

Electronic Thesis and Dissertation Repository

8-4-2023 11:00 AM

Machine Learning Techniques for Improved Functional Brain Parcellation

Da Zhi, *The University of Western Ontario*

Supervisor: Diedrichsen, Jörn, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Computer Science

© Da Zhi 2023

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Computational Neuroscience Commons](#)

Recommended Citation

Zhi, Da, "Machine Learning Techniques for Improved Functional Brain Parcellation" (2023). *Electronic Thesis and Dissertation Repository*. 9436.

<https://ir.lib.uwo.ca/etd/9436>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Brain parcellation studies are fundamental for neuroscience as they serve as a bridge between anatomy and function, helping researchers interpret how functions are distributed across different brain regions. However, two substantial challenges exist in current imaging-based brain parcellation studies: large variations in the functional organization across individuals and the intrinsic spatial dependence which causes nearby brain locations to have a similar function. This thesis presents a series of projects aimed to tackle these challenges from different perspectives by using advanced machine learning techniques.

To handle the challenge of individual variability in building precise individual parcellations, Chapter 3 introduces a novel hierarchical Bayesian brain parcellation framework. This framework learns a brain probabilistic parcellation by integrating across diverse datasets. For single individuals, the framework optimally combines the limited individual data with the group probability map, resulting in improved individual maps. We found that the resultant individual parcellation based on only 10 minutes of imaging scans can achieve an equivalent performance to the ones using 100 minutes of data alone. These improved individual parcellations are essential to accurately capture functional variations across studied populations.

The intrinsic spatial dependence between brain locations poses a significant challenge in both evaluating and generating brain parcellations. To address this, Chapter 2 presents a bias-free method for evaluating different brain parcellations, the distance-controlled boundary coefficient (DCBC). Compared to existing evaluation metrics that bias toward finer and spatial contiguous parcellations due to spatial smoothness, DCBC provides a fair evaluation by controlling the distance of brain location pairs, ensuring a direct comparison of parcellations in different resolutions. To address the intrinsic spatial dependence when learning parcellations, I propose a new model in Chapter 4, the multinomial restricted Boltzmann machine (m-RBM), that can be incorporated into the learning framework in Chapter 3. This model captures spatial structure between brain locations in its architecture. While simulations showed the utility of this type of model in estimating individual parcellations, we could not demonstrate better performance using real imaging data.

Together, this thesis significantly advances the technical toolkit for deriving brain parcellations from functional imaging data. The developments open up new avenues for future research into human brain organization.

Keywords: Brain functional parcellations, Computational Neuroscience, Machine Learning

Summary for Lay Audience

Understanding how the human brain is organized and how different parts of the brain interact with each other is a critical part of neuroscience research. Just like we usually use maps when exploring a new city, brain parcellation subdivides the brain into different regions based on their functional or structural properties, helping scientists make sense of this intricate organ. But, this is not an easy task. Two main challenges stand in the way by the nature of the brain: the fact (1) that every person's brain has a slightly different organization and (2) that brain locations close to each other tend to function more similarly.

This thesis uses advanced machine learning techniques to tackle these issues. Just like how no two cities are alike, no two brains are identical, so it is important that brain parcellations reflect that. To address the first challenge, we developed a hierarchical Bayesian framework that learns the probability that a brain location belongs to a specific map, rather than learning a fixed map. This enables the model to produce individual brain parcellations by combining the group probability map with data from an individual following the Bayes rule, resulting in a more accurate personalized brain map. The framework is also able to learn combined knowledge from different datasets, thereby making full use of different experiments, each studying a different aspect of brain function.

The thesis also addresses the issue of intrinsic smoothness of the brain, the fact that spatially-nearby brain locations have higher functional similarity than the far-away pairs. This biological characteristic is an important consideration, not only when generating but also when evaluating brain parcellations. Without taking it into account, the boundaries between different functional regions in the trained brain maps might be noisy and difficult to identify, or the evaluation methods could unfairly favor those parcellations that are finer-grained and spatially continuous. To solve these problems, we designed a new computational model based on restricted Boltzmann machines (RBM) that explicitly models the spatial smoothness in the brain. The resultant maps capture the spatial structures between brain locations, helping in a more precise mapping of brain functions to individual brain regions. In addition, we also developed a new way of evaluating these brain maps, which reduces the bias in existing evaluation methods caused by the intrinsic spatial smoothness of the brain, making it a bias-free evaluation approach.

In essence, this thesis presents a series of projects to advance the methodological toolkit for producing brain parcellation maps. The tools may lay the foundation for clinical applications in personalized medicine, and be useful for future studies that address fundamental neuroscience questions.

Co-Authorship Statement

This thesis is presented in an integrated article format that contains two published papers and one manuscript in preparation. I would like to acknowledge all co-authors.

Chapter 2: Zhi, D., King, M., Hernandez-Castillo, C. R., Diedrichsen, J. (2022). Evaluating brain parcellations using the distance-controlled boundary coefficient. *Human Brain Mapping*, 43(12), 3706–3720. <https://doi.org/10.1002/hbm.25878>

- Da Zhi: Project conceptualization, software development, experiment analyses, manuscript drafting, and revision of the manuscript
- Maedbh King: Data collection, and revision of the manuscript
- Carlos R. Hernandez-Castillo: Data collection, and revision of the manuscript
- Jörn Diedrichsen: Project conceptualization, software development, and revision of the manuscript

Chapter 3: Zhi, D., Shahshahani, L., Nettekoven, C., Pinho, A. L., Bzdok, D., Diedrichsen, J. (2023). A hierarchical Bayesian brain parcellation framework for fusion of functional imaging datasets. *bioRxiv*, 2023-05. <https://doi.org/10.1101/2023.05.24.542121>

- Da Zhi: Project conceptualization, computational modeling, data pre-processing, manuscript drafting, and revision of the manuscript
- Ladan Shahshahani: Data pre-processing, and revision of the manuscript
- Caroline Nettekoven: Data pre-processing, and revision of the manuscript
- Ana Luísa Pinho: Data pre-processing, and revision of the manuscript
- Danilo Bzdok: Revision of the manuscript
- Jörn Diedrichsen: Project conceptualization, computational modeling, data pre-processing, and revision of the manuscript

Chapter 4: Zhi, D. and Diedrichsen, J. Spatially-informed models for individual brain parcellations. (*In preparation*)

- Da Zhi: Project conceptualization, computational modeling, manuscript drafting, and revision of the manuscript
- Jörn Diedrichsen: Project conceptualization, and computational modeling

Acknowledgements

I would like to express my most profound gratitude to my supervisor, Dr. Jörn Diedrichsen, for his full support and encouragement, insightful advice, and expert guidance throughout my Ph.D. studies and research. Without his incredible knowledge, patience, and counsel, this research would not have been possible.

I wish to say many thanks to all the members of the Diedrichsen Lab (past and present), and more broadly the Sensorimotor Superlab, for being wonderful colleagues and friends, for providing valuable feedback over the years, and for their company throughout my Ph.D. years. I would like to express my special thanks to Ladan Shahshahani for teaching me the key fMRI analysis paradigm and enriching me with the related neuroscience background.

Thanks to Carlos R. Hernandez-Castillo, Maedbh King, Ladan Shahshahani, Caroline Nettekoven, and Ana Luísa Pinho for all the help with the fMRI data collection and pre-processing throughout my research projects.

Thanks to my Ph.D. advisory committee members, Dr. Yalda Mohsenzadeh and Dr. Boyu Wang, for their pivotal mentorship over the years, and for being excellent role models in research.

Thanks to the Brain and Mind Institute (now Western Institute for Neuroscience) community at Western University as a whole for their efforts in cultivating a warm-hearted and invigorating research environment.

Finally, I would like to express my sincere appreciation to all my family members for their unending support throughout my academic journey. My deepest appreciation goes to my parents, Xianqing Kong and Xiangfu Zhi, for everything. Their unbreakable faith in me, even when faced with difficult challenges, has been my motivation to strive for excellence. I dedicate this thesis to them.

Contents

Abstract	ii
Summary for Lay Audience	iii
Co-Authorship Statement	iv
Acknowledgements	v
List of Figures	x
List of Tables	xi
List of Appendices	xii
1 General Introduction	1
1.1 Human Brain Organization	1
1.2 Background	2
1.2.1 Functional magnetic resonance imaging (fMRI)	2
1.2.2 Task-based and resting-state fMRI	4
1.2.3 Computational approaches to Brain parcellation	7
1.2.4 Evaluating brain parcellations	10
1.2.5 State-of-the-art brain parcellation	11
1.2.6 Group and Individual brain parcellations	14
1.3 Challenges of recent human brain organization studies	15
1.4 Thesis Objectives	16
2 Evaluating brain parcellations using the distance-controlled boundary coefficient	18
2.1 Introduction	18
2.2 Methods	21
2.2.1 Overview	21
2.2.2 Evaluation Data	21
2.2.2.1 Task-based dataset (MDTB)	21
2.2.2.2 Resting-state dataset (HCP)	23
2.2.3 Existing evaluation criteria for brain parcellations	24
2.2.4 Measuring spatial distance	25
2.2.5 Distance Controlled Boundary Coefficient (DCBC)	25

2.2.5.1	The problem of spatial smoothness	25
2.2.5.2	Averaging across bins	27
2.2.6	Random Parcellations	28
2.2.7	Random Functional Maps	28
2.2.8	Evaluation of commonly-used group parcellations	29
2.2.9	Parcellation based on the evaluation data	31
2.3	Results	32
2.3.1	Binning reduces the bias introduced by spatial smoothness	32
2.3.2	Adaptive weighted averaging reduces variance and bias	33
2.3.3	Choosing an appropriate bin width	34
2.3.4	DCBC evaluation for real data	34
2.3.5	Resting-state group parcellations predict task-based functional boundaries	36
2.3.6	Comparison to parcellations derived from the evaluation data set	38
2.3.7	Anatomical parcellations do not predict task-based functional boundaries	39
2.3.8	Multi-modal parcellations do not perform better than resting-state parcellations	40
2.3.9	Comparison across different spatial resolutions	40
2.3.10	Open-source toolbox/data support evaluation	41
2.4	Discussion	41
3	A hierarchical Bayesian brain parcellation framework for fusion of functional imaging datasets	47
3.1	Introduction	47
3.2	Results	50
3.2.1	Individual parcellations in the scarce data setting	50
3.2.2	Dataset-specific emission models optimally capture differences in measurement noise	52
3.2.3	Region-specific concentration parameters further improve fusion parcellation	54
3.2.4	Model performance on real data and the choice of atlas resolution K	56
3.2.5	The fusion atlas shows combined strengths across different task-based fMRI datasets	59
3.2.6	Integrating resting-state data into the task-based parcellation	61
3.3	Discussion	62
3.4	Methods	68
3.4.1	A hierarchical Bayesian parcellation framework for data fusion	68
3.4.1.1	EM algorithm for Probabilistic parcellation with data fusion	69
3.4.1.2	Dataset-specific emission models	70
3.4.1.3	The spatial arrangement model	71
3.4.1.4	Message passing and collaborative learning	72
3.4.1.5	Individual and group parcellations	75
3.4.1.6	Initialization and convergence	75
3.4.2	fMRI Datasets	76

3.4.3	Data structure and anatomical normalization	79
3.4.4	Synthetic datasets for simulation	80
3.4.5	Evaluation measures for probabilistic atlas	81
3.4.6	Computational setup	82
3.5	Data availability	82
3.6	Code availability	82
4	Spatially-informed models for individual brain parcellations	83
4.1	Introduction	83
4.2	Results	85
4.2.1	m-RBM model captures spatial smoothness better than the independent model when estimating individual parcellations	85
4.2.2	Convergence problems when training the m-RBM model	89
4.2.3	Smoothing does not improve the estimation of individual functional boundaries on real fMRI data	91
4.3	Discussion	93
4.4	Methods	98
4.4.1	The m-RBM spatial arrangement model	98
4.4.2	Training m-RBM model using stochastic maximum likelihood	100
4.4.2.1	Positive phase: expectation given the data	102
4.4.2.2	Negative phase: expectation given the model	103
4.4.2.3	Update phase: parameter estimation using gradients	104
4.4.3	Algorithm implementation and convergence properties	105
4.4.4	fMRI dataset	107
4.4.5	Synthetic datasets for simulation	108
4.4.6	Evaluation measures for probabilistic atlas and simulation	110
4.4.6.1	DCBC	110
4.4.6.2	Mean expected prediction error	111
4.4.6.3	Mean adjusted expected cosine error	111
4.4.7	Computational setup	113
4.5	Data availability	113
4.6	Code availability	113
5	General Discussion	114
5.1	Summary of the thesis	114
5.2	Using atlases: individual functional localizer	117
5.3	Building atlases: the choice between task-based and resting-state fMRI	119
5.4	Evaluating atlases: aligning functions to brain regions	121
5.5	Interpreting brain organizations as gradient or parcellation	122
5.6	Extensions of current work	123
5.6.1	The optimal number of parcels	124
5.6.2	Hierarchical structure of brain parcellation	124
5.6.3	Deep generative architecture for improved spatial arrangement model	125
5.7	Conclusion	126

Bibliography	127
A Supplementary Materials and Figures for Chapter 3	141
B Supplementary Materials and Figures for Chapter 4	149
Curriculum Vitae	151

List of Figures

2.1	Distance controlled boundary coefficient.	26
2.2	Random cortical parcellations with different number of parcels.	28
2.3	Evaluation results on random parcellations.	32
2.4	Evaluation on the real data sets.	35
2.5	Evaluation of existing group parcellations.	37
3.1	A hierarchical Bayesian parcellation framework for data fusion.	49
3.2	Individual parcellations from the parcellation framework outperform group map.	51
3.3	Simulations of data fusion using two synthetic imaging sessions with similar task activation.	53
3.4	Simulation on two synthetic sessions fusion with different task activation.	55
3.5	DCBC evaluation of brain parcellations learned on two IBC sessions alone compared to the parcellation learned by fusion.	57
3.6	Comparison of cerebellar parcellations learned by type 2 and 3 fusion models using 6 functional task-based datasets.	59
3.7	Performance comparison of the cerebellar parcellations using resting-state data only, task-based data only, or the combination of both.	61
4.1	The multinomial-restricted Boltzmann machine (m-RBM) spatial arrangement model.	84
4.2	Comparison of the m-RBM and the independent arrangement model on simulated data.	87
4.3	Simulation of the convergence issue when training m-RBM model.	90
4.4	Comparison of individual parcellations estimated by the m-RBM vs. independent model using MDTB dataset.	92
A.1	Performance comparison between the parcellations derived from Gaussian Mixture Model (GMM) and von Mises-Fisher Mixture model (VMF).	144
A.2	The synthetic dataset.	145
A.3	Simulation on two synthetic sessions fusion with similar task activation using Type 1, 2, and 3 emission models.	146
A.4	Comparing the performance of Type 2 and 3 models when the number of parcels K used for fitting is different from the true K in the simulation.	147
A.5	The visualization of the learned group maps ($K = 34$).	148
B.1	The synthetic dataset.	150

List of Tables

2.1	Commonly used group-level cortical parcellations.	30
3.1	FMRI datasets used for the functional fusion.	77

List of Appendices

Appendix A: Supplementary Materials and Figures for Chapter 3	141
Appendix B: Supplementary Materials and Figures for Chapter 4	149

Chapter 1

General Introduction

1.1 Human Brain Organization

The study of human brain organization is an important part of neuroscience research. It addresses the fundamental question of how different regions of the brain interact and collaborate to produce a broad range of cognitive and behavioral outcomes. A common approach to understanding brain organization involves subdividing the brain into structurally or functionally distinct regions ([Felleman and Van Essen, 1991](#); [Eickhoff et al., 2018](#)), a process known as brain parcellation. Early atlases of the human brain were based on the cytoarchitectonic organization of the neocortex ([Brodmann, 1909](#); [Zilles et al., 2002](#); [Talairach, 1988a](#)). These anatomical parcellations laid the groundwork for our understanding of the complex structure and functions of the brain and are still contributing to today's neuroscience studies.

In recent years, the advent of neuroimaging technologies, such as functional magnetic resonance imaging (fMRI), offers an alternative approach by acquiring non-invasive brain functional images in large samples of subjects. As a result, a number of fMRI-based brain parcellations based on functional properties have been proposed in recent years ([Yeo et al., 2011](#); [Buckner et al., 2011](#); [Gordon et al., 2016](#); [Schaefer et al., 2018](#); [King et al., 2019](#)). These functional parcellations delineate regions of the brain according to the similarity of their functional connectivity patterns, using resting-state or task-based fMRI data ([Arslan et al., 2018](#); [Eickhoff et al., 2018](#)). In addition, multi-modal parcellations have

recently emerged by integrating various types of data, such as structural, functional, and cytoarchitectonic information, providing a more comprehensive understanding of brain organization (Glasser and Van Essen, 2011; Fan et al., 2016).

Functional parcellations play an important practical role in the study of brain organization. The primary usage of a parcellation is to reduce the complexity of the statistical analysis, from tens of thousands of measurement units (vertices or voxels) to a much smaller set of individual regions. For example, parcellations are commonly used to define regions of interest (ROIs) to summarize functional and anatomical data, or to define the vertices for subsequent connectivity analysis (Sporns, 2011). A widely-accepted parcellation map allows for direct comparison between studies (Arslan et al., 2018). Universally, any parcellation should aim to define regions such that the characterization (whether anatomical measures, connectivity patterns, or task activation) of neural populations within the same region should be maximally similar to each other. In contrast, neural populations assigned to different parcels should be maximally different. Therefore, brain parcellation can be viewed as a clustering problem. As a result, standard machine learning methods can be applied to brain parcellation.

In the rest of this Chapter, I will review the background for the thesis, including the necessary details concerning the fMRI data, especially the difference between task-based and resting-state fMRI. I will then describe computational approaches for deriving brain parcellations and review how they are used in the current state-of-the-art projects. Lastly, I will come to the major challenges for brain organization followed by the thesis objectives.

1.2 Background

1.2.1 Functional magnetic resonance imaging (fMRI)

Functional magnetic resonance imaging (fMRI) is a non-invasive neuroimaging technique that allows researchers to investigate the functions of the human brain. By utilizing the blood oxygenation level-dependent (BOLD) contrast (Ogawa et al., 1990), fMRI measures

changes in blood flow and oxygenation levels in response to neural activity. The BOLD signal arises from the coupling between neural activity and local changes in blood flow, allowing researchers to infer which brain regions are engaged during specific cognitive tasks or resting-state conditions. This technique has revolutionized the field of cognitive neuroscience by providing valuable insights into the functional organization of the brain and its role in various cognitive processes (Logothetis, 2003).

The BOLD signal (or BOLD contrast) in fMRI is based on the principle that when neurons are active, they require more oxygenated blood to meet the metabolic demand compared to when they are inactive. As a result, the brain's vasculature responds by increasing blood flow to the active regions, supplying the necessary oxygen and nutrients. The increased blood flow leads to changes in the ratio of oxygenated to de-oxygenated hemoglobin, causing alterations in the magnetic properties of the blood. These magnetic changes can be detected using specialized MRI techniques (Ogawa et al., 1990), enabling researchers to map brain activity with high spatial resolution. By analyzing the spatiotemporal patterns of the BOLD signal, researchers can identify brain regions involved in specific tasks and gain insights into the neural networks underlying various cognitive functions (Norman et al., 2006).

The BOLD fMRI technique has become an important tool in neuroimaging research. However, it is also crucial to note that fMRI is an indirect measure of neural activity and has certain limitations, preventing direct analysis use. For example, the measurement noise from the scanner, head or body motion during the scanning, and the brain segmentation for the region of interest. Therefore, an fMRI data pre-processing step is often required for subsequent analyses and statistical modeling, including brain segmentation, realignment and registration of functional images, normalization, etc.

The preprocessing of fMRI data is essential for the following neuroimaging studies. The aim of preprocessing is to transform the raw data into a format suitable for statistical analysis. The first stage of preprocessing usually involves the segmentation of the anatomical image into white matter, gray matter, cerebrospinal fluid (CSF), etc, by masking each of these structures. It is because we are only interested in the neural activities within the gray matter, and other structures are masked out as noise. Then

the functional data undergoes motion correction since subjects could move their heads during the scanning, introducing spatial variation in the data that is not related to any neural activity. Therefore, a reference image is selected and all the other functional images are then aligned to this reference image, estimating the movement of each image with respect to the baseline (Friston et al., 1995; Jenkinson et al., 2002). If an image shows a large movement, then it can be excluded from the subsequent steps. After that, the mean functional images need to be co-registered with the segregated anatomical image, aligning the lower-resolution functional image with the underlying higher-resolution anatomical image. Lastly, spatial normalization is another key step, which involves warping the subject's data to match a standard anatomical template, such as the Talairach (Talairach, 1988b), MNI152 (Fonov et al., 2011), or Freesurfer's fsaverage surface space (Fischl, 2012), depending on specific usage. This process ensures that equivalent regions across subjects are compared, allowing for group-level analysis.

1.2.2 Task-based and resting-state fMRI

Traditionally, fMRI was utilized to measure changes in brain activity when participants are performing specific behavioral tasks (Fox and Raichle, 2007; Logothetis, 2008; Raichle and Mintun, 2006). Task-based fMRI is commonly used in early fMRI studies to identify brain functional regions that are active for a specific behavioral task. For example, during a finger-tapping task, the neurons in the motor cortex will be activated and the blood will flow in to deliver more oxygen to this region, resulting in a change in the ratio of oxygenated to deoxygenated blood (Ogawa et al., 1990). Therefore, the motor network associated with finger movement tasks can be located. More typically, task-based fMRI studies compare a task of interest to a well-matched control task to isolate activity due to a specific mental process of interest.

The processed task fMRI data is typically modeled using a general linear model (GLM) framework as a first-level analysis. The GLM is a statistical model that allows researchers to estimate the contribution of different experimental conditions or tasks to the measured fMRI signal. The model includes a set of predictors, representing the expected neural response for each condition, convolved with a canonical hemodynamic

response function (HRF) (Friston et al., 1998a,b) that captures the delayed and prolonged nature of the hemodynamic response. The raw fMRI data is then fit to this GLM using regression analysis, resulting in beta weights that represent the strength of the relationship between each condition and the measured fMRI signal at each voxel. To obtain the beta weights, various estimation methods can be employed, such as ordinary least squares (OLS) or weighted least squares (WLS) (Diedrichsen and Shadmehr, 2005). These methods aim to find the best-fitting parameters that minimize the difference between the predicted response based on the GLM and the actual fMRI signal. These resultant beta weights can then be further analyzed, such as by conducting statistical tests to assess the significance of the observed effects and making inferences about the underlying neural processes involved in the task (Friston et al., 1995).

In contrast, resting-state fMRI (rs-fMRI) measures the BOLD signal when participants are ‘resting’ in the scanner without a specific task (Biswal et al., 1995). The BOLD signal in different brain regions captures the spontaneous fluctuations in neural activity that occur in that state. It has been shown that regions that have high anatomical connectivity with others and are also usually co-activated for specific tasks, tend to show highly correlated fluctuations. Therefore the correlations between the time series of different regions can be used as an estimate of the intrinsic functional connectivity of the brain (Greicius et al., 2003; Fox and Greicius, 2010; Uddin et al., 2010).

Unlike the GLMs for task-based fMRI, the rs-fMRI does not have an independent manipulation that can be used to build a model because there is no task design to provide a regressor for a first-level analysis. Therefore, resting-state functional connectivity (rs-FC) is often used to reflect this synchronized neuronal activity in rs-fMRI by measuring the statistical dependence between BOLD signal time courses in spatially distributed brain regions directly. One common method for analyzing rs-fMRI data is seed-based connectivity analysis (Biswal et al., 1995). In this method, a region of interest (ROI or ‘seed’) is defined, and the BOLD signal time course within this ROI is correlated with the time courses from all other brain voxels to create a connectivity map. This map represents how functionally connected the rest of the brain is with the seed region. Another similar approach derived from the seed-based connectivity analysis is the ROI-

to-ROI-based functional connectivity analysis. However, rather than correlating the time series from one seed region to every other voxel in the brain, in ROI-to-ROI analysis, the average time series of each ROI is correlated with the average time series of every other ROI. This method provides a broader view of the functional relationships between all pairs of selected regions, giving a more holistic picture of brain network interactions. An alternative to the seed-based approaches is based on graph theoretical analysis (Power et al., 2011), where the brain is modeled as a graph network. The voxels or ROIs can be represented as vertices and the correlations between the vertices are the connectivity weights of the edges. One advantage of this approach is to capture and quantify the topological properties of brain networks in the graph, providing a more comprehensive understanding of brain organization compared to conventional seed-based methods. Other data-driven approaches attempt to define the underlying brain regions for rs-FC analyses rely on Independent Component Analysis (ICA), to decompose rs-fMRI data into spatially independent components (Beckmann et al., 2005), or clustering methods, to group voxels into regions based on their similarity measure (Salvador et al., 2005). A detailed survey of methods for measuring resting-state functional connectivity in the literature can be found elsewhere (Lee et al., 2013).

Depending on what goal is in mind, care needs to be taken when we study brain organizations based on which type of functional profile, rs-FC or task activation weights, as they have their own advantages and disadvantages. The former does not require subjects to perform specific tasks, allowing researchers to study even babies or clinical populations that cannot perform a task reliably anymore. It can be acquired relatively quickly and easily, making it suitable for large-scale studies and clinical applications (Lee et al., 2013). On the other hand, task-based fMRI investigates specific cognitive processes by designing tasks, enabling hypothesis-driven investigations to identify brain regions associated with specific functions in task performance. In the context of brain parcellation, this advantage is sometimes also considered a disadvantage: the parcellation will likely depend substantially on the task set that was chosen during fMRI acquisition, as different tasks will activate different brain regions in different patterns. In contrast, the resting state is often considered a “neutral” baseline state that reveals the intrinsic

connectivity of the brain. I will return to this claim and the debate surrounding it in Chapter 5.

1.2.3 Computational approaches to Brain parcellation

Theoretically, functional brain parcellations aim to subdivide the brain into distinct regions based on functional data with higher within-parcel homogeneity, while keeping brain locations with lower similarity in different regions. This goal can be achieved by two different concepts: global similarity (not spatially informed) and local gradient (spatially informed) approaches (Schaefer et al., 2018; Eickhoff et al., 2018). The global similarity approaches seek to group brain-wide voxels or vertices with similar connectivity patterns, the resulting parcels might be highly functionally homogeneous, suggesting that they can better represent the whole brain networks and benefits the further dimensionality reduction analysis. On the other hand, the local gradient approaches seek to identify boundaries between functional areas by capturing sharp changes in connectivity patterns across neighbouring brain locations. These methods, being spatially informed, inherently respect the underlying spatial organization of the brain, capturing local variations and individual differences in brain organization. While both global and local approaches offer valuable insights into brain organization, the choice between them often depends on the specific research questions being addressed.

Among global similarity approaches, the commonly used techniques are based on clustering algorithms from a machine learning perspective. Initial clustering methods for brain parcellation projects utilized the k-means algorithm (Lloyd, 1982) to assign each voxel to a cluster based on the distance measure to its cluster centroid. This algorithm is generally fast and easy to implement but it is sensitive to the initial choice of centroids and biased towards equal-size clusters (Tan et al., 2016). Therefore, several variants of the method have been proposed to deal with these drawbacks of k-means algorithms for brain parcellation problems (Lee et al., 2012; Thirion et al., 2014). Another group of clustering methods for brain parcellations was based on hierarchical-clustering algorithm (Bellec et al., 2010; Mumford et al., 2010) to subdivide the cortex with a bottom-up process, where boundaries derived from the lower resolutions are propagated to higher levels. By

joining parcellated regions at different levels of the hierarchical clustering tree, the global similarity or homogeneity is increased. In contrast, more recent hierarchical clustering-based approaches (Blumensath et al., 2013; Arslan and Rueckert, 2015; Honnorat et al., 2015) improved the resultant parcellations by applying additional (or multiple) clustering layers or combined with other approaches.

Clustering can also be achieved via probabilistic mixture models, which can be considered as “soft k-means”. The statistical modeling assumes the data of voxels or vertices within a functional region follow some statistical distributions, hence, different regions can be clustered by using different parameterized models. In the context of statistical-based clustering approaches, general mixture models are the most popular family for brain parcellation tasks. One basic type of brain functional parcellations was estimated using the Gaussian Mixture model (GMM) (Golland et al., 2007). GMM assumes that the observed fMRI data is generated from a mixture of a finite number of Gaussian distributions, each characterized by its own mean and variance parameters. This model has the advantage of managing complex data distributions and high-dimensional data, enabling simultaneous consideration of diverse and rich signals from fMRI. Recently, the von Mises-Fisher (vMF) mixture model (Banerjee et al., 2005) has emerged as a promising approach for this task as it is particularly well-suited for capturing directional data distributions, such as the rs-FC profile which usually ignores the signal amplitude by the data standardization in brain imaging. Several studies have demonstrated the utility of a mixture of vMF-based approaches in accurately delineating functional brain networks and capturing subtle variations in connectivity patterns (Yeo et al., 2011; Schaefer et al., 2018; Kong et al., 2019). These statistical approaches aim to estimate the parameters of the model, focusing on the generation of each voxel’s functional profile that is assumed to follow a parametric model with unknown parameter values.

Another type of global similarity approach is based on matrix decomposition (factorization) methods, which are tightly correlated to clustering (Ding et al., 2005). The key idea of these methods is that the high-dimensional fMRI data matrix \mathbf{Y} of brain voxels with their functional profiles can be decomposed as the product of different lower-dimensional components \mathbf{V} and \mathbf{W} that carry specific interpretations, such as $\mathbf{Y} = \mathbf{V}\mathbf{W}$.

In the context of brain parcellation, the matrix V might represent functional connectivity patterns of the identified parcels, while W can be viewed as a parcellation of the brain into different regions. The most commonly used matrix decomposition methods for brain parcellation projects are Principal Component Analysis (PCA) and Independent Component Analysis (ICA). A similar algorithm, non-negative matrix factorization (NMF) (Ding et al., 2005), has been frequently used to solve brain parcellation problems (Varikuti et al., 2018; King et al., 2019), which constrains the decomposed components to be strictly non-negative. It enables the division of the brain into localized components that reflect actual brain regions (Sotiras et al., 2015). Recently, spectral clustering (Ng et al., 2001), another matrix decomposition algorithm, based on spectral graph theory (Von Luxburg, 2007) quickly emerged for brain parcellation tasks (Craddock et al., 2012; Shen et al., 2013). This algorithm performs matrix factorization based on the eigenvectors of the matrix of similarity (such as connectivity) between brain locations (voxels or vertices). It keeps the spatial integrity of the parcellations since only vertices sharing the same cluster membership can be connected in the adjacency matrix. But, spectral clustering approaches tend to create similarly sized parcels and impose spatial constraints in order to yield spatially contiguous parcels (Parisot et al., 2016; Arslan et al., 2015).

While these global similarity methods hold great promise for advancing our understanding of brain organization, however, they usually neglect the spatial distance of different brain locations, resulting in some parcels that are spatially disconnected or the neighboring brain locations having different parcellation labels (Homnorat et al., 2015). In contrast, the local gradient approaches delineate functional boundaries in the brain based on the fact that the functional profiles can abruptly change from one spatially contiguous region to a nearby region. This sharp change can be detected by computing the local gradient in functional patterns. The most commonly-used techniques from previous work for delineating such cortical regions include edge detection and boundary mapping (Hirose et al., 2012; Wig et al., 2014). Pioneering work (Cohen et al., 2008) found that rs-fMRI patterns show sharp transitions in correlation patterns and that these putative areal boundaries can be reliably detected in both individual and group data. Other works further utilized boundary mapping approaches for identifying sub-regions in the lateral

parietal cortex (Nelson et al., 2010) and the basal ganglia (Barnes et al., 2010). Furthermore, (Gordon et al., 2016) described a technique for using resting-state functional connectivity to define parcels that represent putative cortical areas by a region-growing algorithm. In general, the local gradient approaches implicitly constrain parcels to be spatially connected, reflecting the nature of spatial smoothness.

To address the limitations that appeared in both approaches, Recent studies have proposed to impose spatial prior from Markov Random Field (MRF) model to the statistical models, which integrated both local gradient and global similarity approaches with resting-state functional connectivity data (Ryali et al., 2013; Schaefer et al., 2018; Kong et al., 2019). The resultant brain parcellation outperformed both pure local and global approaches in terms of the homogeneity of the functional signal within the derived regions. Thus, combining local boundary detection with global clustering is a promising direction for future computational models of brain parcellation.

1.2.4 Evaluating brain parcellations

Using different methods, a large number of competing brain parcellations have been proposed over recent years. In the absence of knowing a ground truth, the question arises of how to evaluate these parcellations. The consistency and reliability between two brain parcellations can be assessed with the Dice coefficient (Dice, 1945) or Adjusted Rand Index (ARI) (Hubert and Arabie, 1985). However, these metrics do not tell us how good a parcellation is, which only provides a measure of how similar a brain parcellation is with regard to another parcellation. On the other hand, assessing the quality and validity of brain parcellations presents an inherent challenge, primarily due to the absence of a universally-accepted parcellation for direct comparison. Thus, the best we can do is to check how well a given parcellation captures functional boundaries by testing it against independent datasets. These data sets can be a broad spectrum of anatomical information, functional resting-state, or task-based fMRI data, depending on what evaluation goal is in mind. For example, if one wants to assess how well a brain parcellation is when predicting task-relevant functional boundaries, then a task-based fMRI dataset should be chosen as the test set as it gives us insights into how brain regions activate in response

to specific tasks.

One popular assessment of the validity of a parcellation is the *Silhouette coefficient* (Rousseeuw, 1987), which compares the average dissimilarity from one vertex to all other vertices in the same parcel, to the average dissimilarity from the same vertex to all the vertices that assigned to neighboring parcels. Similarly, *inter/intra-cluster distance* (Bzdok et al., 2015) compares the Euclidean distance between the cluster centers to the distance between the elements within each cluster. Such criteria describe the goal of brain parcellation to form groups such that brain locations within a group show higher similar connectivity, while the connectivity is different between groups. Another commonly-used evaluation method is *global homogeneity* (Craddock et al., 2012; Gordon et al., 2016). Homogeneity is defined as the average similarity across all pairs of vertices within a parcel, where the similarity measure of two vertices is usually defined as Pearson’s correlation between functional profiles. The global Homogeneity is then simply calculated as the average with-parcel correlation across all parcels, with higher homogeneity suggesting a better parcellation.

1.2.5 State-of-the-art brain parcellation

In this section, I will summarize the state-of-the-art brain parcellations for the human neocortex and cerebellum, ranging from anatomical, functional resting-state or task-based, and multi-modal. The goal of this section was to give an overview of the current status of brain parcellations as important to the topic of this thesis - interested readers will find a more detailed review elsewhere (Arslan et al., 2018; Eickhoff et al., 2018).

The study of brain parcellation has a long tradition, tracing back to the 19th and early 20th centuries. Early studies focused on the spatial distribution of cell types in different layers of the neocortex, a feature discernible after tissue staining. Observations highlighted that these distributions varied significantly across the brain, revealing regions of uniform cytoarchitecture and abrupt changes between regions. An influential subdivision of the brain into discrete areas based on the cellular architecture of the neocortex was provided by Brodmann (Brodmann, 1909). Later work considered different local properties, in particular myeloarchitecture, to define brain areas (Klatzo, 2002). Other

anatomical-based parcellations enriched the field by delineating maps using cortical sulci and gyri (Desikan et al., 2006; Fischl et al., 2004). However, the main drawback of these anatomical brain mapping studies is that they subdivide the brain into parcels based on anatomical landmarks, which cannot reflect the true functional boundaries (King et al., 2019).

Recently, a number of fMRI-based brain parcellations have been proposed in the last decades for the human neocortex and cerebellum. In this category, the vast majority of the studies proposed functional parcellations based on resting-state functional connectivity profiles. For the cerebral cortex, one commonly-used group resting-state parcellation is proposed by Yeo et al. (2011) in 2011 to separate the neocortex into 7 and 17 networks using data from 1000 subjects, each network containing multiple spatial distributed components. Later work built finer-grained group (Schaefer et al., 2018) or individual (Kong et al., 2021) resting-state parcellations aligned to the *Yeo 7* and *Yeo 17* networks with higher resolution (up to 1,000 parcels) in surface representation. Other cortical parcellations include *Craddock 2012* (Craddock et al., 2012) with 10 to 1000 parcels, *Gordon 2016* (Gordon et al., 2016) with 333 parcels, *Bellec 2010* (Bellec et al., 2010) with multiple resolutions from 7 to 444 parcels, *Power 2011* (Power et al., 2011) of 130 networks, and *Shen 2013* (Shen et al., 2013) with 200 parcels covering the whole cortex. In contrast, few recent parcellations highlighted the individual variability in brain organization by proposing individual-specific resting-state parcellations, such as *Wang 2015* (Wang et al., 2015) of 18 networks and *Gordon 2017* (Gordon et al., 2017b) with subject-specific resolutions.

While most previous work concerns the organization of the neocortex, few studies exclusively focused on the cerebellum. For the cerebellar functional organization, one pioneering work of resting-state group parcellations proposed by Buckner et al. (2011) subdivided the cerebellum into 7 and 17 regions from 1000 subjects. In 2013, the release of the Human Connectome Project (HCP) (Van Essen et al., 2013), which is a milestone of resting-state parcellation studies, has made a significant contribution to the field. With an increasing amount of evidence suggesting that the human cerebellum involves many cognitive tasks, cerebellar functional organization studies have become important. Ji

[et al. \(2019\)](#) presented a parcellation of subcortical structures with 10 networks based on its correlation to their cortical networks using the resting state data from HCP S900 release ([Van Essen et al., 2013](#)). Later on, [Xue et al. \(2021\)](#) developed two individual parcellations with 10 parcels based on resting-state data from 31 sessions for each.

Compared to resting-state, task-based fMRI data has been used less for deriving brain parcellations. One reason for this is that brain parcellations depend strongly on the tasks used. Therefore, early task-based approaches used a large number of task-contrast maps assembled over different studies. For example, *Yeo 2015* ([Yeo et al., 2015](#)) derived a parcellation using 10,449 contrast maps across 83 behavioral tasks using a hierarchical Bayesian model. Only recently, a number of research groups have conducted task-based fMRI studies that include a broad range of tasks covering multiple domains in the same subject ([Van Essen et al., 2013](#); [Pinho et al., 2018](#); [Nakai and Nishimoto, 2020](#); [King et al., 2019](#)). This allows the community to reveal a comprehensive brain functional organization by utilizing task-based parcellations.

For the cerebellum, this has led to a new task-based functional parcellation using a multi-domain task battery (MDTB, [King et al. \(2019\)](#)), containing a wide range of social and cognitive tasks, resulting in a task-based parcellation with 10 regions for the cerebellum ([King et al., 2019](#)). In general, task-based parcellation studies are relatively new, it is partially due to the lack of a large and homogeneous task-based dataset across multiple task domains with enough subjects. Even if the task set is deliberately chosen to be broad, each task set will still have some cognitive domains that are not fully covered, potentially resulting in biases in the parcellation. Therefore, a principled way of combining or fusing the knowledge of different task-based datasets would be an important step forward.

In addition, several multi-modal parcellations were proposed using the features from anatomical, resting-state, and task-based fMRI data. Such parcellation combines the strength from different modalities and finds the maximum agreement across all features. For example, a multi-modal group parcellation, *Glasser 2016*, based on rs-FC and cytoarchitectonic information (eg. Brodmann areas and myelin content) from 210 HCP subjects was proposed in [Glasser et al. \(2016\)](#), or *Fan 2016* parcellation was proposed in

Fan et al. (2016) using rs-FC and anatomical information (see Table 2.1 for more details).

1.2.6 Group and Individual brain parcellations

Group-level parcellations aim to create a common template or atlas that represents the average organization of the brain across a group of individuals. These parcellations are constructed by aggregating data from multiple subjects to identify consistent patterns of connectivity or anatomical boundaries. Group parcellations provide a standard reference for comparing brain regions across individuals and can help establish common frameworks for studying brain function and structure in the studied population. These group-level maps have proven invaluable for characterizing the brain’s general functional organization (Buckner et al., 2013) and for comparing different populations, such as healthy controls and patients with a particular neurological disorder (Wang et al., 2007; Barthel et al., 2011). Techniques for group-level parcellation generally involve pooling neuroimaging data from a large number of individuals, then identifying common patterns of connectivity using computational approaches. For example, the “Yeo 2011” parcellation in Figure 2.4 was estimated from the concatenated resting-state fMRI time series across 1000 healthy subjects (Yeo et al., 2011).

In contrast, individual-level parcellations aim to capture the unique structural and functional organization of each person’s brain, as recent studies have shown that inter-individual difference exists for both brain structure and the function (Braga and Buckner, 2017; Gordon et al., 2017a; Kong et al., 2021). These parcellations acknowledge the reality of individual variability in brain organization, which can arise due to a variety of factors, including age, gender, genetics, and experiences. Individual parcellations have been found to provide a more accurate mapping of functional regions for any given individual, which can be especially useful in a clinical context where precision is important. They can have significant implications for understanding individual differences in cognition, behavior, and susceptibility to neurological or psychiatric disorders. The generation of individual parcellations relies on high-quality individual functional profiles, however, creating reliable individual parcellations can be challenging (see 1.3). A key focus in this area is developing methods that can reliably generate individual parcellations from

a limited amount of data.

1.3 Challenges of recent human brain organization studies

In the field of human brain organization research, one of the major challenges is managing the inter-individual variability that characterizes brain organization (Eickhoff et al., 2018; Bijsterbosch et al., 2020). Since human brains are uniquely complex and exhibit substantial variation across different individuals, inter-individual variability challenges brain organization studies in both the generation and evaluation of brain parcellations. In practice, individual parcellations are often desired as individual functional localizers for clinical diagnosis and subsequent analyses. A common way to obtain such individual functional localizers in previous studies is to run a few minutes of individual scans for additional tasks before the main study (Kanwisher et al., 1997; Berman et al., 2010; Lafer-Sousa et al., 2016). However, the resultant individual parcellations are generally poor-quality since a reliable functional localizer usually requires a relatively large amount of individual data. But the acquisition of large amounts of individual data is often prohibitive in current neuroimaging studies (Marek et al., 2018), limiting the usage of related applications. Therefore, how to improve the quality of individual parcellation based on limited data becomes a recent research direction in the field.

The inter-individual variability of human brain organization can also significantly impact the evaluation of brain functional parcellations. With the unique structural and functional differences across individuals, the resultant individual parcellations often show divergent patterns, resulting in any attempt to evaluate brain functional parcellations without considering inter-individual variability that could potentially be biased and inaccurate interpretations. For example, each individual parcellation obtained from one study will get a different evaluation score even using the same evaluation methods.

A second challenge lies in accounting for the intrinsic spatial dependence of brain locations. Recent studies have shown that brain regions possess inherent heterogeneity

to some extent (Van Den Heuvel and Pol, 2010; Van Essen and Glasser, 2018), where the neighboring brain locations exhibit higher functional similarity compared to the spatially far-away locations. This spatial smoothness, or the inherent homogeneity within brain regions, has implications for the way when we parcel brain functional regions (Yeo et al., 2011; Eickhoff et al., 2018). Previous work proposed parcellations using global clustering-based approaches, without considering the spatial connections between brain regions. This disadvantage is more pronounced in individual parcellations compared to group parcellations that have a level of smoothness across subjects. Therefore, it is crucial to have a principled way of generating individual parcellations that combine both spatial proximity and functional similarity.

