description of the setup in Section A.1 of the 
Online Resource.
All simulated 3D images in all simulations are on a cubic grid of $18 \times 18 \times 18$ voxels. Hence, each simulated 3D image consists of $V = 18 \cdot 18 \cdot 18 = 5832$ voxels. Two different settings are considered. In the balanced setting (B), 3D images with equally sized clusters and, in the unbalanced setting (U), 3D images with clusters of three different sizes are simulated.

In a first step, we determine one true parcellation for the balanced setting and one true parcellation for the unbalanced setting. In both settings the true parcellation consists of 27 larger clusters which again split up into two equally sized smaller clusters. Hereby, all these clusters are spatially contiguous. In the balanced setting, the 27 larger clusters are all of size $6 \times 6 \times 6$ and, following, all 54 smaller clusters are of size $6 \times 6 \times 3$. In the unbalanced setting, 9 larger clusters are of size $3 \times 6 \times 6$, 9 larger clusters are of size $6 \times 6 \times 6$ and 9 larger clusters are of size $9 \times 6 \times 6$ which entails that 18 smaller clusters are of size $3 \times 6 \times 3$, 18 smaller clusters are of size $6 \times 6 \times 3$ and 18 smaller clusters are of size $9 \times 6 \times 3$. A visualization of the true parcellations is shown in Figure 1.

In a second step, the voxel intensities are simulated from a multivariate normal distribution. Since 3D images should be simulated, where each image consists of $V = 5832$ voxels, let $\mathbf{Y}$ be a $V$-dimensional random vector which follows a multivariate normal distribution $\mathbf{Y} \sim \mathcal{N}_V(\mathbf{0}, \Sigma)$. Obviously, the expectation vector does not include any cluster information, i.e. the clusters are...
simulated solely via the covariance matrix $\mathbf{\Sigma}$. In order to allow a better interpretability of the clustering results, we choose the covariance matrix to be a correlation matrix with only two different correlation values $1 > \sigma_S \geq \sigma_L \geq 0$. Hereby, $\sigma_S$ is the correlation between voxels from the same smaller cluster and $\sigma_L$ is the correlation between voxels that are in the same larger cluster but not in the same smaller cluster. Moreover, the correlation between voxels from different larger clusters is chosen to be zero. Hence, $\mathbf{\Sigma} = \mathbf{\Sigma}^{(t)}(\sigma_S, \sigma_L)$, where $t \in \{B, U\}$ indicates the setting. In Section A.1 of the Online Resource we show that $\mathbf{\Sigma}^{(t)}(\sigma_S, \sigma_L)$ is positive definite. Thus, $\mathbf{\Sigma}^{(t)}(\sigma_S, \sigma_L)$ is a valid correlation matrix.

In order to analyze whether the clustering accuracy increases for stronger pronounced clusters, i.e. for clusters of higher correlated voxels, three different simulation scenarios are considered for each setting. In the first, second or third scenario of each setting strongly (S), moderately (M) or weakly (W) pronounced clusters are simulated by sampling 3D images from a multivariate normal distribution with covariance matrix $\mathbf{\Sigma}^{(t)}(0.2, 0.1)$, $\mathbf{\Sigma}^{(t)}(0.1, 0.05)$ or $\mathbf{\Sigma}^{(t)}(0.05, 0.025)$, respectively. In each of the six simulation scenarios $H = 25$ data sets are sampled, where each data set consists of $N = 100$ 3D images. Let in the following SimBS, SimBM and SimBW denote the first, second and third simulation scenario from the balanced setting, and let SimUS, SimUM and SimUW denote the first, second and third simulation scenario from the unbalanced setting, respectively. Since structural MR images have only positive intensity values, each data matrix is min-max normalized to be in $[0, 1]^{N \times V}$. The normalized images are smoothed using a multidimensional Gaussian filter (Jones et al., 2005) with full width at half maximum (FWHM) being equal to two. Hereby, the 3D images are extended at the borders by reflecting about the edge of the border voxels. Finally, another optional preprocessing step is
to standardize the columns of the data matrices to have zero mean and unit variance.

### 2.3 Notation

Mathematically, a structural MRI data set consists of two matrices, i.e. a data matrix $X = (x_1, \ldots, x_V) \in \mathbb{R}^{N \times V}$, where $N$ is the number of subjects and $V$ is the number of voxels, as well as a coordinate matrix $Z \in \mathbb{N}_0^{V \times 3}$, where the spatial coordinates of voxel $x_j, j = 1, \ldots, V$, are stored in the $j$-th row $z_j^T = (z_{j1}, z_{j2}, z_{j3})$ of $Z$. The entries of $X$ are positive values representing grey matter volumes. Thus, each row-vector of $X$ is a (preprocessed) structural MR brain image consisting of $V$ grey matter volumes whose spatial locations are stored in $Z$. The columns of $X$ can be centered or standardized. Two voxels $x_j$ and $x_\ell$ are (face touching) neighbors, if these two voxels share a common face, i.e. if $|z_{j1} - z_{\ell1}| + |z_{j2} - z_{\ell2}| + |z_{j3} - z_{\ell3}| = 1$.

### 2.4 Clustering methods for structural brain parcellation

#### 2.4.1 Spatial hierarchical agglomerative clustering

In the beginning of any SHAC algorithm, each voxel forms its own cluster. A sparse $V \times V$ dissimilarity matrix including all pairwise distances between neighboring voxels is determined using a distance metric. We consider the Euclidean distance and the squared correlation distance, which, for two voxels $x_j$ and $x_\ell$, is calculated as

\[
d_{\text{corr}}(x_j, x_\ell) = 1 - \text{corr}(x_j, x_\ell)^2,
\]

where \(\text{corr}(x_j, x_\ell)\) is the Pearson correlation coefficient. In each iteration, the two closest neighbor clusters are merged and the agglomeration method is used to update the dissimilarity matrix, i.e. to calculate the distance of the newly
formed cluster to its neighbor clusters. Two clusters are neighbors, if at least one voxel from the one cluster is a neighbor of at least one voxel from the other cluster. The distance between two clusters is quantified using Euclidean and squared correlation distance based average linkage (SHAC$_{AL, \text{Eucl}}$ and SHAC$_{AL, \text{corr}}$) as well as Euclidean distance based Ward’s minimal variance method (SHAC$_{\text{Ward}}$). This procedure is repeated until all voxels are in the same cluster or until there are no further neighbor clusters. The latter occurs, if not all voxels in the data set belong to one contiguous region. By successively splitting up the last aggregation, a parcellation with any number of clusters can be obtained from the merging hierarchy (Aggarwal and Reddy, 2014). Note that all these SHAC methods are already employed, e.g., by Carvalho et al (2009) to form clusters of Brazilian municipalities based on a set of social-economic characteristics.

We employ SHAC algorithms because of three main reasons. Firstly, spatial constraints can be easily embedded into the hierarchical clustering framework. On the one hand, this property guarantees to obtain spatially connected brain regions. On the other hand, the memory problem, which occurs when calculating the dissimilarity matrix for large data sets, is solved, since the dissimilarity matrix becomes a sparse matrix over the whole run of the algorithm. Hence, by considering spatial constraints, hierarchical clustering algorithms become applicable to structural MRI data at all. Secondly, once the dendrogram is calculated, a parcellation with any number of clusters can be generated in no time. This property makes SHAC algorithms very time-effective for the task of comparing parcellations with different numbers of clusters. Thirdly, different choices for the distance metric and the agglomeration method can be easily compared that are capable to identify a variety of cluster shapes.
In the task of 3D image clustering, the entities to be clustered are voxels, i.e. the features/variables. Vigneau and Qannari (2003) introduce a hierarchical clustering method called clustering of variables around latent components which is especially designed to cluster variables. The idea is to organize highly correlated voxels into clusters, such that for each cluster the within-cluster variance of the voxels in that cluster is well explained by a single latent variable, called latent component. In the agglomeration step of the hierarchical algorithm, the distance between two clusters is the overall loss in explained total variance by all clusters latent components that would be caused, if these two clusters are merged. We modify this method such that only neighbor clusters can be merged in each agglomeration step and call the resulting method SPARTACUS (spatial hierarchical agglomerative variable clustering).

More precisely, let the columns of the data matrix $X$ be centered, and preferably, but not necessarily, standardized. For a fixed number $K$, the goal of the SPARTACUS method is to find the parcellation $C_K = \{C_1, \ldots, C_K\}$ of the voxels into $K$ clusters with corresponding latent components $c_1, \ldots, c_K \in \mathbb{R}^N$ which maximizes

$$T = (N - 1) \sum_{k=1}^{K} \sum_{x_j \in C_k} \text{Cov}(x_j, c_k)^2$$

under the constraints $c_k^T c_k = 1$ (Vigneau and Qannari, 2003) and that all voxels belonging to cluster $C_k$, $k = 1, \ldots, K$, are spatially contiguous. Cov($x_j, c_k$) is the covariance between $x_j$ and $c_k$. Let further $X_k \in \mathbb{R}^{N \times C_k}$ be the matrix which columns consist of the voxels $x_j \in C_k$, i.e. $X_k$ is the data matrix of cluster $C_k$. The latent component of cluster $C_k$ is defined as the first normalized principal component of $X_k$, which, since $X_k$ is centered, is equal to the first
normalized eigenvector of $\frac{1}{N-1} X_k X_k^T$ (Vigneau and Qannari, 2003). Vigneau and Qannari (2003) show that the maximization criterion can be rewritten as

$$T = \sum_{k=1}^{K} \lambda_1^{C_k},$$

where $\lambda_1^{C_k}$ is the first eigenvalue of cluster $C_k$, i.e. the first eigenvalue of $\frac{1}{N-1} X_k X_k^T$.

