Examining the performance of trend surface models for inference on Functional Magnetic Resonance Imaging data

Divya B.

A report presented towards partial fulfillment of requirements for the degree of

Doctor of Philosophy

in

Management and Actuarial Sciences:
Curriculum in Statistics and Actuarial Sciences.

Dipartimento di Scienze Economiche e Statistiche
Universita degli Studi di Udine
Italy

October 31, 2018

Acknowledgments

I would like to thank the University of Udine for providing me with this opportunity to pursue my interests. I am very grateful to my supervisor Prof. Nicola Torelli for his constant support throughout the process.

I would like to express my gratitude to Prof. Christian Beckmann for giving me the opportunity to work at the DCCN, and for his valuable guidance. I am extremely grateful to Prof. Andre Marquand for his careful guidance, immense patience and constant support throughout the difficult process of understanding and working in a new domain with unfamiliar tools.

I am thankful for my encouraging and helpful classmates at Padova, who made the transition into a new field easier.

I am very grateful for Aditya, and for his unwavering support and encouragement, his nuggets of wisdom about life and PhD, and his terrible humor.

Abstract

The current predominant approach to neuroimaging data analysis is to use voxels as units of computation in a mass univariate approach which does not appropriately account for the existing spatial correlation and is plagued by problems of multiple comparisons. Therefore, there is a need to explore alternative approaches for inference on neuroimaging data that accurately model spatial autocorrelation, potentially providing better type I error control and more sensitive inference. In this project we examine the performance of a trend surface modeling (TSM) approach that is based on a biologically relevant parcellation of the brain. We present our results from applying the TSM to both task fMRI and resting-state fMRI and compare the latter to the results from the parametric software, FSL. We demonstrate that the TSM provides better Type I error control, as well as sensitive inference on task data.

Contents

1. Introduction

The development of advanced medical imaging techniques over the last few decades have revolutionized medical diagnosis and treatment. Especially in the case of the brain which is one of the most complex systems existing in nature, the advent of high resolution non-invasive imaging techniques has transformed our understanding of an organ that is otherwise difficult to study. Functional magnetic resonance imaging (fMRI) in particular has proved to be an invaluable tool for the study of the functioning of the brain. Several statistical methods have been developed over the last thirty years to model and draw inferences from fMRI data.

The main goal of fMRI studies is to understand the behavior of neurons in response to presence (or absence) of external stimuli in order to draw conclusions about the areas of the brain that control the performance of a specific task, or to study the connectivity of different brain regions. The fMRI achieves this by capturing the nature of blood flow to different brain regions, based on the understanding that an activated region of the brain consumes higher levels of oxygenated blood. Since blood flow cannot be observed at the neural level the image space of the brain is partitioned into three-dimensional sub-units called voxels. Hence a typical brain scan from fMRI is composed of hundreds of thousands of voxels. This gives fMRI scans high spatial resolution and also high localization power for the observed signals, thus allowing reasonable accuracy in pin-pointing regional activations.

However, there is a downside to the large number of voxels in fMRI – the predominant approach to fMRI data analysis has been the mass-univariate approach, which considers the voxels to be independent and fits a univariate linear model at each voxel. As a result, when performing inference over the whole brain or large regions of the brain one is faced with the problem of multiple comparisons and has to correct for it. Since fMRI is spatial in nature the voxels are spatially correlated, contrary to the assumptions of the mass-univariate approach, thus requiring complex post-hoc correction procedures for the multiple comparisons problem – such as the Random Field Theory. Thus the commonly used fMRI data analysis software such as FSL and SPM are based on the combination of the mass-univariate approach and the Random Field Theory.

In Eklund et al. (2016) [1], the authors examined the statistical validity of SPM and FSL using resting state fMRI data as the real-world null data and found that these packages provide conservative voxelwise inference and invalid clusterwise inference, with a high degree of false positives. They conclude that this is because the spatial nature of the real data violates the assumptions made by the Random Field Theory regarding the spatial autocorrelation of the data. As a result, there is a need for alternative approaches for inference on fMRI data that accurately model its spatial nature and potentially provide optimal Type I error control, as well as sensitive inference.

In this context we examine the performance of a trend surface model proposed by Huertas, I. et al. (2017) [2]. In contrast to the classical approaches, the trend surface model uses a parsimonious set of biologically meaningful basis functions derived from Instantaneous Connectivity Parcellation [3] (typically consisting of a few hundred basis functions) instead of voxels, and a

Bayesian regression model is employed to a find an appropriate linear-weighted sum of these basis functions to fit an imaged brain region of interest. We test this approach on the resting-state fMRI data used by Eklund et al. (2016) in line with their analysis and find that the trend surface model provides optimal inference as opposed to the inflated false positives from FSL (Fig. 4.3). We further test the trend surface model on task fMRI data to confirm that the good specificity is not at the cost of its sensitivity, and we demonstrate that this is indeed the case, using task data from multiple tasks from the Human Conectome Project (Figs. 4.4 – 4.7).

Organization of manuscript

The rest of this manuscript is organized as follows:

Chapter 2 introduces functional magnetic resonance imaging (fMRI). After a brief description of different neuroimaging modalities and the principles of image acquisition in MRI, we describe the process of fMRI signal acquisition, the properties of the signal and the challenges that arise as a result. We also introduce the experimental designs and the pre-processing steps used for fMRI studies and provide information about the statistical software that was used for the research presented in this thesis. The goal of Chapter 2 is to provide a familiarity with the nature of fMRI data in order to enable a better appreciation for the choice of statistical methods to analyze the data, which are presented in Chapter 3.

Chapter 3 gives the theoretical background for the research presented in this thesis. We first provide an overview of the types of statistical models commonly used and the main approaches for inference for fMRI data. Then we delve into more detail for the two approaches that are directly relevant to this thesis – the mass-univariate approach and the trend surface modeling approach. We describe the mass-univariate approach and the assumptions involved and consequently, the challenges that arise in the context of modeling fMRI data, namely, the multiple comparisons problem and the impact of spatial correlation. We then describe the trend surface modeling approach proposed by Huertas, I. et al. (2017) [2] – the performance of which is the focus of research for this thesis. The aim of Chapter 3 is to lay down the theoretical groundwork and explain the state-of-the-art in fMRI data analysis that motivated our research, which is presented in Chapter 4.

Chapter 4 presents our research – the performance of the trend surface model as compared to the popular mass-univariate based software for fMRI data analysis. Our work involves three main parts – (i) the replication of previous work published by Eklund et al. (2016) on the validity of popular software packages for fMRI data analysis with respect to Type I error control, (ii) a similar analysis using the trend surface model on the same resting-state fMRI dataset in order to provide an equitable comparison of the specificity of the trend surface model , and (iii) analyzing task fMRI data from the Human Connectome Project to test the sensitivity of the trend surface model . We provide a detailed description of the resting-state and task fMRI data used in this research, followed by the steps involved in our analysis of this data at the subject and group levels. Finally we present our results to demonstrate that the trend surface model provides optimal Type I error control as well as sensitive inference.

Chapter 5 provides a discussion of our research. We provide a recapitulation for the motivation and relevance of our research and discuss some of the challenges that we faced in the course of our work. We also mention possible solutions to these challenges, which will be the focus of our future research.

2. Introduction to Functional Magnetic Resonance Imaging

An Overview of Neuroimaging

Over the last 30 years neuroimaging techniques have become invaluable tools for studying the brain by virtue of their high spatial resolution and non-invasive nature. The development of several different modalities of imaging the brain such as PET, (f)MRI, DTI and CAT to name a few, have allowed researchers much insight into the structural and functional aspects of the brain.

Structural neuroimaging enables the study of the structural aspects of the brain – such as the contrast between different tissues (cerebrospinal fluid (CSF), grey matter and white matter) and therefore plays a key role in the diagnoses of structural anomalies in the brain, such as tumors, strokes and injury. Computed Tomography (CT) and structural MRI are the most popular structural imaging techniques.

Functional imaging techniques enable the study of the physiology of mental processes by capturing neuronal activity in response to a stimulus, thus making them invaluable in the study of brain function, as well as dysfunction as in the case of psychiatric disorders. The neuronal activity in the brain can be studied in two ways: (i) through electricity, which is a direct measure of neural activity and (ii) through blood flow in the brain, which is an indirect measure of neural activity. Functional techniques such as PET and fMRI utilize the blood flow patterns in the brain while EEG and MEG use the electrical impulses in the brain to measure neuronal activity. In this section we briefly describe structural and functional neuroimaging techniques.

A. Structural neuroimaging techniques

Computed Tomography (CT)

CT is a tomographic modality, that is, it produces cross–sectional images. CT forms an image by sending energy in the form of X-rays into the body and measuring how that energy is altered when passing through the tissues. The amount of energy absorbed by body tissue is calculated from the energy remaining in the beam after the beam passes through the body tissue. The X-ray beam is applied at different angles to generate a cross-sectional image. The amount of energy absorbed is reflected in the observed brightness of tissue. For instance, since bone is one of the densest tissues in the body it absorbs more X-rays and appears bright on the CT scan. Tissues of low-density, such as blood and CSF, absorb much lesser energy and appear darker.

Structural MRI

MRI broadly works similar to CT scans but instead of X-rays, radio frequency (RF) waves are used. As a result the image brightness and contrast, and hence the interpretation of the images,

are based on completely different principles. Brightness in an MRI scan is indicative of the tissue interaction with radio waves inside a magnetic field. The radio waves and magnetic fields used in MRI can be manipulated in many ways to alter the appearance of specific tissues, thus enabling MRI to detect the distinctions between different soft tissues, for example. Since MRI does not use ionizing radiation like the X-rays, there is much lesser risk of tissue damage from repeated scans, compared to CT.

B. Functional neuroimaging techniques

Functional imaging techniques can be classified based on whether they use blood flow (haemodynamic techniques) or neural electrical impulses (electro-magnetic techniques) to measure brain activity. Haemodynamic techniques offer very good spatial resolution ($\sim 2mm$) but poorer temporal resolution ($\sim 1s$) compared to electro-magnetic techniques (< 1 milli-second) and the latter have poor spatial resolution. Therefore the choice of the modality depends on the specific needs of the study at hand.

Electroencephalography (EEG)

Neurons communicate with each other through tiny electrical impulses at the rate of thousands of times per second. In the brain, populations of neurons are organized into connected networks. When these networks fire in sync, the dynamics of the combined electrical activity can be detected and recorded outside the skull. Thus electrical impulses can be measured at milli-second resolution, but it is hard to localize the electrical signal since we do not have realistic models of the head and different tissues have different conductivity properties and hence distort the electrical signal differently, before it is measured outside the skull (Fig. 2.1).

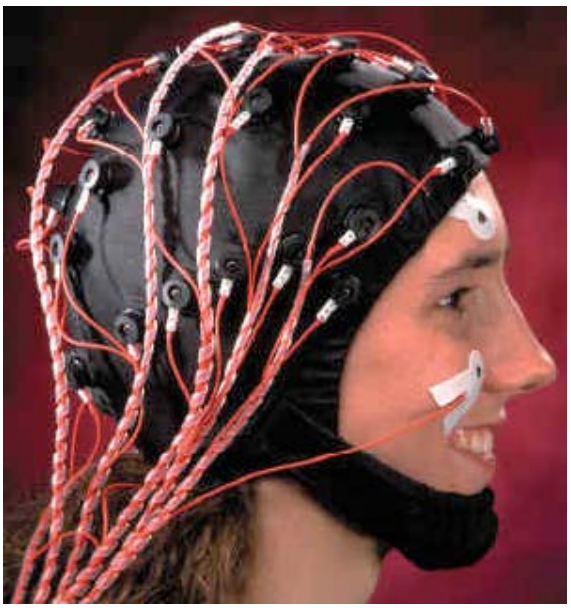

Figure 2.1: Apparatus for EEG. Source: Functional Neuroimaging (http://www.psych.nyu.edu/pylkkanen/Neural_Bases/07_slides/05_Methods.pdf)

Magnetoencephalography (MEG)

MEG is based on the natural phenomenon of the creation of magnetic fields around electric currents. The main advantage of MEG over EEG is that it offers better spatial resolution – in the order of mm for the cortex, and this deteriorates as the source of the signal is located deeper in the brain. This is possible because unlike electrical impulses, magnetic fields pass through

skull and various tissues undistorted. Thus the distribution of the magnetic field around the head provides a good indication of the underlying electrical activity in the brain. The magnetic fields generated by neural activity are tiny – about a hundredth the strength of the Earth's magnetic field and a millionth the strength of the fields in an urban environment. Therefore great care needs to be taken to capture the tiny signals from the brain – this is ensured by placing the MEG apparatus in a magnetically shielded room and using superconductive sensors in the MEG machine (Fig. 2.2).

Figure 2.2: Apparatus for MEG, placed inside a magnetically shielded room. Source: Functional Neuroimaging (http://www.psych.nyu.edu/pylkkanen/Neural_Bases/07_slides/05_Methods.pdf)[4]

Positron emission tomography (PET)

PET works by introducing radioactive energy via a radioactive tracer into the body and measuring the decay of the radioactive element, to form an image of the radioactive quantity and distribution. Injected contrast agents emit tiny particles called positrons that are detected by the scanning device. PET imaging is unique because the radioactive material used can be attached to various substances like glucose. The amount of glucose accumulated in a tissue is indicative of tissue metabolism, and hence normal tissue can be differentiated from cancerous tissue, as the latter has much higher rates of metabolism. PET images are of lower resolution compared to MRI but provide unique information on organ function. There are also hybrid systems, for example PET/CT and PET/MRI, that combine both a high resolution anatomical image and a corresponding image of metabolism in the same exam. Areas of metabolic activity can be directly correlated to anatomic location, which is useful in tumor detection and removal.

Other medical imaging techniques such as nuclear medicine and SPECT also produce images by measuring the decay of injected radioactive tracers, though in these cases the particles emitted are gamma rays.

Functional MRI (fMRI)

When a particular region of the brain is activated, there is an increased flow of oxygenated blood to support the increased metabolic activity in that region. fMRI uses magnetic field generated in an MRI scanner to measure the blood oxygenation level dependent (BOLD) signal based on the difference in magnetic properties of oxygenated and de-oxygenated blood. Therefore unlike PET, no radioactive tracers are required which makes fMRI safer and possibly cheaper than PET. fMRI also provides good spatial resolution and has enough temporal resolution to distinguish between trials (though not enough to distinguish the different stages of stimulus processing, due to the nature of the haemodynamic response, as explained in Section 2.2). The next section elaborates on the properties of fMRI, and the rest of this thesis provides a brief overview of statistical methods and software for fMRI analysis and results from the analysis of fMRI data to test a trend surface model .

Magnetic Resonance Imaging (MRI)

MRI is one of the several non-invasive imaging techniques that produce cross sectional images - known as tomographic images. Other widely used tomographic modalities are Computed Tomography (CT), Single Photon Emission Computed Tomography (SPECT), Positron Emission Tomography (PET) and Ultrasound. Since MRI uses radio waves there is no risk of radiation exposure, unlike CT (which uses X-rays) and PET or SPECT which require injection of a radioactive tracer into the body. Ultrasound also does not involve radiation exposure; however, since ultrasound measure the energy of sound waves and sound does not travel well through air or bone, some areas of the body such as the lungs and the skeleton cannot be properly imaged using ultrasound.

Figure 2.3: Apparatus for MRI. Source: Novel MRI-Based Scoring System Assesses Disease Severity in Juvenile Dermatomyositis, Sheila Jacobs (https://www.rheumatologyadvisor.com/pediatric-rheumatology/jdm-image-score-assesses-disease-severity-in-juvenile-dermatomyositis/article/779149/)

Medical imaging methods form an image by sending energy into the body and measuring how that energy is absorbed or changed when passing through the body's tissues. In MRI, the brightness in the image depicts the differences in how tissues interact with radio waves and is

therefore an indicator of the differences in molecular composition and environment of the tissues. The radio waves and magnetic fields used in MRI can be manipulated to alter the appearance of specific tissues for greater versatility through MRI sequences – which are particular settings of radiofrequency pulses and gradients resulting in a specific type of image. MRI acquisitions can also be made to show differences in blood flow (functional MRI), fat content, diffusion characteristics (diffusion-weighted MRI) and several other traits.

Magnetic resonance imaging is based on the principle of nuclear magnetic resonance. An atom belonging to any element is composed of a nucleus containing equal number of protons (with a positive electric charge) and neutrons (with no electric charge), and electrons (with a negative electric charge) that are equal in number to the protons. The magnetic properties of an element depend on the number and spin of electrons and protons in its atoms. Elements that are *ferromagnetic*, such as iron, are strongly attracted by a magnetic field and can be magnetised, acting as magnets themselves after the removal of the external magnetic field. Elements that have unpaired magnetic spins and are only slightly attracted or repelled by an external magnetic field are called *paramagnetic* (example, oxygen) or *diamagnetic* (example, gold) respectively.

An MRI scanner utilizes a strong magnetic field and radiofrequency pulses to excite the atoms in the body. Since hydrogen is by far the most abundant element in the body due to its presence in water, fat and protein molecules, the MRI scanners are often set up to collect images on the basis of hydrogen atoms. The hydrogen atom is composed of one proton (and one electron), and the charged and spinning proton produces a magnetic field of its own, thus acting as a tiny magnetic dipole. Under ordinary conditions, the magnetic dipoles of the hydrogen nuclei are oriented randomly, resulting in no net magnetization. However when the hydrogen nuclei are place in a strong external magnetic field such as that created inside an MRI scanner (say $\mathbf{B_0}$), the atomic dipoles are forced to align parallel (low energy state) or anti-parallel(high energy state) to the external field such that the number of dipoles aligned parallel is slightly larger than those aligned anti-parallel – producing a net magnetization M in the direction of the external field $\mathbf{B_0}$ (Fig. 2.4). At this stage, the hydrogen nuclei are in a state of equilibrium.

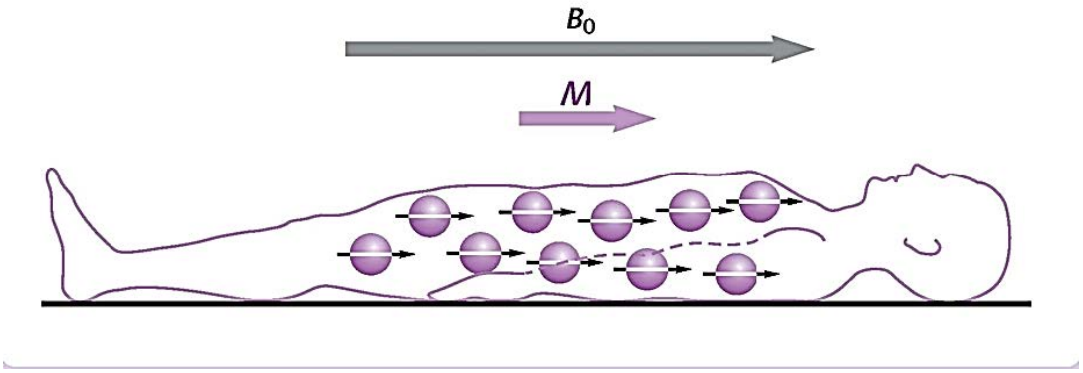

Figure 2.4: Illustration of magnetization of hydrogen nuclei in the body in the presence of a strong external magnetic field. Source: MRI for Technologists, Basic Principles of MRI, Daniel R. Thedens (http://www.icpme.us/)

In order to record images from a scanner, the equilibrium state of the hydrogen nuclei is perturbed by transferring energy to them through externally applied radio frequency (RF) pulses. The appropriate frequency of the RF pulse to be supplied is a function of the element (hydrogen in this case) and the externally applied magnetic field ($\mathbf{B_0}$ from the MRI scanner). When the RF pulse is turned off the excited nuclei release the absorbed energy, which is the source of the signal for the MRI scan images.

