DPhil Thesis

Department of Engineering Science
Submitted Trinity Term 2015

Probabilistic Modelling of Functional Modes in the Human Brain

Sam Harrison

Balliol College
Life Sciences Interface Doctoral Training Centre
Oxford Centre for Functional MRI of the Brain
Oxford Centre for Human Brain Activity

Supervisors

Prof. Steve Smith
Prof. Mark Woolrich

This thesis is submitted to the Department of Engineering Science, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Probabilistic Modelling of Functional Modes in the Human Brain

Sam Harrison

Abstract

It is well established that it is possible to observe spontaneous, highly structured fluctuations in human brain activity from functional magnetic resonance imaging (fMRI) data when the subject is 'at rest'. This activity can be decomposed into groups of spatially distributed brain regions that are consistently temporally co-activated, and this thesis is concerned with developing new methods to identify these functional modes.

We introduce a probabilistic model that allows us to infer functional modes without making restrictive assumptions about the spatio-temporal interactions between them, and furthermore, the model also accounts for the variability of these modes over subjects. Both of these facets of the model represent advances compared to current mode-identification techniques. We use a computationally efficient variational Bayesian approach to make inferences from this model, and this allows us to draw upon the enormous amount of information available from the types of large-scale fMRI data collection initiatives that are becoming the norm.

We demonstrate, using data from over 450 subjects collected as part of the Human Connectome Project, that we can reliably infer a set of modes that captures an enormous degree of spatial variability over subjects and that makes a novel set of predictions about the temporal relationships that modes have with one another. Finally, we highlight the importance of functional registration and bring to light a surprising link between temporal non-stationarities and the fMRI global signal.

DPhil Thesis: Department of Engineering Science, University of Oxford.
Submitted Trinity Term 2015.

Acknowledgements

First and foremost, I would like to thank Mark and Steve for their wonderful supervision over the past three years. Their willingness to let me explore my own interests has made for a fascinating few years of research, and their remarkable generosity with their time and firm grasp of the bigger picture has meant that, hopefully, it has been productive as well. I am enormously grateful for all their efforts.

There are of course many more people with whom I have worked on this project. Right at the beginning of it all, Adrian's enthusiasm, sage advice and uncanny ability to make VB work meant that I could not have got off to a better start. More recently, I would like to thank the two people who, aside from Mark and Steve, I have worked with most closely over the last year: Janine, for reminding me that the childlike pleasure associated with finding things out is what makes research so compelling, and Andy, for his palpable enthusiasm about collaboration, even when we weren't finding things out. I would also like to acknowledge the contribution made by Christian, MJ and Emma, with whom we have had many key discussions about the PFM approach, and thank MJ for all the teaching opportunities I have been afforded.

Furthermore, FMRIB and OHBA have proved to be invigorating places in which to do research, and I would like to thank everyone at those two centres for making my time there so enjoyable. In particular, to Adam, Giles, Zita, Andrew, Malcolm, and everyone else who I have been fortunate enough to spend time in the Lakes

with over the last few years: if me completing this thesis brings the Boot Imaging & Neuroscience Tent even one step closer, then, for that reason alone, I will consider it a success.

I am indebted to all those involved in the Human Connectome Project for providing the data on which this thesis is based, and on a personal level, I would like to thank Matt in particular. Aside from his uncanny ability to spot a mislabelled brain region at twenty paces, his different perspective on this work has been an invaluable addition to the discussion.

I would like to thank the EPSRC for funding me, and the LSI-DTC for pointing me towards an area of research that has fascinated me for the past three years. The lasting friendships I made there have made my time in Oxford all the more memorable. Mike, Mike, Fred, Tom, Marc et al. deserve a special mention, and I mean it as a sincere compliment when I say that the amount I enjoyed our customary four course lunches was only matched by the trepidation I felt when faced with the cycle back up the hill afterwards.

I have lived with a fantastic group of people over the last four years. My thanks to Jay, Laurie, Rahul and everyone else at Balliol who made the Manor such a unique experience. More recently, Rich, Andrea, Jon, Lloyd and James: I could not have wished for a better group of people with whom to plumb the depths of the Oxford housing market.

Liv, your dogged insistence that there is no problem in life that cannot be solved with an appropriately sized Tupperware has helped keep everything in perspective. I am enormously grateful for all your help, and for making the last year such a joy. I can only apologise if this thesis was more stressful for you than it was for me.

Finally, I would like to thank my family, and Mum, Dad and Kate in particular, for all their love over the years. Your unwavering support, and just maybe, your innumerable supply of jigsaws, made this all possible. Thank you.

Technical Acknowledgements

This thesis would not have been possible without the huge amount of neuroimaging data and high-quality software that has been made freely available. All of these projects have my sincere thanks, and I have tried to acknowledge as many of these as possible below.

Software

The code that does the probabilistic inference is written in **C++** (isocpp.org). The bulk of the computation is done by the **Armadillo** linear algebra library (arma.sourceforge.net) and **OpenBLAS** (www.openblas.net). The code is parallelised using **OpenMP** (openmp.org) and also relies on several functions from the **Boost** libraries (boost.org).

Analysis and visualisation of the results is done in **Python** (python.org). This relies on **NumPy** (numpy.org) and **SciPy** (scipy.org), and all plots are generated by **matplotlib** (matplotlib.org/).

Git (git-scm.com) is used for version control.

We use several pieces of software specific to neuroimaging. The inference code interfaces with **FSL** (fsl.fmrib.ox.ac.uk), the visualisation code interfaces with **NiBabel** (nipy.org/nibabel), and cortical surfaces are generated using **Connectome Workbench** (humanconnectome.org/software/connectome-workbench.html).

This thesis is typeset in **LuaTeX** (luatex.org) and uses the **Linux Libertine**

(linuxlibertine.org) and **DejaVu** (dejavu-fonts.org) font families. Extra graphics are provided by the **TikZ** and **PGF** packages (sourceforge.net/projects/pgf). Bibliographies are managed and typeset using **JabRef** (jabref.sourceforge.net) and **BibLaTeX** (ctan.org/pkg/biblatex) respectively.

Data

Data were provided by the **Human Connectome Project**, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Uğurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

The data can be downloaded from humanconnectome.org/data.

Analyses

The work in this thesis is all my own, though I am grateful to Steve and Janine for the sets of modes identified by ICA used in Chapter 5. An earlier version of many of these analyses can be found in our initial publication of the PFM model [Harrison et al. 2015], and much of the interpretation of the PFM model has been informed by the fruitful discussions with the other authors on that paper.

The code used to generate the results in this thesis can be found at git.fmrib.ox.ac.uk/samh/profumo.

Contents

Key Abbreviations & Acronyms

MRI	Magnetic resonance imaging
fMRI	Functional magnetic resonance imaging
rfMRI	Resting-state functional magnetic resonance imaging
BOLD	Blood-oxygen-level dependent [signal]
HRF	Hæmodynamic response function
MEG	Magnetoencephalography
EEG	Electroencephalography
ECoG	Electrocorticography
PCA	Principal component analysis
SVD	Singular value decomposition
ICA	Independent component analysis
sICA	Spatial independent component analysis
tICA	Temporal independent component analysis
TFM	Temporal functional mode
PFM	Probabilistic functional mode
HCP	Human Connectome Project
VB	Variational Bayes

The idea that … thought itself, … emotion and reason, … memories, dreams and reflections should consist of jelly, is simply too strange to understand. *Henry Marsh [2014]*

Chapter 1

Introduction

1.1 A History of Resting-State fMRI

The concept of the resting state represents a radical departure from traditional approaches to neuroimaging. Subjects are scanned without any overt task to perform, simply letting their minds wander[1]. The simplicity of the acquisition belies the complexity of the data that results, and the staggering amount of information about individuals that this conveys.

Here, we chart the development of the resting state, describing how it evolved from more traditional approaches before discussing some of the key findings to date and speculating on what the immediate future holds.

1.1.1 Early days of fMRI

Understanding how the neural circuitry in the brain is harnessed to perform cognitive tasks is a fundamental scientific question, for which dedicated research began in earnest in the 19th century. Initially, studies like the seminal work of Hughlings

[1] Or, surprisingly regularly, sleep [Tagliazucchi and Laufs 2014].

Jackson [1879], Broca [1861] and Wernicke [1874] to name but a few[2], relied upon examination of patients with specific neurological conditions; inferences about function were made by pairing differences in behaviours with observable morphological differences. Only shortly after, Brodmann [1909] published his pioneering work that demarcated brain regions based on differences in the underlying cellular architecture, illustrating that the complex functional behaviours of the brain are underpinned by a similarly complex structure. However, it was the advent of neuroimaging techniques that simultaneously allow detailed examination of the brain's activity and structure that revolutionised the field.

Initially, positron emission tomography (PET) scanning dominated neuroimaging research, with the first proof of principle functional human brain scan taking place in 1976 [Hoffman et al. 1976; Nutt 2002], a whole year before the first complete MR slice through the body was demonstrated [Damadian et al. 1977]. PET directly images brain metabolism by recording the spatial location of radioactively labelled variants of key neural molecules. With these very direct measures of metabolism, PET became the workhorse of functional neuroimaging for almost the next two decades. However, it has several key limitations, including requiring a small radiation dose with each scan and having very limited temporal resolution.