Thus, at iteration $\kappa$ of the SPARTACUS algorithm, $\kappa = 1, ..., V - A$, where $A$ is the number of contiguous brain regions in the data set, the aggregation of two adjacent clusters $C_k^{(\kappa-1)}$ and $C_m^{(\kappa-1)}$, $k \neq m$, $k, m \in \{1, \ldots, V - \kappa + 1\}$, results in a variation in $T$ of

$$T_{\kappa-1} - T_\kappa = \lambda_1^{C_k^{(\kappa-1)}} + \lambda_1^{C_m^{(\kappa-1)}} - \lambda_1^{C_k^{(\kappa-1)} \cup C_m^{(\kappa-1)}}.$$

Furthermore, Vigneau and Qannari (2003) show that

$$\lambda_1^{C_k^{(\kappa-1)}} + \lambda_1^{C_m^{(\kappa-1)}} - \lambda_1^{C_k^{(\kappa-1)} \cup C_m^{(\kappa-1)}} \geq 0.$$

Hence, any cluster aggregation causes the criterion $T$ to decrease. The idea behind the SPARTACUS method is to merge at each iteration the two neighbor clusters that cause the smallest decrease in $T$. Consequently, the SPARTACUS method uses

$$D_{\text{SPARTACUS}} (C_k, C_m) = \lambda_1^{C_k} + \lambda_1^{C_m} - \lambda_1^{C_k \cup C_m}$$

as distance measure between neighbor clusters $C_k$ and $C_m$ in a SHAC algorithm.

2.4.3 Spatial spectral clustering

As comparing method to SPARTACUS and the SHAC methods, we consider SSPEC. Spectral clustering algorithms (Shi and Malik, 2000; Ng and Han,
2002; Von Luxburg, 2007) are one of the most popular modern clustering algorithms, mainly because they are easy to implement, can find clusters of arbitrary nonlinear shapes and often outperform popular clustering algorithms such as $K$-means. Moreover, (spatial) constraints can be easily incorporated. The SSPEC algorithm we employ consists of three steps. First, a spatially constrained adjacency matrix $W = (w_{j\ell})_{j,\ell=1,\ldots,V}$ is calculated by

$$w_{j\ell} = \begin{cases} \exp \left( -\frac{\|x_j - x_\ell\|^2}{N} \right), & \text{if } s_{j\ell} = 1 \\ 0, & \text{otherwise} \end{cases}$$

where $\| \cdot \|^2$ is the squared Euclidean distance and $s_{j\ell} = 1$, if and only if $x_j$ and $x_\ell$ are neighbors or have a common neighbor (compare the BSSC method of Yuan et al (2015), where this neighbor constraint is introduced). Second, let

$$\tilde{a}_j = \sum_{\ell=1}^V w_{j\ell}$$

be the degree of $x_j$ and the diagonal matrix $A = \text{diag}\{\tilde{a}_1, \ldots, \tilde{a}_V\}$ is called the degree matrix. The $K$ smallest eigenvectors corresponding to the $K$ smallest eigenvalues of the symmetric normalized Laplacian matrix

$$L^{\text{sym}} = I - A^{-1/2}W A^{-1/2},$$

where $I$ is the identity matrix, are determined using the ARPACK algorithm (Lehoucq et al, 1998), where $K$ is equal to the number of clusters of the final parcellation. Combining these $K$ smallest eigenvectors in a matrix $F \in \mathbb{R}^{V \times K}$, the $K$-means algorithm (Lloyd, 1982; MacQueen, 1967) is applied to parcellate the rows of $F$ into $K$ clusters $C_1, \ldots, C_K$. Finally, voxel $x_j$ is assigned to cluster $C_k$, if and only if the $j$-th row of $F$ is assigned to cluster $C_k$.

Note that in order to make clustering decisions, SSPEC only uses the distances between voxels that are neighbors or have a common neighbor. In
contrast, the SHAC algorithms consider all pairwise distances between voxels at least once over the run of the algorithm. I.e. the SHAC algorithms use much more information in order to make clustering decisions than the SSPEC algorithms.

2.5 Spatial ensemble clustering

In order to improve the performance of SPARTACUS as well as of the SHAC methods and SSPEC, we employ spatial ensemble clustering (SEC) as an additional layer on top of regular spatial clustering. Like any ensemble clustering method (Boongoen and Iam-On, 2018), this SEC method consists of two steps, i.e. generating a cluster ensemble and employing a consensus function that calculates a new ensemble parcellation based on the cluster ensemble (Yang et al, 2014).

We generate the cluster ensemble via a subsampling approach, i.e. 63.2% of the subjects are randomly drawn without replacement to obtain $B$ subsample data sets $X_{\text{Sub}}^1, \ldots, X_{\text{Sub}}^B$. This specific size is chosen for the subsamples, since the average bootstrap sample includes 63.2% of the original subjects (Alpaydin, 2014). Note that this bootstrap-like subsampling approach achieves a good performance in the simulations conducted by Binder and Schumacher (2008). Next, as base clustering method, a SHAC algorithm is applied to all subsamples to generate $B$ base parcellations $C_K^{(1)}, \ldots, C_K^{(B)}$, each with $K$ clusters ($K$ is fixed). This yields a homogeneous cluster ensemble

$$P_K = \{C_K^{(1)}, \ldots, C_K^{(B)}\},$$

where diversity is solely explained by the subsampling approach. Note that multiple cluster ensembles for multiple numbers of clusters can be obtained
in short time, since, once the dendrograms are calculated for the subsamples, they can be cut to give any number of clusters with low computational cost.

A pairwise similarity based approach is used as consensus function, i.e. to obtain the final ensemble parcellation based on the cluster ensemble. Let $C_K^{(b)}(x_j)$ be the cluster label of $x_j$ due to $C_K^{(b)}$, $b = 1, \ldots, B$. The $V \times V$ connectivity matrix $M_K^{(b)}$ corresponding to $C_K^{(b)}$ is given by

$$M_K^{(b)}(x_j, x_\ell) = \begin{cases} 1, & \text{if } C_K^{(b)}(x_j) = C_K^{(b)}(x_\ell) \\ 0, & \text{otherwise} \end{cases}$$

Then, the ensemble distance between any two voxels is

$$d^E(x_j, x_\ell) = 1 - \frac{1}{B} \sum_{b=1}^{B} M_K^{(b)}(x_j, x_\ell).$$

Using $d^E$ as distance measure between voxels, the final ensemble parcellation with $K$ clusters can be obtained by employing a SHAC algorithm with, e.g., single or average linkage.

### 2.6 Internal validation

We compare the performance of the spatial clustering methods using internal validation measures, i.e. measures that use only information from the underlying data set to evaluate the quality of a parcellation $C_K = \{C_1, \ldots, C_K\}$. Moreover, internal validation measures are used to find interesting numbers of brain regions, as described in the next section.

#### 2.6.1 Silhouette coefficient

A popular internal measure considering the within-cluster compactness and between-cluster separation is the silhouette coefficient (SC) (Rousseeuw, 1987;
Zaki and Meira, 2014). The silhouette width of a single voxel $x_j \in C_k, k = 1, \ldots, K$, is defined as

$$s_j = \frac{b_j - a_j}{\max\{a_j, b_j\}},$$

where $a_j$ is the average distance of $x_j$ to all other voxels in $C_k$ and $b_j$ is the minimum average distance of $x_j$ to all voxels from another cluster. As distance measure we choose the absolute correlation based distance $d_{\text{absCorr}}(x_j, x_t) = 1 - |\text{corr}(x_j, x_t)|$. The SC for $C_K$ is then given by

$$SC = \frac{1}{V} \sum_{j=1}^{V} s_j.$$

SC takes values in $[-1, 1]$, where values close to 1 indicate a high quality parcellation and values close to 0 indicate random clustering.

### 2.6.2 Simplified silhouette coefficient

In order to obtain the SC, all $V(V - 1)/2$ pairwise distances among the $V$ voxels must be calculated. Since with high resolution MRI data the number of voxels is large, the calculation of the SC is computationally very expensive. A variation of SC which is computationally less expensive is the simplified silhouette coefficient (SSC) (Vendramin et al, 2010). In this variation, $a_j$ is the distance of voxel $x_j$ to the centroid of its cluster and $b_j$ is the minimum of the distances of $x_j$ to the centroids of the other clusters. Typically, the distance measure is the Euclidean distance and the centroid of a cluster is the mean over all voxels from that cluster (Vendramin et al, 2010). However, since we are dealing with spatially correlated voxel data, we consider $d_{\text{absCorr}}$ as distance measure and the first principal component of a cluster as its centroid.
2.6.3 Spatial adaptation of (simplified) silhouette coefficient

Both, the SC and SSC ignore the spatial information provided by the data. However, it is, e.g., known that brain regions in one hemisphere may interact with their contralateral regions on the other hemisphere (Davis and Cabeza, 2015). These cross-hemispheric communications may cause similar patterns of grey matter volume in the concerned brain regions. Since usually these brain regions are not spatially connected, they can not be merged by a spatial clustering algorithm and, thus, reduce inter-cluster separation. This results in a worse SC or SSC score. Thus, as spatial adaptation of SC we propose to calculate \( b_j \) as the minimum average distance of \( x_j \in C_k \) to all voxels from a neighbor cluster of \( C_k \), i.e.

\[
b_j = \min_{m \neq k} \min_{s_{km} = 1} \frac{1}{|C_m|} \sum_{x \in C_m} d_{\text{absCorr}}(x_j, x),
\]

where \( s_{km} = 1 \), if and only if \( C_k \) and \( C_m \) are neighbors. Analogously, the spatial adaptation of SSC calculates \( b_j \) as the minimum of the distances of \( x_j \in C_k \) to the centroids of the neighbor clusters of \( C_k \). The spatial adaptations of SC and SSC are referred to as \( SC_{\text{s Spatial}} \) and \( SSC_{\text{s Spatial}} \). Note that \( SC_{\text{s Spatial}} \) and \( SSC_{\text{s Spatial}} \) are not influenced by cross-hemispheric communications and are computationally cheaper than SC and SSC, respectively. Moreover, it is easy to show that \( SC_{\text{s Spatial}} \geq SC \) and \( SSC_{\text{s Spatial}} \geq SSC \).