The release of the absorbed energy by the nuclei to return to their equilibrium state (called relaxation) happens through two events – decay of magnetization in the transverse plane (transverse relaxation) and decay of magnetization in the longitudinal direction (longitudinal relaxation). These two types of relaxation are characterized by the time constants T1 and T2 and are dependent on the particular tissue being observed. T1 describes how rapidly the relaxation process occurs in the longitudinal axis after the RF pulse is turned off. A short T1 value means that the longitudinal magnetization is restored rapidly and a longer T1 value means that the magnetization recovers more slowly.

On the other hand, T2 is a time constant that characterizes the decay of the MR signal after the RF pulse is turned off. A short T2 time means that the transverse magnetization is lost more quickly than it is for a tissue with a longer T2 time. T2 must be shorter than T1 and may be shorter by a factor of 10 or more. Once the transverse magnetization is lost, the signal to be recorded for imaging no longer exists. In a clinical context, these quantities can be used to determine different types of body tissue. The values of T1 and T2 can be manipulated to generate images that reflect subtle differences between soft tissues, resulting in the demonstration of fine anatomical details. Thus images can be acquired under various conditions that highlight (or minimize) the influence of T1 and T2, making MRI a powerful and flexibile technique.

Limitations of MRI

Since MRI uses strong magnetic fields, it cannot be used for patients with ferromagnetic metal in the body such as dental wires, cardiac pacemakers, surgical implants and shrapnel or other metallic foreign bodies, due to safety concerns. Many modern surgical implants such as stents and orthopedic screws are being made from MRI-compatible materials and do not provide a safety risk. However the presence of metal implants in or around the area of scanning may generate artifacts that make the scan partly unreadable. Also, MRI scans are typically longer than that of CT and ultrasound and therefore are prone to image artifacts generated by patient movement. MRI is also more expensive. Additionally, it is not an option for patients who are claustrophobic.

Functional Magnetic Resonance Imaging (fMRI)

fMRI data is used to serve several objectives of brain analysis. Broadly, these can be classified into four categories: (1) detection of activated brain regions (2) to infer brain connectivity (3) prediction of an individual's brain activity or clinical or behavioural response, and (4) imaging genetics. The presence (or absence) of neuronal activity in response to stimuli is used to describe brain function networks, to assess brain development, or to assess impairment related to conditions, such as alcoholism and brain trauma.

The fundamental goal of fMRI studies is to understand the behavior of neurons, most often either to uncover the regions of the brain that are activated by a particular task, or to understand how different regions of the brain are connected. Neural excitation occurs in milliseconds and cannot be directly observed with fMRI. However, neuronal activity is associated with localized changes in metabolism — as a region of the brain becomes active, there is an increase in local oxygen consumption resulting in more oxygen-rich blood flow to the active brain area. The magnetic properties of oxygen can therefore be exploited to measure what is called the blood oxygen level dependent (BOLD) signal contrast, which serves as a proxy for neuronal activation in fMRI studies

The BOLD response is not observed at the neural level. So the image space is partitioned into voxels in a rectangular 3D lattice (about 200,000 voxels for a 3 Tesla scanner) and the BOLD response is observed for each voxel every 2-3 seconds, at several hundred time points during the scan. Thus time series of BOLD responses are produced at every voxel, as the temporal evolution of brain activity at that location (Figure 2.5). As an fMRI experiment on a single subject can yield hundreds of scans in a single session, fMRI data are massive collections of hundreds of thousands of time series arising from spatially distinct locations. The brain image at each time point is captured one slice at a time rather than all at once. In most experiments, 30 to 500 images are collected.

Blood flow, however, is not a discrete on-off process, but instead something that happens

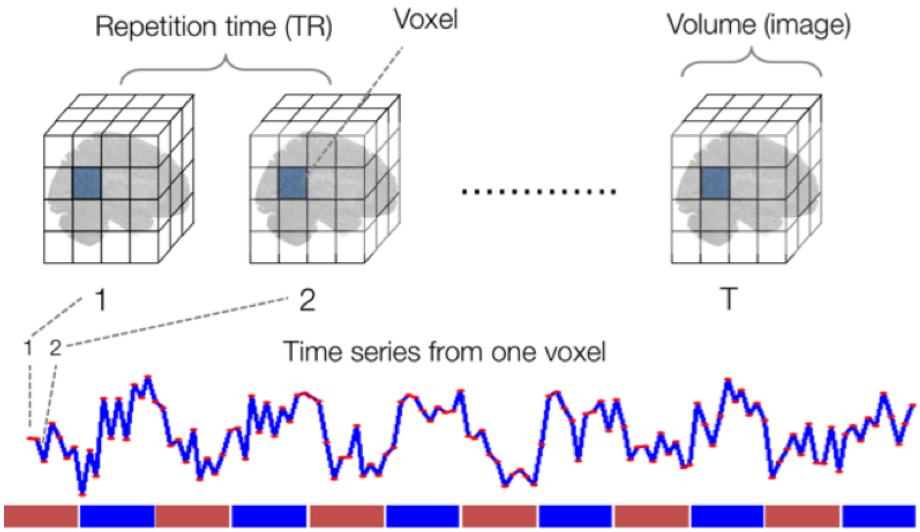

Figure 2.5: Obtaining time series at each voxel. Source: Lindquist, M. A. and Wager, T. D., Principles of fMRI (2015) [5]

continuously and depends on the location of the voxel and the nature of the task. When a neuron becomes active, there is typically a 1âĂŞ2 second delay before an increase in blood flow is detectable. Then blood flow peaks approximately 5âĂŞ8 seconds after neuronal activation and experiences a decay after 10âĂŞ12 seconds, finally returning to baseline levels after 20âĂŞ30 seconds (Fig. 2.6). The nature of dispersion of the BOLD signal both in space and time is described using a hemodynamic response function (HRF). We briefly describe common models for the HRF in this section.

It was observed by Miezin et al. (2000)[7] that the HRF fluctuates to varying degrees due to change in stimulus/task, change in activated brain region and between subjects — with the most variation observed between different brain regions (for the same subject and task). Despite these issues, neuroimaging researchers frequently assume a common HRF across the entire brain for the sake of simplicity during statistical analysis.

Several models for the HRF have been proposed. For instance, Friston et al. (1994)[8] proposed a Poisson function of the form

$$h(t) = \frac{e^{-d_s} d_s^{\tau-1}}{(\tau-1)!}$$

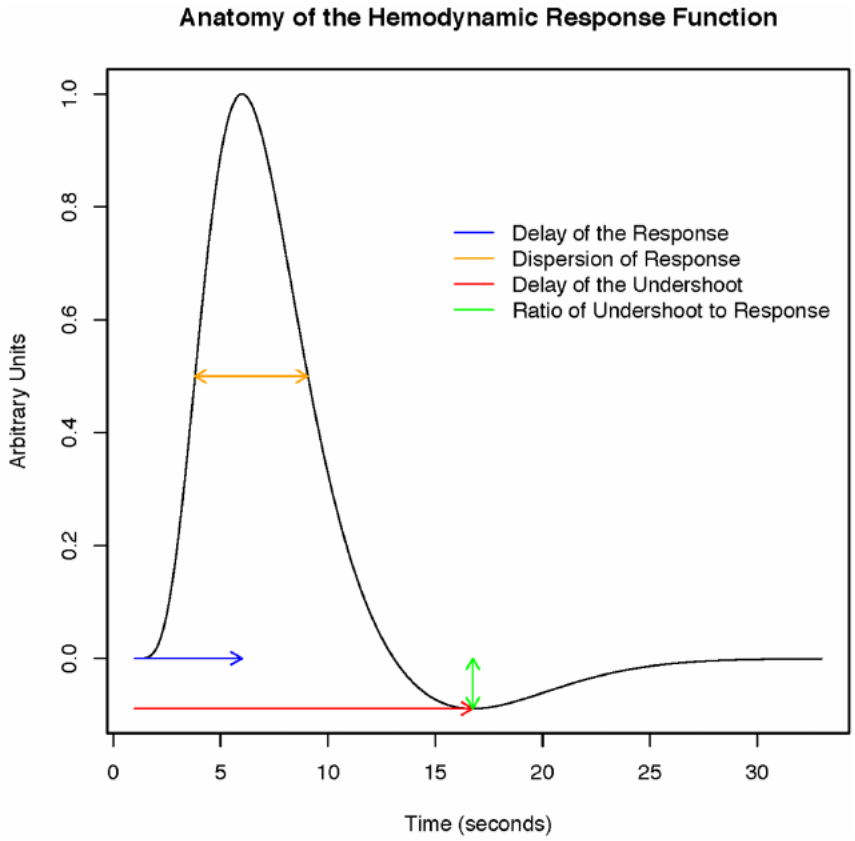

Figure 2.6: The standard canonical model for the HRF used in fMRI data analysis. Source: Analytic Programming with fMRI Data: A Quick-Start Guide for Statisticians Using R [6]

Lange et al (1997)[9] proposed gamma functions of the form

$$h(t) = \theta_2(\theta_2 t)^{\theta_1 - 1} exp\{(-\theta_2 t)/(\Gamma(\theta_1))\}$$

and Friston et al. (1995)[10] proposed Gaussian functions. A canonical HRF modeled using a difference of two gamma densities is given by Friston et al. (1998)[11]

$$h(t) = (t/d_1)^{a_1}\ exp\left\{-\frac{t-d_1}{b_1}\right\}\ -\ c(t/d_2)^{a_2}\ exp\left\{-\frac{t-d_2}{b_2}\right\}$$

Another common choice is the inverse logit function, generated as a superposition of three separate inverse logit functions, proposed by Lindquist et al (2007)[12]

$$h(t) = \alpha_1 L((t-T_1)/D_1) + \alpha_2 L((t-T_2)/D_2) + \alpha_3 L(t-T_3)/D_3), \quad L(x) = 1/(1+e^{-x})$$

The haemodynamic response varies in nature between subjects and between areas of the brain, and therefore there is a need to allow for flexibility in the shape of the HRF function. One approach is to parametrise the shape of the HRF and fit the shape parameters to the data, which involves non-linear fitting. An easier approach is to use linear basis sets to span the space of the expected HRF shapes, such as using linear combinations of temporal derivatives of an explanatory variable.

Constructing fMRI regressors

A voxel-wise fMRI analysis with an additive mean structure is most commonly assumed — since the BOLD signal is affected by many components acting independently of one another. Let $y_{v,t}$

be the BOLD response at voxel v at time t, $a_{v,t}$ be a baseline trend function, $f_{v,t}$ be the activation profile, and $\epsilon_{v,t}$ be the measurement error. Then, the assumed model is give by

$$y_{v,t} = a_{v,t} + f_{v,t} + \epsilon_{v,t}. \tag{2.1}$$

The linear variation of this model, for p stimuli and T_v images (or time series) collected at the v^{th} voxel, is given by

$$y_v = w_T \delta_v + X_v \beta_v + \epsilon_v \tag{2.2}$$

The most common statistical model of a time series of BOLD responses relies on the Gaussian linear model, as first proposed by Friston et al. (1994)[8] . This models the observed fMRI signal as the underlying BOLD response plus a noise component, assuming that the baseline behavior $w_T \delta_v$ is accounted for before statistical analysis.

A widely used model to account for the lapse of time between the stimulus onset and the vascular response considers the BOLD signal to be the convolution of the stimulus pattern with the hemodynamic response function (HRF) (Figure 2.7). This implies that in model (Eq.2.2) each column (task or input stimulus) of X_v is modeled as

$$\mathbf{X_v} = \int_0^1 x(s) h_v(t-s) ds,$$

where $x(s)$ is the known, external time-dependent stimulus function for a particular task.

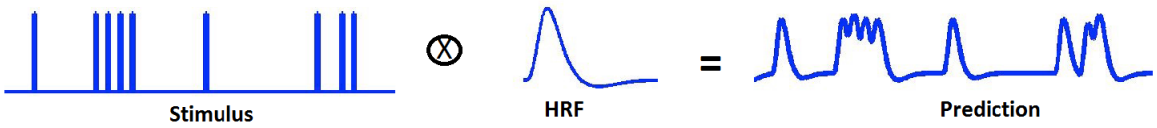

Figure 2.7: Obtaining the predicted response at a voxel. Source: FSL Course (http://fsl.fmrib.ox.ac.uk/fslcourse/) [13]

Spatial and temporal properties of fMRI

Since neuronal activity in the brain unfolds over space and time, the spatial and temporal resolution of fMRI studies have a huge impact on the conclusions that can be drawn from the data. The spatial resolution of fMRI determines its ability to distinguish between adjacent regions of the brain, and the temporal resolution determines the ability to separate brain events in time. The manner in which fMRI data is collected makes it impossible to simultaneously increase both spatial and temporal resolution (due to the limited number of k-space measurements; a topic beyond the scope of this thesis). It is therefore important to balance the need for spatial resolution with that of adequate temporal resolution.

While technology has evolved to allow for high spatial resolution scans through MRI, the time required to acquire such scans is prohibitively high. The potential high resolution of scans is further limited by several factors – such as the need to smooth fMRI data prior to analysis, and normalization of brains scans from different subjects to a common template (as explained in Section 2.2.3). As a consequence of all the blurring effects the activations in small regions may be mis-labelled or entirely missed. Advances in data acquisition and pre-processing can

potentially improve spatial inference in fMRI. Enhanced inter-subject normalization techniques, better smoothing techniques, improvements in signal acquisition (parallel imaging) are some such methods developed in recent years.

Temporal resolution of an fMRI study depends on the time interval between acquisition of each individual image (repetition time, denoted as TR), which typically ranges between 0.5 – 4 seconds. Though the underlying neuronal activity takes place on the order of tens of milliseconds, the BOLD response that is used as a proxy measurement for neuronal activity occurs in the order of seconds and peaks 5–8 seconds after the neural activation. Thus fMRI has a temporal limitation due largely to the latency and duration of the hemodynamic response to a neural event.

fMRI data exhibit short-range serial correlations that have to be modeled in order to obtain an accurate estimate of the degrees of freedom. Due to serial correlations the effective degrees of freedom are lower than in the case of independence, which would lead to a biased estimate of the standard error when calculating the t- or F-statistic and give invalid test results. Thus the covariance matrix has to be estimated assuming non-sphericity. This is achieved through generalized leasts squares estimation by decorrelating (or pre-whitening) the data and then obtaining an ordinary least squares (OLS) estimate. Another approach is to first obtain the OLS estimate assuming the errors are IID, and then perform a post-hoc correction. In addition to the serial correlations in the error, there are also correlations induced in the design matrix through the experimental design, caused by various sources such as heartbeat and breathing.

Noise in fMRI

The data in fMRI is very noisy as it is influenced by a variety of external sources of noise. The noise in fMRI data can be broadly classified into three types – thermal noise, system noise and subject/task-related noise [14]. Thermal and system noise occur in the scanner, while subject- and task- related noise is due to the motion of the subjects in the scanner.

Thermal motion occurs when electrons collide with atoms, and the rate of thermal motion goes up as the temperature of the system (MR scanner in this case) increases and also as the strength of the magnetic field increases. This leads to a distortion of the MR signal over the course of an imaging session. While in theory it is possible to eliminate thermal noise by reducing the temperature in the MR scanner, in practice this is not feasible. Therefore the effects of the thermal noise are mitigated by pre-processing the MR signal, typically by averaging over data points since thermal noise does not have spatial structure.

System noise is introduced due to the fluctuations in the functioning of the MR hardware, commonly due to non-uniform magnetic field and instability in the gradient fields. This causes a drift in the MR signal recorded; that is, over the course of an experiment the signal intensity at any given voxel gradually and systematically changes.

By far the most statistically interesting noise in fMRI data is generated by the subjects themselves. Since the scan records a time-series at each voxel that is typically 2 mm in size, even a small amount of motion is sufficient to displace the recorded activation to neighboring voxels and induce extra spatial correlation. A subject in the scanner is instructed to stay still, but since a scanning session lasts for several minutes subjects often find it hard to remain stationary for so long. In addition, head motion also occurs due to regular activities like swallowing, blinking, breathing and heartbeat, which cannot be avoided. Often, motion is also related to the task being performed by the subject in the scanner – for example, in a simple visual task where the subject has to follow a dot on the screen, there is a natural tendency to move the head along

with the eyes, even if slightly. In this case the head motion is correlated with the task and it is therefore necessary to remove the effects of head motion before studying the signal for task activation.

Pre-processing

Before functional MRI data can be analyzed it is typically subjected to several pre-processing steps to detect and correct for artifacts in the data (caused by the scanner or the subject - such as distortion correction and motion correction) or to prepare the data for statistical analysis (such as smoothing). The preprocessing of fMRI data involves a standard list of methods, though it can vary substantially between different software packages and research groups. We briefly explain some of the preprocessing steps that are employed for functional MRI data analysis, based on [15].

- **Slice timing correction:**

  Since most image-acquisition sequences acquire brain images slice by slice, there can be a difference of 1–3 seconds between the acquisition of the first slice and acquisition of the last slice, as a result of which the data from different slices are shifted in time relative to each other. However, an analysis assumes that all the voxels in an image acquired at a given time point of the time series are acquired at the same time. Therefore there is a need to calculate the signal intensity of all the slices at the same moment in the acquisition period. This is done by interpolating the signal intensity at the chosen time point from the same voxel in previous and subsequent acquisitions using interpolation techniques for signal reconstruction, such as bilinear and sinc interpolations. Event-related experiments require a more precise control over the onset time of the stimulus than block design experiments. Since the stimulus in block design typically lasts many seconds, the interpolation is often not necessary in block designs as there is not much sensitivity lost in all the slices are not collected at the same time.

- **Realignment/Motion Correction:**

  All subjects undergoing a scan move their heads to varying degrees during scanning and this can significantly impact the fMRI data collected, even if it is due to something like swallowing or breathing. This is due to a mismatch of the location of images in the time series obtained during the same scan. Such effects can be corrected by re-aligning the images in the time series to a common reference image using image registration methods and then re-slicing the aligned images to obtain re-aligned versions of the original images collected in the scan. The techniques which correct for simple movement of the head in the images typically model head motion as a rigid body transformation.

- **Smoothing:**

  Spatial smoothing essentially blurs the (high resolution) images obtained from the scan by removing the high-frequency information, which increases the signal-to-noise ratio for signals with larger spatial scales. Since activations in fMRI scans of the brain typically span many voxels, a gain in signal for larger features due to smoothing outweighs the cost of losing finer features. An additional advantage of smoothing is to reduce the mismatch of activation location across individuals in a group study that occurs due to the variability in the spatial location of functional regions across subjects. Smoothing also prepares the data for some analysis methods, such as the Gaussian Random Field Theory that requires a specific degree of spatial smoothness in order to satisfy the assumptions involved in random field theory.

Spatial smoothing is often achieved using the convolution of the three-dimensional image with a three-dimensional Gaussian kernel. The degree of spatial smoothing imposed by the kernel is determined by the full width of the distribution at half the maximum of the distribution (full-width at half-maximum, FWHM).

$$FWHM = 2\sigma\sqrt{2\ln(2)} \approx 2.55\sigma$$

where σ is the standard deviation of the applied kernel.

