4-24-2023 11:00 AM

Estimating the spatial correlation structure of measurement error in functional magnetic resonance imaging (fMRI) to improve multivariate inference

Lingling Lin, *The University of Western Ontario*

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

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Statistics and Actuarial Sciences

Abstract

Multi-voxel pattern analysis (MVPA) provides a powerful framework for making statistical inferences on the information present in brain activity patterns as measured by functional magnetic resonance imaging (fMRI). Many recent studies suggest that MVPA performance benefits from taking into account the spatial voxel-to-voxel correlations in the measurement noise. However, estimating these noise correlations is challenging due to the limited data points and large voxel counts. To address this issue, it is common practice to shrink the empirical correlation estimate towards its identity matrix, which biases the estimate towards the incorrect assumption that voxels are independent. We therefore propose an anatomically-informed model of measurement noise in fMRI, which takes into account the distances of voxels in the measurement volume, their distance on the cortical sheet, and the depth at which they sample the cortex. Our model can predict the noise-correlation structure in new participants and datasets. It improves the noise correlation estimate when used as a shrinkage target, thereby also potentially improving statistical inferences in MVPA.

Keywords: Time series analysis, Representation similarity analysis, noise correlation, multivariate pattern analysis, effective degrees of freedom, regularization, reliability, distance estimates

Summary for Lay Audience

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive imaging technique that allows researchers to measure the activity in the human brain. To compare the fine-grained pattern of brain activation, researchers employ multivariate analysis methods to make inferences on the activity observed in groups of voxels, the 3D analog of image pixels in a brain image. In these analyses, they often assume that the measurement error is independent across voxels. However, that is not the case; fMRI data has a strong and reliable correlation structure across voxels. Mounting evidence shows that, based on this incorrect assumption, researchers are likely to arrive at incorrect decisions. Therefore, the goal of this project is to build a model of the spatial noise correlation structure in fMRI data and use it to improve multivariate inference. We construct potential noise correlation models using different aspects of the anatomical information. For example, two voxels that are close to each other are more correlated than two voxels that are far apart. The spatial distance in the volume hence is one of the anatomical factors that we consider. We also consider the distance along the cortical surface (i.e., taking into account the individual brain folding structure) and whether the voxel measured superficial or deep aspects of the brain. Our results show that the use of our anatomically-informed noise correlation model can lead to better inference than assuming the independence of voxels.

Co-Authorship Statement

I was primarily responsible for the research presented in this dissertation. The concepts presented in all chapters were mainly written by me under the guidance of Dr. Diedrichsen. Six fMRI datasets were preprocessed by Mahdiyar Shahbazi, including calculating anatomical data. Chapter 2.7 fMRI datasets descriptions were written by me and Mahdiyar Shahbazi. Dr. Diedrichsen and Mahdiyar Shahbazi have read this statement and agree with it.

Acknowlegements

There are many people who have provided guidance, assistance, mentorship, and encouragement to the completion of this study. I would like to take this opportunity to name and thank these people. First, I would like many thanks to my supervisor, Dr. Jörn Diedrichsen. When I started my project, it was a big challenge for me. Jörn knew that and provided me with continuous support and patience. He provided many valuable opportunities to help me learn more about neuroscience. I appreciate that you took me to the workshop at Columbia University this January. I met many talented experts in the neuroscience field and learnt much from them. I am very proud of the work we have done during the past two years. Thank you, Jörn, not only for guiding me during this research but also for helping me become a better person.

I would like to next thank my project partner (another "supervisor"), Mahdiyar Shahbazi, for sharing neuroscience resources and guiding me when I face hard problems. As well, thank you for your insightful feedback throughout my dissertation and the encouragement you give me after each meeting. I would also like to thank other lab members: Da Zhi, Ladan Shahshahani, Caro Nettekoven, Jingyu Cui, Giacomo Ariani and Mehrdad Kashefi. You all give me valuable advice on my research and useful feedback for my thesis.

I am also grateful to my thesis examination committee, Dr. Yi, Dr. He, and Dr. Khan, for their help in shaping my research through its various phases. Again, thank you for being one of my thesis examiners and sharing your considerate suggestions.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

1.1 General problem statement

The brain is a complex organ that gives rise to our thoughts, memories, emotions, and movements. As we engage in different tasks, the activity of neurons in the brain changes. Functional magnetic resonance imaging (fMRI), a non-invasive and painless technique, enables us to measure this brain activity by detecting the changes associated with blood flow (see below). Our brain consists of different regions and each of them has a specific functional contribution across diverse tasks. What the computational role of each region is an important open question in system neuroscience. One approach to solving this question is to study exactly how the activity patterns in different regions change across different tasks or experimental conditions. We can interpret neural activity patterns in each region as representing some aspect of the world (Shea [2018]). If we can determine which features of the world are represented in the region, we can gain insight into the computational function of that region. Researchers therefore built brain-computational models that could perform complex tasks (Abbott [2008]) and then compared the representations arising in the models with the brain activity patterns observed in different regions of the brain. But how they can assess whether a representation found in the model is a good match (Kriegeskorte and Diedrichsen [2016])? This is a multivariate inferences prob-

lem, in which we want to compare a set of vectors that describe different states of the world to the multivariate activity patterns observed in the brain. To do so, the research community has adopted three different related approaches: Representational similarity analysis (RSA), Pattern Component Modelling (PCM), and Encoding models (Diedrichsen et al. [2017]). All three approaches ultimately assess a simple prediction: If two task conditions are similar in terms of their model features, they should also be associated with similar activity patterns in the region that represents these features. However, one critical problem in all three approaches is that the measurement noise in fMRI is correlated across different brain locations (or voxels, see below), so when making multivariate inferences, we cannot assume independent and identically distributed noise. This thesis aims to address the spatial dependence of the measurement noise by building a better noise correlations model, ultimately improving the statistical inference in multivariate analysis.

In the next three parts of the introduction, I provide some basic backgrounds of fMRI as well as cerebrum structure, review one specific form of multivariate fMRI analysis, and demonstrate the problem of the spatial noise correlation structure in this analysis.

1.2 Background information: Functional Magnetic Resonance Imaging and cerebrum structure

The brain comprises three parts: the cerebrum, brainstem, and cerebellum. Here, we focus on the cerebrum, as it is the area of most interest for representational fMRI studies. The basic functional unit of the brain is called the neuron. It plays a role in sending and receiving the electrical impulses that communicate messages about sensation, perception, and cognition throughout the brain. There are roughly 100 billion neurons in the brain (Harms and Elias [2013]). The cerebrum comprises two main components: gray matter (the cerebral cortex) and white matter (Fig. 1.1). Gray matter contains the cell bodies of the neurons - this is where information processing happens. The cerebral gray matter forms a 2-dimensional sheet cortex

that is folded in a complex way so that it can be packed into the skull (Harms and Elias [2013]). The grooves are referred to as sulci and the rounded ridges (or elevations) as gyri (Fig. 1.1). White matter mainly consists of myelinated axons, which allow nerve signals to be transmitted between neurons (Mackenzie [2019]). Hence, the main role of white matter is to provide communication between different gray matter areas. The brain is embedded in cerebrospinal fluid (CSF) (Fig. 1.1), which fills the cavities between brain tissue and the skull. Blood is supplied to the cerebrum through vessels that mainly run along the outer surface of the brain, i.e., between the gray matter and the CSF.

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive brain imaging technique that detects neural activity in the working human brain by measuring changes in blood flow over thousands of spatial brain locations. Each element in the fMRI image is a "voxel", the 3-D analog of a pixel (Poldrack et al. [2011]). Each one represents a tidy cube of brain tissue (typically $2 \times 2 \times 2mm$). The fMRI signal therefore provides an activity measure aggregated over hundreds of thousands of neurons. A typical fMRI scan contains (depending on resolution and field-of-view) 50,000 - 200,000 voxels, covering the entire human brain. FMRI relies on the fact that, when the neuron in the area of the brain is activated (or the area of the brain is in use), the amount of blood that flows to the area increases. Interestingly, neuron activity triggers a much larger change in blood flow than is strictly needed, such that the blood becomes ultimately more oxygenated after neural activity increases (Medicine [2023]). The oxygenation level of the blood changes the magnetic properties of hemoglobin, which can be picked by the MRI scanner. The changes in voxels brightness (i.e., the intensity of the signal received by the scanner) are therefore referred to as the blood oxygenation level dependent (BOLD) signal (Poldrack et al. [2011]). Blood flow to a particular region of the brain is significantly sensitive to underlying changes in neural activity. That is, when neural activity increases in a certain region of the brain, the signal also increases by a small amount. This signal change makes it possible to map activity changes related to different experimental tasks.

However, the BOLD signal is relatively weak, and there are many sources of measurement

noise. The main sources of noise include thermal noise, physiological noise, head movements, and fluctuations in neural activity (Liu [2016]). Thermal noise arises from the sensors of the MRI system itself. The receiver of an MRI machine picks up very small fluctuations in the magnetic field. Because the sensors have to work at room temperature, there are random fluctuations induced by the random movement of electrons. Thermal noise is uncorrelated over time (i.e., white noise), but is spatially correlated due to the geometrical properties of modern multi-channel MRI systems. Hemodynamic cardiac and respiratory-cycle fluctuations, such as changes in the rate of blood flow, blood volume over time, and heartbeat, are sources of physiological noise. Physiological noise in fMRI data has been demonstrated to depend on the vascularization of the tissue and vessel size (Liu et al. [2006]), thus it is significantly greater in cortical gray matter regions than in white matter areas (Bodurka et al. [2007]). Head movement in the scanner can only be partially corrected by motion realignment after acquisition. It leads to correlations of measurement noise across the entire brain. These three noises are non-neuronal in nature and are therefore usually considered measurement noise. The response to a stimulus for a subject often changes over time and tasks, and this change leads to variations in trial-to-trial BOLD activity within a participant that is usually unexplained by the experimental design. These neural sources of variability are spatially correlated such that functionally connected regions of the neocortex are more correlated than two regions that do not communicate with each other (Biswal et al. [1995]). Therefore, neuronal sources of measurement noise should introduce a spatial correlation structure that is dictated by the 2D sheet of the neocortex.

1.3 Multivariate fMRI analysis

Originally, most fMRI experiments investigated whether the BOLD signal in a specific brain location increased (or decreased) in response to a stimulus or a task. Statistical inference was performed using the general linear model (GLM) to reveal activated brain areas by searching for voxel time series that match the predicted BOLD signal (Poldrack et al. [2011]). This ap-

proach was performed at each voxel separately and is therefore referred to as "voxel-level" or "mass-univariate" inference (Friston et al. [1994b]). Statistical inference ultimately identified the brain regions that systematically increased activity in response to a specific task or stimulus. In contrast, instead of looking at each voxel individually, or averaging varying signals across voxels within a region of interest (ROI), many researchers are employing analyses that consider patterns of responses across multiple voxels. In general, this class of techniques has become known as multi-voxel pattern analysis (MVPA). MVPA allows researchers to test whether distributed patterns of BOLD activation across multiple voxels differ between experimental conditions (Norman et al. [2006]; Davis et al. [2014]). Since this analysis makes use of activity patterns across voxels, it can detect a broader class of task-related effects than the traditional mass-univariate approach. For example, if individual voxels have different tuning functions over experimental variables, then the cortical area can contain the information regarding the variable even though the mean activation level of the region may be perfectly matched across conditions. In this case, multivariate analysis, MVPA, can detect representations that may not be characterized by the mass-univariate analysis (Diedrichsen et al. [2013]).

There are a number of MVPA approaches, including classification (Cox and Savoy [2003]), encoding models, PCM, and RSA (Kriegeskorte et al. [2008]). Here, we only focus on RSA. Instead of looking directly at the neural activation patterns elicited by the different classes of experimental conditions, RSA examines the relative similarity of the patterns across conditions in small selected brain regions. RSA achieves this by first calculating the distances between estimated activity patterns across all voxels of the regions. The distance between each possible pair of conditions is then arranged into a representational dissimilarity matrix (RDM). In this way, RSA characterizes for which conditions a region is in a similar, and for which conditions it is in a very different state. The empirical RDM is then compared to the RDM predicted by each brain-computational model to test which model best describes the information encoded in a particular brain region.

1.4 Motivation

A critical problem for using RSA (or any multivariate fMRI analysis methods) to draw inferences is that the noise is not independent across voxels. Instead, the covariance of the noise across voxels has a strong and reliable structure. That is, the noise of neighboring voxels is strongly correlated (Fig. 1.2A). The structure of the covariance matrix is also stable across different experimental runs for the same subject (Fig. 1.2). There are two options to address the spatial dependence of the measurement noise. One option is to simply ignore this issue and assume that the noise is independent across voxels after suppressing the high error variance of voxels by its standard deviation from the residuals of the regression model. This procedure is also called "univariate prewhitening". However, the independence assumption is usually violated (de Zwart et al. [2008]; Arcaro et al. [2015];Rosenbaum et al. [2017]); the strong and reliable spatial dependence across nearby voxels leads to sub-optimal or erroneous inferences in multivariate analysis (Domingos and Pazzani [1997]; Kuncheva [2006]; Wald and Polimeni [2017]).