2.7 Model selection

An important aspect when performing brain parcellation is the issue of identifying interesting numbers of brain regions, where these different numbers may reflect different levels of brain organization (Eickhoff et al, 2018b). For this, we employ two different approaches, i.e. a subsampling based clustering stability and a subsampling based clustering quality approach.
2.7.1 Clustering stability

Clustering stability (Von Luxburg, 2010; Vinh and Epps, 2009) is a popular framework to identify interesting numbers of clusters. The idea behind clustering stability is that, if a clustering method with the right set of parameters is applied to several data sets sampled from the same underlying distribution, the clustering solutions will be similar, i.e. stable (Von Luxburg, 2010). More specifically, among sets of parcellations with different numbers of brain regions between the sets those should be most stable, for which the number of brain regions coincides with a number of brain regions reflecting a level of brain organization.

Here we employ a subsampling based clustering stability approach. In this approach, $B$ subsamples are generated, each by drawing 63.2% of the original subjects without replacement. The specific size of the subsamples is chosen identically to the size of the subsamples in the SEC framework. For each number $K = 2, \ldots, K_{\text{max}}$ of brain regions a spatial clustering algorithm is applied to each subsample, generating parcellations $C^K_{b}, b = 1, \ldots, B$. Pairwise stability among two parcellations with the same number of brain regions is measured using the adjusted Rand index (ARI) (Hubert and Arabie, 1985). The overall stability of parcellations with the same number of brain regions is then quantified by the mean over all their pairwise stability scores, i.e.

$$\overline{\text{ARI}}(K) = \frac{2}{B(B-1)} \sum_{b=1}^{B-1} \sum_{b'=b+1}^{B} \text{ARI} \left( C^K_{b}, C^K_{b'} \right).$$

E.g., local maxima of $\overline{\text{ARI}}$ may indicate interesting numbers of clusters.

A major advantage of SHAC algorithms over other clustering algorithms, such as SSPEC, is that a SHAC algorithm only needs to be applied once to each subsample and parcellations with different numbers of clusters are obtained by splitting up the hierarchy at low computational cost. This reduces
the computational complexity of the subsampling based clustering stability approach dramatically, especially, if the goal is to compare many different numbers of clusters.

The subsampling based clustering quality approach has the advantage over the usual approach that the internal validation scores are calculated based on the OOB subjects, i.e. on subjects that are not used for generating the parcellations. Moreover, the mean internal validation curves are smoother and more robust. However, if a sufficient infrastructure for parallelization is not available, its computation takes much longer.

2.7.2 Clustering quality

The usual approach to identify interesting numbers of clusters using internal validation measures is to apply the same clustering algorithm with different numbers of clusters to the input data set (Arbelaitz et al, 2013). Afterwards, an internal validation measure, e.g., the SC or SSC, is applied to each of the parcellations and the validation scores are plotted against the numbers of clusters in a scatterplot. Local maxima (assuming that larger scores indicate larger clustering quality) of that curve indicate interesting numbers of clusters and are considered for further investigation.

However, we employ a subsampling based clustering quality approach, giving a more stable estimate of clustering quality. Like in the subsampling based clustering stability approach, $B$ subsamples of size $[0.632 \cdot N]$ are drawn. For each number $K = 2, \ldots, K_{\text{max}}$ of brain regions, a spatial clustering algorithm is applied to each subsample, generating parcellations $C_{K}^{b}$, $b = 1, \ldots, B$. The quality of each parcellation is evaluated using an internal validation measure on the corresponding out-of-bag (OOB) subjects, i.e. those subjects from the
input data set that are not part of the respective subsample. Clustering quality for each $K$ is then quantified by the mean over all corresponding internal validation scores, i.e.

$$\bar{\text{IVS}}(K) = \frac{1}{B} \sum_{b=1}^{B} \text{IVS}(C_{bK}^b),$$

where $\text{IVS}(C_{bK}^b)$ is the OOB internal validation score of $C_{bK}^b$. Note that IVS is a dummy variable coding for SC, SSC, SC_{spatial} or SSC_{spatial}. Again, local maxima of $\bar{\text{IVS}}$ may indicate interesting numbers of brain regions.

### 2.8 Convergence analysis with existing brain atlases

There exist various atlases of the human brain in the literature which are derived based on different modalities and parcellation methods. A critical and widely accepted idea in this context is that brain organization can be described by distinct brain regions/cortical areas of large within homogeneity and large between heterogeneity with respect to all three modality categories structure, connectivity and function. Hereby, the edges from all modalities should closely match each other (Eickhoff et al, 2018a). According to this idea, as further quality measure of a brain parcellation we consider its convergence to other (well established) atlases, i.e. how well the brain regions of the parcellation converge with those of existing atlases. We measure convergence by mapping both the new brain parcellation and the existing atlas to a common brain template and then apply the ARI. Hereby, we distinguished between convergence with histological mapping, i.e. anatomical atlases, and convergence with alternative parcellations (Varikuti et al, 2018), e.g., resting-state fMRI based parcellations.

We determine the convergence of the 1000BRAINS based parcellations derived in this manuscript with multiple anatomical and alternative brain
**Table 1** Overview over brain atlases considered for comparison.

| Name          | Number of regions | Brain coverage     | voxel resolution  |
|---------------|-------------------|--------------------|-------------------|
| **Anatomical atlases** |                   |                    |                   |
| AAL1          | 116               | Whole brain        | $2 \times 2 \times 2 \text{mm}^3$ |
| AAL3          | 166               | Whole brain        | $1 \times 1 \times 1 \text{mm}^3$ |
| MarsAtlas     | 97                | Cerebrum           | $1 \times 1 \times 1 \text{mm}^3$ |
| **Resting-state fMRI based atlases** |                   |                    |                   |
| Bellec        | 7,12,20,36,64,    | Whole brain        | $3 \times 3 \times 3 \text{mm}^3$ |
|               | 122,197,325,444   |                    |                   |
| Craddock      | 10 to 1000        | Whole brain        | $4 \times 4 \times 4 \text{mm}^3$ |
| Schaefer      | 100 to 1000       | Cerebral cortex    | $1 \times 1 \times 1 \text{mm}^3$ |
| Shen          | 93,184,278        | Whole brain        | $1 \times 1 \times 1 \text{mm}^3$ |
| **Others**    |                   |                    |                   |
| Glasser       | 180               | Cerebral cortex    | $0.5 \times 0.5 \times 0.5 \text{mm}^3$ |
| Varikuti (MIXED) | 25 to 675 | Whole brain        | $1.5 \times 1.5 \times 1.5 \text{mm}^3$ |

Atlases from the literature. Anatomical atlases are the AAL1 atlas ([Tzourio-Mazoyer et al., 2002](#)), the AAL3 atlas ([Rolls et al., 2020](#)) and the MarsAtlas atlas ([Auzias et al., 2016; Brovelli et al., 2017](#)). Resting-state fMRI based atlases are those by [Bellec et al. (2010)](#) registered in the symmetric version of the MNI template, by [Craddock et al. (2012)](#), where similarity is determined based on the temporal similarity between voxels (tcorr) and clustering is performed using a two-level scheme (2level) clustering individual subjects first before combining these individual partitions to one final partition, by [Schaefer et al. (2018)](#) or by [Shen et al. (2013)](#). Moreover, the multimodal brain atlas by [Glasser et al. (2016)](#) and the atlases by [Varikuti et al. (2018)](#) which are generated by applying OPNMF based clustering to the MIXED data set, i.e. to another structural MRI data set, are considered. More information about these atlases is summarized in Table 1.
3 Results

In order to compare the performance of SPARTACUS with the competing methods, i.e. SHAC and SSPEC, and in order to investigate whether using SEC as additional layer improves clustering quality, these methods are applied to the simulation study as well as to the 1000BRAINS data set.

3.1 Analysis of simulation study

In the following, we present the main results of the analysis of the simulation study. A more detailed description of the results is given in Section A.2 of the Online Resource.

3.1.1 Performance comparison of spatial clustering methods

The four SHAC algorithms SPARTACUS, SHAC\textsubscript{Ward}, SHAC\textsubscript{AL\textsubscript{corr}} and SHAC\textsubscript{AL\textsubscript{Eucl}} as well as the SSPEC algorithm are considered for comparison. We are further interested in analyzing whether standardization of the simulated data sets influences clustering quality. Note that SHAC\textsubscript{AL\textsubscript{corr}} is not influenced by standardization, since the correlation is not influenced by standardization. Moreover, when applying the SPARTACUS method, the columns of the simulated data sets must be at least centered but better standardized. Hence, all five clustering methods are applied to the standardized data sets, whereas only the non-correlation based algorithms, i.e. SHAC\textsubscript{Ward}, SHAC\textsubscript{AL\textsubscript{Eucl}} and SSPEC, are also applied to the non-standardized data sets. Thus, in total we consider eight different clustering methods, where we indicate clustering methods that are applied to standardized data sets by \textit{method}^S, e.g., SPARTACUS$^S$.