The intrinsic smoothness also impacts the evaluation of brain parcellations. This can be observed by evaluating brain parcellations using homogeneity-based methods (Rousseeuw, 1987; Craddock et al., 2012; Gordon et al., 2016), where a finer-grained, or even random but spatially contiguous, parcellation can achieve a relatively higher evaluation score. This bias, caused by intrinsic spatial smoothness, makes a direct comparison between parcellations in different resolutions difficult. Therefore, the issue of brain intrinsic smoothness necessitates the exploration of alternative brain parcellation representations and evaluations.

1.4 Thesis Objectives

The primary objective of this thesis is to address some of the challenges reviewed in the previous section. The overarching goal of this research aims to enhance our understanding of brain organization and improve the accuracy and reliability of brain parcellation techniques. The specific objectives of each chapter are described as follows.

Chapter 2 proposes an unbiased criterion to evaluate discrete brain parcellations, called Distance Controlled Boundary Coefficient (DCBC). In contrast to existing evaluation methods, the DCBC takes into account the spatial arrangement of the parcels, thereby correcting biases that arise from the intrinsic smoothness of brain data. We employ DCBC to evaluate existing parcellations of the human neocortex in predicting

functional boundaries for task-based or resting-state fMRI datasets.

In **Chapter 3**, I then propose a hierarchical Bayesian framework to learn brain probabilistic parcellation. This framework addresses an important limitation for human brain parcellation studies: the missing of a large and homogeneous task-based imaging dataset. Instead, the framework allows the fusion of insights across a diverse set of task-based and resting-state datasets, resulting in the trained parcellations having a combined strength. Additionally, the framework also allows the user to derive individual brain parcellations using only 10 minutes of individual data to outperform the performance of group atlases.

Chapter 4 introduces an extension of the framework introduced in Chapter 3. The brain parcellations possess intrinsic spatial dependence, where the nearby brain locations exhibit a higher functional correlation compared to faraway locations. Therefore, a good computational model should account for the spatial structure of brain locations when learning brain parcellations. To this end, I propose a novel computational architecture to model the spatial dependencies between brain locations, called m-RBM. I then use this new model to train individual parcellations in both synthetic and empirical data.

Chapter 2

Evaluating brain parcellations using the distance-controlled boundary coefficient

2.1 Introduction

Neuroscience has a long history of subdividing the human brain into different regions based on differences in histology (Brodmann, 1909). It is commonly understood that brain function arises through the interactions of regions that are structurally and/or functionally distinct (Felleman and Van Essen, 1991; Eickhoff et al., 2018). While early parcellations of the human brain were based on the cytoarchitectonic organisation of the neocortex (Brodmann, 1909; Zilles et al., 2002; Talairach, 1988a), the advent of neuroimaging allowed an in-vivo assessment of brain organisation. In recent years, many parcellations based on task-evoked (Yeo et al., 2015) and resting functional magnetic imaging resonance (fMRI) data (Eickhoff et al., 2015; Arslan et al., 2018; Eickhoff et al., 2018) have been published, along with multi-modal parcellations that also incorporate structural and cytoarchitectonic information (Glasser and Van Essen, 2011; Fan et al., 2016).

In the empirical study of brain function, parcellations play an important practical

role. They are commonly used to define the regions of interest (ROIs) to summarize functional and anatomical data, or to define the nodes for subsequent connectivity analysis (Sporns, 2011). The main function of parcellation is to reduce complexity of the statistical analysis, as the brain-wide data can be summarized with a smaller number of values, each reflecting measurements from a region with high homogeneity. Additionally, widely-accepted parcellations aid the direct comparison between studies (Arslan et al., 2018).

Despite the importance of brain parcellations in human neuroscience research, there is no commonly accepted evaluation criterion to compare different parcellations. The obvious reason for this is that different parcellations are generated with different goals in mind. Specially, some parcellations aim to define regions that have a common anatomical characteristic (Desikan et al., 2006; Fischl et al., 2004), a shared connectivity fingerprint (Yeo et al., 2011; Gordon et al., 2016; Power et al., 2011), or a homogeneous task-activation profile (Yeo et al., 2015). As such, brain parcellations can be evaluated based on different types of data (Arslan et al., 2018).

Universally, however, any parcellation should aim to define regions such that the functional profiles (whether anatomical measures, connectivity patterns, or task activation) of two brain locations in the same region should be maximally similar to each other, whereas two brain locations in different regions should be maximally different. Thus, brain parcellation can be viewed as a clustering problem. As a result, standard machine learning methods to evaluate clustering solutions have been applied to brain parcellation. Two such examples are the measure of global Homogeneity (Gordon et al., 2016; Craddock et al., 2012) and the Silhouette coefficient (Rousseeuw, 1987).

However, these two evaluation criteria have the common problem in that they do not account for the spatial nature of the underlying data. In the case of the human neocortex, the functional correlation between two nodes on the cortical surface depends on their distance, with nearby nodes showing a higher similarity compared to far away ones. This causes even random, but spatially contiguous, parcellations to achieve relatively high global Homogeneity or Silhouette coefficient. To establish whether a parcellation identifies any real functional boundaries at all, Monte-Carlo simulations using random

parcellations are therefore necessary (Arslan et al., 2018). To complicate matters further, both global Homogeneity and Silhouette coefficient tend to be higher for finer parcellations. This makes it difficult to compare between two parcellations with different spatial resolutions.

In this paper, we address this problem by proposing a novel evaluation criterion, the Distance-Controlled Boundary Coefficient (DCBC). As the Silhouette coefficient, it compares within-parcel and between-parcel correlations of the functional profiles. However, the DCBC takes into account the spatial smoothness of the data by only comparing pairs of locations with the same distance on the cortical surface. As we will show, the expected value of the DCBC for a random parcellation is zero. Thus, no simulations with random parcellations are necessary to establish a baseline measurement; we can directly test the DCBC against zero. We also show that this baseline value is invariant to the number of parcels in the random parcellation. This enables us to use the DCBC to directly compare parcellations of different spatial scales.

We then use the DCBC to evaluate a set of common parcellations of the human neocortex (Yeo et al., 2011, 2015; Gordon et al., 2016; Power et al., 2011; Glasser et al., 2016; Schaefer et al., 2018; Fan et al., 2016; Baldassano et al., 2015; Shen et al., 2013; Arslan et al., 2015; Tzourio-Mazoyer et al., 2002; Desikan et al., 2006; Fischl et al., 2004). We performed this evaluation using both a task-based and a resting-state fMRI data set. For the task-based data set, we used the comprehensive Multi-Domain Task Battery (MDTB) (King et al., 2019), which contains functional contrasts across many cognitive domains measured in the same participants. A python toolbox for the efficient computation of the DCBC, as well as a surface-based version of the MDTB data set are publicly available to download.

2.2 Methods

2.2.1 Overview

The DCBC compares the correlation between two brain locations within a parcel to the correlation between two brain locations across a boundary between parcels. Importantly, this comparison is only performed for pairs of brain locations that are separated by the same spatial distance. The calculation of the DCBC proceeds in four steps. First, we require a data set that provides a rich characterization of each brain location. This data set defines the *functional profile* for each brain location. While the DCBC can be applied to any high-dimensional data, such as multi-modal anatomical data, we focus here on task-based fMRI data (the MDTB data set (King et al., 2019), which provides 34 activity estimates across a range of motor, cognitive and social tasks) and resting-state fMRI data (acquired in the Human Connectome Project, HCP, (Van Essen et al., 2013)). Secondly, we need a measure of spatial distance between two brain locations, either defined on the cortical surface, or for subcortical structures, in the volume. Based on these distances, all location pairs are subdivided into a set of spatial bins. The within-parcel and between-parcel correlation is then computed for each spatial bin separately. In the last step, the results are integrated across spatial bins, using an adaptive weighting scheme. To validate the method, we employed random parcellations of the human neocortex using a range of spatial resolutions, as well as sets of smooth artificial functional data sets.

2.2.2 Evaluation Data

2.2.2.1 Task-based dataset (MDTB)

To define the functional profiles for the evaluation, we first used the publicly available MDTB data set (King et al., 2019), which contains a wide range of tasks, quantifying processes required for motor, cognitive, and social function. Each of the 24 participants (16 females, 8 males, mean age=23.8) was scanned four times for 80-minutes, while performing either task set A or B (17 tasks for each, 9 tasks in common). Task set A

was performed in the first two sessions, task set B in the last two sessions. A total of approximately 5.3 h of functional data per participant was collected.

In each imaging run, every task was performed once for 35 s, starting with a 5 s instruction period, followed by a 30 s period of continuous task performance. The task battery included motor (finger tapping, sequence production), working memory (2-back task, math), language (verb generation, reading), social (theory of mind, action observation), executive control (no-go, stroop), attention (visual search), emotion (facial expression, pleasant/ unpleasant pictures), spatial (mental rotations), introspection tasks (spatial and motor imagery), movie-based tasks (cartoon, nature, landscapes), and rest (fixation) (King et al., 2019).

All fMRI data were acquired on a 3T Siemens Prisma at Western University. The imaging parameters were as follows: repetition time = 1 s; field-of-view = 20.8 cm; phase encoding direction P to A; 48 slices; 3 mm thickness; in-plane resolution $2.5 \times 2.5 \text{ mm}^2$. For anatomical localization and normalization, a 5 min high-resolution scan of the whole brain was acquired (see King et al. (2019) for more details).

Data pre-processing was carried out using tools from SPM12 (www.fil.ion.ucl.ac.uk/spm/doc/spm12_manual.pdf), as well as custom-written scripts written in MATLAB. For all participants, an anatomical image (T1-weighted MPRAGE, 1mm isotropic resolution) was acquired in the first scanning session. Functional data were realigned for head motion within each session, and for different head positions across sessions using the six-parameter rigid body transformation. The mean functional image was then co-registered to the anatomical image and this transformation was applied to all functional images. No smoothing or group normalization was applied.

The anatomical image of each of the 24 subjects was processed by standard recon-all pipeline of the freesurfer software (version 5.0) (Fischl, 2012), including brain extraction, white and pial surfaces generation, inflation, and spherical alignment to the new symmetric fsLR-32K template (Van Essen et al., 2012). Individual surfaces were then re-sampled into this standard grid. This resampling led to surfaces with 32,492 vertices that are shared both across participants and across left and right hemisphere.

A General Linear Model (GLM) was fitted to the time series data of each voxel for

each imaging run. Each task was modeled as a 30s regressor and all the preceding 5s instructions were modeled as separate regressors. The regression weights (betas) were estimated for each run independently and then averaged across the 16 runs for each task set.

To combine the activity estimates across the two task sets, we used the mean of the shared tasks as a common reference point. We subtracted this pattern from the average beta estimates for each task set separately, and then concatenated the two vectors of activity estimates. The average beta weights were then divided by the square root of the average mean-square-residual from the first-level GLM to obtain z-scores for each voxel. The resulting functional profiles consisted of 34 pre-whitened activity estimates (set A = 17; set B = 17) for each voxel. Finally, we subtracted the overall mean across all tasks from the functional profile of each voxel.

The functional profiles were then mapped to each individual cortical surface by averaging the value from voxels along the connecting line between the pial and white-gray matter surface, using 5 equally spaced locations between the two surfaces.

2.2.2.2 Resting-state dataset (HCP)

The second data set used in this study was the resting-state fMRI (rs-fMRI) data from the “unrelated 100” subjects (54 female, 46 male adults, aged from 22 to 35), which was made publicly available in the Human Connectome Project (HCP) S1200 release ([Van Essen et al., 2013](#)). The rs-fMRI scans for each subject were collected in two sessions held on different days, including a total four runs of approximately 15 minutes each. During the scans, the subjects were asked to fixate a white cross-hair on a dark background.

The HCP resting-state fMRI time series were acquired using 3T Siemens “Connectome Skyra” scanner with $2 \times 2 \times 2$ mm spatial resolution and a TR of approximately 0.7 s. For more details of the data acquisition parameters, see [Smith et al. \(2013\)](#), and [Uğurbil et al. \(2013\)](#).

All data were pre-processed using the HCP minimal processing pipeline ([Glasser et al., 2013](#)), including structural registration, correction for spatial distortion, head motion, cortical surface mapping, and functional artefact removal ([Smith et al., 2013](#); [Glasser](#)

et al., 2013). For each rs-fMRI run, this resulted in 1200 time points for each of the 32k vertices of the standard fsLR-32K template (Van Essen et al., 2012) per hemisphere. To generate the functional profiles for the HCP data set, we concatenated all 4 runs after mean-centering.

2.2.3 Existing evaluation criteria for brain parcellations

Given that brain parcellation can be viewed as a clustering problem, two common methods used to evaluate the resultant parcels are the global Homogeneity (Craddock et al., 2012; Gordon et al., 2016), and the Silhouette coefficient (Rousseeuw, 1987). Homogeneity is defined as the average similarity across all pairs of vertices within a parcel. As the similarity measure of two vertices, we used the Pearson’s correlation between functional profiles. The global Homogeneity is then simply the average with-parcel correlation across all parcels, with higher homogeneity suggesting a better parcellation.

Another popular evaluation metric for brain parcellations is the Silhouette coefficient (Rousseeuw, 1987), which compares the average dissimilarity (defined as $1-R$, where R represents Pearson’s correlation between functional profiles) from one vertex to all other vertices in the same parcel (w_i), to the average dissimilarity from the same vertex to all the vertices that assigned to neighbouring parcels (b_i) (Yeo et al., 2011; Arslan et al., 2018). For a given a parcellation $\{P_1, P_2, \dots, P_k\}$, w_i and b_i can be defined as:

$$w_i = \frac{1}{m_k - 1} \sum_{j \in P_k, i \neq j} 1 - R(v_i, v_j), \quad (2.1)$$

$$b_i = \frac{1}{N} \sum_{j \in nb(P_k)} 1 - R(v_i, v_j) \quad (2.2)$$

where m_k indicates the number of vertices that within the parcel P_k . N is the total number of vertices in all neighbouring parcels and $nb(P_k)$ represents all neighbouring parcels of P_k .

For each cortical vertex v_i , the Silhouette coefficient is defined as:

$$S_i = \frac{b_i - w_i}{\max(w_i, b_i)} \quad (2.3)$$

Based on this definition, the Silhouette coefficient for each vertex ranges from -1 to 1, 1 indicates that there is a perfect correlation within each parcel ($r = 1$) and on average, not correlations across parcels ($r = 0$). As we will see, both of these measures are biased by the intrinsic smoothness of the functional profiles on the cortical surface.

2.2.4 Measuring spatial distance

To account for the intrinsic smoothness of the data, we require a measure of spatial distance between any pair of brain locations. For subcortical structures, we have used the Euclidean distance between pairs of voxels (King et al., 2019). For the neocortex, however, we ideally would like to use the geodesic distance between vertices on the cortical surface. As an approximation to this distance, we used Dijkstra’s algorithm (Dijkstra et al., 1959) to estimate the shortest paths between each pair of vertices on each individual cortical surface. For this computation we used the mid-cortical layer which is the average of the pial and white-gray matter surface. For computational and memory efficiency we only considered distances up to maximum of 50mm. Inter-vertex distances were then averaged across individuals and hemispheres. This resulted in a matrix that indicates the average cortical distance between nearby brain locations for the atlas brain surface.

2.2.5 Distance Controlled Boundary Coefficient (DCBC)

2.2.5.1 The problem of spatial smoothness

The problem with global Homogeneity and Silhouette coefficient is that they do not take into account that function tends to vary in a smooth fashion across the cortical surface. For instance, if we compute the correlation of vertex pairs across the cortex using task-evoked functional profiles (King et al., 2019) for an individual participant, we can clearly see that the correlation falls off with the spatial distance between vertices (See Figure 2.1a). Note that this smoothness is not an artifact of the data processing; except for motion realignment and mapping onto the surface, no smoothing was applied to the data. Thus, the dependence on spatial distance reflects the intrinsic smoothness of functional specialization on the cortex.

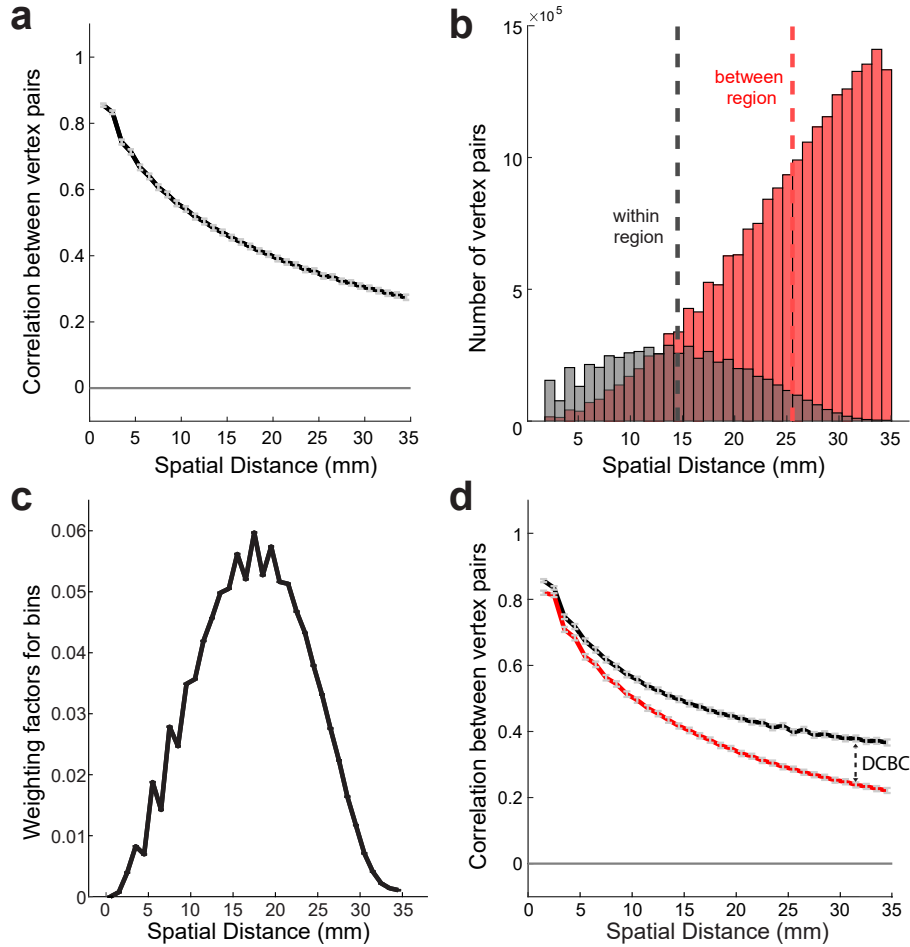


Figure 2.1: **Distance controlled boundary coefficient.** (a) Correlation between task-evoked functional profiles (see methods) of pairs of surface vertices as a function of their spatial distance; (b) Histogram of the number of within and between vertex pairs as a function of spatial distance for a random Icosahedron 162 parcellation. (c) Weighting factor across different bins for the Icosahedron 162 parcellation and binning shown in b. (d) The correlations for within- (black) and between-region (red) vertex pairs as a function of the spatial distance (for Yeo 17 parcellation). The DCBC is defined by the weighted average distance between the two curves.

For the global Homogeneity measure, this property favors parcellations with small parcels, as only close-by vertex pairs will be within the same parcel. Similarly, the spatial smoothness also biases the Silhouette coefficient, as the spatial distance for within-parcel pairs is on average smaller than that for between-parcel pairs. For example in random parcellation Icosahedron 162 (Fig. 2.1b), the average spatial distance of within-parcel pairs is 14.5 mm. Even if we limit the comparison to vertex pairs from spatially adjacent parcels, as is common practice in the evaluation of brain parcellations, the between-parcel

pairs have a substantially larger average distance (25.5mm). This discrepancy results in a higher average correlation of functional profiles for within-parcel pairs compared to between-parcel pairs.

We therefore propose to only compare vertex pairs with a similar spatial distance. For this purpose, the DCBC method bins all vertex pairs according to their spatial distance, and then compares the correlation for within- and between-pairs within each bin. One important practical decision is the choice of bin size, a question that we address in the results section. For our neocortical data, a bin size of 1 mm appears to be adequate.

2.2.5.2 Averaging across bins

Parcellations can be compared by investigating the difference in within- and between-parcels as a function of the spatial distance (see [King et al. \(2019\)](#), Fig. 3,4). However, for many applications we would like a single evaluation criterion for each parcellation, which necessitates the averaging across a range of spatial distances. This raises the question of what range of spatial distances to consider, and how to weight the distances within that range. A rational solution to this problem is to find the weighting that, for any given parcellation, provides us with the best estimate of the average difference between within- and between-parcel correlations, assuming that this difference is constant across the desired range of distances. The variance of the estimate of the correlation difference (d_i) for bin i can be approximated by assuming the independence of the different vertex pairs. In this case, the variance of the estimate depends on the number of within- ($n_{w,i}$) and between-parcel vertex pairs ($n_{b,i}$) in each spatial bin:

$$\text{var}(d_i) = \frac{1}{n_{w,i}} + \frac{1}{n_{b,i}} = \frac{n_{w,i} + n_{b,i}}{n_{w,i}n_{b,i}} \quad (2.4)$$

For averaging, we define a weight that is proportional to the precision (inverse of the variance) of each estimator:

$$w_i = \frac{n_{w,i}n_{b,i}}{n_{w,i} + n_{b,i}} / \sum_j \frac{n_{w,j}n_{b,j}}{n_{w,j} + n_{b,j}} \quad (2.5)$$

For example, figure 2.1c shows the weighting factor for each spatial bin of Icosahedron 162 random parcellation using a 1 mm bin width. The DCBC is then the weighted average of the correlation difference across bins.

2.2.6 Random Parcellations

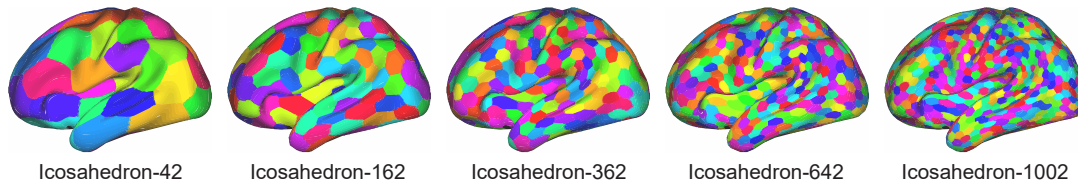


Figure 2.2: **Random cortical parcellations with different number of parcels.**

To evaluate the DCBC for parcellations that on average do not align with real functional boundaries, we generated a set of random parcellations. If our method successfully controls for the spatial smoothness of the functional profiles, the average DCBC for such random parcellations should be zero, i.e there should be no difference between within- and between-parcel correlations. To test this claim for parcellations at different spatial scales, we used a regular hexagonal parcellations of a sphere (Icosahedron) with 42, 162, 362, 642, and 1002 parcels. To generate random alignment of this parcellation with the data, we rotated each map randomly around the x, y, and z axis. We repeated this process 100 times to obtain 100 random parcellations for each spatial scale.

2.2.7 Random Functional Maps

Real data may have functional boundaries that are correlated across participants. As a consequence, some random parcellations will still by chance align with these boundaries and yield systematically positive DCBC values; and other random parcellations will misalign with the real functional boundaries and yield systematically negative values. To test the DCBC on functional maps without a systematic boundary structure across participants, we also randomly generated 100 cortical functional maps with 34 task conditions. These maps then were used in the analysis shown in Fig. 2.3. We drew independent normally distributed values for every condition and vertex for the fsLR-32k

template (Van Essen et al., 2012), and then smoothed these maps on the cortical surface using `-metric-smooth` function provided by Connectome Workbench software (Marcus et al., 2011). The smoothing kernel was set to 12 centimeters, yielding a similar spatial autocorrelation function as in our real data.

2.2.8 Evaluation of commonly-used group parcellations

We then evaluated a set of commonly used group parcellations of the human neocortex (Table 2.1). The majority of the parcellations considered here are based on resting-state functional connectivity (rs-FC) data. The fMRI data is recorded at rest, and the covariance or correlations between the time series of different brain locations is submitted to a clustering or dimensionality reduction approach. Parcellations can also be based on task-evoked activation maps. For example Yeo et al. (2015) derived a parcellation from 10,449 experiment contrasts across 83 behavioural tasks. The anatomical parcellations considered here are based on the macro-anatomical folding structure of the human neocortex, following the major sulci and gyri of the human brain. Finally, we also considered 2 multi-modal parcellations, which combined rs-FC and anatomical criteria.

Within each of these categories, parcellations also differ in the approach used for generation. For instance, several parcellations were obtained based on a two-level approach, where subject-level parcellations are derived in a first step, and then integrated across subjects in a second step using clustering or majority voting. Other parcellations are directly derived by clustering group-averaged data (Arslan et al., 2018).

Because the DCBC evaluation considers only vertex pairs up to a specific spatial distance on the cortical surface, the evaluation is conducted separately for the left and right hemisphere. For many parcellations, the parcels are separated for the two hemispheres. For example, Gordon et al. (2016) used 161 and 172 distinct regions for the left and right hemisphere respectively, totaling 333 regions. Other parcellations use bi-hemispheric parcels. As a consequence the 7 and 17 regions in Yeo et al. (2011), were effectively evaluated as 14 and 34 parcels.

Note that three group-level parcellations (Fan et al., 2016; Shen et al., 2013; Tzourio-Mazoyer et al., 2002) were only available in volume space. These parcellations were

Name	Type	Resolution	Reference	Link	Approach
<i>Yeo</i>	resting state	7 and 17	Yeo et al. (2011)	www.freesurfer.net/fswiki/CerebellumParcellation_Buckner2011	cortical networks spanning both hemispheres; group-averaged rs-FC, spectral clustering
<i>Power</i>	resting state	130	Power et al. (2011)	balsa.wustl.edu/study/show/WG33	65 regions per hemisphere, based on group-average rs-FC; mapping to surface space by Glasser et al. (2013).
<i>Schaefer</i>	resting state	100 to 1000	Schaefer et al. (2018)	github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal	group-averaged rs-FC; gradient-weighted Markov Random Field method; aligned to Yeo 7 and 17 networks
<i>Gordon</i>	resting state	333	Gordon et al. (2016)	sites.wustl.edu/petersenschlaggarlab/resources/	group-averaged rs-FC; agglomerative-based clustering method
<i>Baldassano</i>	resting state	171	Baldassano et al. (2015)	biomedica.doc.ic.ac.uk/brain-parcellation-survey/	individual and group rs-FC data; spectral clustering; volume-surface mapping by Glasser et al. (2013)
<i>Shen</i>	resting state	200	Shen et al. (2013)	www.nitrc.org/frs/?group_id=51	based on HCP S1200 group PCA output (Smith et al., 2014); group-averaged parcellations obtained through group-ICA (Beckmann and Smith, 2004)
<i>ICA</i>	resting state	15 to 300	Beckmann and Smith (2004)	https://db.humanconnectome.org/data/projects/HCP_1200	individual parcellations derived via spectral clustering; integrated into a group parcellation using second-level clustering
<i>Arslan</i>	resting state	50 to 497	Arslan et al. (2015)	biomedica.doc.ic.ac.uk/brain-parcellation-survey/	hierarchical Bayesian model applied to 10,449 task contrasts
<i>Yeo 2015</i>	task-evoked	12	Yeo et al. (2015)	surfer.nmr.mgh.harvard.edu/fswiki/BrainmapOntology_Yeo2015	commonly used brain atlas based on cortical folding; volume-based atlas mapped to surface
<i>AAL</i>	anatomical	82	Tzourio-Mazoyer et al. (2002)	www.gin.cnr.s.fr/en/tools/aal/	individual parcellations based on cortical folding; group parcellations via majority-voting
<i>Desikan</i>	anatomical	70	Desikan et al. (2006)	surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation	individual parcellations based on cortical folding; group parcellation via majority-voting
<i>Destrieux</i>	anatomical	150	Fischl et al. (2004)	fswiki/CorticalParcellation	multi-modal group parcellation based on rs-FC and cytoarchitectonic information (Brodmann areas and myelin content) from 210 HCP subjects
<i>Glasser</i>	multi-modal	360	Glasser et al. (2016)	balsa.wustl.edu/study/show/RWVG	multi-modal parcellation; rs-FC and anatomical information from 40 HCP subjects; mapping to surface by Glasser et al. (2013)
<i>Fan</i>	multi-modal	210	Fan et al. (2016)	atlas.brainnetome.org/bnatlas.html	

Table 2.1: Commonly used group-level cortical parcellations. rs-FC: resting state functional connectivity (fMRI).

mapped to HCP standard fsLR-32k cortical surface using the volume-to-surface pipeline described in [Van Essen et al. \(2012\)](#) and [Arslan et al. \(2018\)](#). All parcellations in this study are available as a collection in fsLR-32k surface space at ([Zhi and Diedrichsen, 2021](#)).

2.2.9 Parcellation based on the evaluation data

To estimate how well a group parcellation could theoretically subdivide the neocortex into functionally distinct regions, we derived a parcellation from the MDTB dataset. We estimated 12 cortical parcellations with 14 to 1000 parcels, using group-averaged MDTB functional profiles. For evaluation on the MDTB data, this parcellation has the unfair advantage that the individuals used in the evaluation is also contained within the training set, providing an upper-bound estimate of the noise ceiling ([Nili et al., 2014](#)). To estimate a lower bound of the noise ceiling, we used a leave-one-out cross validation approach: We derived a group parcellation from the averaged data from 23 participants, and then evaluated it on the remaining subject. We then averaged the DCBC across the 24 different parcellations.

To derive the MDTB group parcellation we used spectral clustering. We first down-sampled the surface data from 32K vertices to 4002 vertices. Then we performed spectral clustering on the affinity matrix between the vertices of the down-sampled map. After clustering, we then restored the map to the original resolution of the surface. The lower resolution ensured that the resulting parcellations were spatially contiguous. We consider the MDTB group parcellation as a potential lower bound (see Results 3.4 and Discussion) of how well a group parcellation can perform on the MDTB data set.

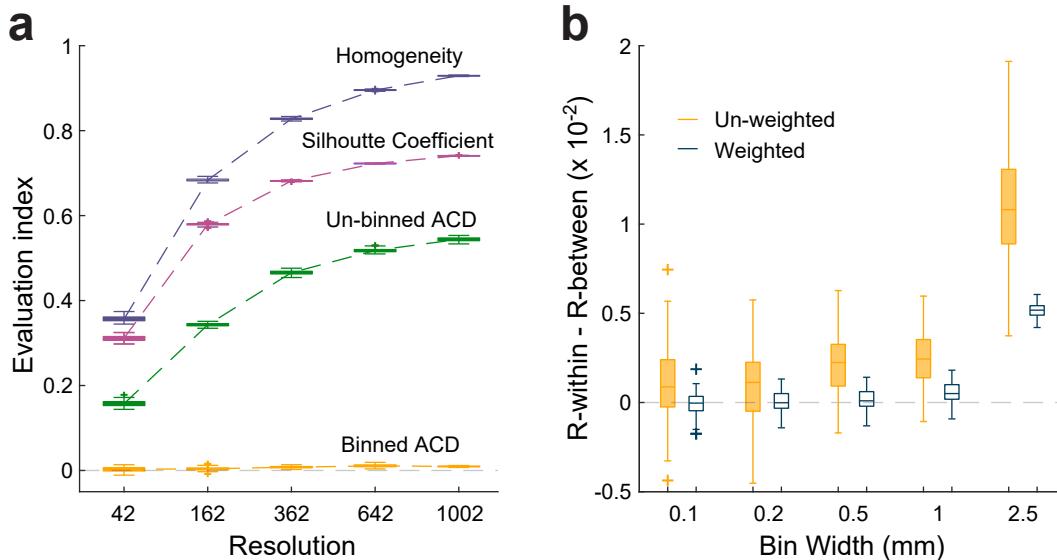


Figure 2.3: **Evaluation results on random parcellations.** (a) Results of four evaluation methods for the random functional maps simulation in different resolutions: Homogeneity, Silhouette Coefficient, the un-binned average correlation difference (ACD) of within-parcel and between-parcel, and the binned (bin width = 2.5 mm) ACD of within-parcel and between-parcel; (b) The weighted and un-weighted DCBC evaluation results for 100 random functional maps in different bin widths of random parcellations Icosahedron 642.

2.3 Results

2.3.1 Binning reduces the bias introduced by spatial smoothness

Existing evaluation methods for brain parcellations have the problem of being biased by the natural smoothness of functional brain maps. To demonstrate this effect, we first tested various evaluation methods using random functional maps and random parcellations of different spatial scales. As can be seen in Fig. 2.3a, both the Homogeneity and Silhouette coefficient show highly significant positive values, even for these random maps. Furthermore, the values for both methods increase when the parcellation increases in spatial resolution (i.e. have more and smaller parcels). This makes direct comparisons of different parcellations in different spatial scales difficult, and necessitates the use of randomisation analyses for each parcellation to determine the baseline value expected by

random chance (Arslan et al., 2018).

A similar problem can also be seen when using the difference between the correlations of within- and between-parcel pairs of vertices (un-binned average correlation difference, ACD). This is caused by the tendency that vertices that are closer together show higher functional correlations (Fig. 2.1a), combined with the fact that within-parcel vertex pairs are on average closer to each other than between-parcel pairs (Fig. 2.1b). To remove this bias, the DCBC calculation involves the binning of vertex pairs according to their spatial distance. We then calculate the difference between the average correlation between within- and between-parcel pairs within each spatial bin, thereby only comparing vertex pairs of matched spatial distance.

To ascertain that this approach provides an approximately unbiased and stable evaluation criterion, we first applied the suggested technique on the simulated functional data. As can be seen (Fig. 2.3a, binned ACD), even using a relatively coarse spatial binning of 2.5 mm, this approach nearly removes all bias caused by the spatial smoothness. For the finest parcellation, an Icosahedron with 1002 regions, the binned difference between correlations (0.009) was approximately 60 times lower than the mean of the difference calculated in each bin (0.544). This shows that the binning approach dramatically reduces the bias caused by spatial smoothness.

2.3.2 Adaptive weighted averaging reduces variance and bias

After binning, we often want to integrate the results across bins to arrive at a single evaluation criterion. This can be done by simply averaging the differences in correlation across bins. However, this approach leads to a summary measure with high variability (Fig. 2.3b). This is caused by the fact all bins have equal influence on the average, even though some bins contain very few vertex pairs. This can be addressed by taking the number of within- and between-parcel pairs in each bin into account in an adaptive weighting scheme (see methods). Indeed, the standard deviation of the weighted DCBC in the simulation is 2.8 times lower than for the un-weighted version for 1 mm bins, and 8.1 times lower for the 2.5 mm bins. Furthermore, the weighted DCBC also shows smaller residual bias than the unweighted DCBC.

2.3.3 Choosing an appropriate bin width

An important practical issue in the DCBC calculation is to choose a specific bin width. This choice is subject to a fundamental trade-off. If the bin width is too wide, the DCBC may still be biased as a result of the spatial smoothness of the functional profiles. This is because within each bin, the average spatial distance for within-parcel pairs is still slightly smaller than for the between-parcel pairs, inducing a systematic difference between the correlations within each spatial bin. Even though this bias is small, it can remain highly significant when tested across the 100 simulations presented in Fig. 2.3b for a bin width of 2.5mm. Choosing a smaller bin width reduces this bias. For bins of size 0.1 mm and 0.2 mm, the same 100 simulated data sets no longer show a significant difference against zero ($p = 0.327$ and 0.202 , respectively).

Choosing a very small bin width, however, also comes with some disadvantages. First, the computational cost of the DCBC calculation increases linearly with the increasing number of bins. More importantly, if a very small bin is chosen, it can result in noisier DCBC estimate, as very few vertex pairs will fall within each bin. In the extreme case, there would either be no within- or between- vertex pair in a bin, such that the bin could not be used for the difference calculation. For neocortical data projected to the fsLR-32k template (Van Essen et al., 2012) a bin width of 1mm appears to offer a good balance between bias, accuracy and computational speed.

2.3.4 DCBC evaluation for real data

Using a task-based data set (MDTB) and a resting-state data set (HCP), we evaluated 15 commonly used group-level cortical parcellations (Fig. 2.4a, Table 2.1). These parcellations relied either on anatomical criteria (cortical folding), task-evoked activation, or functional resting-state connectivity. Two multi-modal parcellations (Glasser et al., 2016) relied on a combination of anatomical and functional features. Each of the parcellations was evaluated per hemisphere and the global DCBC of a subject was then averaged across hemispheres.

For the MDTB data set, the difference between the within-parcel and between-parcel

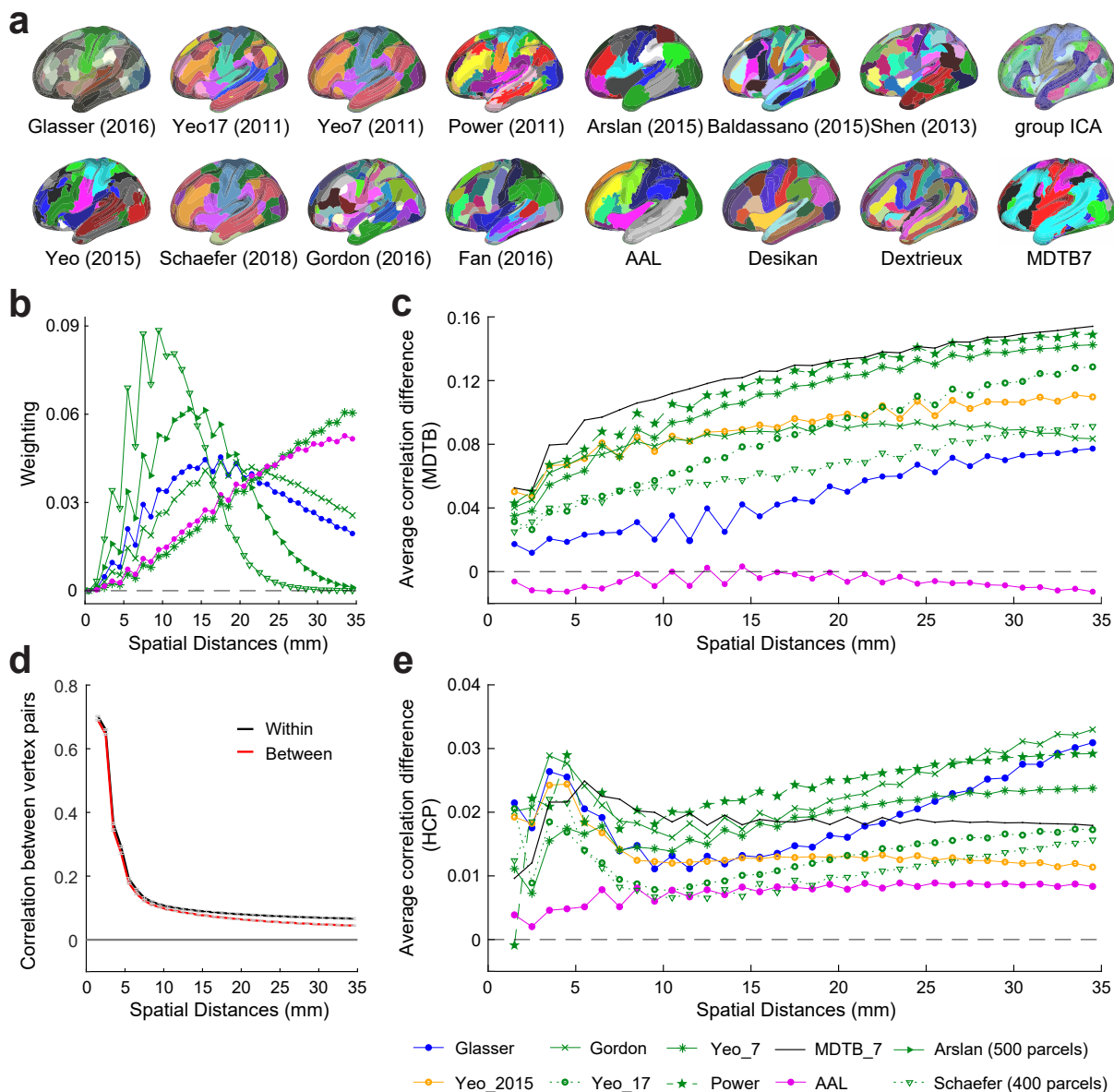


Figure 2.4: **Evaluation on the real data sets.** (a) The left hemisphere of 15 commonly used cortical parcellations and the MDTB cortical parcellation with 7 regions; (b) weighting factors of selected parcellations with different resolutions within range of 0 to 35mm; (c) the correlation difference as a function of the spatial distance for selected parcellations evaluated on the task-based data set; (d) the correlation for within- and between-parcel vertex pairs as a function of the spatial distance (for Yeo 17 parcellation) evaluated on the resting-state dataset; (e) the average correlation difference as a function of the spatial distances for selected parcellations evaluated on the resting-state data set.

correlations across range of spatial distances (0-35mm) is shown in Fig. 2.4c. While the difference increased with increasing distance, the relative ordering of the parcellations was relatively consistent: Independent of the exact spatial distance considered, the Power and

the MDTB correlation appear to outperform the other parcellations.

For the HCP data set (Fig. 2.4e), the difference between the within- and between-parcel correlation was substantially smaller. This is mostly caused by the fact that the raw Pearson’s correlation of the time series (Fig. 2.4d) were lower than the correlations for the MDTB data set (Fig. 2.1a,d). The correlations for the rs-fMRI data also fell off more rapidly with increasing distance, reaching values of $r < 0.1$ for distances over 8mm (Fig. 2.4d). The lower DCBC values for this data set, therefore, are partly due the fact that correlations between full fMRI time series are usually lower than correlations between task-based activity estimates. For the HCP data set the difference in correlations were relatively stable across the range of spatial distances considered ($< 35mm$).

To obtain a minimum-variance estimate of the correlation difference when averaging across spatial distances, we weighted the difference curves with the parcellation-specific weighting function (Fig. 2.4b). This procedure is certainly justified if the differences between correlation curves is stable across spatial distances. For parcellation where the differences between within- and between-parcel correlations vary with distance, small biases may arise. For example we may expect for the MDTB data set could give a small advantage to coarser parcellations. We will return to this issue in the discussion.