An alternative option is to take into account the noise correlation structure between voxels by prewhitening the data using an estimate of the noise correlation matrix across voxels. This "multivariate prewhitening" requires the inversion of the estimated correlation matrix, which can become unstable when the number of voxels in the residuals is close to or even larger than the number of time points (i.e., $T \leq P$), which causes the estimated correlation to be rank-deficient. To address this issue, Ledoit and Wolf [2004] proposed to regularize the correlation estimates towards its diagonal (the identity matrix in this case) with the optimal shrinkage parameter $0 \leq \lambda \leq 1$. The optimal parameter explicitly minimizes the mean squared error of the resultant correlation estimator. Most multivariate analysis methods use such a regularized estimate as the noise correlation estimate to prewhiten the data. However, the shrinkage target, the identity matrix, assumes that the noise is independent across voxels, which is not the case. This raises the question of whether we can develop a model of the fMRI spatial correlations structure, which could act as a more appropriate shrinkage target (or prior) for the regularization of

the correlation estimator.

1.5 About this thesis

Therefore, the first goal of this present study is to construct better anatomically-informed noise correlation models. We first considered to what degree the spatial distance in the volume (3d-distance, Fig. 1.1A), the spatial distance along the cortical surface (2d-distance, Fig. 1.1A), and whether the voxel samples superficial or deep aspects of the cortical sheet (cortical depth, Fig. 1.1B) determine the correlation between two voxels. We then used these anatomical characteristics of voxel pairs to develop several noise correlation models and compared these models based on their effective number of voxels. Next, we determined how to estimate the model parameters by comparing the inference in the best model based on the parameters estimated using cross-validation within the individual, across individuals, or across datasets. In the second part of the thesis, I assessed the impact of the developed model on multivariate inference. I therefore compared the statistical inference between regularizing the empirical correlation matrix towards our best model and regularizing towards the identity matrix with the shrinkage factor.

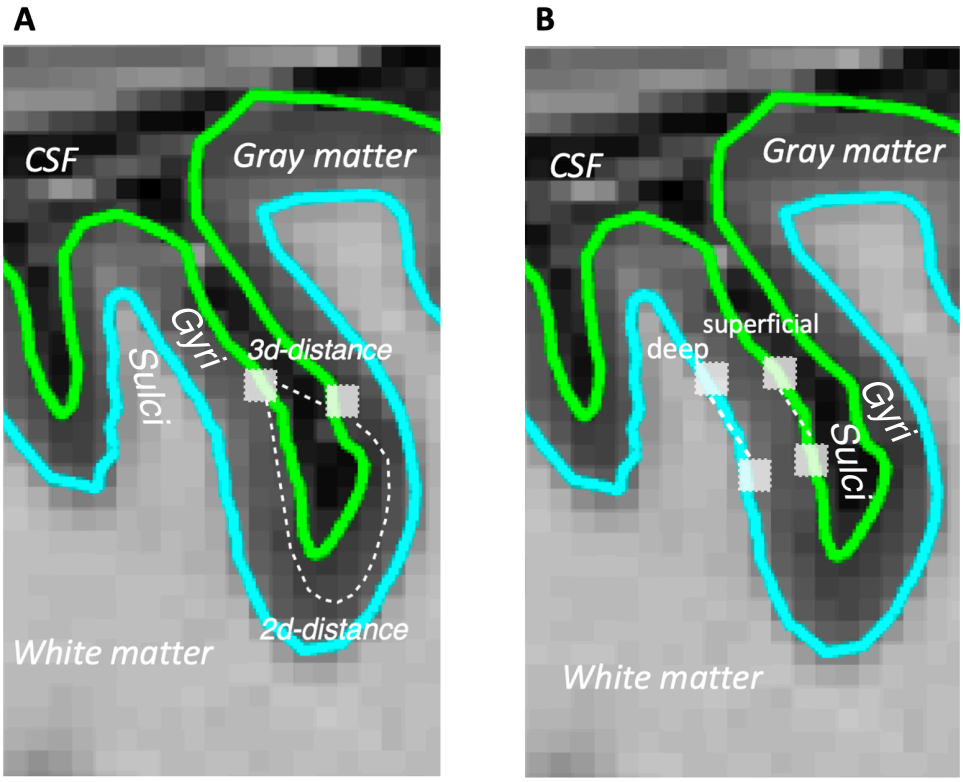

Figure 1.1: A cut through the human brain: a part of the brain's cerebral cortex with gray matter, white matter, and cerebrospinal fluid (CSF). *A*: 3d-distance and 2d-distance between the voxel pair. *B*: Voxel pairs in different depth levels: either both are superficial or deep. Both plots show sulci, the grooves, and gyri, the folds or ridges.

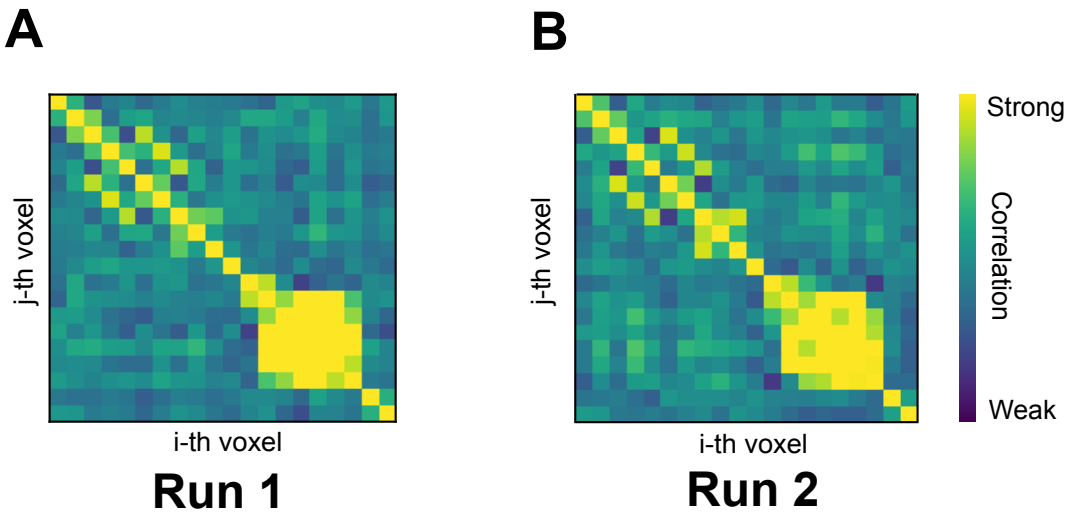

Figure 1.2: Empirical covariance matrix between each pair of voxels for the same subject but different experimental runs.

Chapter 2

Materials and methods

2.1 FMRI datasets

We used datasets from six different fMRI experiments for this project. All experiments differed considerably in task paradigm (length and repetition of each trial type and task ranging from cognitive to motor domain), amount of data (the number of participants and functional runs), scanner parameters (repetition time (TR), and voxel size), and functional regions considered. General information for each dataset is shown in Fig. 2.1.

Dataset 1: MDTB

The full study is described in King et al. [2019]. Twenty-four healthy right-handed participants were engaged in a diverse set of motor (such as finger press) and cognitive (such as working memory and theory of mind) tasks. All fMRI data were acquired on a 3T Siemens Prisma with a 32-channel head coil. Each participant's data contained 8 functional runs of 598 volumes using a 2D echo-planar imaging sequence (TR=1s) at a voxel size of 3 mm isotropic. For this study, we used the data from the first session of this experiment. Each block consisted of 29 trials (17 trial types, each repeated 1 to 3 times) and each was 30 s long. We analyzed the data from 8 ROIs located in the left hemisphere: the primary motor cortex (M1), the primary sensory cortex (S1), the dorsal premotor area (PMd), the ventral premotor area (PMv), the sup-

plementary motor area (SMA), the primary visual cortex (V1), anterior and posterior portion of the superior parietal lobule (SPLa & SPLp). The average number of voxels across ROIs was 451 (standard deviation 161 voxels).

Dataset 2: Online Planning

In this unpublished experiment, twenty-one participants underwent scanning while performing either single finger presses or chords of three simultaneous finger presses. Imaging data were acquired on a 7T Siemens Megnetom MRI scanner with a 32-channel head coil. 10 functional runs of 366 volumes, each using a 2D echo-planner imaging sequence (TR = 1.0 s) at 2.3mm isotropic resolution voxel size, were recorded for each participant. Each run consisted of 192 trials (96 with 750 ms and 96 with 1.3 s stimulus time) of 10 trial types; each repeated 15-20 times. We used the data from 8 ROIs (M1, S1, PMd, PMv, SMA, SPLa, SPLp, V1) in the left hemisphere (contra-lateral to the performing hand)

Dataset 3: Motor Sequence

The full study is described in Wiestler et al. [2014]. Forty-two healthy and right-handed participants underwent a scanning session while performing 4 finger sequences with their right hand and 4 finger sequences with their left hand. Finger presses were executed against an MRI-compatible keyboard comprised of 10 piano-style keys and measured by a force transducer mounted underneath each key. Imaging data were acquired using a 3T Siemens Trio MRI scanner with a 32-channel head coil. Eight functional runs of 156 volumes per run, each using a 2D echo-planar sequence (TR = 2.72 s) at 2.3mm isotropic resolution voxel size were recorded for each subject. Each run consisted of 3 repetitions of each sequence-hand combination (24 trials), with three executions of each sequence in each trial (13.5 s). We used the data from M1, S1, PMd, PMv, SMA, SPLa, and SPLp of the left hemisphere.

Datatset 4: Passive Pattern

The full study is described in Arbuckle et al. [2022]. The focus of this study was located at M1 and S1 of the left hemisphere (contra-lateral to the engaged hand). Ten participants underwent scanning while all 31 combinations of the 5 fingers of their right hand were stimulated.

Imaging data were acquired on a Siemens 7T Magnetom with a 32-channel head coil. Participants completed 11 total functional runs, yielding 682 total trials (31 combinations x 2 repeats x 11 runs), across two separate scanning days, with six runs on the first day and five runs on the second day. Each run was 614s. During each run, 410 functional images were obtained using a multiband 2D echo-planar imaging sequence (TR =1.5s,) with a voxel size of 1.4mm isotropic.

Datatset 5: Cognition

The full study is described in Nakai and Nishimoto [2020]. Six healthy participants performed 104 cognitive tasks. During the first session, subjects were engaged in 12 training runs of 281 volumes per run. The number of repetitions for each task was four, except for one task (FeedbackPos) that repeated 66 times. Each run contained 77-83 trials with a duration of 6-12 *s* per trial. To ensure all runs had the same length, a 2-*s* feedback for the preceding task was presented 9–13 times per run. The duration of a single run was 556 *s*. Imaging data were acquired using a 3 T Siemens Trio with a 32-channel head coil. 72 interleaved axial slices were scanned using a T2-weighted gradient echo multiband echoplanar imaging sequence (TR = 2s, TE = 30ms, 2.0mm thick without a gap) at 2.0 mm isotropic resolution voxel size. Similarly, we used data from M1, S1, PMd, PMv, SMA, V1, SPLa, and SPLp of the left hemisphere.

Datatset 6: Working memory

The full study is described in Shahshahani et al. [2023]. Sixteen participants underwent a scanning session while they were performing a working memory task (forward and backward remembering of numbers in a sequence with different lengths). Each run, therefore, consisted of 12 conditions and each repeated for five 13-second-long trials (6 seconds for encoding and 7 seconds for retrieval). MRI data were acquired on a 3T Siemens Prisma. Five functional runs of 409 volumes (412 volumes in total but 3 volumes are discarded because they are dummy scans), each using an echo-planar imaging sequence (TR = 1 s, TE = 30 ms) at voxel 2.5×2.5×3 mm were recorded for each participant. We used the data from M1, S1, PMd, PMv, SMA, V1, SPLa, and SPLp of the left hemisphere.

	Data set (W)					
	1 (24)	**2** (21)	**3** (42)	**4** (10)	**5** (6)	**6** (16)
Study type	motor/cogniton /social/affecive	motor	motor	motor	cognition	working memory
Number of subjects	24	21	42	10	6	16
Number of ROIs	8	8	7	2	8	8
Averaged number of voxels across ROIs	456	628	523	1168	836	451
Number of timepoints per run	598	366	156	408	281	409
Number of runs	16	10	8	11	12	5
Number of conditions	29	11	8	32	104	12

Table 2.1: Summary of six fMRI datasets used in the analysis.