A rich new avenue of exploration was opened up in 1990 when Ogawa and colleagues discovered the blood-oxygenation-level dependent (BOLD) signal in MRI [Ogawa and Lee 1990; Ogawa et al. 1990a; b]. It had been known since 1936 that the magnetic properties of hæmoglobin depend on whether or not oxygen is bound to the molecule [Pauling and Coryell 1936]; Ogawa's key observation was that these subtle differences give an observable change in the MRI signal which is dependant on the concentration of oxygen in the blood. This observation, that MRI could be indirectly used to measure metabolic processes in the brain, paved

[2]The book by Eling [1994] ties together much of this early work, and more information on Hughlings Jackson can be found in the book by Critchley and Critchley [1998].

the way for the new field of fMRI[3].

1.1.2 Task fMRI

MRI scanners had been commercially available for a decade when the BOLD signal was discovered, so the relative maturity of the technology meant that it was not long before fMRI was producing images of brain activity in humans. In 1992 three fMRI studies were published, reporting activation in either the visual or motor systems dependant upon the tasks performed by the subjects [Bandettini et al. 1992; Kwong et al. 1992; Ogawa et al. 1992].

These three studies established the blueprint for the task based experiments that have given enormous insights into the functional organisation of the brain. Typically, a group of subjects are scanned while undertaking a cognitive, sensory or motor task. Crucially, the task is performed multiple times, with task repetitions being interspersed with a suitable control condition. Then, the observed fMRI response in each voxel is compared to the timings at which the task was performed, after suitable adjustment to account for the hæmodynamic processes. The standard inference is that any voxels which have a statistically similar response to the task presentations—that is to say voxels that are more activated in the task than control conditions—are part of the functional network that is recruited for the cognitive process under examination[4].

This approach has been enormously successful, yielding some remarkable insights into brain function through the application of a staggeringly diverse set of tasks. For example, the task paradigm has been used to research a set of topics as varied as the phenomenal flexibility of the neural architecture [Amedi et al. 2003; Goyal et al. 2006], the brain's internal representation of space [Doeller et al. 2010],

[3]For a good overview of the development of fMRI see Huettel et al. [2004].

[4]With the normal, though frequently omitted, caveats about correlation and causation [Van Horn and Poldrack 2009].

the neural substrate of the pain response both within the brain and the spinal cord [Sprenger et al. 2015], and why we found magical stories so appealing [Hsu et al. 2015]. However, all these investigations share the same fundamental analysis technique which, conceptually, remains essentially identical from task to task.

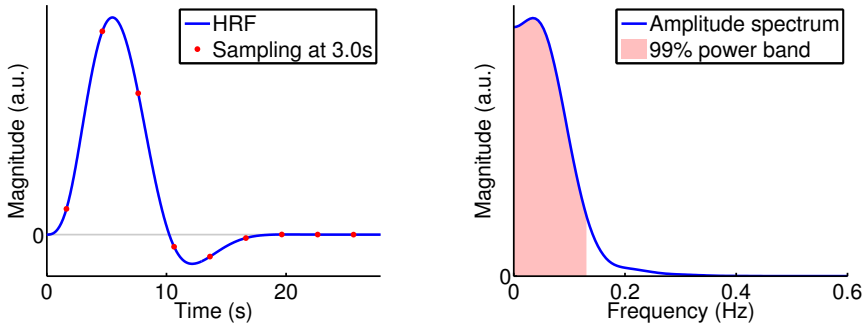
Implications of the hæmodynamic response function (HRF)

It became apparent very early on that the dynamics of blood flow in the brain dominate the observed responses in fMRI. There is a complex sequence of physiological events that link neural activity to changes in the BOLD signal, and this needs to be accounted for before relating the task presentations to the fMRI data. While complex biophysical models can be formulated [Buxton et al. 1998], most models just capture two key properties: relative to the neuronal signal, the BOLD signal is both delayed and temporally blurred.

The standard approach is to approximate the BOLD signal as a linear convolution of the underlying neural activity. An example of a hæmodynamic response function (HRF) that is commonly used to perform this convolution is shown in Figure 1.1(a), and the delay and blurring actions of this filter are clear. In the frequency domain, the hæmodynamic processes act as a low-pass filter. Figure 1.1(b) shows the frequency response of the HRF, and in this case 99 % of the power is admitted at frequencies of less than 0.15 Hz—theoretically, this tells us that we can sample once every three seconds and still comfortably capture all of that signal.

Naturally, the HRF is much more complex than a simple linear filter. It has been shown to vary across both subjects and brain regions [Aguirre et al. 1998; Handwerker et al. 2004] and to have strongly coupled spatio-temporal interactions [Kriegeskorte et al. 2010], amongst many other complex properties. There also appear to be intriguing non-linearities, as data acquired at higher sampling rates have been shown to contain much richer temporal dynamics [Smith et al. 2012].

However, the fundamental point remains—fMRI data is dominated by the low-



(a) Example hæmodynamic response function, randomly drawn from a linear basis set [Woolrich et al. 2004].

(b) Frequency response of the hæmodynamic response function, along with the 99 % low-pass power band [Heinzel et al. 2002].

Figure 1.1: Temporal characteristics and frequency content of a hæmodynamic response function.

frequency blurring of the HRF and any analysis techniques must be robust in light of those properties, whatever the experimental paradigm.

Alternative task analysis techniques

While the traditional, hypothesis-driven approach to task analysis was, and continues to be, enormously successful, it requires the experimenter to propose an explicit model for the neural response to the task in question[5]. Traditionally, the proposed neural response is simply that the neural populations of interest will be more active during the task than during the control condition.

However, there are many classes of tasks where this simple model is unrealistic. For example, problems could arise either if the task is complex enough that it induces a non-trivial sequence of neural responses, or if habituation changes the response across repeated stimulus presentations, or in patient populations, or for any number of other reasons[6]. In these cases, it would be useful to be able to

[5]Of course, this is simply the (Popperian) scientific process, and the ability to refute hypotheses is arguably the greatest strength of the whole procedure [Friston 1998].

[6]Note that we define the task paradigm as one in which a stimulus is presented, and inferences are made by comparing the responses to a pre-specified model. This allows for much more complex procedures than the simple task-versus-control example given above. For example, it is possible to

interrogate the data in a more free-form way, even if only to 'provide insights that allow better models to be elaborated' [Friston 1998].

In 1998, Terry Sejnowski's group provided a template for doing exactly that. Their three hugely influential papers from that year introduced independent component analysis (ICA) as a method for analysing task fMRI data [McKeown et al. 1998a; b; McKeown and Sejnowski 1998]. Rather than looking for correlations with a pre-specified time course, ICA seeks to decompose the data into a set of maximally independent components; in the case of fMRI data, a component consists of a time course, capturing when that component was active during the scan, and a spatial map, encoding which voxels in the brain contained that time course (see Chapter 2 for a much more thorough discussion of ICA and related approaches). Crucially, the ICA pipeline proceeds without any knowledge of the task, and the model is inferred in a completely data-driven way—the interpretation of the results is normally informed by a post hoc comparison between the component time courses and the task itself.

The methods community seized on this new development, and within a few years ICA-based techniques had been developed to perform group analyses [Calhoun et al. 2001], automatically infer the number of components of interest [Beckmann and Smith 2002], separate physiological signals from artefacts [Kiviniemi et al. 2003], and many more besides [Calhoun et al. 2003].

In reality, these data-driven analyses have had a relatively minor impact on analysis of task fMRI data. As Friston [1998] noted at the time, 'the hypothesis based scientific process … has arisen … because it is so efficacious' and in that sense was always unlikely to be usurped. However, as we shall see later, it was the validation of techniques that used the fMRI data itself to define an internal,

model the effects of habituation on neural responses, referred to as repetition suppression, and this has allowed, for example, some genuinely astonishing investigations into the neural substrates of the brain's internal representational space [Doeller et al. 2010; Barron et al. 2013].

rather than an external, model that would prove crucial.

1.1.3 Emergence of the resting state

In their eloquent synthesis of the history of the resting state, Callard and Margulies [2011] ascribe the emergence of the new discipline to the coalescence of two distinct lines of research: the efforts to characterise the spontaneous fluctuations observed in fMRI, and the attempts to understand what should be considered an appropriate baseline for task fMRI. While the two strands now form a mutually dependent whole, they have shaped, and continue to shape, the study of rest in distinct ways.

Here, we follow the structure of Callard and Margulies' account; we give a brief summary of the initial bodies of work, before describing some of the important contributions to the nascent composite discipline.

Structured fluctuations and functional connectivity

The first line of research was primarily concerned with the characterisation of the spontaneous fluctuations observed in fMRI data. For task fMRI, these fluctuations had long been considered as noise; while problematic, the assumption was that the fluctuations would be 'sufficiently random … to be averaged out in statistical analysis' [Deco et al. 2011].

The key observation came from Biswal et al. [1995], whose study showed that these spontaneous fluctuations exhibited a spatial structure that was strikingly similar to patterns of task-induced activity. The concept was a simple one: they looked for regions that correlated with the time course of a voxel in the left motor cortex of a subject scanned at rest. They observed that the pattern of inter- and intra-hemispheric correlations within the motor cortex at rest was remarkably similar to the regions activated by a simple finger-tapping task. Their figure 3, in particular, is a beautifully concise summary of what was a revolutionary concept.

This introduced a new way of characterising relationships between brain

regions, namely *functional connectivity*. This is simply defined as 'statistical dependencies among remote neurophysiological events' [Friston 2011], so states that we can consider regions with similar patterns of activity to be connected in some sense. This has generated its fair share of controversy[7], but nonetheless was a profound reimagining of the way interactions between brain regions are characterised.