2.3.5 Resting-state group parcellations predict task-based functional boundaries

We then calculated the averaged weighted DCBC across all parcellation (Fig. 2.5). The first insight is that the nine parcellations that are based solely on functional resting state data (Yeo et al., 2011; Power et al., 2011; Gordon et al., 2016; Arslan et al., 2015; Baldassano et al., 2015; Shen et al., 2013; Schaefer et al., 2018; Smith et al., 2014) predicted the functional boundaries in the task-based data set substantially better than chance (Fig. 2.5a). For example, the within- and between-parcel correlations for the Yeo 17 parcellation (Fig. 2.1d) differed by approximately 0.1 across spatial bins, reflected in an average DCBC value of 0.1020 (SE=0.0053) across the 24 subjects. Other resting-state parcellations also showed clear differences between the within- and between-parcel

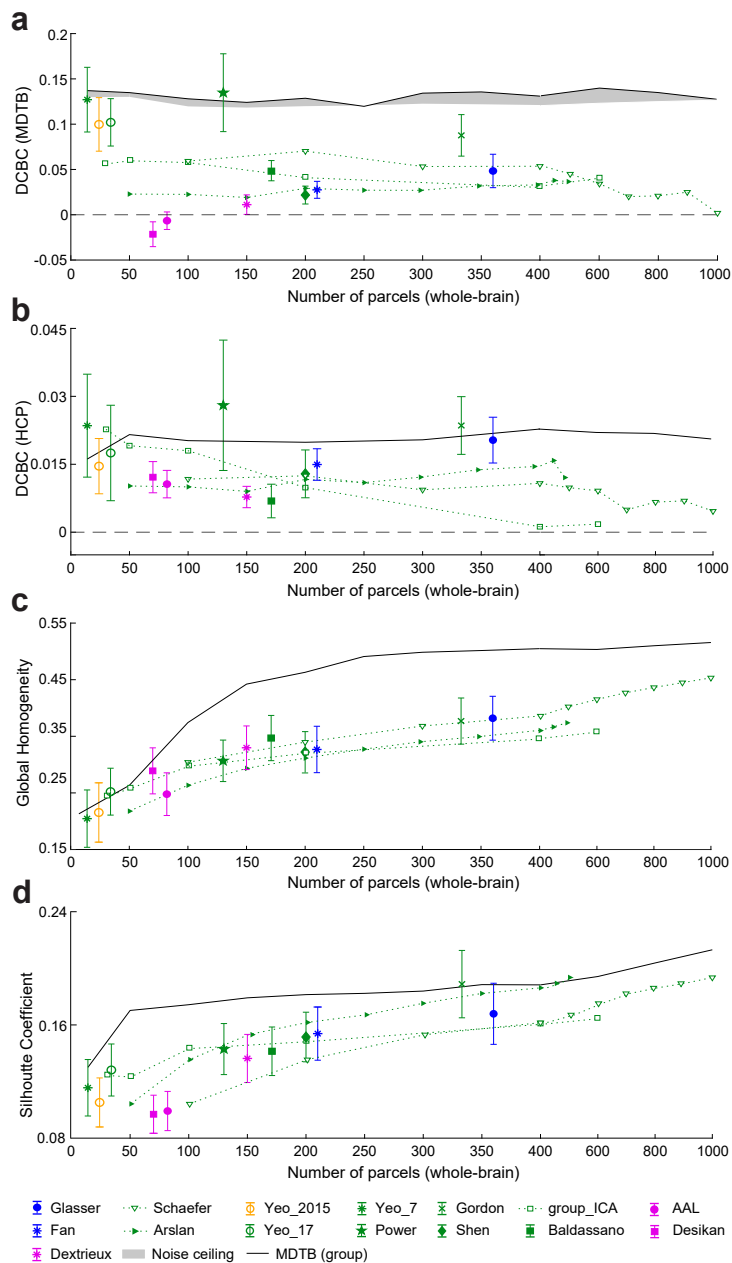


Figure 2.5: **Evaluation of existing group parcellations.** (a) The DCBC evaluation of the existing parcellations on the task-based dataset. The evaluation of the MDTB parcellations is indicated by the gray area (noise ceiling). The upper and lower bounds are the non-crossvalidated and cross-validated evaluations respectively; (b) the DCBC evaluation of the existing parcellations on HCP resting-state dataset; (c) the global homogeneity measures for all parcellations using the MDTB task-based dataset. Error bars indicate the standard deviation of DCBC across the 24 evaluation subjects. (d) The Silhouette Coefficient of the parcellations using the MDTB task-based dataset.

correlations, especially Power 2011 (DCBC=0.1334, SE=0.0085), Yeo 7 (DCBC =0.1271, SE=0.0073), and Gordon 2016 (DCBC=0.0876, SE=0.0047). This finding confirms that resting-state group parcellations generally predict task-relevant functional boundaries significantly better than chance (Tavor et al., 2016).

When evaluating these resting-state parcellations on resting state data (see Fig. 2.5b), we obtain consistent results. Even though the overall DCBC was substantially lower than for the task-based data, the best-performing parcellations were based on resting-state data, including the Yeo 7 (DCBC=0.0213, SE=0.0021), Power (DCBC=0.0261, SE=0.0025), and Gordon (DCBC=0.0236, SE=0.0018) parcellations.

2.3.6 Comparison to parcellations derived from the evaluation data set

How well do these group-based resting-state parcellations predict task-based functional boundaries, relative to what would be possible? Given the inter-individual variability of boundaries, and the fact that even individual boundaries are not perfectly sharp, there is an upper limit to the highest achievable DCBC on our evaluation data set. To obtain an idea of this “noise ceiling” (Nili et al., 2014), we derived a set of clustering solutions from the MDTB data itself (see methods, Fig. 2.4a), spanning the range from 14 to 1000 parcels.

While we cannot determine the noise ceiling directly, we can obtain a lower and upper performance estimate. For the lower estimate, we derived the parcellation on 23 of the participants, and evaluated it on the remaining, left-out participant. For the upper estimate, we over-fitted the data by deriving and evaluating the parcellation on all 24 participants. The gap between these two performance curves indicates how much of the performance advantage of the MDTB parcellation is due to the over-fitting to the particular set of subjects.

As expected, the MDTB-based parcellations (gray area in Fig. 2.5a) generally outperformed other group parcellations on this data set. Nonetheless, some existing resting-state parcellations showed performance very close or even slightly higher than the MDTB

parcellation (Yeo et al., 2011; Power et al., 2011).

When evaluated the task-based MDTB parcellations on the resting-state data (Fig. 2.5b), it performed remarkably well, and was only outperformed by 3 resting-state parcellations. This again demonstrates the consistency of functional boundaries across task- and resting-state data.

2.3.7 Anatomical parcellations do not predict task-based functional boundaries

We then evaluated 3 commonly used anatomical group parcellations of the human neocortex: The Desikan parcellation (Desikan et al., 2006), the Dextrieux atlas (Fischl et al., 2004), and the Automated Anatomical Labelling (AAL) atlas (Tzourio-Mazoyer et al., 2002). On the task-based data (Fig. 2.5a), the averaged correlation between any vertex pairs within a predefined anatomical parcel was not much greater than the correlation between vertex pairs across anatomical boundaries, resulting in very low DCBC values (Dextrieux: DCBC=0.0112, SE=0.0022; AAL: DCBC=-0.0066, SE=0.002). The Desikan parcellation (DCBC=-0.0215, SE=0.0028) even showed significantly negative DCBC values across the 24 subjects. These parcellations, based on the macroanatomical folding structure of the neocortex, therefore define boundaries that are systematically misaligned with the functional subdivisions in task-evoked activity profiles.

A similar pattern emerged when anatomical parcellation were evaluated on resting-state data. All anatomical parcellations showed relatively low performance (average DCBC = 0.0066). In contrast to the task-based evaluation, the DCBC of all three parcellations was significantly positive, when tested against zero (all $t_{23} > 9.4466, p < 2.21 \times 10^{-9}$), implying that they aligned with the boundaries of the resting-state networks slightly better than chance.

2.3.8 Multi-modal parcellations do not perform better than resting-state parcellations

We also applied DCBC evaluation to two multi-modal parcellations (Glasser et al., 2016; Fan et al., 2016) to determine whether combining anatomical and functional data is superior to unimodal parcellations. The Glasser parcellation had a higher DCBC score (DCBC=0.0483, SE=0.0038) as compared to the Fan parcellation (DCBC=0.0275, SE=0.0019). However, both were lower than the average DCBC across the resting-state parcellations (0.0766). It therefore appears that the combination of multiple modalities does not systematically lead to a better prediction of task-relevant function boundaries than parcellations that are derived on resting-state data alone.

2.3.9 Comparison across different spatial resolutions

For simulated random functional maps, we have shown that the expected value of the DCBC is zero, no matter how fine the parcellation (Fig. 2.3b). In contrast, the value of the global Homogeneity and Silhouette coefficient increases for finer parcellations even for random maps (Fig. 2.3a).

This bias can also be observed for real parcellations. The value of the global Homogeneity (Fig. 2.5c) and Silhouette coefficient (Fig. 2.5d) when calculated on the task-based evaluation data set clearly increases for finer parcellations, whereas there is no strong relationship between the DCBC and the number of parcels (Fig. 2.5a and b).

In this context, the set of Schaefer 2018 parcellations (Schaefer et al., 2018) is especially interesting, as it provides a nested set of subdivisions with an increasing number of parcels, all aligned with Yeo 7 or 17 networks (we use the one aligned with Yeo 7 networks in the experiment). To statistically evaluate the change in evaluation metric with parcel size, we conducted a repeated-measures analysis of variance (ANOVA) across the 10 Schaefer parcellations, ranging from 100 to 1000 parcels. As expected, both the Homogeneity ($F_{9,207} = 1730.6, p = 1.55 \times 10^{-189}$) and the Silhouette coefficient ($F_{9,207} = 667.6, p = 1.11 \times 10^{-147}$) clearly showed significant increases for an increasing number of parcels. Given that such increases were also found for random functional

maps and parcellations, it is not clear whether the finer parcellations identified functional boundaries better, the same, or worse than coarser parcellations.

In contrast, the unbiased DCBC shows that the Schaefer parcellation reaches the best performance around 200 parcels, and then slowly declines for finer parcellations (Fig. 2.5a). The ANOVA showed a significant change with number of parcels ($F_{9,207} = 189.4576, p = 8.19 \times 10^{-95}$). Indeed, for the finest parcellation (1000 parcels), performance did not differ significantly from chance ($t_{23} = 1.0253, p = 0.3159$). One possible reason for this is that when defining more than 200 functional parcels, the new boundaries do not consistently predict discontinuities in the functional organisation at the group level anymore.

In summary, the application of the novel DCBC criterion to known cortical parcellations allowed for the following conclusions: (1) anatomical parcellations based on cortical folding do not align with functional boundaries in the neocortex; (2) resting-state parcellations predict task-relevant functional boundaries, outperforming other types of cortical parcellations; (3) multi-modal parcellations did not improve performance as compared to pure resting-state parcellations.

2.3.10 Open-source toolbox/data support evaluation

The code for DCBC evaluation is published as an open-source toolbox written in Python at (Zhi and Diedrichsen, 2021). The package also contains the pre-processed contrast maps for all task conditions of the MDTB data set (n=24 subjects), sampled to the standard fsLR-32k template.

2.4 Discussion

In this study, we introduce a novel evaluation criterion for brain parcellations, the Distance Controlled Boundary Coefficient (DCBC). The method takes into account the spatial smoothness of the data by controlling the distance of the vertex pairs when comparing within- and between-parcel correlations. We used an earlier form of this approach for volume-based data (using the Euclidean distance instead of a surface-based distance)

to evaluate cerebellar parcellations (King et al., 2019). Here, we further improve the approach by adaptively weighting over the spatial distance bins, resulting in a global measure with low bias and variance. Our evaluation on simulated smooth data shows that the new criterion overcomes the size- and shape-dependent bias of other homogeneity-type evaluation criteria (Craddock et al., 2012; Rousseeuw, 1987). The advantage of the DCBC is twofold: 1) it enables a direct comparison between group brain parcellations that have different spatial resolutions, and 2) it provides a direct test of how well a parcellation subdivides the brain into homogeneous regions than expected by chance.

One important caveat is that DCBC only removes the bias completely if the difference between within- and between correlations is stable across spatial distances. This is because different parcellations use different weighting across spatial distances (Fig. 2.4b). There are a number of practical solutions to ameliorate this problem. First, by choosing a maximal distance for vertex distances (here 35mm) the evaluation is constrained to be relatively local in all cases. While future users of the method may want to choose a different range of spatial distances, we believe that 35mm provides a good compromise for cortical parcellations. Secondly, it is always useful to plot the DCBC as a function of the spatial distance before averaging (see King et al. (2019), Fig. 4c,e) to investigate whether different parcellations may behave differently across the distances considered. Finally, if the DCBC varies substantially across spatial distances, one could use a common, averaged weighting for all parcellations, or simply decide on a more specific set of spatial distances. Nonetheless, the biases from differential weighting were relatively small for our evaluation data sets - and the DCBC successfully removed the main biasing influence of parcel size (Fig. 2.5a vs. c,d).

We used the DCBC to evaluate a range of existing cortical surface-based parcellations in their ability to predict functional boundaries on task-based and resting-state data. We found that the parcellations derived from resting-state fMRI data largely succeed in predicting task-evoked activity boundaries, replicating earlier work (Tavor et al., 2016; King et al., 2019). These results demonstrate again the practical utility of resting state data in identifying brain networks that work together during active task performance. Even though the correlation structure across the cortex does clearly change in a task-

dependent fashion (Hasson et al., 2009a; Salehi et al., 2020a,b), our results emphasize the existence of a basic scaffold that determines functional specialization across a wide range of tasks, as well as during rest. In the opposite direction, parcellations derived from a rich task-based battery (MDTB) also achieved relatively high DCBC values when evaluated on rs-fMRI data, further confirming that structure of neural fluctuations during rest aligns with co-activation across tasks.

In contrast, anatomical parcellations (Desikan et al., 2006; Fischl et al., 2004; Tzourio-Mazoyer et al., 2002) did not perform better than chance to predict functional boundaries in the task-based data, and only slightly better than random for rs-fMRI data. The Desikan parcellation even showed a negative DCBC score on the MDTB dataset. This finding corroborates previous work that shows a misalignment between macroanatomical folding structure and functional boundaries in the neocortex (Arslan et al., 2018) and the cerebellum (King et al., 2019). An inspection of the differences between functional and anatomical parcellations (Fig. 2.4a) suggest an explanation of why this may be the case. Cortical motor areas, for example, are subdivided in all anatomical parcellations along the central sulcus, which separates the primary motor cortex (M1) from primary somatosensory cortex (S1). In this case, the macro-anatomical folding roughly aligns with the cyto-architectonic boundaries between the two regions (Brodmann, 1909; Amunts and Zilles, 2015; Fischl et al., 2008). In contrast, functional parcellations typically separate the motor regions along a ventral-dorsal axis, that is, into foot, hand and mouth regions. Within each body zone, these parcellations leave M1, S1, and premotor regions in the same parcel, likely reflecting the high functional correlations between regions responsible for the control of a body part. Similar observations can be made in the visual system - with functional parcellations separating areas associated with foveal and peripheral vision, rather than subdividing known cytoarchitectonic regions (V1, V2, V3). This anti-correlation of functional and anatomical boundaries partly explains why the Desikan atlas showed a significantly negative DCBC.

It is therefore also unsurprising that multi-modal parcellations that combine functional and anatomical criteria did not outperform the pure resting-state parcellations. For example, Glasser et al. (2016) used resting-state connectivity, intra-cortical myelin

signal, and cortical folding, thereby improving alignment with cytoarchitectonically defined areas. The inclusion of anatomical information likely also led to the division of functionally correlated brain regions. This does not imply that cytoarchitectonic parcellations of the neocortex are of lesser value. Instead, our finding simply illustrates that the goal of isolating anatomically consistently organised regions is unrelated to, and in some cases conflicts with, the goal of defining brain regions with homogeneous functional profiles.

Therefore, the evaluation results in our study would have a different look if we used anatomical data rather than task-evoked activity profiles as an evaluation data set. It is worth stressing, however, that the DCBC as a method to control for the influence of spatial smoothness can be used with any suitable data set. For instance, anatomical data, such as myelin measures (Tozer et al., 2005) or anatomical connectivity fingerprints derived from diffusion data (Behrens et al., 2003; Johansen-Berg et al., 2004) could be used to evaluate how well parcellations isolate anatomically homogeneous regions.

In the current study we focused on task-evoked and resting-state fMRI data as evaluation targets. While the two datasets led to a similar pattern of results when comparing parcellations, the overall DCBC values for rs-fMRI data were substantially lower than the DCBC values for task-based fMRI data. This is likely due to the lower signal-to-noise level for fMRI time-series, as compared to task activity estimates, which are averaged over 16 runs. Note that the high correlation for vertex pairs with small spatial distances (<5mm) are likely driven by the interpolation across neighboring voxels (motion realignment, surface mapping, minimal smoothing) inherent in both pre-processing pipelines. One advantage of evaluating parcellations on task-activity data is the obvious face validity of the results: If the goal of the the brain parcellation is to define regions with a homogeneous response across a wide range of tasks and mental states, then it should be best to evaluate the parcellation that way, rather than relying on the possibly more restricted mental states during rest.

A possible extension of the current work is to develop a parcellation algorithm that explicitly optimizes the DCBC. Given the nature of the DCBC, such an algorithm would have to be a local, rather than a global parcellation algorithm (Schaefer et al., 2018), such

as a region-growing algorithm proposed in [Gordon et al. \(2016\)](#) or [Salehi et al. \(2020a\)](#). The algorithm to be proposed would very likely improve the DCBC beyond what was achieved by spectral clustering, which does not consider the spatial arrangement of the vertices.

Even more substantial improvement in the quality of the parcellations can be expected when moving from a group to an individual parcellation. Recent results have shown that the inter-individual differences in the exact spatial location of functional boundaries are substantial ([Salehi et al., 2020b](#); [King et al., 2019](#)). Of course, individual parcellations can suffer from having too little individual data to reliably establish the parcellation. Optimal ways of fusing group and individual-level data ([Buckner et al., 2013](#); [Kong et al., 2019](#)), which also makes parcels comparable across subjects ([Salehi et al., 2020a](#)), clearly provides a promising future addition to the neuroimaging toolkit. In these efforts, the DCBC can provide a bias-free and reliable evaluation criterion that can be calculated without computationally expensive simulation studies.

When developing, using, and evaluating brain parcellations, it is of course important to consider the much more fundamental issue of whether this form of representation ([Bijsterbosch et al., 2020](#)) is a valid description of brain organisation. In our mind, it remains an open question to what degree segmenting the brain into discrete regions with hard boundaries ([Van Essen and Glasser, 2018](#)) is a useful description at all, or to what degree this functional organisation is better explained through a set of smooth gradients ([Tononi et al., 1994](#); [Dohmatob et al., 2021](#); [Huntenburg et al., 2018](#); [Guell et al., 2018](#)). Either way, we believe that the DCBC evaluation provides an useful tool to advance this debate. If brain functions only varied in smooth gradients across the cortical surface, the DCBC should not be systematically above zero, at least not when evaluated on a novel set of tasks. However, most resting-state parcellations identified boundaries that also aligned with more rapid changes in the active functional response. Thus, as for the human cerebellum ([King et al., 2019](#)), this demonstrates the existence of task-invariant functional boundaries on the cortical surface. On the other hand, not all boundaries are equally strong, and not all boundaries are equally stable across tasks. The ability of DCBC to evaluate individual boundaries, as done in [King et al. \(2019\)](#), therefore

provides an important tool to evaluate both functional segregation, as well as continuous functional integration ([Eickhoff et al., 2018](#)) in a region-specific way.

A Python-based software toolbox for the evaluation of surface-based parcellations on the MDTB activity maps is made publicly available at ([Zhi and Diedrichsen, 2021](#)). The toolbox is also designed to allow users to evaluate parcellations on other types of data. We hope that the new evaluation criterion will support and facilitate researchers in understanding the functional compartmentalization of the human brain.

Chapter 3

A hierarchical Bayesian brain parcellation framework for fusion of functional imaging datasets

3.1 Introduction

The application of machine learning to functional Magnetic Resonance Imaging (fMRI) data promises better models of brain organization. Brain parcellations, which subdivide the brain into a discrete set of functionally distinct regions, are one important type of model with many practical applications. A number of such parcellation schemes have been derived from large resting-state fMRI datasets (Yeo et al., 2011; Buckner et al., 2011; Power et al., 2011; Schaefer et al., 2018; Ji et al., 2019). Previous studies have shown that functional boundaries detected during resting-state are indeed predictive of functional boundaries during task performance (Cole et al., 2014; Laumann et al., 2015; Tavor et al., 2016). However, there is also increasing evidence for systematic differences in the functional organization measured during the task and rest setting (Hasson et al., 2009b; Cole et al., 2014; Greene et al., 2020). It is therefore important to consider task-based datasets in deriving brain parcellations (King et al., 2019), foreshadowing a comprehensive understanding of the dynamic nature of the brain’s functional organiza-

tion.

In recent years, an increasing number of high-quality task-based fMRI datasets that sample a broad range of tasks have become available (King et al., 2019; Nakai and Nishimoto, 2020; Pinho et al., 2018, 2020). Nonetheless, compared to the large and homogeneous resting-state datasets (Van Essen et al., 2013), task-based datasets usually only contain a small to medium number of individuals and are always limited in the tasks that they cover. It would be therefore highly desirable to have a principled way of combining evidence across many datasets into a single model. This is especially important as functional brain organization may not only differ between task and rest, but also between different task sets.

A second important practical problem is that functional brain organization shows considerable inter-individual variations even after anatomical variability is accounted for (Mueller et al., 2013), limiting the usefulness of functional group atlases. This problem could be potentially addressed by including individual resting-state (Wang et al., 2015) or task-based data (King et al., 2019; Pinho et al., 2018, 2020) as a functional localizer to derive individual brain parcellation maps. But a reliable characterization of brain organization requires an extensive amount of individual functional data (Marek et al., 2018), which in practice is often too costly to acquire.

In this paper, we addressed both of these problems by developing a hierarchical Bayesian parcellation framework (Fig. 3.1), which could be efficiently trained on a range of fMRI datasets, $\mathbf{Y}^{s,n}$, recorded in different sessions (n) from different subjects (s). The model assigns each of the possible brain locations in each individual to one of K functional regions (here referred to as parcels). The parcel assignments are collected in the matrices \mathbf{U}^s , with $U_{k,i}^s = 1$ if the i^{th} brain location is assigned to the k^{th} parcel. The model estimates the expected value of these latent variables, $\langle \mathbf{U}^s \rangle$, which provides a probabilistic parcellation for that individual (see Methods 3.4.1.1 for details). The model consists of a spatial *arrangement model*, $p(\mathbf{U}^s | \boldsymbol{\theta}_A)$, the probability of how likely a parcel assignment is within the studied population, and a collection of dataset-specific *emission models*, $p(\mathbf{Y}^{s,n} | \mathbf{U}^s; \boldsymbol{\theta}_{En})$, the probability of each observed dataset given the individual brain parcellation. This distributed structure allows the parameters of the model, $(\boldsymbol{\theta}_A, \boldsymbol{\theta}_{E1}, \dots)$ to

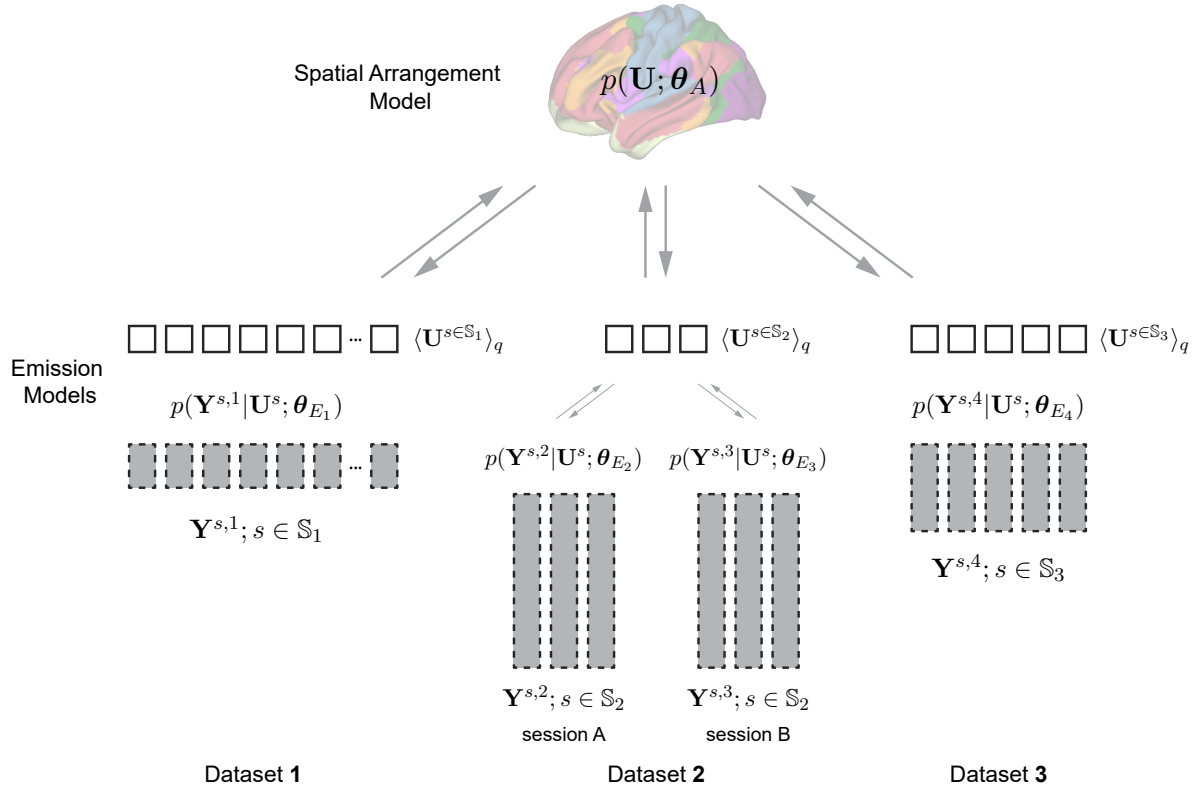


Figure 3.1: **A hierarchical Bayesian parcellation framework for data fusion.** Three datasets are shown. Data from each participant are indicated as a gray box. The height of the box indicates the amount of data per participant. Dataset 2 contains two sessions from the same set of participants ($s \in \mathbb{S}_2$), and the subject-specific data likelihood is integrated across the two sessions for each subject. The central quantity of the model is the estimated individual brain organization \mathbf{U}^s . The spatial arrangement model provides the population-wide probability of observing a specific brain organization.

be estimated using a message-passing algorithm between the different model components (Methods 3.4.1.4).

We applied the new framework to a collection of seven task-based fMRI datasets (Table 3.1), with four of them containing a wide range of task conditions and three others related to specific functional domains, including executive function and motor movements. Starting within a single dataset, we show that our framework optimally integrates data from a single individual subject with the group-based arrangement model, resulting in substantially improved individual brain parcellations. Then, we compare different approaches to estimate a unified group-based arrangement model across datasets, using both simulations and real data. We show that both group and individual parcellations

learned by datasets fusion using our proposed framework outperform the parcellations trained on each single dataset alone.

3.2 Results

3.2.1 Individual parcellations in the scarce data setting

Given the substantial inter-individual functional variability, it is often desirable to derive parcellations for single subjects. An important limitation, however, is that obtaining a reliable individual parcellation requires a substantial amount of data (Marek et al., 2018). A central feature of our model is that we do not only obtain a parcellation based on the individual data, $p(\mathbf{U}^s|\mathbf{Y}^s)$, and a parcellation based on the learned group parameters, $p(\mathbf{U}^s|\boldsymbol{\theta}_A)$, but also an optimal integration of individual and the group-level probability map (Methods 3.4.1.5). We first sought to determine how much improvement this integrated individual parcellation offers. For this purpose, we first trained a group parcellation (17 parcels) on the first task set of the multi-domain task battery dataset (MDTB, (King et al., 2019)). Individual parcellations were derived using between 1-16 imaging runs (10-160 min) of individual training data only. We compared the performance of these "data-only" parcellations with the group parcellation, and with individual parcellations learned in our framework by Bayesian integration of individual data and group map. All probabilistic parcellations were first converted into hard parcellations using a winner-take-all approach. We then evaluated how well the parcel boundaries corresponded to functional boundaries on the second task set (also 16 runs) acquired on the same subject. For this, we computed the distance-controlled boundary coefficient (DCBC, Zhi et al. (2022), the difference of the within-parcel and the between-parcel correlation of the functional profiles for each spatial distance (see Methods 3.4.5).

The individual parcellations based on 10 min of imaging data (without using the group probability map, Fig. 3.2a) performed generally poorly, with an average DCBC of 0.088 (Standard Error of the Mean, SEM = 0.009). Indeed, the individual parcellation performed worse than the group map $t_{23} = -7.786, p = 6.815 \times 10^{-8}$ (Fig. 3.2d, dashed

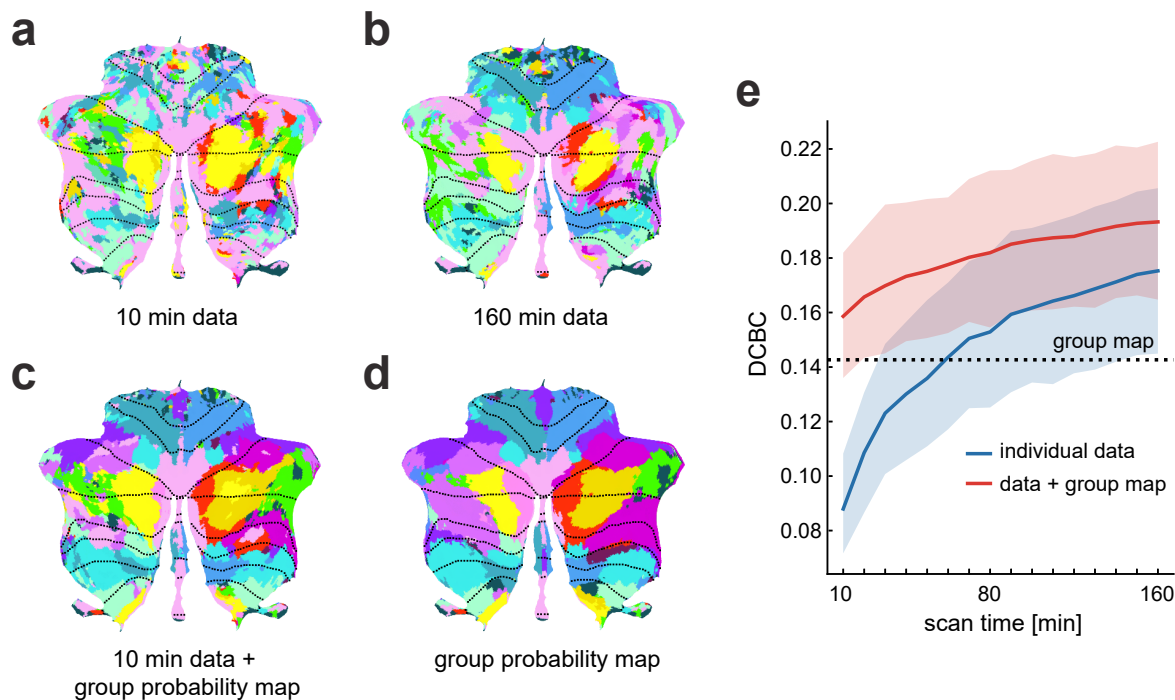


Figure 3.2: **Individual parcellations from the parcellation framework outperform group map.** (a) An estimated individual parcellation based on 10 min (1 run) of imaging data, using only the individual data. (b) An estimated individual parcellation of the same subject based on 160 min (16 runs), using only the individual data. (c) The estimated individual parcellation using 10 min of individual data and the group probability map learned by the arrangement model. (d) The group probability map from the arrangement model. (e) The DCBC value (higher = better) of the parcellations tested on the independent second session of the MDTB data set. Each individual parcellation was estimated using only the individual data (blue curve) or using the individual data and the learned group probability map (red curve). The x-axis indicates the length of the imaging time series (10 min = 1 run) used to estimate the individual parcellations. The error bars represent the standard error of the mean across all 24 subjects.

line in Fig. 3.2e). The individual parcellation improved continuously when using more data (Fig. 3.2b), reaching an average DCBC value of 0.175 (SEM = 0.016) for 160 min of data, ultimately outperforming the group map ($t_{23} = 3.286, p = 0.003$). This indicates that there are replicable differences in brain organization across individuals. Individual parcellations can capture these differences, leading to significantly better prediction performance than a group probability map on independent test data.

Although individual parcellations were superior to the group map using more data (blue line in Fig. 3.2e), our results suggest that more than 110 min of individual imaging

data is required to obtain a brain parcellation that is significantly better than the group probability map ($t_{23} = 2.190, p = 0.039$). At 60 min of imaging, an individual parcellation map is only just about as predictive as the group probability map. Acquiring this amount of individual data for functional localization is rarely feasible in basic and clinical functional imaging studies.

Our framework, however, automatically integrates the individual data with the group probability map, leading to dramatically improved performance of individual parcellations. For only 10 min of individual data, the DCBC was now significantly higher than the group probability map alone ($t_{23} = 3.123, p = 0.005$). Using 10 min of imaging data and our model led to individual parcellation performance that was roughly equivalent to using 100 min of individual imaging data without the model.

The resultant individual parcellation map (Fig. 3.2c) constitutes an optimal fusion of the individual data and the knowledge learned from the entire group. Even when there was a large amount of individual data available, such as 160 minutes, the integration with the group map led to a significant improvement relative to using only the individual data ($t_{23} = 5.562, p = 1.171 \times 10^{-5}$). Another advantage of the integration of group and individual data is that it naturally deals with missing individual brain data. For brain locations (voxels) where the individual data is missing, the group probability map will dictate the parcel assignment.

3.2.2 Dataset-specific emission models optimally capture differences in measurement noise

Different imaging datasets, or even sessions within a single dataset, often show different signal-to-noise ratios. For instance, two different imaging sessions of the IBC data set (Fig. 3.3a, Methods 3.4.2) show quite different levels of within-subject reliability, indicative of different levels of measurement noise. A simple approach to modeling different sessions from a single individual is to concatenate the data and model the two sessions with a single emission model (Type 1 model, Fig. 3.3b). In this scenario, however, it is possible that the second, noisier session will make the integrated model worse than

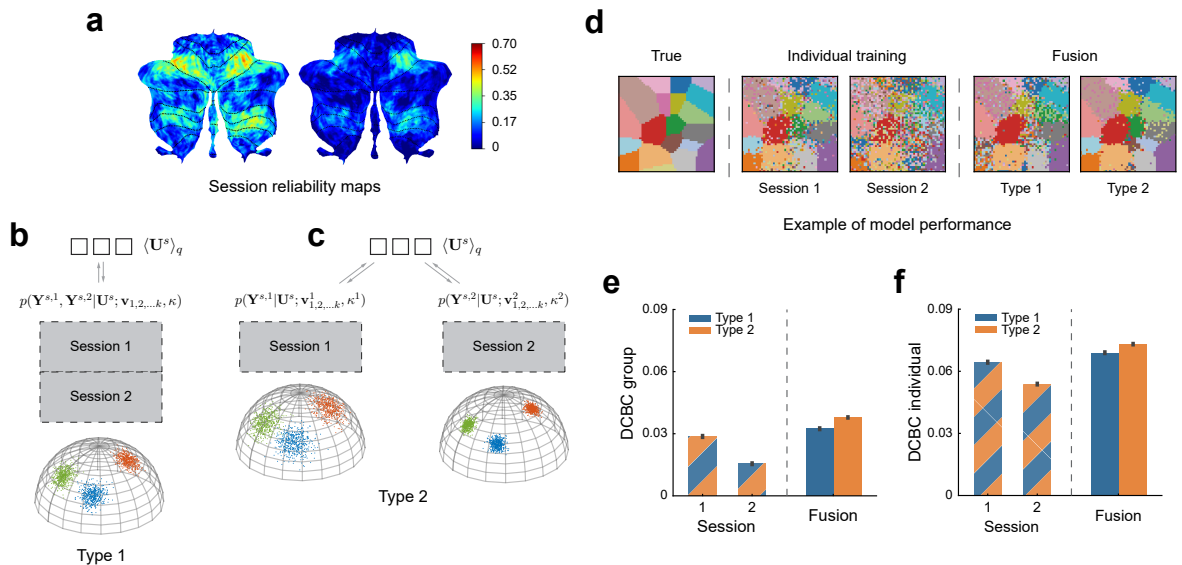


Figure 3.3: **Simulations of data fusion using two synthetic imaging sessions with similar task activation.** (a) The reliability map of two imaging sessions from the IBC dataset with similar task sets (*hcp1* and *archi*). (b) Type 1 model: sessions are concatenated and will be learned in a single emission model. (c) Type 2 model: sessions are separated and modeled using two separate emission models. (d) Reconstruction of the true parcellation map using synthetic data, using Session 1 or 2 alone vs. the fusion of both sessions using either model type 1 or 2. (e) The mean DCBC value of the group map (using model type 1 or type 2) learned from Session 1 or 2 alone or from the fusion of both sessions. (f) The mean DCBC value of individual parcellations. Error bars indicate SEM (standard error of the mean) across 100 simulations.

the first session alone. Therefore, in a different version of the model (Type 2), each imaging session was modeled with a separate emission model. This allows differences in variability to be captured by a session-specific concentration parameter (e.g. κ^1 for session 1 and κ^2 for session 2 in Fig. 3.3c). As long as the κ s are estimated accurately, the subsequent Bayesian integration will ensure the optimal weighting across the different sessions. Therefore, even the addition of a low-quality dataset should never lead to decreases in the quality of the integrated model.

To test for this behavior of the dataset-specific (Type 2) model, we generated two synthetic datasets (sessions) sampled from the same set of subjects with similar task activation but different overall noise variances, σ_k^2 (Methods 3.4.4). The measurement noise was set to $\sigma_k^2 = 0.5$ for synthetic session 1 and to $\sigma_k^2 = 0.8$ for session 2. We then learned group and individual parcellations using Type 1 or Type 2 models, either using

each session alone or fusing both sessions. We tested the performance of all models on an independent simulated test set (Methods 3.4.4), repeating the simulation 100 times.

The visual inspection of the group parcellations (Fig. 3.3d) suggests that the group map trained on session 1 alone approximates the true map more accurately than using session 2. The fusion of both sessions improved the group reconstruction, especially when using separate emission models (Type 2). We evaluated the parcellation performances quantitatively using the DCBC measure on the test set (Fig. 3.3e, 3.3f). Indeed, both group and individual maps learned from session 1 (DCBC group=0.029, DCBC individual=0.064) showed better performance averaged across 100 simulations than the one using session 2 (DCBC group=0.016, DCBC individual=0.054). When we evaluated the fusion parcellations, the DCBC value of the group and individual map learned by the Type 1 fusion model improved by 0.004 ($SD=3.752 \times 10^{-3}$) and 0.005 ($SD=3.781 \times 10^{-3}$) compared to dataset 1 alone, respectively. The parcellation performance of Type 2 fusion further improves compared to Type 1 by 0.005 ($SD=4.006 \times 10^{-3}$) for the group DCBC and 0.004 ($SD=4.666 \times 10^{-3}$) for the individual DCBC. These simulations demonstrate that session-specific emission models allow for better fusion when the signal-to-noise level differs across sessions or datasets.

3.2.3 Region-specific concentration parameters further improve fusion parcellation

In empirically observed task-based fMRI data, the signal-to-noise level does not only differ between sessions or datasets, but also between different regions within the same session or dataset. Some sessions or datasets provide a better signal-to-noise ratio for some functional regions and a lower signal-to-noise ratio for others. For example (Fig. 3.4a), the “*Preference*” session of the IBC dataset provided high within-subject reliability in the motor areas, whereas the perspective taking “*TOM*” session had high reliability in language-related areas. Ideally, a probabilistic framework should account for these differences and optimally combine the region-specific strengths of each dataset. To this end, we introduced a third variant of our emission model (Type 3), which has a separate

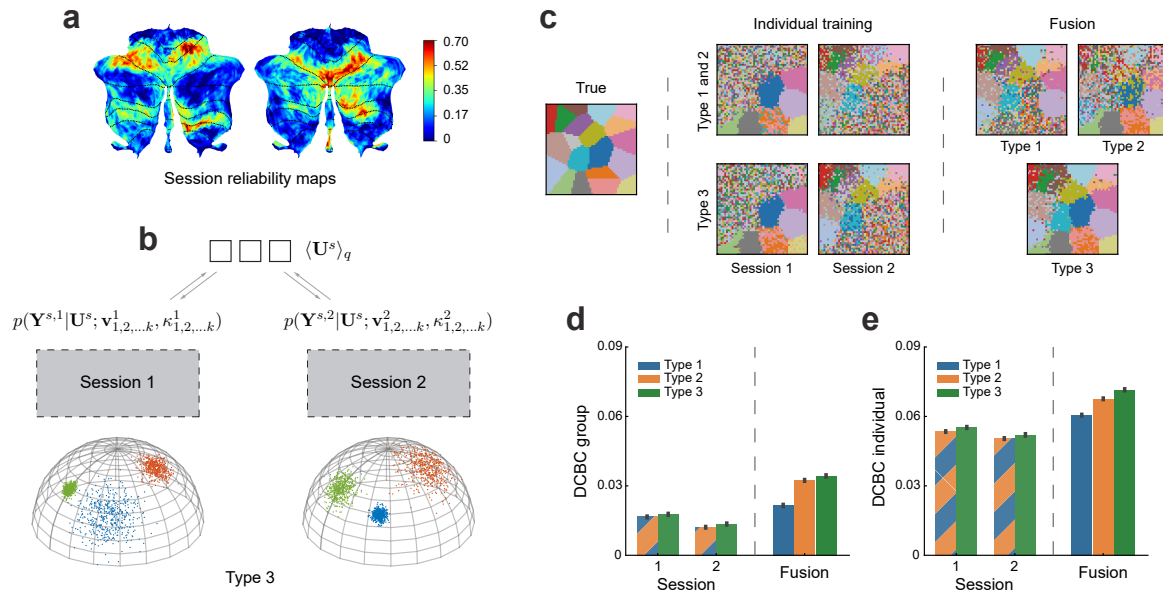


Figure 3.4: **Simulation on two synthetic sessions fusion with different task activation.** (a) The reliability map of two imaging sessions from the IBC dataset with different tasks (*Preference* and *TOM*). (b) Type 3 model: different sessions are modeled using different emission models, and furthermore the concentration parameters $\kappa_{1,2,\dots,k}$ are estimated separately for each parcel. (c) The comparison of reconstruction performance when leaned on synthetic session 1 or 2 alone vs. learned by data fusion using type 1, 2, or 3 models. (e) The mean DCBC value of the group map across sessions and model types. (f) The mean DCBC value of individual maps across sessions and model types. Error bars indicate SEM across 100 times simulation.

concentration parameter for each region and session (e.g. $\kappa_{1,2,\dots,k}^1$ for session 1 and $\kappa_{1,2,\dots,k}^2$ for session 2 in Fig. 3.4b).

To test the ability of this model to pool information across distinct datasets with different types of information, we conducted a second simulation by randomly dividing all functional regions into two groups. Instead of a common signal-to-noise level for all regions, we first created synthetic data in which one session had good signal-to-noise in the first group and poorer signal-to-noise in the other (Methods 3.4.4). We reversed the assignment for the second synthetic session. When we trained the model on Session 1 or 2 alone, there was high uncertainty of the cluster assignment in the area with low signal-to-noise level (Fig. 3.4c – *Individual training*). This is no surprise, as the activation here was too weak to detect the boundaries reliably.

Importantly, when combining the two sessions, the functional boundaries that were

not detected based on single sessions became visible (Fig. 3.4c – *Fusion*). However, both Type 1 and Type 2 models needed to compromise: when using session 1 to achieve parcellation of the lower right corner, the same weighting was applied to the upper left regions, decreasing the quality of the parcellation here. In contrast, model Type 3 allowed different weightings in different parcels, using mostly information from session 1 for the lower right parcels and mostly information from session 2 for the upper left regions. The quantitative evaluation (Fig. 3.4d, 3.4e) suggests a clear advantage of model Type 3 for both the group (improved 0.002, $SD=3.324 \times 10^{-3}$) and individual parcellation (improved 0.004, $SD=3.831 \times 10^{-3}$). We also verified the Type 3 model did not perform worse than Type 2 when two sessions had the same signal-to-noise level across all functional regions (see Supplementary Fig. A.3). Overall, the model with region-specific concentration parameters showed clear advantages when aggregating across sessions that differ not only in their overall signal-to-noise level, but also in what regions they specifically provide information for.