2.2 Problem Statement

Let $\boldsymbol{Y} = [\boldsymbol{y}_1, ..., \boldsymbol{y}_p] \in \mathbb{R}^{T \times p}$ be the fMRI time series data (i.e., BOLD response measurements) that comprises the activation patterns over T time courses and p voxels. Each row of the data matrix $\boldsymbol{Y}$ is the measured activity pattern over voxels, and each column corresponds to the activity profile of a certain voxel across time. The first step in fMRI analysis is to relate the measured activity to the experimental paradigm. Generally, such paradigms consist of Q different experimental conditions, and each condition repeats n times to the participant (i.e., the number of repetitions or runs = n). Thus, the classical experiment will have $N = Q \times n$ trials. To estimate the activity patterns for each trial, it is standard to conduct a so-called "first-level" general linear model (GLM):

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i \tag{2.1}$$

where $\mathbf{y}_i$ is an $T \times 1$ observed time-series vector of fMRI signal intensities for each voxel i (with $1 \leq i \leq p$), $\mathbf{X} \in \mathbb{R}^{T \times N}$ is the design matrix that represents the temporal evolution of the predicted BOLD response in to each of the N trials of the experiments. $\boldsymbol{\beta}_i \in \mathbb{R}^{N \times 1}$ is a vector of regression coefficients that represent the brain activity for voxel i across N trials, reflecting the relative contribution of each experimental condition to the fMRI signal. $\boldsymbol{\epsilon}_i$ is the true error term of length T for voxel i. The standard assumption of $\boldsymbol{\epsilon}_i$ is that it follows multivariate

normal distribution with $\boldsymbol{\epsilon}_i \sim MVN(0, \sigma_i^2 \boldsymbol{\Sigma}_T)$, where $\boldsymbol{\Sigma}_T$ is a $T \times T$ temporal correlation matrix (Friston et al. [1994a]). This analysis estimates the values of the vector $\boldsymbol{\beta_i}$ that best fit the observed fMRI signal intensities $\mathbf{y}_i$, using least squares or maximum likelihood estimation. The resulting estimates of $\boldsymbol{\beta_i}, \hat{\boldsymbol{\beta}}_i$, then can be tested for statistical significance, allowing researchers to identify brain regions that show significant responses to the experimental conditions. The estimate of $\boldsymbol{\epsilon}_i$, $\boldsymbol{e}_i$, describes the difference between $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i$ that are obtained for the study (Friston et al. [1994a]). $\boldsymbol{e}_i$ then can be used to form the residuals $\mathbf{R} = [\boldsymbol{e}_1, ..., \boldsymbol{e}_p] \in \mathbb{R}^{T \times p}$. This first-level model is "mass-univariate", i.e. it is typically conducted independently for each voxel in the brain. Overall, the estimation of the model results in the $N \times p$ matrix of activity estimates $\mathbf{B} = [\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_p]$.

Here we are concerned with the second-level GLM (multivariate analysis), which then tries to understand the relationship between the different activity patterns across the voxels in a region of interest. Like the first-level model, this model is formulated for each participant separately. We assume a multivariate random effects model, where the activity patterns $\mathbf{B}$ are modeled using a design matrix $\mathbf{Z} \in \mathbb{R}^{N \times Q}$, which relates each trial to an experimental condition. Each column of $\mathbf{Z}$ consists of indicator variables, either 0 or 1, denoting whether the Q^{th} experimental conditions occurred on the respective trial. Now, we can consider the familiar random effect model that assumes the data $\mathbf{B}$ are generated as a linear combination of Q regressors with some noise $\mathbf{E}$.

$$\boldsymbol{B} = \boldsymbol{Z}\boldsymbol{U} + \boldsymbol{E} \tag{2.2}$$

where $\mathbf{U} \in \mathbb{R}^{Q \times p}$ is a matrix of true activity patterns (we treat $\mathbf{U}$ across different subjects as the random effects). The rows of the matrix $[\boldsymbol{u}_1, ..., \boldsymbol{u}_Q]$ are the activity patterns associated with the Q experimental conditions. $\mathbf{E} \in \mathbb{R}^{N \times p}$ is the noise matrix.

We assume that across p voxels, the rows of $\mathbf{U}$ are distributed with mean 0 and variance-covariance matrix $\mathbf{G} \in \mathbb{R}^{Q \times Q}$. This covariance matrix determines the overlap, or similarity of different conditions, and is the central quantity of interest for representational models (Diedrichsen et al. [2017]). Hypothesis are then tested by analyzing these covariance matrices

across participants.

Typically, multivariate models make the (implicit) assumption that the true activity patterns do not have a specific spatial structure, and that the spatial structure is the same across all conditions. Thus, **U** is assumed to have a matrix normal (MN) distribution with

$$\boldsymbol{U} \sim MN(\mathbf{0}, \boldsymbol{G}, \boldsymbol{I}) \tag{2.3}$$

Many commonly used multivariate analysis techniques, such as RSA, aim to test the hypotheses about the distribution of activity patterns specified by different representational models, each of which predicts a different variance-covariance matrix **G**.

A critical problem here is that the noise **E** is correlated across voxels. Therefore, we have to assume that the columns of **E** are distributed normally with mean **0** and variance-covariance matrix $\boldsymbol{V}_p$. In contrast, the rows of **E** can be assumed to be independent and identically distributed (i.i.d.), as long as the temporal (auto-) correlation of the measurement noise has already been taken into account in the first-level GLM (Eq. 2.1). Together with a normal assumption, we therefore have

$$\boldsymbol{E} \sim MN(\mathbf{0}, \boldsymbol{I}, \boldsymbol{V}_p) \tag{2.4}$$

In practice, there are two options to deal with the spatial correlation of the measurement noise. The first option is to ignore this issue and assume that the noise is independent after normalizing each voxel by its standard deviation from the residual of the time series model (Eq. 2.6). To this end, we first employ the residuals **R**, a T (number of time points) $\times\, p$ (number of voxels) matrix, from the regression model to obtain an empirical covariance matrix $\hat{\boldsymbol{V}}_p$

$$\hat{\boldsymbol{V}}_p = \frac{1}{T}\boldsymbol{R}^T\boldsymbol{R} \tag{2.5}$$

Then, we use $\hat{\boldsymbol{D}}_p$, the diagonal of matrix $\hat{\boldsymbol{V}}_p$, to transform or prewhiten the data $\mathbf{B}$,

$$\tilde{\boldsymbol{B}} = \boldsymbol{B}\hat{\boldsymbol{D}}_p^{-1/2} \tag{2.6}$$

leading to an approximately equal variance across voxels in $\tilde{\boldsymbol{B}}$.

The second option not only suppresses the variance of voxels but also considers the noise correlation structure between voxels. In this case, an estimate of noise correlation (known as the empirical correlation, $\hat{\boldsymbol{\Sigma}}_p$) can be obtained from the empirical covariance matrix:

$$\hat{\boldsymbol{\Sigma}}_p = \hat{\boldsymbol{D}}_p^{-1/2}\hat{\boldsymbol{V}}_p\hat{\boldsymbol{D}}_p^{-1/2} \tag{2.7}$$

which can be then used to prewhiten the data

$$\tilde{\boldsymbol{B}} = \boldsymbol{B}\hat{\boldsymbol{D}}_p^{-1/2}\hat{\boldsymbol{\Sigma}}_p^{-1/2} \tag{2.8}$$

Note that when the number of voxels p exceeds the number of time points T, the empirical correlation estimate (Eq. 2.7) will become rank deficient and hence non-invertible. Even if $p < T$, the matrix may contain many small eigenvalues, which makes the inverse of the matrix unstable and causes the pre-whitened activity estimates to have a large variance.

One common practice that has been proposed in the field of financial statistics is to shrink the empirical correlation towards its diagonal (i.e., shrinkage target is identity matrix in this case) with the shrinkage factor $0 \leq \lambda \leq 1$ (Ledoit and Wolf [2004]).

$$\tilde{\boldsymbol{\Sigma}}_p = \lambda\mathbf{I} + (1-\lambda)\hat{\boldsymbol{\Sigma}}_p \tag{2.9}$$

This regularization biases the estimate towards the assumption that all voxels are independent, which ignores the strong noise correlation across neighboring voxels in fMRI data (Friston et al. [1994a]; Zarahn et al. [1997]) and therefore leads to sub-optimal inference. Hence, we

aim to seek a more realistic shrinkage target that is informed by the underlying anatomy to improve the inference in multivariate analysis.

2.3 Anatomically-informed spatial noise correlation

In this paper, we consider several potential spatial noise correlation models based on three important characteristics of the anatomical location of the voxel pairs in question: the spatial distance in the volume space (3d-distance), the spatial distance along the cortical surface (2d-distance), and whether the voxel samples superficial or deep aspects of the cortical sheet (cortical depth). We also define a baseline model for comparison.

Exp3d Model: It is well known that the noise correlations structure between a pair of voxels are determined by their spatial distance in the volume space. In particular, two neighboring voxels are higher correlated than voxel pairs that are far apart. We first considered a model in which the noise correlation falls off exponentially with the 3d-distance between the voxel pairs (i.e. Euclidean distance):

$$M(i, j) = \gamma_1 \exp(-\alpha d_{ij}) \tag{2.10}$$

where d_{ij} is the pairwise Euclidean distance between voxel i and j, and $M(i, j)$ the predicted correlation by the model. This equation specifies that the noise correlation between two voxels diminishes as their 3d-distance increases. The parameters $\alpha > 0$ and $\gamma_1 > 0$ together control the magnitude of such decay. Note that we do not model the correlation of the voxel with itself, all $M(i, i)$ are assumed to be 1 for all models.

Double Exp3d model: Instead of using a single exponential function, another way to model such decay is to use a double exponential function

$$M(i, j) = \gamma_1 \exp(-\alpha d_{ij}) + \gamma_2 \exp(-\beta d_{ij}^2) \tag{2.11}$$

where d_{ij}^2 is the squared Euclidean distance between the i-th and the j-th voxels. The parameters $\alpha > 0$ and $\gamma_1 > 0$ together adjust the scale of the first exponential function, while $\beta > 0$ and $\gamma_2 > 0$ control the magnitude of the decreasing pattern captured by the second (squared) exponential function. An important characteristic of voxel-wise noise correlation is that it rapidly declines for a short distance, but much more slowly over a long distance (see below). The key difference between the Exp3d and the Double Exp3d model is that the Double Exp3d model can capture the different decay behaviors between a short and a long distance, but the Exp3d fails to do so.

Double Exp2d model: In addition to 3d-distance, we also explored to what degree the noise correlations are determined by spatial distance along the cortical surface (2d-distance). Similar to Double Exp3d, we can model noise correlations between voxel pairs as a double exponential function of 2d-distance (measured by the geodesic distance along the cortical surface):

$$M(i, j) = \gamma_1 \exp(-\alpha g_{ij}) + \gamma_2 \exp(-\beta g_{ij}^2) \tag{2.12}$$

where g_{ij} is the geodesic distance between the voxel pair (i, j).

Double Exp2d3d model: To investigate whether a combination of 3d- and 2d-distance would better model the noise correlations structure, we can try a simple additive combination, i.e. Double Exp3d + Double Exp2d:

$$M(i, j) = \gamma_1 \exp(-\alpha_1 d_{ij}^2) + \gamma_2 \exp(-\alpha_2 g_{ij}^2) + \gamma_3 \exp(-\beta_1 d_{ij}) + \gamma_4 \exp(-\beta_2 g_{ij}) \tag{2.13}$$

$\gamma_3, \beta_1, \gamma_4, \beta_2$ are parameters that adjust the shape of the exponential kernel related to 3d-distance, and $\gamma_1, \alpha_1, \gamma_2, \alpha_2$ are scaling parameters for 2d-distance.

Double Scaled Exp2d3d model: Another simplified way for combining 3d- and 2d-distance

is to share the same parameters for squared 3d-distance (d_{ij}^2) and squared 2d-distance (g_{ij}^2):

$$M(i,j) = \gamma_1 \exp(-\alpha d_{ij}^2) + \gamma_1 \exp(-\alpha g_{ij}^2) + \gamma_2 \exp(-\beta_1 d_{ij}) + \gamma_3 \exp(-\beta_2 g_{ij}) \tag{2.14}$$

This model implicitly assumes that the slope between noise correlation and short 3d-distance is same as that of short 2d-distance. This is also the main difference between the Double Exp2d3d and the Double Scaled Exp2d3d model.

Double Exp3d Depth model: Importantly, we also found that noise correlations structure depend on the location of the gray matter voxels (cortical depth). This model combines two important determinants of noise correlations: 3d-distance and cortical depth:

$$M(i,j) = \gamma_1 \exp\{-\alpha d_{ij} - \theta((h_i - \mu)^2 + (h_j - \mu)^2)\} + \gamma_2 \exp(-\beta d_{ij}^2) \tag{2.15}$$

where h_i, h_j are the depth of i, j-th voxel, and μ is the estimated center of the circle formed by the correlation of voxels pairs between h_i and h_j (see Fig.3.4A). The parameter θ controls how quickly the correlation falls off as the cortical depth increases.