The smoothness of the image depends on the smoothing applied to the image during preprocessing, as well as the smoothness intrinsic to the image due to the correlated nature of fMRI activations across voxels. The resulting smoothness of the image is given by

$$FWHM = \sqrt{FWHM^2_{\text{intrinsic}} + FWHM^2_{\text{applied}}}.$$

- **Normalization:**

  MRI data of a single individual is of interest in the cases when it is necessary to understand the structural and functional aspects of that particular individual's brain, such as when planning a surgery to excise a tumor. However, most of the time fMRI data is collected from a group of individuals in order to draw inferences that can be extended to the human species in general. This requires that we should be able to integrate the data from different individuals to obtain summary statistics at the group level. However, even though human brains are very consistent in their overall structure across individuals, there exists a great deal of variability in the finer aspects of brain structure from subject to subject (such as in the size and the shape of the head). To allow for meaningful group level analyses it is necessary to perform spatial normalization to spatially transform the brain images from each individual to reduce the variability between individuals.

  Different types of reference frames are used to align individuals – three-dimensional Cartesian co-ordinate spaces (Talairach space), atlases (Talairach atlas) and templates (MNI templates).

Experimental Design

The experimental design of an fMRI study has to account not only for the standard issues present in psychological experiments, but also for issues related to data acquisition and stimulus presentation. Apart from the number of trials of a given set of conditions, the spacing and ordering of events is also a crucial aspect of the design. An optimal experimental design takes into consideration the psychological nature of the task and the ability of the fMRI signal to track changes introduced by the task over time. Additionally, it is necessary to consider the implications of the design for subsequent statistical analyses, as the efficiency of the statistical analysis is directly related to the experimental design. Therefore a good experimental design aims to maximize both the statistical power and psychological validity of the experiment.

Commonly used designs for neuroimaging studies are block design, event-related design and mixed block/event-related designs. This section focuses on block and event-related designs, which are the designs used in this research project. Both the designs have their benefits and drawbacks, and the choice of design is made based on the task being performed. For example, block designs are more efficient in detecting differences (such as pain vs. no pain), and event–related designs are more suitable to link neuronal activity to particular events.

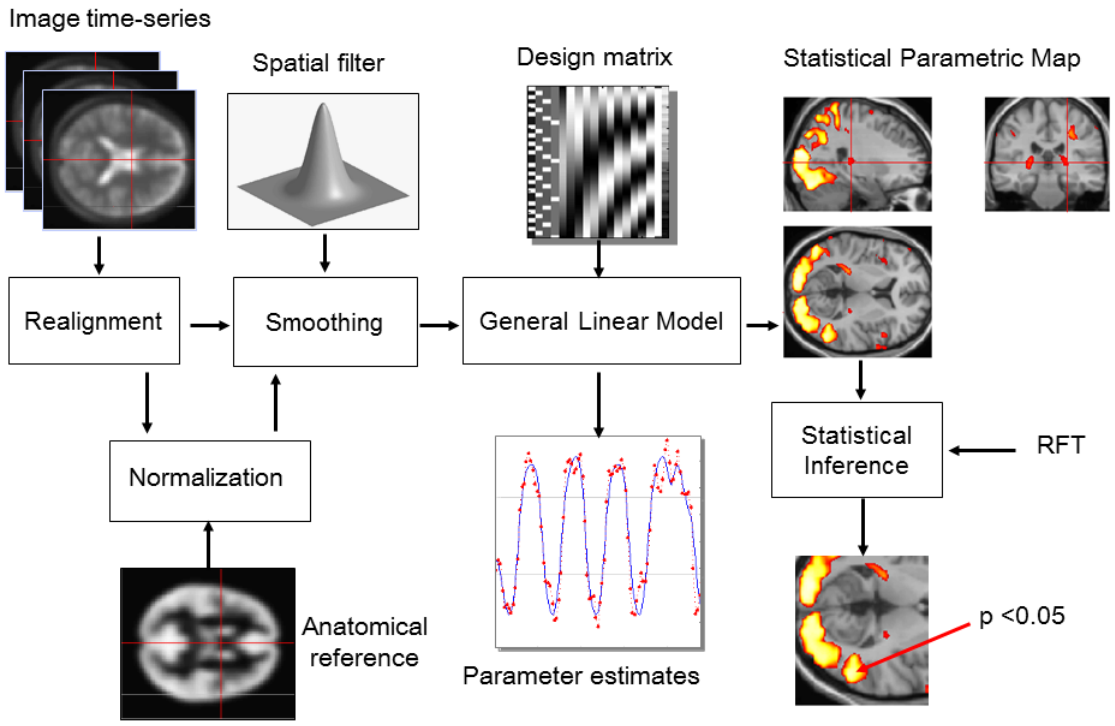

Figure 2.8: Typical steps for image processing and analysis. The primary data are commonly pre-processed, e.g. realigned, anatomically normalized, and filtered. A statistical model is chosen, determined by the experimental design and the previous model selection procedures. A test statistic is chosen and model parameters are estimated. Finally, statistical inference represented as a statistic image is obtained, taking into account multiple non-independent comparisons and possible temporal autocorrelation. (Source: Karl Friston, SPM workshop (May 2011))

Block design

Block design experiments utilize blocks of identical stimulus. For example, the tasks of interest (A and B) are repeated in alternating blocks, separated by a rest block. The comparison between the blocks is then used to compare the differences in signal between the two conditions (A - B) or between one of the conditions and rest. Increasing the length of each block leads to a larger evoked response and a greater separation in the two types of signal, which leads to increased power. Thus block design experiments are efficient since they aggregate across many trials to attain an adequate signal-to-noise ratio. They are also innately suited for detecting regions of interest (ROI) for particular tasks. Additionally, block designs are robust to uncertainties in the specification of the HRF, since the predicted BOLD signal depends on total activation caused by a series of identical stimuli rather than variations in shape of responses to individual stimuli.

However, the block design can not distinguish between trial types within a block (for example, correct versus error trials), nor can they identify interesting within trial or across trial events. They do not account for the transient responses at the beginning and end of task blocks. It is assumed that the same mental processes are evoked throughout each block, therefore long block lengths that lead to fatigue or boredom should be avoided so that this assumption is not violated.

Event-related design

Event-related designs are more complex task paradigms that allow the extraction of information regarding underlying neuronal activity from the BOLD signal. The stimuli (events) are presented briefly in random order (for example, flashing lights), with a random inter-stimulus interval (unlike the fixed interval in block designs, Fig. 2.9). The inter-stimulus interval should be short

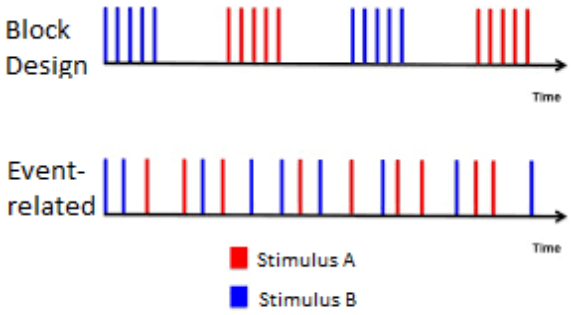

Figure 2.9: An illustration of a block design and an event-related design based on the same stimuli.

but not shorter than 2 – 5 s, in order to allow for the delay in the heamodynamic response. Event related designs are based on the assumption that neural activity will occur for short and discrete intervals. The random order of the stimuli rather than an alternating pattern offers more flexibility in experimental design. Such designs also allow for the estimation of key features of the HRF (such as onset and width). Another advantage of event-related designs is that the effects of fatigue and boredom can be avoided.A drawback of the event-related design is a decrease of signal-to-noise leading to less power than block designs of similar timing.

Statistical inference in Neuroimaging data

In neuroimaging, the frequentist approach to inference has been employed predominantly, as discussed in Chapter 3. Here we introduce some basics of frequentist inference.

Hypothesis Testing

Hypothesis testing plays a key role in limiting the subjective interpretation of a statistical model, by laying down rules and procedures to quantify decision errors. In order to test the occurrence of a particular phenomenon, it is observed under precise experimental conditions designed for the purpose. Under the null hypothesis (H_0), the observation of the phenomenon occurs by chance and is not related to the conditions of the experiment. What is an acceptable occurrence under chance (denoted by α) is decided beforehand, typically based on consensus in the field. If the observed occurrence exceeds this measure, then it is concluded that the phenomenon is not a spurious observation and the null hypothesis 'is rejected'.

When several measurements are made under specific experimental conditions, a summary statistic based on the multiple measurements is calculated as the test statistic that can be compared to its null distribution to decide how plausible such a statistic is under the null hypothesis. In neuroimaging, the F-statistic and the t-statistic are most commonly used. The p-value – defined as the chance of observing a test statistic as extreme or more extreme than the observed test statistic – quantifies the evidence against the null hypothesis for any test statistic.

When p-values are used to judge a null hypothesis two types of errors are possible. The rejection of H_0 when there is no effect is the false positive error or the Type I error, and the failure to reject the H_0 when there is indeed an effect is the false negative error or the Type II error. The ability of a statistical test to guarantee that false rejections of the null hypothesis are bound by the chosen $\alpha\%$ is the Type I error control of the test, while its ability to reject the null hypothesis when there is actually an effect is the Type II error control or power of the test. A statistical test is said to be 'valid' if the chance of a Type I error is less than or equal to α, and among such valid tests we look for the tests with the maximum power.

In neuroimaging, the false positive risk over the entire region of interest is measured mostly through the familywise error rate (FWER) and to some extent the more lenient false discovery rate (FDR). The FWER is defined as the chance of one or more false positives in the image. The FDR is the ratio of the false positives to all the positives detected in the image.

Challenges in Neuroimaging

In the absence of ground truth, it is difficult to choose the best procedure to analyze data. Neuroimaging analyses are largely comprised of group level analyses carried out to study behavioral or genetic variables to extend inference to a particular population. The main difficulty in such studies is due to the existence of inter-subject variability in brain shape and structure and the typical approach to solve this issue is to register and normalize the subject brains to a common reference space. Since it is not possible to achieve perfect voxel-to-voxel correspondence, additional smoothing is also carried out to reduce the effects of anatomical variability.

The main approach in neuroimaging is to fit and test a model at each voxel - the univariate voxel-wise approach - that results in several hypothesis tests being simultaneously performed, leading to the multiple comparisons problem (which is dealt with in more detail in section 3.2.1). In order to account for this multiplicity, measures of error such as the familywise error rater (FWER) and the false discovery rate (FDR) are defined, and controlled by procedures such as the Bonferroni correction and the Benjamini-Hochberg procedure, respectively. Accounting for the spatial correlation existing in the data and additionally introduced by pre-processing means that the above corrections are too conservative, since the voxels are no longer independent. This calls for more complex correction methods such as the Random Field Theory, which is explained in more detail in section 3.2.2.

Another challenge is to identify an appropriate function of time to represent the haemodynamic response, namely, the haemodynamic response function (HRF). While the qualitative character of the HRF is well known, the exact relationship between the quantity of interest, the neuronal activity and the BOLD response is unknown. The exact nature of the BOLD response (especially the delay until response) varies between subjects, and across the brain within a subject.

Yet another challenge lies in accounting for the temporal autocorrelation of fMRI data. Many approaches to modeling the BOLD response treat the response as a linear time-invariant system, which means that the predicted haemodynamic response is the convolution of a fixed HRF with the waveform of the experimental paradigm. The time invariance implies that the HRF is independent of time and previous responses, that is, it is stationary. However, it is known that the transformation over the experimental paradigm, the neural activity and the BOLD response to obtain the measured fMRI signal is not linear or time-invariant. It is to be noted that linearity assumption of the BOLD response is distinct from the linearity of the statistical model. Several recent approaches have been developed to take into consideration the non-linear time-invariant structure of the BOLD response.

The GLM is based on the assumption that the residual errors are mean zero, with constant variance and independent. However, since fMRI time-series display autocorrelation, estimates of variability are biased and as a result the calculation of the significance of effects and the false positive rates is altered. The linear model approaches to handle this problem are the generalized least squares (GLS) and the ordinary least squares (OLS) with adjustment for correlated errors. The GLS decorrelates or whitens the data and then applies the OLS, as shown in section 3.2. When using OLS on the correlated data, the null distribution of the test statistics is approximated by using the 'effective degrees of freedom' instead of the conventional degrees of freedom. In the

context of a mass univariate model, where the same regressors are used at every voxel in case of independent data, the presence of temporal autocorrelation means that the same model for the autocorrelation function should also be used at each voxel.

Some neuroimaging and data analysis softwares

This section focuses on the softwares that have been used in this thesis, which are mainly tools for visualization, image processing and statistical analysis of Neuroimaging data.

FSL

The FMRIB Software Library (FSL) [16] is a library of tools for statistical analyses and image processing for functional, structural and diffusion MRI data. The research presented in this thesis chiefly uses the FEAT software tool to perform model-based fMRI data analysis. FEAT analysis can be set-up using the GUI as well as a processing script. The tool automates many of the analysis decisions by setting defaults, though it is important to note that the defaults may change across versions and hence result in different results. For the results presented in this thesis, we used FSL 5.0.9 for analysis and FSLeyes from FSL 5.0.11 for visualization and images.

The FEAT software tool in FSL offers both fixed effects (FE) and mixed effects (ME) higher-level modeling. Two different kinds of ME modeling are available:

1. Ordinary Least Squares (OLS) ignores all subject-level variance and applies a simple higher-level model over the subjects in the group. This is a fast estimation technique but it is also the least accurate.

2. FMRIB's Local Analysis of Mixed Effects (FLAME) is a two-stage estimation process that uses Bayesian modeling and estimation. This technique allows for separate modeling of variances in different subject groups and forces the random error variance to be positive. Most of the times it is sufficient to use just the first stage (FLAME 1) to obtain significantly more accurate results compared to OLS, at a comparable speed. The second stage takes the voxels that are shown in the first stage to be close to the threshold and performs a full MCMC-based analysis at those points. However FLAME 1+2 improves the accurately only slightly over FLAME 1 and takes much longer. FLAME 1+2 provides a significant value when running a higher-level analysis on a small group of subjects (say, $n < 10$).

Python

We mainly use the Python libraries NumPy, SciPy and NiBabel for modeling and statistical tests and Matplotlib for visualization. Since these packages are open-source softwares written in Python that do not depend on the original code base, they are easily shareable and reproducible.

NumPy

The NumPy library supports large multi-dimensional arrays and matrices, and enables high-level mathematical operations on the arrays. The core functionality of NumPy is its n-dimensional array data structure ("ndarray"). Besides its numeric capabilities, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined, which allows NumPy to seamlessly integrate with a wide variety of databases. Additionally, some linear algebra, Fourier transform and random number capabilities are also available.

We use NumPy to load data and perform vector and matrix operations on the data.

SciPy

SciPy library is used for scientific and technical computing and contains modules for linear algebra, integration, ODE solvers, special functions and signal and image processing, to name a few. The basic data structure used by SciPy is a multi-dimensional array provided by NumPy.

We mainly use the *scipy.stats* module which contains a large number of probability distributions as well as statistical functions.

NiBabel

NiBabel provides read and write access to some common medical and neuroimaging file formats, such as NIFTI. The various image format classes give full or selective access to header (meta) information and access to the image data is made available via NumPy arrays. A nibabel image is the association of three things: (i) a 3D or 4D array of image data, (ii) an affine array that conveys the position of the image array data in a reference space and (iii) image metadata describing the image, usually in the form of an image header.

Matplotlib

Matplotlib is part of the NumPy stack of libraries and provides an object-oriented API for embedding plots into applications. It is a Python 2D plotting library which produces plots of publication quality in a variety of hardcopy formats and interactive environments, across platforms.

3. Statistical Methods in Neuroimaging

Introduction

During the last thirty years a significant body of work has been developed to provide a framework to investigate functional neuroimaging data and enable scientifically sound inference. Statistical model building and model fitting are essential to characterize the nature of the signals present in the data, while inferential methods are necessary to test hypotheses and determine the confidence of the estimates obtained from the models that determine the predictability of the observed effects. Accurate statistical inference needs a well fitting statistical model.

In this chapter, we first provide a brief overview of the major classes of approaches to modeling neuroimaging (specifically, fMRI) data and the types of inferential approaches[17]. In the later sections, we elaborate on the specific models and inferential procedures that directly pertain to the research and analysis presented in this thesis – namely, the mass univariate approach and the Bayesian approach to modeling neuroimaging data, and the classical approach to inference.

Models

In the construction of statistical models, explicit and implicit assumptions are made about the data, and it is important that the assumptions and approximations are sufficiently fulfilled by empirical data. It is also crucial, that in the case where the assumptions are not fully met, the models still perform robustly.

Several approaches have been proposed to tackle the complex nature of neuroimaging data, and we briefly introduce the broad types of popular approaches.

A. Mass-univariate models

The neuroimaging data from scans is typically composed of images that are subdivided into three-dimensional cubic units called voxels, with a unit measure of 2mm – 4mm depending on the resolution of the scanner. The most common neuroimaging analysis strategy is to fit identical univariate models at each voxel – the mass-univariate approach. One motivation for this approach is that there are usually far more voxels ($\sim$ 200,000 for a 3T scanner) than observations (number of scans, typically in the 10s or 100s) which prevents the use of a standard multivariate structure since the covariate matrix is singular. Another reason is that treating each voxel individually retains the localizing power and allows for statistical inference at specific voxels.

Univariate modeling estimates the relationship between the known effects (such as the experimental conditions, subject and performance) and the observed data. The estimated effects are used to account for the systematic variability in the data and the residual variability is used to estimate the variance parameters of the model. Hence determining a univariate model involves a trade-off between parsimony and including all conceivable effects. Including too many explanatory variables can over-fit the data and degrade the generalizability of the results. In

an over-parametrized model each additional effect modeled consumes degrees of freedom (dof) while not explaining much additional variance in the model, thus reducing the power. This is particularly so in the case where there are few dof, such as in random effects models (as explained in more detail in the next section), where additional effects will, by reducing the dof, increase uncertainty in the estimation of model variance and decrease the significance of a given effect (unless the additional effect significantly reduces the residual variability).

On the other hand, not accounting for a relevant effect will inflate the residual variability and bias the estimate of model variance. The unmodeled effects will introduce dependencies in the residuals, which violates the assumptions of standard inference procedures, namely, that the residuals are independent and identically distributed. However in a reduced model, there are more dof available to estimate the model, decreasing the uncertainty in the estimation of model variance and increasing the significance of the effects – this beneficial effect is negligible when the model has high dof and a few additional dof are unlikely to make much impact.

Another source of random variation is introduced in the case of univariate models that are modeled with mixed or random effects – the between-subject variation as a consequence of considering the subjects to be drawn randomly from a population (for the purposes of population level inference). A fixed-effects model only considers the fixed and unknown effects influenced by within-subject variability and as such is concerned only with the residual error variation. However, the inference from a fixed effects model only pertains to the specific sample and cannot be extended to the population of interest.

The general linear model is a framework that encompasses all basic univariate fixed effects models, and methods have been developed to incorporate hierarchical mixed or random effects to enable population level inference. This will be considered in more detail in section 3.2.

B. Multivariate models

While the mass univariate approach provides reliable identification of task-dependent signal changes at the voxel-level (that is, has good localizing power), it explicitly precludes the possibility that responses arise from the co-ordinated dynamics between different areas of the brain. Conversely, multivariate analyses take advantage of the spatial and temporal dependencies among image elements, thus enabling inference across space and time.