3.2.4 Model performance on real data and the choice of atlas resolution K

We then attempted to validate the performance of the models on real imaging data. Here, we first used the IBC dataset. This dataset is ideal to test the integration of data from different sessions across the same participants, as it consists of 14 sessions some of which have similar tasks while others do not (Pinho et al., 2018, 2020). We tested the different model types, each time fusing two IBC sessions ($C_{14}^2 = 91$ combinations) to learn a probabilistic parcellation model with 17 parcels. The learned models were then evaluated on the six other functional task-based fMRI datasets (see Tabel 3.1) in terms of their group and individual parcellations. To evaluate the ability of the model to provide individual parcellations, we split each evaluation dataset into two halves. The first half was used to infer the individual parcellations \mathbf{U}^s for the participants of the test set. The other half was used to calculate the DCBC value. We then reversed the role of the two halves and averaged performance across the two cross-validation folds.

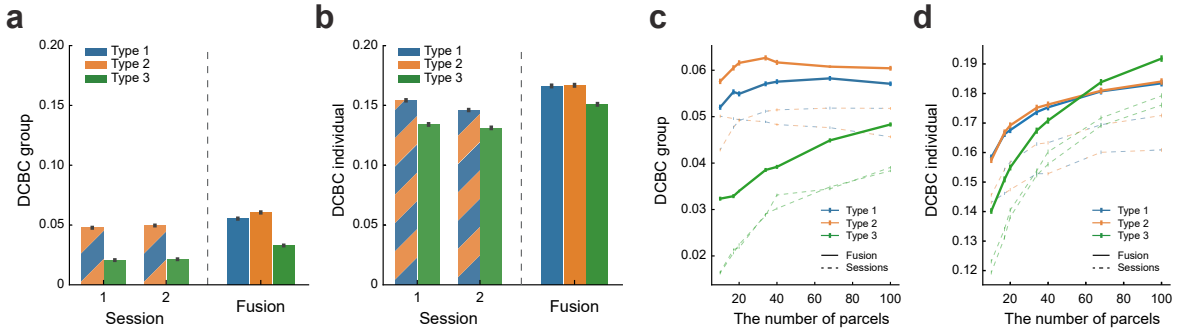


Figure 3.5: **DCBC evaluation of brain parcellations learned on two IBC sessions alone compared to the parcellation learned by fusion.** (a) Mean DCBC value of the group map across all 91 two-session combinations when training 17-regions parcellations. (b) Mean DCBC value of individual maps across all 91 two-session combinations when training 17-regions parcellations. (c) Mean DCBC value of the group map as a function of the number of parcels. (d) Mean DCBC value of the individual maps across all 91 two-session combinations as a function of the number of parcels. All error bars indicate the SEM across the evaluation subjects across the 6 task-based fMRI datasets.

We first confirmed that our probabilistic parcellation framework optimally learns group parcellation across sessions when comparing the performance to the parcellation learned from a single session. Specifically, all fusion parcellations showed substantial improvement (Fig. 3.5a) over the best one learned on a single session (Type 1: $t_{98} = 12.282, p = 1.513 \times 10^{-21}$, Type 2: $t_{98} = 18.749, p = 3.485 \times 10^{-34}$, Type 3: $t_{98} = 15.594, p = 2.698 \times 10^{-28}$). This improvement also held for individual parcellations (Fig. 3.5b, Type 1: $t_{98} = 15.283, p = 1.100 \times 10^{-27}$, Type 2: $t_{98} = 14.198, p = 1.624 \times 10^{-25}$, Type 3: $t_{98} = 9.353, p = 3.079 \times 10^{-15}$). Additionally, we found the group parcellations learned using session-specific emission models (Type 2) showed significantly better performance than the ones learned by concatenating the data (Type 1) ($t_{98} = 13.287, p = 1.196 \times 10^{-23}$).

Against our expectations, however, model Type 3 performed substantially worse on real data when compared to model Type 2 for both group ($t_{98} = -16.765, p = 1.521 \times 10^{-30}$) and individual ($t_{98} = -6.269, p = 9.807 \times 10^{-9}$) parcellations. This behavior differed markedly from our simulation results (Fig. 3.4), where model Type 3 performed consistently better. Further simulations suggested that this behavior can be explained

by the choice of the number of parcels (K): when K was close to or higher than the true number of parcels, model Type 3 outperformed model Type 2. If, however, K was chosen to be smaller than the true K , model Type 3 started to yield inferior performance (Supplementary Fig. A.4). In such cases, one parcel in model Type 3 typically had a very low concentration parameter, effectively capturing all voxels that are unexplained by the model. Model Type 2 constrains all functional regions to have the same concentration parameter, preventing the model from developing a “residual” parcel.

This idea suggests that model Type 3 should improve or even outperform model Type 2 when K increases and approaches the true number of parcels. Unlike the simulation, the true number of parcels in real data is unknown. We therefore estimated the fusion models on every pair of two IBC sessions using $K = (10, 17, 20, 34, 40, 68, 100)$. The evaluation results (Fig. 3.5c,d) indicated that the performance of the model Type 3 indeed improved with increasing K . This improvement was also clearly observed in individual parcellations (Fig. 3.5d), where the DCBC evaluation of the model Type 3 became as good as model Type 2 around $K = 60$ and showed a significant advantage at $K = 100$ ($t_{98} = 4.115, p = 8.059 \times 10^{-5}$). A similar pattern exists in the group map evaluation where the averaged DCBC value of 100 parcels substantially improved compared to the ones with only 10 parcels ($t_{98} = 28.191, p = 8.215 \times 10^{-49}$). For up to 100 parcels, the fusion parcellation from model Type 3 did not appear to be superior to the one from model Type 2 in group evaluation, however, we found this to be the case when considering more datasets (see Fig. 3.6e).

Overall, across analysis scenarios, we confirm that estimating separate concentration parameters for each session (Type 2) leads to better data fusion on real fMRI data. Additionally allowing a region-specific concentration parameter (Type 3) has both advantages and disadvantages: If the model assumes a large number of parcels, parcellations can improve. If, however, the assumed number of parcels is low, performance appears to be better when constraining the concentration parameter to be the same across regions.

3.2.5 The fusion atlas shows combined strengths across different task-based fMRI datasets

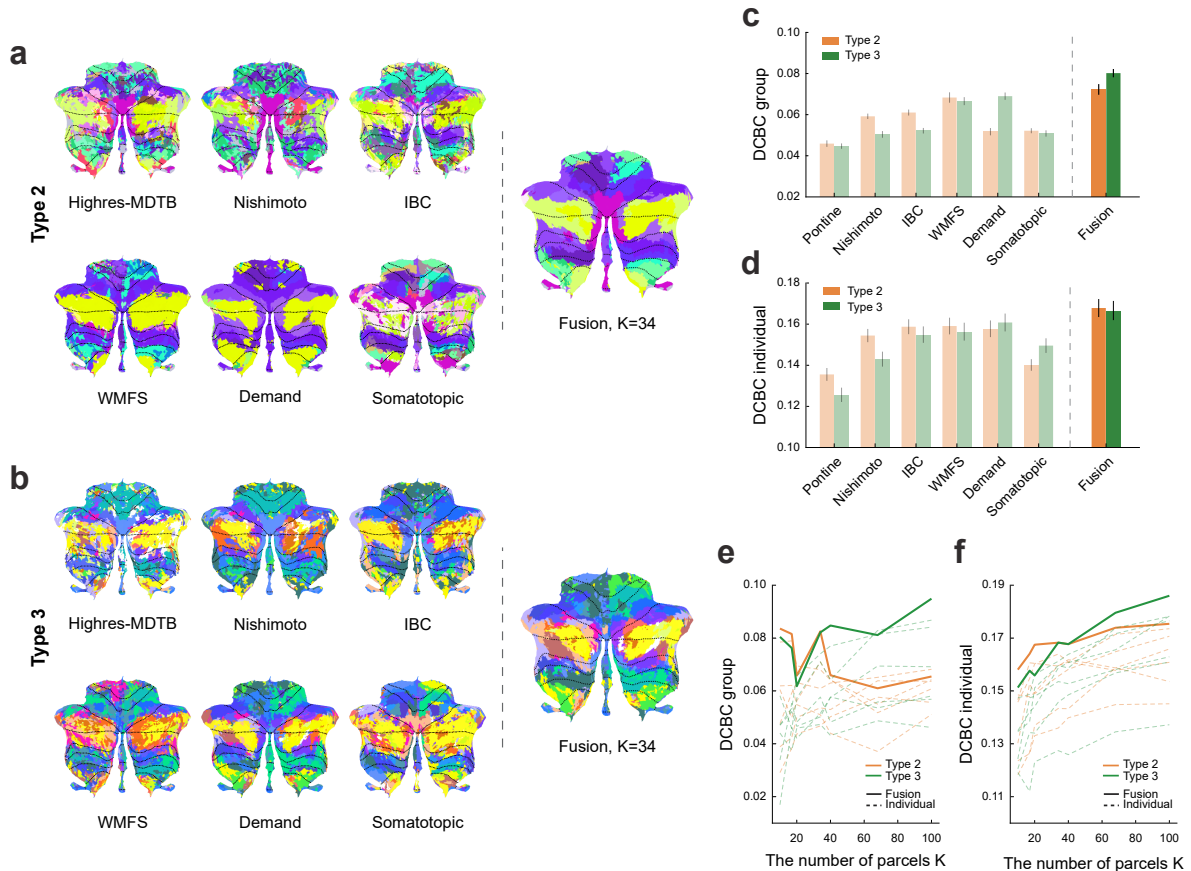


Figure 3.6: **Comparison of cerebellar parcellations learned by type 2 and 3 fusion models using 6 functional task-based datasets.** (a) The group parcellation maps learned on each individual dataset alone or datasets fusion with 34 parcels using the type 2 model. (b) Same as (a), but using the type 3 model. (c) Mean DCBC value of the group parcellation maps across subjects in the test dataset. (d) Mean DCBC value of the individual parcellation maps across subjects in the test dataset. (e) Mean DCBC value of the group map for $K = 10$ to $K = 100$. (f) Mean DCBC value of the individual map across for $K = 10$ to $K = 100$.

Finally, we trained our fusion model on 6 of the 7 task-based fMRI datasets (Table 3.1). We reserved the MDTB dataset as a test set. The resultant group maps of both models Type 2 and 3 showed the combined strength of the maps trained on individual datasets. For example, only the group map derived from the *Somatotopic* dataset delineated the foot region of the cerebellum (hemispheric lobule IV), while the ones derived

from other datasets did not. The Fusion maps (Fig. 3.6a,b) veridically retained this region. In contrast, the parcellation based on the *Somatotopic* dataset did not show a good parcellation of lobules Crus I and II, but here the fusion map used information from other datasets.

To evaluate the parcellations quantitatively, we calculated the DCBC on the left-out MDTB dataset (Fig. 3.6c,d). Averaged across all K s, all parcellations showed positive DCBC values, which means that the functional boundaries learned from any of the datasets generalized to some degree to the MDTB dataset. The best DCBC among parcellations trained on a single dataset was for the *WMFS* dataset for model Type 2 and for the *Demand* dataset for model Type 3. When we evaluated the fusion parcellations, we found considerable improvements for both models Type 2 and 3 compared to the best individual parcellation. For the fused parcellation using the model Type 2, both the group DCBC ($t_{23} = 2.339, p = 2.840 \times 10^{-2}$) and the individual DCBC ($t_{23} = 3.173, p = 4.248 \times 10^{-3}$) were considerably better than for *WMFS*. Similar improvement could be observed for model Type 3, where the fused parcellation significantly outperformed the best single-dataset parcellation (*Demand*) both in terms of the group ($t_{23} = 7.049, p = 3.503 \times 10^{-7}$) and individual ($t_{23} = 3.219, p = 3.800 \times 10^{-3}$) DCBC value.

Finally, we compared the fusion across the six task-based fMRI datasets directly between models Type 2 and 3. For $K = 10$, both averaged group and individual DCBC (Fig. 3.6e,f) were higher for model Type 2 than for model Type 3 (group: $t_{23} = 0.726, p = 0.475$; individual: $t_{23} = 1.842, p = 0.078$). But when K increased to 100, the fusion parcellation for model Type 3 became substantially better than model Type 2 (group: $t_{23} = 4.551, p = 1.426 \times 10^{-4}$; individual: $t_{23} = 2.468, p = 2.144 \times 10^{-2}$). The cross-over occurred somewhere around $K = 34$, where models performed equivalently (group: $t_{23} = 0.210, p = 0.835$; individual: $t_{23} = -0.009, p = 0.993$).

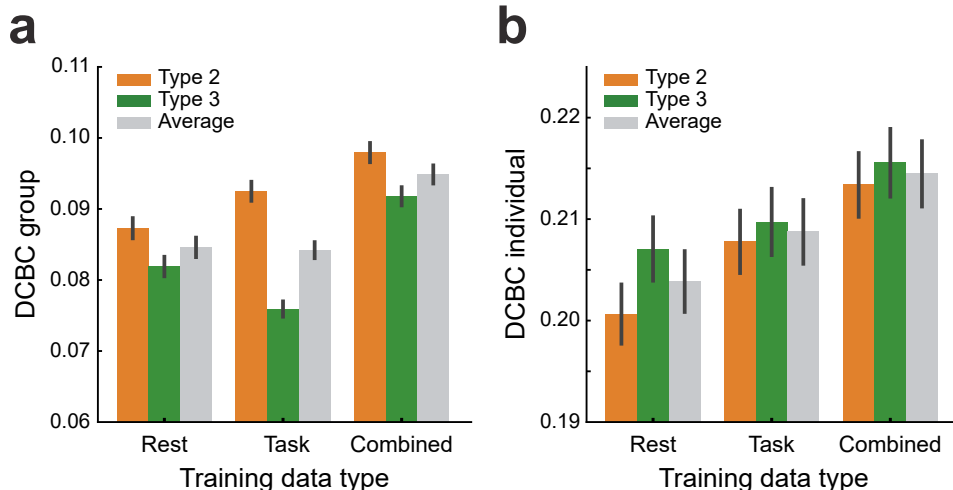


Figure 3.7: **Performance comparison of the cerebellar parcellations using resting-state data only, task-based data only, or the combination of both.** Probabilistic parcellations were learned using Type 2 (orange) or 3 (green) models. The gray bar indicates the averaged performance across the two models. **(a)** Mean group DCBC, and **(b)** mean individual DCBC evaluated on the task-based datasets in a leave-one-dataset out fashion. Error bar indicates the SEM across all 111 subjects of the 7 task datasets

3.2.6 Integrating resting-state data into the task-based parcellation

Lastly, we investigated the fusion of resting-state and task-based data into a single parcellation atlas. To do so, we used the cortical connectivity profile for each cerebellar voxel derived for 50 participants from the HCP *Unrelated 100* dataset (see Method 3.4.2). As we wanted to evaluate performance on a large range of task-based datasets, we trained the model on 6 out of 7 task datasets and evaluated the performance on the left-out task dataset. We then repeated this scheme for all 7 task datasets. The combined and resting-state parcellations were also trained and evaluated in a similar approach (for the resting-state parcellation, no dataset had to be left out).

Averaging the DCBC evaluations across models Type 2 and 3, the models trained on the combination of resting-state and task-based datasets outperformed the ones trained on resting-state or task-based datasets alone. For the group parcellation (Fig. 3.7a), the combined model was significantly better than the one trained on resting-state ($t_{110} =$

6.349, $p = 4.983 \times 10^{-9}$), and task-based datasets ($t_{110} = 3.886$, $p = 1.745 \times 10^{-4}$). Similar results were found for individual parcellations (Fig. 3.7b, vs. resting state alone: $t_{110} = 7.625$, $p = 9.287 \times 10^{-12}$, vs. task-based alone $t_{110} = 7.254$, $p = 6.027 \times 10^{-11}$).

3.3 Discussion

We developed a hierarchical Bayesian framework to learn probabilistic brain parcellation by fusing data from both functional task-based and resting-state fMRI datasets. Our work introduces two main innovations: First, by dividing the problem into a common spatial arrangement model and a set of dataset-specific emission models, we are able to optimally integrate information across many, quite heterogeneous, datasets. Second, because the framework directly models individual differences in brain organization, it provides not only a probabilistic group atlas, but also allows the user to obtain an optimal estimate of brain organization for new individuals.

Learning functional brain parcellations across datasets. While most of the current brain parcellations are generated using functional resting-state fMRI data, a number of studies (King et al., 2019; Cole et al., 2014) suggest that boundaries derived using resting-state data can differ systematically from those measured during task performance. One possible interpretation of this finding is that the boundaries of functional regions truly shift depending on the task the person performs (Salehi et al., 2020a). However, given that there is a basic common organization that is stable across rest and different tasks (King et al., 2019; Tavor et al., 2016), an alternative interpretation is that the boundaries stay the same, but are more or less visible depending on the task sets or mental states (such as rest) during which they are measured. This is obviously true for task sets that emphasize one specific aspect of mental function (See Fig. 3.4a), but also applies to resting-state data. For example, in resting-state data, the left and right-hand regions are usually highly correlated and often end up in the same parcel. However, when using a task set that contains both left and right unimanual movements, the two regions are readily dissociated (King et al., 2019). Therefore, the integration of data from a large array of tasks promises a more representative map of brain organization.

Because there is no single, large task-based dataset that would cover all of the mental functions, we developed here a framework that allows us to fuse data from a growing number of deep-phenotyping task-based datasets with fewer participants (King et al., 2019; Pinho et al., 2018, 2020; Nakai and Nishimoto, 2020; Assem et al., 2022). To make data fusion feasible in a Bayesian framework, we deployed a series of emission models, each one learns the specific characteristics of the corresponding dataset, including the expected response for each brain region and their variability. The integration across datasets is achieved through a common spatial arrangement model, which characterizes the variability of the functional organization across individuals. As shown in the simulations (Results 3.2.2 and 3.2.3), this allows us to integrate the strength of different datasets without inheriting their weaknesses. We can now deploy this framework to an increasing number of real datasets, namely the “wide” datasets with many participants (King et al., 2019), and “deep” datasets with only a few participants but a detailed characterization of each studied individual (Nakai and Nishimoto, 2020; Pinho et al., 2018, 2020). Given the message-passing algorithm (Methods 3.4.1.4), the individual datasets do not necessarily need to be hosted on the same server, but each dataset and emission model can be housed separately. This architecture will promote the scaling of the approach as it allows for distributed computing across many sites, which will be necessary to finally approach a “big-data” regime for learning complex models of functional brain organization. The distributed nature of data storage also makes the framework more suitable for clinical data, which for data privacy reasons often needs to remain within a dedicated server infrastructure.

Individual vs. group parcellation maps. Group parcellation maps identify patterns of functional organizations that are common and consistent across individuals. Group parcellations are in common use, as they provide a consistent framework to analyze and report functional imaging data, and can be applied using only the anatomical image from the individual. However, the boundaries between functional regions vary substantially across individual brains (Braga and Buckner, 2017; Gordon et al., 2017a; Kong et al., 2021), possibly biasing subsequent analysis (Bijsterbosch et al., 2018, 2019). Recent studies suggest that the inter-individual difference may be even more pronounced

in the human cerebellum (Marek et al., 2018). Therefore, using individual brain parcellations has the potential to improve the precision and quality of subsequent analyses. A major limitation, however, is that a substantial amount of individual data is necessary to derive an individualized map of sufficient quality (Marek et al., 2018). In our study, we found that 60 minutes of individual data were required to reach the same performance as the group map, and more than 110 minutes were necessary to substantially outperform it (see Results 3.2.1). For most studies, acquiring this amount of data for an individual functional localizer would be prohibitive, explaining the persistent popularity of group maps.

Different from previous approaches to leverage the group and individual parcellations (Salehi et al., 2018; Zhang et al., 2021), our approach performs the combination in a principled (Bayesian) way, weighting each part according to the respective uncertainty. Even when using a very short functional localizer (10min), the resultant individual parcellation outperforms the group map. Relative to parcellation built on individual data only, we found that the integrated estimate had a performance equivalent to using 100 min. Finally, the Bayesian approach also automatically deals with missing data from individuals due to lack of coverage or signal dropout.

While we derived and tested the individual parcellations for participants included in our training set, the model can also be used to derive individual parcellations on completely new participants. This would only require researchers to estimate a new emission model for the specific task set used for the functional localization data. In this process, the parameters of the arrangement model, which was trained across datasets, can be frozen. Therefore, an efficient estimation can be achieved even for small groups of participants, and results can be interpreted in the framework of established atlases. This approach makes individual functional localization in larger studies feasible, which is important for practical and clinical applications.

Comparing dataset-specific and regions-specific uncertainty parameters.

The concentration parameter (κ) in each emission model dictates how strongly the respective dataset is weighted, both when learning to determine the group parcellation map, and when deriving an individual parcellation. In this paper, we tested three ways

of estimating this concentration parameter: (a) we simply concatenated all sessions for each subject, giving the entire dataset a single concentration parameter (Type 1); (b) we used a separate emission model and, therefore, a separate concentration parameter for each session (Type 2); and (c) we used a separate concentration parameter for each session and region (Type 3).

We first showed that the model Type 2 performed better than model Type 1 in capturing different levels of measurement noise from different sessions in both simulation and real data (Results 3.2.2, 3.2.4). However, when we compared Type 2 (dataset-specific) and Type 3 (region-specific) models, we found that each had specific advantages, which is also dependent on the choice K , the number of parcels (Results 3.2.4). When allowing separate concentration parameters for each session and region (Type 3), we can account for the fact that some sessions may contain tasks that provide signals in some areas, while other sessions may highlight other areas. This example clearly the case in the IBC dataset (Fig. 3.4a). While in simulation, model Type 3 led to superior performance, on real data it often performed worse than model Type 2. In model Type 3, we found that when the assumed number of parcels (K) was smaller than the true number of parcels, one region would be estimated to have a very low concentration parameter, such that it could model all the residual, non-explained regions. Such a residual region led to a more fragmented group parcellation (Fig. 3.6b) and an impaired evaluation of the independent data.

However, the constraint of equal concentration parameters across all regions (model Type 2) prevented this from happening. This led to compact clusters regardless of the choice of K (Fig. 3.6a). Nonetheless, for large K , model Type 3 could outperform model Type 2. The choice of emission model (Type 2 or Type 3) therefore will depend on desired granularity of the parcellation and likely also the amount and quality of data available. Our framework offers both implementations, allowing the user to choose the correct algorithm in a context-specific manner.

Choice of datasets: task-based vs. resting-state fMRI. Our evaluation of task-based and resting-state parcellations (Fig. 3.6) shows that both can predict the functional boundaries measured in a left-out task set well above chance. A visual inspec-

tion of the two parcellations (Supplementary Fig. A.5a,b), however, also reveals some systematic differences (King et al., 2019; Cole et al., 2014). One important decision when applying our framework is therefore which datasets to include. Our Bayesian framework weights each dataset according to its reliability. Because different datasets will emphasize different sets of functional boundaries, and because the true number of functional regions is likely larger than the number of assumed parcels, each dataset will bias the final parcellation in a specific direction. A single large dataset could dominate the group map, possibly reducing the predictive performance of other datasets. It is therefore important to achieve a good balance between resting-state and task-based datasets highlighting different cognitive domains (Salehi et al., 2020a). Where this balance lies, or whether it is preferable to have different brain parcellations for different functional states, remains a research question that demands further attention.

Limitations and further developments. Being able to leverage an increasing number of datasets will hopefully also allow the development of models that can learn regularities in the spatial arrangement of functional regions in the human brain. In this paper, we have used only the independent spatial arrangement model, which in essence learns a probabilistic group atlas. In our framework, however, we can also use models that make assumptions about the intrinsic smoothness of individual functional parcellations, such as a Markov Random Field (MRF) spatial prior (Ryali et al., 2013; Schaefer et al., 2018; Kong et al., 2019) with coupling parameters. As a further extension, deep generative models, such as a deep Boltzmann machine (Salakhutdinov and Hinton, 2009), provide a promising avenue to actually learn the complex short- and long-range dependencies in functional brain organization. While the emission models would remain the same, both the E-step and M-step for the spatial arrangement model would rely on an approximation through sampling. Training such models will require a large amount of data, and our framework takes a critical step in this direction by enabling the integration of a wide range of varied datasets.

Another direction for possible improvements is to explore other forms of emission models. Here, we used a mixture of vMF distributions (Methods 3.4.1.2), which for both resting-state and task-based data has been shown to perform considerably better

than a mixture of multivariate Gaussians (Røge et al. (2017), Supplementary Fig. A.1). In contrast to resting-state data, task-based data provide not only a direction of the functional profile, but also a signal amplitude, as the experimental paradigm allows for separate estimation of signal and noise. The signal amplitude could be potentially used to distinguish between reliable and noisy brain locations: we found that the functional profile at brain locations with larger signal magnitudes tended to be closer to the mean response vector for that region. Nonetheless, the vMF model ignores this information as each profile is standardized to unit length. During the development of the model, we therefore experimented with weighted vMF models, in which the emission log-likelihood from each brain location was weighted by its respective signal-to-noise level. In practice, however, we found the final performance of the model did not improve enough to outweigh the possible instabilities in the estimation of the weights. We then decided to stay with the standard vMF mixture model for this paper. But, an emission model with voxel-, region-, and subject-specific signal-to-noise parameters might be useful for certain applications.

Conclusion. This article designs and evaluates a hierarchical Bayesian parcellation framework for data fusion across heterogeneous data sources. In conjunction with a collection of task-based and resting-state datasets which were preprocessed and stored in a consistent manner, the framework enables optimal estimation of functional brain organization across a range of diverse datasets. Here, we have applied the framework to derive new functional maps of the human cerebellum - however, the same process could be repeated nearly effortlessly for any other brain structure.

We anticipate that this framework will be useful for two reasons. First, the model can provide individual functional parcellations for new subjects using very limited individual data. While normally individual parcellations require an extensive amount of data (Marek et al., 2018), our framework makes it feasible to derive an individual region definition of considerably better quality than a group map with 10 min of functional localizer data. Secondly, the framework allows the optimal fusion of functional insights using a range of different task-based datasets, thereby overcoming the limitation that current task-based datasets are restricted both in terms of the breadth of their task

battery and the number of subjects. The framework accurately quantifies the different signal-to-noise levels across sessions and datasets, thereby providing an optimal weighting for each. The resultant maps possess a combined strength in detecting the detailed functional boundaries, outperforming the parcellations trained by single datasets.

3.4 Methods

3.4.1 A hierarchical Bayesian parcellation framework for data fusion

We introduce here a hierarchical Bayesian framework that can be used to learn a probabilistic brain parcellation across multiple fMRI datasets. The framework (Fig. 3.1) consists of a group-based brain parcellation model (the spatial arrangement model), and a series of dataset-specific emission models. The two parts of the framework are connected by a message-passing and collaborative-learning process, making learning and inference computationally efficient.

The framework is able to learn parcellations from a collection of data $\mathbf{Y}^{s,n}$ recorded from different subjects (s) during different experiments or sessions (n). \mathbb{S}_n is the set of subjects for the n -th experiment or session, and $\mathbb{S} := \{\mathbb{S}_1 \cup \mathbb{S}_2 \cup \dots \cup \mathbb{S}_n\}$ is the entire set of unique subjects. The parcellation model assigns each of the P possible brain locations in each individual s to one of K functional regions (here referred to as parcels). The parcel assignment for the i -th brain location is denoted in the one-hot encoded vector \mathbf{u}_i^s , and collected into the $K \times P$ matrix \mathbf{U}^s . This individual brain organization is the central latent variable in the model. The model estimates the expected value, $\langle \mathbf{U}^s \rangle$, which provides a probabilistic parcellation for that individual - specifically $\langle \mathbf{u}_{i,k}^s \rangle$ is the probability that brain location i is part of the functional region k . Note that we use $\langle \cdot \rangle$ to denote the expected value throughout.

The arrangement model provides a probabilistic group model of how likely a certain parcel assignment to brain locations is across individuals, $p(\mathbf{U}; \boldsymbol{\theta}_A)$. This probability depends on a set of (to-be-estimated) parameters of arrangement model ($\boldsymbol{\theta}_A$). In this

paper, we use a spatial arrangement model that estimates these probabilities for each brain location independently (Methods 3.4.1.3), and therefore effectively learns a group-based probabilistic brain atlas (see discussion for further extensions).

Each emission model specifies the likelihood of observed data given the individual brain parcellation, $p(\mathbf{Y}^{s,n}|\mathbf{U}^s; \boldsymbol{\theta}_E)$. For each dataset/session, we introduce a separate emission model with a separate set of emission-model parameters ($\boldsymbol{\theta}_E$). This allows us to model the different task sets with different signal-to-noise levels inherent in each experiment/session.

3.4.1.1 EM algorithm for Probabilistic parcellation with data fusion

We used an *Expectation Maximization* (EM) algorithm to optimize the parameters ($\boldsymbol{\theta}$) of the hierarchical Bayesian model. For such models, the computation of the log-likelihood of the data, $\log p(\mathbf{Y}^s; \boldsymbol{\theta})$, is unfeasible given a large number of latent variables in the model (here, these are the individual brain parcellations \mathbf{U}^s).

The key idea in EM is to introduce a proposal distribution over the latent variables $q(\mathbf{U})$, and then to optimize the *Evidence Lower Bound* (ELBO) of the model (Wainwright et al., 2008; Blei et al., 2017). The ELBO provides a lower bound to the full likelihood (over all datasets and subjects) that we want to optimize:

$$\sum_{s,n} \log p(\mathbf{Y}^{s,n}; \boldsymbol{\theta}) \geq \sum_{s,n} \langle \log p(\mathbf{Y}^{s,n}, \mathbf{U}^s; \boldsymbol{\theta}) \rangle_q - \langle \log q(\mathbf{U}^s) \rangle_q \quad (3.1)$$

The first term of the ELBO is the expected complete log-likelihood \mathcal{L} . Given the model structure, this quantity can be further split into the expected emission log-likelihoods \mathcal{L}_{E_n} for each experiment or session and the expected arrangement log-likelihood \mathcal{L}_A as:

$$\begin{aligned} \mathcal{L} = \sum_{s,n} \langle \log p(\mathbf{Y}^{s,n}, \mathbf{U}^s; \boldsymbol{\theta}) \rangle_q &= \sum_{s \in \mathcal{S}_1} \langle \log p(\mathbf{Y}^{s,1} | \mathbf{U}^s; \boldsymbol{\theta}_{E1}) \rangle_q + \sum_{s \in \mathcal{S}_2} \langle \log p(\mathbf{Y}^{s,2} | \mathbf{U}^s; \boldsymbol{\theta}_{E2}) \rangle_q \\ &+ \dots + \sum_s \langle \log p(\mathbf{U}^s; \boldsymbol{\theta}_A) \rangle_q \triangleq \mathcal{L}_{E1} + \mathcal{L}_{E2} + \dots + \mathcal{L}_A \end{aligned} \quad (3.2)$$

where the parameters are subdivided into those for the arrangement model, $\boldsymbol{\theta}_A$, and those for each of the emission models $\{\boldsymbol{\theta}_{E1}, \boldsymbol{\theta}_{E2}, \dots\}$. This division makes the parameter

updates that can be performed independently for the arrangement and emission models.

In the expectation step, the ELBO is increased by updating the proposal distribution $q(\mathbf{U}^s)$ to the approximate posterior distribution, given the current set of parameters as

$$\begin{aligned} q(\mathbf{U}^s) &= p(\mathbf{U}^s | \mathbf{Y}^{s,1}, \mathbf{Y}^{s,2}, \dots; \boldsymbol{\theta}) \\ &\propto p(\mathbf{Y}^{s,1} | \mathbf{U}^s; \boldsymbol{\theta}_{E1}) \times p(\mathbf{Y}^{s,2} | \mathbf{U}^s; \boldsymbol{\theta}_{E2}) \times \dots \times p(\mathbf{U}^s; \boldsymbol{\theta}_A). \end{aligned} \quad (3.3)$$

This step also allows us to calculate the expectation of the latent variables, resulting in an estimate of the individual brain parcellations $\langle \mathbf{U}^s \rangle_q$. In the maximization step, we update these parameters using these estimated individual brain parcellations.

3.4.1.2 Dataset-specific emission models

One commonly-used choice for a model of fMRI data across regions is the *Gaussian Mixture Model* (GMM) (Golland et al., 2008). However, the amplitude of fMRI brain signals \mathbf{y}_i (whether or not they are normalized by the measurement noise) vary greatly between datasets, participants, and brain locations. That is, two voxels in the same region may have highly correlated signals, but the amplitude of one may be twice as large as the other. Therefore, an increasing number of modeling approaches for resting-state fMRI data use a mixture of *von Mises-Fisher* (vMF) distributions (Banerjee et al., 2005; Ryali et al., 2013; Schaefer et al., 2018; Lashkari et al., 2010; Yeo et al., 2011). It has been demonstrated that such a directional distribution outperforms the GMM in modeling resting-state fMRI data (Røge et al., 2017). Here, we confirmed this is also the case for task-based fMRI data: the vMF mixture model performed better in model evaluation than the GMM (Supplementary Fig. A.1). We thus used the vMF mixture as our primary emission model.

The probability density function of a N -dimensional ($N \geq 2$) vMF distribution for a data point \mathbf{y}_i ($\|\mathbf{y}_i\| = 1$) is defined as:

$$p_N(\mathbf{y}_i | \mathbf{v}, \kappa) = c_N(\kappa) \cdot \exp(\kappa \mathbf{v}^\top \mathbf{y}_i) \quad (3.4)$$

where \mathbf{v} denotes the mean direction ($\|\mathbf{v}\| = 1$), κ indicates the concentration parameter ($\kappa \geq 0$). The normalizing constant $c_N(\kappa)$ is given by:

$$c_N(\kappa) = \frac{\kappa^{\frac{N}{2}-1}}{(2\pi)^{\frac{N}{2}} I_{\frac{N}{2}-1}(\kappa)} \quad (3.5)$$

where $I_r(\cdot)$ refers to the modified Bessel function of the r order. In a k -classes vMF mixture, each of the $1 \leq k \leq K$ parcels is specified with the parameters $\{\mathbf{v}_k, \kappa_k\}$, where κ_k is the concentration parameter and \mathbf{v}_k is the mean direction vector. Because any spatial dependency in the data is modeled through the arrangement model, these emission log-likelihoods can be computed separately for each brain location i . For each subject s and emission model n , we can calculate the data log-likelihood for each i brain location as:

$$\ell_{i,k}^{s,n} = \log p(\mathbf{y}_i^{s,n} | \mathbf{u}_i^s(k) = 1; \boldsymbol{\theta}_{En}) = \log c_N(\kappa_k) + \kappa_k \mathbf{v}_k^\top \mathbf{y}_i^s \quad (3.6)$$

We explored three variants of this model: (a) **Type 1** model assumes that all sessions of a single subject are concatenated and modeled with a single emission model; (b) **Type 2** model uses different emission models for different sessions from the same participant (Fig. 3.1, Dataset 2). Evidence from different sessions of the same subject is combined during the message passing (eq. 3.3). Different sessions have different concentration parameters κ , providing the possibility of adaptive weighting across sessions. The concentration parameter, however, is assumed to be the same across all parcels; (c) **Type 3** model is identical to the Type 2 model, but employs a different concentration parameter, κ_k , for each of the parcels. In the maximization step, the emission model parameters $\boldsymbol{\theta}_E := \{\mathbf{v}_k, \kappa_k\}$ are updated by maximizing the expected emission log-likelihood \mathcal{L}_E (Supplementary materials A.2).

3.4.1.3 The spatial arrangement model

The arrangement model aims to provide a (possibly not normalized, see discussion) probability measure $p(\mathbf{U}; \boldsymbol{\theta}_A)$ for each unique individual s ($s \in \mathbb{S}$) in the studied population

over a set of latent variables \mathbf{u}_i^s , which indicates the affiliation of a certain brain location i to a specific functional region k . We considered here the most basic architecture for the spatial arrangement model, namely the **independent arrangement model**, where different brain locations are considered to be mutually independent. In this case, the spatial arrangement model learns the group probability at each location i across all subjects belonging to a parcel k , denoted as $p(\mathbf{u}_i(k))$. We parameterize this model using a group log-probability parameter $\eta_{i,k}$ for each brain location i and parcel k :

$$p(\mathbf{u}_i(k)) = \frac{\exp(\eta_{i,k})}{\sum_j \exp(\eta_{i,j})} \quad (3.7)$$

This arrangement model can be estimated using the EM algorithm for inference. In the E-step, we calculate the posterior $\langle \mathbf{u}_i^s(k) \rangle_q$ for each individual by integrating the log evidence from the data and the group probability map $\eta_{i,k}$:

$$\begin{aligned} \langle \mathbf{u}_i^s(k) \rangle_q &= p(\mathbf{u}_i^s = k | \mathbf{y}_i^s; \boldsymbol{\theta}_A, \boldsymbol{\theta}_E) \\ &= \frac{\exp(\log p(\mathbf{y}_i^s | \mathbf{u}_i^s = k; \boldsymbol{\theta}_E) + \eta_{i,k})}{\sum_j \exp(\log p(\mathbf{y}_i^s | \mathbf{u}_i^s = j; \boldsymbol{\theta}_E) + \eta_{i,j})}. \end{aligned} \quad (3.8)$$

The arrangement model parameters $\boldsymbol{\theta}_A := \{\eta_{i,k}\}$ are then updated in the M-step (Supplementary materials A.6).

3.4.1.4 Message passing and collaborative learning

Since the full model breaks into different parts (Fig. 3.1), the learning algorithm can be partitioned into separate E-steps and M-steps for arrangement and emission models (Algorithm 1). The two parts then communicate through a message-passing process. Specifically, if there are multiple emission models in the framework, each of the n emission models calculates $\ell_{i,k}^{s,n}$ (eq. 3.6) for each individual s .

During *message-passing*, the evidence for a single subject s is integrated across any experiment/session that is available for this subject (e.g. Dataset 2 in Fig. 3.1),

$$\ell_{i,k}^s = \sum_n \ell_{i,k}^{s,n}. \quad (3.9)$$

Algorithm 1: EM algorithm of the fusion framework

Input: K , fMRI data for subject s and experiment/session n $\{\mathbf{Y}^{s,n}, \dots\}$, initial emission model parameters $\boldsymbol{\theta}_E^{(0)}$, initial arrangement model parameters $\eta_{i,k}^{(0)}$

Output: the final estimated parameters $\eta_{i,k}^{(t)}$ and $\langle \mathbf{u}_i^s(k) \rangle_q^{(t)}$

- 1 Initialize: $t = 0$, $t_{max} = 200$, $\Delta = 0.01$
- 2 **while** $t \leq t_{max}$ **do**
- 3 calculate emission log-likelihoods eq.3.6 for each experiment/session:
- 4 **for** $n = 1$ to N **do**
- 5 emission E-step for each available subject s in $\mathbf{Y}^{s,n}$ using eq.A.2:
- 6 $\ell_{i,k}^{s,n(t)} = \log p(\mathbf{y}_i^{s,n} | \mathbf{u}_i^s(k) = 1; \boldsymbol{\theta}_{En}^{(t)})$
- 7 **end**
- 8 sum emission log-likelihoods across experiments/session for each subject:
- 9 $\ell_{i,k}^s(t) = \sum_n \ell_{i,k}^{s,n(t)}$
- 10 arrangement E-step using Supplementary eq.A.4:
- 11
$$\langle \mathbf{u}_i^s(k) \rangle_q^{(t)} = \frac{\exp(\ell_{i,k}^s(t) + \eta_{i,k}^{(t)})}{\sum_j \exp(\ell_{i,j}^s(t) + \eta_{i,j}^{(t)})}$$
- 12 calculate expected complete log-likelihood by summing up eq.3.10 and eq.3.11:
- 13
$$\begin{aligned} \mathcal{L}^{(t)} &= \mathcal{L}_A^{(t)} + \sum_n \mathcal{L}_{En}^{(t)} \\ &= \sum_{s \in \mathbb{S}} \sum_i \sum_k \langle \mathbf{u}_i^s(k) \rangle_q^{(t)} \cdot \eta_{i,k}^{(t)} + \sum_n \sum_{s \in \mathbb{S}_n} \sum_i \sum_k \langle \mathbf{u}_i^s(k) \rangle_q^{(t)} \cdot \ell_{i,k}^{s,n(t)} \end{aligned}$$
- 14 check converge criterion:
- 15 **if** $t \geq 1$ and $\mathcal{L}^{(t)} - \mathcal{L}^{(t-1)} < \Delta$ **then**
- 16 **return** $\eta^{(t)}$, $\langle \mathbf{u}_i^s(k) \rangle_q^{(t)}$
- 17 **end**
- 18 arrangement M-step using Supplementary eq.A.6:
- 19 $\eta_{i,k}^{(t+1)} \leftarrow \log \sum_s \langle \mathbf{u}_i^s(k) \rangle_q^{(t)}$
- 20 **for** $n = 1$ to N **do**
- 21 emission M-step by eq.A.7, and eq.A.8 (Type 1, 2) or A.10 (Type 3)
- 22 $\boldsymbol{\theta}_{En}^{(t+1)} \leftarrow \operatorname{argmax}_{\boldsymbol{\theta}_{En}} \mathcal{L}_{En}^{(t)}(\boldsymbol{\theta}_{En})$
- 23 **end**
- 24 $t \leftarrow t + 1$
- 25 **end**

These combined emission log-likelihoods $\ell_{i,k}^s$ are then collected and passed to the arrangement model. The arrangement model then computes the posterior expectation $\langle \mathbf{u}_i^s(k) \rangle_q$ (eq. 3.8) of the parcel assignment in each subject s , which are collected into a $K \times P$ matrix $\langle \mathbf{U}^s \rangle_q$. These quantities are then used to calculate the expected emission log-likelihoods \mathcal{L}_{En} and the expected arrangement log-likelihood \mathcal{L}_A . In case of an independent arrangement model, the expected arrangement log-likelihood \mathcal{L}_A can be computed in closed form:

$$\mathcal{L}_A = \sum_{s \in \mathbb{S}} \langle \log p(\mathbf{U}^s; \boldsymbol{\theta}_A) \rangle_q = \sum_{s \in \mathbb{S}} \sum_i \sum_k \langle \mathbf{u}_i^s(k) \rangle_q \cdot \eta_{i,k}. \quad (3.10)$$

Similarly, the expected emission log-likelihood is calculated by multiplying the data log-likelihood in eq. 3.6 with the posterior expectation (eq. 3.8) and summing these quantities over subjects, brain locations, and parcels:

$$\mathcal{L}_{En} = \sum_{s \in \mathbb{S}_n} \langle \log p(\mathbf{Y}^{s,n} | \mathbf{U}^s; \boldsymbol{\theta}_{En}) \rangle_q = \sum_{s \in \mathbb{S}_n} \sum_i \sum_k \langle \mathbf{u}_i^s(k) \rangle_q \cdot \ell_{i,k}^{s,n}. \quad (3.11)$$

The sum of these expected log-likelihoods \mathcal{L} (in eq. 3.2) is then used as an objective function to check the convergence.

In the implementation, the algorithm takes inputs of the fMRI datasets $\mathbf{Y}^{s,n}$ ($n = 1, 2, \dots, N$ and $s \in \mathbb{S}$) with the initial arrangement and emission model parameters, $\eta_{i,k}^{(0)}$ and $\theta_E^{(0)}$. The parameters $\eta_{i,k}^{(0)}$ were initiated randomly from a normal distribution. For the initial emission model parameters, the mean direction vectors $\mathbf{v}_k^{(0)}$ were also drawn from a normal distribution and normalized to be unit vectors. The initial concentration parameters $\kappa_k^{(0)}$ were randomly drawn from a uniform distribution between 10 to 150, as we wanted to start with a “medium-sized” directional variance. After convergence, the algorithm returns the estimated group parameters for arrangement and emission models, as well as the posterior expectation $\langle \mathbf{u}_i^s(k) \rangle_q^{(t)}$ from the last iteration (t).