Null model: As a baseline for model comparison, we defined the identity matrix as our null model that assume no correlation between voxels pairs.

$$M(i,j) = 0, \forall i \neq j \tag{2.16}$$

For all models, we consider both the model prediction alone (based only on the anatomical information from that subject) and any mixture with the empirical correlation estimate ($\hat{\mathbf{\Sigma}}_p$, Eq. 2.7)

$$\tilde{\mathbf{M}} = \lambda \mathbf{M} + (1 - \lambda)\hat{\mathbf{\Sigma}}_p \tag{2.17}$$

2.4 Evaluation criterion: Effective number of voxels

After building noise correlation models, the next important issue is to find an appropriate metric that measures the quality of our estimate. While many distance measures between correlation matrices exist, such as the Riemannian distance (Shahbazi et al. [2021]), we would like to use a discrepancy measure that directly evaluates the goal achieved with our estimate. Our goal is to make the measurement noise, after prewhitening it with our estimate, as close to independent across voxels as possible (Eq. 2.8). That is, we expect $\mathbf{\Sigma}_R$, the correlation matrix of $\tilde{\mathbf{B}}$ across voxels, is as close to the identity matrix ($\mathbf{I}_p$) as possible, where

$$\mathbf{\Sigma}_R = \boldsymbol{M}_p^{-1/2} \mathbf{\Sigma}_p \boldsymbol{M}_p^{-1/2} \tag{2.18}$$

$\boldsymbol{M}_p$ is the predicted anatomically-informed spatial noise correlation matrix, and $\mathbf{\Sigma}_p$ is the true correlation matrix. Therefore, rather than measuring the distances between the estimated and true correlation matrix, we measure the deviation of the resulting correlation matrix (i.e., $\mathbf{\Sigma}_R$), after prewhitening the true noise correlation matrix with our estimate, from the identity matrix.

Our metric is inspired by the notion of the effective degrees of freedom (dof) proposed by Worsley and Friston [1995], which can be calculated using Satterthwaite approximation (Satterthwaite [1946]). The latter is a moment-based approximation for the mean and variance of a linear combination of squared normal random variables. If the normal random variables are independent, the sum of squared of those independent normal random variables (the noise terms across voxels in our case) follows χ^2 distribution. Here the Satterthwaite approximation deals with the situation of correlated random variables and provides an effective dof, which refers to the number of independent observations that would lead to a χ^2 distribution of a similar shape.

We want our metric to consider the spatial correlation between neighboring voxels after prewhitening. Therefore, I, analogy, define a metric that we call the *effective number of voxels* (ν) to estimate the number of independent voxels in the measurement noise, which would have

given rise to sums-of-squares with similar mean and variance:

$$\nu = \frac{trace(\mathbf{\Sigma_R})^2}{trace(\mathbf{\Sigma}_R\mathbf{\Sigma}_R)} = \frac{p^2}{trace(\mathbf{\Sigma}_R\mathbf{\Sigma}_R)}, \text{ where } 1 \leq \nu \leq p \tag{2.19}$$

The equal sign holds because $\mathbf{\Sigma}_R$ is a $p \times p$ correlation matrix (Eq. 2.18). When $\mathbf{\Sigma}_R = \mathbf{I}_p$, meaning the predicted matrix exactly equals to the true correlations (i.e., $\boldsymbol{M}_p = \mathbf{\Sigma}_p$), ν will reach to p. In contrast, if $\mathbf{\Sigma}_R$ deviates largely from $\mathbf{I}_p$ such that the measurement noise is still highly correlated after pre-whitening, then $trace(\mathbf{\Sigma}_R\mathbf{\Sigma}_R)$ will close to p^2 and hence ν will close to 1.

This metric determines the variance of the second moment matrix across conditions $\mathbf{G}$, the central quantity that multivariate analysis methods make inferences about. Under the null hypothesis that there is no difference between true activity pattern (i.e., $\mathbf{U} = 0$) (see Appendix A.2 for details):

$$Var(\hat{\mathbf{G}}_{ij}) \propto 1/\nu$$

To compare this measure across different regions with a different number of voxels, we can normalize ν by the number of voxels p. This results in the normalized effective number of voxels (NENoV), denoted by $|\nu|$:

$$|\nu| = \frac{\nu}{p}, \text{where } 0 < |\nu| \leq 1$$

Since we expect $\mathbf{\Sigma}_R$ to be as close to $\mathbf{I}_p$ as possible and also expect the variance of our quantity $\hat{\mathbf{G}}$ to be small, the model with the highest NENoV was considered to be the best one.

2.5 Model fitting and evaluation

We compared three different ways of estimating the model parameters.

2.5.1 Individual fit

The first option is to fit the models separately within each individual, ROI and dataset. To evaluate these models in an unbiased fashion, we then employed a leave-one-run-out cross-validation procedure. We used the residuals $\mathbf{R}$ of a single imaging run as a test set, fitting the parameters to empirical correlation estimate from the remaining runs (Eq. 2.7) . We then used those parameters and the individual anatomical information to predict the model $\hat{\boldsymbol{M}}_p$. By plugging the empirical correlation $\hat{\boldsymbol{\Sigma}}_p$ from residuals and the predicted model $\hat{\boldsymbol{M}}_p$ into Eq. 2.19, we obtained the effective number of voxels. After normalizing ν by the number of voxels p, we obtained a normalized effective number of voxels $|\nu|$.

2.5.2 Group fit

As an alternative option, we fitted the models on all the participants in each ROI and dataset. To evaluate the models, we performed cross-validation across participants. We first estimated the parameters using the data from all but one participant (training set). The parameters, combined with the anatomical information from the left-out subject, were then used to predict the correlation matrix of each run in the test set. As for the individual model fitting and evaluation, the normalized effective number of voxels was then averaged across runs.

2.5.3 Across-dataset fit

Finally, we considered obtaining one global set of parameters across all datasets and ROIs. For evaluation, we applied a cross-validation across datasets to estimate the model performance on a novel dataset, in an arbitrary cortical ROI. To avoid the challenge of fitting a large number of correlation matrices, we chose here a computational more efficient variant. For each ROI within each dataset, we fitted the parameters across participants. We then predicted the noise correlation model for a new subject within the left-out dataset with each of these models and then averaged the predictions across models.

2.6 RDM reliability analysis

RSA is an important tool for studying brain representation. A key step in RSA is to compute the representational dissimilarity matrix (RDM), which measures the dissimilarity, or distance, between activity patterns associated with different experimental conditions. One way to evaluate different forms of deriving these distances on real data is to evaluate the reliability of RDMs across independent halves of the data within each ROI and individual (Walther et al. [2016]). Here, we evaluated the impact of noise covariance estimates on the reliability of RDMs. To calculate the reliability, we used split-half reliability estimates. We divided the data into two independent partitions (odd and even runs), pre-whitening the activity patterns of each partition using the estimate derived from that half of the data. The dissimilarity measures were then computed in each split. We chose the cross-nobis distance estimate, an unbiased estimate of the squared Mahalanobis distance between two activity patterns (Diedrichsen et al. [2020]). Ultimately, we obtained two $1 \times K$ vectors of dissimilarities, where $K = Q(Q-1)/2$ is the number of the pairwise distance between Q conditions. To assess the reliability between those two RDMs, we computed their cosine similarity between the two vectors of dissimilarities. Best regularization parameters λ were found using a simple grid search.

2.7 Statistical Analysis

We compared model performance either within each dataset separately, or across all datasets. For the latter, we considered participants across all six datasets ($n = 119$) as a random effect, therefore conducing paired t-test with one estimate per participant. That is, we treat each subject in each dataset independently. This is reasonable because the experiments and the subjects are different in different datasets, and this analysis helps improve the generalizability of the findings beyond the specific dataset being studied. To evaluate the replicability of these results across datasets, we also performed a similar analysis within each dataset (reported in tables), again considering the participant as a random effect. All t-tests were paired two-sided.

Data distribution was assumed to be normal, but this was not formally tested.

Chapter 3

Result

3.1 Three anatomical factors that determine noise correlations

3.1.1 Spatial distance in the volume (3d-distance)

It is well-known that there is a strong dependence between the noise correlations and the spatial distance of voxels, with voxels that are close together being much more correlated than voxels that are far apart (Wiestler et al. [2011]). As can be seen for a selected ROI and dataset, the correlation is quite high for directly neighboring voxels but then rapidly falls off as 3d-distance increases (Fig. 3.1, Empirical). To model this fall-off, we first tried a single exponential model (Exp3d, Eq. 2.10). We estimated the parameters of the model for each participant, leaving one run out, and evaluated the model by assessing the normalized effective number of voxels (NENoV) on that left-out run (see *Individual fit*). We then compared this model fit with the null model that assumed no correlation between two voxels (iid, Eq. 2.16). The Exp3d model showed a clear improvement in the NENoV compared to the null model (Fig. 3.2, iid and Exp3d). This advantage was highly significant when tested across all individuals and datasets (two-sided paired t-test: $t_{118} = -18.471$, $p = 1.3 \times 10^{-36}$) and also significant in each dataset

across all individuals (Table 3.1, iid vs. Exp3). These results clearly show that correlations between voxels depend strongly on 3d-distance.

While the single exponential model can capture some drop-off of the correlation with 3d-distance, the model does not fit the empirical data very well (Fig. 3.1, Exp3d). This is because the correlation decreases rapidly for short distances but much more slowly for long distances. Therefore, the relationship between 3d-distance and noise correlation may be better captured by a double exponential model (Double Exp3d, Eq. 2.11). The first term models the correlation when the distance is large, while the second part captures the association when the distance is small. The Double Exp3d model showed an increase in the NENoV compared to the Exp3d model (Fig 3.2). This improvement was significant across all individuals and datasets ($t_{118} = -7.018$, $p = 1.5 \times 10^{-10}$) and was also significant for each data, except one, in isolation (Table 3.1, Exp3d vs. Double Exp3d). Thus, the rapid reduction of noise correlations for short distances, and the persistent correlations even over long distances can be better captured by a double exponential model.

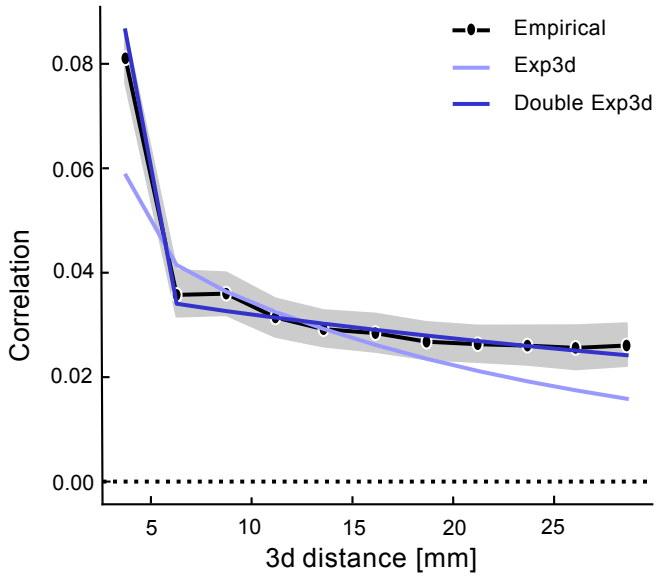

Figure 3.1: Correlation of voxel pairs as a function of the 3d-distance of the pair in the MDTB dataset. The average empirical correlation (Eq. 2.7) is shown in the solid black line in comparison to the single (Exp3d) and double exponential (Double Exp3d) model. The shaded area indicates the between-participant standard error of the empirical correlation.

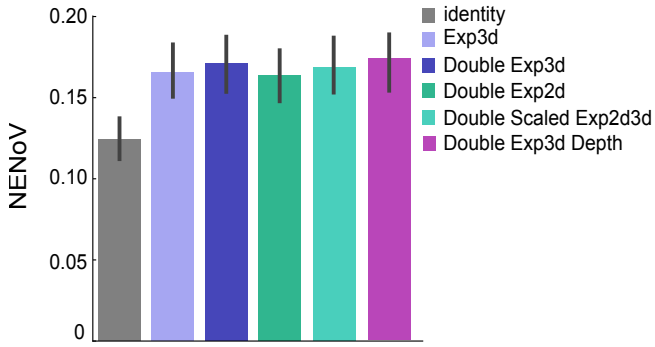

Figure 3.2: The normalized effective number of voxels (NENoV) based on the individual cross-validated fit for each model. Error bars indicate standard errors between participants across datasets. Each color corresponds to different noise correlation models.