Multivariate analyses can be broadly divided into two kinds - exploratory and confirmatory. Exploratory techniques are primarily used to identify robust patterns of co-varying neural activity (such as principal component analysis (PCA) and independent component analysis (ICA)) and to possibly relate these patterns to design variables (Canonical variate analysis (CVA), canonical correlation analysis(CCA), partial least squares (PLS)). These methods do no need the specification of an explicit hypothesis about the contribution of different brain regions (or about the differentiations of conditions or groups in case of experimental studies) and are usually data driven. On the other hand, confirmatory analyses are used to test specific hypotheses and an explicit model of regional interactions is formulated and its performance tested on the data (such as structural equation modeling (SEM), dynamic causal modeling(DCM)). For the purposes of this thesis, we focus on the exploratory multivariate analyses techniques of PCA and ICA, and provide a brief overview below.

- **Principal Component Analysis (PCA)** factorizes a data matrix with many variables into a new set of variables called principal components, each of which is a linear combination of the original variables such that they are all mutually uncorrelated, and the first principal component captures the greatest variance in the data and the last one captures

the least. Thus in neuroimaging PCA can be used to summarize a typical dataset of hundreds of thousands of voxels into a much smaller number of principal components (which is limited by the rank of the matrix and hence often the same as the number of scans per subject or the number of subjects in a group).

PCA is performed by first using singular value decomposition (SVD) to deconstruct a data matrix $\mathbf{X}_{n\times p}$ (n observations and p variables – voxels in voxel-based approaches) into its basic structure

$$\mathbf{X} = \mathbf{UBF^T}$$

where $\mathbf{U_{n\times n}}$ and $\mathbf{F_{n\times p}}$ are the orthonormal matrices composed of the left and right singular vectors, respectively, and $\mathbf{B_{n\times n}}$ is a diagonal matrix of singular values. At the subject-level in a voxel-based approach, the number of observations n refers to the number of scans obtained for the subject (which is typically about a hundred) and the number of variables p is the number of voxels (typically in the hundreds of thousands). Therefore in this scenario, the rank of $\mathbf{X}$ is given by the lesser of n and p, that is, the number of scans. The vector $\mathbf{F}(i)$ contains the weights that indicate the degree to which each voxel contributes to the i-th principal component – that is, it is a spatial image of that principal component. The vector $\mathbf{U}(i)$ contains weights that indicate the dgree to which the i-th principal component is expressed in each scan n – that is, the eigentimeseries. The i-th squared singular value from $\mathbf{B}$ is proportional to the portion of variance accounted for by the i-th principal component.

Since PCA concentrates as much variance as possible in as few components as possible, it can be used to reduce dimensionality further, such as by selecting the first k components that account for a desired percentage of overall variance. Also, PCA aggregates variance in components by appropriately weighing original variables that tend to co-vary and as a result, it isolates prominent patterns of regional covariation – that represent functional interactions. These functional networks can then be tested for task or group significance using methods like the CVA. We employ PCA in our analysis to generate orthogonal basis functions based on a biologically relevant brain atlas, as we explain in section **4.1**.

- **Independent Component Analysis (ICA)** also seeks to reduce the dimensionality of the data like PCA by choosing components with maximum inter-dependencies, but unlike PCA that produces components that are mutually uncorrelated spatially and temporally, ICA produces components that are maximally statistically independent only in one domain (spatial or temporal, not both). ICA assumes that independent components were mixed to give rise to the observed signal – in a neuroimaging context, ICA assumes that the neural activity measured in different voxels is a linear cobination of a small number of underlying independent sources. Like the PCA, ICA also creates a new set of axes to represent the data in, but instead of the axes being orthogonal they are oriented such that the projection of data points onto the axes is maximally non-Gaussian. The model for ICA can be represented as

$$\mathbf{X^T} = \mathbf{AS}$$

where $\mathbf{X}_{n\times p}$ is the observed data matrix with n scans (observations) and p voxels (variables), $\mathbf{S}_{r\times n}$ is the source matrix with r sources and their activity in each of the n scans, and $\mathbf{A}_{p\times r}$ is the mixing matrix that represents how the sources were combined to produce the observed signals in $\mathbf{X}$.

Both spatial ICA (independence in spatial domain) and temporal ICA (independence in temporal domain) are employed in neuroimaging , based on the requirements of the data

and the problem at hand. For instance, spatial ICA is more commonly used for fMRI data since task-related activations are assumed to be relatively sparse in a volume consisting of thousands of voxels and independent components can isolate networks of regions that overlap minimally. Spatial ICA assumes that the rows of $\mathbf{S}$ are independnet, while temporal ICA assumes that the columns of $\mathbf{S}$ are independent. ICA tries to simultaneously estimate both $\mathbf{A}$ and $\mathbf{S}$ through an iterative process, by maximizing the non-Gaussianity of either the rows or the columns of $\mathbf{S}$. The ICA finally projects the original data into source space:

$$\mathbf{A^{-1}X = S}.$$

$\mathbf{A^{-1}}$ is called the unmixing matrix whose rows are spatial maps that represent the participation of each brain region.

ICA cannot be used to estimate the number of underlying sources in the data and in fact requires that the number of sources be specified before the analysis. Additionally, since ICA is an iterative algorithm that is a computationally expensive, in practice the typical approach is to first perform PCA and choose the first k components that explain the desired percentage of variability and then perform ICA on these principal components to obtain maximally independent components. To generate the Instantaneous Connectivity Parcellation atlas that we employ in our analysis, van Oort, E.S.B. et al. (2016) employ a spatial ICA algorithm (Section **3.3**).

Given the nature of functional neuroimaging data multivariate approaches would seem to be a natural fit for data analysis. Several voxel based multivariate approaches have been adapted to suit the needs of functional neuroimaging data. These approaches attempt to characterize the overall distribution pattern of the changes in brain activity without regard to the location, and therefore lack localizing power.Additionally, the high dimensionality of functional neuroimaging data relative to the number of observations (scans) rules out the straighforward application of standard multivariate statistics, as the estimated covariance structure of the data would be singular. The presence of spatial correlations in the neuroimaging data and regional changes in error variance lead to non-sphericity in the error terms, which would require a large number of parameters to be estimated for each voxel.

C. Bayesian models

Bayesian methods can be used to enforce soft constraints on the parameter estimates by incorporating prior knowledge through the specification of priors. In a frequentist approach, the observed data are considered as a realization of a random process and the parameters of the model are fixed but unknown quantities. The Bayesian approach on the other hand regards the parameters as random variables as well. Before an experiment is performed, the parameters of interest are considered to have a prior distribution that is based on existing knowledge in the field. After the experiment is performed, the information gathered through the experiment is used to 'update' the prior knowledge, to give a posterior distribution which is used to address queries about the parameters of interest. As the number of observations from the experiment (the information) increases, the importance of the prior decreases and the Bayesian results converge to the frequentist result.

As we shall see in Section 3.2, fMRI data is often described and analyzed using a hierarchical approach, which makes the Bayesian framework a natural fit for the analysis of fMRI data. Spatio-temporal models are another class of fMRI models that also lend themselves well to the Bayesian approach. As fMRI methods have become more widespread, a wealth of information

has been accumulated about brain activation, which can be used to build prior distributions for a Bayesian analysis. The main hindrance to the adoption of Bayesian methods in fMRI has been the cost of computation, given the large and high-dimensional datasets and the complex relationships within and among voxel time-series. Early attempts at Bayesian models relied on simplified assumptions and computational shortcuts to make the analyses feasible, as we shall see in Section 3.3. However, with rapid advances in computational capacity, the field of fMRI is fast adopting Bayesian models.

D. Functional networks and connectivity analysis

The mass-univariate, multivariate and Bayesian models study changes in brain activity and how these changes covary with the experminetal paradigm. Research has suggested that higher cognitive functions of the brain are the result of the network interactions between different brain regions, thus necessitating the study of interactions between brain regions. Friston (1994) defined functional connectivity as the observed correlations over time between different brain regions that are independent of the sources of these correlations. Considering that brain regions that are part of a functional network have correlated activities, the interactions between the regions are studied based on the covariance pattern observed in functional neuroimaging data. However, this poses several challenges. For instance, the actual sources of inter-regional covariances are unknown. Several sources for the same have been proposed, but these might be confounded with correlations due to effective connectivity (the influence that one neural system exerts over another, such as fatigue, attention drift etc.). The presence of spurious correlations biases the results unless they are removed or properly accounted for. One approach to study inter-regional correlations is to examine the partial correlation coefficients between pairs of pre-selected regions of interest. Another approach is based on selecting a reference region or voxel and studying the correlations of this region or voxel with the rest of the brain (or select regions of interest). These solutions involve problems with multiple comparisons, and some methods have been proposed to tackle this issue (Petersson et al. 1999b).

Inference

Inference in neuroimaging has largely been restricted to classical inference based on statistical parametric maps. Statistical parametric maps are constructed using general linear model and a thresholding procedure such as the random field theory to test hypotheses about regionally specific effects. There are multiple ways to summarize a statistic image, and here we define voxel-level, cluster-level and set-level approaches.

In a voxel-level approach, the test statistic at each and every voxel is individually tested for evidence against the null hypothesis. If the statistic at a particular voxel exceeds a pre-determined threshold u, then the voxel is classified as 'active' or 'significant'. Thus voxel-wise inference allows for a highly spatially specific inference, provided the threshold u is chosen properly. However, such a voxel-by-voxel inference does not take into account the spatially extended nature of fMRI signals – often the activated brain regions are much larger than the dimensions of a voxel, and additionally, during pre-processing fMRI data is often spatially smoothed and oversampled during spatial normalization.

Cluster-wise inference allows to take advantage of this spatial nature of fMRI data by making inference about clusters of activated voxels rather than individual voxels. The most common approach to cluster-level inference involves two stages. First, clusters of voxels are defined by applying a 'cluster-defining threshold' u_c to the statistic image and identifying groups of con-

tiguous [†] voxels above u_c. In the second stage, the clusters thus identified are deemed significant if their size (in voxels) is above a certain threshold k. Cluster-level inference has greater power than voxel-level inference, given its ability to better detect a signal that is larger in scale than the smoothness of the nose in the data. However, this greater power comes at the lack of spatial specificity – it is not possible to pin-point which voxels in a cluster are activated; it is only possible to conclude that at least one voxel in a significant cluster has evidence against the null hypothesis. While this is not a problem when the cluster size is small, when the clusters are very large it may lead to no useful inference. In such instances the cluster forming threshold u_c (which can be arbitrarily chosen, in principle) can be increased to obtain smaller clusters, but this in turn leads to a multiple testing problem (which is discussed in section 3.2). Also, assigning significance to clusters based on their size ignores the statistic values within the cluster.

Set-level inference provides an overall test to see if there are any significant signals anywhere in the brain, without indicating the location of the signals with any degree of specificity. For arbitrarily chosen cluster-defining threshold u_c and cluster size threshold k, the test statistic for set-level inference is the number of significant clusters. A significant set-level p-value indicates that there are an unusually large number of significant clusters, but provides no information regarding the location of the significant clusters.

The thresholds k (for cluster-level inference) and u (for voxel-level inference) are chosen using thresholding procedures described in section 3.2. In this thesis we mainly use voxel-level inference as a basis to explain the concepts of inference, unless otherwise specified.

Classical approach: The General Linear Model

Mass univariate models

All classical analyses of functional neuroimaging data are based on the general linear model. Analysis consists of model specification, parameter estimation and inference. We first look at the model specification in this section.

Let Y_i denote a response variable measured (a random variable), for instance, during an experiment. Let $x_{ij}(j = 1, \dots p)$ denote the explanatory variables which may be continuous covariates, functions of covariates or dummy variables. A general linear model explains the relationship between the response variable Y_i and the explanatory variables as

$$Y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + \epsilon_i \tag{3.1}$$

where β_j $(j = 1, \dots, p)$ are the unknown parameters that are to be estimated and ϵ_i are the errors that are independent and identically distributed (iid) normal random variables with zero mean and variance σ^2. The equivalent matrix notation is

$$Y = X\beta + \epsilon$$

where Y is the column vector of N observations, ϵ the column vector of error terms corresponding to each of the N observations, and β the column vector of parameters $\beta = [\beta_1, \dots, \beta_p]^T$. The $N \times p$ matrix X is the design matrix which provides near complete explanation of the model; the remaining assumptions are about the distribution of errors.

[†]What is considered 'contiguous' depends on how a neighborhood is defined. In 3D two voxels above the threshold u_c can be considered connected based on a 6-connectivity (when only connection of faces is considered) or 18-connectivity (when edges are also considered) or 26-connectivity (when corners are also considered)

Parameter estimation: Ordinary Least Squares (OLS)

The simultaneous equations implied by the general linear model (considering $\epsilon = 0$) cannot be solved since the number of parameters p is typically less than the number of observations N. This calls for a method to estimate the parameters to obtain a best fit of the data – such as the ordinary least squares method (OLS).

Let the parameter estimates be denoted by $\hat{\beta} = [\hat{\beta}_i, \ldots, \hat{\beta}_p]^T$ that give the predicted responses $\hat{Y} = [\hat{Y}_1, \ldots, \hat{Y}_N]^T = X\hat{\beta}$ with the residual errors $e = y - \hat{Y} = y - X\hat{\beta} = [e_1, \ldots, e_N]^T$, where y is the value of Y observed during the experiment. Then the residual sum of squares obtained as $S = \sum_{i=1}^{n} e_i^2 = e^T e$ is the sum of the square differences between the actual and fitted values, and measures the fit of the model provided by these parameter estimates. The least squares estimates are the parameter estimates that minimize the residual sum of squares:

$$\begin{aligned} S = e^T e &= (y - X\hat{\beta})^T (y - X\hat{\beta}) \\ &= y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta} \end{aligned}$$

which is minimised * when

$$\begin{aligned} & \frac{\partial S}{\partial \hat{\beta}} = 0 \\ \implies & -2X^T y + 2X^T X\hat{\beta} = 0 \\ \implies & X^T X\hat{\beta} = X^T y \end{aligned}$$

If the inverse of $X^T X$ exists, then by premultiplying the above equation on both sides with $(X^T X)^{-1}$ we obtain the estimate

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

which is the best linear unbiased estimator (BLUE) for the general linear model .

If X has linearly dependent columns (the model is overparametrized or overdetermined) then it is rank deficient and $(X^T X)$ is singular and therefore has no inverse. In this case a set of least squares estimates can be found by imposing constraints on the estimates, or by obtaining a pseudo-inverse of $(X^T X)$, such as the Moore-Penrose pseudo-inverse denoted by $(X^T X)^-$. Then the set of least squares estimates is given by

$$\hat{\beta} = (X^T X)^- X^T y$$

Using the pseudo-inverse for parameter estimation in overdetermined models does not allow testing for the linear combinations of effects for which there exist an infinite number of solutions. However, it does allow for the estimation of unique mixtures without changing how X is specified.

Non-sphericity

In neuroimaging the presence of spatial and temporal non-sphericity can have significant impact. Sphericity is the assumption that the error terms are independent and identically distributed. If the error terms are not identically distributed then the observations are heteroschedastic, and correlation among error terms indicate dependence. The presence of spatial correlations in the neuroimaging data and regional changes in error variance lead to non-sphericity in the error terms, which would require a large number of parameters to be estimated for each voxel.

*This is indeed a minimum: on taking the second derivative of S with respect to $\hat{\beta}$ we obtain $2X^T X$ which is a positive definite matrix as long as X is full rank.

Parameter estimation: Generalized Least Squares (GLS)

In the context of non-sphericity, the Gauss-Markov assumptions are violated (the errors are no longer uncorrelated) and therefore the OLS estimates are no longer optimal. This calls for the transformation of the general linear model to be able to obtain parameter estimates.

Let Y represent the smoothed time series and ϵ the error term such that $\epsilon \sim \mathrm{N}(0, V),\ V \neq \sigma^2 I$.Since V is a positive semi-definite matrix, it can be decomposed as $V = K^T K$. For the general linear model

$$Y = X\beta + \epsilon$$

the optimum estimator of β is obtained by unsmoothing the data by multiplying by K^{-1} and applying least squares to the uncorrelated data Then

$$K^{-1}Y = K^{-1}X\beta + K^{-1}\epsilon,\ \ Var(\epsilon) = K^{-1}VK$$

or $\tilde{Y} = \tilde{X}\beta + \tilde{\epsilon}$ where $\tilde{\epsilon}$ is now uncorrelated and the original β is still the linear parameter, so the OLS can now be applied to obtain

$$\hat{\beta}_{GLS} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{Y} = (X^T V^{-1}X)^{-1}X^T V^{-1}Y \tag{3.2}$$

and it can be shown that $E[\hat{\beta}_{GLS}] = \beta$ and $Var[\hat{\beta}_{GLS}] = (X^T V^{-1}X)^{-1}$. The variance V is estimated using a restricted maximum likelihood method (ReML) and the expectation maximization (EM) algorithm.

Inference

If X is full rank then the parameter estimates $\hat{\beta} = (X^T X)^{-1}X^T y$ are normally distributed as $\hat{\beta} \sim \mathrm{N}(\beta, \sigma^2(X^T X)^{-1})$. For a column vector c containing p weights,

$$c^T\hat{\beta} \sim \mathrm{N}(c^T\beta, \sigma^2 c^T(X^T X)^{-1}c)$$

A hypothesis $\mathrm{H}_0 : c^T\beta = d$ can be assessed by calculating

$$t = \frac{c^T\beta - d}{\sqrt{\sigma^2 c^T(X^T X)^{-1}c}} \tag{3.3}$$

and obtaining a p-value by comparing t with a t-distribution having $N-p$ degrees of freedom[18].

General Linear Models for fMRI data

Since fMRI data represent time-series, there is serial correlation involved and temporal non-sphericity needs to be modeled. Additionally, the data are a result of dynamical processes and require a convolution model for their modeling. Parametric softwares such as FSL employ voxel based methods to fit models to neuroimaging data. Each voxel carries a time varying signal which is measured and compared to the predicted response from a particular stimulus to detect activation at that voxel. The predicted response is obtained from a convolution of the applied stimulus signal with the heamodynamic response function (HRF) as discussed in the previous chapter.

Single subject analysis

Regressors are constructed from each type of stimulus provided during an experiment and these are then fit to the observed signal at each voxel using a linear model. Thus a set of regression

parameters (effect sizes) and a residual function is obtained at each voxel (Figure 3.1). Therefore at each voxel we have,

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e} \tag{3.4}$$

where $\mathbf{y}$ contains the signal recorded at the voxel, $\mathbf{X}$ contains the regressors/predicted responses from each stimulus supplied during the experiment, β is the vector of parameters to be estimated (that is specific to each voxel) and $\mathbf{e}$ is the residual at each voxel.