In a first step, we apply all eight spatial clustering methods to all six simulation scenarios (SimBS, SimBM, SimBW, SimUS, SimUM and SimUW)
Table 2: The mean over 25 ARI and SSC scores for each of eight clustering methods based on SimBW and SimUW, where each ARI value compares a predicted parcellation with $K = 27$ or $K = 54$ clusters with the respective true parcellation and each SSC score evaluates the quality of a predicted parcellation on the training data.

|          | SimBW          |          | SimUW          |          |
|----------|----------------|----------|----------------|----------|
|          | ARI K=27 K=54  | SSC K=27 | ARI K=27 K=54  | SSC K=27 |
| SPARTACUS | 0.930 0.861    | 0.450 0.407 | 0.875 0.865    | 0.435 0.405 |
| SHACWard  | 0.907 0.846    | 0.441 0.401 | 0.854 0.853    | 0.424 0.400 |
| SHACWard  | 0.923 0.852    | 0.448 0.404 | 0.869 0.856    | 0.431 0.402 |
| SHACWard  | **0.936** 0.721 | **0.452** 0.430 | **0.941** 0.733 | **0.448** **0.428** |
| SHACAL, corr | 0.004 0.143 | -0.139 0.142 | 0.005 0.130 | -0.136 0.113 |
| SHACAL, Eucl | 0.929 0.710 | 0.450 **0.430** | 0.936 0.709 | 0.447 0.424 |
| SSPEC     | 0.630 0.278    | 0.342 0.111 | 0.403 0.295    | 0.144 0.132 |
| SSPEC     | 0.831 0.345    | 0.430 0.183 | 0.547 0.420    | 0.235 0.221 |

From these tables it can be concluded that the standardized methods achieve a higher quality than their corresponding non-standardized methods, especially if the clusters are weakly pronounced. While this quality increase caused by standardization is severe for SHACAL, Eucl and SSPEC, its only
marginal for SHAC_{Ward}. Moreover, these tables reveal, that the SHAC algorithms outperform the SSPEC algorithm, especially if $K = 54$ or if the true parcellations are unbalanced. Among the standardized SHAC methods, SPARTACUS$^S$ performs similarly to SHAC$^S_{Ward}$ and SHAC$^S_{AL, corr}$ performs similarly to SHAC$^S_{AL, Eucl}$ in all scenarios. SPARTACUS$^S$ and SHAC$^S_{Ward}$ more stably predict the true parcellations with $K = 54$ in both settings. However, SHAC$^S_{AL, corr}$ and SHAC$^S_{AL, Eucl}$ better identify the true parcellation with $K = 27$ in the second setting. Interestingly, the parcellations due to SHAC$^S_{AL, corr}$ and SHAC$^S_{AL, Eucl}$ achieve slightly better SSC scores than SPARTACUS$^S$ and SHAC$^S_{Ward}$ in all scenarios, even though SPARTACUS$^S$ and SHAC$^S_{Ward}$ achieve higher ARI scores for $K = 54$.

By visualizing the parcellations generated by SSPEC$^S$ or SHAC$^S_{AL, Eucl}$ (see Section A.2.1 of the Online Resource) it can be observed that SSPEC$^S$ tends to generate cubical-shaped clusters of equal size, regardless of the underlying data structure. If and only if the data is organized in a cubical, equally sized fashion as for $K = 27$ in the first setting, the underlying data structure is found. However, already in the presence of non-cubic clusters as for $K = 54$ in the first setting, SSPEC$^S$ continues to produce cubical clusters and, therefore, fails in finding any true cluster. Moreover, SHAC$^S_{AL, Eucl}$ tends to form a few large clusters and many very small clusters, if the simulated effect is small. These observations explain the poor performance of SSPEC$^S$ and SHAC$^S_{AL, Eucl}$.

### 3.1.2 Performance of spatial ensemble clustering

In a second step, we investigate whether SEC methods are able to improve clustering quality. Homogeneous cluster ensembles are generated by drawing $B = 50$ subsamples (randomly drawing 63.2% of the $N = 100$ subjects without replacement) and considering each of the four standardized SHAC methods as well as SSPEC$^S$ as base clustering methods. We only consider standardized
### Table 3

The mean over 25 ARI and SSC scores for each of nine SEC methods based on SimBW and SimUW, where each ARI value compares a predicted parcellation with $K = 27$ or $K = 54$ clusters with the respective true parcellation and each SSC score evaluates the quality of a predicted parcellation on the training data.

| SEC, method | SimBW   | SimUW   |
|-------------|---------|---------|
|             | ARI K=27 | SSC K=27 | ARI K=54 | SSC K=54 | ARI K=27 | SSC K=54 | ARI K=54 | SSC K=54 |
| SEC_AL      |         |         |         |         |         |         |         |         |
| SPARTACUS$^S$ | 0.993   | 0.972   | 0.467   | 0.443   | 0.948   | 0.968   | 0.450   | 0.440   |
| SHAC$^S_{Ward}$ | 0.993   | 0.969   | 0.467   | 0.443   | 0.956   | 0.962   | 0.450   | 0.439   |
| SHAC$^S_{AL, corr}$ | 0.980   | 0.721   | 0.463   | 0.433   | 0.988   | 0.721   | 0.461   | 0.432   |
| SHAC$^S_{AL, Eucl}$ | 0.972   | 0.715   | 0.460   | 0.431   | 0.983   | 0.714   | 0.459   | 0.432   |
| SSPEC$^S$   | 0.832   | 0.337   | 0.425   | 0.177   | 0.546   | 0.415   | 0.233   | 0.220   |

| SEC_SL      |         |         |         |         |         |         |         |         |
|-------------|---------|---------|---------|---------|---------|---------|---------|---------|
| SPARTACUS$^S$ | 0.765   | 0.674   | 0.361   | 0.376   | 0.761   | 0.674   | 0.350   | 0.368   |
| SHAC$^S_{Ward}$ | 0.739   | 0.669   | 0.353   | 0.368   | 0.767   | 0.666   | 0.351   | 0.360   |
| SHAC$^S_{AL, corr}$ | 0.761   | 0.638   | 0.367   | 0.393   | 0.791   | 0.636   | 0.367   | 0.391   |
| SHAC$^S_{AL, Eucl}$ | 0.760   | 0.633   | 0.364   | 0.391   | 0.792   | 0.632   | 0.367   | 0.391   |
| SSPEC$^S$   | 0.191   | 0.007   | 0.165   | -0.205  | 0.049   | 0.021   | -0.041  | -0.169  |

Clustering methods as base clustering methods, since these methods outperformed the non-standardized clustering methods in the previous step. In the consensus function step, a SHAC algorithm with single linkage (SL) and average linkage (AL) is employed to obtain the final ensemble parcellation from the cluster ensemble. Hence, $(4 + 1) \cdot 2 = 10$ SEC methods are applied to all scenarios from the simulation study, where, e.g., SEC$_{AL}(SPARTACUS^S)$ refers to the SEC method using SPARTACUS$^S$ as base clustering method and average linkage based SHAC as consensus function. We employ the same two approaches as in the first step in order to evaluate the performance of the SEC methods.

The results presented in Table 3 (for SimBW and SimUW) and in Section A.2.2 of the Online Resource reveal that SEC$_{AL}$ is able to improve clustering quality and robustness of the corresponding SHAC methods, if these SHAC methods do not already achieve perfect clustering. E.g., in SimUW the mean ARI (SSC) scores for $K = 27$ and $K = 54$ increase from 0.875 and 0.865
(0.435 and 0.405) based on \textsc{spartacus}$^S$ to 0.948 and 0.968 (0.451 and
0.440) based on \textsc{sec}$^A_l$(\textsc{spartacus}$^S$), respectively. \textsc{sec}$^A_l$ performs best
in combination with \textsc{spartacus}$^S$ or \textsc{shac}$^S_{\text{Ward}}$ as base clustering method.
\textsc{sec}$^A_l$(\textsc{spartacus}$^S$) and \textsc{sec}$^A_l$(\textsc{shac}$^S_{\text{Ward}}$) stably identify even weakly
pronounced clusters (smallest mean ARI value over all scenarios is 0.9483).
Moreover, they more stably identify the parcellations with $K = 54$ clus-
ters than \textsc{sec}$^A_l$(\textsc{shac}$^S_{\text{AL, corr}}$) and \textsc{sec}$^A_l$(\textsc{shac}$^S_{\text{AL, Eucl}}$). However, \textsc{sec}$^A_l$
does not improve the performance of \textsc{sspec}$^S$.

In contrast, \textsc{sec}$^S_{\text{L}}$ fails as SEC method, if the corresponding \textsc{shac}
method does not already achieve nearly perfect clustering. E.g., in SimUW
the mean ARI (SSC) scores for $K = 27$ and $K = 54$ decrease from 0.875 and
0.865 (0.435 and 0.405) based on \textsc{spartacus}$^S$ to 0.761 and 0.674 (0.350
and 0.368) based on \textsc{sec}$^S_{\text{L}}$(\textsc{spartacus}$^S$), respectively. Moreover, \textsc{sec}$^S_{\text{L}}$
clearly decreases the performance of \textsc{sspec}$^S$.