	Data set (W)					
Model comparison	**1** (24)	**2** (21)	**3** (42)	**4** (10)	**5** (6)	**6** (16)
Normalized effective number of voxels (metric)						
iid vs. Exp3d	**-14.549** **4.33e-13**	**-8.921** **2.08e-08**	**-13.266** **2e-16**	**-10.279** **2.84e-6**	**-7.700** **5.91e-4**	**-8.257** **5.82e-7**
Exp3d vs. Double Exp3d	**-10.312** **4.27e-10**	0.830 0.42	**-2.062** **0.046**	**-5.014** **7.25e-4**	**-3.300** **0.02**	**-5.238** **1e-4**
Double Exp2d vs. Double Exp3d	**-12.231** **1.51e-11**	-1.065 0.30	**-3.746** **5.53e-4**	**-7.929** **2.38e-5**	**-4.850** **0.005**	**-6.894** **5.11e-6**
Double Scaled Exp2d3d vs Double Exp3d	**-6.687** **8.03e-7**	-2.064 0.052	**-3.014** **0.004**	0.085 0.93	-1.357 0.23	**-4.525** **4.02e-4**
Double Exp3d vs. Double Exp3d Depth	**-6.069** **3.44e-6**	-0.069 0.95	-0.756 0.45	**-5.002** **7.36e-4**	0.516 0.63	**-5.324** **8.51e-5**

Table 3.1: Model comparisons in each dataset across subjects. The table shows t-values (above) and p-values (below) for comparisons of the NENoV between the two models. Here the degree of freedom $(df) = W$(the number of subjects) $- 1$ within each dataset. Each row lists a comparison between two noise correlation models. Each column lists one of the six fMRI datasets. Values were computed using two-sided paired t tests. Dark black: the second model performed significantly better than the first model for a significance threshold of 0.05.

3.1.2 Spatial distance along the cortical surface (2d-distance)

FMRI signal suffers from multiple sources of noise. While variability may originate from measurement error (non-neural noise), it can also arise from trial-by-trial fluctuations of the neural activity that we seek to measure. Non-neuronal variability arises from the fMRI measurement process, head motion, and from physiological processes, such as heartbeat and breathing (Liu

[2016]). All these processes would lead to noise correlations that are either uniform across the brain (such as head motion) or fall off with the proximity of two voxels in the volume (3d-distance). In contrast, neural variability arises from the spontaneous activity of the human brain. These interactions rely on communication between different neuronal elements. Because nearby neurons generally exchange more signals with each other than neurons that are far apart, we can assume that these correlations would fall off with the distance along the cortical surface (2d-distance) (Rosenbaum et al. [2017]). In places where the neo-cortex is highly folded (Fig. 1.1A), such as in the primary visual cortex (V1), a model that predicts noise correlations from the 3d-distance or from the 2d-distance would make very different predictions. Here we aim to understand whether the structure of fMRI variability is determined by the 2d-distance between voxels on the cortical surface, the 3d-distance in the volume, or both.

We first asked whether the distance along the cortical surface is a determinant of noise correlations between two voxels after we account for their spatial distance in the volume. We therefore compared voxel pairs that were equally far apart in 3d-distance, but one pair had a short 2d-distance while another had a long 2d-distance (Fig. 1.1A). This scenario often happens in places where the cortex is folded inwards (Sulci). To the degree that noise correlations depend on neuronal fluctuations, voxels that are located on the same side of the sulcus (and hence have a small 2d-distance) should be more correlated with each other than voxel pairs that are located on opposite sides (and hence have a large 2d-distance). To test this hypothesis, we selected voxel pairs within a small range of 3d-distance and studied the influence of the 2d-distance on the correlation within each bin. Fig 3.3 illustrates an example of the primary visual cortex of the MDTB dataset. In most bins, we found a slight decrease in the correlation with increasing 2d distance. To quantify this effect, we first the evenly spaced 3d-distance in each ROI and dataset into 20 bins. For each level of 3d-distance, subject, and ROI, we fitted a linear regression between correlations and 2d-distance. We then averaged the slope across bins and ROIs in each dataset. If 2d-distance does have a significant effect on noise correlations, the expected value of the slope should be smaller than zero. For each dataset, we therefore

tested the slope against zero across subjects. Although all datasets showed a negative average slope, only 2 out of 6 datasets reached significance (Table 3.2). These results suggest that the 2d-distance between voxels has an influence on noise correlations, but its effect is weak.

In highly folded areas of the neo-cortex, distances in the volume and along the cortical surface can be quite different. Overall, however, the two are generally quite correlated with each other (mean correlation= 0.644 across datasets, SD = 0.136). This raises the question of which one is a better predictor for the correlation structure of the fMRI noise. To answer this question, we built a model that only contained 2d-distance (Double Exp2d, Eq. 2.12) and compared it with a model with only 3d-distance (Double Exp3d). The model with only 3d distance showed a significantly higher NENoV than the 2d model across all participants and datasets ($t_{118} = -9.105, p = 2.61 \times 10^{-15}$; Fig. 3.2). When tested in an individual dataset, this comparison also revealed that the 3d-distance model had a larger contribution to noise correlations (Table 3.1, Double Exp2d vs. Double Exp3d) that the distance along the surface. These results suggest that the main source of measurement errors in fMRI arises from non-neuronal sources.

Finally, we tested whether the combination of 3d- and 2d-distance would allow a better prediction of the noise correlation structure using the 3d-distance alone. To test this, we first studied different ways of combining the 2d and 3d-distance models. It turned out that a simple additive combination (with 8 parameters, Eq. 2.13) performed worse than a model in which the parameters for the squared 3d-distance (d_{ij}^2) and squared 2d-distance (g_{ij}^2) were shared. This shared parameter model (Eq. 2.14) captures how quickly the correlation falls off over short distances, regardless of whether measured in 3d or 2d. When we compared Double Scaled Exp2d3d with Double Exp3d, we found that overall, the 3d-distance model outperformed the combined 3d/2d model ($t_{118} = -7.121, p = 9.07 \times 10^{-11}$; Fig. 3.2). This difference was significant in 3 out of 6 datasets, and the combined 2d/3d-model performed numerically better only in a single dataset (Table 3.1, Double Scaled Exp2d3d vs. Double Exp3d). Because we previously showed that the distances along the surface had some influence on the noise

correlations, this result suggests that the added complexity of the combined 2d and 3d-model does not translate into enough benefit to improve prediction performance.

Data set	t-value	degree of freedom (df)	p-value
1	-7.123	23	**2.96e-7**
2	-3.770	20	**0.001**
3	-1.734	41	0.09
4	-1.476	9	0.17
5	-2.367	5	0.06
6	-1.772	15	0.10

Table 3.2: Summary statistics for the slope between 2d-distance and correlation across subjects in each dataset. Dark black: 2d-distance shows significance after averaging the slope across distance bins and ROIs in each dataset for a significance threshold of 0.05.

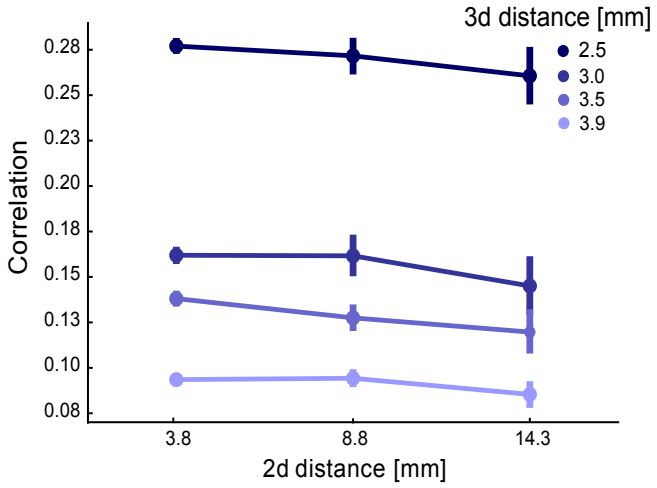

Figure 3.3: The relationship between 2d-distance and empirical correlation for specific 3d-distance across all subjects in V1 of MDTB dataset. Error bars indicate the standard error between subjects.

3.1.3 Cortical depth

Some voxels sample areas relatively deep in the cortical surface, mixing gray-matter signals with those from the underlying white matter (Fig. 1.1B, deep), while other voxels sample relatively superficial areas of the cortical surface, mixing gray-matter signals with cerebral-spinal-

fluid (Fig. 1.1B, superficial). We strongly suspected that noise correlations between voxels would depend on their depth within this surface. In terms of neural fluctuations, it has been shown that V1 neurons in the deeper (granular) layers were weakly correlated, while neurons in the superficial (supragranular) layers were strongly correlated with each other. This layer dependence is due to heterogeneous intracortical inputs to different layers of the cortex (Hansen et al. [2012]). Here we aim to understand across ROIs if the structure of noise correlations can be better captured by taking into account the cortical depth.

To investigate how noise correlations vary as cortical depth, we plotted the correlation of voxel pairs with a 3d-distance ($5mm \leq d_{ij} \leq 7mm$) as a function of the depth of both voxels (Fig. 3.4A). We can clearly see that voxel pairs tend to have a higher correlation if both voxels are superficial. When at least one of the voxels is located in deeper layers of the cortical surface, the correlation drops off substantially. We then investigated how the effects of depth interacted with the spatial distance between two voxels. We performed a median-split based on the voxel depth, and then assigned all voxel pairs to three distinct levels : (both voxels superficial), (one superficial, one deep), (both voxels deep). We then plotted the correlation as a function of the 3d-distance of the voxel pair (Fig. 3.4B). For each spatial distance, the superficial-superficial voxel pairs clearly showed a much higher noise correlation as compared to the other pairs. The effect of depth also appeared to interact multiplicatively (rather than additively) with the effect on 3d-distance. Based on these insights, we created a model that can capture the dependence of noise correlations on depth and 3d-distances. Fig. 3.4A suggests a function in which the noise correlation falls off with the depth from a specific center point. This feature can be captured with the formula $corr(i, j) = (h_i - \mu)^2 + (h_j - \mu)^2$, where h_i, h_j are the depth of voxel i, j, and μ is the center of the circle. In addition, the noise correlation for each level of pairs falls off as a double exponential of the 3d-distance (Fig. 3.4B). This led us to Eq. 2.15 (Double Exp3d Depth), which performed better than other alternative ways of capturing the dependence of noise correlation of depth and distance. We then tested whether this model would be able to predict noise correlations better than the 3d-distance model without

depth. In 3 out of 6 datasets, we found that the Double Exp3d Depth yielded significantly better performance than the 3d-distance model (Table 3.1, Double Exp3d vs. Double Exp3d Depth). Although the difference between the Double Exp3d Depth model and the Double Exp3d model across individuals and datasets was not significant ($t_{118} = -1.132, p = 0.26$; Fig.3.2), we still prefer the Double Exp3d model, as it on average performed better than the 3d-model. Together, these results suggest that noise correlations indeed vary with the depth of the voxel within the cortical surface. Incorporating cortical depth into the model provided, at least for some datasets, a clear improvement in the prediction of the noise correlation.

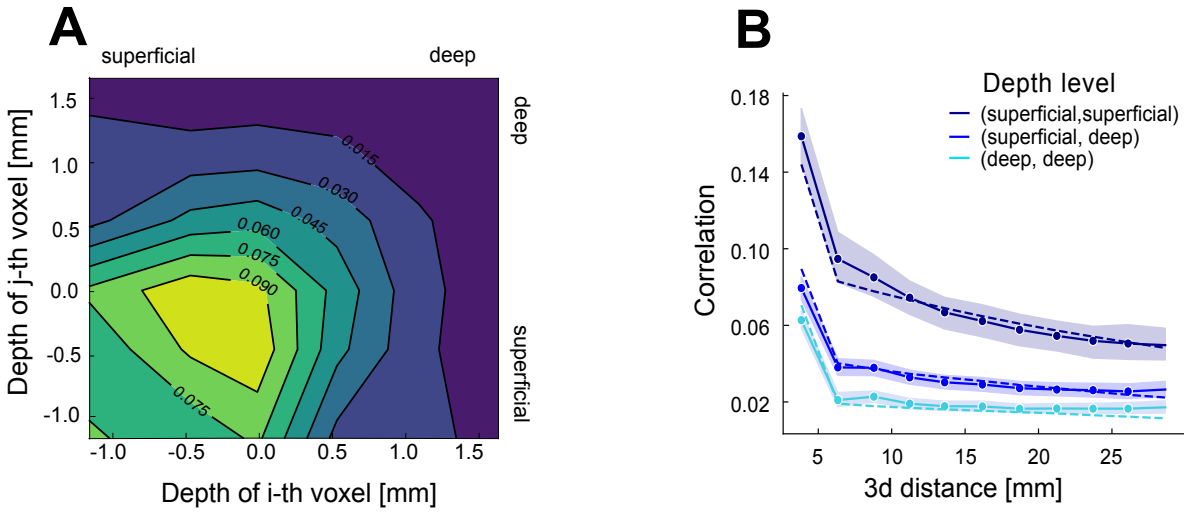

Figure 3.4: Influence of depth on noise correlation. *A*: Contour plot of the relationship between the depth of *i*-th voxel, *j*-th voxel and their corresponding empirical correlations for voxel pairs with 5 − 7mm Euclidean distance. A depth of 0 indicates that the center of the voxels is on the surface between gray-matter and CSF, and a depth of 1 indicates that the center of the voxel is located on the surface between gray-matter and white matter. Because the value of a voxel is determined by the average MRI signal within it, even voxels with centers located outside the cortical ribbon will reflect some gray-matter signals. *B*: Correlation for voxel pairs after a median-split of the voxel-depth values. Pairs can belong to three categories: (superficial, superficial), (superficial, deep), and (deep, deep). Empirical correlation (solid line) and predicted correlation by Double Exp3d Depth Model (dashed line). Shaded areas represent across subjects.