However, this approach was not immediately accepted. Three years later, Ogawa et al. [1998] covered this in a review and, while obviously excited by the findings, were concerned that 'it is yet to be shown that the origin of the correlation is neurophysiological'. Even five years later, Cordes et al. [2000] showed that Biswal's findings could be extended to multiple brain regions, and that the functional connectivity between regions was dominated by low-frequency fluctuations that seemed to be distinct from cardiac and respiratory artefacts [Cordes et al. 2000; 2001]. What is interesting is that methodologically, this was essentially a replication of Biswal's original study; even at this stage there was still a concerted effort to confirm the legitimacy of the functional connectivity approach.

The hunt for a baseline and the re-emergence of the self

The second line of work was primarily concerned with whether or not there is a baseline brain state, against which the activity induced by task demands can be reliably compared. As described earlier, the task paradigm works by contrasting the activity during a cognitive task of interest with the observed activity in a control state. Initially at least, the control state was often either the subject maintaining visual fixation on a cross on a screen, but being otherwise unoccupied, or a passive condition, typically observing the same set of stimuli used for the task but with

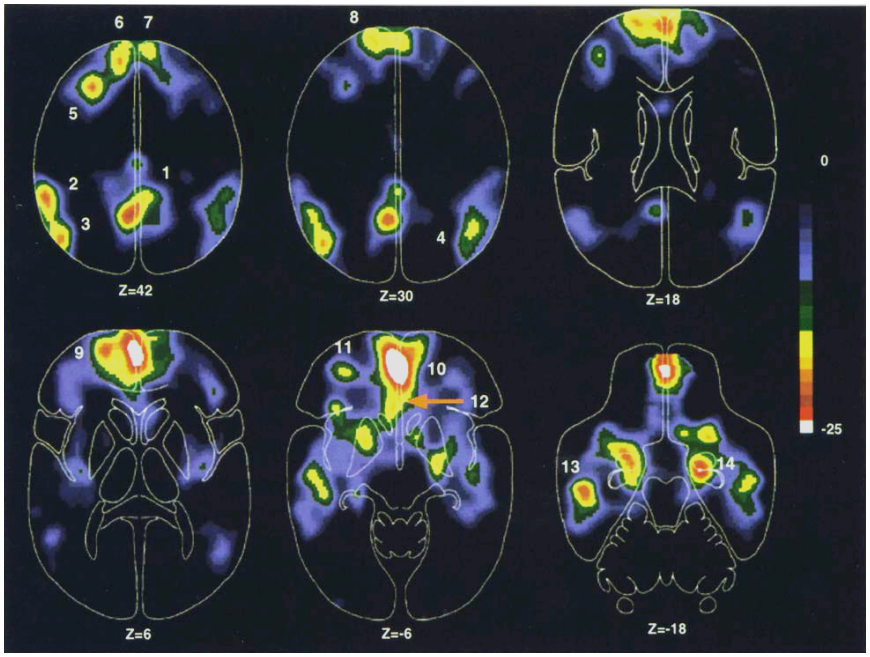
[7] In essence because functional connectivity is descriptive rather than explanatory. The alternative is *effective connectivity*, which 'refers explicitly to the influence that one neural system exerts over another' [Friston 2011] and is intrinsically linked to an underlying model of neural activity.

no engagement required. Despite an 'uneasiness in interpreting passive task conditions' [Buckner et al. 2008], this approach had served the field well.

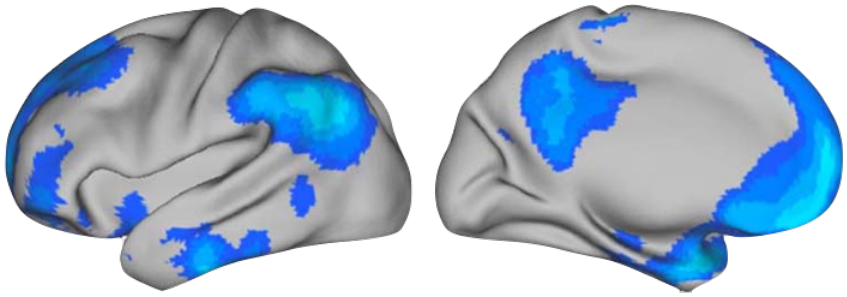
In 1997, the debate about what a suitable control state would look like was suddenly brought to the fore by a pair of meta-analyses. Shulman et al. [1997a] re-analysed nine different PET studies incorporating a diverse set of tasks, hoping to identify regions that were activated by all the tasks, and as such might be thought of as providing a common neural substrate for task responses. They found no such cortical regions. However, in the second analysis, Shulman et al. [1997b] observed that there was a set of regions that seemed to be more active during the control state than during the task state, and what's more, this set of regions could be consistently identified from each of the different tasks. These results are shown in Figure 1.2.

The consistency of this result over a wide range of tasks was striking, but the interpretation of their observation was much less clear. The task paradigm is predicated on the comparison between conditions, but this introduces an ambiguity. The first possible explanation is that these observations reflected an increase in activity induced by the control state itself, or in other words, 'any control state, no matter how carefully it is selected, is just another task state with its own unique areas of activation' [Raichle et al. 2001]. The second interpretation is antithetical to the first, and simply ascribes the relatively higher activity in the control state to a decrease in activity during the task presentations, or in other words, this is a set of functional regions that is suppressed by externally focused, cognitively demanding tasks.

In 2001, Marcus Raichle's group proposed a solution to this problem. In the first of three seminal papers they would publish that year, they hypothesised that the activation/deactivation ambiguity could be resolved by examining the brain's metabolism in more detail [Raichle et al. 2001].



(a) Brain regions that exhibited a decrease in blood flow during task conditions, as identified by Shulman et al. [1997b]. The image was generated from a meta-analysis of nine PET studies.



(b) The data from panel (a) projected onto the cortical surface of the left hemisphere by Buckner et al. [2008]. Blue regions exhibit decreases in activity during tasks.

Figure 1.2: The set of brain regions identified by Shulman et al. [1997b] that exhibit consistent decreases in activity during tasks. These would go on to form the basis of the default mode network as described by Greicius et al. [2003].

From a metabolic point of view, the brain is enormously expensive. In adults, it accounts for 2 % of body mass but its energy use accounts for 20 % of the resting metabolic rate [Kety 1957; Rolfe and Brown 1997]. Even more remarkably, Kuzawa et al. [2014] showed that the brain's energy consumption peaks at 66 % of the resting metabolic rate in children[8]. However, these phenomenal energy requirements are only slightly modulated by conscious mental activity [Sokoloff et al. 1955], with changes in metabolic demand 'usually in the range of 10 % or less, hardly a serious physiologic challenge' [Shulman et al. 1997b].

Raichle's crucial insight was that this abnormally high and consistent metabolic load could be use to define a physiological baseline. In their detailed study of the brain's metabolism using PET, Raichle et al. [2001] examined subjects at rest, with their eyes open. What they noticed was that while the blood flow to different brain regions varied considerably, the fraction of available oxygen used by different tissues was strikingly consistent across the whole brain. Their interpretation was that this consistency was indicative of a general physiological equilibrium, and therefore that the observed changes in activity induced by different tasks represented fluctuations around this baseline. This recasting of the resting state as an absolute baseline weighs in favour of the second interpretation of the observations made by Shulman et al. [1997b], namely that there is a set of 'brain function[s] that [are] suspended during specific goal- directed behaviors' [Raichle et al. 2001].

More was to follow, as 2001 proved to be something of an *annus mirabilis* for Marcus Raichle. While they had proposed sound reasons to consider the resting state as a baseline, a fundamental question remained: what does this baseline level of activity represent? In their second paper, Gusnard et al. [2001] showed

[8]Humans have exceptionally slow growth rates in the period between weaning and puberty [Tanner 1962], with the rate of growth in body mass falling to levels more commonly associated with reptiles than mammals [Walker et al. 2006]. The suggestion from the work of Kuzawa et al. [2014] is that the staggeringly high metabolic cost of the brain during this phase of development is one potential explanation of this phenomenon—there simply isn't any spare energy for growth!

that activity in the medial prefrontal cortex, one of the regions exhibiting strong task decreases, could in fact be modulated by tasks that required self-referential judgements[9]. They went on to speculate that self-referential activity and emotional processing were critical to understanding the brain at rest, and furthermore, that further examination may elucidate 'aspects of the multifaceted concept known as the "self"' [Gusnard et al. 2001].

Taken together, the two papers from Raichle's group had pulled together work from two disparate fields to form an inchoate, but nevertheless consistent, scaffold upon which further work could build. However, the radical, and lasting, legacy of this work was to completely reimagine the purpose of studying the brain at rest. Their final paper [Gusnard and Raichle 2001] was part review, part call to arms, and finished the process they had started with their speculations on the self in their previous work. In it, they fleshed out their ideas about the role of the default mode, and combined with their previous contention that 'a useful way to explore the neurobiology of the self is to explore the nature of default state activity' [Gusnard et al. 2001], the resting state had suddenly been recast with the study of the self as its raison d'être.

Coalescence

Raichle's revolutionary idea was to link the resting state and the self, but their speculations about functions had relied on reverse-inference from task data. It may seem obvious now, but it was by no means self-evident how data acquired at rest could be best utilised to answer the fundamental questions that the nascent field was starting to explore. In other words, it had been demonstrated that understanding the brain at rest was vital, but was the best way to do that to use

[9]Note that this is also another validation of Raichle's baseline hypothesis—the activity in these regions can be modulated up as well as down. While this may seem counter-intuitive, this is consistent with other brain regions. For example, activity in the visual system can be modulated around the eyes-open baseline Raichle defined: it can be reduced by closing the eyes or increased by performing a visual task.

the task paradigm to manipulate the resting state, in the way that Raichle's group had demonstrated?