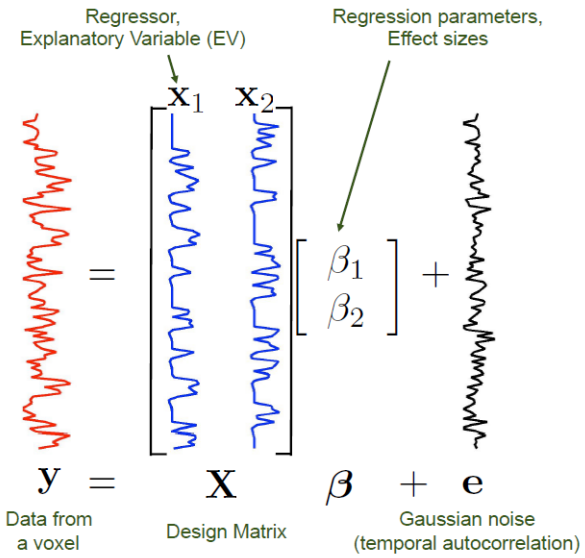

Figure 3.1: The linear modeling framework at each voxel. Source: FSL Course (http://fsl.fmrib.ox.ac.uk/fslcourse/)[13]

The actual parameter estimate at each voxel is divided by the error in the estimation of that parameter to obtain a t-value (as in Eq. 3.3, with $d = 0$), which can be converted to a t- or z-statistic. Images made up of t-values (or z- values) show how strongly each voxel is related to each explanatory variable (EV). Additionally, the parameter estimates can be compared to see if one EV is more relevant than others, by developing contrast maps, by choosing the appropriate vector for c in Eq. 3.3. These contrast maps can be used to query the data and test for significant effects. In order to control false positives across the whole brain, a threshold is determined for voxel signal intensities such that only 5% (typically) or less of these contrast maps show a voxel above the determined threshold (as detailed in Section 2.2.4).

Multi-subject analysis

Research studies using fMRI typically aim to answer questions about the activation effects in a population of subjects, and therefore involve data from multiple subjects analyzed to allow for inference at the group level. For a group-level analysis it is simple to formulate a complete single-level general linear model relating the parameters of interest at the group level to the full set of time-series data (for example, Cnaan et. al, 1997 [19]). However, since the computational costs involved in fMRI are relatively high it becomes cumbersome to re-analyze individual subject level data for group level inference. In this case it is desirable to be able to make group-level inferences using only the results of the subject-level analysis. Beckmann et. al (2003)[20] demonstrated that the single-level GLM can be decomposed into an equivalent two-level model so that group analyses can be performed using only the parameter estimates and their covariances from the lower-level. The group level inference can be obtained through a fixed effects approach, or

through a mixed effects model.

For a group of size n, for a subject k the fixed effects model is represented as

$$Y_k = X_k\beta_k + \epsilon_k \tag{3.5}$$

$$\beta_k = X_g\beta_g, \quad \beta_g = \frac{1}{n}\sum_{k=1}^{n}\beta_k, \quad X_g = [1]_{n\times 1} \tag{3.6}$$

Mixed effect modeling, by accounting for the between-subject variance in the group (ϵ_g), allows for inference that is more representative of the population.

$$Y_k = X_k\beta_k + \epsilon_k \tag{3.7}$$

$$\beta_k = X_g\beta_g + \epsilon_g, \quad \beta_g = \frac{1}{6}\sum_{k=1}^{n}\beta_k, \quad X_g = [1]_{n\times 1} \tag{3.8}$$

The inferences from a fixed effects analysis are limited to the cohort of subjects studied, while those from random effects analyses are used to make inferences about the population sampled. The only variance modeled in a fixed effects model (of repeated observations on a group of subjects) is the within-subject variability within the specific experimental condition, that is, the variability from scan to scan of the same condition within an individual. This represents measurement variability confounded with physical, physiological and cognitive effects. The random effects analyses also account for the between-subject variability, which is particularly important for group-level comparisons especially for fMRI where the within-subject variability is most certainly smaller than the between-subject variability – as a result of which the significance of between-group differences will be overestimated in a fixed-effects analysis.

Multi-level general linear model

Let N be the number of subjects and for each subject i let X_i denote the design matrix and β_i denote the corresponding subject-level parameter estimates. A single-level general linear model relates the first-level parameters to the to the individual datasets as

$$Y_i = X_i\beta_i + \epsilon_i$$

where ϵ_i are the single-subject residuals such that $E[\epsilon_i] = 0, Cov(\epsilon_i) = V_i$.

A two-level model for this situation can be specified as

$$\begin{aligned} Y &= X\beta + \epsilon \\ \hat{\beta} &= X_G\beta_G + \eta \end{aligned} \tag{3.9}$$

where X_G is the group-level design matrix,β_G is the vector of group-level parameters and η gives the residuals of the group activation. $E[\eta] = 0, Cov(\eta) = V_G$ and $Cov(\epsilon) = V$ denote the block-diagonal form of the first level covariance matrices V_i.
The BLUE can be calculated using the generalized least squares (GLS) approach (Eq. 3.2), giving at the first level [20]:

$$\begin{aligned} \hat{\beta} &= (X^TV^{-1}X)^{-1}X^TV^{-1}Y \\ Cov(\hat{\beta}) &= (X^TV^{-1}X)^{-1} \end{aligned} \tag{3.10}$$

and at the second level:

$$\begin{aligned} \hat{\beta_G} &= (X_G^T V_G^{-1} X_G)^{-1} X_G^T V_G^{-1} \hat{\beta} \\ Cov(\hat{\beta_G}) &= (X_G^T V_G^{-1} X_G)^{-1} \end{aligned} \tag{3.11}$$

However in practice, only the estimates of the first-level parameters, $\hat{\beta}$, are used as inputs for the second-level. So the second level in Eq. 3.9 should instead be given as

$$\hat{\beta} = X_G \beta_G + \eta'$$

where $Cov(\eta') = V_{G2}$. Therefore the BLUE at the second level is now

$$\begin{aligned} \hat{\beta_G} &= (X_G^T V_{G2}^{-1} X_G)^{-1} X_G^T V_{G2}^{-1} \hat{\beta} \\ Cov(\hat{\beta_G}) &= (X_G^T V_{G2}^{-1} X_G)^{-1} \end{aligned} \tag{3.12}$$

Now substituting the value of the first-level parameter estimate $\hat{\beta}$ gives

$$\hat{\beta_G} = (X_G^T V_{G2}^{-1} X_G)^{-1} X_G^T V_{G2}^{-1} (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

The two-level model can be collapsed into a single-level model by substitution, giving

$$\begin{aligned} Y &= X X_G \beta_G + \gamma \ \text{ where,} \\ \gamma &= X\eta + \epsilon \\ E(\gamma) = 0,\ Cov(\gamma) &= W = X V_G X^T + V \end{aligned} \tag{3.13}$$

The BLUE for this general linear model is obtained as

$$\begin{aligned} \hat{\beta_G} &= (X_G^T X^T W^{-1} X X_G)^{-1} X_G^T X^T W^{-1} Y \\ Cov(\hat{\beta_G}) &= (X_G^T X^T W^{-1} X X_G)^{-1} \end{aligned} \tag{3.14}$$

which relates the group level parameter estimate to the full data vector Y for N subjects, thus involving the solution for matrices of greatly increased size.

In order to employ the computationally more efficient approach of using only the parameter estimates from the first-level for the second-level inference, the estimated group parameter in Eq. 3.12 should be equal to that in Eq. 3.13. Using the Sherman–Morrison–Woodbury formula to write

$$W^{-1} = V^{-1} - V^{-1} X (V_G^{-1} + X^T V^{-1} X)^{-1} X^T V^{-1}$$

the equivalence relation can be obtained as

$$V_{G2} = V_G + (X^T V^{-1} X)^{-1}$$

which expresses the second-level covariance as the sum of the group covariance from the single-level model (Eq. 3.13) and the parameter covariance from the first-level of the two-level model (Eq. 3.12). The group level parameter estimate can now be re-written as

$$\hat{\beta_G} = (X_G^T (V_G + (X^T V^{-1} X)^{-1})^{-1} X_G)^{-1} X_G^T (V_G + (X^T V^{-1} X)^{-1})^{-1} \hat{\beta}$$

that is, as a function of the first level parameter estimates $\hat{\beta}$ and their covariances $(X^T V^{-1} X)^{-1}$.

Inference

For inference at the group level, hypothesis tests such as one-sample t-test, two-sample t-test, paired t-test and one-way ANCOVA are used. For our research we use one-sample and two-sample t-tests, which we elaborate on here [18].

One-sample t-test

The one-sample t-test can be used to test the null hypothesis that the mean of n subjects is zero. The model for this can be specified as

$$Y = x\beta_1 + \epsilon$$

where x is a vector of ones and $\epsilon \sim \mathrm{N}(0, \sigma^2 I)$, the null hypothesis is H: $\beta_1 = 0$. The t-value is computed as:

$$T = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2/n}} \sim t_{n-1}$$

where $\hat{\sigma}^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2/(n-1),\ \hat{Y}_i = (x_i\hat{\beta}_1)_i = \hat{\beta}_1$.

Two-sample t-test

Two-sample t-tests are used to test the null hypothesis that the means of two groups are equal. The design matrix consists of three columns: the first two columns encode the group membership of each subject and the third models a common constant across subjects of both groups. This model is overdetermined by one degree of freedom, i.e. the sum of the first two regressors equals the third regressor.

If the groups considered have n_1 and n_2 subjects, then the first regressor consists of n_1 ones followed by n_2 zeros, the second regressor consists of n_1 zeroes followed by n_2 ones and the third regressor consists of n_1+n_2 ones. In order to test the null hypothesis H: $\beta_1 - \beta_2 = 0$ the contrast vector required is $c = [1, -1, 0]^T$. The matrix

$$X^T X = \begin{pmatrix} n_1 & 0 & n_1 \\ 0 & n_2 & n_2 \\ n_1 & n_2 & n_1 + n_2 \end{pmatrix}$$

is rank deficient (since it is overdetermined) and therefore the pseudo-inverse $(X^T X)^-$ is used to compute the t-statistic. Applying the contrast c for the desired hypothesis,

$$c^T (X^T X)^- c = 1/n_1 + 1/n_2$$

giving the t-statistics as

$$T = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\hat{\sigma}^2/(1/n_1 + 1/n_2)}} \sim t_{(n_1+n_2-2)}$$

where $\hat{\sigma}^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2/(n_1+n_2-2)$, assuming equal variance in both groups. This assumption would be violaed in the case when one group consists of healthy subjects and the other is composed of patients. Typically in research studies, $n_1 = n_2$.

Thresholding

In this section the case of voxel-wise inference - which needs a height threshold - is used to elaborate on the process of choosing a threshold for statistical parametric maps. In general there are also cluster-wise and set-level inference which require height threshold as well as spatial extent

thresholds to be chosen.

Under the null hypothesis, the voxel values in a statistical parametric map (SPM)[18] are distributed according to a known probability function, typically the t- or F-distributions. Every voxel is analyzed using a univariate statistical test for activation on a particular explanatory variable and the resulting statistical parameters are assembled to obtain the SPM. Unexpected deviations in the SPM are interpreted as regional effects as a result of the experiment the subject has undergone. In order to decide if the observed statistic at a voxel is the result of a signal or reflects underlying noise, the statistic is compared to its assumed value under a null distribution (H_0) – which is typically that there is no activation in the brain region of interest as a result of the external stimulus. An acceptable Type I error level is chosen (usually $\alpha = 0.05$) and for a one-sided test the null hypothesis is rejected if $P(T > t_0|H_0) < \alpha$, where t_0 is the value of the test statistic corresponding to α.

Therefore an important step in the inference is to choose an appropriate t-value to threshold the statistical map. A high threshold value ensures good specificity, but risks falsely concluding there is no signal in the data. Likewise, a low threshold enables detection of more of the signal but risks falsely concluding there is an effect when there is none (Fig. 3.2). In many cases it is not known where in the brain the activation will arise, which means that the entire volume of the brain has to be taken into account and the familywise error controlled across the whole brain. Since the whole brain volume comprises typically of 100,000 voxels or higher (depending on the resolution of the scanner) there is a need to control for multiple comparisons across voxels.

Standard hypothesis tests are designed to control the error rate only per test, and not meant to be used repeatedly for multiple related tests. Therefore for fMRI data, the false positive risk needs to be measured over the entire image, and this is most commonly accomplished using the Familywise error rate.

a. Familywise error rate

The Familywise error rate (FWE) is the chance of one or more false positives anywhere in the image. For a valid procedure with α_{FWE}, there is at most a $\alpha_{FWE}\%$ chance of any false positives across the statistical map. In voxel-wise inference, for each voxel a 'corrected p-value' is calculated, which is the smallest α_{FWE} that allows detection of that voxel. Several procedures are available to obtain valid corrected p-values, three of which we briefly present here.

(i) Bonferroni correction

The 'corrected p-value' at each voxel is obtained as follows:

$$
\begin{aligned}
&P(\text{Type 1 error at voxel}) = \alpha_v \\
&\implies P(\text{no Type 1 error across n independent voxels}) = (1-\alpha_v)^n \\
&\implies P(\text{atleast one Type 1 error across n independent voxels}) = 1-(1-\alpha_v)^n \\
&\implies FWER = 1-(1-\alpha_v)^n \approx n\alpha_v
\end{aligned}
$$

using the Binomial expansion for large n and small α_v, where α_v is the threshold at each voxel and FWER is the familywise error rate to be controlled for over the entire volume of voxels. In order to control for a particular FWER across the brain, the t-statistic at each voxel should be chosen corresponding to an $\alpha_v = FWER/n$. Thus if the brain region is composed of n=1000 voxels and the required FWER over the brain region is chosen as 0.05, then the threshold t-statistic at each voxel is computed from $\alpha_v = 0.05/1000 = 0.00005$. After the statistical parametric map for this region has been thresholded, if there are more than 5% of voxels with t-statistic values

greater than the threshold, then the null hypothesis is rejected and the brain region is concluded to show activation in response to the stimulus.

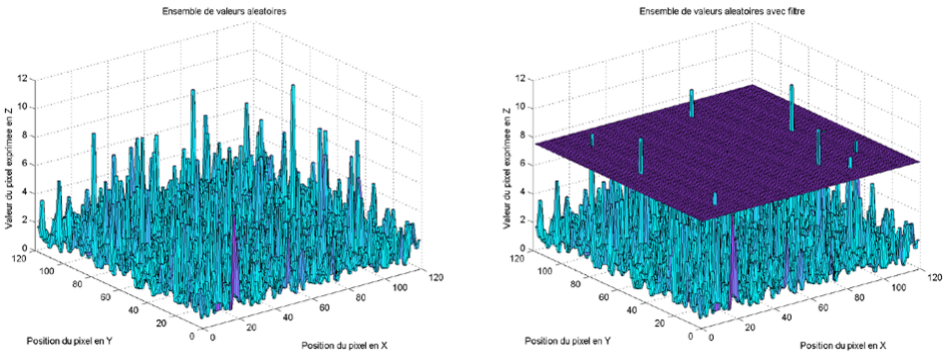

Figure 3.2: Choosing an appropriate threshold for inference. (Source: Presentation, RFT for Dummies - Part 1 (2009), Lea Firmin and Anna Jafarpour)

In reality, due to the nature of brain activations in fMRI there is spatial correlation between adjacent voxels. Pre-processing steps that are typically carried out such as spatial normalization, realignment and smoothing introduce further spatial correlation in the data. Therefore the Bonferroni correction is usually strongly conservative and there is a need for correction procedures that account for the dependency among the voxels.

(ii) Random Field Theory (RFT)

The random field theory [21] is used to obtain a more accurate threshold for the statistical maps by first estimating the smoothness of the statistical map and then using this value to obtain the expected value of the Euler Characteristic, E[EC] corresponding to the chosen tolerance for Type 1 error.

The smoothness of an SPM is estimated from the spatial correlation in the images. The calculated smoothness is used to compute the value of a Resel, which is defined as the block of pixels that are the same size as the FWHM (full width at half maximum) value of the smoothness in the SPM. Thus, for a two-dimensional image with pixels of size 2 mm and smoothness of 8 mm, a Resel is $(8/2)^{dimension} = 4 \times 4 = 16$ pixels. If the entire image is composed of 256 pixels, then the number of resels in the image is given by $256/16 = 16$ resels.

The Euler Characteristic (EC) is approximately given by the number of "blobs" above the threshold in a thresholded SPM. At high threshold, the FWE rate can be approximated by the E[EC] [21]. For a two-dimensional image, the expected value of the Euler Characteristic is computed by

$$E[EC] = R\frac{4\ln(2)}{\sqrt{((2\pi)^3)}} z_t \exp(-\frac{z_t^2}{2})$$

where R is the number of resels and z_t is the desired z-score threshold. By setting E[EC] = FWE, the required z-score threshold for the statistical parametric map is obtained.

Assumptions of Random Field Theory Any statistical inference that uses RFT to correct for multiple comparisons also needs to conform to the assumptions [21] of the Random Field Theory, namely

(i) The error fields are "reasonable lattice approximations" to an underlying random field with a multivariate Gaussian distribution.

(ii) The random fields are continuous with twice-differential autocorrelation function.

(iii) Nonparametric approaches

An alternative approach is to use the data itself to obtain empirical null-distribution of the test statistic of interest, rather than making parametric assumptions about the data to obtain approximate p-values. The bootstrap and the permutation test are the most popular non-parametric options, of which the former is an asymptotic method while the latter provides exact control of the false positive risk.

The main idea behind a permutation test is that under the null hypothesis, the group labels are irrelevant and therefore by repeatedly shuffling the assignment of the group labels and analyzing the data for each shuffle gives a distribution of statistic values that would be expected under the null hypothesis. For fMRI data the permutation test is difficult to apply for a single subject's data since the temporal autocorrelation makes the time series autocorrelated (and hence not exchangeable) and computationally demanding procedures are required to decorrelate the data. However for group level fMRI analysis, the permutation is much easier to apply.

Through repeated permutation, a distribution for the maximal statistic (for voxel-wise inference, this is the largest intensity in the statistic image) is constructed and the FWE corrected p-value is calculated as the proportion of maxima in the permutation distribution that is as or more extreme than the observed statistic value.

The main disadvantage of permutation methods is their computational intensity – where RFT methods take seconds, a typical permutation analysis takes a few minutes to an hour on modern computing hardware.

b. False Discovery Rate (FDR)

An alternative measure of error in neuroimaging that is more lenient than the FWE is the FDR. The FDR measures the proportion of false positives among all rejected tests. If TP is the number of true positives (the null hypothesis is rejected when it is not true) and FP the number of false positives or false discoveries (the null hypothesis is rejected when it is true), then the FDR is defined as

$$FDR = \frac{FP}{TP + FP}$$

The FDR is equivalent to the FWER when all the null hypotheses are true, and any procedure that controls the FWER will also control the FDR. Therefore a procedure that controls for the FDR instead of the FWER is less stringent and leads to greater power. Since by definition the FDR depends only on the p-values rather than the actual test statistics, it is applicable to any valid statistical test, as opposed to the FWER based on RFT that requires the test statistics to follow a known distribution.

The greater sensitivity of the FDR comes with greater false positive risk. Additionally, voxel-level FDR lacks spatial specificity – in a statistical map of FDR - significant voxels, it is only possible to state that on average not greater than $x\%$ (where x is the FDR level) of voxels on average are false positives, and it is not possible to pick out significant voxels individually.

Drawbacks of parametric methods

In their recent work Eklund et al. (2016) [1] examined the statistical validity of the popular fMRI software packages SPM, FSL and AFNI using resting state fMRI data as the null data. The authors found that all three statistical packages lead to conservative voxelwise inference and

invalid clusterwise inference for both one- and two-sample t-tests, with a high degree of false positives (up to 70%, compared to the nominal 5%) for clusterwise inference. They found that the choice of smoothing and cluster defining threshold affect the familywise error (FWE) rate of the parametric methods. They concluded that the principal cause for the invalid clusterwise inference is that the spatial autocorrelation in the data violates the assumptions made by parametric methods used to correct for multiple comparisons, principally Gaussian Random Field Theory.