3.4.1.5 Individual and group parcellations

Once the model is trained, the group probability map can be derived from the estimated parameters of the spatial arrangement model. For the independent arrangement model, the k -long vector of probabilities at each brain location is $\langle \mathbf{u}_i \rangle = \text{softmax}(\eta_i^{(t)})$. To obtain a hard parcellation for later evaluation (Methods 3.4.5), we applied a winner-takes-all approach assigning each brain location i to the parcel with the highest probability ($\arg \max_k \langle \mathbf{u}_i(k) \rangle$).

Individual parcellations can be obtained even for the individuals that were not part of the model training by applying a single E-step, using the trained parameters and their data. This procedure effectively integrates the individual data likelihood with the group probability map. Since we assume an independent spatial arrangement model, the posterior expectation for a location i of subject s can be exactly computed as,

$$\langle \mathbf{u}_i^s \rangle = \text{softmax}(\log p(\mathbf{y}_i^s | \mathbf{u}_i^s; \boldsymbol{\theta}_E) + \eta_i). \quad (3.12)$$

Similarly, a hard individual parcellation was then again obtained by assigning i to the region with the highest probability, $\arg \max_k \langle \mathbf{u}_i^s(k) \rangle$. For the comparison reported in Results 3.2.1, we also derived a parcellation only based on data likelihood without taking the group probability into account:

$$\langle \mathbf{u}_i^s \rangle = \text{softmax}(\log p(\mathbf{y}_i^s | \mathbf{u}_i^s; \boldsymbol{\theta}_E)). \quad (3.13)$$

3.4.1.6 Initialization and convergence

As for most other complex non-convex optimization problems, local minima and slow convergence also constitute a problem during learning in our framework. While each emission model quickly learns a set of mean vectors \mathbf{v}_k that reasonably approximates the respective dataset, the different parcels are not necessarily aligned across the datasets. This is especially the case when the emission models are randomly and independently initialized. As the arrangement model receives conflicting information from different emission models, it can take a long time to bring the different emission models into

alignment.

To solve this problem, it is sufficient to start the algorithm with a single down-pass of information from the (randomly initialized) arrangement model to all emission models. That is, during the first iteration of the loop, we skipped the calculation of the emission log-likelihood (line 3-9) of the Algorithm 1, setting all $\ell_{i,k}^s$ to zeros. This “pre-training” helps to align the corresponding parcel assignments across all datasets.

A further technique to address the slow convergence is to initialize the model from many different random starting points, and only perform a few learning iterations. After this initial phase of learning, we picked the model with the highest expected log-likelihood, and only completed the training until the likelihood increased less than ($\Delta = 0.01$) in a single step. We used 50 initializations, each trained for an initial 30 steps.

Finally, we repeated this entire process a minimum number of 50 times and then continued until the solution with the highest likelihood was found at least 10 times in independent learning runs. This increased our confidence that we indeed had found a solution that could constitute a global maximum.

3.4.2 fMRI Datasets

In this project, we considered seven task-based and one resting-state fMRI datasets (see Table 3.1). The task-fMRI datasets refer to: (1) the *Multi-Domain Task Battery* (MDTB, King et al. (2019)); (2) a high-resolution version of the MDTB (*High-res MDTB*; not yet published); (3) the *Nakai & Nishimoto* dataset (Nakai and Nishimoto, 2020); (4) a subset of the *Individual Brain Charting (IBC)* dataset (Pinho et al., 2018, 2020); (5) the *Shahshahni* dataset (Shahshahani et al., 2023); (6) the *Multi-Demand* dataset (Assem et al., 2022); and (7) the *Somatotopic* dataset (Saadon-Grosman et al., 2022). The first four datasets of the list include a broad range of task conditions from the perceptual, cognitive, motor, and social domains. In the first three datasets, tasks were randomly intermixed in each imaging session. In the *IBC* dataset, individual runs comprised only one task or a few tasks pertaining to a specific cognitive domain. The three last datasets of the list probe a more circumscribed array of functions: the *Shahshahni* dataset includes verbal working memory tasks (with forward and backward recall) and finger tapping

Name	Subjects	Unique task conditions	Functional scan time (min/subject)	Voxel size (mm)	Description	DOI	Reference
<i>MDTB</i>	24	62	240	3T, 3mm	Cognitive, motor, perceptual, social	10.18112/openneuro.ds002105.v1.1.0	King et al. (2019)
<i>Highres-MDTB</i>	7	9	120	7T, 1.5mm	Cognitive, motor, perceptual, social	N/A	N/A
<i>Nishimoto</i>	6	103	162	3T, 2mm	Cognitive, motor, perceptual, social	10.18112/openneuro.ds002306.v1.0.3	Nakai and Nishimoto (2020)
<i>IBC</i>	12	208	822	3T, 1.5mm	Cognitive, motor, perceptual, social	10.18112/openneuro.ds002685.v1.3.1	Pinho et al. (2018); Pinho et al. (2020)
<i>WMFS</i>	16	17	65	3T, 3mm	Motor and working memory task	N/A	Shahshahani et al. (2023)
<i>Multi-Demand</i>	37	12	100	3T, 2mm	Executive Tasks	N/A	Assem et al. (2022)
<i>Somatotopic</i>	8	6	96	3T, 1.8/2.4mm	Motor	N/A	Saadon-Grosman et al. (2022)
<i>HCP-Unrelated 100</i>	100	none	60	3T, 2mm	Resting-state	https://www.humanconnectome.org/study/hcp-young-adult/data-releases	(Van Essen et al., 2013)

Table 3.1: **FMRI datasets used for the functional fusion.** All datasets but the last one are task-based, all of them together covering a wide range of psychological domains. The last dataset is a subset of the HCP resting-state data.

tasks; the *Multi-Demand* dataset includes three executive function tasks (n-back, task-switch, a no-go); and the *Somatotopic* dataset probes foot, hand, glutes, and tongue movements. Lastly, as a resting-state fMRI dataset, we used the *Unrelated 100* subjects, which made publicly available in the *Human Connectome Project (HCP)* S1200 release (Van Essen et al., 2013).

The task-based datasets were pre-processed using either the *SPM12* software package (Wellcome Department of Imaging Neuroscience, London, UK) or the *FSL* library (Analysis Group, FMRIB, Oxford, UK). For every participant, an anatomical MRI image (T1-weighted MPRAGE, 1mm isotropic resolution) was acquired in one scanning session. FMRI data (time series acquired with Echo-Planar Imaging, T2*-weighted sequence us-

ing Blood-Oxygenation-Level-Dependent contrast) were realigned for head motion within each session, and for different head positions across sessions using the six-parameter rigid body transformation (Friston et al., 1995; Jenkinson et al., 2002). The mean functional image was then co-registered onto the anatomical image and this transformation was applied to all functional images (Ashburner and Friston, 1997; Greve and Fischl, 2009). No smoothing or group normalization was applied.

In parallel, the individual anatomical volumes were segmented into different tissue types (Ashburner and Friston, 2005), and the whole-brain plus gray-matter masks were derived from this segmentation. Each anatomical image was submitted to the standard `recon-all` pipeline from the *FreeSurfer* software (Fischl, 2012) to obtain a reconstruction of the individual cortical surfaces. Similarly, each anatomical image was processed using `SUIT` (Diedrichsen, 2006), which provided cerebellar segmentation and normalization. The cerebellar mask was derived from this segmentation and hand-corrected, whenever necessary, to ensure that voxels from occipital and inferior temporal cortices were not included.

A mass-univariate General Linear Model (GLM) was then fitted to the realigned functional data to estimate brain activation per imaging run. Each task condition was modeled as a boxcar function according to the onsets and duration of the given task condition. The corresponding boxcar function was then convolved with the canonical Hemodynamic Response Function (HRF) (Friston et al., 1998a,b). The whole-brain mask was applied to the realigned functional volumes to restrict the GLM to voxels inside the brain. Coefficients of the GLM were divided by the root-mean-square error (RMSE) for each voxel, resulting in individual volume-based maps of normalized activity estimates. These functional derivatives, obtained for each task condition and imaging run served as input to the fMRI dataset integration framework (see Section 3.4.3).

The resting-state data were pre-processed using the HCP minimal processing pipeline (Glasser et al., 2013), including structural registration, correction for spatial distortion, head motion, cortical surface mapping, and functional artifact removal (Smith et al., 2013; Glasser et al., 2013). For each imaging run, this resulted in 1200 time points of processed time series for each voxel of the standard MNI152 template (Van Essen et al.,

2012) in the cerebellum. To generate the resting-state functional connectivity (rs-FC) fingerprint of the cerebellar voxels from the HCP data set, a group Independent Component Analysis (ICA) was applied. We first concatenated the preprocessed functional data temporally across subjects, sessions, and runs to create a single matrix. Then we used the group-ICA implemented in FSL’s MELODIC (Jenkinson et al., 2012) with automatic dimensionality estimation, resulting in 1072 group-level components. Sixty-nine signal components were identified from the first 300 ICA components as resting-state networks. Lastly, we regressed the 69 group network spatial maps into the subject-and-run-specific cortical time series, resulting in 69 cortical network time courses. The cerebellar rs-FC fingerprints were calculated as Pearson’s correlations of the cerebellar voxel time series with each cortical network time course.

3.4.3 Data structure and anatomical normalization

One important barrier to integrating task contrasts across different fMRI datasets is that these derivative measures are often stored in different atlas spaces (e.g. MNI, fsLR) and with different naming conventions, requiring specialized code for each dataset. To address this problem, we specified a data structure for fMRI derivatives using BIDS-derivative naming convention and file standards (Gorgolewski et al., 2016). For each dataset, we imported the task contrasts (estimates) for each subject, run, and condition that were estimated from minimally pre-processed, non-normalized, and un-smoothed, fMRI data (see Method 3.4.2). We then developed a toolbox that allowed the automatic and fast extraction of this data in any desired atlas space (surface- or volume-based), at any desired level of smoothing and aggregation across runs.

After extraction the resulting functional files are stored using the CIFTI format, resulting in fast and efficient loading times. For the current project, we extracted the cerebellar data in 3mm resolution, aligned to the MNI152NLin2009cSym template (Ciric et al., 2022), resulting in 5446 voxel locations in group space. The sampled functional data of all datasets were smoothed using a Gaussian kernel of 2mm standard deviation, except the *Somatotopic* dataset that used a 3mm smoothing kernel. The proposed file structure and code are available in a public repository (<https://github.com/DiedrichsenLab/>

[Functional_Fusion](#)). The parcellations were visualized using a surface-based representation of the cerebellum ([Diedrichsen and Zotow, 2015](#)).

3.4.4 Synthetic datasets for simulation

To validate the proposed framework, we ran several simulations (Results [3.2.2](#), [3.2.3](#)) on synthetic datasets. To generate individual brain organization maps (\mathbf{U}^s), we used a Markov random field of rectangular 50×50 grid with a 4-neighbor connectivity scheme. Each grid point represented a brain location and could take one of K values (a.k.a Potts Model ([Wu, 1982](#))). We first generated an artificial smooth group probability map (Fig. [A.2a](#)) by selecting K centroids μ_k at random locations, and assigning the bias parameters of the spatial arrangement model $\eta_{i,k}$ for the node a location x_i to be:

$$\eta_{i,k} = -\frac{|x_i - \mu_k|_2^2}{2\sigma_\mu^2} \quad (3.14)$$

where σ_μ^2 controls the smoothness of the group map (see Supplementary Fig. [A.2b](#)).

The individual maps \mathbf{U}^s were then sampled from the Potts model where the local probability $\psi_{i,j}$ between two vertices i and j was set to

$$\psi_{i,j} = \exp(\theta_w \cdot \mathbf{u}_i^\top \mathbf{u}_j \cdot w_{i,j}), \quad \text{where } \mathbf{u}_i^\top \mathbf{u}_j = \begin{cases} 1; & \text{if } \mathbf{u}_i = \mathbf{u}_j \\ 0; & \text{otherwise.} \end{cases} \quad (3.15)$$

The pairwise weight of two vertices $w_{i,j}$ ($w_{i,j} = w_{j,i}$) indicates whether i and j are neighbouring vertices ($w_{i,j} = 1$ if i and j are neighbours; $w_{i,j} = 0$ otherwise). The temperature parameter θ_w controls how strong the spatial co-dependence between neighbouring vertices is. A higher θ_w encourages that the two neighbouring nodes are more likely to be assigned to the same parcel, enforcing the overall local smoothness of the map (Supplementary Fig. [A.2c](#)). Ultimately, the individual maps were generated using vertex-wise Gibbs sampling after a burn-in of 20 iterations across all vertices.

We then generated synthetic functional data \mathbf{Y}^s for each participant based on their individual parcellation maps. Rather than using a von Mises-Fisher distribution, we

wanted to generate data that had both an amplitude and direction. Additionally to the region-specific mean direction of the response v_k , we therefore introduced a non-negative region-specific signal strength parameter, λ_k . The data for each voxel was generated from:

$$\mathbf{y}_p = \lambda_k \mathbf{v}_k + \boldsymbol{\epsilon} \quad (3.16)$$

where $\boldsymbol{\epsilon}$ was a normal random vector with variance $\mathbf{I} \cdot \sigma_k^2$. These parameters allowed us to control the signal and noise levels in each region separately. After normalization of the data to unit length, the generated data conformed approximately to a von Mises-Fisher distribution with mean \mathbf{v}_k and concentration $\kappa_k = \lambda_k^2 / \sigma_k^2$. Ultimately, a synthetic dataset consisting of N task observations was generated for P brain locations and S subjects.

For the simulation in Results 3.2.2 and 3.2.3, the bias terms for the Potts model were generated with $\sigma_\mu^2 = 120$, while the true number of parcels in the group map and fitting model were both set to $K = 20$. Then, we sampled 10 individual maps \mathbf{U}^s from the group map with local connection weights $w_{i,j} = 1.5$. These individual maps are further used to sample synthetic data from two sessions $\mathbf{Y}^{s,1}$ (session 1, $N = 40$ tasks), $\mathbf{Y}^{s,2}$ (session 2, $N = 20$ tasks) and a test set $\mathbf{Y}^{s_{test}}$ ($N = 120$ tasks) with equal signal strength $\lambda_k = 1.1$ for all functional regions. The λ_k might be changed depending on specific simulations (see Results 3.2.2 and 3.2.3).

3.4.5 Evaluation measures for probabilistic atlas

The distance-controlled boundary coefficient (DCBC, Zhi et al. (2022)) is an unbiased evaluation criterion for brain parcellation, which allows the direct comparison of brain maps generated from different modalities (i.e resting-state, task-based, and anatomical) and different number of parcels. The coefficient controls for the intrinsic smoothness of brain data, which is biased in other evaluation metrics (Gordon et al., 2016; Rousseeuw, 1987). The DCBC method solves this problem by binning all vertex pairs based on their spatial distance and only comparing the Pearson’s correlation for within-parcel pairs and between-parcel pairs for the same distance. Then, the DCBC value is calculated as

the average correlation differences, weighted by reliability across distances. The spatial distance is calculated as the Euclidean distance between the center of each voxel pair in the atlas volume space. The underlying functional profiles for calculating the correlations of voxel pairs are the associated betas weights in a task-based dataset. A higher DCBC value of a parcellation indicates that this parcellation predicts the functional boundaries well on the tasks of the dataset being used.

3.4.6 Computational setup

Model training and evaluations were performed on either an NVIDIA 1080Ti GPU with Python 3, CUDA 11.3, and PyTorch 1.10.2 or on NVIDIA GRID A100-10C GPU with Python 3, CUDA 11.6, and PyTorch 1.13.1. For the fMRI datasets, all data were pre-processed and extracted on an Intel i7-8700 CPU with NumPy 1.24.0, NiBabel 4.0.2, neuroimagingtools 0.5.0. Other detailed requirements and parameters used for the data processing pipeline are available in the respective repositories (see [Code availability](#)).

3.5 Data availability

The raw data for the fMRI studies used in this project are publicly available on <https://openneuro.org/> for the studies listed in Table 3.1.

3.6 Code availability

The code for the hierarchical Bayesian parcellation framework is publicly available as the GitHub repository <https://github.com/DiedrichsenLab/HierarchBayesParcel>. The organization, file system, and code for managing the diverse set of datasets is available in a separate repository https://github.com/DiedrichsenLab/Functional_Fusion. The paper-specific code for generating the functional probabilistic parcellations for the cerebellum, as well as running the simulation presented in this paper is available at <https://github.com/DiedrichsenLab/FusionModel>.

Chapter 4

Spatially-informed models for individual brain parcellations

4.1 Introduction

Traditional parcellation approaches primarily rely on group-level analyses (Yeo et al., 2011; Shen et al., 2013; Glasser et al., 2016), which emphasize shared patterns across individuals. These approaches, however, obscure the significant individual variations in functional organization (Eickhoff et al., 2018; Bijsterbosch et al., 2020). These differences, echoing an individual’s unique functional pattern, are critical for understanding the spectrum of individual cognitive functioning. However, the individual neuroimaging data is often insufficient and compounded by inherent measurement noise. These limitations frequently restrict our ability to obtain a reliable and accurate individual parcellation, capturing the unique variability at the individual level. As a result, understanding and exploring individual variability becomes a significant challenge in brain parcellation studies.

Another challenge for individual brain organization study is intrinsic spatial dependence (Laumann et al., 2015; Glasser et al., 2016; Kong et al., 2019). The intrinsic spatial dependence or smoothness of the functional organization arises because nearby brain locations are more likely connected to each other than far-away brain locations. This is due to the fact that short-range neuronal connections are metabolically and developmen-

tally much cheaper than long-range connections, and are therefore much more frequent in the human brain. As a consequence, the neighboring brain locations tend to exhibit a higher functional similarity compared to far-away pairs (Zhi et al., 2022). However, the influence of this intrinsic smoothness is not adequately accounted for in most parcellation methods, which leads to a generally poor and noisy reconstruction of individual parcellations. Therefore, a comprehensive approach to individual brain parcellation should not only account for the unique functional patterns of variability across individuals but also appropriately incorporate the spatially smoothed nature of the functional organization.

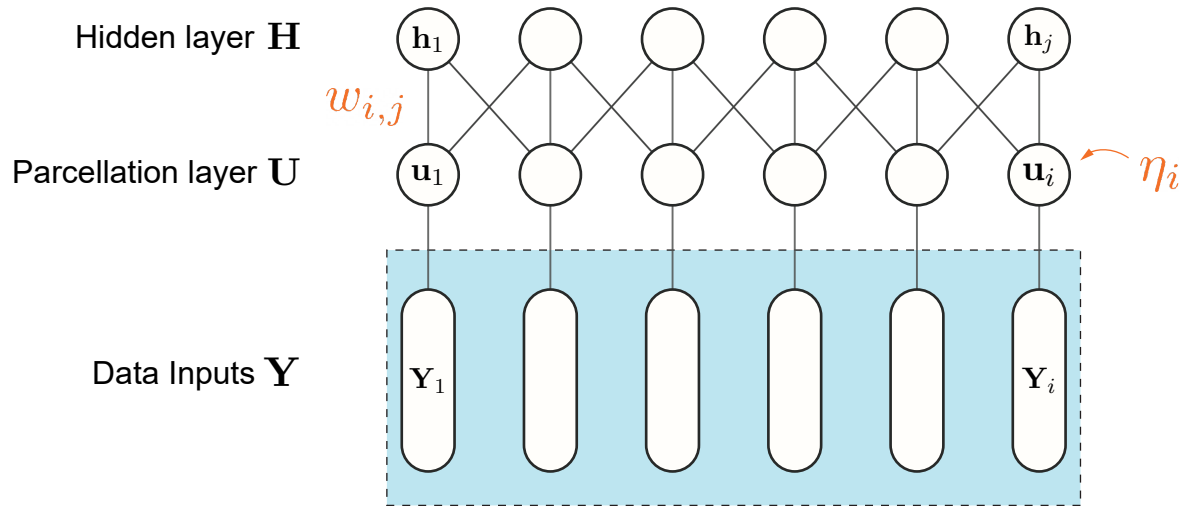


Figure 4.1: **The multinomial-restricted Boltzmann machine (m-RBM) spatial arrangement model.**

In Chapter 3, I have presented a Bayesian framework that results in improved individual parcellations (Zhi et al., 2023). The improvement is achieved by a Bayesian integration of the (noisy and limited) individual data and a spatial group map. Here in Chapter 4, I extend this framework to account for the spatial co-dependency between brain locations. For this, I will replace the independent spatial arrangement model in the hierarchical Bayesian framework with a novel computational architecture: the multinomial-restricted Boltzmann machine (m-RBM) (Fig. 4.1). This structure captures the intrinsic spatial dependence between brain locations. The m-RBM is a multinomial extension of the deep restricted Boltzmann machine (RBM, Methods 4.4.1). Instead of

binary units, the latent nodes used to simulate brain locations (voxels) are designed to be multinomial variables to account for a probabilistic measure. Compared to previous work model spatial dependencies through a Markov Random Field model (Ryali et al., 2013; Schaefer et al., 2018), the m-RBM model has an advantage in computational speed, as we can use layer-wise Gibbs sampling and variational approximation. This advantage arises from the fact that the nodes are conditionally independent within a layer (Discussion 4.3).

Then, we first validated the m-RBM model through an exhaustive simulation and found a significant advantage of the m-RBM model compared to the independent model in estimating individual parcellations with intrinsic spatial dependencies. We also show that the performance of predicted individual parcellations cannot be improved by imposing smoothness directly on the data. Lastly, we tested the ability of the m-RBM model to improve the quality of individual parcellations on the neocortical data from the openly available fMRI dataset.

4.2 Results

4.2.1 m-RBM model captures spatial smoothness better than the independent model when estimating individual parcellations

To demonstrate the characteristics of the model, we first sampled brain organization maps from an independent arrangement model and from the new m-RBM (Fig. 4.2a). Both used a smooth group probability map with 5 regions. The samples from the independent arrangement model were - despite the smooth group map - spatially highly fragmented (Fig. 4.2a, *Independent*). In contrast, samples from the m-RBM model showed contiguous regions (Fig. 4.2a, *m-RBM*). Note that in this simulation, the marginal probability between the two models was matched (Fig. 4.2a, *Marginal probabilities*). In the next step, we sought to determine whether the m-RBM model outperforms the independent model in improving the performance of individual parcellations when the data is generated from

the spatially dependent (m-RBM) model.

To this end, we performed a simulation on the synthetic training and test dataset sampled from the same set of subjects by a modified restricted Boltzmann Machine (RBM) (Methods 4.4.5). We first trained the learning framework using the independent model (Zhi et al., 2023) with the ground truth emission model on the synthetic dataset (Methods 4.4.5). During the learning, we froze the emission model to only provide the true data log-likelihood for each training iteration to avoid introducing unnecessary uncertainty. We then calculated the individual parcellations from the proposed m-RBM model with true parameters using the same routine. Lastly, we evaluated the performance of reconstructed individual parcellations estimated by different models on an independent test set for comparison. The evaluation metric for this simulation includes DCBC (Methods 4.4.6.1), the mean prediction error between the true and reconstructed parcellations $\bar{\epsilon}_{|\mathbf{U}-\hat{\mathbf{U}}|}$ (Methods 4.4.6.2), and mean adjusted expected cosine error $\bar{\epsilon}_{Acosine}$ (Methods 4.4.6.3). We repeated this simulation 100 times.

The simulation first confirmed our results in Chapter 3 (Results 3.2.1): the predicted individual parcellations estimated by the independent model are substantially better than the ones based on data only. This can be seen in a visual inspection of a reconstructed individual map derived from the independent model (Fig. 4.2b, *Independent + unsmoothed data*) compared to using data only (Fig. 4.2b, *Data only*), where the former map shows more clear reconstruction to the true map (Fig. 4.2b, *True*). The evaluation of the mean expected prediction error (Fig. 4.2c) for the individual parcellations also shows a significantly lower value than the “data only” parcellation when compared to the true map ($t_{99} = -1734.161, p = 1.039 \times 10^{-223}$).

Importantly, however, the parcellations are further improved using the m-RBM model. When visually comparing a reconstructed parcellation derived from the m-RBM model (Fig. 4.2b, *m-RBM + unsmoothed data*) to the one derived from the independent model (Fig. 4.2b, *Independent + unsmoothed data*), we can see not only a better estimation to the true map, but its associated probability map also shows less uncertainty between the boundaries. Quantitatively (Fig. 4.2c), the estimated individual parcellations from the m-RBM model show significantly lower prediction error compared to the ones from

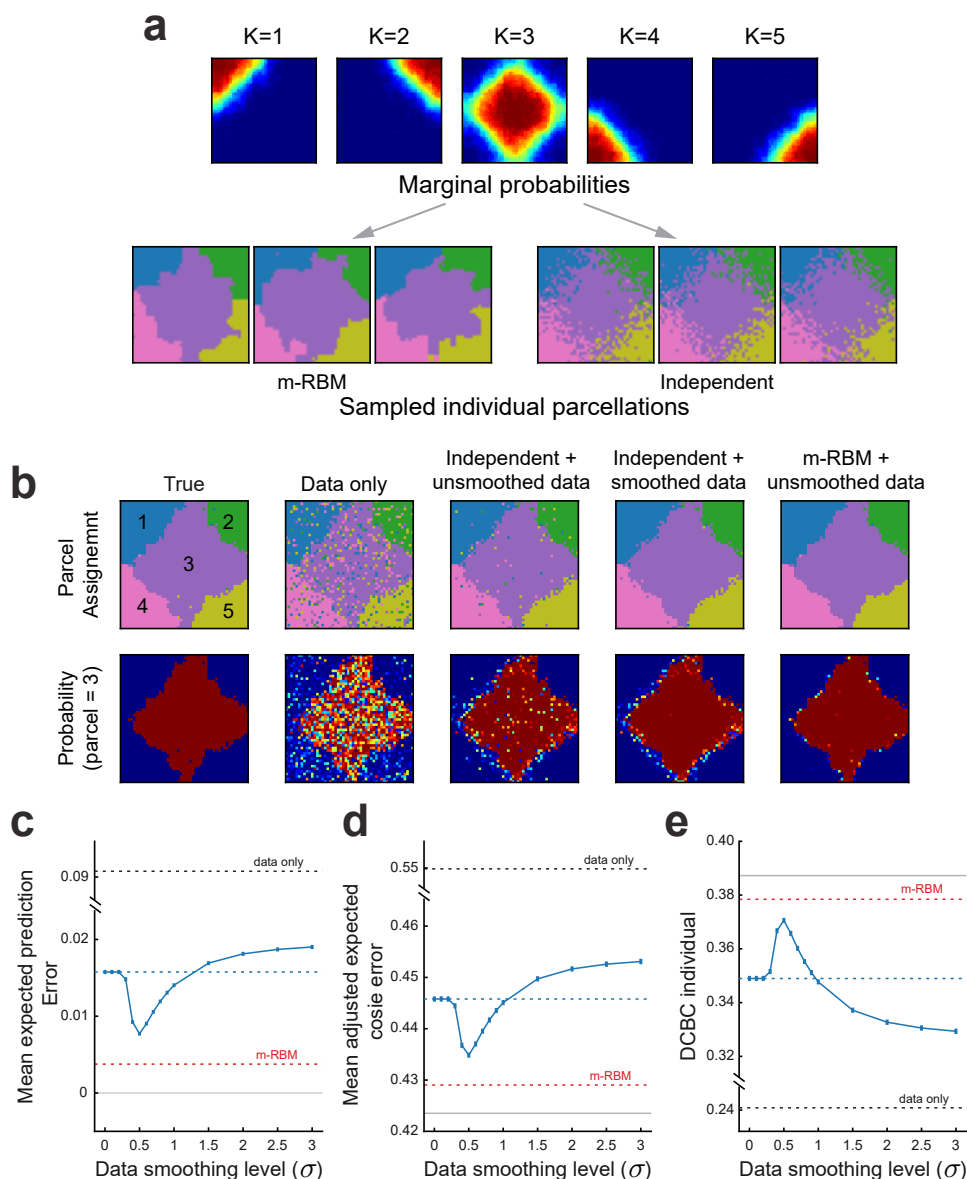


Figure 4.2: **Comparison of the m-RBM and the independent arrangement model on simulated data.** (a) Sampled individual parcellations from both m-RBM and independent model with the same marginal probability or parcel assignment. (b) The comparison of reconstructed individual parcellation (top row) and probability map of parcel 3 (bottom row) to the true map when it learned from data only, independent model with or without the smoothed data, and m-RBM model. (c) The mean absolute reconstruction error (lower=better), (d) the mean expected weighted cosine error (lower=better), and (e) the DCBC (higher=better) of the individual parcellation derived from the independent model (blue dotted line) and m-RBM model (red dotted line) with unsmoothed data, or by the independent model with smoothed data (blue solid line). As reference the performance of the data only (black dotted) and the noise floor (gray solid line) are shown. The error bar indicates the SEM across 100 simulations.

the independent model ($t_{99} = -316.112, p = 1.519 \times 10^{-150}$). This advantage over the independent model is also confirmed by the other two evaluation methods (Fig. 4.2d,e), the mean adjusted expected cosine error ($t_{99} = -298.376, p = 4.589 \times 10^{-148}$) and the DCBC ($t_{99} = 220.495, p = 4.439 \times 10^{-135}$), respectively.

Data smoothing is a commonly-used strategy to deal with the problem of lower quality of individual parcellation by enhancing the signal-to-noise ratio of the data to reveal the underlying patterns. Therefore, we trained the independent model on different levels of smoothed synthetic data (Methods 4.4.5), and then estimated the individual parcellations using the same data. While the reconstruction accuracy of the individual maps was improved by using smoothed data, the peak performance was still lower than the one estimated by the m-RBM model using un-smoothed data. In simulation results (Fig. 4.2c), we found that a light smoothing of data improved the performance of individual parcellations derived from the independent model and reached its best performance at around $\sigma = 0.5$, which was significantly better than using un-smoothed data ($\langle \bar{\epsilon}_{|\mathbf{U}-\hat{\mathbf{U}}|} \rangle_q$: $t_{99} = -255.688, p = 1.956 \times 10^{-141}$). However, the estimated individual parcellations from the m-RBM model with un-smoothed data still substantially outperformed this best performance of the independent model ($\langle \bar{\epsilon}_{|\mathbf{U}-\hat{\mathbf{U}}|} \rangle_q$: $t_{99} = -163.642, p = 2.708 \times 10^{-122}$). A similar pattern can be observed in both mean adjusted expected cosine error ($\bar{\epsilon}_{Acosine}$: $t_{99} = -171.720, p = 2.335 \times 10^{-124}$) and DCBC ($t_{99} = -96.923, p = 6.422 \times 10^{-100}$) evaluations (Fig. 4.2d,e).

When we increase the smoothing level of the data, the reconstructed accuracy of the parcellations by the independent model continuously decreased and becomes even worse than the ones using un-smoothed data at around $\sigma = 1.5$ ($\langle \bar{\epsilon}_{|\mathbf{U}-\hat{\mathbf{U}}|} \rangle_q$: $t_{99} = 31.583, p = 1.688 \times 10^{-53}$; $\bar{\epsilon}_{Acosine}$: $t_{99} = 62.525, p = 2.216 \times 10^{-81}$; DCBC: $t_{99} = -64.926, p = 5.796 \times 10^{-83}$) (Fig. 4.2c,d,e). Overall, this simulation shows a systematic advantage of the m-RBM model over the independent arrangement model in predicting individual parcellations when the data has an intrinsic spatial structure. A moderate smoothing of data can enhance the performance of maps derived from the independent model, but the outcomes are still substantially behind those achieved using the m-RBM model. Moreover, overly-smoothed data can lead to opposite results, producing parcellations

that are even worse than those derived from un-smoothed data.

4.2.2 Convergence problems when training the m-RBM model

In the previous simulation, we fixed the parameters of the m-RBM model to the ground truth to demonstrate that it can outperform the independent model in the best-case scenario. However, in the real fMRI data, we do not have the luxury of knowing the ground true model parameters. Thus, the convergence of model training has to be validated through the simulation. Here, we relaxed the two parameters of the m-RBM model, $\theta_{\text{m-RBM}} := \{\eta_k, \theta_w\}$, can be freely estimated during each learning epoch, and compared the estimated parameters whether converge to the ground truth. To this end, we repeated the simulation 10 times with the parameters η_k and θ_w learned from random initialization, where η_k were initialized from the normal distribution with a zero mean and θ_w was uniformly generated between 0 to 5, respectively.

Surprisingly, the simulation shows that the 10 learning epochs converged to very different values (Fig. 4.3a,b). Additionally the mean $\hat{\theta}_w$ s overestimates the the true θ_w . However, when we monitored the model marginals, the estimated marginal probabilities converged to the true value across all learning epochs, indicating the m-RBM model has successfully learned the underlying data distribution. This can be validated by calculating their L2 norms during the learning, which converged to a place close to zero with an expected bias (Fig. 4.3c). The irreducible bias is caused by the noise in the data observations since the inputs to the model are the synthetic functional data sampled from emission models instead of the true individual parcellations. We then compared the estimated marginals to the ones learned by the independent model using the same data (Fig. 4.3c, red dotted line), and found that the marginals converge to the same place.

Therefore, this learning behavior in the simulation revealed a convergence issue of the m-RBM model training. This is a common issue when training complex deep neural networks based on contrastive divergence (Sutskever and Tieleman, 2010) and can be illustrated by a flat curve of the objective function (Fig. 4.3d, solid line). The learning process will then converge to multiple combinations of parameters θ_w and η , each corresponding to a different local minimum of the error surface (Fig. 4.3d, red dots).

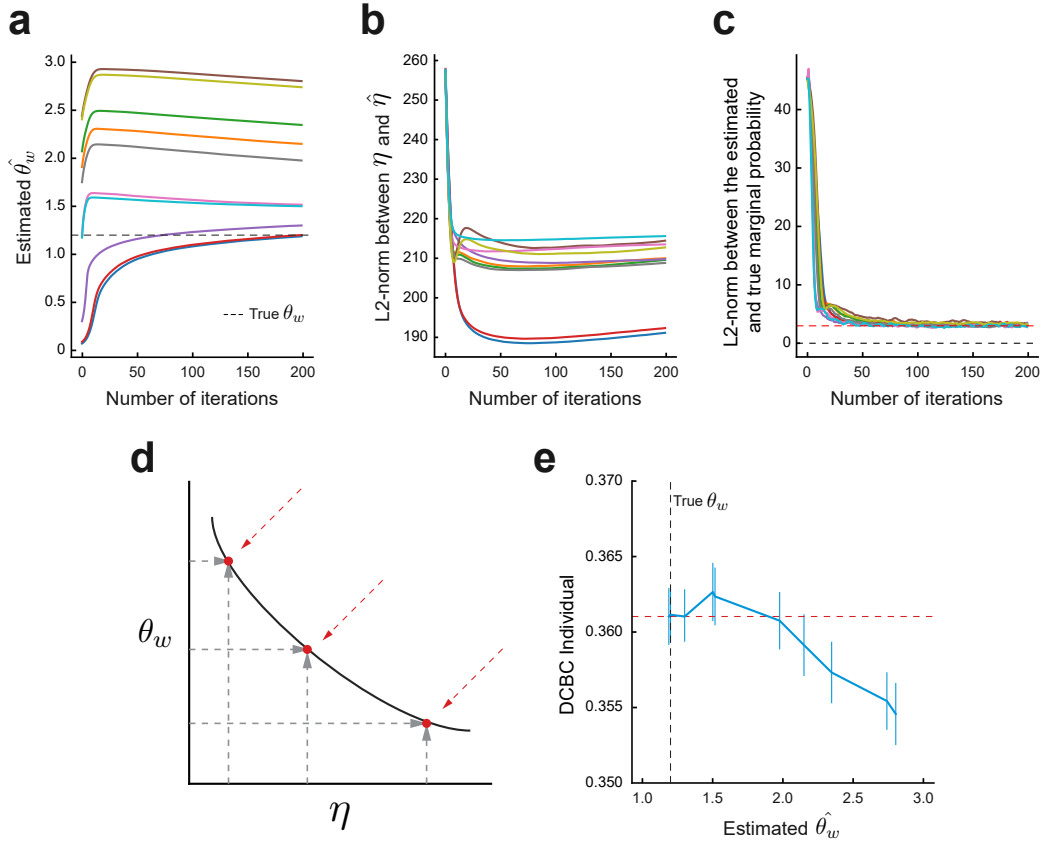


Figure 4.3: **Simulation of the convergence issue when training m-RBM model.** (a) The estimated model parameter $\hat{\theta}_w$ for all 10 simulations learned from random initializations compared to the ground truth $\theta_w = 1.2$ (black dotted line). (b) The L2 norm between the estimated group probability parameter $\hat{\eta}$ for all 10 simulations learned from random initializations to the true η . (c) The L2 norm between the estimated marginal probabilities of the 10 simulations to the true marginal probability. (d) The illustration of the convergence issue, where each estimated $\hat{\theta}_w$ has a corresponding η_k on the learning curve. (e) The DCBC evaluation of the estimated individual parcellations from all 10 simulations with different estimated $\hat{\theta}_w$ on an independent test set. The true $\theta_w = 1.2$ (black dotted line). The red dotted line is the averaged DCBC value of individual parcellations estimated from the m-RBM with the ground-truth parameters. The error bars indicate the standard error of the mean (SEM) across 30 individuals.

We then investigated to what extent this model convergence issue influences the performance of the estimated individual parcellations, especially the overestimation of the overall connection strength θ_w . To this end, we calculated the DCBC values of the estimated individual parcellations across all 10 simulations with “freely learned” parameters to the ones learned from the model with fixed true parameters on the independent test

set (Methods 4.4.5). The result (Fig. 4.3e) suggested that there are no significant differences in performance between the two if the estimated $\hat{\theta}_w$ is close to the true one ($t_{29} = 0.209, p = 0.836$, in case of $\hat{\theta}_w = 1.3$). In contrast, this performance of individual parcellations experienced a consistent decrease with an overestimated $\hat{\theta}_w$ and became significant starting around $\hat{\theta}_w = 2.15$ ($t_{29} = 2.697, p = 1.154 \times 10^{-2}$) and further dropped with larger estimation. Together in this simulation, the training of the m-RBM model is subject to a convergence issue. Therefore, we decided to fix the parameter θ_w by a pre-defined range of linear search to make the training of the m-RBM model successful on the empirical fMRI data (Results 4.2.3).

4.2.3 Smoothing does not improve the estimation of individual functional boundaries on real fMRI data

With the validation in the previous simulation, we sought to estimate individual parcellations using the m-RBM model on the real fMRI data to determine whether we can get improvement compared to the ones derived from the independent model. Here we used the MDTB (King et al., 2019) task-based fMRI dataset (Methods 4.4.4), which contains two independent task sets, and deployed the hierarchical Bayesian framework (Zhi et al., 2023). We first obtained the individual parcellations on the first task test using the m-RBM model by fixing the parameter θ_w within a range of 0.1 to 4 to find the best training performance (Methods 4.4.3). Then, we estimated the individual parcellations of the same task set using the independent model on the un-smoothed or smoothed data with different smoothing levels (Methods 4.4.4). Lastly, we tested the performance of these estimated individual parcellations in predicting the functional boundaries of the same subjects in the second independent task set of the MDTB dataset by calculating its DCBC evaluation metric. We then reversed the role of the two task sets and averaged performance across the two cross-validation folds.

We first confirmed that the performance of individual parcellations derived from the independent model using the real fMRI data (MDTB dataset) constantly decreased by increasing the data smoothing level (Fig. 4.4b, *blue line*). To statistically evaluate the

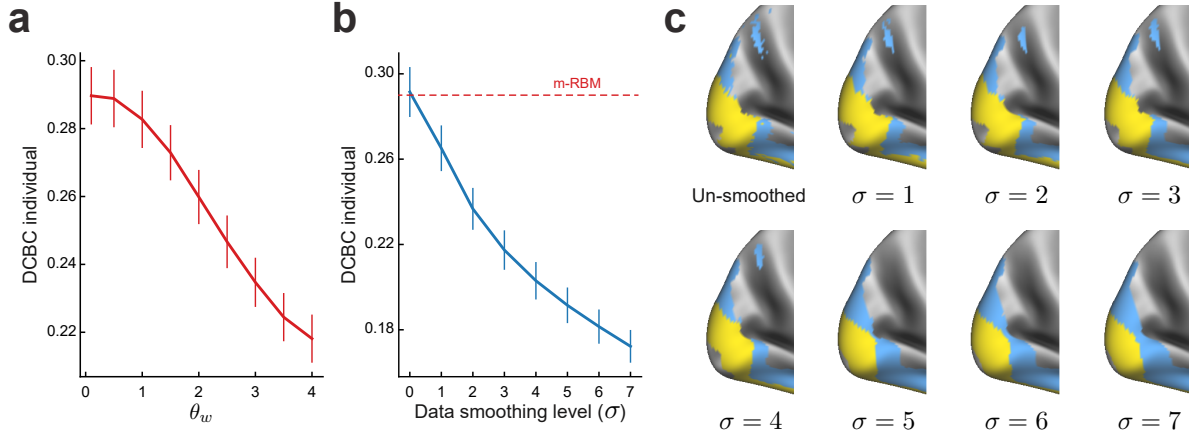


Figure 4.4: **Comparison of individual parcellations estimated by the m-RBM vs. independent model using MDTB dataset.** (a) The DCBC evaluation of individual parcellations derived from the m-RBM model using different temperature parameters θ_w from 0.1 to 4. (b) The DCBC evaluation of individual parcellations derived from the independent model using data with different levels of smoothing, ranging from 0 to $\sigma = 7$. (c) The visualization of the estimated individual parcellations where two neighbouring functional regions are colored. The error bars in (a) and (b) indicate the SEM of DCBC across the 24 evaluation subjects.

decreasing parcellations performance with respect to the smooth levels, we conducted a repeated-measures analysis of variance (ANOVA) across the 8 smoothing levels, ranging from un-smoothed $\sigma = 0$ to $\sigma = 7$ (Methods 4.4.4). Similar to what we observed in the simulation, the DCBC values of the 24 individual parcellations show a significant decrease for the larger data smoothing levels ($F_{161}^7 = 447.55, p = 4.861 \times 10^{-102}$). A visual inspection of the estimated functional boundaries between two example regions (Fig. 4.4c, *blue and yellow parcels*) also shows that the boundaries tend to get smoothed and shifted with larger data smoothing level. Although this experiment did not observe a performance boost due to light smoothing, as has been shown in the previous simulation, we sought a possible explanation that could relate to pre-processing steps for the surface-based cortical data. These steps have already applied a significant amount of smoothing for the MDTB dataset. As a result, the processed un-smoothed data has likely reached or passed its peak performance. This could explain why additional light smoothing does not enhance the performance further (Discussion 4.3).

In contrast, we sought to determine whether the advantage of the m-RBM model

shown in the simulation can be replicated when training individual parcellations using real data. To this end, we tended to find the best m-RBM training model with the highest performance of their estimated individual parcellation. Since the model training is subject to a convergence issue (Result 4.2.2), we fixed the parameter θ_w to a set of values during the training. Within the range of $\theta_w = 0.1$ to 4, we found that the performance of individual parcellations derived from the m-RBM model continuously decreased with a larger temperature parameter θ_w (Fig. 4.4a). The peak performance of the estimated individual parcellations is at the minimum value $\theta_w = 0.1$ with a mean DCBC of 0.290 (SEM=0.008). However, when comparing this peak performance of individual parcellations from the m-RBM model (Fig. 4.4b, red dotted line) to the ones estimated by the independent model using un-smoothed data (Fig. 4.4b, blue error-bar at $\sigma = 0$), we found the performance of individual parcellations were not improved ($t_{23} = -5.065, p = 3.972 \times 10^{-5}$). Therefore, we did not observe the expected advantage of the m-RBM model when training individual parcellations over the independent model, using the pre-processed MDTB cortical data.