In fact, it was two years before a remarkable paper by Greicius et al. [2003] linked the default mode and functional connectivity[10]. The set of regions that were deactivated by tasks was well established, but it was unclear whether the different regions were linked, or whether they were independently suppressed by task demands. The breakthrough was to examine the regions at rest, with the key finding being that they were strongly functionally connected.

Greicius et al. [2003] had provided 'compelling evidence ... for the existence of a cohesive *default mode network*' [emphasis added]. This was every bit as revolutionary as Raichle's work three years earlier and marks the point at which the default mode network, as we currently understand it, was first formally defined[11]. Suddenly, it was possible to consider the 'default [mode] network [as] a brain system much like the motor system or the visual system' [Buckner et al. 2008], aligning it with more traditional notions of neural organisation. However, in this case the crucial inference was driven by the resting-state data itself, and functional connectivity had answered a question for which the task paradigm had yielded no clear answers.

[10]Greicius et al. [2003] were the first to make the link explicit, but those examining functional connectivity often came agonisingly close. For example, Calhoun et al. [2002] published an elegant examination of the default mode using functional connectivity—they identified a set of networks with ICA, and showed that the default mode was far more active during rest than task—except the link to Raichle's work was never made, and the paper itself is a middle-of-the-road study on the neural correlates of driving.

[11]While Raichle's work on the default mode has been canonised, there are examples of other researchers providing almost prophetic insights. In a remarkable paper that has often been overlooked, Binder et al. [1999] posited that the task-induced decreases identified by Shulman et al. [1997b] were due to suppression of a '"conceptual processing" network', a full four years before Greicius et al. [2003].

1.1.4 Resting-state fMRI comes of age

Initially, the relationship between functional connectivity and the default mode was essentially symbiotic. The default mode, and the aspects of the self it purported to embody, suddenly furnished the resting state with a distinct purpose and crucial legitimacy as a self-contained field of research. Functional connectivity had provided a way to analyse a system that was inherently unamenable to task manipulations, and with the relative maturity of alternative task analysis techniques, promised much more.

Much of the early work on resting-state data remained focused on the default mode network. Almost immediately after identifying it, Greicius et al. [2004] showed that there seemed to be systematic differences in the default mode network, as identified from resting-state data, that distinguished patients with Alzheimer's disease from healthy controls. Next, and almost simultaneously, Fox et al. [2005] and Fransson [2005] showed that the default mode network was anti-correlated with regions of the brain that were activated by a wide range of tasks. This was another validation of the functional connectivity approach, demonstrating that the dichotomy between externally and internally oriented attention, and the associated brain regions, was also a feature of the dynamics of the brain at rest. Later, Greicius et al. [2009] demonstrated that the functional connections that had been used to define the default mode network reflected the underlying anatomical connections[12], in one of the first papers to compare structural and functional connectivity.

However, while the default mode network had lent an enormous amount of credibility to functional connectivity, functional connectivity would slowly erode the hegemony of the default mode network as the resting-state field matured. It was immediately apparent that it was possible to identify far more functional systems

[12] Greicius et al. [2009] motivated their paper by stating that 'considerable skepticism remains as to whether resting-state functional connectivity maps reflect neural connectivity', a mere fourteen years after the original paper from Biswal et al. [1995]!

than the default mode network using functional connectivity and resting-state data, and a flurry of papers seized upon the earlier task-focussed work on ICA and sought to characterise a more complete set of *resting-state networks* [Beckmann and Smith 2004; Beckmann et al. 2005; Damoiseaux et al. 2006; De Luca et al. 2006][13]. As more and more studies were published, a familiar set of networks began to emerge that were consistently identified, irrespective of different subjects, scan protocols or analysis techniques. But how many functional systems is it possible to identify from the resting state in this way?

In a beautiful demonstration of the staggering richness of the dynamics present in resting-state data, Smith et al. [2009] audaciously proposed that the question actually had a simple answer: all functional systems would be identifiable at rest. They analysed an online repository of results from task fMRI investigations containing, at that stage, the reported activations from over 1,500 journal articles, and involving nearly 30,000 subjects. What they neatly demonstrated was that the key functional systems that could be identified from this task meta-analysis, by virtue of their re-occurrence in a wide variety of experiments, showed an almost exact correspondence with a set of resting-state networks. Not only did the resting-state analysis demonstrate that the organisation of the brain, as revealed by functional connectivity, was consistent with the insights from the previous seventeen years of task fMRI, but it also offered an almost improbably parsimonious way of replicating the task findings—the resting-state networks were recovered from a mere three-and-a-half-hours of data collected from 36 subjects.

This engendered yet another reimaging of the resting state: the default mode network was now just one of an entire suite of resting-state networks, where 'the full repertoire of functional networks [utilised] by the brain in action is

[13]It is apparent from some of these papers quite how quickly the resting state became a preoccupation of those who had previously been developing alternative task analysis techniques—note the change between Beckmann and Smith [2002] and Beckmann et al. [2005], for example.

continuously and dynamically "active" even when at "rest"' [Smith et al. 2009].

1.1.5 Connectomics

Of course, resting-state research was not proceeding *in vacuo*. As Smith et al. [2009] ushered in a more holistic view of the dynamics of the brain at rest, both the quality and quantity of resting-state data being collected was increasing markedly. Parallel acquisition techniques were providing the impetus for considerable increases in the spatial and temporal resolution of fMRI data [Blaimer et al. 2004; Larkman and Nunes 2007], while the simplicity of resting-state acquisitions was simultaneously enabling 'data-sharing across institutions on an unprecedented scale' [Callard and Margulies 2011][14]. Here, we will briefly discuss the ways cutting-edge research is making use of this opulence and how this has facilitated yet another shift in the emphasis of resting-state research.

Access to more and better data was a boon for those analysing the resting state, as it both made new analysis techniques possible [Varoquaux et al. 2011; Smith et al. 2012] and yielded more nuanced results from existing procedures [Kiviniemi et al. 2009]. Crucially, this meant it became possible to partition large-scale resting-state networks into their functional subunits[15]. This began pushing the field towards an interpretation of resting-state data that more clearly acknowledges the distinction between functional segregation and functional integration [Friston 2011].

Functional segregation captures the idea that different processes are localised to different areas of cortex. Early attempts to parcellate the cortex were based on cytoarchitecture[16] [Brodmann 1909], but even with sophisticated modern tech-

[14]The ENIGMA consortium is a prime example of the power of this data sharing approach, and is a fascinating look at neuroimaging data with a geneticist's view of sample size [Thompson et al. 2014].

[15]For example, the default mode network, as shown in Figure 1.2, can be separated into several spatially distinct regions. These regions are normally considered as part of a functional system because they exhibit very strong temporal correlations with one another.

[16]A situation that persisted until the middle of the 20th century, when a proliferation of poor quality architectonic maps prompted a withering critique from Lashley and Clark [1946]: 'An

niques [Schleicher et al. 2005] this is still a painstaking process [Eickhoff et al. 2006; Scheperjans et al. 2008]. These difficulties meant that most of the early modern work in this area was based on monkey studies, like the iconic map of connections within the macaque visual system by Felleman and Van Essen [1991][17]. The advent of fMRI made functional mapping in humans a possibility, and the exquisite maps of visual cortex generated by retinotopic studies are arguably the best example of this [Sereno et al. 1995; Pitzalis et al. 2010]. However, as alluded to earlier, there are large swathes of association cortex for which an appropriate set of task manipulations remains elusive. Again, this opened the door for the resting state. Defining parcels as areas with similar resting-state activity naturally leads to a parcellation of the entire cortex[18], and this has received much attention [Cohen et al. 2008; Yeo et al. 2011; Craddock et al. 2012; Blumensath et al. 2013; Gordon et al. 2014].

Functional integration is concerned with how these localised processing units can be dynamically yoked together to support different functions. This examines the connections between functional subunits and there have been two major lines of investigation here. The first deals with characterising the complex network of interactions between brain regions [Bullmore and Sporns 2009], and as such can be considered as a natural evolution of the more holistic interpretation of the resting state Smith et al. [2009] precipitated. The second tries to properly characterise the complex, non-stationary[19] dynamics of these interactions and the transient nature

analysis of the literature on cytoarchitecture revealed generally such inadequacies of description, such uncritical use of differentiating criteria, and such inconsistencies in the results of different investigators in describing and charting the same regions that it seemed essential to reexamine the assumptions and principles upon which the parcellation of the cortex has been carried out.' Figure 2 from the recent review by Amunts and Zilles [2012] gives an idea of this variability.

[17] 'In many ways the famous Felleman and Van Essen model of the cortical hierarchy is the best map we have of the cortex.' [Markov et al. 2011]

[18] With the caveat that a resting-state parcellation can make no claims about the functions of regions, merely that they are functionally distinct.

[19] The field has never quite worked out what it means by non-stationarities, and the term as it is currently used conflates a number of different concepts. One suspects this is a source of wry amusement for ~~Karl~~ those studying effective connectivity, for whom there is a clear distinction

of functional connectivity [Hutchison et al. 2013].