These results call for a need to explore alternative approaches for inference on neuroimaging data that accurately model spatial autocorrelation, potentially providing better type I error control and more sensitive inference. In contrast to RFT, the trend surface modeling approach models spatial correlation using a set of biologically informed basis functions derived from resting state fMRI, then performs inference over the coefficients for these basis functions. Since this approach involves only a few hundred basis functions the penalty for multiple comparisons is greatly reduced and there is no longer a need for complex corrections and a simpler measure such as the Bonferroni correction can be used instead.

Voxel based features also not suitable for multivariate approaches such as pattern-recognition (which can model the correlations between the brain regions) since they lead to highly ill-posed problems on account of the large number ($\sim 100,000$) of features versus much fewer observations (scans or subjects; ~ 10 or 100). Therefore for mass-univariate as well as multivariate approaches it is more optimal to find parsimonious representations of brain function (or structure) that would better represent the underlying signal.

Bayesian spatial statistics approach: Trend surface model

Bayesian methods for fMRI

As mentioned earlier, most early Bayesian models for fMRI relied on simplifying assumptions and computational short-cuts to make Bayesian analysis feasible for the large and complex datasets. Friston et a. (2002)[22] and Friston and Penny (2003)[23] proposed an empirical Bayes approach, where in the parameters of the priors are estimated from the data rather than depending on hyper-priors (which is the case in Full Bayes approach). The first Fully Bayes approaches to fMRI data were proposed by Genovese(2000)[24] and Kershaw et al. (1999)[25], where they use a non-informative Jeffrey's prior for all parameters to avoid computational issues. However, these methods still assumed the time-series at each voxel to be independent and do not account for the spatial properties of fMRI. This was quickly remedied by other research works: Hartvig (2002)[26] use a marked point process to describe the spatial activation pattern. Woolrich et al. (2004)[27] propose a model for the HRF that includes a spatial noise component. Spatial knowledge was also incorporated in the form of priors using Markov random fields (Gössl et al. 2000[28], 2001[29]) and mixture models (Hartvig and Jensen, 2000[30]).

Another impediment to adoption of Bayesian methods is the computational complexity in computing analytically intractable posterior distributions using expensive procedures like the Markov Chain Monte Carlo (MCMC) simulations. A popular approach to avoid complex simulations is to obtain analytically tractable posteriors, such as by using conjugate priors. In this case a Variational Bayes approach can be used to approximate the true posterior distribution by estimating it using a posterior factorized over subsets of the model parameters and minimizing the difference between the true and approximate distributions using the conjugate prior. When the models are not conjugate, as is often the case, then further approximations are needed, such as by using the Approximate Variational Bayes approach. Another approach to simplify

Bayesian inference is a hybrid of frequentist and Bayesian approaches – for instance, Neumann and Lohmann (2003)[31] use the ordinary general linear model without priors for the subject level analysis and only incorporate a Bayesian scheme at the second level.

The Bayesian approach enables the exploration of questions beyond those related to localization of the BOLD signal, because once the posterior distributions of the parameters are available they can be used to answer a wide range of questions about the parameters. For instance, Genovese (2000)[24] explored whether the strength of the response to a stimulus increases with the difficulty of the task – which is a very tough question to answer based on frequentist methods.

Applying principles of spatial statistics

Several approaches have been proposed to account for the spatial dependence in fMRI data, based on the principles of spatial statistics. Discrete spatial models that provide local smoothing for the parameter estimates from mass-univariate analysis were proposed by Penny et al (2005)[32] and Woolrich et al (2004)[27]. However, these models cannot accommodate long range spatial dependencies that are characteristic of neuroimaging data. The spatial mixed model, that accounts for spatial dependencies through a continuous (typically Gaussian) spatial random field, is a more accurate approach. In this model the covariance matrix of the spatial random effect describes the spacial correlations between the statistical units. However, in order to obtain the estimates it is necessary to invert the covariance matrix, which poses a computational burden when used on the whole brain space and is therefore more useful when examining restricted regions of interest in the brain. Another option is to use data reduction techniques to approximate the underlying spatial process, as demonstrated by Hyun et al (2014)[33] and Zhu et al (2014)[34].

Instead of working with high dimensional spatial processes, a more efficient approach is to obtain low rank models by approximating the covariance matrix using a reduced number of basis functions, as shown by Cressie and Johannesson (2008)[35]. The basis functions are typically non-linear functions such as radial basis functions, b-splines or wavelets. Another recent approach proposed by Gershman et al (2011)[36] models fMRI data as a superposition of image sources constructed from adaptive radial basis functions. Thus they abstract away from the voxels, but the basis functions do not clearly map on to the biology.

Huertas, I. et al. (2017) [2] propose a spatial model based on a data-driven parcellation of the brain that is derived from measures of brain function as proposed by van Oort, E.S.B. et al. (2016) [3] (Instantaneous Connectivity Parcellation), and a Bayesian linear model is used for model fitting. This model is motivated by recent evidence of temporally independent, spatially overlapping subnetworks within functional networks in the human brain that are believed to represent fine-scale computational units. This model is generic and can be adapted for many different brain regions. It offers the benefit of a substantial reduction in the number of parameters to be estimated and clinical interpretability of the inferences. Additionally, the Bayesian framework offers a quantification of uncertainity in the parameter estimates. The model proposed by Huertas, I. et al. (2017) [2] is employed for the research presented in this thesis and is discussed in more detail in the next section.

Trend surface model

A classical geostatistical model [37] is represented as

$$\mathcal{Y}(\mathbf{s}) = \mu(\mathbf{s}) + e(\mathbf{s}) \tag{3.15}$$

where $\mathcal{Y}(\mathbf{s})$ is the continuous spatial random field, $\mu(\mathbf{s}) \equiv E[\mathcal{Y}(\mathbf{s})]$ is the mean function (assumed deterministic and continuous) and $e(\cdot) \equiv \{e(\mathbf{s}) : \mathbf{s} \in D\}$, where $\mathbf{s} \in D$ and $D \subset \mathcal{R}^d (d = 2 \text{ or } 3)$

is the study region of interest, is a zero-mean random error process satisfying the stationarity condition. The mean function $\mu(\cdot)$ accounts for large scale spatial variation and $e(\cdot)$ accounts for small-scale spatial variation as well as measurement error, which can be further broken down into these components as $e(\mathbf{s}) = \eta(\mathbf{s}) + \epsilon(\mathbf{s})$, where $\eta(\mathbf{s})$ is the spatially dependent component and $\epsilon(\mathbf{s})$ is the measurement error.
The most commonly used parametric mean model is a linear function given by

$$\mu(\mathbf{s}; \beta) = \mathbf{X}(\mathbf{s})^T \beta \tag{3.16}$$

where $\mathbf{X}(\mathbf{s})$ is the vector of covariates observed at $\mathbf{s}$ and β is an unrestricted parameter vector. The covariates often include the attributes of interest at the spatial co-ordinate $\mathbf{s}$. However, if data on such potentially useful attributes is not available then the mean functions is expressed as a polynomial function of the spatial co-ordinate alone; such models are called trend surface models.

A standard method of fitting a linear mean function to spatial data is ordinary least squares (OLS). However due to the sensitivity of OLS estimators to outliers and their potential to have large variances due to the possible multicollinearity of the covariate in a trend surface model, there is a need for procedures that allow for more robust estimates.

Huertas, I. et al. (2017) [2] employ a spatial statistical modelling framework that models neuroimaging data using data-driven, biologically relevant basis functions obtained through instantaneous correlation parcellations ([3]). A bayesian regression method is used to automatically find a linear weighted sum of basis functions to fit an imaged brain region of interest. We briefly explain the trend surface model based on the presentation in Huertas, I. et al. (2017) [2].

Consider a dataset involving S subjects. The preprocessed three dimensional data voxels from each of the S subjects is collected into a vector $\mathbf{y}_s$ of length V. The objective is to predict the data in $\mathbf{y}_s$ using a set of basis functions $\{\mathbf{\Phi}_m(\mathbf{x})\}_{m=1}^M$ that vary over the spatial domain $\mathbf{x}$.

$$\mathbf{y}_s = \sum_{m=1}^{M} \mathbf{w}_m \phi_m(\mathbf{x}) + \epsilon_s \tag{3.17}$$

where, M is the total no. of basis functions, $\epsilon_s \sim N(0, \theta^{-1})$, θ is the noise precision and $\mathbf{w}_s = [w_{1,s}, \ldots, w_{M,s}]^T$ is an M dimensional weight vector of regression coefficients.

Empirical Bayes

This approach assumes that subjects are independent realizations of the same distribution.

$$p(\mathbf{Y}, \mathbf{\Phi}, \mathbf{W} \mid \alpha, \theta) = \prod_{s=1}^{S} p(\mathbf{y}_s \mid \mathbf{\Phi}, \theta, \mathbf{w}_s) p(\mathbf{w}_s \mid \alpha) \tag{3.18}$$

where $\mathbf{\Phi}$ is a $V \times M$ matrix of basis functions, $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$ is an $M \times S$ matrix containing the weight vectors for each subject and $\mathbf{Y}$ is a $V \times S$ matrix of the neuroimaging data for all subjects.
A Gaussian prior $p(\mathbf{w}_s \mid \alpha) = \mathcal{N}(\mathbf{w}_s \mid \mu, \mathbf{\Lambda}_\alpha^{-1})$ is assumed over the weights for each subject, where μ is assumed to be zero. $\mathbf{\Lambda}_\alpha$ is the inverse covariance matrix, or the precision matrix, which is shared across all the subjects and has $\alpha = [\alpha_1, \ldots, \alpha_m]^T$ as hyperparameters. The precision matrix $\mathbf{\Lambda}_\alpha$ is diagonal, with an independent parameter α_m for each basis function that acts as an automatic relevance determination (ARD) prior to emphasize informative basis functions

consistently across all subjects. The precision parameters α are estimated from the data using an empirical Bayes approach, as shown below. For fixed α and β the posterior distribution over all parameters of interest ($\mathbf{W}$ in this case) can be computed in closed form, since the posterior obtained from the combination of a Gaussian prior and a Gaussian likelihood is also Gaussian. Therefore,

$$p(\mathbf{W} \mid \mathbf{Y}, \mathbf{\Phi}, \alpha, \theta) = \prod_s p(\mathbf{w}_s \mid \mathbf{y}_s, \mathbf{\Phi}, \alpha, \theta) \tag{3.19}$$

and

$$\prod_s p(\mathbf{w}_s \mid \mathbf{y}_s, \mathbf{\Phi}, \alpha, \theta) = \mathcal{N}(\mathbf{w}_s \mid \bar{\mathbf{w}}_s, \mathbf{A^{-1}}) \tag{3.20}$$

where $\mathbf{A} = \theta \mathbf{\Phi}^T \mathbf{\Phi} + \mathbf{\Lambda}_\alpha$ and $\bar{\mathbf{w}}_s = \theta \mathbf{A^{-1}} mb \Phi^T \mathbf{y_s}$.

Full Bayes

The model specified in eq.(3.18) does not account for the uncertainity in estimation of all the parameters, nor does it explicitly account for spatial correlations between basis functions. A fully bayesian approach is proposed to remedy these drawbacks.

$$p(\mathbf{Y}, \mathbf{\Phi}, \mathbf{W}, \mathbf{\Lambda}_\alpha, \theta \mid \gamma_\theta, \gamma_\alpha) = p(\theta \mid \gamma_\theta) p(\mathbf{\Lambda}_\alpha \mid \gamma_\alpha) \prod_{s=1}^{S} p(\mathbf{y}_s \mid \mathbf{X}, \theta, \mathbf{w_s}) p(\mathbf{w_s} \mid \mathbf{\Lambda}_\alpha) \tag{3.21}$$

This model now accommodates spatial correlations between basis functions by allowing off-diagonal entries in $\mathbf{\Lambda}_\alpha$. The prior over the weights is the same as in the empirical bayes case and additionally the other parameters are also assigned weights as follows:

The ARD (automatic relevance determination) precision matrix is assigned a Wishart prior:

$$p(\mathbf{\Lambda}_\alpha \mid \gamma_\alpha) = \text{Wish}(\mathbf{\Lambda}_\alpha \mid N, \mathbf{P}), \tag{3.22}$$

where N is degrees of freedom and $\mathbf{P}$ is the precision of the prior. The prior over the regression coefficients has a Gamma distribution

$$p(\theta \mid \gamma_\theta) = \text{Gamma}(\theta \mid a, b) \tag{3.23}$$

where a, b are the shape coefficients. This choice of priors once again utilizes the conjugacy of the distributions and simplifies the inference. A blocked Gibbs sampling algorithm samples from the full conditional distribution of each block of variables conditioned on the current estimates of all others, thus reducing a high-dimensional distribution to simpler low-dimensional distributions to be sampled from.

For each of the $t = 1, \ldots, T$ iterations in the Markov chain samples are drawn from the full conditional distributions of $\mathbf{W}, \theta$ and $\mathbf{\Lambda}_\alpha$ based on current estimates for other parameters:

$$\begin{aligned}
p(\mathbf{W}^{(t+1)} \mid {\mathbf{\Lambda}_\alpha}^{(t)}, \theta^{(t)}, \mathbf{Y}) &= \prod_{s=1}^{S} \mathcal{N}\left(\mathbf{w}_s^{(t+1)} \mid \bar{\mathbf{w}}_s^{(t+1)}, (\mathbf{A}^{-1})^{(t+1)}\right) \\
p(\theta^{(t+1)}) \mid {\mathbf{\Lambda}_\alpha}^{(t)}, \mathbf{W}^{(t)}, \mathbf{Y}) &= \text{Gamma}\left(\theta^{(t+1)} \mid a + \frac{SV}{2}, b + \frac{1}{2} \sum_s (\mathbf{y}_s - \mathbf{\Phi}\mathbf{w}_s^{(t)})^T (\mathbf{y}_s - \mathbf{\Phi}\mathbf{w}_s^{(t)})\right) \\
p({\mathbf{\Lambda}_\alpha}^{(t+1)} \mid \mathbf{W}^{(t)}, \theta^{(t)}, \mathbf{Y}) &= \text{Wish}\left({\mathbf{\Lambda}_\alpha}^{(t+1)} \mid N + S, \mathbf{P} + \sum_s \mathbf{w}_s^{(t)} (\mathbf{w}_s^{(t)})^T\right)
\end{aligned} \tag{3.24}$$

Huertas, I. et al. (2017) found that while Empirical Bayes approach is relatively efficient for small number of basis functions, it does not scale well for models involving a larger number of basis functions and that the MCMC approach of Full Bayes is usually an order of magnitude faster. The Full Bayes approach is also more robust to overfitting in the case of high numbers of basis functions (with strong correlations). The Full Bayes approach quantifies the uncertainty across all modeled paramters and propagates that uncertainty through to the predictions – which is necessary for applications where predictive uncertainty is important. In our implementation, we use the trend surface model with the Full Bayes approach, and adjust the priors to be non-informative.

Benefits of the TSM approach

This spatial approach provides an elegant alternative to the classical mass-univariate approach and has several advantages. Firstly, the number of statistical tests needed for inference is greatly reduced, leading to greater statistical power. Secondly, the computational units are biologically meaningful unlike the arbitrary voxels, which enables more directly interpretable clinical outcomes. Thirdly, the hierarchical structure of the basis functions allows for the capturing of long range interactions as well as local dependencies. Finally, there is no commitment to a specific scale of parcellation, as a result of which the trend surface model can be applied to regions with different scales of resolution.

Instantaneous Connectivity Parcellation (ICP)

Most existing parcellations of the brain employ a bottom-up strategy of creating a parcellation by clustering smaller units, such as voxels. Such parcellations are dependent on the initial definition of their smallest atoms and any inaccuracies in the atom level parcellation can propagate through to the highest levels. Voxels, which when convolved with the spatial smoothing kernel form the smallest possible entities in MRI, have no inherent biological validity and are therefore not the ideal starting point for parcellation.

ICP [3] adopts a top-down approach to parcellating the brain - a large-scale brain region of interest (ROI) that is based on prior anatomical or functional knowledge is divided into smaller sub-regions that are considered functionally homogenous, based on their temporal signature. van Oort, E.S.B. et al. (2016) apply independent component analysis to resting state fMRI data from the Human Connectome Project, after the data was transformed using a modified Pearson correlation to amplify minor temporal differences between individual voxels within the larger brain ROI.

For our research we use a basis set that employs soft parcellations, as opposed to the common practice of using hard partitioning of the brain through clustering techniques. Soft parcellation allows a spatial unit to be involved in different networks, and for a more gradual transition between regions. This also mitigates the risk of mixing signals from different brain regions if a spatial parcel has been misspecified.

Inference

We employ classical inference on the estimates from the trend surface model in order to enable a direct comparison with the results of Eklund et al. (2016) [1]. We fit the trend surface model to the contrast maps of the parameter estimates and their variances (that are results from subject-level analysis using FSL), obtaining parameter estimates and variances in the ICP - based parcel space, at the subject-level.

For group-level inference we create random groups of size N from the list of available subjects. At each group we compute the group-level t-statistic at each parcel. Let $\hat{\beta}_{ki}$ denote the estimate of the parameter at parcel i, $i = 1 \dots P$ for subject k where $k = 1 \dots N$ and P is the total number of parcels (or basis functions; in the order of a few hundred) in the basis set being used. Let $\hat{\gamma}_{ki}$ denote the mean of the variance of parameter estimates, and $\hat{\sigma^2}_{ki}$ denote the variance in the variance of parameter estimates for subject k and parcel i. Then for subject k at each parcel i an F-statistic[[20]] is computed

$$f_{ki} = \frac{\hat{\gamma}^2_{ki}}{\hat{\sigma^2}_{ki}}$$

and the group-level pooled variance at each parcel is computed as

$$F_i = \frac{\sum_{k=1}^{N} f_{ki}}{N-1}$$

and the t-statistic as

$$t_i = \frac{\sum_{k=1}^{N} \hat{\beta}_{ki}/N}{\sqrt{F_i/N}}$$

We obtain a corrected p-value using the Bonferroni correction (since in this particular work we employ uncorrelated basis functions)

$$\alpha_{parcel} = 0.05/P$$

for a nominal false positive rate of 5%, which is then used to test for the FWER across all parcels.

4. Data Analysis and Results

Introduction

We have seen so far an overview of the state of the art in the statistical analysis and inference of fMRI data. While the classical approach of mass-univariate analysis offers flexibility in the modeling of fMRI data, due to the large number of tests that have to be performed simultaneously for hundreds of thousands of voxels there is a need for multiple testing correction. The presence of spatial correlations in the data mean that straightforward correction procedures such as the Bonferroni correction are no longer applicable, and instead complex post-hoc correction procedures such as the random field theory need to be applied (Section 3.2). It has recently been demonstrated by Eklund et al. (2016) that popular fMRI analysis packages such as FSL, SPM and AFNI lead to very high false positives for cluster-wise inference and conservative false positives for voxel-wise inference, when tested using resting-state fMRI data as real-world null data. Eklund et al. (2016) inferred that the primary cause for such performance was the violation of one of the two underlying assumptions of random field theory , namely, that the spatial autocorrelation in the data has a squared exponential structure.

In light of these results, there is a need to explore alternative modeling approaches to fMRI data that do not involve the same drawbacks as the mass-univariate approach. Huertas, I. et al. (2017) [2] propose one such approach that overcomes the multiple testing problem by using a biologically meaningful, parsimonious basis set that divides the brain into a few hundred parcels using instantaneous correlation parcellations [3]. Huertas, I. et al. (2017) [2] use a Bayesian regression model that automatically finds an appropriate linear-weighted sum of the parcels to fit an imaged brain region of interest.