3.2 Common parameters that can be applied to any dataset

In the previous analysis, we constructed several models based on the dependence of noise correlation on anatomical information and concluded that the Double Exp3d Depth model provides the most accurate predictions across runs. If we want to apply this model to improve inference on multivariate activity patterns, we need to address the important practical issue of how we estimate the model parameters. In our model comparison in the last sections, we fitted the models to residuals from the same subject, brain area, and dataset as the left-out test data. While this is technically possible, the computational expense is considerable, especially if we want to draw multivariate inferences within many different small regions of the cortex. Additionally, the individual fit may suffer from high parameter uncertainty, given the relatively restricted training data for each individual. We therefore considered two alternative ways to estimate the parameters. First, we fitted the parameters across all subjects in a dataset (group fit) within each ROI. This would yield more stable estimates, while still retaining the possible dataset- and ROI-specific variation in the optimal parameters. Secondly, we fitted the models across all subjects, ROIs and datasets (across-datasets fit). While this approach ignores any possible variation across dataset or ROI, it provides the most practical solution, as the model would not have to be re-fit when they are applied to improve multivariate inference.

To compare these three approaches and determine which method to implement, we compared the inference in the best model based on the parameters estimated using cross-validation within the participant, across participants, or across datasets. For the first method, we took one run of the data as a test set, fitted parameters to data from the remaining runs, and then evaluated the NENoV in the left-out run. In group (or across subjects) cross-validation, we estimated the parameters from all but one participant and then evaluated them from the left-out subject. The first two methods were performed within each ROI and dataset. The last method was performed across ROI and dataset: we implemented cross-validation from all but one dataset and evaluated them in the left-out dataset (See *Model fitting and evaluation*). Relative to the differences between models, the differences between individual, group, and across-datasets were rather

small (Fig. 3.5). There was an overall significant difference between individual and group fit ($t_{118} = 3.837, p = 0.0002$), with individual fitting being significantly better than group fitting in 3 out of 6 datasets (Table 3.3, individual vs. group). In contrast, individual fits did not significantly differ from across-datasets fits ($t_{118} = 0.603, p = 0.54$; Fig.3.5, individual and across datasets). For 3 out of 6 datasets (Table 3.3, individual vs. across datasets), we found significant improvement, but for the largest dataset, the individual fit performed slightly worse than the across-dataset fit. These results indicate that the relationship between the anatomical location of voxels and their noise correlation shows modest subject-to-subject and ROI-to-ROI variation. Overall, however, a single set of parameters estimated on completely separate datasets performed nearly equivalently. This implies that practically, we can use the common parameters estimated in our study to apply to a completely new dataset.

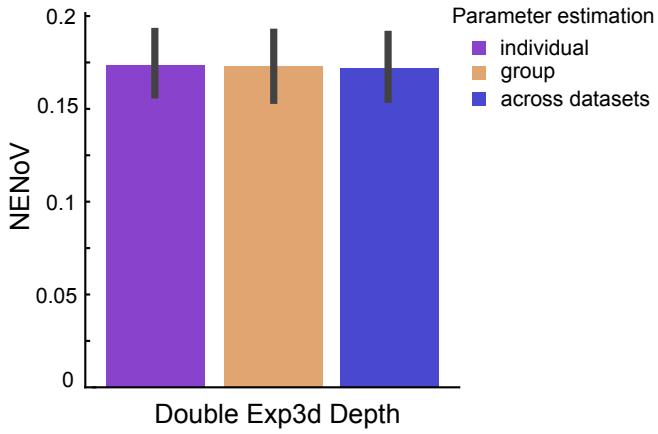

Figure 3.5: The normalized effective number of voxels based on individual, group, across datasets cross-validated fit in Double Exp3d Depth. Error bars indicate standard errors between-participant across datasets.

	Data set (W)					
Model comparison	**1** (24)	**2** (21)	**3** (42)	**4** (10)	**5** (6)	**6** (16)
Parameter Fit						
individual vs. group	**3.375**	**2.164**	0.338	2.127	2.043	**2.329**
	0.003	**0.04**	0.737	0.06	0.10	**0.03**
individual vs. across datasets	**6.637**	0.05	-1.019	1.434	**3.722**	**3.018**
	9.01e-7	0.96	0.31	0.19	**0.01**	**0.009**

Table 3.3: Comparison of the normalized effective number of voxels in the best model (Double Exp3d Depth) based on cross-validated individual, group, or across-dataset fit. Dark black: the former parameter estimation method performs significantly greater than the latter.

3.3 Using model-based shrinkage improves statistical inference

Mounting evidence suggests that the incorrect assumption of independence between variables leads to sub-optimal or erroneous multivariate inference. One of our main motivations for developing a better model of voxel-to-voxel noise correlations in fMRI was to address this very issue. Indeed, the effective number of voxels in the Double-Exp3d-Depth model was substantially higher compared with the identity matrix (Fig. 3.2, identity and Double Exp3d Depth). This suggests that using our model provides a better assumption of noise correlations than independence and therefore results in a better multivariate inference.

However, most multivariate analysis methods for fMRI do not simply use the identity matrix as a noise correlation estimate. Instead, it is common practice to obtain an empirical estimate of the covariance matrix from the residuals of the regression model applied to the fMRI time series data. To stabilize this estimate, it is then shrunken towards the identity matrix (Ledoit and Wolf [2004]), using the shrinkage factor λ (Eq. 2.9). The regularized estimate ($\tilde{\mathbf{\Sigma}}_p$) then is used as the noise correlation estimate for pre-whitening (Brier et al. [2015]; Walther et al. [2016]; Rahim et al. [2019]).

Therefore, instead of comparing our model with the independence assumption, we want

to investigate what happens if we shrink the empirical estimate of the correlation matrix towards our model prediction, rather than towards the identity matrix. To examine this issue, we regularized the empirical correlation estimate towards the Double Exp3d Depth model with λ (Eq. 2.17) and compared the result with regularizing towards the identity matrix (Eq. 2.9). We explored the best setting for λ on the range of [0, 1] in steps of 0.1 across individuals for each dataset. Using only univariate prewhitening (i.e., using the independence assumption) acts here as the baseline or Null-model. λ is the tuning parameter that decides how much we regularize the empirical estimates towards the shrinkage target. When $\lambda = 0$ (i.e., no regularization), the noise correlation estimate is simply the empirical estimate. When $\lambda = 1$, we would use only the shrinkage target (either identity matrix or Double Exp3d Depth model). In this analysis, the model parameters are taken from the across-dataset fit over all 6 datasets and regions.

As can be seen in Fig. 3.6, when $\lambda = 1$ the identity matrix (*univariate prewhitenning*) had a lower normalized effective number of voxels than our model (*Double Exp3d Depth*), which was consistent with our previous result. Moreover, when we had enough data (i.e., $T >> P$), the empirical correlation estimate performed well. In this case, the shrinkage factor that achieves the highest value is $\lambda = 0.1$ for the independent model and 0.2 for our model. However, in typical fMRI datasets, we are seldom in a situation in which T is larger than P by a factor of nearly 20. More typically, the ratio is closer to 1. Fig. 3.6B shows an example of this situation. Here we had to regularize more towards the shrinkage target in order to achieve the peak performance, with the optimal $\lambda = 0.5$ for the independent target and 0.6 when using our model.

In both cases, however, the use of our model as a shrinkage target led to significantly better performance than the identity matrix, when using the respective optimal λ. This also held true for all the other datasets (Table 3.4). Notice that the optimal λ for our model is usually higher than that for the identity matrix. This makes sense as the closer a shrinkage target is to the true correlation structure, the more we can reduce the variance in the estimate without having to pay with an increased bias. Together, these results demonstrate that the use

of an anatomically informed noise covariance model as the shrinkage target has the potential to improve multivariate inference.

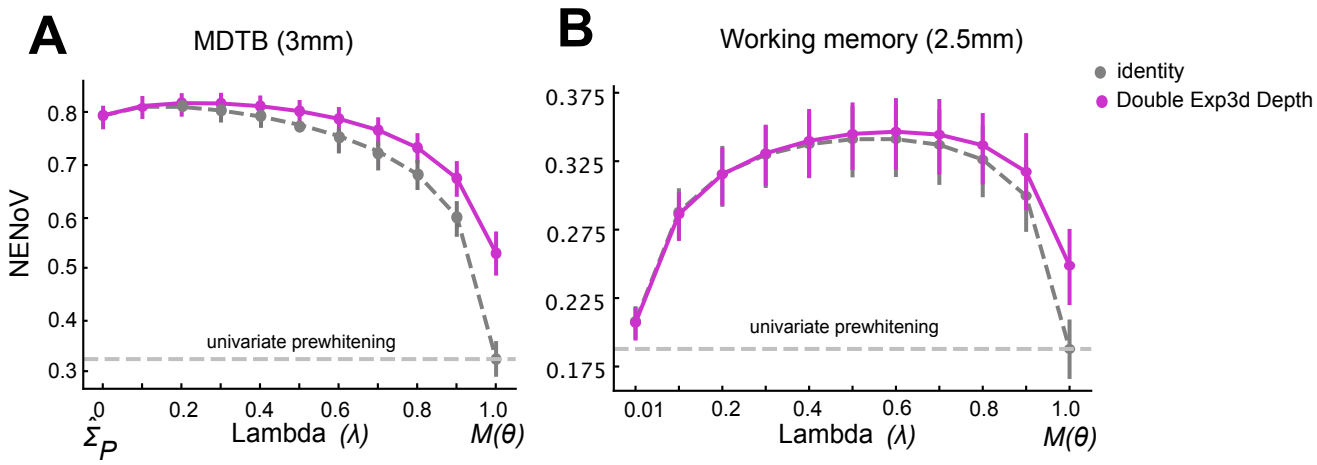

Figure 3.6: The normalized effective number of voxels in two fMRI datasets after regularizing the empirical estimate ($\hat{\mathbf{\Sigma}}_p$) towards the identity model (grey dashed line) or our Double Exp3d Depth model. *A*: Comparison results in dataset 1 with T (# of time points = 4784) >> $\bar{p}$ (averaged # of voxels ≈ 456). *B*: Comparison results in dataset 6 with $T = 409 < \bar{p} \approx 451$. In this case, $\hat{\mathbf{\Sigma}}_p$ (when $\lambda = 0$) will be rank-deficient, so we start from $\lambda = 0.01$.

Data set	Opt-λ-iid	Opt-λ-M	df	t-value	p-value
1	0.1	0.2	23	-15.346	**1.42e-13**
2	0.3	0.3	20	-16.412	**4.53e-13**
3	0.3	0.3	41	-10.83	**1.35e-13**
4	0.5	0.7	9	-9.150	**7.46e-6**
5	0.4	0.4	5	-13.131	**4.57e-5**
6	0.4	0.5	15	-7.718	**1.34e-6**

Table 3.4: Comparison between $\tilde{\boldsymbol{M}}_p$ and $\tilde{\mathbf{\Sigma}}_p$ in the normalized effective number of voxels with optimal λ. The table shows optimal λ in identity model (opt-λ-iid)/3d-distance-depth model (opt-λ-M), t-values, degree of freedom (df), and p-values across subjects in each dataset.

3.4 Reliability of distance measures also depends on the spatial structure of the signal

The effective number of voxels, theoretically, determines the variance of distance estimates (see *Evaluation criterion: effective number of voxels*). Pre-whitening the data using a model that results in a higher effective number of voxels should therefore increase the reliability of the distance estimates across replications of the experiments. Indeed, the split-half reliability of the matrix of distances is often used as a proxy measure for the quality of the noise-correlation estimate (Walther et al. [2016]) for the empirical dataset. Intuitively, the reliability with multivariate noise normalization should perform similarly with the effective number of voxels. However, that is not always the case.

To directly compare the behaviour of RDM reliability and the effective number of voxels, we estimated based on half of the empirical data in both cases. The model parameters were obtained from a group fit on the respective data set (see *RDM reliability analysis*). Fig. 3.7A shows an exemplary ROI and dataset, in which the reliability behaves similarly with the effective number of voxels: Our model outperforms the identity matrix across all values of λ. On the other hand, Fig. 3.7B illustrates an opposite example. Here, our model led to a higher effective number of voxels than the identity matrix but showed lower reliability for all λ, except $\lambda = 0$ and 1.