Meanwhile, as the resting state was undergoing its own data-driven revolution, diffusion MRI was causing a similarly drastic reimagining of the study of the brain's anatomical connectivity. The in vivo analysis of the fibre pathways in the brain in large numbers of subjects had allowed the investigation of whole brain models of connectivity and, inspired by the genomics revolution, the field bullishly adopted its own *-omics* suffix [Sporns et al. 2005]. The field of connectomics was born, and the concomitant grandiloquent claims abounded: 'You are your connectome' [Seung 2012].

The resting-state community was quick to align itself with this approach, and the functional connectome—that is to say, the network of interactions between brain regions as inferred using functional connectivity—emerged as a complementary summary of brain connectivity to the structural connectome [Castellanos et al. 2013; Smith et al. 2013b; Deco and Kringelbach 2014]. This confidence in this approach has been infectious, and a myriad of large-scale projects aiming to map the connectomes have begun[20]. As such, the field stands on the brink of another revolution, with large-scale data sharing initiatives and complex, multimodal models of brain connectivity poised to transform our understanding of the macro-scale organisation of the brain once again.

between stationarity at the model level and the potentially non-stationary output dynamics.

[20]For example, the 1000 Functional Connectomes Project [Biswal et al. 2010] and the Human Connectome Project [Van Essen et al. 2013] are already yielding new insights into brain function, and the UK Biobank [Sudlow et al. 2015] is quite frankly breathtaking in its scope. For a particularly interesting counterpoint to these multi-subject projects, see the MyConnectome Project [Laumann et al. 2015].

1.2 Complementary Modalities

1.2.1 MRI-based techniques

One of the key strengths of MRI has always been that it is possible to perform different types of scans to quantify different aspects of brain structure and function, using the same hardware and within the same scanning session. This gives a multimodal description of the brain which is far broader in scope than one any other modality can muster. We have already covered functional MRI, so in this section we briefly describe insights into the brain's structure that have been derived from structural and diffusion MRI.

Structural imaging

Structural MRI looks at gross brain anatomy, so is primarily used to distinguish different tissue types, segment cortical and subcortical structures, and examine variations in morphology. In recent years, as higher resolution structural scans acquired at higher field strengths have become available, there has been a push to expand the repertoire of structural MRI by using these data to examine microstructural properties in vivo [Eickhoff et al. 2005; Glasser and Van Essen 2011; Sereno et al. 2013].

However, as well as being of interest in its own right, high quality structural information is vital for fMRI as it is needed for accurate registration—the process whereby the brains of different subjects are brought into correspondence. This is an essential step in any group analysis of fMRI data, but presents a formidable challenge as the structural variability across subjects is enormous. For example, in primary visual cortex—an area of cortex that exhibits some of the lowest levels of intra-subject functional variability [Mueller et al. 2013]—there can be a three-fold

variation in the areal size of functional regions across subjects[21] [Andrews et al. 1997]. Furthermore, there can even be pronounced differences in the shape and location of large-scale anatomical features due to variations in cortical folding patterns, even in identical twins [Bartley et al. 1997; Hasan et al. 2011].

Diffusion imaging

Diffusion MRI can detect fibre pathways in the brain, so is used to investigate anatomical connectivity. We have already touched upon how developments in this field have been seized upon by those examining the structural connectome, so we will only briefly describe two ways in which this can be used in conjunction with fMRI data.

Firstly, one of the great strengths of functional connectivity is that it can be used to augment diffusion data in known problem areas. Diffusion MRI struggles where many fibre tracts cross or concentrate, so interhemispheric and cortico-cerebellar connections are particularly problematic. In these regions, functional connectivity can be used to fill in some of the blanks, as O'Reilly et al. [2010] and Buckner et al. [2011] have shown. However, functional connectivity is sensitive to both direct and indirect connections, meaning diffusion data may be used to disambiguate the two.

Secondly, however, there is much more to be learned from the relationship between structural and functional connectivity than simply using one to complete the other. There is a huge amount of fascinating work looking at the way observed patterns of functional connectivity relate to the backbone of structural connectivity [Deco et al. 2011]. These approaches often use complex biophysical models to explain some of the emergent properties of functional data, be that the gross patterns of functional connectivity [Deco et al. 2009; Honey et al. 2009] or more

[21] And these size differences may not even be correlated across functionally related regions within subjects [Dougherty et al. 2003].

recently, their non-stationary dynamics [Hansen et al. 2015].

1.2.2 Alternative functional modalities

Finally, we touch on non-MRI based techniques that complement functional MRI by providing alternative descriptions of neural activity. We will outline a few that more directly record neural activity than fMRI, and hence give a much better temporal resolution than the BOLD signal affords.

Electroencephalography (EEG) records the voltage fluctuations at the scalp caused by neural currents, giving a signal that is very temporally rich. However, the low conductivity of the skull spatially blurs these signals, and this means that reconstructing the spatial locations within the brain where these currents originated is a particular challenge [Aydin et al. 2015]. Magnetoencephalography (MEG) non-invasively records the magnetic fields induced by these neural currents, and crucially, these can be spatially resolved—albeit at a relatively low spatial resolution relative to fMRI. This can be used to investigate the neural basis of fMRI resting-state networks [Brookes et al. 2011], and has shown that these patterns of functional connectivity are present at surprisingly high frequencies [Baker et al. 2014].

Finally, there are several invasive techniques that provide detailed recordings of neural activity from electrodes in direct contact with the brain, though there is necessarily relatively little data available from humans. However, when implanted for medical reasons they generate a prodigious amount of high quality data, and like MEG, this type of data is a window onto the high-frequency neural oscillations that give rise to the BOLD signal [Hacker et al. 2015].

1.2.3 Multi-modal approaches

Taken together, all these modalities offer a fantastically diverse description of brain structure and function, at a range of temporal and spatial scales. The challenge for the field is to work out how best to assimilate the different facets of neural organisation that the different modalities are sensitive to. As mentioned previously, detailed biophysical models are showing genuine promise here, as they provide a principled framework within which it is possible to incorporate the full spectrum of information from these different modalities. It seems highly probable that investigations along these lines will foster new models of the brain's architecture and dynamics in the near future.

1.3 Thesis Overview

In the next chapter, we formally define functional modes and review existing methods for identifying these from resting-state fMRI data.

In Chapter 3 we describe our probabilistic model for functional modes, and in Chapter 4 we describe how we make inferences on this.

Chapter 5 gives an overview of the properties of the functional modes we identify. In it, we both describe our attempts to validate these modes, and offer some interpretations of the novel properties that these have.

In Chapter 6 we go beyond simply examining the modes themselves, and describe two ways in which we have used our model to explore the more general properties of resting-state data.

Finally, in Chapter 7, we offer up some conclusions and suggest several potential lines of enquiry for future work.

Chapter 2

Literature Review

As should hopefully be clear from the Introduction, the resting state has been a remarkably protean construct, already encapsulating three or four differentiable concepts throughout its short life to date. Consequently, the field is an etymological minefield. With that in mind, we explicitly define a few key terms before progressing any further.

2.1 Definitions

2.1.1 Network

Firstly, we maintain the formal mathematical definition of a *network* as a collection of *nodes* and *edges*, and keep this as an abstract concept, distinct from any neuroscientific terminology. Furthermore, we will assume that the terms network and graph are synonymous: the current phraseology tends to refer to graph theory as a tool for analysing the brain network. This allows us to describe functional connectomics as the analysis of the whole brain network, where edges are defined in terms the functional connectivity between regions. But how do we define the nodes of this network?

The beauty of maintaining an abstract definition of a network is that we can use almost anything for network nodes, providing it is possible to sensibly define the connectivity between them. In this thesis, we are concerned with the analysis of resting-state fMRI (rfMRI) data so we are looking to identify a set of 'functional subunits', a term that we are leaving deliberately vague. It turns out that in order to construct a functional connectome, we only need to stipulate that it must be possible to define a single representative time course for each functional subunit. Then, if we take the subunits as the network nodes, the edges can be derived by computing any measure of functional connectivity between each pair of time courses. Historically, there have been two major ways of defining these subunits and we will give explicit definitions for each of these in turn.

2.1.2 Mode

We define a *mode* as any spatial distribution over the brain that shares a common time course. The lack of restrictions means that multiple modes can be highly overlapping, and individual modes can include anti-correlated regions (meaning that some regions within the mode have a negative spatial weight and others have a positive one). This is a deliberately broad definition, as it is designed to encompass the more traditional concept of a 'resting-state network'. From now on, it can be assumed that references to modes in a historic context are referring what would then have been termed resting-state networks.

As we outlined in the Introduction, the concept of modes naturally arose from attempts to understand rfMRI data in its own right. As it became more generally accepted that these were neuronal in origin, the idea that modes 'reflect functional systems supporting core perceptual and cognitive processes' [D. M. Cole et al. 2010] slowly coalesced. This means that modes lie somewhere between functional segregation and integration, neither reflecting the local processing

units nor capturing the full extent of whole-brain integration. Rather, they reflect collections of functional subunits that seem to be preferentially yoked together for broad, though not completely distinct, categories of cognitive processes.