Our research evolved in this context – to examine the performance of the trend surface model presented by Huertas, I. et al. (2017) [2] in line with the analysis performed by Eklund et al. (2016) [1] in order to demonstrate that the trend surface model performs optimally with regard to Type I error, without sacrificing sensitivity. To this end, we first successfully replicate the results obtained by Eklund et al. (2016) for FSL, using one of the datasets. Then we test for the specificity of the trend surface model on the same resting-state fMRI data and show that it performs optimally. We further test for the sensitivity of the trend surface model to demonstrate that the specificity of the model does not come at the cost of its power – and we achieve this by testing the trend surface model on task fMRI ('activated') data.

In this chapter, we first describe in detail the data used for our research, followed by the steps involved in our analysis of the data at the subject and group levels. We present the results obtained in our research to demonstrate the performance of trend surface model .

Data

Based on the nature of activation, fMRI data consists of two kinds - resting state fMRI and task fMRI data. Resting-state fMRI measures spontaneous, low frequency fluctuations in the BOLD signal and is obtained by scanning subjects who are told not to perform any cognitive, language or motor tasks. Such data has been used to identify spatially distinct areas of the brain that demonstrate synchronized BOLD fluctuations at rest (called Resting State Networks, RSN). Some examples of RSNs are the Default Mode Network (DMN), the somatosensory network, the visual network and the auditory network.

Task fMRI on the other hand, is obtained using task based or stimulus driven paradigms and has been critical to our current understanding of brain function. Certain areas of the brain are inferred to be activated based on relative changes from baseline in the BOLD signal in those regions during the performance of a task or in response to a stimulus by the subject in the scanner. Typical tasks used for MRI experiments are timed experimental events such as finger-tapping, listening to sounds, or viewing faces. Every confounding parameter is controlled so that only the performance of the task (or the state of rest in case of resting state fMRI data) is supposed to create the measured BOLD signal fluctuations (for example, the subject can be asked to close their eyes while performing a listening task, to avoid spurious visual events causing undesired activations in the brain). More complex confounding factors such as breathing or heart beat can be recorded and handled in image pre-processing (as explained in section 2.2.3).

We use publicly shared anonymized fMRI datasets that are subject to the ethics guidelines of the review boards local to the collection sites, that approved the experiments and the dissemination of the datasets. We use resting state fMRI data from the 1000 Functional Connectomes Project as 'null' data to test the specificity of the trend surface model, and task fMRI data from the Human Connectome Project to test for the sensitivity of the trend surface model .

Resting state fMRI data

The resting state fMRI data from 198 healthy controls was downloaded from the 1000 Functional Connectomes Project [38] Cambridge dataset. This dataset was chosen because it was one of the datasets used by Eklund et al. (2016) [1] for their analysis. The Cambridge dataset contains data from 75 male and 123 female subjects between the ages of 18 – 30 years (mean 21.03, SD 2.31). The data was collected using 3 Tesla (T) scanners with a TR of 3 seconds, consisting of 119 time points per subject and $72 \times 72 \times 47$ voxels of size $3 \times 3 \times 3$ mm^3.

A note on the suitability of resting state fMRI data as null data [1]

Resting-state data should not contain any systematic changes in brain activity such as those observed in task fMRI data. However, while a subject does not perform any tasks during the collection of resting state fMRI data, the brain may still be engaged in many forms of mental activity and the data may also be affected by consistent trends or transients (due to the delay of the hemodynamic response function) at the start of the scanning session. In order to counter these possible effects, we follow the example of Eklund et al. (2016) and assume multiple activity paradigms (two block based and two event related) to induce randomness and destroy any systematic effects present in the data. Also the residuals in task fMRI data have been found to have resting-state networks (Fair et al (2007)), suggesting a similarity in the covariance structure of task and resting-state fMRI data – which means that resting-state data and task noise are similar in nature. This is confirmed by Eklund et al. (2016) , who further concluded that such autocorrelation is not brain related and is instead due to MR acquisition. Eklund et al. (2016)

also performed a two-sample t-test on task fMRI data by comparing two groups of randomly chosen subjects from a set of homogenous subjects (that should effectively destroy any activation patterns in the data) that produced a similar rate of false positives as the resting state data, confirming the poor specificity of the software packages examined.

Task fMRI data

The task fMRI data used for the sensitivity analysis was from the 500 Subjects Release (2014) of the Human Connectome Project, that includes behavioral and 3T MR imaging data from over 500 healthy adult participants. We use data from four tasks (one contrast for each task) : working memory(2BK-0BK), gambling (reward–punishment), emotion (faces – shapes) and language (story – math). From a total of 339 unrelated subjects in the HCP database, we chose the subset of subjects that were processed using the r227 image reconstruction algorithm.

Description of tasks

We describe each of the tasks used in our analysis on the basis of information provided by Barch et al. (2014) [39]

1. **Working Memory task (2BK - 0BK)**:
   An N-back working memory task involves identifying whether the current stimulus presented is the same as that presented N stimuli before. Within each run, four different stimulus types (pictures of faces,places, tools and body parts) are presented in separate blocks within the run. Within each run, half of the blocks use a 2-back working memory task (respond âĂŸtargetâĂŹ whenever the current stimulus is the same as the one two back) and half use a 0-back working memory task (a target cue is presented at the start of each block, and the person must respond âĂŸtargetâĂŹ to any presentation of that stimulus during the block). A 2.5 s cue indicates the task type (and target for 0-back) at the start of the block. Each of the two runs contains 8 task blocks (10 trials of 2.5 s each, for 25 s) and 4 fixation blocks (15 s each). On each trial, the stimulus is presented for 2 s, followed by a 500 ms inter-trial interval. Each block contains 10 trials, of which 2 are targets, and 2âĂŞ3 are non-target lures (for example, repeated items in the wrong n-back position, either 1-back or 3-back).

2. **Gambling (Reward - Punishment):**
   Participants play a card guessing game where they are asked to guess the number on a mystery card (represented by a "?") in order to win or lose money. They are told that potential card numbers range from 1âĂŞ9 and to indicate if they think the mystery card number is more or less than 5 by pressing one of two buttons on the response box. Feedback is the number on the card which is generated by the program as a function of whether the trial was a reward, loss or neutral trial. The feedback is either: (i) a green up arrow with "$1" for reward trials (ii) a red down arrow next to - $0.50 for loss trials or (iii) the number 5 and a gray double headed arrow for neutral trials. The "?" is presented for up to 1.5 s followed by feedback for 1.0 s. There is a 1.0 s inter-trial interval with a "+" presented on the screen. The task is presented in blocks of 8 trials that are either mostly reward (6 reward trials pseudo randomly interleaved with either 1 neutral and 1 loss trial, 2 neutral trials, or 2 loss trials) or mostly loss (6 loss trials interleaved with either 1 neutral and 1 reward trial, 2 neutral trials, or 2 reward trials). In each of the two runs, there are 2 mostly reward and 2 mostly loss blocks, interleaved with 4 fixation blocks (15 s each). All participants are provided with money as a result of completing the task, though it is a standard amount across subjects.

3. **Emotion "Hariri Task":**
   In this task developed by Hariri and colleagues (Hariri et al., 2002), participants are pre-

sented with blocks of trials that either (i) ask them to decide which of two *faces* presented on the bottom of the screen match the *face* at the top of the screen, or (ii) which of two *shapes* presented at the bottom of the screen match the *shape* at the top of the screen. The faces shown have either angry or fearful expressions. Trials are presented in blocks of 6 trials of the same task (face or shape), with the stimulus presented for 2 s and a 1 s inter-trial interval. Each block is preceded by a 3s task cue (âĂIJshapeâĂİ or âĂIJfaceâĂİ), so that each block is 21 s long, including the cue. Each of the two runs includes 3 face blocks and 3 shape blocks.

4. **Language (Story - Math):**
The task consists of two runs that each interleave four blocks of a story task and four blocks of a math task. The lengths of the blocks vary (average of approximately 30 s), but the task was designed so that the math task blocks match the length of the story task blocks, with some additional math trials at the end of the task to complete the 3.8 minute run as needed. The story blocks present participants with brief auditory stories (5âĂŞ9 sentences) adapted from AesopâĂŹs fables, followed by a two-alternative forced-choice question that asks participants about the topic (or 'moral') of the story. The math task also presents trials auditorily and requires subjects to complete addition and subtraction problems and choose the correct answer from the two choices provided following each arithmetic operation. The goal of including the math blocks was to provide a comparison task that was demanded attention, was similar in auditory and phonological input, and unlikely to generate activation of anterior temporal lobe regions involved in semantic processing, though likely to engage numerosity related processing in the parietal cortex.

Data Analysis

The data analysis involved two steps. A first -level analysis of the subjects was performed using FSL to pre-process the data and to fit a general linear model to obtain contrasts of parameters estimates and their variances at the subject-level. A second level analysis then used these summary statistics from the subject-level in order to conduct group-level inference. This section elaborates on the details of the two levels of analyses performed on the resting state fMRI and task fMRI data.

First Level Analysis

Resting state fMRI (Cambridge) data

In order to replicate the results obtained by Eklund et al. (2016) we reproduce their first-level analysis [1]. We use a processing script (set up using the FSL FEAT GUI) to perform a first level analysis for each subject to obtain brain activation maps in a standard brain space (Montreal Neurological Institute (MNI)). The pre-processing involved normalization to a brain template, motion correction and four different levels of smoothing (4 mm, 6 mm, 8 mm, 10 mm full width half maximum (FWHM)). Additionally, a general linear model was applied to the preprocessed data using two block designs (boxcar10 (B1), boxcar30 (B2)) and two event-related designs (E1, E2) as in Eklund et al. (2016) . The spatial normalization to the brain template (MNI152 T1 2mm brain.nii.gz) was performed as a two step linear registration using the function FLIRT. One fMRI volume was aligned to the anatomical volume using the boundary based registration) (BBR) option in FLIRT. The anatomical volume was aligned to MNI space using a linear registration with 12 degrees of freedom , and the two transforms were finally combined. The first level models were fit in the subject space after spatial smoothing, and the contrasts and their variances were then transformed to the atlas space.

Task fMRI (HCP) data

The downloaded HCP data consists of precomputed fixed effects analysis to estimate the average effects across runs within-participants . The motion correction, distortion correction, and registration to MNI standard space for each individual task run were carried out in accordance with the minimal preprocessing pipeline set forth in Glasser et al (2013) [40]. The volume data for individual task runs was processed in a Level 1 analysis that included high-pass filtering at 200s, and smoothing of 4 mm FWHM in volume space. FSL 'film_gls' was employed for model estimation and temporal autocorrelation. The two runs for each task and subject were then combined in a Level 2 fixed-effects analysis to obtain the task fMRI analysis dataset for each task and subject.

Group Level Analysis

As a result of the first-level analyses using FSL, we obtained the contrasts of parameter estimates ('copes') and their variances ('varcopes') for each of the conditions in the task data as well as each smoothing-level/study-design combination for the resting state fMRI data. The group level analyses use these contrasts from the first level analyses for inference. The group-level analyses on resting state fMRI data were performed using FSL FLAME 1 to replicate the results of Eklund et al. (2016) , and using the trend surface model to test the specificity of TSM. To test the sensitivity of the trend surface model , we perform a group level analysis on the precomputed first level HCP task fMRI data.

FSL FLAME 1

The group analyses were performed using FSL FLAME 1 that employs random effects group analysis, i.e., it uses both the beta weight and the variance of each subject in the group and estimates a between subject variance. A total of 1000 groups with 20 randomly selected subjects each were generated and tested for group activity using a one sample t-test. A particular group analysis was considered to give a significant result if any cluster had a FWE-corrected p-value < 0.05. Since the subjects were not performing any task and all are healthy and of a similar age, the expected number of analyses with one or more significant effects should follow the nominal rate (5 % in our analysis).

Trend surface model (TSM)

We first perform principal component analysis (PCA) on the original set of basis functions that were generated using instantaneous correlation parcellations to obtain an orthogonal set of basis functions. We choose the first n principal components that explain 90% of variance in the model (4.1). The subject level (first level) contrasts and their variances are fit with the trend surface model to these n principal components to obtain parameter estimates and variances in the principal component space at the subject level.

Then for the group level analyses, a total of 1000 groups of 20 subjects were randomly selected from the list of subjects (the Cambridge dataset in case of the resting state fMRI data and the list of unrelated HCP subjects in the case of the task fMRI data). We obtain group-level t-statistics at each principal component using the subject level trend surface model estimates and on the basis of these t-statistics the groups were tested for activation at the group level using a one-sample t-test. A particular group analysis is considered to give a significant result if any principal component has a Bonferroni-corrected p-value < 0.05. Since the principal components are orthogonal, the Bonferroni-corrected p-value would not be too conservative. The principal components that show significance in a majority (i.e., in $> 50\%$) of the random group tests are

mapped to the MNI space to produce a visualization and compared to the results published by Barch et al. (2014) .

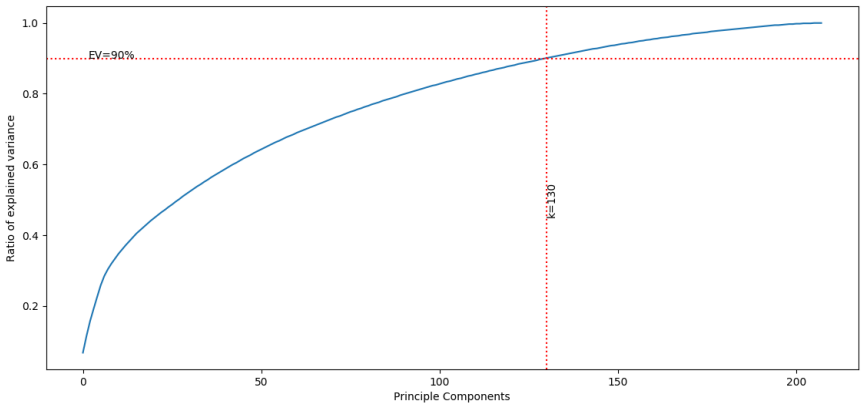

Figure 4.1: Selecting the principal components that explain 90 % of variance in the model

Results

Reproduction of results from Eklund et al. (2016)

Eklund et al. (2016) found that among the parametric software packages they tested, FSL FLAME 1 produced much lower FWE compared to the other packages, that was often valid (under 5%). We therefore reproduce the analysis performed by Eklund et al. (2016) for the FLAME 1 package of FSL and use it as a benchmark to compare our results from the trend surface model. We analyzed Cambridge dataset using four different activity paradigms (two block based (10s and 30s on-off) and two event related (E1: 2s activation-6s rest and E2: 1-4s activation/3-6s rest, randomised), four levels of subject-level smoothing (4mm, 6mm, 8mm, 10mm) and two cluster defining thresholds (z=2.3,3.1), and we were able to successfully reproduce the results obtained by Eklund et al. (2016) (Figure 4.2).

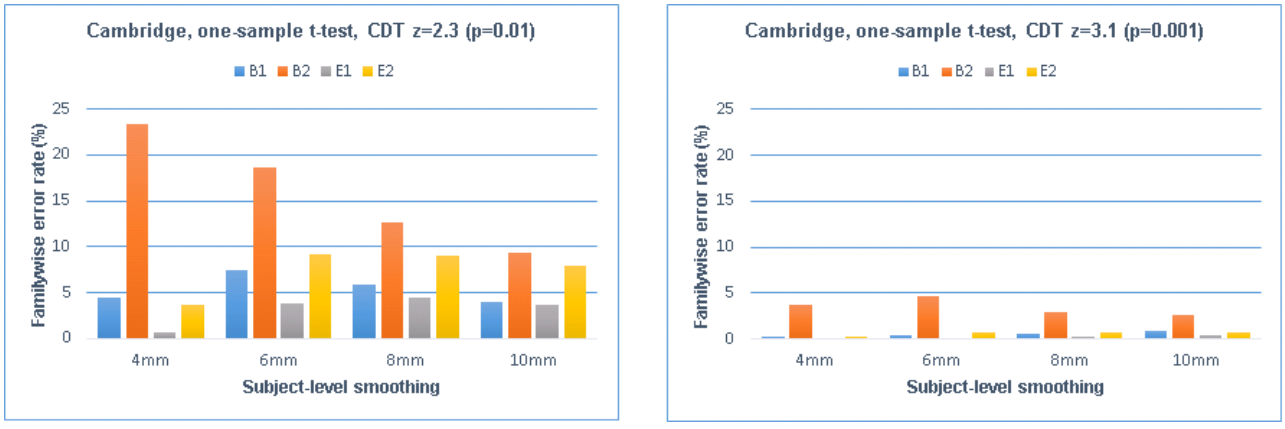

Figure 4.2: FLAME 1 results for one-sample t-test showing estimated FWE rates for Cambridge data across all four smoothing levels and activity paradigms, using FWE-corrected threshold of P=0.05 and cluster defining thresholds (CDT) of P=0.01 and P=0.001. The groups of 20 subjects were chosen based on the permutations generated by Eklund et al. (2016) and 1,000 such random group analyses were performed for each smoothing-level and activity paradigm.

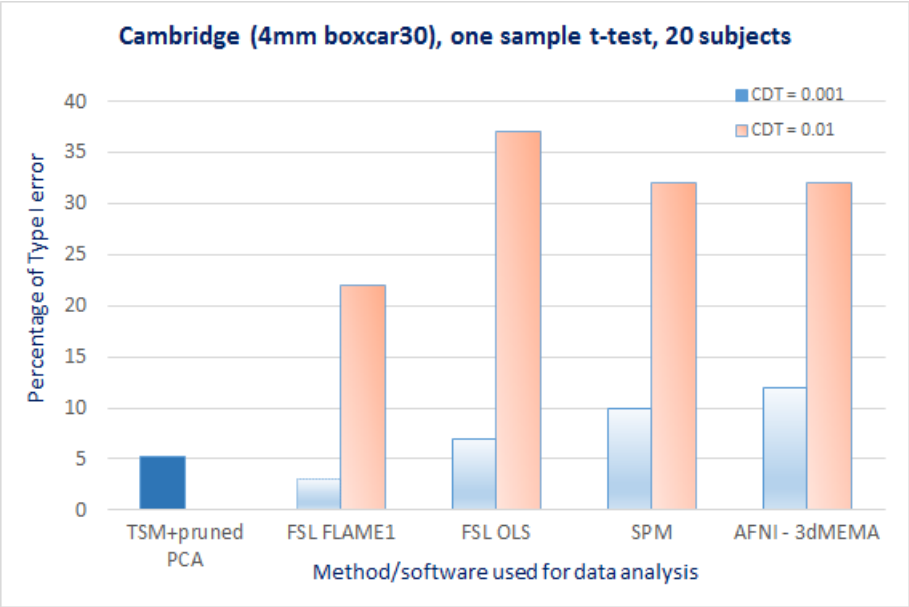

Figure 4.3: False positive rate obtained from TSM, in comparison to the software packages FSL, SPM and AFNI for two cluster defining thresholds – P = 0.01 and P = 0.001 – as shown by Eklund et al. (2016)

Specificity analysis of the TSM

We analyzed data from the Cambridge dataset for a smoothness of 4mm and the boxcar30 block design using trend surface model and performed a one-sample t-test, as explained in Section 4.3.2. We obtained a false positive rate of between 5 - 6 % (variation due to random tests) as opposed to over 20 % false positive rate from FSL FLAME1, as shown in Fig. 4.3.