3.1.3 Performance of methods to find interesting numbers of
brain regions

In a third step, we analyze whether the subsampling based clustering stability
approach and the subsampling based clustering quality approach are able to
identify the true numbers of clusters of the simulation study. Both approaches
are applied to each data set from the simulation study. In both approaches we
consider all eight \textsc{shac} and \textsc{sspec} methods as spatial clustering algorithms,
choose $K_{\text{max}} = 100$ and draw $B = 50$ subsamples. Moreover, in the sub-
sampling based clustering quality approach we consider four different internal
validation measures, i.e. \textsc{sc}, \textsc{ssc}, \textsc{sc}_{\text{spatial}} and \textsc{ssc}_{\text{spatial}}. For SimBW and
SimUW the mean over the resulting $H = 25$ ARI or \textsc{ssc}_{\text{spatial}} scores is plotted
for each $K = 2, \ldots, 100$ in Figure 2. The remaining results can be found
in Section A.2.3 of the Online Resource.
Fig. 2 For SimBW and SimUW the mean over $H = 25$ ARI or $SSC_{\text{spatial}}$ scores generated by the subsampling based clustering stability or quality approach, respectively, is plotted for each $K = 2, \ldots, 100$

From these figures it can be observed that the spatial adaptations $SC_{\text{spatial}}$ and $SSC_{\text{spatial}}$ generate nearly identical mean OOB internal validation curves compared to SC and SSC, respectively. Moreover, the SC and SSC curves progress very similar to each other, even though SSC produces larger scores than SC. Thus, we conclude that SSC, $SC_{\text{spatial}}$ and $SSC_{\text{spatial}}$ are valid simplifications/modifications of SC for both finding interesting numbers of brain regions from structural MRI data and evaluating the quality of parcellations. Apart from this, both approaches produce similar results that lead to the same conclusions. As expected, they generate larger scores, the more the simulated
clusters are pronounced. Both approaches perform best in combination with
SPARTACUS\textsuperscript{S}, SHAC\textsubscript{Ward} \textsuperscript{S} or SHAC\textsubscript{Ward}. In all scenarios of the first setting
they correctly identify the true numbers of clusters. Also in the scenarios of
the second setting they always correctly produce a maximum at \( K = 54 \). How-
ever, \( K = 27 \) can not clearly be identified in the second setting. In contrast,
in combination with SHAC\textsubscript{AL, corr} \textsuperscript{S} and SHAC\textsubscript{AL, Eucl} \textsuperscript{S}, both approaches iden-
tify \( K = 27 \) as interesting number of clusters in all scenarios but \( K = 54 \) is
hardly found in any scenario. The worst performance is achieved in combina-
tion with the SSPEC algorithms and with SHAC\textsubscript{AL, Eucl}. Not only do they
achieve a clearly lower quality and stability, but they also hardly identify any
true number of clusters.

3.2 Analysis of 1000BRAINS data set

3.2.1 Application of spatial clustering algorithms

In order to compare the performance of SPARTACUS\textsuperscript{S}, SHAC\textsubscript{Ward},
SHAC\textsubscript{Ward} \textsuperscript{S}, SHAC\textsubscript{AL, corr} \textsuperscript{S}, SHAC\textsubscript{AL, Eucl} \textsuperscript{S} and SHAC\textsubscript{AL, Eucl} \textsuperscript{S}
for different num-
bers of brain regions, we apply these methods to the 1000BRAINS data set
and split up each hierarchy to obtain parcellations with \( K = 2, \ldots, 1000 \) brain
regions. The quality of each parcellation is evaluated on the 1000BRAINS data
set using SSC\textsubscript{spatial}.

The results of this analysis are presented in Figure 3 and reveal that
the standardized SHAC methods produce parcellations of similar quality,
where SPARTACUS\textsuperscript{S} and SHAC\textsubscript{Ward} \textsuperscript{S} show a slightly better performance than
SHAC\textsubscript{AL, corr} \textsuperscript{S} and SHAC\textsubscript{AL, Eucl} \textsuperscript{S}. In contrast, the non-standardized SHAC
methods SHAC\textsubscript{Ward} and SHAC\textsubscript{AL, Eucl} \textsuperscript{S} perform poorly. By visualizing the par-
cellations due to SHAC\textsubscript{Ward} and SHAC\textsubscript{AL, Eucl} \textsuperscript{S} (see Section B.1 of the Online
Fig. 3 The $\text{SSC}_{\text{spatial}}$ scores evaluating the quality of the parcellations with $K = 2, \ldots, 1000$ brain regions generated by the six SHAC methods based on the 1000BRAINS data set.

Resource) we observe that SHAC$\text{Ward}$ produces central clusters which are surrounded by multiple thin cluster rings. This patterning is not established in the field of brain parcellation and is also not observed in the simulation study. As in the simulation study, SHAC$\text{AL, Eucl}$ produces a few very large brain regions and a large number of very small brain regions, i.e. it assigns outlier voxels to singleton brain regions. Therefore, it is not very surprising that the quality of these parcellations is low.

### 3.2.2 Selecting numbers of brain regions

Interesting numbers of brain regions are searched for using the subsampling based clustering stability approach and the subsampling based clustering quality approach. In both approaches we employ the four standardized SHAC methods as spatial clustering methods, draw $B = 100$ subsamples and choose $K_{\text{max}} = 1000$. As internal validation measures in the subsampling based clustering quality approach we consider $\text{SSC}_{\text{spatial}}$.

By looking at the results generated by the subsampling based clustering quality approach in Figure 4 we observe that the $\text{SSC}_{\text{spatial}}$ curves progress...
Fig. 4 The $\overline{SSC}_{\text{spatial}}$ curves generated by the subsampling based clustering quality approach with $K = 2, \ldots, 1000$ and using the four standardized SHAC methods as spatial clustering methods similarly to the respective curves from (the lower plot of) Figure 3, i.e. they increase monotonic until they reach a plateau approximately at $K = 600$, from where on the curve’s slopes are close to zero. Hence, by looking for local maxima of these curves no interesting numbers of clusters can be deduced.

However, the first derivatives of these curves might show some interesting slope patterns. Therefore, we display the first derivative of $\overline{SSC}_{\text{spatial}}$, denoted by $\frac{\partial}{\partial K} \overline{SSC}_{\text{spatial}}$, together with the respective $\overline{ARI}$ curves generated by the subsampling based clustering stability approach in Figure 5 (for SPARTACUS$^S$ and SHAC$^S_{\text{AL, corr}}$) and in Section B.2 of the Online Resource (for SHAC$^S_{\text{Ward}}$ and SHAC$^S_{\text{AL, Eucl}}$).

Interestingly, it is observed from these figures that there is some agreement between the $\frac{\partial}{\partial K} \overline{SSC}_{\text{spatial}}$ and $\overline{ARI}$ curves. Consistently, maxima in the $\overline{ARI}$ curves are accompanied by changes, e.g., elbow points, in the $\frac{\partial}{\partial K} \overline{SSC}_{\text{spatial}}$ curves. E.g., at $K = 154$ or $K = 146$ there is an elbow point in the $\frac{\partial}{\partial K} \overline{SSC}_{\text{spatial}}$ curve and a local maximum in the $\overline{ARI}$ curve based on SPARTACUS$^S$ or SHAC$^S_{\text{AL, corr}}$, respectively. Therefore, at least three interesting numbers of brain regions can be deduced, i.e. $K \approx 70$, $K \approx 150$. 
Fig. 5 The first derivative of the $\overline{\text{SSC}_{\text{spatial}}}$ curve (blue) generated by the subsampling based clustering quality approach and the $\overline{\text{ARI}}$ curve (green) generated by the subsampling based clustering stability approach, both with $K = 2, \ldots, 1000$ and using SPARTACUS$^S$ or SHAC$^S_{\text{AL}, \text{corr}}$ as spatial clustering algorithm.

and $K \approx 600$. When only looking at the curves due to SHAC$^S_{\text{AL}, \text{corr}}$ and SHAC$^S_{\text{AL}, \text{Eucl}}$, $K \approx 300$ is another interesting number of brain regions.

3.2.3 Application of spatial ensemble clustering

In order to obtain higher quality parcellations, we employ a SEC method, where we consider the same $B = 100$ subsamples as in the previous analysis and use the two standardized SHAC methods SPARTACUS$^S$ and SHAC$^S_{\text{AL}, \text{corr}}$ as base clustering methods to obtain homogeneous cluster ensembles with $K = 2, \ldots, 500$ brain regions. Only numbers of brain regions up to 500 are considered (and not up to 1000), since the SEC methods are very expensive to compute and almost all interesting numbers of brain regions identified in the previous analysis are included in this smaller range. SEC$^S_{\text{AL}}$ is used as
The SSC$_{\text{spatial}}$ scores evaluating the quality of the parcellations with $K = 2, \ldots, 500$ brain regions generated by the four SEC$_{\text{AL}}$ methods and, for comparison, by the SPARTACUS$^S$ method based on the 1000BRAINS data set consensus function to calculate the final ensemble parcellations from the cluster ensembles. The quality of the ensemble parcellations is evaluated on the 1000BRAINS data set using SSC$_{\text{spatial}}$. The results presented in Figure 6 show that all four SEC methods are able to improve clustering quality of the standardized SHAC methods. For $K > 88$, the SSC$_{\text{spatial}}$ scores by both SEC methods are larger compared to SPARTACUS$^S$ which achieves the largest SSC$_{\text{spatial}}$ scores among the SHAC methods (compare Figure 3). The overall best performance is achieved by SEC$_{\text{AL}}$(SPARTACUS$^S$).