The reason behind this inconsistency lies likely in the fact that pre-whitening the pattern estimates not only changes the spatial correlation of the measurement noise but also the spatial structure of the signal (see Discussion). Depending on the structure of our final correlation estimate, we will de-emphasize specific aspects of the differences between conditions. Whether this will improve or impair the quality of the inferences will depend on the particular paradigm and dataset. It, however, brings attention to the fact that better assumptions about the noise covariance structure do not translate automatically into more reliable inferences - it also changes the inference we are making.

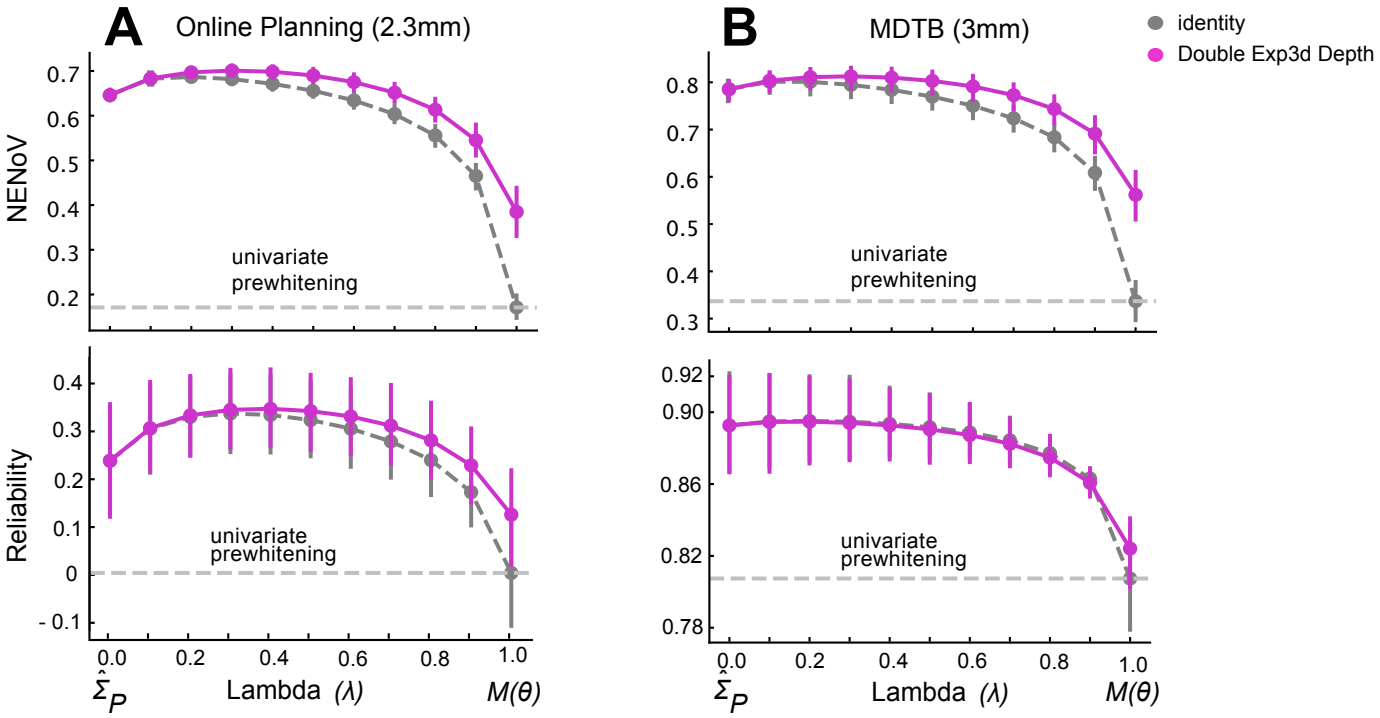

Figure 3.7: The normalized effective number of voxels and reliability in two fMRI datasets. *A*: an example graph that two metrics perform similarly. *B*: an example graph that those two metrics perform differently

Chapter 4

Discussion

fMRI can provide unique insights into the human brain - it enables us to observe the changing patterns of neural activity across the entire neocortex. Multi-variate analysis methods, which analyze the relationship of these patterns with each other, provide an opportunity to test hypotheses about the information that is represented and processed in different brain areas (Kriegeskorte and Diedrichsen [2019]). However, the fMRI signal generally has a low signal-to-noise ratio, and the measurement noise has a strong and reliable spatial correlation structure. The fact that we are missing a stable estimate of this structure (due to the limited number of data points and a large number of voxels) complicates multivariate statistical inferences on such patterns. Here we therefore develop a model of the spatial noise correlations in fMRI. Our model takes into account anatomical information, such as 3d-, 2d-distance, and voxel depth, and we show that these variables can be used to better predict the noise correlations.

Sources of fMRI measurement noise: Modelling the different factors that determine the correlation structure provides some insights into the sources of fMRI measurement noise. Head motion is one of the dominant sources of noise in fMRI studies that can only be partially corrected by motion realignment during pre-processing. Power et al. [2012] have demonstrated that movement of the head during scans can cause systematic bias to connectivity estimates by creating spurious noise correlations throughout the brain. In particular, head movement can

produce changes in the BOLD signal throughout the brain in a regionally specific manner, such that the signal increases in some parts of the brain and decreases in others. This effect can introduce either correlations or anti-correlations of fMRI signals over long distances, with nearby voxels often more correlated. Head movements can therefore account for the noise correlations across the entire brain (large 3d-distances), but also for the increased noise correlation between nearby voxels (small 3d-distances).

Physiological effects (Pulsation, breathing, changes in heart rate) will have a similar effect on noise correlations. Some of the effects will induce correlations in measurement noise over large distances. However, because the fMRI signal is dominated by smaller vessels, many of these correlation effects will be local, increasing the correlations of nearby voxels. Therefore, similar to head movements, physiological noise will induce correlations that fall off with the 3d-distance between two voxels. In contrast to head movements, however, physiological noise will vary systematically with cortical depth, because superficial aspects of the cortex have substantially more and larger vessels than deeper layers or white matter. The largest blood vessels lie along the pial surface, and the smallest micro-vessels and capillaries are situated deeper in the parenchyma (Polimeni et al. [2018]). FMRI voxels in superficial layers of the cortex are therefore more likely to be susceptible to physiological artifacts. In short, the depth dependence of physiological noise sources is expected due to the systematic change in vascular density and vessel size with cortical depth.

In addition to measurement noise arising from MRI data acquisition processes (non-neural noise), variability can also originate from trial-by-trial variation of neuronal responses that is usually unexplained by the experimental design. Our result showed that correlations between voxels fall off as the 2d-distance increases. Such dependency is expected, as neuronal fluctuation should be more correlated across two cortical patches if these are exchanging more signals. While cortico-cortical connections are complex, it tends to be true that regions that are located nearby on the cortical surface do communicate more with each than regions that are located in a separate location on the cortical surface (2d-distance). The fact that 3d-distance predicts

the structure of noise correlations better than the 2d-distance supports the idea that the major source of fMRI measurement noise is non-neuronal. Overall, the main contributions of fMRI measurement errors are head motion and physiological noise. Neuronal noise has some effect on noise correlations, but this effect is relatively weak.

Use of the model for the multivariate inference: The improvements in multi-variate prewhitening when using an anatomically informed shrinkage target suggest that it would be beneficial to apply this approach when conducting multivariate analysis for regions on the cortical surface (Oosterhof et al. [2011]). To apply our approach to new data, the user would need to specify which ROIs they are trying to analyze on the individually reconstructed cortical surface. The function that selects the voxels for such a cortical ROI already calculates the relevant anatomical variables: the 3d-distance between voxel pairs and the cortical depth of each voxel from the subject in the specified ROI. These variables would then be entered into the Double Exp3d Depth model $\hat{\boldsymbol{M}}_p$ in Eq. 2.15 to obtain a model prediction for the specific ROI. Notably, based on our result, it is not necessary to refit model parameters because the common parameters estimated from our study are stable across different ROIs and datasets. Thus, users can directly apply the pre-trained model parameters (see Appendix A for parameter values), increasing the ease of use. Additionally, we require the empirical correlation estimate from the current data. These are obtained by applying Eq. 2.5 followed by Eq. 2.7 to the residuals from the first-level GLM. This results in an empirical correlation estimate $\hat{\boldsymbol{\Sigma}}_p$. Finally, the empirical estimate needs to be integrated with the predicted correlation matrix using the shrinkage factor λ (Eq. 2.17). One strategy is to tune the hyper-parameter, shrinkage factor λ, on the range of [0, 1] in steps of 0.1 using grid-search. Ultimately, users will obtain the $p \times p$ regularized correlation matrix, which can then be used to prewhiten the data and improve multivariate inferences.

Consequences for Inferences: We found that our model led to more independent activity estimates across voxels after prewhitening, as can be seen by the increased effective number of voxels measure. However, we also found that the reliability of distance estimates did not

always increase with multivariate noise normalization. Therefore, the model comparative inference may not uniformly improve. Why is this the case? Theoretically, the variance of distance estimates should directly depend on the effective number of voxels. The explanation for this effect lies in the fact that prewhitening does not only make the measurement noise more independent, but it also changes the aspect signal itself. Eq. 4.1 illustrates the process of using the estimated noise correlation matrix $\hat{\mathbf{M}}$ to prewhiten the data $\mathbf{B}$.

$$\begin{aligned}\tilde{\boldsymbol{B}} &= \boldsymbol{B}\hat{\boldsymbol{M}}_p^{-1/2} \\ &= (\boldsymbol{ZU} + \boldsymbol{E})\mathbb{V}\mathbb{D}^{-1/2}\mathbb{V}^T \qquad \text{(By eigendecomposition)} \\ &= \boldsymbol{ZU}(\mathbb{V}\mathbb{D}^{-1/2}\mathbb{V}^T) + \boldsymbol{E}(\mathbb{V}\mathbb{D}^{-1/2}\mathbb{V}^T)\end{aligned} \tag{4.1}$$

If the noise $\mathbf{E}$ has spatial correlations structure and the model $\hat{\mathbf{M}}$ is exactly true noise correlation, then after prewhitening, the noise correlation structure will become the i.i.d. across voxels (the second term of Eq. 4.1). However, the prewhitening also impacts on signal (the first term of Eq. 4.1). This impact can be understood as projecting the signal onto the eigenvectors of the noise correlation matrix (columns of $\mathbb{V}$), and reweighting them by the inverse square root of the eigenvalue ($\mathbb{D}^{-1/2}$). These reweighted vectors are then projected back into voxel space. Thus, prewhitening de-emphasizes spatial modes with strongly correlated noise.

If the signal has no particular spatial structure (i.i.d.) then on average each spatial mode should carry equal amounts of signal variance - hence signal variance should not be changed by prewhitening. Thus, for the i.i.d signal, the reliability increases linearly with the effective number of voxels. However, if the signal is not the i.i.d., prewhitening may attenuate the data along spatial mode where there is a strong signal. For example, if the estimated noise correlations are large and constant even for voxels with large spatial distances, the spatial model corresponding to the mean activity will have a high eigenvalue and will be therefore suppressed. If differences between conditions are mostly due to overall activity differences (across all voxels), this type of prewhitening will have potentially lower reliability.

Thus, it is important to keep in mind that spatial prewhitening of the data not only deals

with correlations in the measurement noise but also changes the spatial modes of the signal that we are studying. For example, our model will tend to de-emphasize the superficial voxel of the cortex (as they are more correlated to their neighbours) and emphasize deeper voxels. How this changes the multivariate inference, i.e., what information is carried in different spatial modes of the activity patterns, is an important topic for future research. A better model of spatial noise correlation, as developed here, is an important tool for this investigation.

Limitation: One limitation of our study is that we only focused on neocortical data to construct our model. It is unclear which factors determine the noise correlation in other brain areas, such as the cerebellum, hippocampus, and thalamus. It is likely that in those areas both the importance of the different factors and the parameters associated with these factors will be different from that in neocortical areas. An important future step is therefore to train our model also with data from these areas.

Conclusion: This study proposes a new model for spatial noise correlation for fMRI data from the human neocortex. We trained and tested the model in a wide range of regions and datasets, using different field strengths and spatial resolutions. We anticipate this work will be useful for two important applications. First, the model itself provides valuable insight into the structure of noise correlations. It can help to understand which source of noise dominates the fMRI measurement error. Second, the novel anatomically-informed noise correlation model provides the potential to improve multivariate inferences. Our analyses clearly show that the use of the Double Exp3d Depth model as the shrinkage target provides better inference than the existing independent prior, the identity matrix. Moreover, given that our acquisition parameters covered different voxel sizes, regions of the neocortex, and fMRI experiments, the model constructed in this study can be applied in any neocortical area and at any spatial resolution.