Sensitivity analysis of the TSM

We evaluate the performance of the TSM on task fMRI by comparing the resulting brain maps to the results presented by Barch et al. (2014) . Based on this comparison we conclude that the trend surface model captures all the networks of activation that are expected to arise under the particular task.

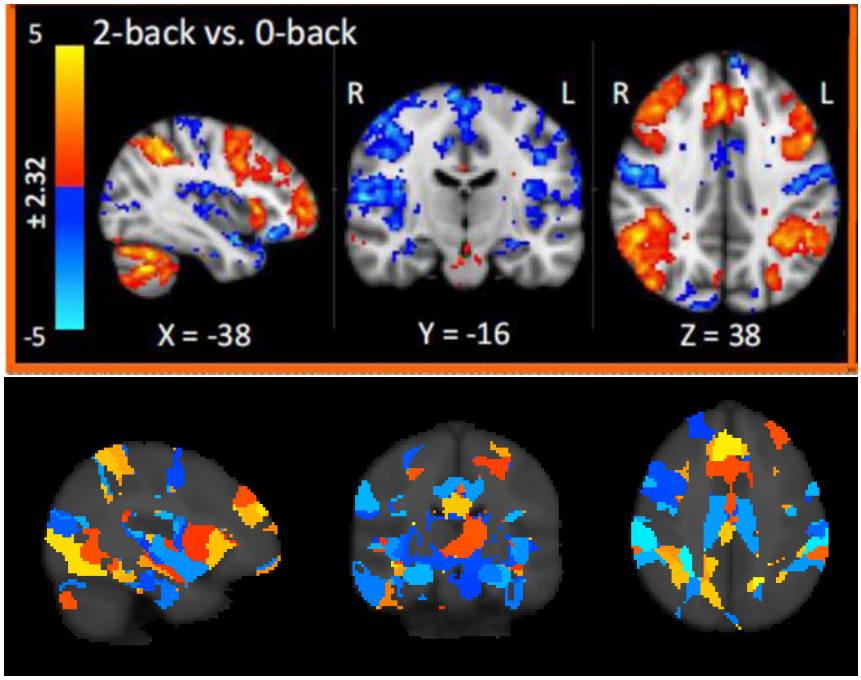

Figure 4.4: Brain activation under the Working Memory task from Barch et al. (2014) [39] (top) and that obtained using TSM (bottom).

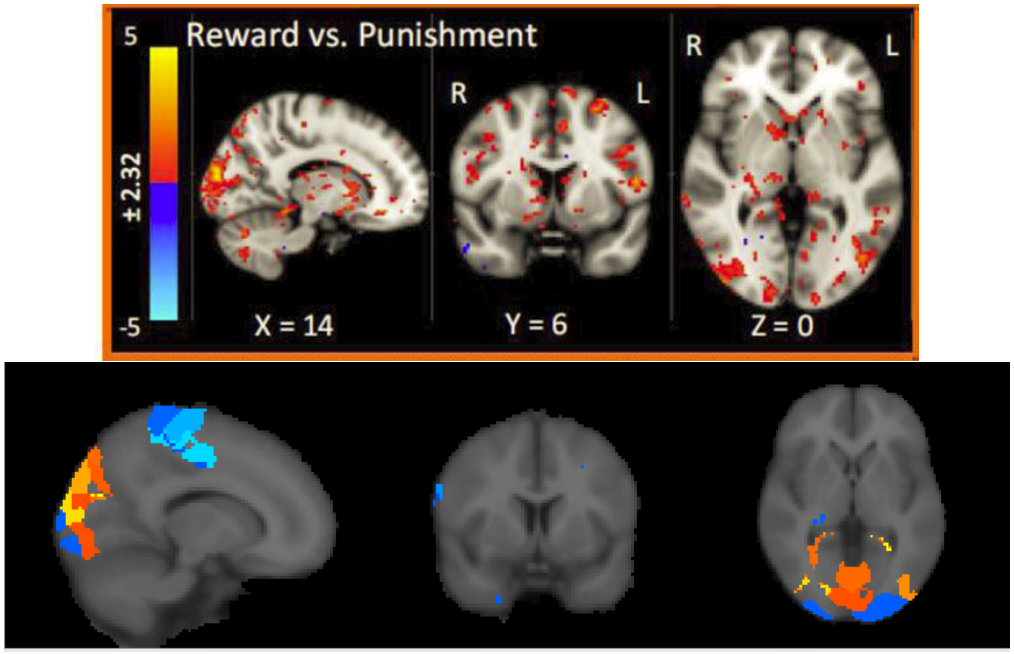

Figure 4.5: Brain activation under the Gambling task from Barch et al. (2014) [39] (top) and that obtained using TSM (bottom)

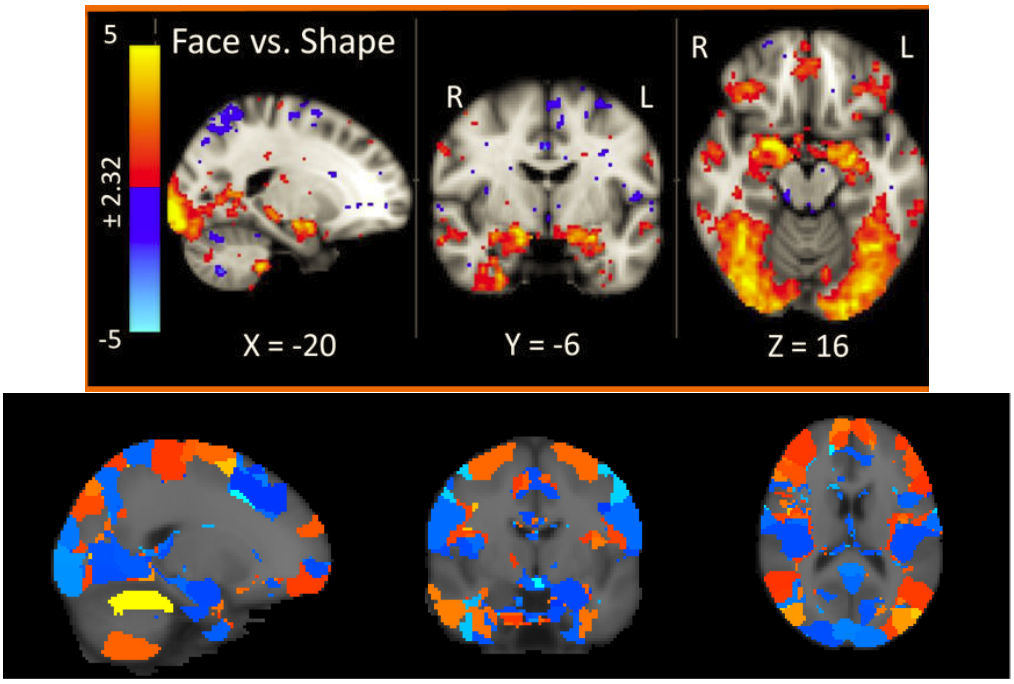

Figure 4.6: Brain activation under the Emotion task from Barch et al. (2014) [39](top) and that obtained using TSM (bottom).

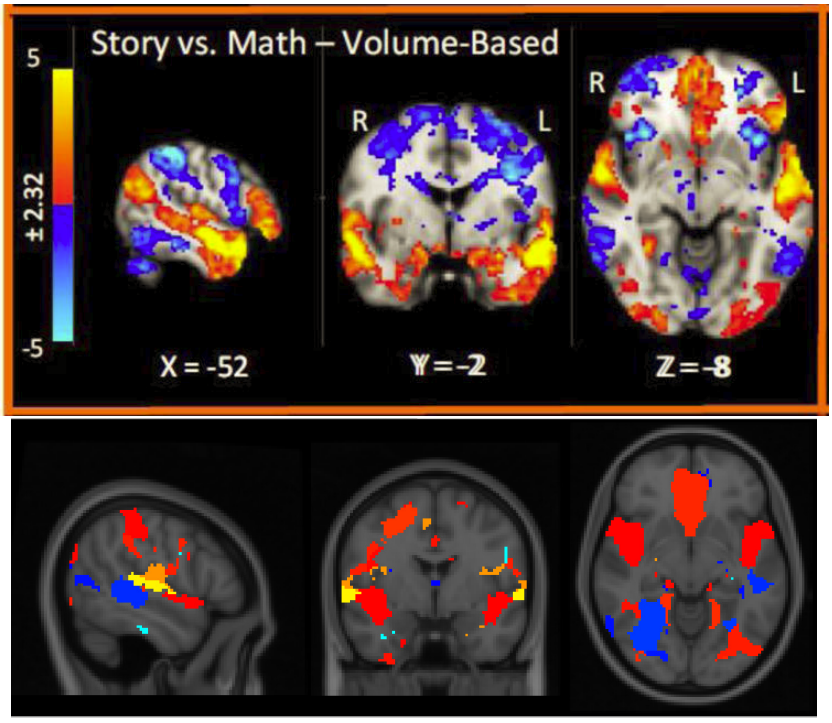

Figure 4.7: Brain activation under the Language task from Barch et al. (2014) [39] (top) and that obtained using TSM (bottom).

5. Discussion

The classical approach to analyzing fMRI data that involves fitting independent linear models at each of the hundreds of thousands of voxels requires a correction for multiple comparisons. Given the spatial correlation in the data as a result of the spatial nature of fMRI as well as the pre-processing steps involved, a Bonferroni correction is too conservative and this calls for complex post-hoc correction procedures such as the Random Field Theory. It has been shown by Eklund et al. (2016) that popular parametric softwares FSL and SPM that employ the random field theory result in a high degree of false positives for cluster-wise inference. They conclude that this occurs because the spatial correlation in the data violates the random field theory assumptions regarding the spatial auto-correlation function.

This necessitates the exploration of other approaches to analyze fMRI data, and we choose to examine the performance of the model set forth by Huertas, I. et al. (2017) [2]. Huertas, I. et al. (2017) propose a trend surface model based on biologically relevant parcellation that overcomes the multiple comparisons problem by regressing on parcels of the brain that are temporally correlated. In this work, we test the performance of this approach on resting state fMRI and task fMRI data to test for specificity and sensitivity, respectively. By using a subset of the data used by Eklund et al. (2016) we show that the trend surface model performs optimally - achieving the nominal familywise error rate while being sensitive to all activated brain regions in the task fMRI data.

While voxel-based methods provide the advantage of greater spatial specificity, a voxel (whose dimensions are often arbitrarily chosen) is of no biological significance and hence, unlike the parcels in the TSM, does not provide a biologically meaningful interpretation. In addition, due to the smoothness that is characteristic of fMRI data and also a consequence of the pre-processing steps, the voxel is no longer well-defined. The TSM also offers the advantage of highly reduced number of computational units, and hence a reduction of the multiplicity involved in the inference. As a consequence, there is no longer a need for complex correction procedures such as the RFT. While permutation testing provides optimal inference, considering the computational intensity of the procedure, the adequacy of simple correction procedures for inference using the TSM make the latter an attractive option.

For our analysis we used a hierarchical basis set with two levels of hierarchy that was generated using the Instantaneous Connectivity Parcellation method. An exploratory analysis of the basis set being used showed that there were significant correlations among particular sets of basis function in the second-level (each correlated set belonged to the same higher level parcel). This was causing a high rate of false positives when the trend surface model was tested on the resting state fMRI data. Therefore we had to perform principal components analysis (PCA) on the original basis set to obtain uncorrelated regressors, and we chose the least noisy PCs that explained 90% of the variance. A trend surface model based on these modified orthogonal regressors provides us with the nominal false positive rate – that is expected from the advantages provided by the trend surface model . However, using this modified basis set means that the spatial biological structure of the brain is not reflected in the regressors anymore.

Therefore, as part of our future work, we plan to develop a new basis set that is biologically relevant, while being composed of uncorrelated basis functions. We will test the performance of the trend surface model on the same datasets using this basis set, and we expect optimal performance. We will further expand the analysis to all the datasets analyzed by Eklund et al. (2016) . This would also allow us to examine the performance of the trend surface model in the presence of different levels of spatial smoothing in the data. Since the trend surface model does not use the random field theory or incorporate assumptions regarding the spatial correlation, we expect that the model would perform equally well on all levels of smoothing.

Bibliography

[1] Anders Eklund, Thomas E. Nichols, and Hans Knutsson. Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 113(28):7900–7905, 2016.

[2] Ismael Huertas, Marianne Oldehinkel, Erik S.B. van Oort, David Garcia-Solis, Pablo Mir, Christian F. Beckmann, and Andre F. Marquand. A bayesian spatial model for neuroimaging data based on biologically informed basis functions. *NeuroImage*, 161:134 – 148, 2017.

[3] Erik S. B. van Oort, Maarten Mennes, Vinod J. Kumar, Wolfgang Grodd, and Christian F. Beckmann. Human brain parcellation using time courses of instantaneous correlations. *NeuroImage*, 2017.

[4] Liina Pylkkanen. Functional Neuroimaging (http://www.psych.nyu.edu/pylkkanen/Neural_Bases /07_slides/05_Methods.pdf).

[5] Martin Lindquist and Tor D. Wager. *Principles of fMRI*. Leanpub, 2015.

[6] Ani Eloyan, Shanshan Li, John Muschelli, Jim J. Pekar, Stewart H. Mostofsky, and Brian S. Caffo. Analytic programming with fmri data: A quick-start guide for statisticians using R. *PLoS One*, 9(2), 2014.

[7] Maccotta L. Ollinger J. M. Petersen S. E. Miezin, F. M. and R. L. Buckner. Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage*, 11:735–759, 2000.

[8] K. J. Friston, P. Jezzard, and R. Turner. Analysis of functional mri time-series. *Human Brain Mapping*, 1(2):153–171, 1994.

[9] Nicholas Lange and Scott L. Zeger. Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(1):1–29, 1997.

[10] K.J. Friston, A.P. Holmes, J-B. Poline, P.J. Grasby, S.C.R. Williams, R.S.J. Frackowiak, and R. Turner. Analysis of fmri time-series revisited. *NeuroImage*, 2(1):45 – 53, 1995.

[11] K.J. Friston, P. Fletcher, O. Josephs, A. Holmes, M.D. Rugg, and R. Turner. Event-related fmri: Characterizing differential responses. *NeuroImage*, 7(1):30 – 40, 1998.

[12] Martin A. Lindquist and Tor D. Wager. Validity and power in hemodynamic response modeling: a comparison study and a new approach. *Human brain mapping*, 28 8:764–84, 2007.

[13] FMRIB Analysis Group and MGH Boston (Multiple Authors). FSL course. *https://fsl.fmrib.ox.ac.uk/fslcourse/*, 2017.

[14] M. A. Lindquist. The Statistical Analysis of fMRI Data. *ArXiv e-prints*, June 2009.

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

[16] Mark Jenkinson, Christian F. Beckmann, Timothy E.J. Behrens, Mark W. Woolrich, and Stephen M. Smith. Fsl. *NeuroImage*, 62(2):782 – 790, 2012. 20 Years of fMRI.

[17] Karl Magnus Petersson, Thomas E. Nichols, Jean-Baptiste Poline, and Andrew P. Holmes. Statistical limitations in functional neuroimaging. I. Non-inferential methods and statistical models. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354(1387):1239–1260, 1999.

[18] K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, and W. D. Penny. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2006.

[19] Avital Cnaan, Nan M. Laird, and Peter Slasor. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, 16(20):2349–2380.

[20] Christian F. Beckmann, Mark Jenkinson, and Stephen M. Smith. General multilevel linear modeling for group analysis in fmri. *NeuroImage*, 20(2):1052 – 1063, 2003.

[21] R. S. J. Frackowiak, C. D. Frith, C. J. Price, J. T. Ashburner, K.J. Friston, R. J. Dolan, S. Zeki, and W. Penny. *Human Brain Function*. Academic Press, second edition, 2004.

[22] K.J. Friston, D.E. Glaser, R.N.A. Henson, S. Kiebel, C. Phillips, and J. Ashburner. Classical and bayesian inference in neuroimaging: Applications. *NeuroImage*, 16(2):484 – 512, 2002.

[23] K.J. Friston and W.D. Penny. Posterior probability maps and SPMs. *NeuroImage*, 19(3):1240–1249, 2003.

[24] C. R. Genovese. A bayesian time-course model for functional magnetic resonance imaging data. *Journal of the American Statistical Association*, 95(451):691âĂŞ703, Sep 2000.

[25] J. Kershaw, B. A. Ardekani, and I. Kanno. Application of bayesian inference to fmri data analysis. *IEEE Transactions on Medical Imaging*, 18(12):1138–1153, 1999.

[26] Niels VÃęver Hartvig. A stochastic geometry model for functional magnetic resonance images. *Scandinavian Journal of Statistics*, 29(3):333–353, 2002.

[27] Mark W. Woolrich, Timothy E.J. Behrens, Christian F. Beckmann, Mark Jenkinson, and Stephen M. Smith. Multilevel linear modelling for fmri group analysis using bayesian inference. *NeuroImage*, 21(4):1732 – 1747, 2004.

[28] C. GÃűssl, D.P. Auer, and L. Fahrmeir. Dynamic models in fmri. *Magnetic Resonance in Medicine*, 43(1):72–81, 2000.

[29] C. GÃűssl, D. P. Auer, and L. Fahrmeir. Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, 57(2):554–562, 2001.

[30] Niels VÃęver Hartvig and Jens Ledet Jensen. Spatial mixture modeling of fmri data. *Human Brain Mapping*, 11(4):233–248, 2000.

[31] Jane Neumann and Gabriele Lohmann. Bayesian second-level analysis of functional magnetic resonance images. *NeuroImage*, 20(2):1346 – 1355, 2003.

[32] W. D. Penny, N. J. Trujillo-Barreto, and K. J. Friston. Bayesian fmri time series analysis with spatial priors. *NeuroImage*, 24:350–362, 2005.

[33] Li Y. Gilmore J. H. Lu Z. Styner M. Hyun, J. W. and H. Zhu. Sgpp: spatial gaussian predictive process models for neuroimaging data. *NeuroImage*, 89(2):70–80, 2014.

[34] Hongtu Zhu, Jianqing Fan, and Linglong Kong. Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association*, 109(507):1084–1098, 2014. PMID: 25435598.

[35] Noel Cressie and Gardar Johannesson. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(1):209–226, 2008.

[36] Samuel J. Gershman, David M. Blei, Francisco Pereira, and Kenneth A. Norman. A topographic latent source model for fmri data. *NeuroImage*, 57(1):89 – 100, 2011.

[37] S. Banerjee B. P. Carlin and A. E. Gelfand. *Hierarchical modeling and analysis for spatial data*. CRC Press, 2014.

[38] Multiple authors. 1000 Functional Connectomes Project (http://fcon_1000.projects.nitrc.org).

[39] Deanna M. Barch, Gregory C. Burgess, Michael P. Harms, Steven E. Petersen, Bradley L. Schlaggar, Maurizio Corbetta, Matthew F. Glasser, Sandra Curtiss, Sachin Dixit, Cindy Feldt, Dan Nolan, Edward Bryant, Tucker Hartley, Owen Footer, James M. Bjork, Russ Poldrack, Steve Smith, Heidi Johansen-Berg, Abraham Z. Snyder, and David C. Van Essen. Function in the human connectome: Task-fmri and individual differences in behavior. *NeuroImage*, 80:169 – 189, 2013. Mapping the Connectome.

[40] Matthew F. Glasser et al. The minimal processing pipelines for the human connectome project. *NeuroImage*, 80:105–124, 2013.