The final ensemble parcellations with $K = 70$ and $K = 150$ brain regions due to SEC$_{\text{AL}}$(SPARTACUS$^S$) and SEC$_{\text{AL}}$(SHAC$_{\text{AL, corr}}^S$) are visualized in Figure 7 and in more detail in Section B.3 of the Online Resource. From these figures it can be observed that SEC$_{\text{AL}}$(SPARTACUS$^S$) generates brain regions of similar sizes, whereas SEC$_{\text{AL}}$(SHAC$_{\text{AL, corr}}^S$) has a slight tendency to produce a few large and multiple smaller brain regions.
3.2.4 Method comparison with spectral and geometric clustering

We further compare the performance of $\text{SEC}_{\text{AL}}(\text{SPARTACUS}^S)$ with the performance of $\text{SSPEC}^S$. Moreover, as reference for comparison, similar to Thirion et al. (2014) we consider a geometric clustering approach that uses only the spatial coordinates and ignores the grey matter volumes. More precisely, we perform $\text{SHAC}_{\text{AL}, \text{Eucl}}$ clustering of the coordinate matrix $Z$, i.e. using $Z$ as data matrix and as coordinate matrix. Parcellations with $K = 2, \ldots, 500$ brain regions are generated by all three methods and these parcellations are evaluated on the 1000BRAINS data set using the SSC. The results of this analysis are presented in Figure 8 and reveal that for all numbers of brain regions $\text{SEC}_{\text{AL}}(\text{SPARTACUS}^S)$ outperforms $\text{SSPEC}^S$. Moreover, $\text{SSPEC}^S$ achieves a better performance than geometric clustering. Interestingly, the internal validation curves corresponding to $\text{SEC}_{\text{AL}}(\text{SPARTACUS}^S)$ and $\text{SSPEC}^S$ progress similar to the internal curves corresponding to geometric clustering, i.e. all three curves increase monotonically with increasing number of brain regions.
In order to get a visual impression of the parcellations due to SSPEC$^S$ and geometric clustering and to better understand the clustering behaviour of these methods, the respective partitions with $K = 150$ are shown in Section B.4 of the Online Resource. These visualizations show, as already observed in the simulation study, that SSPEC$^S$ tends to produce spherical shaped brain regions of equal size. As expected, geometric clustering produces equally sized brain regions that reflect the spatial structure of the brain.

### 3.2.5 Similarity comparison with existing brain atlases

As further metric of the quality of the parcellations with $K = 2, \ldots, 500$ brain regions belonging to the SEC$_{AL}$(SPARTACUS$^S$) or the SSPEC$^S$ family, where, e.g., the SEC$_{AL}$(SPARTACUS$^S$) family consists of all parcellations with $K = 2, \ldots, 500$ brain regions obtained by applying SEC$_{AL}$(SPARTACUS$^S$) to the 1000BRAINS data set, we compare the convergence of some of these parcellations with the anatomical and the alternative (algorithmic) atlases presented at the end of the methods section. The convergence of any two brain atlases, regardless whether they are obtained from the same modality
or from different modalities, should be more similar to each other than to a random parcellation, such as a geometric parcellation. Therefore, as reference for comparison, we also determine the convergence of geometric parcellations with the anatomical and alternative atlases. The convergence between any two atlases with similar numbers of brain regions is quantified by the ARI.

However, some issues occur when quantifying the convergence between atlases, i.e. the brain atlases neither are registered to the same brain template in MNI space, nor have the same voxel resolution, nor have the same brain coverage, nor have the same numbers of brain regions. Thus, we employ some preprocessing steps in order to match atlases from two different atlas families. These preprocessing steps are described in Section B.5 of the Online Resource. After pairwise matching, those atlases from two matched families are paired that have the most similar numbers of brain regions, but only, if the absolute difference between their numbers of brain regions is less than or equal to 20.

Using the ARI, the convergence of the parcellations belonging to the \( \text{SEC}_\text{AL}(\text{SPARTACUS}^S) \), \( \text{SSPEC}^S \) or geometric family with the three anatomical atlases is presented in Table 4 and with the parcellations from the six alternative atlas families is shown in Figure 9 and in Section B.6 of the Online Resource.

Table 4 reveals that the parcellations from the \( \text{SEC}_\text{AL}(\text{SPARTACUS}^S) \) and \( \text{SSPEC}^S \) family converge equally well and above chance (ARI \( \in (0.25, 0.36) \)) with established anatomical brain atlases. However, this convergence is only

|                     | AAL1 (115) | AAL3 (153) | MarsAtlas (97) |
|---------------------|------------|------------|----------------|
| \( \text{SEC}_\text{AL}(\text{SPARTACUS}^S) \) | 0.2883     | 0.2682     | 0.3284         |
| \( \text{SSPEC}^S \) | 0.3020     | 0.2530     | 0.3524         |
| geometric           | 0.2808     | 0.2510     | 0.3176         |
Fig. 9 Each subplot shows the ARI based convergence of parcellations from the Varikuti (MIXED) or Craddock family with the parcellations from the SEC\textsubscript{AL}(SPARTACUS\textsuperscript{S}), SSPEC\textsuperscript{S} and geometric family. The values on the \(x\)-axis are the numbers of clusters of the parcellations from the alternative atlas family after matching slightly better than the convergence of parcellations from the geometric family with the anatomical atlases.

From the figures we observe that parcellations from all 1000BRAINS families converge above chance (ARI \(\geq 0.2\)) with parcellations from all alternative atlas families. The parcellations from the SEC\textsubscript{AL}(SPARTACUS\textsuperscript{S}) family converge well with parcellations from the Varikuti (MIXED) family derived based on the same modality. This convergence is clearly better compared with parcellations from both the geometric and the SSPEC\textsuperscript{S} family. However, SEC\textsubscript{AL}(SPARTACUS\textsuperscript{S}) parcellations converge worse with resting-state fMRI based parcellations than SSPEC\textsuperscript{S} and geometric parcellations. Thus, parcellations from the SEC\textsubscript{AL}(SPARTACUS\textsuperscript{S}) family show a good within-modality and a rather poor between-modality convergence. In contrast, SSPEC\textsuperscript{S} parcellations show a poor within-modality and good between-modality convergence, achieving the largest convergence with resting-state fMRI based parcellations,
but the worst convergence with parcellations from the Varikuti (MIXED) family (even worse than geometric parcellations). The overall lowest convergence of parcellations from any 1000BRAINS family is achieved with the multimodal atlas from Glasser et al (2016).

4 Discussion

The generation and application of human brain parcellations is a fundamental concept in the field of neuroscience in order to understand brain organization. While a large number of different parcellations exist in the literature, these parcellations differ mainly from each other by the modalities that they are derived from and the parcellation techniques. This article contributes by extensively investigating the performance and stability of the newly proposed SPARTACUS method as well as of popular SHAC algorithms, in application to one specific modality, i.e. grey matter volume deduced from T1-weighted structural MRI scans. The results of this investigation give a better orientation on which clustering method to choose for brain parcellation and provide an additional view on human brain organization, thus, helping to improve the understanding of the mechanisms of the human brain.

4.1 Suggestions for structural MRI parcellation

Based on the results of our performance analysis conducted on simulated data and on the 1000BRAINS data, we suggest to standardize the voxels of the input data to have a mean of zero and a standard deviation of one prior to clustering. Standardization clearly improves clustering quality for all considered data sets. E.g., \( \text{SHAC}_{\text{AL, Eucl}} \) hardly finds any simulated clusters based on non-standardized data but stably identifies the correct parcellation based on standardized data. The reason is that, based on the non-standardized data,
SHAC$_{AL, Eucl}$ tends to form a few large clusters and assigns outlier voxels to singleton or small clusters. This behavior of average linkage hierarchical clustering is well known in the literature (Senbabaoğlu et al., 2014). Standardizing the data seems to reduce the impact of outlier voxels, such that SHAC$_{AL, Eucl}$ much more stably finds the true parcellation.

For structural brain parcellation we recommend to employ the SPARTACUS$^S$ or the SHAC$^S_{Ward}$ method. Both methods generate parcellations consisting of similar sized brain regions, while at the same time being sensitive to the structural data. Thus, these methods consider a good balance between spatial and structural information. In contrast, SSPEC$^S$ mainly considers spatial information by tending to produce spherical shaped brain regions of equal size, while having a low sensitivity to the underlying structural information. Note that similar observations are made by Thirion et al. (2014) on task-based functional MRI data. Moreover, SPARTACUS$^S$ and SHAC$^S_{Ward}$ outperform SHAC$^S_{AL, corr}$ and SHAC$^S_{AL, Eucl}$ in the simulation study. On the 1000BRAINS data set SHAC$^S_{AL, corr}$ and SHAC$^S_{AL, Eucl}$ tend to produce a few large and multiple smaller brain regions even after standardization, which is an unwanted effect.

The quality of brain parcellations generated by a SHAC algorithm can be further improved using the SEC$_{AL}$ approach but not using the SEC$_{SL}$ approach. The SEC$_{AL}$ approach stably identifies even weakly pronounced clusters in the simulation study and improves clustering quality of the respective 1000BRAINS based parcellations. In contrast, SEC$_{SL}$ decreases clustering quality in the simulation study. A main reason for the bad performance of SEC$_{SL}$ is its (well-known) chaining tendency, generating parcellations with a few large and many very small clusters. We, therefore, recommend to employ SEC$_{AL}$ in order to further improve the quality of a parcellation. However, this
comes at the cost of increased computational complexity. A rough estimate of
the computation time of SEC_{AL} is that it takes twice as long compared with
SHAC_{AL}, since the generation of the cluster ensemble can be parallelized and
a separate ensemble hierarchy must be computed based on the cluster ensem-
ble, where both the computation of the cluster ensemble and of the ensemble
hierarchy take roughly as much computation time as SHAC_{AL}.