2.1.3 Parcel

The recent interest in using rfMRI to derive cortical parcellations, as a distinct line of work from characterising modes, means that we will also define a parcel. Attempts to subdivide the cortex predate the emergence of the resting state, and historically, 'cortical areas have ... been identified using information from one or more of four general approaches: architectonics, topographic organization, functional specificity, and connectivity'[1] [Van Essen and Glasser 2014]. However, the multi-modal approach above relies on procedures that tend to be painstaking or invasive, and it is only recently that neuroimaging techniques have matured enough to allow serious attempts to parcellate the whole-brain in vivo.

rfMRI-derived parcellations have emerged as they purport to allow subdivision of the whole brain, with a relatively simple data collection procedure. Parcels derived in this manner often represent a 'hard' parcellation of grey matter into multiple non-overlapping, contiguous regions [Rubinov and Sporns 2010; Craddock et al. 2012; Gordon et al. 2014]. However, we relax this definition slightly to allow both blurred boundaries and bilaterally paired regions[2] [Yeo et al. 2011]. Therefore, in the spatial domain, a *parcel* is defined as a set of probabilities, with the spatial weights representing the probability that a given location is a member of that parcel. The expectation is that, within a parcel, these spatial weights will be arranged as a limited number of contiguous regions, with minimal overlap between different parcels. As an aside, note that a mode—as an extended spatial distribution having common temporal dynamics—can be defined either in terms of a spatial voxelwise

[1] This definition is inherently multi-modal, but not necessarily consistent across all modalities.
[2] Or indeed, cortico-cerebellar and cortico-subcortical groupings.

map, or as a weighted set of spatial parcels.

Clearly, functionally homogeneous parcels are a useful simplification of rfMRI data, but it is not clear how to reconcile these purely rfMRI-based parcels with those derived from the more holistic multi-modal definition given above. It is well known that cortical areas are not necessarily functionally homogeneous: for example, primary visual cortex consists of a set of regions, each containing a complete map of the visual hemifield and having clear architectonic boundaries, that form a visual processing hierarchy. However, functional connectivity is strongly modulated by position in the visual field, to the extent that it is often observed that the coupling across different levels of the hierarchy is stronger than the connectivity within a region [Buckner and Yeo 2014], and as such, rfMRI-based schemes tend to partition visual cortex differently to architectonics. It will be interesting to see whether new techniques for deriving rfMRI parcellations can resolve these differences, or whether more general notions of functional connectivity get assimilated into the more traditional multimodal framework for delineating cortical regions.

Summary

The above definitions have been motivated by a desire to separate the concept of a network from any neuroscientific constructs, as it is important to have a clear framework within which it is possible to unambiguously discuss these complex notions of brain connectivity. As stated earlier, it is possible to use either parcels or modes as the nodes to examine in a network analysis, and while a parcel is just a mode with additional spatial restrictions, maintaining a semantic distinction between them is important nevertheless. Finally, it is worth reiterating that both modes and parcels are also interesting in their own right—that is after all the point of this thesis—and that they have already yielded important insights into brain function without recourse to network analyses.

2.2 Modes or Parcels?

In this thesis, we aim to come up with a new method for characterising rfMRI data. Clearly, the first decision to make is whether to identify modes or parcels. Our choice is modes.

While parcels are a useful description of the data, we have several concerns about exclusively rfMRI-based schemes. For example, at the scale at which parcellations are consistent with existing descriptions of cortical areas, there is genuine concern about the ability to guarantee functional consistency across subjects with rfMRI data. However, simply working with larger parcels does not assuage our concerns, as there is reason to believe that large-scale parcels represent an oversimplification of the variable strengths of inter- and intra-regional functional connectivity. Our view is that parcels are currently best defined and identified in a multi-modal framework, and that, for now, modes still represent a useful way of investigating rfMRI data in its own right.

2.3 Traditional Analysis Techniques

In this section, we will review some of the currently used techniques for decomposing rfMRI data into modes. Our main focus is on independent component analysis (ICA) and related approaches, as they have been most successful historically.

Seed-based analyses

The original functional connectivity paper by Biswal et al. [1995] used a seed-based analysis. While this approach has largely been superseded, and technically neither identifies modes nor parcels, we cover it here as it introduces some important concepts. The technique itself is remarkably simple: a time course is extracted from a seed region, and the whole-brain spatial pattern of functional connectivity

is extracted by examining the correlation between the seed time course and the time course of each voxel in turn. This raises two points which we wish to expound upon.

Firstly, this demonstrates that for fMRI data, zero-lag correlation is an adequate description of functional connectivity. It is well known that zero-lag connectivity does not capture the full complexity of neural dynamics [Canolty and Knight 2010], and that the delay induced by the HRF exhibits variability across brain regions [Handwerker et al. 2004]. However, the strong temporal blurring of neuronal responses by the hæmodynamics dominates these temporal misalignments, to the extent that it is only just becoming possible to reliably investigate anything other than zero-lag connectivity [Mitra et al. 2015].

Secondly, if seed-based techniques allow examination of functional connectivity at the voxel level, then the obvious question is why not perform all analyses at that spatial resolution? One of the main reasons is practical: if we can identify modes or parcels that are genuinely neuronal in origin, then they give a very parsimonious representation of the data—thousand-fold reductions in the number of entities to examine are not uncommon. Furthermore, '[a]veraging data within predefined areas … improve[s] signal-to-noise and reduce[s] multiple comparison problems in statistical testing' [Gordon et al. 2014]. Similarly, this approach can reduce biases caused by large differences between the areal size of regions, though it is possible that a distinct set of biases will be introduced by estimating time courses from different amounts of data [Wig et al. 2011]. Finally, enforcing functional consistency across subjects with registration is a challenge at the level of regions, and is therefore practically impossible at the voxel level.

2.3.1 Matrix factorisations and the outer-product model

Most algorithms for identifying modes are based on matrix factorisation models. The idea is to decompose the V voxels by T time points data matrix, $\boldsymbol{D} \in \mathbb{R}^{V \times T}$, into the product of two matrices, $\boldsymbol{P} \in \mathbb{R}^{V \times M}$ and $\boldsymbol{A} \in \mathbb{R}^{M \times T}$ (Equation 2.1). We can think of $\boldsymbol{P}$ and $\boldsymbol{A}$ as describing a set of M modes: for mode m, each column of $\boldsymbol{P}$ represents a spatial map, $\boldsymbol{P}_m \in \mathbb{R}^{V \times 1}$, while each row of $\boldsymbol{A}$ represents a time course, $\boldsymbol{A}_m \in \mathbb{R}^{1 \times T}$. The spatial map is a set of weights, encoding the strength of each voxel's involvement in the mode. This tends to look something like the maps shown in Figure 1.2. The time course simply captures the temporally varying level of activity of the mode over the scan.

$$\begin{aligned} \boldsymbol{D} &= \boldsymbol{P}\,\boldsymbol{A} \\ &= \sum_{m=1}^{M} \boldsymbol{P}_m\,\boldsymbol{A}_m \end{aligned} \tag{2.1}$$

This is often referred to as the outer-product model in the fMRI literature[3], as each mode's contribution to the data can be expressed as the vector outer-product of the mode's spatial map and time course. It should be clear that this outer-product interpretation satisfies the definition of a mode given earlier. However, by itself this model is not identifiable. Additional constraints need to be introduced in order for a unique solution to exist. Several families of algorithms have been proposed that do exactly this, each with a different way of constraining and then solving this model.

[3] In the signal processing literature this is known as a blind source separation problem, as neither $\boldsymbol{P}$ nor $\boldsymbol{A}$ are known *a priori* [Beckmann 2012].

Interactions between modes

Before describing specific solutions of this model, we will clarify some terminology. We have so far talked in vague terms about the interactions between modes, but the outer-product model formalism now allows us to be more specific.

In the spatial domain we will use two terms in particular: 'correlation' and overlap. For the former, we quantify the relationship between mode spatial maps using $\hat{\rho}$, and this takes the form

$$\hat{\rho}_{i,j} = \frac{\boldsymbol{P}_i^{\mathrm{T}} \boldsymbol{P}_j}{\sqrt{\boldsymbol{P}_i^{\mathrm{T}} \boldsymbol{P}_i} \sqrt{\boldsymbol{P}_j^{\mathrm{T}} \boldsymbol{P}_j}} \tag{2.2}$$

As this resembles the equation for the Pearson product-moment correlation coefficient, from now on we will refer to modes for which $\hat{\rho}_{i,j} \neq 0$ as being 'correlated', though this does not represent a correlation in the technical sense—rather, if each spatial map is considered as a vector in a high-dimensional space then $\hat{\rho}_{i,j}$ represents the angle between any two spatial map vectors. Note that under this definition if two modes are orthogonal they are also uncorrelated. This is relevant as some solutions to the matrix factorisation model identify components that are orthogonal by construction.

Overlap is a term which describes the extent to which modes have non-zero spatial weights in the same locations. If we denote this as $\bar{\rho}$, and $\mathbb{1}_0(x)$ represents the indicator function for x being equal to zero, then

$$\bar{\rho}_{i,j} = \frac{1}{V} \sum_{v=1}^{V} \Big(1 - \mathbb{1}_0\big(\boldsymbol{P}_{vi}\big)\Big)\Big(1 - \mathbb{1}_0\big(\boldsymbol{P}_{vj}\big)\Big) \tag{2.3}$$

However, the modes inferred by most commonly used techniques contain lots of very small values that are not exactly equal to zero, so a simple thresholding step would need to be applied for this measure to be meaningful. Therefore, we will

tend to use overlap in a qualitative sense, whereby if we describe two modes as overlapping then we mean that they both contain high spatial weights in some brain regions.