4.3 Discussion

In this project, we developed a novel computational architecture, namely the m-RBM arrangement model, based on a restricted Boltzmann machine to estimate individual brain functional parcellations. Importantly, this architecture models the spatial co-dependence between neighboring brain locations by between-layer connection, rather than imposing smoothness directly to the functional data. We confirmed the advantage of using the m-RBM model for learning accurate individual parcellations compared to the previous model that assumes brain locations are spatially independent through a simulation. Our experiment results on both simulation and empirical data also suggested that the performance of individual parcellations cannot be improved by learning largely-smoothed data.

Studying intrinsic spatial dependence in the neocortex using surface representation. Intrinsic spatial dependence The intrinsic spatial dependence between

brain locations reflects the biological nature of the brain, primarily shaped during neural connectivity development. Throughout this process, spatially nearby neurons typically establish more robust communication pathways with each other, contributing to a higher degree of functional similarity compared to far-away neuron pairs. However, this phenomenon can manifest differently when examining the brain using fMRI imaging. For instance, voxel pairs that are spatially close to other but separated by structures could show a low functional correlation, such as one voxel in the neocortex and another one belonging to the cerebellum. Therefore, in the exploration of the brain's intrinsic smoothness, it's important to take into account different brain structures.

Among all brain structures, the neocortex is the largest part of the human cerebral cortex with billions of neurons, handling most of the functions in the brain. When we look at the cortical anatomy, the gray matters (neurons) live in a highly folded surface structure. Notably, compared to the gray matter in the cerebellum forms thin and tightly folded layers, the cortical neurons are organized into a thick and convoluted sheet between the pial and white surface. By unfolding this cortical sheet, we can map the neurons of the cortex to a surface, resulting in surface-based representations that have been developed to benefit the subsequent analyses (e.g. fsaverage cortical mesh, [Fischl \(2012\)](#)). Compared to the volumetric image, the surface-based representation measures the spatial dependencies of cortical neurons more precisely. For example, two spatially nearby voxels between cortical sulci are expected to show higher functional connectivity. But they could exhibit lower similarity because, in fact, they have relatively large distances along the cortical sheet. Together, these form our motivation and the importance of studying the spatial dependence of the neocortex using an unfolded and smoothed surface-based representation.

Considering spatial smoothness for estimating individual parcellations. By considering spatial dependencies, we can identify spatially contiguous functional regions, leading to a more meaningful brain subdivision ([Schaefer et al., 2018](#)). This is especially important for estimating individual-specific networks due to individual variations. Previous studies have reached a consensus that a practically-useful brain parcellation should be spatially contiguous ([Kaas, 1987](#); [Amunts and Zilles, 2015](#)), while recent works ([Yeo](#)

et al., 2011; King et al., 2019; Zhi et al., 2022) also show that neighboring brain regions often share similar functions or contribute to the same cognitive processes. But if the estimated individual parcellations only emphasize functionally similar regions without showing intrinsic smoothness, it may lead to fragmented and discontinuous parcels that undermine the interpretability and overlook the spatial continuity of neuronal circuits. Hence, the individual variability in brain functional organization, which is a key feature of the human brain, might not be properly captured. Therefore, when estimating individual parcellations, it is important to take into account the spatial proximity of regions to capture their functional relationships accurately.

To this end, a computational architecture to estimate individual parcellations that incorporates the spatial dependence of the brain locations helps for accounting for inter-individual variability (Bijsterbosch et al., 2018; Cachia et al., 2018; Kong et al., 2019). Unlike the group parcellation is able to inherit the spatial smoothness across the studies population, the individual parcellation methods usually suffer from limited individual data to reconstruct an accurate functional localizer (Marek et al., 2018; Zhi et al., 2023). Some pioneering works (Kong et al., 2021; Zhi et al., 2023) improved individual parcellation by imposing spatial arrangement information as a prior within a hierarchical Bayesian framework, resulting in better performance and behavioral prediction as an individual functional localizer. However, these imposed spatial priors are usually pre-computed (Schaefer et al., 2018) or learned across subjects by assuming the brain locations are spatially independent (Zhi et al., 2023), prohibiting adaptive learning of the spatial arrangement topography to benefit the individual parcellation improvements. Therefore, a novel computational model that simultaneously addresses intrinsic spatial dependence and individual variability demands further development.

Computational efficiency of the m-RBM model. Our proposed m-RBM model was developed from an RBM architecture so that it inherits the advantage of RBMs in computational speed. This advantage is pronounced when compared to other computational architectures in brain parcellation tasks, such as Markov Random Field (MRF) Ryali et al. (2013); Schaefer et al. (2018). Although we did not run a direct speed comparison in this project, the m-RBM indeed takes much less time than the MRF in training

the same dataset from a theoretical point of view. Traditional RBMs are energy-based models with a bipartite structure (Hinton et al., 2006), comprising a visible layer and a hidden layer, where no intra-layer connections are allowed. In the m-RBM model, these two layers have all become hidden layers without intra-layer relations. The intrinsic spatial dependencies are incorporated through the connectivity weights and modeling using the latent variables. This architecture enables efficient layer-wise Gibbs sampling, where all nodes within a layer can be updated simultaneously. This makes sampling in RBMs highly parallelizable, which can considerably speed up the computation when leveraging modern hardware architectures, such as GPUs.

In contrast, MRF (Kindermann and Snell, 1980) is a single-layer architecture that integrates local spatial structures by modeling the dependency between neighboring nodes, aligning with the inherent characteristics of brain organization. But, it often necessitates a more computationally intensive node-wise sampling approach. In MRFs, each node is directly influenced by its neighboring nodes, requiring the consideration of the overall network state when updating each node and preventing parallelization. Consequently, MRFs typically involve a higher computational cost and longer convergence times. This makes RBMs more suited for large-scale and high-dimensional problems, such as brain parcellation tasks, where computational efficiency and scalability are of utmost importance.

The m-RBM model does not improve the individual parcellations using MDTB dataset. In the simulation (Results 4.2.1), we showed that a “lightweight” data smoothing helps in improving the performance of the estimated individual parcellations. This is a simple and commonly-used approach to increase the signal-to-noise level in the limited amount of individual data, especially useful when the model for training parcellations is not equipped with spatial dependence. However, we think this strategy is only effective for higher-resolution fMRI imaging data, and less pronounced or even invalid for a lower data resolution. For example, the synthetic data in the simulation was generated from a Markov random field of 50×50 grid, in which each point on the grid simulates a most basic brain location unit (or even can be treated as a single neuron). Therefore, applying a small level of smoothness to the sampled synthetic functional data

indeed improves the signal-to-noise from the sampling noise. However, this does not hold in the real fMRI data, where the recorded signals are already averaged within each of the voxels containing many neurons. For example in our case, the MDTB data were acquired on a 3T scanner with $3 \times 3 \times 3 \text{ mm}^3$ in-plane resolution (voxel size). This assumption can be proved by comparing the results of the real data experiments (Results 4.2.3) with simulation (Results 4.2.1), where we did not observe a performance increase but a continuous decrease with a larger smoothing level on the MDTB data (Fig. 4.4b).

A relatively coarse fMRI resolution not only makes the efforts of data smoothing trivial, but also diminishes the performance of the individual parcellations derived from the m-RBM model. This can be seen in the performance comparison of the individual parcellations, where the maps derived from the m-RBM model have significantly lower performance than the ones by the independent model using un-smoothed data (Fig. 4.4b). It is because applying such a model equipped with spatial dependence structures in the learning is equivalent to adding another layer of smoothness to the data, making the performance of the estimated individual parcellations decrease. Therefore, when estimating individual parcellations from the fMRI data of a 3 mm^2 voxel size level, an independent model that assumes brain locations are spatially independent might be the optimal choice. Another potential reason that causes this unexpected m-RBM model performance is the excessive amount of training data (160 mins imaging scans per subject), making the effects of spatial smoothness less pronounced. Further investigations are needed to determine which factor is more significant.

Limitations and further development. In this paper, we have developed an m-RBM model to estimate individual brain parcellations, which is essential to understand the spatial dependence between brain locations. We demonstrated several advantages and the importance of using such a model to deal with intrinsic smoothness and individual variability in individualized brain organization studies. However, like any energy-based model, m-RBM suffers from the intractable partition function problem, which prohibits the obtaining of true posterior calculation. In this project, we applied a persistent contrastive divergence (p-CD) (Tieleman, 2008) algorithm to approximate the posterior expectation, which relies on mean-field approximation in the positive phase.

The mean-field approximation for energy-based model (Welling and Hinton, 2002) assumes each node is treated as being influenced by an “average” of the other variables, rather than considering the specific influence of each node, resulting in the proposal distribution q is factorized (eq. 4.1). However, this approach can lead to an oversimplification and inaccurate approximation of the probability of the whole model given the data. With the recent development of learning algorithms that more accurately approximate the partition function, such as Score Matching (Hyvärinen and Dayan, 2005) and Noise Contrastive Estimation (Gutmann and Hyvärinen, 2010), we seek an improvement in the training of the m-RBM model.

4.4 Methods

In this work, we used the hierarchical Bayesian parcellation framework to learn individual brain organization (Chapter 3, Zhi et al. (2023)). The framework and the emission model remain the same as presented in Chapter 3 (Methods 3.4.1.1, 3.4.1.2), but instead of using an independent spatial arrangement model (Methods 3.4.1.3), we use a newly designed arrangement model which models the spatial dependency between brain locations, called m-RBM model. Due to the intractability of the m-RBM model, we used the stochastic maximum likelihood instead of EM to train the model 4.4.2.

Another difference in the implementation here is that we only study individual parcellations using a single fMRI dataset rather than a number of different datasets for simplicity. It is because the main goal of this project is to compare the performance of different spatial arrangement models. Other mathematical notations remain the same as we explained in Chapter 3 (Methods 3.4) for consistency and would not repeat in the following sections unless new notations are introduced.

4.4.1 The m-RBM spatial arrangement model

As a spatial arrangement model, the m-RBM (multinomial-restricted Boltzmann machine) model provides a group probability measure $p(\mathbf{U}; \boldsymbol{\theta}_A)$ for each unique individual s over a set of latent variables \mathbf{u}_i^s (Methods 3.4.1.3). These latent variables \mathbf{u}_i^s quantify how

likely a brain location i is assigned to a functional region k . Compared to the independent arrangement model that assumes different brain locations are spatially independent, the m-RBM model imposes a spatial dependence structure on the group probability measure. This spatial structure was modeled through a modified three-layer restricted Boltzmann machine (RBM) (Fig. 4.1). In contrast to traditional RBMs, the hidden units in the m-RBM model are now multinomial variables rather than binary states from the Bernoulli distribution.

In contrast to the binary units in traditional RBMs, the latent variables \mathbf{u}_i^s for each node in the parcellation layer is now a multinomial vector with a length of K , which is the reason we named this model “m-RBM”. We use \mathbf{U} to represent a collection of these one-hot encoded latent variables into a $K \times P$ matrix. In this layer, the latent nodes are also associated with a group probability parameter $\eta_{i,k}$ (bias term) for each location i and parcel assignment k . From a traditional two-layer RBMs viewpoint, this parcellation layer \mathbf{U} should observe the training data (visible layer), however in the m-RBM model, the data is now modeled as the data layer \mathbf{Y} (Fig. 4.1 shaded area). Then, the data layer is connected to the parcellation layer by one-to-one mapping, which calculates the location-specific data likelihood and passed them to the corresponding brain locations in \mathbf{U} .

The hidden layer \mathbf{H} consists of J multinomial latent nodes h_j , which are connected to the parcellation layer \mathbf{U} with connectivity weights $w_{i,j}$ (Fig. 4.1). The pairwise connectivity weights $w_{i,j}$ ($w_{i,j} = w_{j,i}$) are used to model the spatial dependencies between neighbouring brain locations i and j ($w_{i,j} = 1$ if i and j are neighbours; $w_{i,j} = 0$ otherwise). In practice, we also introduced a non-negative temperature parameter θ_w to control the strength of the pairwise connection. Hence, with J hidden nodes, \mathbf{H} is a $K \times J$ matrix with one-hot encoded states of the hidden variables. The connectivity weights $w_{i,j}$ were then collected into a $J \times P$ matrix \mathbf{W} .

4.4.2 Training m-RBM model using stochastic maximum likelihood

This model architecture permits inference using variational methods and stochastic maximum likelihood in the model training (Salakhutdinov and Hinton, 2012). For updating parameters, we applied contrastive divergence (CD) (Hinton et al., 2006) algorithm to calculate the gradients of the un-normalized expected arrangement model likelihood (Methods 4.4.2) with respect to each model parameter. Therefore, in the positive phase (see Methods 4.4.2.1), the data distribution q is now approximated using a mean-field approximation:

$$q(\mathbf{U}^s, \mathbf{H}^s | \mathbf{Y}^s) \approx \prod_i p(\mathbf{u}_i^s | \mathbf{Y}^s) \prod_j p(\mathbf{h}_j^s | \mathbf{U}^s) \quad (4.1)$$

In the negative learning phase, we alternate sampling from the distributions $p(\mathbf{Y}^s | \mathbf{U}^s)$, $p(\mathbf{H}^s | \mathbf{U}^s)$, and $p(\mathbf{U}^s | \mathbf{H}^s, \mathbf{Y}^s)$ (Methods 4.4.2.2). The m-RBM model parameters $\boldsymbol{\theta}_A := \{\eta, \theta_w\}$ are then updated in the parameters update phase (Methods 4.4.2.3).

Since the m-RBM model cannot provide a close-formed posterior expectation, the previous integrated EM algorithm for training the model is not appropriate. Here, we applied a stochastic maximum likelihood algorithm with mini-batch learning (Algorithm 2) to replace the line 10 to 19 in Algorithm 1 for training the m-RBM model in the message-passing and collaborative learning process (Methods 3.4.1.4), so that the E- and M-steps for the emission model remain unchanged. The term n that represents different experiments or sessions will be dropped in the algorithm pseudocode for simplicity since we used a single dataset in this project. Here, we define $\text{softmax}(\cdot)$ as the activation function for each multinomial node, where the `softmax` operation will be applied along the dimension of nodes vector if it takes matrix input.

The stochastic process takes the input of subject-specific data log-likelihoods of all subjects in \mathbb{S} . After the full learning, the approximated posterior expectation $\langle \mathbf{u}_i^s(k) \rangle_q$ (or $\langle \mathbf{U}^s \rangle_+$ in this case) for each subject was calculated from the model, the unnormalized

Algorithm 2: The stochastic maximum likelihood algorithm as the optimization procedure for training the m-RBM model. This is to replace lines 10-19 in Algorithm 1.

Input: a collection of S subjects' data likelihood $\mathbb{L} = \{\ell_{i,k}^{1(t)}, \dots, \ell_{i,k}^{s(t)}\}, s \in \mathbb{S}, S$ Gibbs chains for negative phase $\tilde{\mathbf{U}}^{s(0)}$

Output: $\langle \mathbf{U}^s \rangle_+^{(t)}$, a matrix of estimated posterior values of $\langle \mathbf{u}_i^s(k) \rangle_q^{(t)}$; $\tilde{\mathbf{U}}^{s(n)}$, the updated Gibbs chains from negative phase

- 1 **Set** learning rate α to a small positive value. Mini-batch size B , $m = 5$ times of mean-field approximation in the positive phase, $n = 10$ of Gibbs steps in the negative phase.
- 2 **Initialize** counter=1, two empty $S \times K \times P$ tensors $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{H}}$ for recording samples in the negative phase, $\langle \mathbf{H} \rangle_+^{(0)} = 0$ for positive phase, if no persistent sampling chains for the negative phase are given, sample $\mathbb{S} \tilde{\mathbf{U}}^{(0)}$ from $\text{softmax}(\eta^{(t)})$
- 3 **while** counter $\leq \lceil S/B \rceil$ **do**
- 4 Randomly sample a minibatch of \mathbb{B} from \mathbb{L} without replicates. Remove \mathbb{B} from \mathbb{L} .
- 5 (1) positive phase for m times mean-field approximation by eq.4.4 and 4.5:
- 6 **for** $a = 0$ to $m - 1$ **do**
- 7 $\forall s \in \mathbb{B}, \langle \mathbf{U}^s \rangle_+^{(a)} = \text{softmax}(\ell^{s(t)} + \eta^{(t)} + \langle \mathbf{H}^s \rangle_+^{(a)} \mathbf{W} \cdot \theta_w^{(t)})$
- 8 $\forall s \in \mathbb{B}, \langle \mathbf{H}^s \rangle_+^{(a+1)} = \text{softmax}(\langle \mathbf{U}^s \rangle_+^{(a)} \mathbf{W}^\top \cdot \theta_w^{(t)})$
- 9 **end**
- 10 $\langle \mathbf{U}^s \rangle_+^{(t)} \leftarrow \langle \mathbf{U}^s \rangle_+^{(m-1)} (s \in \mathbb{B}), \quad \langle \mathbf{H}^s \rangle_+^{(t)} \leftarrow \langle \mathbf{H}^s \rangle_+^{(m)} (s \in \mathbb{B})$
- 11 (2) negative phase for n Gibbs mixing steps using eq.4.7 and 4.6:
- 12 **for** $b = 0$ to $n - 1$ **do**
- 13 $\forall s \in \mathbb{B}, \tilde{\mathbf{Y}}^{s(b)}$ sampled from $p(\mathbf{Y}^s | \tilde{\mathbf{U}}^{s(b)}; \boldsymbol{\theta}_E^{(t)})$
- 14 $\forall s \in \mathbb{B}, \tilde{\mathbf{H}}^{s(b)}$ sampled from $\text{softmax}(\tilde{\mathbf{U}}^{s(b)} \mathbf{W}^\top \cdot \theta_w^{(t)})$
- 15 $\forall s \in \mathbb{B}, \tilde{\mathbf{U}}^{s(b+1)}$ sampled from $\text{softmax}(\tilde{\ell}^{s(b)} + \eta^{(t)} + \tilde{\mathbf{H}}^{s(b)} \mathbf{W} \cdot \theta_w^{(t)})$
- 16 **end**
- 17 $\langle \mathbf{U}^s \rangle_-^{(t)} \leftarrow \tilde{\mathbf{U}}^{s(n)} (s \in \mathbb{B}), \quad \langle \mathbf{H}^s \rangle_-^{(t)} \leftarrow \tilde{\mathbf{H}}^{s(n-1)} (s \in \mathbb{B})$
- 18 $\tilde{\mathbf{U}}^{s(0)} \leftarrow \langle \mathbf{U}^s \rangle_-^{(t)} (s \in \mathbb{B})$ // Reset $\tilde{\mathbf{U}}^{s(0)}$ for persistent-CD in next iteration
- 19 (3) Parameters update phase using eq.4.10 and 4.11:
- 20 $\theta_w^{(t)} = \theta_w^{(t)} + \alpha \cdot \frac{1}{B} \mathbf{W} \sum_{s \in \mathbb{B}} (\langle \mathbf{H}^s \rangle_+^{(t)\top} \langle \mathbf{U}^s \rangle_+^{(t)} - \langle \mathbf{H}^s \rangle_-^{(t)\top} \langle \mathbf{U}^s \rangle_-^{(t)})$
- 21 $\eta^{(t+1)} = \eta^{(t)} + \alpha \cdot \frac{1}{B} \sum_{s \in \mathbb{B}} (\langle \mathbf{U}^s \rangle_+^{(t)} - \langle \mathbf{U}^s \rangle_-^{(t)})$
- 22 counter \leftarrow counter +1
- 23 **end**
- 24 **return** $\langle \mathbf{U} \rangle_+^{(t)}, \tilde{\mathbf{U}}^{(0)}$

expected arrangement log-likelihood $\tilde{\mathcal{L}}_A$ of the m-RBM model can be represented as:

$$\begin{aligned}\tilde{\mathcal{L}}_A &= \sum_{s \in \mathbb{S}} \langle \log \tilde{p}(\mathbf{U}^s; \boldsymbol{\theta}_A) \rangle_q \\ &= \sum_{s \in \mathbb{S}} \sum_{i,k} \langle \mathbf{u}_i^s(k) \rangle_q \cdot \eta_{i,k} + \sum_{s \in \mathbb{S}} \sum_{i,j,k} \theta_w \cdot w_{i,j} \langle \mathbf{u}_i^s(k) \rangle_q \langle \mathbf{h}_j^s(k) \rangle_q\end{aligned}\quad (4.2)$$

Then, the learning of the m-RBM model follows the gradient of $\tilde{\mathcal{L}}_A$ with respect to each of the parameters of $\boldsymbol{\theta}_A$. To approximate the gradients, we applied persistent contrastive divergence (PCD, or PCD- k to indicate the use of k Gibbs steps per update) (Tieleman, 2008) and split the learning into a positive phase (Methods 4.4.2.1) and negative phase (Methods 4.4.2.2). The parameters of $\boldsymbol{\theta}_A$ were then updated in the parameters update phase (Methods 4.4.2.3).

4.4.2.1 Positive phase: expectation given the data

In the positive step, the proposal data distribution was calculated by the mean-field approximation (eq. 4.1). Here for consistency, we use $\langle \cdot \rangle_+$ to indicate the expected value with respect to data distribution in the positive phase.

At the t learning iteration, the posterior expectation $\langle \mathbf{u}_i^s(k) \rangle_+$ for a brain location i of individual s given the data is calculated as:

$$\langle \mathbf{u}_i^s(k) \rangle_+^{(m)} = p(\mathbf{u}_i^s = k | \mathbf{y}_i^s; \boldsymbol{\theta}_A^{(t)}, \boldsymbol{\theta}_E^{(t)}) \quad (4.3)$$

$$= \frac{\exp(\langle \log p(\mathbf{y}_i^s | \mathbf{u}_i^s = k; \boldsymbol{\theta}_E^{(t)}) \rangle_+ + \eta_{i,k}^{(t)} + \langle \mathbf{H}_{k,\cdot}^s \rangle_+^{(m)} \mathbf{W}_{\cdot,i} \cdot \theta_w^{(t)})}{\sum_l \exp(\langle \log p(\mathbf{y}_i^s | \mathbf{u}_i^s = l; \boldsymbol{\theta}_E^{(t)}) \rangle_+ + \eta_{i,l}^{(t)} + \langle \mathbf{H}_{l,\cdot}^s \rangle_+^{(m)} \mathbf{W}_{\cdot,i} \cdot \theta_w^{(t)})} \quad (4.4)$$

where m indicates the number of mean-field inference loop and the $\langle \mathbf{H}^s \rangle_+^{(m)}$ represent the expected value of hidden nodes at the m^{th} mean-field iteration. For the beginning of the mean-field calculation, we initialize the values in hidden layer $\langle \mathbf{H}^s \rangle_+^{(0)} = 0$. Then the expected values $\langle \mathbf{h}_j^s(k) \rangle_+$ in $\langle \mathbf{H}^s \rangle_+$ can be updated for the next iteration given the

values of $\langle \mathbf{u}_i^s(k) \rangle_+^{(m)}$ as:

$$\langle \mathbf{h}_j^s(k) \rangle_+^{(m+1)} = p(\mathbf{h}_j^s = k | \langle \mathbf{U}^s \rangle_+, \boldsymbol{\theta}_A^{(t)}) = \frac{\exp(\langle \mathbf{U}_{k,\cdot}^s \rangle_+^{(m)} \mathbf{W}_{j,\cdot}^\top \cdot \boldsymbol{\theta}_w^{(t)})}{\sum_l \exp(\langle \mathbf{U}_{l,\cdot}^s \rangle_+^{(m)} \mathbf{W}_{j,\cdot}^\top \cdot \boldsymbol{\theta}_w^{(t)})} \quad (4.5)$$

where $\langle \mathbf{U}^s \rangle_+$ is the matrix representation of $\langle \mathbf{u}_i^s(k) \rangle_+$. We then alternatively update $\langle \mathbf{U}^s \rangle_+$ and $\langle \mathbf{H}^s \rangle_+$ through m times mean-field approximation. After the calculation is finished, the $\langle \mathbf{U}^s \rangle_+^{(m)}$ and $\langle \mathbf{H}^s \rangle_+^{(m+1)}$ from the last update iteration will be the estimated posterior expectation at the t learning epoch for positive phase. We then record them as sufficient statistics, $\langle \mathbf{U}^s \rangle_+^{(t)} = \langle \mathbf{U}^s \rangle_+^{(m)}$ and $\langle \mathbf{H}^s \rangle_+^{(t)} = \langle \mathbf{H}^s \rangle_+^{(m+1)}$, for the gradient calculation.

4.4.2.2 Negative phase: expectation given the model

In the negative phase of the learning, we approximate the posterior expectations under the model distribution by n steps of Gibbs sampling. To distinguish from the positive phase, here we use $\langle \cdot \rangle_-$ to indicate the expected value under the model distribution p . For simplicity, we use the matrix notation in the calculation. At the t learning epoch, the mixing of the samples $\tilde{\mathbf{U}}^s$ and $\tilde{\mathbf{H}}^s$ for a subject s between the two layers was started by employing a persistent multinomial Markov chain (Tieleman, 2008) for the parcellation layer. If no Markov chain was given to the model, we initialize $\tilde{\mathbf{U}}^{s(0)}$ by sampling from the group probability map $\eta^{(t)}$.

Then in the n^{th} step of Gibbs updates, the full mixing process includes three steps: **(1)** The data $\tilde{\mathbf{Y}}^{s(n)}$ was first sampled from $p(\mathbf{Y}^s | \tilde{\mathbf{U}}^{s(n)}; \boldsymbol{\theta}_E^{(t)})$, and its associated data log-likelihoods $\tilde{\ell}^{s(n)}$ will be calculated by the emission model (see eq. 3.6). **(2)** $\tilde{\mathbf{H}}^{s(n)}$ can be sampled by only conditioning on the \mathbf{U} layer due to the RBM structure, as:

$$\tilde{\mathbf{H}}^{s(n)} \sim p(\mathbf{H}^s | \tilde{\mathbf{U}}^{s(n)}; \boldsymbol{\theta}_A^{(t)}) = \text{softmax}(\tilde{\mathbf{U}}^{s(n)} \mathbf{W}^\top \cdot \boldsymbol{\theta}_w^{(t)}) \quad (4.6)$$

where $\text{softmax}(\tilde{\mathbf{U}}^{s(n)} \mathbf{W}^\top \cdot \boldsymbol{\theta}_w^{(t)})$ ($K \times J$ matrix) is a collection of multinomial distributions for each hidden node. The expected value for each node $\tilde{\mathbf{h}}_j^{s(n)}$ was then sampled from the corresponding distribution independently and collected into $\tilde{\mathbf{H}}^{s(n)}$. **(3)** Lastly, we

sample the expected value $\tilde{\mathbf{U}}^{s^{(n+1)}}$ of the parcellation layer based on data and hidden nodes sampled in (1) and (2) for the next Gibbs sampling step, as:

$$\tilde{\mathbf{U}}^{s^{(n)}} \sim p(\mathbf{U}^s | \tilde{\mathbf{Y}}^{s^{(n)}}, \tilde{\mathbf{H}}^{s^{(n)}}; \boldsymbol{\theta}_A^{(t)}) = \text{softmax}(\tilde{\ell}^{s^{(n)}} + \eta^{(t)} + \tilde{\mathbf{H}}^{s^{(n)}} \mathbf{W} \cdot \theta_w^{(t)}) \quad (4.7)$$

Similarly $\text{softmax}(\tilde{\ell}^{s^{(n)}} + \eta^{(t)} + \tilde{\mathbf{H}}^{s^{(n)}} \mathbf{W} \cdot \theta_w^{(t)})$ is a $K \times P$ matrix of multinomial distributions for each parcellation node, and their expected value $\tilde{\mathbf{u}}_i^{s^{(n)}}$ were sampled from these distributions independently.

In the last Gibbs sampling update, we use the probability instead of sampling a multinomial value for both $\tilde{\mathbf{U}}^{s^{(n+1)}}$ and $\tilde{\mathbf{H}}^{s^{(n)}}$ to avoid unnecessary sampling noise (Hinton, 2012). We then record the two expected values as sufficient statistics at the t learning epoch of the negative phase for the calculation of the gradient in the update phase, where $\langle \mathbf{U}^s \rangle_-^{(t)} = \tilde{\mathbf{U}}^{s^{(n+1)}}$ and $\langle \mathbf{H}^s \rangle_-^{(t)} = \tilde{\mathbf{H}}^{s^{(n)}}$. The sampled $\tilde{\mathbf{U}}^{s^{(n+1)}}$ will be stored for the next $(t + 1)$ learning epoch for persistent contrastive divergence (Tieleman, 2008).

4.4.2.3 Update phase: parameter estimation using gradients

The m-RBM model parameters are updated in its update phase following the gradients of the un-normalized expected arrangement log-likelihood in eq. 4.2. Given the expectation of the hidden and latent variable for the positive and negative phase calculated in Methods 4.4.2.1 and 4.4.2.2, the gradients with respect to the parameters $\boldsymbol{\theta}_A := \{\eta_{i,k}, \theta_w\}$ at the (t) learning iteration are:

$$\nabla \theta_w^{(t)} = \frac{1}{S} \mathbf{W} \sum_s (\langle \mathbf{H}^s \rangle_+^{(t)\top} \langle \mathbf{U}^s \rangle_+^{(t)} - \langle \mathbf{H}^s \rangle_-^{(t)\top} \langle \mathbf{U}^s \rangle_-^{(t)}) \quad (4.8)$$

$$\nabla \eta^{(t)} = \frac{1}{S} \sum_s (\langle \mathbf{U}^s \rangle_+^{(t)} - \langle \mathbf{U}^s \rangle_-^{(t)}) \quad (4.9)$$

Therefore, the two parameters are updated for the $t + 1$ iteration by:

$$\theta_w^{(t+1)} = \theta_w^{(t)} + \alpha \cdot \nabla \theta_w^{(t)} \quad (4.10)$$

$$\eta^{(t+1)} = \eta^{(t)} + \alpha \cdot \nabla \eta^{(t)} \quad (4.11)$$

where α is the learning rate that is set to a small positive value.

4.4.3 Algorithm implementation and convergence properties

Since the m-RBM model was developed from the RBM family, it inherits the limitations in the training using a contrastive divergence procedure (Hinton, 2002). Therefore, a certain amount of implementation settings and practical experience are required to successfully train the model. In this section, we want to explicit details about the algorithm implementation and several techniques that we used to improve the training performance of the m-RBM model.

The first choice we made about the general training framework is to use stochastic maximum likelihood (SML), a.k.a persistent contrastive divergence (p-CD) (Tieleman, 2008), equipped with mini-batch learning. One big advantage of SML is its training efficiency compared to standard contrastive divergence (CD) when dealing with more complex models. It continuously updates Markov chain throughout the learning process, rather than regenerating at each gradient step in CD, leading to faster and better convergence by exploring more possible space of model configurations (Tieleman, 2008). By integrating the mini-batches, the learning inherits stochasticity from choosing different training subsets to help the model escape from shallow local minima, aiding the optimization process. Another benefit of mini-batches is the computation can be performed much faster with less memory requirement since the gradients are computed over a subset of the entire dataset. This advantage is especially important in our project when dealing with large amounts of imaging data using GPU matrix multiplication acceleration.

We also applied several commonly-used technical settings (Hinton, 2012) for improved m-RBM training, including batch size, learning rate, number of learning epochs, momentum, and Gibbs sampling. In practice, the batch size, learning rate, and number of epochs are highly related and influenced by each other. Due to the limited GPU memory, we first sought to determine the optimal loadable batch size B by finding the maximum divisor of the total training subjects, and repeatedly finding the second maximum if the batch samples cannot be loaded in GPU. At the beginning of a learning epoch, the algorithm randomly sampled B subjects and extracted their data likelihood for the following

computation. When all computations are finished for the current batch, the algorithm returns the subject-specific sufficient statistics to the corresponding subject index. The sampled subjects will be excluded from the next batch sampling, making all subject's data have been used in a single learning epoch. The total number of training epochs was set to $t = 200$ and used as the only stopping criteria for the training due to the lack of a reliable convergence property (see below paragraph). The base learning rate was set to a relatively small value $\alpha = 0.01$ to avoid learning to potentially pass the optima, but as a trade-off, the learning speed may significantly be decreased. Therefore, in the implementation, we used the momentum method (Sutskever et al., 2013) for increasing the learning speed when the objective function contains long, narrow, and straight ravines. Specifically, the momentum method updates the model parameters by a *velocity* term \mathbf{v} learned from the last learning iteration, resulting in the update rule in eq. 4.10 and eq. 4.11 to be changed as:

$$\boldsymbol{\theta}_A^{(t+1)} = \boldsymbol{\theta}_A^{(t)} + \mathbf{v}^{(t+1)}, \quad \text{where } \mathbf{v}^{(t+1)} = \gamma \mathbf{v}^{(t)} - \alpha \cdot \nabla \boldsymbol{\theta}_A^{(t)} \quad (4.12)$$

where γ represents the momentum coefficient that controls how quickly the previous gradients decayed in contributing to the current update, which was set to $\gamma = 0.9$. Lastly, we set the number of mean-field approximation for the positive phase to $m = 5$ and $n = 10$ Gibbs sampling updates for the negative phase respectively (Methods 4.4.2.1, 4.4.2.2) to obtain enough burn-in iterations for reaching an equilibrium.

To monitor the learning progress of the full model, we exclusively built a function \mathcal{J} , which was printed out during the learning. Since the m-RBM model cannot provide a closed-form posterior, the function \mathcal{J} summed up expected emission log-likelihood \mathcal{L}_E with a negative cross-entropy measure $\mathcal{H}(\cdot)$ between $\langle \mathbf{U} \rangle_+^{(t)}$ and $\langle \mathbf{U} \rangle_-^{(t)}$. For a single dataset or session, the \mathcal{J} is calculated as:

$$\mathcal{J}^{(t)} = - \sum_{s \in \mathbb{S}} \mathcal{H}(\langle \mathbf{U}^s \rangle_+^{(t)}, \langle \mathbf{U}^s \rangle_-^{(t)}) + \sum_{s \in \mathbb{S}} \langle \log p(\mathbf{Y}^s | \mathbf{U}^s; \boldsymbol{\theta}_E^{(t)}) \rangle_q \quad (4.13)$$

where the first term is used to penalize the reconstructed latent variables $\langle \mathbf{U}^s \rangle_-$ that

have a larger difference to the $\langle \mathbf{U}^s \rangle_+$. If $\langle \mathbf{U}^s \rangle_+$ and $\langle \mathbf{U}^s \rangle_-$ is perfectly matched, the value of the first term will become zero, and \mathcal{J} restores as \mathcal{L}_E . Overall, an increasing trend of $\mathcal{J}^{(t)}$ during the learning progress indicates the gradients are likely updated on the right track.

In addition, the training of the m-RBM model is subject to a convergence issue, which leads to multiple parameter estimations (Methods 4.2.2). To avoid this problem in the learning of real fMRI data, we fixed the temperature parameter θ_w to a set of positive values within a linearly spaced range from 0.1 to 4, and estimated group probability η only. To further improve the training efficiency, we also froze the learning of the vMF emission model with pre-trained parameters $\boldsymbol{\theta}_E$ trained by the EM algorithm (Methods 1). Besides, we repeated the entire learning process a minimum number of 10 times with each containing 200 epochs. We then picked the estimated parcellations from the trained models with the highest DCBC evaluation score as our final results. By doing all these steps, we have a higher confidence that the resultant parcellations are valid and model training converged to the correct place with a higher probability.

4.4.4 fMRI dataset

In this project, we only considered a single openly available task-based fMRI dataset, MDTB (King et al., 2019), as our main purpose is to investigate the intrinsic smoothness of individual brain organization but not the dataset fusion. The MDTB dataset covers a broad range of task conditions from the perceptual, cognitive, motor, and social domains. A total of approximately 240 minutes of functional scans were collected for each of the 24 participants (16 females, 8 males, mean age=23.8) while performing task set A and B (17 tasks for each, 9 tasks in common). Details of data acquisition parameters, anatomical image registration, and functional data preprocessing can be found elsewhere in King et al. (2019). For the functional profiles, we combined the activity estimates (betas weights) across two task sets and used the mean of the shared tasks as a common baseline. Then the functional profiles were then mapped to each individual cortical surface by averaging the value from voxels along the connecting line between the pial and white-gray matter surface, using 5 equally spaced locations between the two surfaces (Zhi et al.,

2022). No smoothing or group normalization was applied.

To test the performance of the estimated individual parcellations derived from the independent model using smoothed data (Results 4.2.3), we smoothed the cortical functional profiles using `-cifti-smoothing` function provided by Connectome Workbench software (Marcus et al., 2011). Specifically, we gave multiple smoothing levels using a Gaussian kernel from 1 mm to 7 mm standard deviation ($\sigma = 1, 2, 3, 4, 5, 6, 7$). The cortical vertices where the individual data is missing were filled with zeros before the smoothing to prevent the NaN values spread out to the neighboring vertices. After data smoothing, these “missing data” vertices are mapped back to NaN values for the subsequent model training since the framework dealt with missing data naturally (Zhi et al., 2023).

4.4.5 Synthetic datasets for simulation

To validate the proposed spatial arrangement model, we ran several simulations (Results 4.2.1, 4.2.2) on the synthetic datasets. Here, we mainly carried out using the synthetic dataset generation process described in Zhi et al. (2023) with some changes to adapt this project. A four-neighbor schemed Markov Random Field of 50×50 grid is applied to generate the ground-truth individual brain parcellation. Each vertex on the grid represents a brain location that can be assigned to one of the K parcel assignments. In this project, we fixed the $K = 5$ throughout.

Firstly, we generated the artificial group probability map from five centroids μ_k ($k = 1, 2, \dots, 5$) at four corners and one central location on the MRF grid (Appendix B, Fig. B.1a). Then the group probability parameters associated with each vertex for the k^{th} parcel will be represented as the bias term $\eta_{i,k}$ at a location x_i as:

$$\eta_{i,k} = -\frac{|x_i - \mu_k|_2^2}{2\sigma_\mu^2} \quad (4.14)$$

where σ_μ^2 controls the smoothness strength of the group map (Appendix B, Fig. B.1b). Therefore, the group probability $p(\mathbf{u}_i)$ at each location i is then normalized using `softmax` function over the k -long vector, `softmax(η_i)`, as a discrete multinomial distribution.

The parcel assignment for each vertex was then sampled from the group probability $p(\mathbf{u}_i)$ and concatenated as initial individual maps \mathbf{U}^s . Then, this initial \mathbf{U}^s will be passed in a modified RBM model to perform layer-wise Gibb sampling, resulting in the final individual maps after burn-in of m iterations. Specifically, the RBM model contains a visible layer and a hidden layer with the same number of nodes (the total number of vertices on the MRF grid), which are linked by pairwise connection weights $w_{i,j}$. The weights $w_{i,j}$ ($w_{i,j} = w_{j,i}$) indicates whether i and j are neighbouring vertices ($w_{i,j} = 1$ if i and j are neighbours; $w_{i,j} = 0$ otherwise). In practice, $w_{i,j}$ is multiplied by an overall temperature parameter θ_w to control how strong the spatial co-dependence between neighbouring vertices is. A higher θ_w encourages that the two neighbouring nodes to be assigned to the same parcel, enforcing the overall local smoothness of the map (Appendix B, Fig. B.1c). The bias term for both visible and hidden layers is dropped for simplicity in the synthetic data generation. Ultimately, the final stable individual maps were then sampled alternatively between visible and hidden layers after m iterations (Appendix B, Fig. B.1b).

We then generated synthetic functional data \mathbf{Y}^s for each participant based on their individual parcellation maps (Appendix B, Fig. B.1d). Here, we used a standard N -dimensional GMM with normalized mean \mathbf{v}_k ($\|\mathbf{v}_k\| = 1$) of the k^{th} component. The data for each vertex i was generated from:

$$\mathbf{y}_i = \mathbf{v}_k + \boldsymbol{\epsilon} \quad (4.15)$$

where $\boldsymbol{\epsilon}$ was a normal random vector with variance $\mathbf{I} \cdot \sigma_k^2$. Ultimately, a synthetic dataset consisting of N task observations was generated for P brain locations and S subjects.

For the simulation in Results 4.2.1, we smoothed the synthetic functional data in different smoothing levels using a 5×5 Gaussian kernel with different standard deviation σ to convolve the data on the MRF grid. The temperature parameter that controls the overall strength of spatial connection for neighbouring vertices was set to $\theta_w = 1.2$. The bias terms $\eta_{i,k}$ for the group probability were generated with $\sigma_\mu^2 = 240$, while the true number of parcels in the group map and fitting model were both set to $K = 5$. Then,

we sampled 30 individual maps \mathbf{U}^s from the group map with local connection weights $\theta_w \dot{w}_{i,j}$ after $m = 10$ burn-in iterations. The individual maps are further used to sample synthetic training dataset \mathbf{Y}^s ($N = 10$ tasks), and an independent test set \mathbf{Y}^s_{test} ($N = 10$ tasks).

4.4.6 Evaluation measures for probabilistic atlas and simulation

4.4.6.1 DCBC

The distance-controlled boundary coefficient (DCBC, [Zhi et al. \(2022\)](#)) is an unbiased evaluation criterion for brain parcellation, which allows the direct comparison of brain maps generated from different modalities or in different resolutions. This coefficient controls for the intrinsic smoothness of brain data, which is biased in other evaluation metrics ([Gordon et al., 2016](#); [Rousseeuw, 1987](#)). The DCBC method solves this problem by binning all brain location pairs based on their spatial distance and only comparing Pearson’s correlation for within-parcel pairs and between-parcel pairs for the same distance. Then, the DCBC value is calculated as the average correlation differences, weighted by reliability across distances.

In this project, the spatial distance between each pair of brain locations is calculated as the geodesic distance between vertices on the cortical surface under the standard fsLR-32k template ([Van Essen et al., 2012](#)) mesh. This distance metric computation was carried out using the `-surface-geodesic-distance` function provided by Connectome Workbench command ([Marcus et al., 2011](#)) up to 50 mm distance. The underlying functional profiles for calculating the correlations of vertex pairs are the associated betas weights in a task-based dataset. A higher DCBC value of a parcellation indicates that this parcellation predicts the functional boundaries well on the tasks of the test dataset being used. A zero value of DCBC indicates the given parcellation is a random parcellation, while a negative DCBC value represents that the given parcellation predicts the functional boundaries even worse than chance.

4.4.6.2 Mean expected prediction error

With the ground-truth sampled individual parcellation, a second evaluation metric for the simulation is to calculate the absolute error between the true parcellation \mathbf{U} and the predicted $\hat{\mathbf{U}}$ which is inferred from the training data. It is defined as:

$$\langle \bar{\epsilon}_{|\mathbf{U}-\hat{\mathbf{U}}|} \rangle_q = \frac{1}{PS} \sum_s^S \sum_i^P |\mathbf{u}_i^s - \langle \mathbf{u}_i^s \rangle_q| \quad (4.16)$$

where the \mathbf{u}_i^s represents the true parcel assignment label of location i in subject s and $\langle \mathbf{u}_i^s \rangle_q$ is the expected parcel assignment under the expectation q . Both are multinomial encoded vectors. In practice, this calculation of the mean expected prediction error is not subject to the permutation of the parcellation since the parcel assignment vector $\langle \mathbf{u}_i^s \rangle_q$ is aligned with the true assignment \mathbf{u}_i^s . Therefore, looping over all possible permutations and finding the minimum error is not applied.