As recommended by Thirion et al (2014), we identify interesting numbers
of brain regions in a data-driven fashion employing a clustering stability and
a clustering quality approach. Both approaches perform equally well in the
simulation study, being able to stably identify the correct numbers of clus-
ters. In application to the 1000BRAINS data set, interesting numbers could be
identified by peaks in the clustering stability curves. Interestingly, these max-
ima are accompanied by changes, e.g., elbow points in the first derivatives of
the clustering quality curves. However, the clustering quality curves increase
monotonically and show no peaks. Nonetheless, we advise to employ both
approaches. Especially if a number of clusters is identified by both approaches,
the corresponding parcellation more confidently reflects a true level of brain
organization.

4.2 Neurobiological findings

The convergence of parcellations between different modalities, i.e. the question
of how well the borders identified based on one modality match those detected
based on a different modality, is a very important research topic (Eickhoff et al,
2018a). Eickhoff et al (2018a) argue that parcellations generated based on one
modality transfer to another modality, but do not reach the quality of parcell-
lations that are directly derived based on the other modality. Also Varikuti
et al (2018) suggest that parcellations generated based on one modality are to some extend transferrable to another modality for data reduction.

From our analysis we observe that the 1000BRAINS parcellations generated by SEC_{AL}(SPARTACUS\textsuperscript{S}) achieve a good within-modality, but a poor between-modality convergence. More precisely, the convergence with structural parcellations (Varikuti et al, 2018) is clearly better than chance. However, the convergence with established anatomical brain atlases or resting-state fMRI based parcellations is hardly better or worse than chance, respectively. Based on this observation we argue that parcellations derived from structural MRI data capture both spatial and structural patterns. However, mainly the spatial information is transferrable to other modalities. Apart from that, other modalities seem to capture different aspects of brain organization.

As the human brain is assumed to be organized in a multi level fashion, a single true number of brain regions is unlikely to exist. In our analysis based on the 1000BRAINS data set we identified $K \approx 70$, $K \approx 150$, $K \approx 300$ and $K \approx 600$ as interesting numbers of brain regions. These numbers may reflect different levels of brain organization.

However, our estimations can be only partly associated with granularity estimations made in the literature. E.g., the granularity of anatomical atlases is typically coarse, ranging around $K = 100$ (Tzourio-Mazoyer et al, 2002; Auzias et al, 2016). I.e. the first two numbers $K \approx 70$ and $K \approx 150$ are only roughly in the same order of magnitude. Other estimations evidently suggest that the number of cortical areas ranges around $K = 180$ (Amunts and Zilles, 2015; Glasser et al, 2016) and, e.g., the Julich-Brain atlas differentiates 248 cytoarchitectonic cortical areas and subcortical nuclei (Amunts et al, 2020). I.e. these estimations are larger then our whole brain estimation of $K \approx 150$. A possible explanation is that the voxel resolution of the 1000BRAINS data
set is not fine enough. Common granularity estimations based on data-driven approaches are $K \in [200,500]$, where these estimations are often based on reproducibility or prediction performance (Schaefer et al, 2018; Varikuti et al, 2018; Thirion et al, 2014; Van Essen et al, 2012). E.g., Thirion et al (2014) recommend a granularity of $K \in [200,500]$ based on reproducibility and Varikuti et al (2018) obtain the best age prediction results for 300 to 500 structural components. Thus, our estimation $K \approx 300$ is associated with these estimations. Finally, $K \approx 600$ is not established as interesting granularity in the literature.

These observations illustrates a major issue of granularity estimation, i.e. interesting numbers of brain regions are variable with respect to, e.g., data sets, modalities or resolutions. Thus we suggest to make granularity decisions individually adapted to the respective data scenario.

4.3 Future challenges

A (sparse) voxelwise statistical/machine learning analysis is suffering from a series of drawbacks. Both memory consumption and computational complexity are large. The number of features in a structural MRI data set typically exceeds the number of samples many times over. This problem is known in the literature as ”curse of dimensionality” and machine learning models derived from such data are prone to overfitting (Mwangi et al, 2014). Moreover, the interpretability of important features identified by such approaches is poor (Varikuti et al, 2018), since these features are isolated voxels embedded in a highly correlated spatial structure. More precisely, because of their small spatial size the voxels are too variable between subjects and because of their high correlation with other voxels, e.g., regularization methods such as LASSO regression (Tibshirani, 1996) can not perform a reliable feature selection (Varikuti et al, 2018). Hence, it is commonly agreed upon that there
should be some dimensionality reduction before developing a predictive model (Mwangi et al, 2014).

The SHAC algorithms parcellate the human brain into neurobiologically meaningful regions that allow for a good interpretability. These regions can be used as features in a subsequent machine learning analysis by representing each region, e.g., by the mean grey matter volume over all voxels in this region or the region’s first principal component. Inspired by Jiang et al (2020), another idea is to train a convolutional neural network (CNN) (LeCun et al, 2015) to each brain region using the grey matter volumes from this region as input. Afterwards, a weighted average of the predictions can be determined, allowing for an ensemble prediction and a ranking of regions according to their relevance for the final prediction represented by the weights.

We showed that standardizing the voxels prior to clustering has a positive effect on clustering quality. For future work it is interesting to investigate whether standardizing subjects instead of voxels has a similar effect.

4.4 Implementation in Python package

Python implementations of all procedures presented in this article are publicly available on Pypi (https://pypi.org/project/SPARTACUS10) and Github (https://github.com/totie10/SPARTACUS10). Note that SHAC_Ward is also implemented in scikit-learn’s (Pedregosa et al, 2011) AgglomerativeClustering function, producing the same results in less computational time. This observation suggests that our implementations have some potential for run-time optimization. Moreover, the code used to generate the results as well as the parcellations due to SEC_A (SPARTACUS^S) for K = 2, 4, . . . , 500 are available at Github (https://github.com/totie10/SPARTACUS10).
5 Conclusions

In conclusion, our study reveals some promising results that should be build upon in future studies. SHAC algorithms turned out to be stable and powerful methods for structural brain parcellation, outperforming SSPEC and geometric clustering. They parcellate the brain into spatially contiguous and, therefore, neurobiologically meaningful regions. Among the SHAC methods, SPARTACUS and SHAC\textsubscript{Ward} achieve the best performance on standardized data. By employing an average linkage based SHAC algorithm as consensus function in an ensemble clustering framework, we could further improve parcellation quality. Moreover, SHAC methods are well suited for identifying interesting granularities, since, once the hierarchy is computed, parcellations with any numbers of clusters are obtained at low computational cost. Using SHAC algorithms in a clustering stability and a clustering quality approach, we could identify four interesting numbers of brain regions from the 1000BRAINS data set. By reporting these interesting numbers as well as the corresponding parcellations we provide an additional view on the organization of the aging brain.

Finally, we observe that the 1000BRAINS parcellations generated by SEC\textsubscript{AL}(SPARTACUS\textsuperscript{S}) show a good within-, but a poor between modality convergence in relation to (random) geometric parcellations. Against the general assumption that patterns derived from one modality transfer to another modality, this observation suggests that mainly spatial patterns are transferrable between structural MRI and another modality, whereas structural patterns are only transferrable to structural MRI data but not to a different modality.
References

Aggarwal CC, Reddy CK (2014) Data clustering. Algorithms and applications. Chapman&Hall/CRC Data mining and Knowledge Discovery series, Londra, doi:10.1201/9781315373515

Alpaydin E (2014) Introduction to machine learning. MIT press

Amunts K, Zilles K (2015) Architectonic mapping of the human brain beyond Brodmann. Neuron 88(6):1086–1107. doi:10.1016/j.neuron.2015.12.001

Amunts K, Mohlberg H, Bludau S, et al (2020) Julich-brain: A 3d probabilistic atlas of the human brain’s cytoarchitecture. Science 369(6506):988–992. doi:10.1126/science.abb4588

Arbelaitz O, Gurrutxaga I, Muguerza J, et al (2013) An extensive comparative study of cluster validity indices. Pattern Recognition 46(1):243–256. doi:10.1016/j.patcog.2012.07.021

Arslan S, Ktena SI, Makropoulos A, et al (2018) Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. Neuroimage 170:5–30. doi:10.1016/j.neuroimage.2017.04.014

Auzias G, Coulon O, Brovelli A (2016) MarsAtlas: a cortical parcellation atlas for functional mapping. Human Brain Mapping 37(4):1573–1592. doi:10.1002/hbm.23121

Bellec P, Rosa-Neto P, Lyttelton OC, et al (2010) Multi-level bootstrap analysis of stable clusters in resting-state fMRI. Neuroimage 51(3):1126–1139. doi:10.1016/j.neuroimage.2010.02.082
Binder H, Schumacher M (2008) Adapting prediction error estimates for biased complexity selection in high-dimensional bootstrap samples. Statistical Applications in Genetics and Molecular Biology 7(1). doi:10.2202/1544-6115.1346

Blumensath T, Jbabdi S, Glasser MF, et al (2013) Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. Neuroimage 76:313–324. doi:10.1016/j.neuroimage.2013.03.024

Boongoen T, Iam-On N (2018) Cluster ensembles: A survey of approaches with recent extensions and applications. Computer Science Review 28:1–25. doi:10.1016/j.cosrev.2018.01.003