In the temporal domain we characterise the interactions between modes based on both their correlations and partial correlations. Temporal correlation is defined in an analogous manner to spatial correlation, and the partial correlation between a pair of modes can be thought of as representing the correlation between the two time series after accounting for the influence of any other inferred modes. If we have a set of M modes, and are interested in the partial correlation between modes i and j, then we first have to 'regress out' the time series relating to the other $M-2$ modes, $\hat{\boldsymbol{A}}_{\neg ij}$. Formally, if we denote the time series after this regression as $\hat{\boldsymbol{A}}_i$ and the (right) pseudoinverse of $\boldsymbol{A}$ as $\boldsymbol{A}^+$, then this can be calculated as

$$\hat{\boldsymbol{A}}_i = \boldsymbol{A}_i - \left(\boldsymbol{A}_i \boldsymbol{A}^+_{\neg ij}\right)\boldsymbol{A}_{\neg ij} \ . \tag{2.4}$$

The partial correlation is simply the correlation between $\hat{\boldsymbol{A}}_i$ and $\hat{\boldsymbol{A}}_j$. Partial correlation is frequently used for rfMRI as it less sensitive to indirect connections than full correlation [Smith et al. 2011].

Matrix factorisation models for fMRI data

Many generic matrix factorisation algorithms have been tried on fMRI data over the years. Principal component analysis (PCA) of functional neuroimaging data predates even fMRI [Moeller and Strother 1991], but was almost immediately supplanted when ICA was introduced. More recent attempts include non-negative matrix factorisation, which solves the problem by introducing a set of positivity constraints [Lee et al. 2011], while generalisations of PCA allow fairly broad classes of sparsity or smoothness regularisation terms [G. I. Allen et al. 2014]. However, none of these have been particularly successful, and ICA has been the dominant

approach for the past decade.

2.3.2 Independent component analysis

When operating on rfMRI data, ICA is used to identify a set of modes that are either spatially or temporally independent. The ICA algorithm is based on a fundamentally simple principle, arising from the central limit theorem. If the conditions of the central limit theorem hold, then it can be shown that the distribution of any linear mixture of non-Gaussian random variables will be more Gaussian than the distributions that make up the mixture. If we assume that we are looking for non-Gaussian sources—and there are many classes of problems where, for example, sparsity is a desirable property—then the above property of mixtures of non-Gaussian random variables suggests a natural algorithm for recovering the original sources: we simply find the set of components that are maximally non-Gaussian under a certain metric [Comon et al. 1991; Jutten and Herault 1991; Comon 1994; Hyvärinen and Oja 2000; Hyvärinen et al. 2001b].

In practice, the generality of the above description means that there are a variety of ways of calculating an ICA decomposition, though for fMRI only FastICA [Hyvärinen 1999] and Infomax [Bell and Sejnowski 1995] are widely used [Correa et al. 2007].

Group analyses

Of course, we usually wish to analyse groups of subjects with rfMRI, but as it is classically formulated ICA works on a set of equivalent observations. A variety of ways to extend ICA to groups of subjects have been proposed [Calhoun et al. 2001; Lukic et al. 2002; Svensén et al. 2002; Schmithorst and Holland 2004; Beckmann et al. 2005; Esposito et al. 2005; Damoiseaux et al. 2006; Guo and Pagnoni 2008; Varoquaux et al. 2010; Erhardt et al. 2011], though only two are widely used.

The first is particularly simple: as registration has brought all subjects into

a common space, the data from all subjects is simply temporally concatenated. It is standard to look for a small number of modes, relative to the number of voxels or time points, so most ICA approaches use PCA as an initial dimensionality reduction step. Therefore, temporal concatenation is equivalent to 'projecting each dataset onto a common PCA eigenbasis' before running ICA [Beckmann et al. 2005].

The second approach is to decompose each subject's data using PCA, and then combine the individual subspaces, usually with another PCA projection, before running ICA [Calhoun et al. 2001; Varoquaux et al. 2010].

While each approach has its advantages and disadvantages [Smith et al. 2014], in practice, they often recover similar group-level modes [Erhardt et al. 2011].

Spatial ICA

The only further decision is whether ICA should look for modes with independent spatial maps or independent time courses. Due to the large numbers of voxels and relatively few time points of most imaging studies, and because the blurring action of the HRF reduces the relative strength of temporal non-Gaussianities, spatial ICA (sICA) tends to give the most robust decompositions and therefore became the dominant approach. If a relatively small number of components are inferred, these represent the classic set of spatially-distributed modes [Beckmann et al. 2005].

While the identification of these modes was a pivotal moment in the history of the resting state, concerns about sICA were raised almost as soon as it was introduced. Given that '[distinct] large-scale neuronal dynamics can share a substantial anatomical infrastructure' [Friston 1998; Smith et al. 2012], it is unclear how well sICA will decompose extended modes that are spatially correlated.

These concerns were allayed to some extent by Beckmann et al. [2005], who showed that, in the presence of noise, ICA components can still contain strong residual dependencies. They demonstrated that, even if the ICA components

were orthogonal and therefore uncorrelated by construction, strongly overlapping spatial maps can be recovered by a simple thresholding step. Intuitively, the sub-threshold mode weights can be arranged such that the inferred modes become uncorrelated without affecting the gross spatial pattern. What is perhaps less clear is what, if any, biases are introduced when the inferred spatial maps have very low noise, as this will limit the ability of ICA to counteract the effects of strongly correlated underlying modes with sub-threshold spatial weights. This situation will arise when the data has a high signal-to-noise ratio or when large groups are analysed, and these are both cases that are becoming increasingly common.

As touched upon in the Introduction, the current trend is to use components from a high-dimensional sICA for connectivity analyses [Kiviniemi et al. 2009; Smith et al. 2013b; E. A. Allen et al. 2014] as these tend to represent distinct, spatially compact functional subunits, and this further allays some of the initial concerns about overlap. This is often referred to as a type of parcellation, however, under our definitions, this still results in a set of modes[4].

Temporal ICA

The obvious alternative to sICA is to look for temporally independent modes. This has only recently become possible, as the aforementioned blurring action of the HRF means that temporal ICA (tICA) requires a large number of time points to operate robustly. The current spate of studies of large cohorts acquired at high sampling rates has therefore been the driving force behind making this a possibility [Smith et al. 2012].

The tICA approach allows spatially overlapping modes to be identified, at the expense of placing restrictions on the global temporal dynamics; as well as this

[4]Note how the very definition of parcels precludes their discovery with spatial ICA: in a given voxel, a high probability for membership of one parcel necessitates low probabilities for the others, which clearly violates the independence assumption. Practically, even high-dimensional components can still contain anti-correlations [Smith et al. 2013b] and this approach is not guaranteed to give full coverage of the cortex in a readily interpretable manner.

being a concern in and of itself, this restriction will also limit any subsequent network analyses of the mode time courses. As Smith et al. [2012] discuss, temporally independent functional modes (TFMs) are forced to have orthogonal timecourses, meaning that further analysis of the temporal interactions between different modes is not straightforward.

Summary

While each ICA strategy has its own advantages, the fundamental issue with all ICA-based approaches is that 'it is not clear that, from a neuroscientific point of view, independence is the right concept to isolate brain networks, as no functional system is fully segregated' [Varoquaux et al. 2010]. What is perhaps surprising is the empirical observation that ICA approaches work well, even when their central assumptions are violated [Hyvärinen 2013]; indeed, forms of ICA have been developed that explicitly incorporate information derived from the residual statistical dependencies between components, illustrating that an inability to strictly enforce independence is not too deleterious [Hyvärinen and Hoyer 2000; Hyvärinen et al. 2001a]. However, while the ICA approach has been successful in practice, moving away from a set of assumptions which are fundamentally irreconcilable with our understanding of modes as a set of entities with complex spatio-temporal interactions is the obvious next step for new methods.

2.3.3 Extraction of subject-specific information

The majority of rfMRI analyses make their inferences from a group of subjects, as this gives the analyses a huge amount of power, and, in general, this also makes it easier to subsequently compare the modelling results across subjects. Mode extraction techniques are, for the most part, no different, and the majority of ICA approaches result in a single set of group-level mode maps. In order to investigate individual subjects, one could simply extract a set of time courses from the subject-

level rfMRI data using the group-level modes, and analyse these. This would, for example, allow the construction of subject-level functional connectomes.

However, a weakness of this approach is that it completely relies on the registration process to enforce consistency across subjects. As we have mentioned previously, structural registration does not guarantee functional correspondence, and as '[i]nconsistent or imprecise node definitions can have a major impact on subsequent analyses' [Fornito et al. 2013] this is an approach which can end up conflating several different types of variation.

It seems reasonable to assume that the enormous amount of residual information at the subject level could be harnessed to inform the specification of functional regions. With that in mind, there are a number of algorithms that take group-level results and try to extract more meaningful information at the subject level.

Dual regression

Perhaps the conceptually simplest method for extracting subject-specific information is dual regression [Beckmann et al. 2009]. The group-level mode spatial maps are regressed into each subject's data, resulting in a set of subject-specific time courses. The second step then regresses those time courses back into the data, and this now results in a set of subject-specific spatial maps that hopefully more closely resemble the 'true' modes[5].

Several variants of this approach have been proposed, many in the context of ICA specifically, but they often required projection via specific subspaces [Calhoun et al. 2001]. While there have been some suggestions that these alternatives give better performance [Erhardt et al. 2011], a major advantage of dual regression is

[5]One can of course continue this argument *ad infinitum*. If the subject-level spatial maps identified by dual regression are indeed a better description of the modes in each individual than the group maps, as they purport to be, then it seems reasonable to assume that the time courses obtained by regressing the subject maps into the data are more accurate than those obtained by regressing the group maps into the data. That new set of time courses would then be ideal for identifying a new set of spatial maps, and so on... The stability of this iterative procedure, especially in the presence of noise, does not appear to have been systematically investigated.

its simplicity. It can be applied to any template set of modes, on any data set, and it does appear to be sensitive to subject variability [Filippini et al. 2009; Zuo et al. 2010].