Bibliography

L. Abbott. Theoretical neuroscience rising. *Neuron*, 60(3):489–495, 2008. ISSN 0896-6273. doi: https://doi.org/10.1016/j.neuron.2008.10.019. URL https://www.sciencedirect.com/science/article/pii/S0896627308008921.

S. A. Arbuckle, J. A. Pruszynski, and J. Diedrichsen. Mapping the integration of sensory information across fingers in human sensorimotor cortex. *Journal of Neuroscience*, 42(26): 5173–5185, 2022.

M. J. Arcaro, C. J. Honey, R. E. Mruczek, S. Kastner, and U. Hasson. Widespread correlation patterns of fmri signal across visual cortex reflect eccentricity organization. *Elife*, 4:e03952, 2015.

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.

J. Bodurka, F. Ye, N. Petridou, K. Murphy, and P. A. Bandettini. Mapping the mri voxel volume in which thermal noise matches physiological noise—implications for fmri. *Neuroimage*, 34 (2):542–549, 2007.

M. R. Brier, A. Mitra, J. E. McCarthy, B. M. Ances, and A. Z. Snyder. Partial covariance based functional connectivity computation using ledoit–wolf covariance regularization. *NeuroImage*, 121:29–38, 2015.

D. D. Cox and R. L. Savoy. Functional magnetic resonance imaging (fmri)"brain reading": detecting and classifying distributed patterns of fmri activity in human visual cortex. *Neuroimage*, 19(2):261–270, 2003.

T. Davis, K. F. LaRocque, J. A. Mumford, K. A. Norman, A. D. Wagner, and R. A. Poldrack. What do differences between multi-voxel and univariate analysis mean? how subject-, voxel-, and trial-level variance impact fmri analysis. *Neuroimage*, 97:271–283, 2014.

J. A. de Zwart, P. v. Gelderen, M. Fukunaga, and J. H. Duyn. Reducing correlated noise in fmri data. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 59(4):939–945, 2008.

J. Diedrichsen, T. Wiestler, and N. Ejaz. A multivariate method to determine the dimensionality of neural representation from population activity. *Neuroimage*, 76:225–235, 2013.

J. Diedrichsen, E. Berlot, M. Mur, H. H. Schütt, M. Shahbazi, and N. Kriegeskorte. Comparing representational geometries using whitened unbiased-distance-matrix similarity. *arXiv preprint arXiv:2007.02789*, 2020.

J. Diedrichsen et al. Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS computational biology*, 13(4):e1005508, 2017.

P. Domingos and M. Pazzani. On the optimality of the simple bayesian classifier under zero-one loss. *Machine learning*, 29:103–130, 1997.

K. J. Friston, P. Jezzard, and R. Turner. Analysis of functional mri time-series. *Human brain mapping*, 1(2):153–171, 1994a.

K. J. Friston, K. J. Worsley, R. S. Frackowiak, J. C. Mazziotta, and A. C. Evans. Assessing the significance of focal activations using their spatial extent. *Human brain mapping*, 1(3): 210–220, 1994b.

B. Hansen, M. Chelaru, and V. Dragoi. Correlated variability in laminar cortical circuits. *Neuron*, 76(3):590–602, 2012. ISSN 0896-6273. doi: https://doi.org/10.1016/j.neuron.2012.08.029. URL https://www.sciencedirect.com/science/article/pii/S0896627312007660.

V. Harms and L. Elias. *Brain, Tissue*, pages 262–263. Springer New York, New York, NY, 2013. ISBN 978-1-4419-1005-9. doi: 10.1007/978-1-4419-1005-9_1105. URL https://doi.org/10.1007/978-1-4419-1005-9_1105.

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.

N. Kriegeskorte and J. Diedrichsen. Inferring brain-computational mechanisms with models of activity measurements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1705):20160278, 2016.

N. Kriegeskorte and J. Diedrichsen. Peeling the onion of brain representations. *Annual review of neuroscience*, 42(1):407–432, 2019.

N. Kriegeskorte, M. Mur, and P. A. Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, page 4, 2008.

L. I. Kuncheva. On the optimality of naïve bayes with dependent binary features. *Pattern Recognition Letters*, 27(7):830–837, 2006.

O. Ledoit and M. Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004. ISSN 0095-4918. doi: 10.3905/jpm.2004.110.

C.-S. J. Liu, A. Miki, J. Hulvershorn, L. Bloy, E. E. Gualtieri, G. T. Liu, J. S. Leigh, J. C. Haselgrove, and M. A. Elliott. Spatial and temporal characteristics of physiological noise in fmri at 3t. *Academic Radiology*, 13(3):313–323, 2006. ISSN 1076-6332.

doi: https://doi.org/10.1016/j.acra.2005.10.018. URL https://www.sciencedirect.com/science/article/pii/S1076633205009165.

T. T. Liu. Noise contributions to the fmri signal: An overview. *NeuroImage*, 143:141–151, 2016.

R. J. Mackenzie. Gray matter vs white matter, 2019. URL https://www.technologynetworks.com/neuroscience/articles/gray-matter-vs-white-matter-322973. [Online; accessed 23-March-2023].

U. S. D. S. Medicine. What is fmri, 2023. URL http://fmri.ucsd.edu/Research/whatisfmri.html.

T. Nakai and S. Nishimoto. Quantitative models reveal the organization of diverse cognitive functions in the brain. *Nature communications*, 11(1):1142, 2020.

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.

N. N. Oosterhof, T. Wiestler, P. E. Downing, and J. Diedrichsen. A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage*, 56(2):593–600, 2011.

R. A. Poldrack, J. A. Mumford, and T. E. Nichols. *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.

J. R. Polimeni, V. Renvall, N. Zaretskaya, and B. Fischl. Analysis strategies for high-resolution uhf-fmri data. *NeuroImage*, 168:296–320, 2018. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2017.04.053. URL https://www.sciencedirect.com/science/article/pii/S1053811917303713. Neuroimaging with Ultra-high Field MRI: Present and Future.

J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen. Spurious but systematic correlations in functional connectivity mri networks arise from subject

motion. *NeuroImage*, 59(3):2142–2154, 2012. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2011.10.018. URL https://www.sciencedirect.com/science/article/pii/S1053811911011815.

M. Rahim, B. Thirion, and G. Varoquaux. Population shrinkage of covariance (posce) for better individual brain functional-connectivity estimation. *Medical image analysis*, 54:138–148, 2019.

R. Rosenbaum, M. A. Smith, A. Kohn, J. E. Rubin, and B. Doiron. The spatial structure of correlated neuronal variability. *Nature neuroscience*, 20(1):107–114, 2017.

F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics bulletin*, 2(6):110–114, 1946.

M. Shahbazi, A. Shirali, H. Aghajan, and H. Nili. Using distance on the riemannian manifold to compare representations in brain and in models. *NeuroImage*, 239:118271, 2021.

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.

N. Shea. *Representation in cognitive science*. Oxford University Press, 2018.

L. L. Wald and J. R. Polimeni. Impacting the effect of fmri noise through hardware and acquisition choices–implications for controlling false positive rates. *Neuroimage*, 154:15–22, 2017.

A. Walther, H. Nili, N. Ejaz, A. Alink, N. Kriegeskorte, and J. Diedrichsen. Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage*, 137:188–200, 2016.

T. Wiestler, D. J. McGonigle, and J. Diedrichsen. Integration of sensory and motor representations of single fingers in the human cerebellum. *Journal of neurophysiology*, 105(6):3042–3053, 2011.

T. Wiestler, S. Waters-Metenier, and J. Diedrichsen. Effector-independent motor sequence representations exist in extrinsic and intrinsic reference frames. *Journal of Neuroscience*, 34 (14):5054–5064, 2014.

K. J. Worsley and K. J. Friston. Analysis of fmri time-series revisited—again. *Neuroimage*, 2 (3):173–181, 1995.

E. Zarahn, G. Aguirre, and M. DEsposito. Empirical analyses of bold fmri statistics. 1. spacially unsmoothed data collected under null-hypothesis conditions (vol 5, pg 179, 1997), 1997.

Appendix A

A.1 Pre-trained model parameters for the Double Exp3d Depth:

$$\gamma_1 = -1.780, \alpha = -3.386, \gamma_2 = -0.974, \beta = -1.628, \mu = -0.241, \theta = -0.664$$

A.2 Derivation of variance of G

Under the null hypothesis that there is no difference between true activity pattern (i.e., $\mathbf{U} = 0$), the variance of $\hat{\boldsymbol{G}}_{ij}$ can be derived from following:

$$
\begin{aligned}
Var(\hat{\boldsymbol{G}}_{ij}) &= Var\left(\frac{\tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_j^T}{p}\right) \\
&= \frac{1}{p^2}Var(\tilde{\mathbf{e}}_i\tilde{\mathbf{e}}_j^T) \text{ (under the null hypothesis)} \\
&= \frac{1}{p^2}Var(\sum_{m=1}^{p}\tilde{e}_{im}\tilde{e}_{jm}) \\
&= \frac{1}{p^2}\left(\mathbb{E}(\sum_{m=1}^{p}\tilde{e}_{im}\tilde{e}_{jm})^2 - \left(\mathbb{E}(\sum_{m=1}^{p}\tilde{e}_{im}\tilde{e}_{jm})\right)^2\right) \\
&= \frac{1}{p^2}\left(\mathbb{E}(\sum_{m=1}^{p}\sum_{n=1}^{p}\tilde{e}_{im}\tilde{e}_{jm}\tilde{e}_{in}\tilde{e}_{jn}) - \left(\sum_{m=1}^{p}\mathbb{E}(\tilde{e}_{im}\tilde{e}_{jm})\right)^2\right) \\
&= \frac{1}{p^2}\left(\sum_{m=1}^{p}\sum_{n=1}^{p}\mathbb{E}(\tilde{e}_{im}\tilde{e}_{jm}\tilde{e}_{in}\tilde{e}_{jn}) - \left(\sum_{m=1}^{p}\mathbb{E}(\tilde{e}_{im}\tilde{e}_{jm})\right)^2\right) \\
&= \frac{1}{p^2}\left(\sum_{m=1}^{p}\sum_{n=1}^{p}\mathbb{E}(\tilde{e}_{im}\tilde{e}_{in})\mathbb{E}(\tilde{e}_{jm}\tilde{e}_{jn})\right) \text{ (because } \mathbb{E}(\tilde{e}_i) = 0\,\forall i \text{ and assume } \tilde{e}_i \text{ is independent of } \tilde{e}_j) \\
&= \frac{1}{p^2}\sum_{m=1}^{p}\sum_{n=1}^{p}(\boldsymbol{V}_R)_{mn}(\boldsymbol{V}_R)_{mn} \\
&= \frac{1}{p^2}\sum_{m=1}^{p}\left(\sum_{n=1}^{p}(\boldsymbol{V}_R)_{mn}(\boldsymbol{V}_R)_{nm}\right) \\
&= \frac{1}{p^2}\sum_{m=1}^{p}\left(\sum_{n=1}^{p}\sigma^2(\boldsymbol{\Sigma}_R)_{mn}\sigma^2(\boldsymbol{\Sigma}_R)_{nm}\right) \\
&= \frac{\sigma^4}{p^2}\sum_{m=1}^{p}\left(\sum_{n=1}^{p}(\boldsymbol{\Sigma}_R)_{mn}(\boldsymbol{\Sigma}_R)_{nm}\right) \\
&= \frac{\sigma^4}{p^2}\sum_{m=1}^{p}(\boldsymbol{\Sigma}_R\boldsymbol{\Sigma}_R)_{mm} \\
&= \frac{\sigma^4}{p^2}trace(\boldsymbol{\Sigma}_R\boldsymbol{\Sigma}_R) \\
&\propto \begin{cases} 1/\nu \text{ (if } \boldsymbol{\Sigma}_R \neq \mathbf{I}_p) \\ 1/p \text{ (if } \boldsymbol{\Sigma}_R = \mathbf{I}_p) \end{cases}
\end{aligned}
$$

where $\tilde{\mathbf{b}}_i \in \mathbb{R}^{1\times p}$ is i-th row of $\tilde{\mathbf{B}}$ across voxels in Eq. 2.8, $\tilde{\mathbf{b}}_i \sim MVN(\mathbf{0}, \mathbf{V}_R)$ where $\mathbf{V}_R = \sigma^2\boldsymbol{\Sigma}_R$ under equal variance assumption. $\tilde{\mathbf{b}}_i$ is independent of $\tilde{\mathbf{b}}_j$ for $1 \leq i, j \leq Q$. $\tilde{\mathbf{e}}_i \in \mathbb{R}^{1\times p}$ is the i-th

row of $\tilde{\mathbf{E}}$ across voxels.

Curriculum Vitae

Name:	Lingling Lin
Post-Secondary Education and Degrees:	University of Victoria Victoria, BC, Canada 2016 - 2021 B.Sc Western University London, ON, Canada 2021 - 2023 M.Sc
Related Work Experience:	Teaching Assistant Western University 2021 - 2023 Research Assistant Western University 2021 - 2023