4.4.6.3 Mean adjusted expected cosine error

A third method to evaluate the simulation results is the mean adjusted expected cosine error. Because the data is generated or modeled using a directional statistical model (e.g. vMF or normalized GMM), the averaged cosine error $\bar{\epsilon}_{\text{cosine}}$ becomes a natural choice of evaluation measure between the reconstructed data $\hat{\mathbf{y}}_i$ to the true data \mathbf{y}_i . In the simplest version, one possibility is to use for each brain location the most likely predicted mean direction,

$$\bar{\epsilon}_{\text{cosine}} = \frac{1}{P} \sum_{i=1}^P \left(1 - \mathbf{v}_{\text{argmax}_k}^\top \frac{\mathbf{y}_i}{\|\mathbf{y}_i\|} \right) \quad (4.17)$$

where $\|\mathbf{y}_i\|$ is the length of the data at brain location i , and $\mathbf{v}_{\text{argmax}_k}$ represents the \mathbf{v}_k with the maximum expectation for that location. We then compute the mean cosine distance across all P brain locations for a single subject.

However, we can also compute the mean expected cosine error for a subject based on the difference between the observed activity profiles \mathbf{y}_i of a location i from the inde-

pendent test set \mathbf{Y}_{test} and the predicted mean directions from the model \mathbf{v}_k , under the expectation of a brain location belongs to which parcel $\langle \mathbf{u}_i \rangle_q$, which defined as below:

$$\langle \bar{\epsilon}_{\text{cosine}} \rangle_q = \frac{1}{P} \sum_i^P \sum_k \langle \mathbf{u}_i(k) \rangle_q (1 - \mathbf{v}_k^\top \frac{\mathbf{y}_i}{\|\mathbf{y}_i\|}) \quad (4.18)$$

where $\langle \mathbf{u}_i(k) \rangle_q$ is the inferred expectation on the training data using the fitted model.

But a possible problem with the cosine error is that brain locations that have very weak signal-to-noise count as much as the ones with strong signal strength due to the normalization. To make it a fair evaluation, we adjusted each expected cosine error in eq. 4.18 by the squared length of the data vector, calculated as $\langle \bar{\epsilon}_{\text{Acosine}} \rangle_q$:

$$\langle \bar{\epsilon}_{\text{Acosine}} \rangle_q = \frac{1}{\sum_i^P \|\mathbf{y}_i\|^2} \sum_i \sum_k \langle \mathbf{u}_i(k) \rangle_q (\|\mathbf{y}_i\|^2 - \mathbf{v}_k^\top \mathbf{y}_i \|\mathbf{y}_i\|) \quad (4.19)$$

We then compute the mean adjusted cosine distance across all P brain locations for a single subject as our third evaluation metric for the simulation. The full mathematical derivations of the mean adjusted expected cosine error $\langle \bar{\epsilon}_{\text{Acosine}} \rangle_q$ (eq. 4.19) and the proof of its equivalent to calculating $1 - R^2$ can be found in Appendix B.

Conceptually, the general idea of mean adjusted expected cosine error is to split the individual data into training $\mathbf{Y}_{\text{train}}$ and test set \mathbf{Y}_{test} . Then it makes a prediction of test data $\hat{\mathbf{Y}}_{\text{test}}$ and determines how well this prediction reconstructs the true test data by calculating the cosine distance between $\hat{\mathbf{Y}}_{\text{test}}$ and \mathbf{Y}_{test} . This prediction is calculated by combining the estimated parcellation from individual training data with the estimated mean task response for K brain regions (eq. 4.19). By doing so, $\langle \bar{\epsilon}_{\text{Acosine}} \rangle_q$ is able to provide a measure of how good a region of the estimated individual parcellation in responding to the tasks from the test set of the same subject. If $\hat{\mathbf{Y}}_{\text{test}}$ is close to the true test data \mathbf{Y}_{test} , a lower $\langle \bar{\epsilon}_{\text{Acosine}} \rangle_q$, it means the estimated individual parcellation and its associated mean task response vector align well with the task in the test.

4.4.7 Computational setup

Model training and evaluations were performed on either an NVIDIA 1080Ti GPU with Python 3, CUDA 11.3, and PyTorch 1.10.2 or on an NVIDIA 4080 GPU with Python 3, CUDA 11.6, and PyTorch 2.0.1. For the fMRI datasets, all data were preprocessed and extracted on an Intel i7-8700 CPU with NumPy 1.24.0, NiBabel 4.0.2, neuroimagingtools 1.0.0. Other detailed requirements and parameters used for the data processing pipeline are available in the respective repositories (see [Code availability](#)).

4.5 Data availability

The raw MDTB data for the fMRI studies used in this project are publicly available on <https://openneuro.org/datasets/ds002105/versions/1.1.0>.

4.6 Code availability

The code for the hierarchical Bayesian parcellation framework is publicly available as the GitHub repository <https://github.com/DiedrichsenLab/HierarchBayesParcel>. The organization, file system, and code for managing the diverse set of datasets is available in a separate repository https://github.com/DiedrichsenLab/Functional_Fusion. The paper-specific code for generating the functional probabilistic parcellations for the neocortex, as well as running the simulation presented in this paper is available at <https://github.com/DiedrichsenLab/FusionModel>.

Chapter 5

General Discussion

5.1 Summary of the thesis

The overarching goal of my thesis is to enhance our understanding of brain functional organization through machine learning techniques. One important approach to human brain mapping is brain parcellation - the definition of a set of distinct regions that can be linked to unique functions. In this thesis, I present a sequence of research projects that addresses some of the challenges and open questions in the context of human brain parcellation.

In **Chapter 2**, we proposed an unbiased criterion to evaluate discrete brain parcellations, called Distance Controlled Boundary Coefficient (DCBC). Recently, numerous brain parcellations have been proposed, using structural, or functional Magnetic Resonance Imaging (fMRI) data. However, the intrinsic smoothness of brain data poses a problem for current methods seeking to compare different parcellations. For example, criteria that simply compare within-parcel to between-parcel similarity provide even random parcellations with a high value. Furthermore, the evaluation is biased by the spatial scale of the parcellation. To address these problems, the proposed DCBC method takes into account spatial smoothness by only comparing pairs of brain locations with the same distance. We then employed this new criterion to evaluate existing parcellations of the human neocortex in their power to predict functional boundaries for an fMRI data set with many different tasks, as well as for resting-state data. Using this improved

criterion, we found that common anatomical parcellations do not perform better than chance, suggesting that task-based functional boundaries do not align well with sulcal landmarks. In contrast, the parcellations based on resting-state fMRI data perform well. Finally, multi-modal parcellations that combine functional and anatomical data perform substantially worse than those based on functional data alone, indicating that functionally homogeneous regions often span major anatomical landmarks. Overall, the DCBC advances the field of functional brain mapping by providing an unbiased metric that directly compares the performance of different brain parcellations to define brain regions that are functionally maximally distinct.

In **Chapter 3**, we addressed one important barrier in the development of complex models of human brain organization: the lack of a large and comprehensive task-based neuroimaging dataset. Because of this limitation, atlases of functional brain organization are currently mainly based on single resting-state datasets. To address this limitation, we proposed a hierarchical Bayesian framework that can learn a probabilistically defined brain parcellation across numerous task-based and resting-state datasets, exploiting their combined strengths. The framework is partitioned into a spatial arrangement model that defines the probability of a specific individual brain parcellation, and a set of dataset-specific emission models that defines the probability of the observed data given the individual brain organization. We showed that the framework optimally combines information from different datasets to achieve a new population-based atlas of the human cerebellum. Furthermore, we demonstrated that, using only 10 min of individual data, the framework is able to generate individual brain parcellations that outperform group atlases.

In **Chapter 4**, we introduced an important extension for the hierarchical Bayesian framework described in Chapter 3. Since the brain organizations are intrinsically smoothed, where nearby regions show higher functional similarity, we sought to further improve the performance of individual parcellations by taking spatial dependence into account. To achieve this goal, we proposed a novel computational architecture called the multinomial-restricted Boltzmann machine (m-RBM), designed to capture the intrinsic spatial dependence between brain locations while accounting for individual variations. We then

integrated the m-RBM model into the hierarchical Bayesian parcellation framework in Chapter 3 to estimate individual parcellations. The simulation results showed the m-RBM model has significant advantages in estimating individual parcellations in the presence of spatial dependencies. On our available dataset, however, we could not find an improvement in brain parcellations over what was achieved with the model from Chapter 3. Nonetheless, further refinements of the m-RBM model could result in improved individual brain parcellations.

Taken together, this series of projects addresses two major challenges in current imaging-based brain parcellation studies, namely individual variability and intrinsic spatial dependence. To account for the individual differences when estimating brain parcellations, Chapters 3 and 4 adopted a hierarchical Bayesian perspective, in which the individual variability is explicitly represented in the model. The advantage of this architecture comes in two folds: (1) individual differences are explicitly taken into account during the learning process, where the individual parcellations are estimated from the subject-specific data rather than average or concatenate all subject's data. (2) When estimating individual parcellations the framework optimally combines the data from an individual subject with the group probability map, thereby improving the quality of individual parcellations. The second advantage is especially important when there is little individual training data, or when it is of poor quality. In this situation, the group probability, derived from the fusion of datasets, plays a critical role in refining uncertain regions, consequently enhancing the quality of individual brain parcellations. Hence, the individual functional differences became more pronounced and comparable across the studied population. For evaluating individual parcellations, the DCBC evaluation method (Chapter 2) quantifies how well a parcellation performs when it is used to predict the functional boundaries in an independent test set. These subject-specific DCBC values provide a measure of how individual functional boundaries vary across different tasks, effectively capturing the differences across studied populations.

Second, to account for the problem of intrinsic spatial dependence between brain locations when evaluating parcellations, the DCBC in Chapter 2 also provides an unbiased evaluation. The DCBC largely reduced the impact of spatial smoothness in the data by

binning brain location pairs based on their spatial distances, and only comparing the within- and between-parcel correlations within each bin. Compared to existing evaluation methods that are biased toward finer and spatially contiguous parcellations, the DCBC provides an unbiased evaluation, making a direct comparison between parcellations in different resolutions possible. To deal with intrinsic spatial dependence from generating perspective, Chapter 4 extended the learning framework in Chapter 3 with the m-RBM spatial arrangement model. This model is designed to handle the spatial structure between brain locations by its computational architecture, resulting in improved performance of the estimated individual parcellations in a simulation. Although we did not observe the expected advantage of the individual parcellations estimated by the m-RBM model on real fMRI data, it is possible that it may still perform better on higher-resolution data. Overall, this set of studies emphasizes the importance of taking into account individual variability and spatial dependence across brain locations to better understand the functional boundaries of brain regions.

5.2 Using atlases: individual functional localizer

In personalized brain studies, one important usage of individual parcellations is to localize a certain functional region in an individual’s brain, namely individual functional localizer (Friston et al., 2006). Traditional brain organization studies rely on the group-averaged atlas to localize functional regions (or ROIs) for individuals. But, recent studies suggest the group-based parcellations may not align well with the functionally unique regions in individual brains or adequately capture the individual variability, which can cause inconsistencies and inaccuracies when individual ROIs are being used (Harrison et al., 2015; Glasser et al., 2016; Kong et al., 2019). Therefore, when we want to localize individual functional regions, we should use individual parcellations. But, the obtaining of reliable individual localizers is often limited by the insufficient amount of individual data, resulting in these individual parcellations being generally noisy and cannot be used for delineating individual ROIs.

To address this problem, the learning framework detailed in Chapter 3 offers a novel

approach to improve the precision of individual parcellations based on individual functional localizing data. This quality improvement in estimated individual parcellations is achieved by combining individual data with a group probability map. It results in the estimated individual parcellation based on only 10 min data to have an equivalent performance of using 100 min data alone. Our evaluation results also consistently showed an improvement when the group probability map is combined with different amounts of individual functional localizing data (Chapter 3, Fig. 3.2).

To estimate the individual parcellations using the framework, we need a pre-trained emission model and a group probability map (see eq. 3.12). The former provides a functional response for each region and will be used to infer the subject-specific data likelihood from the individual functional localizing data. Then, these individual data likelihoods are integrated with the group map to infer the individual parcellation. Therefore, the quality of these parcellations to localize the individual ROIs will be subject to the choice of the group map and individual data. For instance, if there is a constraint that allows only 10 minutes of imaging data to be collected for a specific subject, and the goal is to derive a reliable functional localizer for this subject based on the limited data using our framework, what would be the optimal data to acquire?

Several options exist for acquiring this limited amount of individual localizing data, encompassing a spectrum of tasks such as motor, language, and cognitive activities, or even a ten-minute resting scan. However, the optimal choice of data collection should depend on the specific goal of the users since the estimated individual parcellations can vary substantially based on the type of data being used (Salehi et al., 2020a). Therefore, a careful selection of the behavioral paradigm for the functional localizer becomes important. For example, a few minutes of motor-related task data is preferred if the goal is to study the individual motor region, or language-related tasks should be used in an individual language ROI study. Although individual ROIs can be detected by using limited data, the precision in accurately identifying the functional locations as compared to using larger datasets still remains uncertain. Therefore, it is equally important to choose an appropriate group probability map, as it ensures the estimated individual parcellations can be used as reliable functional localizers. Such a group map should provide useful

information for those uncertain areas, resulting in improved performance of individual functional localizers.

Overall, individual parcellation offers a personalized approach to studying brain organization. However, depending on the specific goal, the effectiveness of individual parcellations lies in balancing the acquisition of individual data with the utilization of a group probability map. This combined approach promises a robust and reliable identification of ROIs and, ultimately, a more accurate understanding of individual brain organization.

5.3 Building atlases: the choice between task-based and resting-state fMRI

Recent brain parcellation studies primarily rely on resting-state functional connectivity (Yeo et al., 2011; Gordon et al., 2016; Schaefer et al., 2018), which captures the intrinsic network pattern based on spontaneous fluctuations of neuronal activity across multiple regions. However, such an approach does not consider the systematic differences observed in the functional organization during various tasks and at rest (Hasson et al., 2009b; Cole et al., 2014; Greene et al., 2020). For example, the hand motor regions are considered as a single parcel in resting-state cerebellar parcellations (Buckner et al., 2011). In contrast, using a task-based parcellation, the left- and right-hand regions can be separated, because some of the tasks involved unimanual left-hand responses and other tasks involved unimanual right-hand responses (King et al., 2019).

On the other hand, traditional task-based fMRI studies usually focus on one type of task, resulting in the data only characterizing a few functional regions well, hence, preventing a comprehensive comparison between task and resting-state brain organizations in different regions. Recently, a number of “deep-phenotyping” task-based fMRI datasets have become openly available (Pinho et al., 2018; Nakai and Nishimoto, 2020), which offers the opportunity to systematically study the brain organization during tasks, revealing functional boundaries across multiple task domains (King et al., 2019). Therefore, it is important to consider task-based datasets in deriving brain parcellations (King

et al., 2019), leveraging the strength across different datasets.

This systematic difference in the brain functional organization leads to an open question: what type of data to use for building brain functional atlases, task-based or resting-state fMRI? In Chapter 3, we addressed this question by comparing the parcellations based only on task-based data with parcellations based only on resting-state fMRI data. We found that both types of parcellation can predict functional boundaries across other task-based datasets, however, we still observed some systematic differences between the two (Appendix A, Fig. A.5). This is because different datasets will emphasize different sets of functional boundaries. On the other hand, task-based and resting-state parcellations do retain some common patterns, as seen in the superior performance of the resulting fused parcellation compared to those created exclusively from either task-based or resting-state data (Chapter 3, Fig. 3.7). This finding implies that integrating both task-based and resting-state fMRI data can lead to better brain parcellations.

When integrating different types of data, we also need to decide how much data to use from each source. What is the optimal balance between task-based and resting-state data usage? This is an important decision to make when using our fusion framework because a large dataset could dominate the group map, possibly reducing the predictive performance of other datasets. In Chapter 3, we demonstrated that the fused parcellations have better performance when integrating 50 subjects' resting-state data with the task-based data of 110 subjects. But, it remains uncertain whether this advantage in the fused parcellation would persist when we use different amounts of data from each modality. If changing the balance between task and resting data results in a reduction in the performance of the fused parcellation, it could suggest that creating different brain parcellations is better to use different data modalities. Therefore, finding an optimal balance between task-based and resting-state data usage forms a research question that demands further investigation.

5.4 Evaluating atlases: aligning functions to brain regions

Evaluation methods for assessing brain parcellation can be categorized into consistency and validity measures (Chapter 1, 1.2.4). The former measures the similarity between parcellations, while the latter aims to evaluate the ability of a parcellation to segregate brain locations into functionally homogeneous regions. In Chapter 2, we introduced the DCBC as a method in the second category. This criterion evaluates how well a brain parcellation predicts the local functional boundaries in an independent test set. However, there is a common issue with these methods: the evaluation is blind to parcel assignments, which means the evaluation score will remain the same if the functional profiles are flipped between two regions. Therefore, the DCBC evaluation provides a good measure of how well a parcellation is for matching functional boundaries, but if parcellations are estimated with a functional response for each parcel (functional localizer), the DCBC cannot determine how well the functional response in a parcel is consistent across subjects.

To address this limitation, our proposed mean adjusted expected cosine error (Chapter 4, 4.4.6.3) could be used for evaluating individual parcellation as a supplementary measure to the DCBC. First, this measure is suited to assess predictions of brain functional responses, where the profiles are learned from subjects other than the one being examined. This aspect is critical to gain insights into the differences across individual brains and how the brain’s response patterns differ from one person to another. Second, the mean adjusted expected cosine error only yields lower values if the brain responses in a specific region are consistent across subjects. Lastly, by employing the expected value, this method allows for the evaluation of probabilistic (soft) brain parcellations, largely expanding the applicability of brain parcellations in different types. Therefore, an individual parcellation, that simultaneously possesses a higher DCBC value and lower value of mean adjusted expected cosine error, detects the functional boundaries well and predicts the functional profiles of parcels more consistently across individuals at the same time. Overall, the mean adjusted expected cosine error provides an alternative view of

evaluating brain parcellations for future studies and advancements in brain parcellation.

5.5 Interpreting brain organizations as gradient or parcellation

There is an ongoing debate in recent functional brain organization studies: should the brain be interpreted as soft, smoothed gradients or a hard, discrete parcellation (Tononi et al., 1994; Friston, 2011; Bijsterbosch et al., 2020)? While the segmentation of the brain into discrete regions is the most common approach to describe its functional organization, some studies have interpreted brain organization as a functional gradient (Cohen et al., 2008; Margulies et al., 2016; Haak et al., 2018; Tian et al., 2020): the gradual transitions along a functional spectrum. This gradient perspective is based on the observation that many functional regions are not separated by an abrupt functional boundary, but that the transition between them occurs smoothly. The functional gradient perspective embraces the idea that cognitive functions result from the integrated activity of networks, emphasizing no brain region is sufficient to perform a particular function by itself.

Although this thesis work was built upon the concept of functional parcellation, Chapter 2 addressed this open debate from an evaluation perspective. If the brain functions are assumed to be solely varied in smooth gradients across the cortical surface, the expected DCBC value when tested on a novel task set should be around zero, which means no boundaries are aligned. However, in our study, we observed systematic positive DCBC values for functional parcellations (Chapter 2, Fig. 2.5). This demonstrates clearly that there are functional boundaries in the human brain, where the functional specialization changes abruptly across tasks. On the other hand, these predicted functional boundaries are not equally strong, as can be seen in the noisy or uncertain areas in the estimated parcellation (Appendix A, Fig. A.5), or as evidenced in boundary-wise evaluation presented in King et al. (2019). With boundaries of decreasing importance, the true number of functionally distinct parcels, denoted as K , is not clearly defined. This can be seen in our evaluation results with the performance not peaking at a specific value for K , but

slowly reaching an asymptote (Chapter 3, Fig. 3.6f). If we use a different set of tasks to evaluate the parcellations, some boundaries are not stable and may shift across different tasks. This suggests some parts of the brain are better explained by a gradual transition from one function to the other. Overall, our results argue that the two descriptions, discrete parcellation vs. continuous gradients, both capture important aspects of the underlying brain organization, revealing that the two interpretations may not be mutually exclusive.

These together raise interesting further methodological questions of whether we can find a way to integrate the two interpretations into a common framework. To this end, further development can build brain parcellations that simultaneously possess hard boundaries and the feature of smooth gradients, such as probabilistic brain parcellations. In such parcellations, the parcel assignment of a brain location is defined through a probability distribution over all possible K regions, resulting in a flexible version between the two concepts. For instance, in regions where the function transitions gradually, the probability distribution would be more evenly spread across multiple parcels. In contrast, in regions with sharp functional boundaries, the distribution would be heavily skewed towards a single parcel. This kind of approach could help bridge the gap between the gradients and parcellations, offering a better understanding of the complex functional organization.

5.6 Extensions of current work

In previous discussions, I have dissected this thesis work from three common scenarios for brain organization studies, such as when using (Discussion 5.2), building (Discussion 5.3), and evaluating (Discussion 5.4) brain atlases. Within each scenario, I also discussed the implications or potential research directions from a broader view. Here in this section, I will address some useful extensions that can be applied to my work in the near future from a computational modeling perspective.

5.6.1 The optimal number of parcels

The number of parcels in a brain parcellation is a critical factor, influencing both the resolution and performance of the resultant brain maps. Without knowing the ground-truth number of parcels K , most brain parcellation methods use a fixed number throughout or a predefined range to learn a set of parcellations in different resolutions. This is also the case in our fusion framework, where K was defined from 10 to 100 for estimating cerebellar parcellations. However, this strategy has certain limitations as the true number of parcels vary across different training dataset or subject. Even using the same dataset, the performance of the derived parcellations can significantly vary depending on the choice of K (Chapter 3, Fig. 3.6). Hence, it is important to develop a parcellation learning algorithm capable of adaptively determining the optimal number of regions or parcels.

To adaptively estimate the number of parcels K in our learning framework (Chapter 3), one potential solution is to leverage the idea from Bayesian model selection methods, such as Bayesian information criterion (BIC) (Schwarz, 1978). It provides a measure of the trade-off between the likelihood of a model given the data and the complexity of the model (complexity increases with K rising), helping to avoid over-fitting. In this case, we can add an outer loop to iteratively decrease K from a relatively larger number and incorporate a term related to BIC into the model likelihood calculation for each iteration, effectively balancing model complexity (K) and goodness of fit during the learning process. This would allow the algorithm to “decide” on an optimal number of parcels during the learning process rather than having to specify it beforehand.

5.6.2 Hierarchical structure of brain parcellation

Another interesting and practically useful extension to our framework is to build a hierarchical version of brain parcellation. The resulting parcellation can be represented as a hierarchical tree, where on top of the tree contains the information of a coarse parcellation. This coarse parcellation separates the brain into a few large regions by strong and stable functional boundaries, emphasizing the parcellation perspective of brain organization. On the other hand, users have the flexibility to “cut” this hierarchical tree

from a relatively bottom position, resulting in a finer resolution parcellation. In this finer parcellation, the smaller size parcels are nested inside the coarse parcels separated by weak boundaries. The instability of the boundaries for these small regions can be detected by the shifting when K changes, reflecting the gradient perspective of the brain organization.

This hierarchical structure of brain parcellation addresses the open debate between functional gradient and parcellation. Hierarchical parcellation, by its nature, would capture both the fine-grained and coarser elements of brain organization, thus bridging the gap between functional gradient and parcellation views. When building such parcellation under our fusion framework in Chapter 3, we can start the learning algorithm with a very fine resolution (high K) and keep merging the neighboring parcels bottom-up, until the algorithm meets the converge criterion or reaches a desired lower resolution. In this case, the development of such convergence properties has become important for addressing both interpretations of brain organization and remained a practical question for further investigation.

5.6.3 Deep generative architecture for improved spatial arrangement model

In Chapter 4, we proposed an m-RBM model to estimate individual brain parcellations, which is essential to model the spatial structure between brain locations. The m-RBM model showed several advantages in modeling intrinsic smoothness but it still suffers from the intractability of the partition function like any energy-based model. Although advanced training algorithms for energy-based models exist and have better performance and convergence compared to traditional contrastive divergence (Song and Kingma, 2021), they are still subject to certain limitations. These limitations may include computational efficiency, susceptibility to local minima, sensitivity to hyperparameter choices, and their ability to effectively learn complex, high-dimensional distributions, which is particularly important when dealing with neuroimaging data.

Thus, future extensions for improved spatial arrangement models could explore the

utilization of deep generative models, such as variational autoencoders (Kingma and Welling, 2013), generative adversarial networks (Goodfellow et al., 2020), or normalizing flow (Rezende and Mohamed, 2015) that have shown great promise in handling complex and high-dimensional data. These models could offer a more efficient and stable training process, and they also provide an explicit model of the data distribution, which is advantageous for understanding and visualizing the estimated individual brain parcellations. Furthermore, the latent space learned by these models could offer a rich, lower-dimensional representation of the complex spatial arrangement of brain regions, enabling more precise mapping of individual brain functions.

5.7 Conclusion

In this thesis, I have developed a number of methodological innovations that address important challenges in the field of brain parcellation. These developments have allowed us to create better models that capture the intrinsic spatial dependence and individual variability present in human brain structures, thus pushing our knowledge of personalized brain studies forward. This progress doesn't only improve our understanding of brain functions but also carries significant implications for clinical neuroscience. Improved individual parcellations, as produced by the methods presented in this thesis, could pave the way for enhanced clinical studies and open new doors in personalized medicine. By providing more accurate mapping of individual brain functions, these advancements can also potentially facilitate more precise diagnoses and more targeted treatments for neurological disorders. The journey to fully comprehend the human brain complexity is far from over, but with the help of machine learning, we are one step closer to unraveling its mysteries.

Bibliography

- K. Amunts and K. Zilles. Architectonic mapping of the human brain beyond brodmann. *Neuron*, 88(6):1086–1107, 2015.
- S. Arslan and D. Rueckert. Multi-level parcellation of the cerebral cortex using resting-state fmri. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 47–54. Springer, 2015.
- S. Arslan, S. Parisot, and D. Rueckert. Joint spectral decomposition for the parcellation of the human cerebral cortex using resting-state fmri. In *International Conference on Information Processing in Medical Imaging*, pages 85–97. Springer, 2015.
- S. Arslan, S. I. Ktena, A. Makropoulos, E. C. Robinson, D. Rueckert, and S. Parisot. Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *Neuroimage*, 170:5–30, 2018.
- J. Ashburner and K. Friston. Multimodal image coregistration and partitioning—a unified framework. *Neuroimage*, 6(3):209–217, 1997.
- J. Ashburner and K. J. Friston. Unified segmentation. *Neuroimage*, 26(3):839–851, 2005.
- M. Assem, S. Shashidhara, M. F. Glasser, and J. Duncan. Basis of executive functions in fine-grained architecture of cortical and subcortical human brain networks. *bioRxiv*, 2022.
- C. Baldassano, D. M. Beck, and L. Fei-Fei. Parcellating connectivity in spatial maps. *PeerJ*, 3:e784, 2015.
- A. Banerjee, I. S. Dhillon, J. Ghosh, S. Sra, and G. Ridgeway. Clustering on the unit hypersphere using von mises-fisher distributions. *Journal of Machine Learning Research*, 6(9), 2005.
- K. A. Barnes, A. L. Cohen, J. D. Power, S. M. Nelson, Y. B. Dosenbach, F. M. Miezin, S. E. Petersen, and B. L. Schlaggar. Identifying basal ganglia divisions in individuals using resting-state functional connectivity mri. *Frontiers in systems neuroscience*, 4: 18, 2010.
- H. Barthel, H.-J. Gertz, S. Dresel, O. Peters, P. Bartenstein, K. Buerger, F. Hiemeyer, S. M. Wittmer-Rump, J. Seibyl, C. Reininger, et al. Cerebral amyloid- β pet with florbetaben (18f) in patients with alzheimer’s disease and healthy controls: a multicentre phase 2 diagnostic study. *The Lancet Neurology*, 10(5):424–435, 2011.

- C. F. Beckmann and S. M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23(2):137–152, 2004.
- C. F. Beckmann, M. DeLuca, J. T. Devlin, and S. M. Smith. Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):1001–1013, 2005.
- T. E. Behrens, H. Johansen-Berg, M. Woolrich, S. Smith, C. Wheeler-Kingshott, P. Boulby, G. Barker, E. Sillery, K. Sheehan, O. Ciccarelli, et al. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature neuroscience*, 6(7):750–757, 2003.
- P. Bellec, P. Rosa-Neto, O. C. Lyttelton, H. Benali, and A. C. Evans. Multi-level bootstrap analysis of stable clusters in resting-state fmri. *Neuroimage*, 51(3):1126–1139, 2010.
- M. G. Berman, J. Park, R. Gonzalez, T. A. Polk, A. Gehrke, S. Knaffla, and J. Jonides. Evaluating functional localizers: the case of the ffa. *Neuroimage*, 50(1):56–71, 2010.
- J. Bijsterbosch, S. J. Harrison, S. Jbabdi, M. Woolrich, C. Beckmann, S. Smith, and E. P. Duff. Challenges and future directions for representations of functional brain organization. *Nature neuroscience*, 23(12):1484–1495, 2020.
- J. D. Bijsterbosch, M. W. Woolrich, M. F. Glasser, E. C. Robinson, C. F. Beckmann, D. C. Van Essen, S. J. Harrison, and S. M. Smith. The relationship between spatial configuration and functional connectivity of brain regions. *elife*, 7:e32992, 2018.
- J. D. Bijsterbosch, C. F. Beckmann, M. W. Woolrich, S. M. Smith, and S. J. Harrison. The relationship between spatial configuration and functional connectivity of brain regions revisited. *eLife*, 8, 5 2019.
- B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.
- D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.
- T. Blumensath, S. Jbabdi, M. F. Glasser, D. C. Van Essen, K. Ugurbil, T. E. Behrens, and S. M. Smith. Spatially constrained hierarchical parcellation of the brain with resting-state fmri. *Neuroimage*, 76:313–324, 2013.
- R. M. Braga and R. L. Buckner. Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. *Neuron*, 95(2):457–471, 2017.
- K. Brodmann. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, 1909.

- R. L. Buckner, F. M. Krienen, A. Castellanos, J. C. Diaz, and B. T. Yeo. The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(5):2322–2345, 2011.
- R. L. Buckner, F. M. Krienen, and B. T. Yeo. Opportunities and limitations of intrinsic functional connectivity mri. *Nature neuroscience*, 16(7):832–837, 2013.
- D. Bzdok, A. Heeger, R. Langner, A. R. Laird, P. T. Fox, N. Palomero-Gallagher, B. A. Vogt, K. Zilles, and S. B. Eickhoff. Subspecialization in the human posterior medial cortex. *Neuroimage*, 106:55–71, 2015.
- A. Cachia, M. Roell, J.-F. Mangin, Z. Y. Sun, A. Jobert, L. Braga, O. Houde, S. Dehaene, and G. Borst. How interindividual differences in brain anatomy shape reading accuracy. *Brain Structure and Function*, 223:701–712, 2018.
- R. Ciric, W. H. Thompson, R. Lorenz, M. Goncalves, E. E. MacNicol, C. J. Markiewicz, Y. O. Halchenko, S. S. Ghosh, K. J. Gorgolewski, R. A. Poldrack, et al. Templateflow: Fair-sharing of multi-scale, multi-species brain models. *Nature Methods*, 19:1568–1571, 2022.
- A. L. Cohen, D. A. Fair, N. U. Dosenbach, F. M. Miezin, D. Dierker, D. C. Van Essen, B. L. Schlaggar, and S. E. Petersen. Defining functional areas in individual human brains using resting functional connectivity mri. *Neuroimage*, 41(1):45–57, 2008.
- M. W. Cole, D. S. Bassett, J. D. Power, T. S. Braver, and S. E. Petersen. Intrinsic and task-evoked network architectures of the human brain. *Neuron*, 83(1):238–251, 2014.
- R. C. Craddock, G. A. James, P. E. Holtzheimer III, X. P. Hu, and H. S. Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.
- R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.
- L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- J. Diedrichsen. A spatially unbiased atlas template of the human cerebellum. *Neuroimage*, 33(1):127–138, 2006.
- J. Diedrichsen and R. Shadmehr. Detecting and adjusting for artifacts in fmri time series data. *Neuroimage*, 27(3):624–634, 2005.
- J. Diedrichsen and E. Zotow. Surface-based display of volume-averaged cerebellar imaging data. *PloS one*, 10(7):e0133402, 2015.

- E. W. Dijkstra et al. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.
- C. Ding, X. He, and H. D. Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proceedings of the 2005 SIAM international conference on data mining*, pages 606–610. SIAM, 2005.
- E. Dohmatob, H. Richard, A. L. Pinho, and B. Thirion. Brain topography beyond parcellations: local gradients of functional maps. *NeuroImage*, 229:117706, 2021.
- S. B. Eickhoff, B. Thirion, G. Varoquaux, and D. Bzdok. Connectivity-based parcellation: Critique and implications. *Human brain mapping*, 36(12):4771–4792, 2015.
- S. B. Eickhoff, B. T. Yeo, and S. Genon. Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience*, 19(11):672–686, 2018.
- L. Fan, H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A. R. Laird, et al. The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral cortex*, 26(8):3508–3526, 2016.
- D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex*, 1(1):1–47, 1991.
- B. Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.
- B. Fischl, A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, et al. Automatically parcellating the human cerebral cortex. *Cerebral cortex*, 14(1):11–22, 2004.
- B. Fischl, N. Rajendran, E. Busa, J. Augustinack, O. Hinds, B. T. Yeo, H. Mohlberg, K. Amunts, and K. Zilles. Cortical folding patterns and predicting cytoarchitecture. *Cereb Cortex*, 18(8):1973–1980, 2008.
- V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, D. L. Collins, B. D. C. Group, et al. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*, 54(1):313–327, 2011.
- M. D. Fox and M. Greicius. Clinical applications of resting state functional connectivity. *Frontiers in systems neuroscience*, page 19, 2010.
- M. D. Fox and M. E. Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews neuroscience*, 8(9):700–711, 2007.
- K. J. Friston. Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36, 2011.
- K. J. Friston, C. D. Frith, R. S. Frackowiak, and R. Turner. Characterizing dynamic brain responses with fmri: a multivariate approach. *Neuroimage*, 2(2):166–172, 1995.

- K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. Rugg, and R. Turner. Event-related fmri: characterizing differential responses. *Neuroimage*, 7(1):30–40, 1998a.
- K. J. Friston, O. Josephs, G. Rees, and R. Turner. Nonlinear event-related responses in fmri. *Magnetic resonance in medicine*, 39(1):41–52, 1998b.
- K. J. Friston, P. Rotshtein, J. J. Geng, P. Sterzer, and R. N. Henson. A critique of functional localisers. *Neuroimage*, 30(4):1077–1087, 2006.
- M. F. Glasser and D. C. Van Essen. Mapping human cortical areas in vivo based on myelin content as revealed by t1-and t2-weighted mri. *Journal of Neuroscience*, 31(32):11597–11616, 2011.
- M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. Van Essen, and M. Jenkinson. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105–124, 2013.
- M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.
- P. Golland, Y. Golland, and R. Malach. Detection of spatial activation patterns as unsupervised segmentation of fmri data. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29–November 2, 2007, Proceedings, Part I 10*, pages 110–118. Springer, 2007.
- Y. Golland, P. Golland, S. Bentin, and R. Malach. Data-driven clustering reveals a fundamental subdivision of the human cortex into two global systems. *Neuropsychologia*, 46(2):540–553, 2008.
- I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- E. M. Gordon, T. O. Laumann, B. Adeyemo, J. F. Huckins, W. M. Kelley, and S. E. Petersen. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral cortex*, 26(1):288–303, 2016.
- E. M. Gordon, T. O. Laumann, B. Adeyemo, A. W. Gilmore, S. M. Nelson, N. U. Dosenbach, and S. E. Petersen. Individual-specific features of brain systems identified with resting state functional correlations. *Neuroimage*, 146:918–939, 2017a.
- E. M. Gordon, T. O. Laumann, A. W. Gilmore, D. J. Newbold, D. J. Greene, J. J. Berg, M. Ortega, C. Hoyt-Drazen, C. Gratton, H. Sun, et al. Precision functional mapping of individual human brains. *Neuron*, 95(4):791–807, 2017b.

- K. J. Gorgolewski, T. Auer, V. D. Calhoun, R. C. Craddock, S. Das, E. P. Duff, G. Flandin, S. S. Ghosh, T. Glatard, Y. O. Halchenko, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, 3(1):1–9, 2016.
- A. S. Greene, S. Gao, S. Noble, D. Scheinost, and R. T. Constable. How tasks change whole-brain functional organization to reveal brain-phenotype relationships. *Cell reports*, 32(8):108066, 2020.
- M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the national academy of sciences*, 100(1):253–258, 2003.
- D. N. Greve and B. Fischl. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1):63–72, 2009.
- X. Guell, J. D. Schmahmann, J. D. Gabrieli, and S. S. Ghosh. Functional gradients of the cerebellum. *Elife*, 7:e36652, 2018.
- M. Gutmann and A. Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 297–304. JMLR Workshop and Conference Proceedings, 2010.
- K. V. Haak, A. F. Marquand, and C. F. Beckmann. Connectopic mapping with resting-state fmri. *Neuroimage*, 170:83–94, 2018.
- S. J. Harrison, M. W. Woolrich, E. C. Robinson, M. F. Glasser, C. F. Beckmann, M. Jenkinson, and S. M. Smith. Large-scale probabilistic functional modes from resting state fmri. *NeuroImage*, 109:217–231, 2015.
- U. Hasson, H. C. Nusbaum, and S. L. Small. Task-dependent organization of brain regions active during rest. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26):10841–10846, jun 2009a.
- U. Hasson, H. C. Nusbaum, and S. L. Small. Task-dependent organization of brain regions active during rest. *Proceedings of the National Academy of Sciences*, 106(26):10841–10846, 2009b.
- G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.
- G. E. Hinton. A practical guide to training restricted boltzmann machines. *Neural Networks: Tricks of the Trade: Second Edition*, pages 599–619, 2012.
- G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

- S. Hirose, T. Watanabe, K. Jimura, M. Katsura, A. Kunimatsu, O. Abe, K. Ohtomo, Y. Miyashita, and S. Konishi. Local signal time-series during rest used for areal boundary mapping in individual human brains. *PLoS One*, 7(5):e36496, 2012.
- N. Honnorat, H. Eavani, T. D. Satterthwaite, R. E. Gur, R. C. Gur, and C. Davatzikos. Grasp: geodesic graph-based segmentation with shape priors for the functional parcellation of the cortex. *Neuroimage*, 106:207–221, 2015.
- K. Hornik and B. Grün. movmf: An r package for fitting mixtures of von mises-fisher distributions. *Journal of Statistical Software*, 58(10):1–31, 2014.
- L. Hubert and P. Arabie. Comparing partitions. *Journal of classification*, 2:193–218, 1985.
- J. M. Huntenburg, P.-L. Bazin, and D. S. Margulies. Large-scale gradients in human cortical organization. *Trends in cognitive sciences*, 22(1):21–31, 2018.
- A. Hyvärinen and P. Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.
- M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.
- M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.
- J. L. Ji, M. Spronk, K. Kulkarni, G. Repovš, A. Anticevic, and M. W. Cole. Mapping the human brain’s cortical-subcortical functional network organization. *Neuroimage*, 185:35–57, 2019.
- H. Johansen-Berg, T. Behrens, M. Robson, I. Drobnjak, M. Rushworth, J. Brady, S. Smith, D. Higham, and P. Matthews. Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences*, 101(36):13335–13340, 2004.
- J. H. Kaas. The organization of neocortex in mammals: implications for theories of brain function. *Annual review of psychology*, 38(1):129–151, 1987.
- N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, 17(11):4302–4311, 1997.
- R. Kindermann and J. L. Snell. *Markov random fields and their applications*, volume 1. American Mathematical Society, 1980.
- M. King, C. R. Hernandez-Castillo, R. A. Poldrack, R. B. Ivry, and J. Diedrichsen. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nature neuroscience*, 22(8):1371–1378, 2019.

- D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- I. Klatzo. *Cecile and Oskar Vogt: the visionaries of modern neuroscience*, volume 80. Springer Science & Business Media, 2002.
- R. Kong, J. Li, C. Orban, M. R. Sabuncu, H. Liu, A. Schaefer, N. Sun, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, et al. Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. *Cerebral cortex*, 29(6):2533–2551, 2019.
- R. Kong, Q. Yang, E. Gordon, A. Xue, X. Yan, C. Orban, X.-N. Zuo, N. Spreng, T. Ge, A. Holmes, et al. Individual-specific areal-level parcellations improve functional connectivity prediction of behavior. *Cerebral Cortex*, 31(10):4477–4500, 2021.
- R. Lafer-Sousa, B. R. Conway, and N. G. Kanwisher. Color-biased regions of the ventral visual pathway lie between face-and place-selective regions in humans, as in macaques. *Journal of Neuroscience*, 36(5):1682–1697, 2016.
- D. Lashkari, E. Vul, N. Kanwisher, and P. Golland. Discovering structure in the space of fmri selectivity profiles. *Neuroimage*, 50(3):1085–1098, 2010.
- T. O. Laumann, E. M. Gordon, B. Adeyemo, A. Z. Snyder, S. J. Joo, M.-Y. Chen, A. W. Gilmore, K. B. McDermott, S. M. Nelson, N. U. Dosenbach, et al. Functional system and areal organization of a highly sampled individual human brain. *Neuron*, 87(3):657–670, 2015.
- M. H. Lee, C. D. Hacker, A. Z. Snyder, M. Corbetta, D. Zhang, E. C. Leuthardt, and J. S. Shimony. Clustering of resting state networks. *PloS one*, 7(7):e40370, 2012.
- M. H. Lee, C. D. Smyser, and J. S. Shimony. Resting-state fmri: a review of methods and clinical applications. *American Journal of neuroradiology*, 34(10):1866–1872, 2013.
- S. Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.
- N. K. Logothetis. The underpinnings of the bold functional magnetic resonance imaging signal. *Journal of Neuroscience*, 23(10):3963–3971, 2003.
- N. K. Logothetis. What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–878, 2008.
- D. Marcus, J. Harwell, T. Olsen, M. Hodge, M. Glasser, F. Prior, M. Jenkinson, T. Laumann, S. Curtiss, and D. Van Essen. Informatics and data mining tools and strategies for the human connectome project. *Frontiers in neuroinformatics*, 5:4, 2011.
- S. Marek, J. S. Siegel, E. M. Gordon, R. V. Raut, C. Gratton, D. J. Newbold, M. Ortega, T. O. Laumann, B. Adeyemo, D. B. Miller, et al. Spatial and temporal organization of the individual human cerebellum. *Neuron*, 100(4):977–993, 2018.

- D. S. Margulies, S. S. Ghosh, A. Goulas, M. Falkiewicz, J. M. Huntenburg, G. Langs, G. Bezgin, S. B. Eickhoff, F. X. Castellanos, M. Petrides, et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, 113(44):12574–12579, 2016.
- S. Mueller, D. Wang, M. D. Fox, B. T. Yeo, J. Sepulcre, M. R. Sabuncu, R. Shafee, J. Lu, and H. Liu. Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3):586–595, 2013.
- J. A. Mumford, S. Horvath, M. C. Oldham, P. Langfelder, D. H. Geschwind, and R. A. Poldrack. Detecting network modules in fmri time series: a weighted network analysis approach. *Neuroimage*, 52(4):1465–1476, 2010.
- T. Nakai and S. Nishimoto. Quantitative models reveal the organization of diverse cognitive functions in the brain. *Nature communications*, 11(1):1–12, 2020.
- S. M. Nelson, A. L. Cohen, J. D. Power, G. S. Wig, F. M. Miezin, M. E. Wheeler, K. Velanova, D. I. Donaldson, J. S. Phillips, B. L. Schlaggar, et al. A parcellation scheme for human left lateral parietal cortex. *Neuron*, 67(1):156–170, 2010.
- A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001.
- H. Nili, C. Wingfield, A. Walther, L. Su, W. Marslen-Wilson, and N. Kriegeskorte. A toolbox for representational similarity analysis. *PLoS Comput Biol*, 10(4):e1003553, 2014.
- K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby. Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends in cognitive sciences*, 10(9):424–430, 2006.
- S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.
- S. Parisot, S. Arslan, J. Passerat-Palmbach, W. M. Wells III, and D. Rueckert. Group-wise parcellation of the cortex through multi-scale spectral clustering. *NeuroImage*, 136:68–83, 2016.
- A. L. Pinho, A. Amadon, T. Ruest, M. Fabre, E. Dohmatob, I. Denghien, C. Ginisty, S. Becuwe-Desmidt, S. Roger, L. Laurier, et al. Individual brain charting, a high-resolution fmri dataset for cognitive mapping. *Scientific data*, 5(1):1–15, 2018.
- A. L. Pinho, A. Amadon, B. Gauthier, N. Clairis, A. Knops, S. Genon, E. Dohmatob, J. J. Torre, C. Ginisty, S. Becuwe-Desmidt, et al. Individual brain charting dataset extension, second release of high-resolution fmri data for cognitive mapping. *Scientific Data*, 7(1):353, 2020.