Brovelli A, Badier JM, Bonini F, et al (2017) Dynamic reconfiguration of visuomotor-related functional connectivity networks. Journal of Neuroscience 37(4):839–853. doi:10.1523/JNEUROSCI.1672-16.2016

Carvalho AXY, Albuquerque PHM, de Almeida Junior GZ, et al (2009) Spatial hierarchical clustering. Revista Brasileira de Biometria 27(3):411–442

Caspers S, Moebus S, Lux S, et al (2014) Studying variability in human brain aging in a population-based German cohort—rationale and design of 1000BRAINS. Frontiers in Aging Neuroscience 6:149. doi:10.3389/fnagi.2014.00149

Craddock RC, James GA, Holtzheimer III PE, et al (2012) A whole brain fMRI atlas generated via spatially constrained spectral clustering. Human Brain Mapping 33(8):1914–1928. doi:10.1002/hbm.21333

Davis SW, Cabeza R (2015) Cross-hemispheric collaboration and segregation associated with task difficulty as revealed by structural
and functional connectivity. Journal of Neuroscience 35(21):8191–8200.
doi:10.1523/JNEUROSCI.0464-15.2015

Desikan RS, Ségonne F, Fischl B, et al (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31(3):968–980.
doi:10.1016/j.neuroimage.2006.01.021

Eickhoff SB, Constable RT, Yeo BTT (2018a) Topographic organization of the cerebral cortex and brain cartography. Neuroimage 170:332–347.
doi:10.1016/j.neuroimage.2017.02.018

Eickhoff SB, Yeo BTT, Genon S (2018b) Imaging-based parcellations of the human brain. Nature Reviews Neuroscience 19(11):672–686.
doi:10.1038/s41583-018-0071-7

Glasser MF, Coalson TS, Robinson EC, et al (2016) A multimodal parcellation of human cerebral cortex. Nature 536(7615):171–178.
doi:10.1038/nature18933

Gordon EM, Laumann TO, Adeyemo B, et al (2016) Generation and evaluation of a cortical area parcellation from resting-state correlations. Cerebral Cortex 26(1):288–303. doi:10.1093/cercor/bhu239

Hubert L, Arabie P (1985) Comparing partitions. Journal of Classification 2(1):193–218. doi:10.1007/BF01908075

Jiang H, Lu N, Chen K, et al (2020) Predicting brain age of healthy adults based on structural MRI parcellation using convolutional neural networks. Frontiers in Neurology 10:1346. doi:10.3389/fneur.2019.01346
Jones DK, Symms MR, Cercignani M, et al (2005) The effect of filter size on VBM analyses of DT-MRI data. Neuroimage 26(2):546–554. doi:10.1016/j.neuroimage.2005.02.013

Kelly C, Uddin LQ, Shehzad Z, et al (2010) Broca’s region: linking human brain functional connectivity data and non-human primate tracing anatomy studies. European Journal of Neuroscience 32(3):383–398. doi:10.1111/j.1460-9568.2010.07279.x

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. doi:10.1038/nature14539

Lehoucq RB, Sorensen DC, Yang C (1998) ARPACK users’ guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods. SIAM, doi:10.1137/1.9780898719628

Lloyd S (1982) Least squares quantization in PCM. IEEE Transactions on Information Theory 28(2):129–137. doi:10.1109/TIT.1982.1056489

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5-th Berkeley Symposium on Mathematical Statistics and Probability, vol 1. University of California Press, pp 281–297

Moghimi P, Dang AT, Netoff TI, et al (2021) A review on mr based human brain parcellation methods. arXiv preprint arXiv:210703475

Monti S, Tamayo P, Mesirov J, et al (2003) Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Machine Learning 52(1-2):91–118. doi:10.1023/A:1023949509487
Mugler III JP, Brookeman JR (1990) Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). Magnetic Resonance in Medicine 15(1):152–157. doi:10.1002/mrm.1910150117

Mwangi B, Tian TS, Soares JC (2014) A review of feature reduction techniques in neuroimaging. Neuroinformatics 12(2):229–244. doi:10.1007/s12021-013-9204-3

Ng RT, Han J (2002) CLARANS: A method for clustering objects for spatial data mining. IEEE Transactions on Knowledge and Data Engineering 14(5):1003–1016. doi:10.1109/TKDE.2002.1033770

Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research 12:2825–2830

Pravitasari AA, Iriawan N, Fithriasari K, et al (2020) A Bayesian neo-normal mixture model (nenomimo) for MRI-Based brain tumor segmentation. Applied Sciences 10(14):4892. doi:10.3390/app10144892

Rolls ET, Huang CC, Lin CP, et al (2020) Automated anatomical labelling atlas 3. Neuroimage 206:116,189. doi:10.1016/j.neuroimage.2019.116189

Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20:53–65. doi:10.1016/0377-0427(87)90125-7

Schaefer A, Kong R, Gordon EM, et al (2018) Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cerebral Cortex 28(9):3095–3114. doi:10.1093/cercor/bhx179
Senbabaoğlu Y, Michailidis G, Li JZ (2014) Critical limitations of consensus clustering in class discovery. Scientific Reports 4(1):1–13. doi:10.1038/srep06207

Shen X, Tokoglu F, Papademetris X, et al (2013) Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. Neuroimage 82:403–415. doi:10.1016/j.neuroimage.2013.05.081

Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8):888–905. doi:10.1109/34.868688

Sotiras A, Resnick SM, Davatzikos C (2015) Finding imaging patterns of structural covariance via non-negative matrix factorization. Neuroimage 108:1–16. doi:10.1016/j.neuroimage.2014.11.045

Strehl A, Ghosh J (2002) Cluster ensembles — a knowledge reuse framework for combining multiple partitions. The Journal of Machine Learning Research 3(Dec):583–617. doi:10.1162/153244303321897735

Thirion B, Varoquaux G, Dohmatob E, et al (2014) Which fMRI clustering gives good brain parcellations? Frontiers in Neuroscience 8:167. doi:10.3389/fnins.2014.00167

Tibshirani R (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58(1):267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical
parcellation of the MNI MRI single-subject brain. Neuroimage 15(1):273–289. doi:10.1006/nimg.2001.0978

Van Essen DC, Glasser MF, Dierker DL, et al (2012) Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. Cerebral Cortex 22(10):2241–2262. doi:10.1093/cercor/bhr291

Van Oort ESB, Mennes M, Schröder TN, et al (2018) Functional parcellation using time courses of instantaneous connectivity. Neuroimage 170:31–40. doi:10.1016/j.neuroimage.2017.07.027

Varikuti DP, Genon S, Sotiras A, et al (2018) Evaluation of non-negative matrix factorization of grey matter in age prediction. Neuroimage 173:394–410. doi:10.1016/j.neuroimage.2018.03.007

Vendramin L, Campello RJGB, Hruschka ER (2010) Relative clustering validity criteria: A comparative overview. Statistical Analysis and Data Mining 3(4):209–235. doi:10.1002/sam.10080

Vigneau E, Qannari EM (2003) Clustering of variables around latent components. Communications in Statistics - Simulation and Computation 32(4):1131–1150. doi:10.1081/SAC-120023882

Vinh NX, Epps J (2009) A novel approach for automatic number of clusters detection in microarray data based on consensus clustering. In: 2009 Ninth IEEE International Conference on Bioinformatics and BioEngineering, IEEE, pp 84–91, doi:10.1109/BIBE.2009.19

Von Luxburg U (2007) A tutorial on spectral clustering. Statistics and Computing 17(4):395–416. doi:10.1007/s11222-007-9033-z
Von Luxburg U (2010) Clustering stability: an overview. Now Publishers Inc, doi:10.1561/2200000008

Ward Jr JH (1963) Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58(301):236–244. doi:10.1080/01621459.1963.10500845

Yang F, Li X, Li Q, et al (2014) Exploring the diversity in cluster ensemble generation: Random sampling and random projection. Expert Systems with Applications 41(10):4844–4866. doi:10.1016/j.eswa.2014.01.028

Yeo BTT, Krienen FM, Sepulcre J, et al (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology 106(3):1125–1165. doi:10.1152/jn.00338.2011

Yuan S, Tan PN, Cheruvelil KS, et al (2015) Constrained spectral clustering for regionalization: Exploring the trade-off between spatial contiguity and landscape homogeneity. In: 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), IEEE, pp 1–10, doi:10.1109/DSAA.2015.7344878

Zaki MJ, Meira W (2014) Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press, doi:10.1017/CBO9780511810114

Statements and Declarations

Funding

SBE acknowledges funding by the European Union’s Horizon 2020 Research and Innovation Program (grant agreements 945539 (HBP SGA3) and 826421 (VBC)).
**Competing Interests**

The authors have no relevant financial or non-financial interests to disclose.

**Author Contributions**

All authors contributed to the study conception and design. The data was provided by Simon B. Eickhoff. Investigation, methodology, software, data analysis, validation and visualization were performed by Tobias Tietz. Supervision and project administration were performed by Simon B. Eickhoff and Holger Schwender. The first draft of the manuscript was written by Tobias Tietz. All authors reviewed and commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Data Availability**

The 1000BRAINS dataset analyzed during the current study is not publicly available since a dissemination is not allowed. It is only available from the corresponding author on reasonable request.

**Ethics approval**

The local ethics committees of the university hospital Aachen and Heinrich-Heine-University Düsseldorf have approved the original study protocol of the data used here.

**Consent to participate**

Informed consent was obtained from all individual participants included in the study. The current data were analyzed anonymously.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryinformation.pdf