- J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, et al. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011.
- M. E. Raichle and M. A. Mintun. Brain work and brain imaging. *Annu. Rev. Neurosci.*, 29:449–476, 2006.
- D. Rezende and S. Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.
- R. E. Røge, K. H. Madsen, M. N. Schmidt, and M. Mørup. Infinite von mises–fisher mixture modeling of whole brain fmri data. *Neural Computation*, 29(10):2712–2741, 2017.
- P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- S. Ryali, T. Chen, K. Supekar, and V. Menon. A parcellation scheme based on von mises-fisher distributions and markov random fields for segmenting brain regions using resting-state fmri. *Neuroimage*, 65:83–96, 2013.
- N. Saadon-Grosman, P. A. Angeli, L. M. DiNicola, and R. L. Buckner. A third somato-motor representation in the human cerebellum. *Journal of Neurophysiology*, 128(4):1051–1073, 2022.
- R. Salakhutdinov and G. Hinton. Deep boltzmann machines. In D. van Dyk and M. Welling, editors, *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 448–455, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 16–18 Apr 2009. PMLR.
- R. Salakhutdinov and G. Hinton. An efficient learning procedure for deep Boltzmann machines. *Neural computation*, 24(8):1967–2006, 2012.
- M. Salehi, A. Karbasi, X. Shen, D. Scheinost, and R. T. Constable. An exemplar-based approach to individualized parcellation reveals the need for sex specific functional networks. *Neuroimage*, 170:54–67, 2018.
- M. Salehi, A. S. Greene, A. Karbasi, X. Shen, D. Scheinost, and R. T. Constable. There is no single functional atlas even for a single individual: Functional parcel definitions change with task. *NeuroImage*, 208:116366, mar 2020a.
- M. Salehi, A. Karbasi, D. S. Barron, D. Scheinost, and R. T. Constable. Individualized functional networks reconfigure with cognitive state. *NeuroImage*, 206:116233, feb 2020b.
- R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. Bullmore. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral cortex*, 15(9):1332–1342, 2005.

- A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, 28(9):3095–3114, 2018.
- G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.
- L. Shahshahani, M. King, C. Nettekoven, R. Ivry, and J. Diedrichsen. Selective recruitment: Evidence for task-dependent gating of inputs to the cerebellum. *bioRxiv*, pages 2023–01, 2023.
- X. Shen, F. Tokoglu, X. Papademetris, and R. T. Constable. Groupwise whole-brain parcellation from resting-state fmri data for network node identification. *Neuroimage*, 82:403–415, 2013.
- S. M. Smith, C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, et al. Resting-state fmri in the human connectome project. *Neuroimage*, 80:144–168, 2013.
- S. M. Smith, A. Hyvärinen, G. Varoquaux, K. L. Miller, and C. F. Beckmann. Group-pca for very large fmri datasets. *Neuroimage*, 101:738–749, 2014.
- Y. Song and D. P. Kingma. How to train your energy-based models. *arXiv preprint arXiv:2101.03288*, 2021.
- A. Sotiras, S. M. Resnick, and C. Davatzikos. Finding imaging patterns of structural covariance via non-negative matrix factorization. *Neuroimage*, 108:1–16, 2015.
- O. Sporns. The human connectome: a complex network. *Annals of the New York Academy of Sciences*, 1224(1):109–125, 2011.
- I. Sutskever and T. Tieleman. On the convergence properties of contrastive divergence. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 789–795. JMLR Workshop and Conference Proceedings, 2010.
- I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR, 2013.
- J. Talairach. 3-dimensional proportional system; an approach to cerebral imaging. co-planar stereotaxic atlas of the human brain. *Thieme*, pages 1–122, 1988a.
- J. Talairach. Co-planar stereotaxic atlas of the human brain. *3-D proportional system: An approach to cerebral imaging*, 1988b.
- P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to data mining*. Pearson Education India, 2016.

- I. Tavor, O. Parker Jones, R. B. Mars, S. M. Smith, T. E. Behrens, and S. Jbabdi. Task-free MRI predicts individual differences in brain activity during task performance. *Science*, 352(6282):216–220, 2016.
- B. Thirion, G. Varoquaux, E. Dohmatob, and J.-B. Poline. Which fmri clustering gives good brain parcellations? *Frontiers in neuroscience*, 8:167, 2014.
- Y. Tian, D. S. Margulies, M. Breakspear, and A. Zalesky. Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nature neuroscience*, 23(11):1421–1432, 2020.
- T. Tieleman. Training restricted boltzmann machines using approximations to the likelihood gradient. In *Proceedings of the 25th international conference on Machine learning*, pages 1064–1071, 2008.
- G. Tononi, O. Sporns, and G. M. Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, 91(11):5033–5037, 1994.
- D. J. Tozer, G. Davies, D. Altmann, D. Miller, and P. Tofts. Correlation of apparent myelin measures obtained in multiple sclerosis patients and controls from magnetization transfer and multicompartamental t2 analysis. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 53(6):1415–1422, 2005.
- N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.
- L. Q. Uddin, K. Supekar, and V. Menon. Typical and atypical development of functional human brain networks: insights from resting-state fmri. *Frontiers in systems neuroscience*, page 21, 2010.
- K. Uğurbil, J. Xu, E. J. Auerbach, S. Moeller, A. T. Vu, J. M. Duarte-Carvajalino, C. Lenglet, X. Wu, S. Schmitter, P. F. Van de Moortele, et al. Pushing spatial and temporal resolution for functional and diffusion mri in the human connectome project. *Neuroimage*, 80:80–104, 2013.
- M. P. Van Den Heuvel and H. E. H. Pol. Exploring the brain network: a review on resting-state fmri functional connectivity. *European neuropsychopharmacology*, 20(8):519–534, 2010.
- D. C. Van Essen and M. F. Glasser. Parcellating cerebral cortex: how invasive animal studies inform noninvasive mapmaking in humans. *Neuron*, 99(4):640–663, 2018.
- D. C. Van Essen, M. F. Glasser, D. L. Dierker, J. Harwell, and T. Coalson. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral cortex*, 22(10):2241–2262, 2012.

- D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.
- D. P. Varikuti, S. Genon, A. Sotiras, H. Schwender, F. Hoffstaedter, K. R. Patil, C. Jockwitz, S. Caspers, S. Moebus, K. Amunts, et al. Evaluation of non-negative matrix factorization of grey matter in age prediction. *Neuroimage*, 173:394–410, 2018.
- U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17:395–416, 2007.
- M. J. Wainwright, M. I. Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.
- D. Wang, R. L. Buckner, M. D. Fox, D. J. Holt, A. J. Holmes, S. Stoecklein, G. Langs, R. Pan, T. Qian, K. Li, et al. Parcellating cortical functional networks in individuals. *Nature neuroscience*, 18(12):1853–1860, 2015.
- K. Wang, M. Liang, L. Wang, L. Tian, X. Zhang, K. Li, and T. Jiang. Altered functional connectivity in early alzheimer’s disease: A resting-state fmri study. *Human brain mapping*, 28(10):967–978, 2007.
- M. Welling and G. E. Hinton. A new learning algorithm for mean field boltzmann machines. In *ICANN’02: Proceedings of the international conference on artificial neural networks*, pages 351–357, 2002.
- G. S. Wig, T. O. Laumann, and S. E. Petersen. An approach for parcellating human cortical areas using resting-state correlations. *Neuroimage*, 93:276–291, 2014.
- F.-Y. Wu. The potts model. *Reviews of modern physics*, 54(1):235, 1982.
- A. Xue, R. Kong, Q. Yang, M. C. Eldaief, P. A. Angeli, L. M. DiNicola, R. M. Braga, R. L. Buckner, and B. T. Yeo. The detailed organization of the human cerebellum estimated by intrinsic functional connectivity within the individual. *Journal of neurophysiology*, 125(2):358–384, 2021.
- B. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106:1125–1165, 2011.
- B. T. Yeo, F. M. Krienen, S. B. Eickhoff, S. N. Yaakub, P. T. Fox, R. L. Buckner, C. L. Asplund, and M. W. Chee. Functional specialization and flexibility in human association cortex. *Cerebral cortex*, 25(10):3654–3672, 2015.
- J. Zhang, D. Xu, H. Cui, T. Zhao, C. Chu, and J. Wang. Group-guided individual functional parcellation of the hippocampus and application to normal aging. *Human brain mapping*, 42(18):5973–5984, 2021.

- D. Zhi and J. Diedrichsen. DCBC toolbox: A Python toolbox for brain parcellation evaluation, May 2021. URL <https://github.com/DiedrichsenLab/DCBC/tree/v1.0.0>.
- D. Zhi, M. King, C. R. Hernandez-Castillo, and J. Diedrichsen. Evaluating brain parcellations using the distance-controlled boundary coefficient. *Human Brain Mapping*, 43(12):3706–3720, 2022.
- D. Zhi, L. Shahshahani, C. Nettekoven, A. L. Pinho, D. Bzdok, and J. Diedrichsen. A hierarchical bayesian brain parcellation framework for fusion of functional imaging datasets. *bioRxiv*, pages 2023–05, 2023.
- K. Zilles, N. Palomero-Gallagher, C. Grefkes, F. Scheperjans, C. Boy, K. Amunts, and A. Schleicher. Architectonics of the human cerebral cortex and transmitter receptor fingerprints: reconciling functional neuroanatomy and neurochemistry. *European Neuropsychopharmacology*, 12(6):587–599, 2002.

Appendix A

Supplementary Materials and Figures for Chapter 3

Parameter estimation of full model

In this section, we provide details of model parameter estimation for the full EM algorithm. The complete expected log-likelihood $\sum_s \langle \log p(\mathbf{Y}^s, \mathbf{U}^s; \boldsymbol{\theta}) \rangle_q$ can be decomposed into expected emission log-likelihood \mathcal{L}_E and expected arrangement log-likelihood \mathcal{L}_A , where $\langle \cdot \rangle_q$ denotes the expectation with respect to distribution q . Similarly, the model parameter θ can be subdivided into θ_E and θ_A and can be estimated within their corresponding models (Methods 3.4.1). This unique model structure yields the following learning EM process:

Emission model E step. Suppose for a single dataset \mathbf{Y}^n is a $S \times N \times P$ tensor for S subjects (S is the number of subjects in \mathbb{S}) of N data observations across P voxels. The brain activation of a voxel for a single subject \mathbf{y}_i^s is a N -long vector. If the task design has repeated measurements of the same M conditions (e.g. in a single imaging run), the user can specify this over a $N \times M$ design matrix X (M is the number of unique task conditions). To account for the situation that \mathbf{y}_i^s consists of multiple partitions, which could be imaging sessions or runs, we used an N -dimensional partition vector to divide N observation into J independent partitions. Therefore, if we combine the data across repeated measurements and different partitions, the resultant data $\tilde{\mathbf{y}}_i^s$ would be a sum of

normalized data in each partition j as,

$$\tilde{\mathbf{y}}_i^s = \sum_j^J \|(\mathbf{X}_j^\top \mathbf{X}_j)^{-1} \mathbf{X}_j^\top \mathbf{y}_{i,j}^s\| \quad (\text{A.1})$$

However, we can also treat the different repetitions as independent observations, meaning that the resultant data is normalized to length 1 across J independent partitions. This is also the case with the Type 1 model, in which the imaging sessions are simply concatenated. Hence, the expected emission likelihood \mathcal{L}_E of a mixture of k -classes vMF distribution in eq. 3.11 is modified and updated at $(t + 1)$ iteration by:

$$\mathcal{L}_E^{(t+1)} = SPJ \sum_k \log c_M(\kappa_k^{(t)}) + \sum_{s \in \mathbb{S}} \sum_i^P \sum_k^K \langle \mathbf{u}_i^s(k) \rangle_q^{(t)} \kappa_k^{(t)} \mathbf{v}_k^{(t)\top} \tilde{\mathbf{y}}_i^s \quad (\text{A.2})$$

As a sufficient statistic, it should be noticed that the resultant summed vectors $\tilde{\mathbf{y}}_i^s$ become a M -dimensional vector but its magnitude is not 1 anymore. Therefore, the normalizing constant will be computed in M -dimensional correspondingly, denoted as $\log c_M(\kappa_k)$.

Arrangement model E step. Expanding eq. 3.10 and 3.7, the expected posterior under the proposal distribution q at $(t + 1)$ iteration are updated as,

$$\langle \mathbf{u}_i^s(k) \rangle_q^{(t+1)} = p(\mathbf{u}_i^s = k | \mathbf{y}_i^s; \theta_A^{(t)}, \theta_E^{(t)}) \quad (\text{A.3})$$

$$= \frac{\exp(\langle \log p(\mathbf{y}_i^s | \mathbf{u}_i^s = k; \theta_E^{(t)}) \rangle_q + \eta_i^k(t))}{\sum_j \exp(\langle \log p(\mathbf{y}_i^s | \mathbf{u}_i^s = j; \theta_E^{(t)}) \rangle_q + \eta_i^j(t))} \quad (\text{A.4})$$

where η_i^k is defined in Methods.

Arrangement model M step. Expanding the expected arrangement log-likelihood in eq.3.10, we obtain the derivatives with respect to the parameters $\theta_A := \{\eta_{i,k}\}$:

$$\frac{\partial \mathcal{L}_A}{\partial \eta_{i,k}} = \frac{\partial \sum_s \sum_i \langle \mathbf{u}_i^s(k) \rangle_q \cdot \eta_{i,k}}{\partial \eta_{i,k}} \quad (\text{A.5})$$

By setting this derivative to zero, we can obtain the following parameter updates:

$$\eta_{i,k}^{(t+1)} = \log \sum_s \langle \mathbf{u}_i^{s(k)} \rangle_q^{(t)} \quad (\text{A.6})$$

Emission model M step. To update the parameters θ_E of the vMF mixture in the M-step, we need to maximize \mathcal{L}_E in respect to the parameters in vMF mixture $\theta_k = \{\mathbf{v}_k, \kappa_k\}$. First, we update the mean direction \mathbf{v}_k , where we get the intuitive update :

$$\mathbf{v}_k^{(t+1)} = \frac{\tilde{\mathbf{v}}_k}{r_k}, \quad \text{where } \tilde{\mathbf{v}}_k = \sum_s \sum_i \langle \mathbf{u}_i^{s(k)} \rangle_q^{(t)} \cdot \tilde{\mathbf{y}}_i^s; \quad r_k = \|\tilde{\mathbf{v}}_k\| \quad (\text{A.7})$$

The updates of the concentration parameters κ_k are more difficult in particular for high dimensional problems, since it involves the inverting ratio of two Bessel functions. Therefore, we here use an approximate solution suggested by [Banerjee et al. \(2005\)](#) and [Hornik and Grün \(2014\)](#). In our specific case, we want to integrate the evidence across $s = 1, \dots, S$ subjects and $i = 1, \dots, P$ voxels, with each subject and voxel may have J_i^s partitions. Under this assumption, we can **(1)** learn a common κ across classes by restricting κ_k to be equal, as:

$$\kappa^{(t+1)} \approx \frac{\bar{r}M - \bar{r}^3}{1 - \bar{r}^2} \quad (\text{A.8})$$

$$\bar{r} = \frac{\sum_k \|\sum_s \sum_i \langle \mathbf{u}_i^{s(k)} \rangle_q^{(t)} \cdot \tilde{\mathbf{y}}_i^s\|}{\sum_s \sum_i J_i^s} \quad (\text{A.9})$$

which is used in Type 1 and Type 2 model learning.

Alternatively, we can **(2)** learn k -class specific kappa κ_k by relaxing the constraint as:

$$\kappa_k^{(t+1)} \approx \frac{\bar{r}_k M - \bar{r}_k^3}{1 - \bar{r}_k^2} \quad (\text{A.10})$$

$$\bar{r}_k = \frac{\|\sum_s^S \sum_i^P \langle \mathbf{u}_i^{s(k)} \rangle_q^{(t)} \cdot \tilde{\mathbf{y}}_i^s\|}{\sum_s^S \sum_i^P \langle \mathbf{u}_i^{s(k)} \rangle_q^{(t)} \cdot J_i^s} \quad (\text{A.11})$$

which will be used as the parameter estimates for the Type 3 regions-specific emission model.

Supplementary Figures

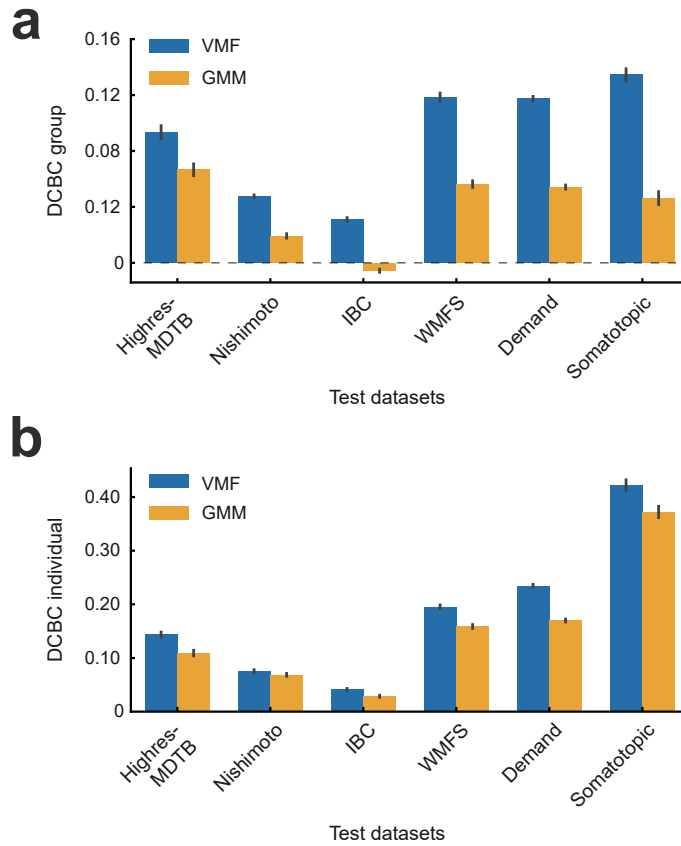


Figure A.1: **Performance comparison between the parcellations derived from Gaussian Mixture Model (GMM) and von Mises-Fisher Mixture model (VMF).** (a) The averaged DCBC value of the group parcellation maps trained by GMM or VMF mixture model across subjects in the test dataset. (b) The averaged DCBC value of the individual parcellation maps trained by GMM or VMF mixture model across subjects in the test dataset.

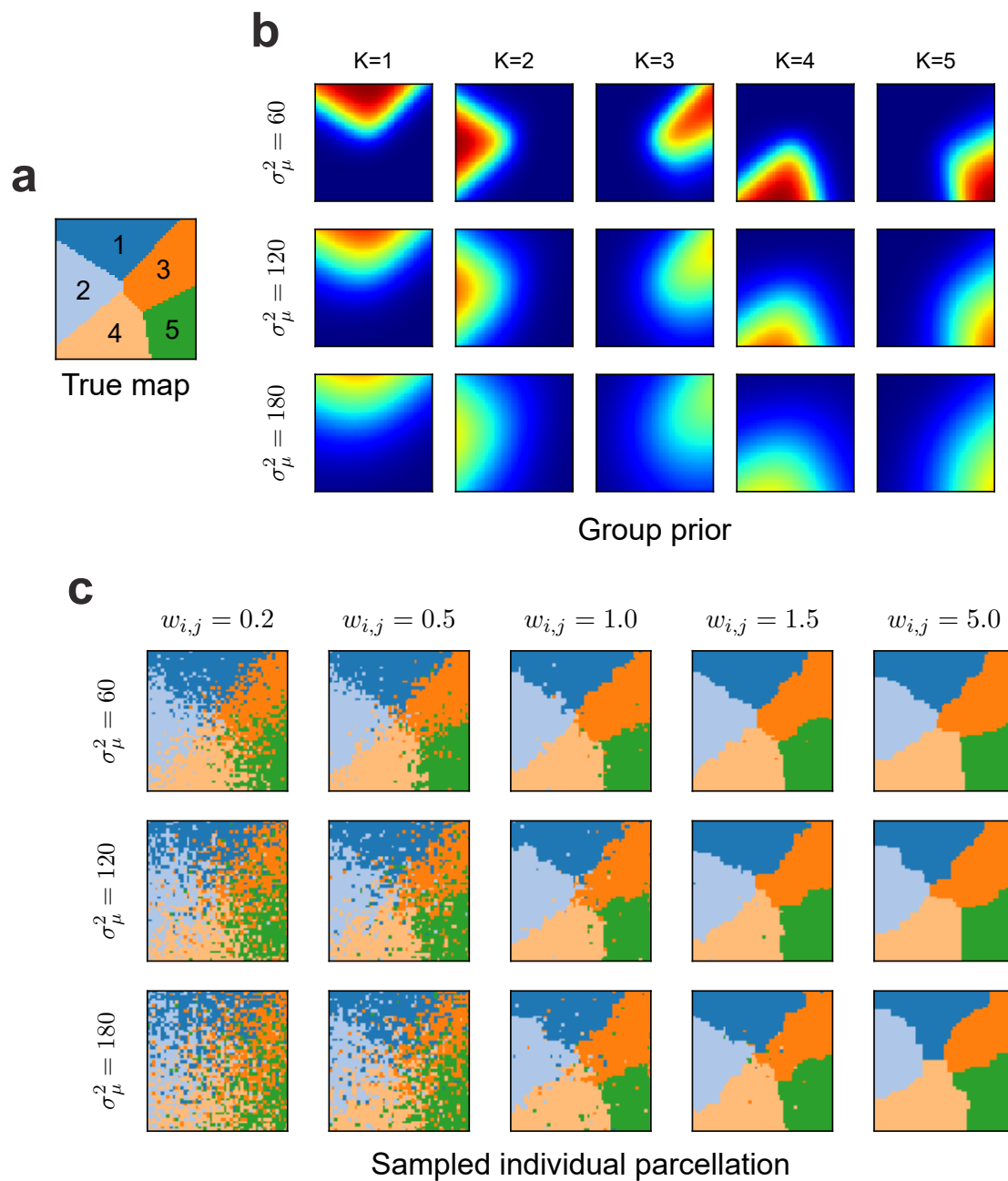


Figure A.2: **The synthetic dataset.** (a) The random true group map with 5 parcels. (b) The group prior controlled by smoothing kernel at different levels for all 5 classes. (c) The example individual parcellation maps generated by different parameters

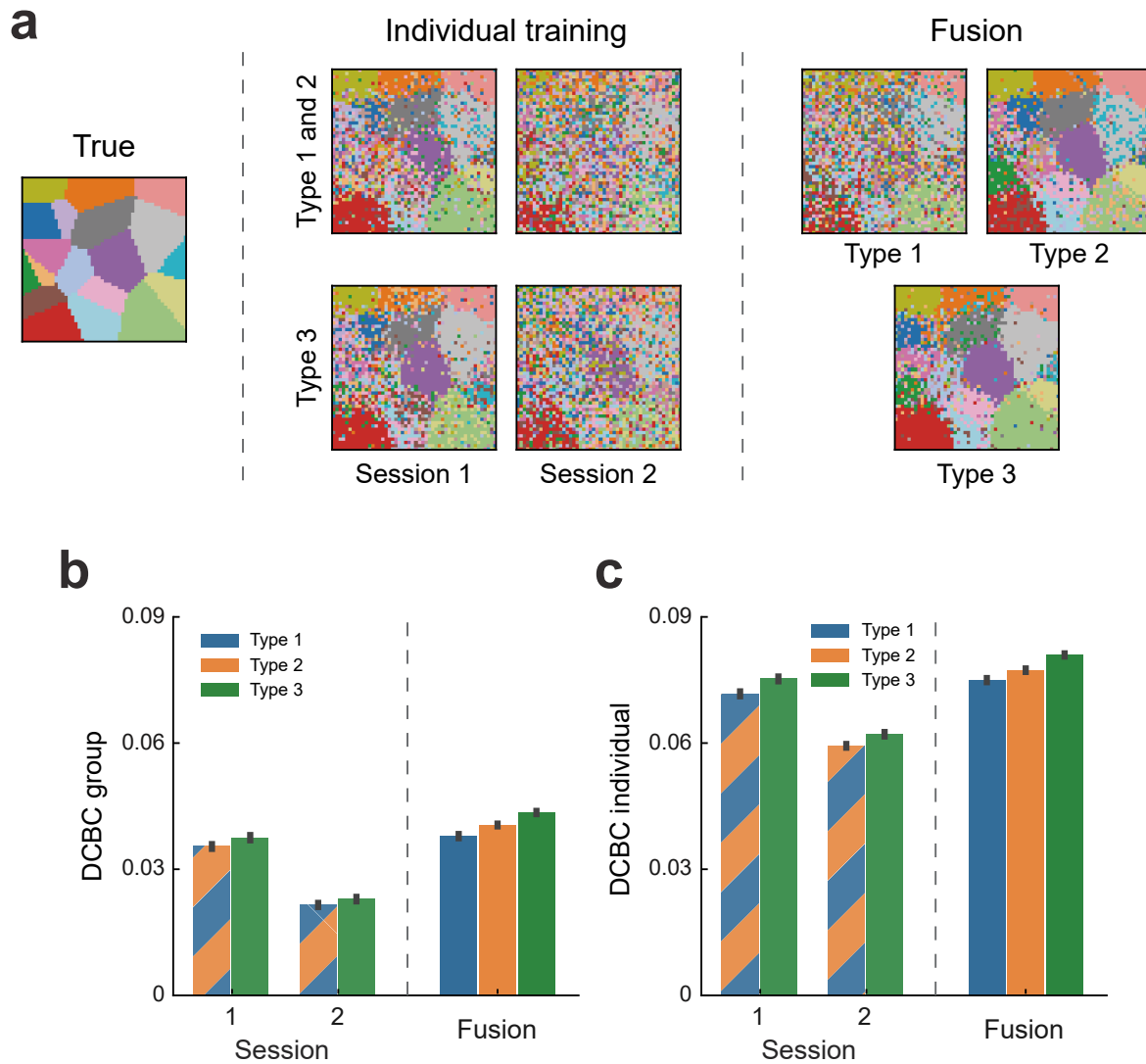


Figure A.3: **Simulation on two synthetic sessions fusion with similar task activation using Type 1, 2, and 3 emission models.** (a) The comparison of model reconstruction performance of group parcellations learned on synthetic session 1 or 2 standalone vs. the ones learned fusion using type 1, 2, or 3 models. (b) The mean DCBC value of the group map learned from session 1 or 2 only or learned by fusion using type 1, 2, or 3 models. (c) The mean DCBC value of individual maps across all participants when learned from session 1 or 2 only or learned by fusion using type 1, 2, or 3 models. Error bars indicate SEM across 100 times simulation.

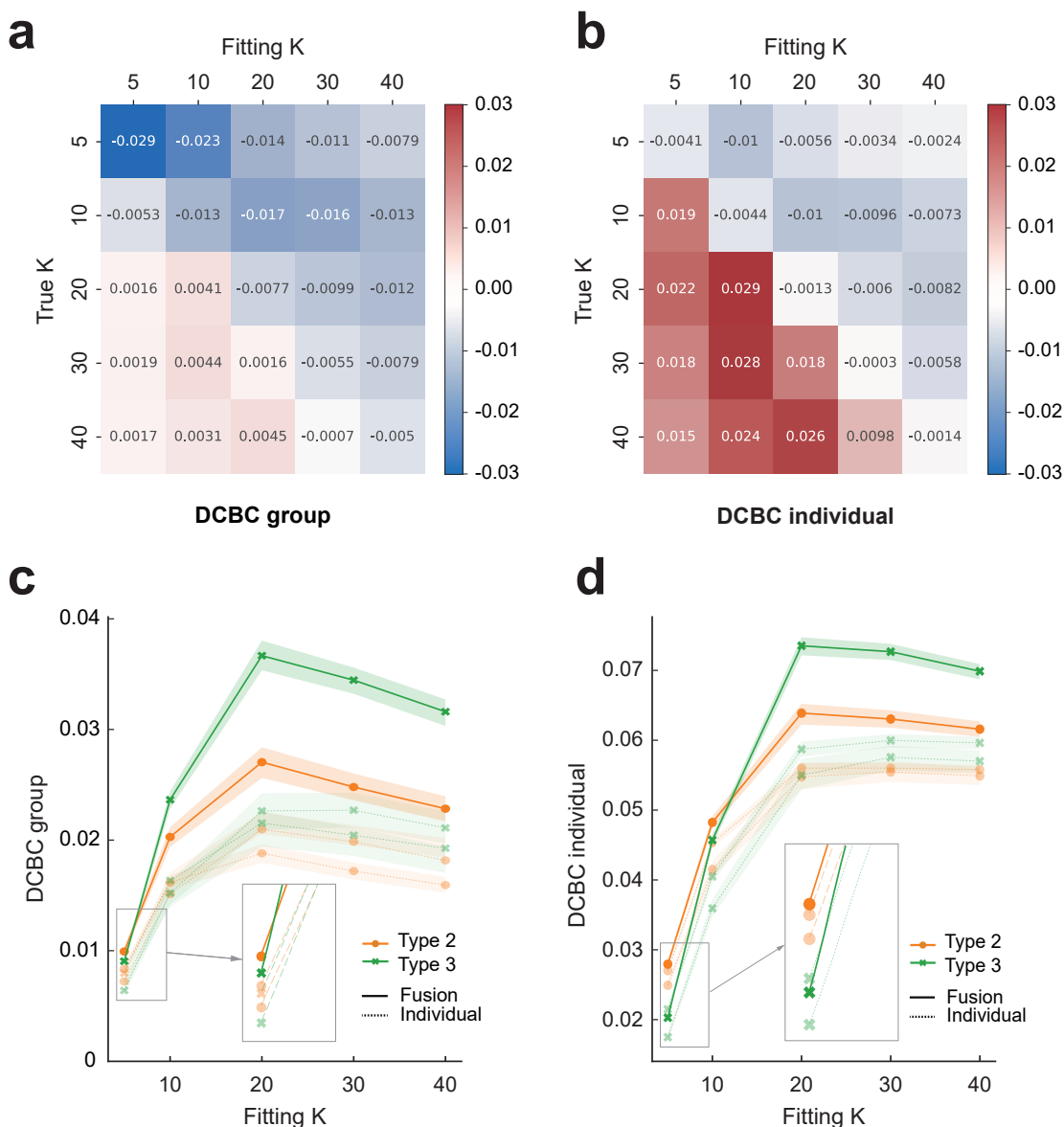


Figure A.4: **Comparing the performance of Type 2 and 3 models when the number of parcels K used for fitting is different from the true K in the simulation.** (a) The difference of the mean DCBC value between the group map trained on a synthetic dataset using Type 2 and Type 3 models with different fitting K and ground true K , which tested on an independent synthetic test set. A positive value on the grid indicates the Type 2 model outperforms the Type 3 model, while negative values mean the opposite. (b) The difference of the mean DCBC value between the individual maps. (c) The mean DCBC value for the group map learned from individual synthetic datasets only or learned by fusion using the Type 2 or 3 models for $K = 5$ to $K = 40$ when the true $K = 20$. The error shade indicates the standard error across 100 simulations. (d) The mean DCBC value for the individual maps.

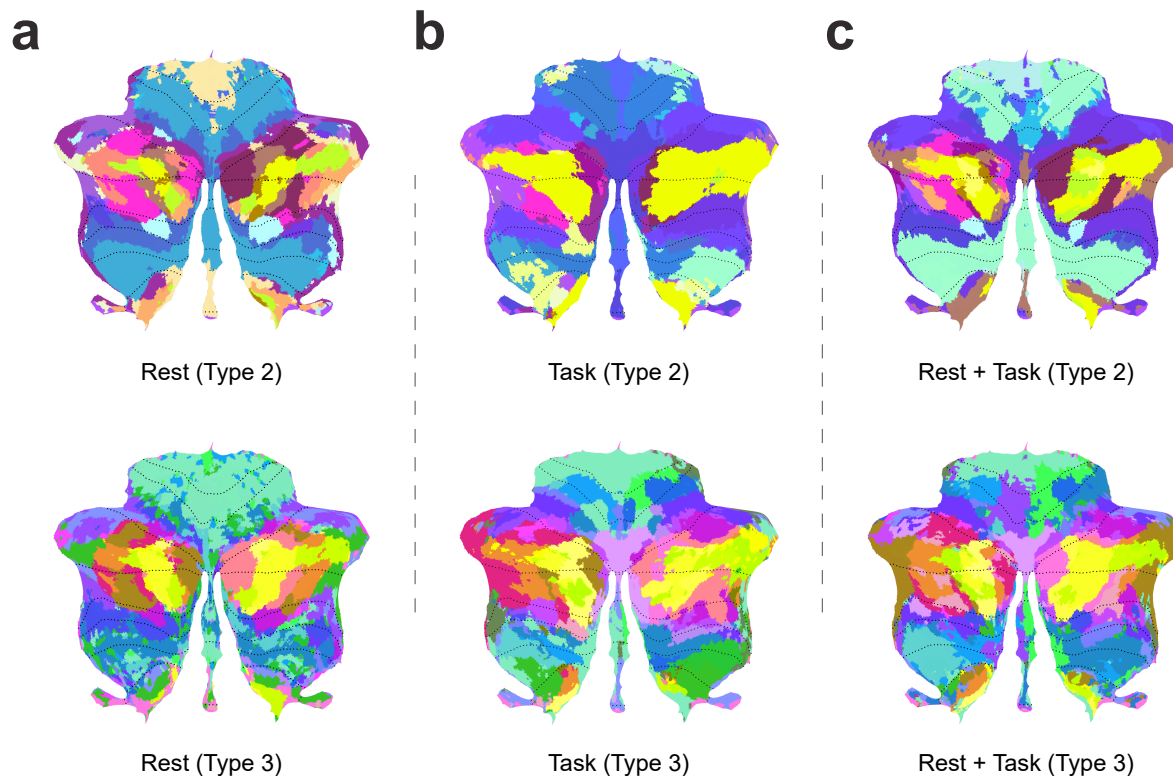


Figure A.5: **The visualization of the learned group maps ($K = 34$).** (a) The maps were trained on a pure resting-state dataset *HCP-Unrelated 100* using the Type 2 or 3 fusion model. (b) The maps were purely trained on task-based datasets using the Type 2 or 3 fusion model. The task datasets are *MDTB*, *Highres-MDTB*, *Nakai&Nishimoto*, *IBC*, *WMFS*, *Demand*, *Somatotopic*. (c) The maps were trained on the combination of resting-state and all task-based datasets. The colors for the parcels are aligned in each type of model, where two similar colors in RGB space indicate the two parcels have similar task activation responses on average.

Appendix B

Supplementary Materials and Figures for Chapter 4

Proof of the expected weighted cosine error is equivalent to $1 - R^2$

Weighting the error by the length of the vector effectively calculates the squared error between \mathbf{y}_i and the prediction scaled to the amplitude of the data ($\mathbf{v}_k \|\mathbf{y}_i\|$). For simplicity, we use \mathbf{v}_k to represent the most likely predicted mean direction $\mathbf{v}_{\text{argmax}_k}$ for each voxel in the following proof. $1 - R^2$ between \mathbf{y}_i and the prediction scaled to the amplitude of the data ($\mathbf{v}_k \|\mathbf{y}_i\|$) is defined as:

$$\begin{aligned} 1 - R^2 &= \frac{RSS}{TSS} \\ &= \frac{1}{\sum_i \|\mathbf{y}_i\|^2} \sum_i (\mathbf{y}_i - \mathbf{v}_k \|\mathbf{y}_i\|)^2 \\ &= \frac{1}{\sum_i \|\mathbf{y}_i\|^2} \sum_i (\mathbf{y}_i^\top \mathbf{y}_i - 2\mathbf{y}_i^\top \mathbf{v}_k \|\mathbf{y}_i\| + \mathbf{v}_k^\top \mathbf{v}_k \|\mathbf{y}_i\|^2) \\ &= \frac{1}{\sum_i \|\mathbf{y}_i\|^2} \sum_i (\|\mathbf{y}_i\|^2 - 2\mathbf{y}_i^\top \mathbf{v}_k \|\mathbf{y}_i\| + \|\mathbf{y}_i\|^2) \\ &= \frac{2}{\sum_i \|\mathbf{y}_i\|^2} \sum_i (\|\mathbf{y}_i\|^2 - \mathbf{y}_i^\top \mathbf{v}_k \|\mathbf{y}_i\|) \end{aligned} \tag{B.1}$$

By equation 4.19, we can see that $1 - R^2 = 2\bar{\epsilon}_{Acosine}$, and similarly we can easily proof below equation:

$$\langle \bar{\epsilon}_{MSE} \rangle_q = \frac{1}{\sum_i \|\mathbf{y}_i\|^2} \sum_i \sum_k \langle \hat{\mathbf{u}}_i^{(k)} \rangle (\mathbf{y}_i - \mathbf{v}_k \|\mathbf{y}_i\|)^2 = 2\langle \bar{\epsilon}_{Acosine} \rangle_q \quad (\text{B.2})$$

where $\langle \hat{\mathbf{u}}_i^{(k)} \rangle$ is the inferred expectation on the training data using the fitted model.

Supplementary Figures

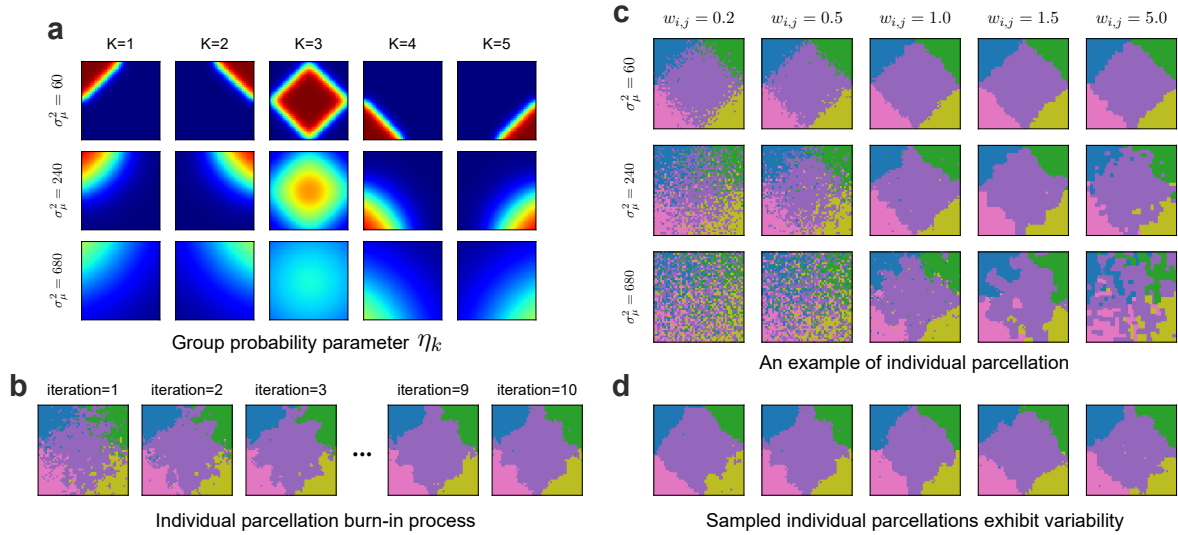


Figure B.1: **The synthetic dataset.** (a) The group probability prior is controlled by smoothing kernel at different levels for all 5 parcels. (b) The burn-in process for generating the individual parcellation. (c) The example individual parcellation maps generated by different parameters

Curriculum Vitae

Name: Da Zhi

Post-Secondary Education and Degrees: University of Western Ontario
London, ON
2018 - 2023 Ph.D.

University of Ottawa
Ottawa, ON
2015 - 2018 M.Sc.

Tianjin University
Tianjin, China
2011 - 2015 B.Eng.

Honours and Awards: Western Graduate Research Scholarship (WGRS)
2018-2022

Related Work Experience: Lecturer
The University of Western Ontario
2020 - 2021

Teaching Assistant
The University of Western Ontario
2018 - 2023

Teaching Assistant
University of Ottawa
2018

Publications:

Zhi, D., Shahshahani, L., Nettekoven, C., Pinho, A. L., Bzdok, D., Diedrichsen, J. (2023). A hierarchical Bayesian brain parcellation framework for fusion of functional

imaging datasets. *bioRxiv*, 2023-05.

Zhi, D., King, M., Hernandez-Castillo, C. R., Diedrichsen, J. (2022). Evaluating brain parcellations using the distance-controlled boundary coefficient. *Human Brain Mapping*, 43(12), 3706–3720.

Zhi, D., de Oliveira, T. E. A., da Fonseca, V. P., Petriu, E. M. (2018, June). “Teaching a Robot Sign Language using Vision-Based Hand Gesture Recognition.” In 2018 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), pp. 1-6. IEEE

da Fonseca, V. P., Kucherhan, D. J., de Oliveira, T. E. A., **Zhi, D.**, Petriu, E. M. (2017). “Fuzzy controlled object manipulation using a three-fingered robotic hand,” In Systems Conference (SysCon), 2017 Annual IEEE International, pp. 1-6. IEEE

Selected Oral Presentations:

Zhi, D., Shahshahani, L., Nettekoven, C., Pinho, A. L., Bzdok, D., Diedrichsen, J. (2023). Learning brain parcellation models across functional imaging datasets. Oral presentation at Organization for Human Brain Mapping, Montreal, Canada.

Zhi, D.. (2022). Learning human brain organization across functional imaging datasets. Oral presentation at the University of Western Ontario Research in Computer Science (UWORCS), London, Canada.

Poster Presentations:

Zhi, D., Shahshahani, L., Nettekoven, C., Pinho, A. L., Bzdok, D., Diedrichsen, J. (2023). Learning brain parcellation models across functional imaging datasets. Poster presentation at Organization for Human Brain Mapping, Montreal, Canada.

Zhi, D., King, M., Hernandez-Castillo, C. R., Ivry, R. B. Diedrichsen, J. (2020). New Evaluation criterion for functional brain parcellation methods using a multi-domain task battery. Poster presentation at Organization for Human Brain Mapping, Virtual.

Diedrichsen, J., **Zhi, D.**, King, M., Hernandez-Castillo, C. R., Ivry, R. B. (2020). An online atlas of the human cerebellum. Poster presentation at Organization for Human Brain Mapping, Virtual.

Zhi, D., King, M., Hernandez-Castillo, C. R., Ivry, R. B. Diedrichsen, J. (2019). Evaluation of functional brain parcellation methods using a multi-domain task battery. Poster presentation at Society for Neuroscience (SfN), Chicago, USA.