Group priors

An alternative approach is to use the group results as a prior on the subject-level decompositions. In theory, this allows a slightly more flexible trade-off between group and subject information. For example, Dhillon et al. [2014] and Du and Fan [2013] proposed variants of PCA and ICA respectively that were regularised by pre-specified group information. Despite a certain degree of heuristic tuning, this is a principled way to introduce group information and has the nice property that the subject information can be derived under the same, for example independence, constraints as the group maps.

Supervised Classification

Finally, Hacker et al. [2013] developed the idea of recasting the extraction of subject information as a supervised learning problem. They trained a multi-layer perceptron (MLP) to associate seed-based correlation maps with a set of modes. Once trained, the perceptron is then able to classify voxels at the individual subject level solely based on their functional connectivity maps.

This results in a more spatially specific classification of voxels than dual regression, which they ascribe to 'high-dimensional non-linear classification boundaries allow[ing] the MLP to extract arbitrary features from a large input space' [Hacker et al. 2013]. However, compared to dual regression there does need to be an extensive pre-processing pipeline and training procedure. Additionally, it is not guaranteed that all parcels identified at the group level will be found in all subjects, although that may be an accurate reflection of reality.

Summary

There are now several plausible techniques for post hoc extraction of subject information from group-level modes, and they all seem able to detect neuroscientifically interesting differences between subjects. For now, dual regression seems to be the most widely used, but as large, consistently pre-processed data sets become available the 'per-unit' costs of training more complex procedures will be reduced dramatically, potentially hastening their adoption. In that sense, the work of Hacker et al. [2013] is particularly interesting as it opens up the problem to the wealth of existing literature on supervised learning.

However, these post hoc approaches mean that the flow of information is almost exclusively from group to subject, and this is potentially a major limitation of current techniques. Intuitively, if we can extract genuinely meaningful subject information then we should be able to utilise this to generate richer descriptions of modes that, for example, encode typical patterns of variability. One would hope that there is a virtuous cycle here: as well as being interesting in and of itself, one suspects that more detailed group-level descriptions could also be used to infer subject information in a more principled manner. Again, the potentially mutually beneficial relationship between group- and subject-level descriptions is an obvious line of investigation for new mode identification techniques, especially as such large amounts of data are being made publicly available.

2.4 New Approaches

The outer-product model—and ICA solutions of this in particular—has allowed the identification of a now well characterised set of modes from multiple independent data-sets, populations and modalities. At this stage, the field is just starting to explore approaches that advance our understanding of modes by moving beyond

these traditional techniques, and we cover two of these here.

2.4.1 Multi-subject dictionary learning

An obvious way of improving techniques based upon the outer-product model is to more overtly base the constraints, that are necessary to make the model identifiable, on the properties of rfMRI data. The multi-subject dictionary learning (MSDL) approach of Varoquaux et al. [2011] is perhaps the best example of a model that specifically looks for modes, rather than seeking to solve the problem in a general sense. Explicitly motivated by some of the properties of modes discovered with ICA and related approaches, their algorithm contains a hierarchical model for spatial subject variability, a constraint favouring simultaneously smooth and sparse spatial distributions as well as the ability to capture the temporal correlations between modes.

Here, we give a brief description of the most recent version of their model, as described by Abraham et al. [2013]. Their spatial model at the group level is detailed, simultaneously enforcing non-negativity, sparsity and spatial contiguity. The subject maps are modelled by including a set of additive, Gaussian-distributed deviations from the group maps. Their time series model specifies that there should be a consistent between-mode correlation structure, but does not restrict the form of the time series; therefore, it does not model any hæmodynamic processes. Finally, these constraints are combined with a noise model, and the resulting cost-function governing their decomposition is solved with a computationally efficient stochastic gradient descent approach.

This represents a laudable attempt to capture more of the properties of modes within a consistent modelling framework, and produces interesting decompositions. However, this approach places a huge amount of emphasis on the properties of the group-level mode maps, while the spatial subject variability and temporal model-

ling is slightly rudimentary by comparison. In essence, the model acknowledges that individual subjects have their own spatial maps, but it does not parameterise the observed variability in any way[6]. Therefore it only completes half the cycle we referred to earlier: subject maps are inferred, and these improve estimation at the group level; however, the group level does not contain any parameters beyond the mean maps that can be used to constrain the subject maps.

2.4.2 Beyond the outer-product model

Finally, there have been some tentative attempts to move beyond the simple, linear description of modes given by the outer product model. It is to be expected that more complex models of rfMRI data, like the two outlined below, will come under increasing scrutiny in the near future.

Hjelm et al. [2014] used a restricted Boltzmann machine to identify a mapping between the full rfMRI data and a set of mode time courses, and while the simple approach they used remains similar to the outer-product model, their work raises the intriguing prospect of using this approach as a building block in much more complex models [Hinton and Salakhutdinov 2006; Plis et al. 2014].

Eavani et al. [2013] used a hidden Markov model to decompose rfMRI data[7] into a sequence of temporally distinct states, each with its own spatial covariance structure. While it takes a certain amount of mathematical gymnastics to express their results as a set of modes, there is a strong conceptual link, and this formulation offers an interesting counterpoint to more traditional descriptions of brain activity.

However, something we have only mentioned in passing, but is coming under increasing scrutiny, is the computational efficiency of different algorithms. The flurry of large-scale data-collection initiatives means that it is now expected that

[6] In fact, the parameter that specifies the variance of the subject maps around the group average is specified *a priori*.

[7] Interestingly, this approach had been used on task data much earlier, and with a more sophisticated model [Højen-Sørensen et al. 1999].

any new technique will be able to simultaneously analyse thousands of subjects. Note that this is what necessitated the restrictions on the model Hjelm et al. [2014] presented, and they were only analysing a relatively small number of subjects. Therefore, it is perhaps this restriction which will have the strongest influence on the adoption of these more complex techniques in the short term.

Summary

Both of the approaches mentioned above show promise, but are also necessarily inchoate. This is therefore an area where there is the potential for new techniques to rapidly further our ability to capture, and hopefully understand, modes in the near future.

In the short term, we, like Varoquaux et al. [2011], suspect that we are better served explicitly modelling the properties of rfMRI data than trying to move beyond linearity. In part this reflects the pragmatic stance that, given the enormous data sets on the horizon, it makes sense to utilise techniques that have already demonstrated their computational suitability for rfMRI data. However, in part this also reflects a more principled standpoint, in that we feel we are best served by building models that are explicitly based around what we know. This is the approach we outline in the following chapter.

Chapter 3

Modelling Approach

In general terms, our aim is to identify extended functional modes which are not restricted to being orthogonal, or non-overlapping, to each other in space or in time. At the mode level, we make use of the outer-product model, but explicitly base our constraints on the properties of rfMRI data.

In short, we define a hierarchical model that allows us to flexibly capture the spatial variation of modes across subjects, while still keeping track of key properties at the group level. Simultaneously, we enforce that the temporal characteristics of modes must relate to the hæmodynamics that drive the BOLD signal, and we explicitly model a hierarchical set of inter-modal correlations.

As the modes we infer are defined by the generative model outlined below, we will refer to them as *probabilistic functional modes* (PFMs).

3.1 Probabilistic Model

Our model is built on the same matrix factorisation approach that underpins PCA, ICA, non-negative matrix factorisation and several other of the well established methods for extracting modes of rfMRI. For completeness, we reintroduce the notation for these models we introduced in Section 2.3.1, before introducing our

extensions.

The fMRI data are acquired in V voxels and at T time points, giving a data matrix $\boldsymbol{D} \in \mathbb{R}^{V \times T}$ for a single run. If we infer a set of M PFMs then we look for spatial maps, $\boldsymbol{P} \in \mathbb{R}^{V \times M}$, and time courses $\boldsymbol{A} \in \mathbb{R}^{M \times T}$. In general we infer a small number of PFMs relative to V and T, which gives a parsimonious set of modes. However, this means that the factorisation will not be exact, so we express the data as the contribution from the PFMs and a noise term, $\boldsymbol{\varepsilon}$, to give

$$\boldsymbol{D} = \boldsymbol{P}\,\boldsymbol{A} + \boldsymbol{\varepsilon}\ . \tag{3.1}$$

However, we rarely wish to model a single run in isolation. In order to infer both the group-level properties and any interesting subject variability, we explicitly account for the full set of all runs, $\mathscr{D}$. This contains data from each subject within the set of subjects S, and furthermore, we allow for the possibility that each subject, s, will have been scanned multiple times, resulting in a set of runs R_s.

$$\mathscr{D} = \left\{ \left\{ \boldsymbol{D}^{(sr)} \right\}_{r \in R_s} \right\}_{s \in S} \tag{3.2}$$

The problem we are faced with is defining an extension to the standard matrix factorisation approach (Equation 3.1) to account for these multiple data. In essence, this forces us to more formally define our model for subject variability. Noise processes and limited amounts of data per subject will naturally cause the observation of spurious variation, but there are several genuine sources of variability that are potentially interesting neuroscientfically, and these are sketched in outline below.

In the spatial domain, we expect residual variability in the locations of functional regions, even after registration [Mueller et al. 2013]. Loosely speaking, misalignments can arise for three reasons: there will naturally be some errors in the registration process, resulting in structural features that are not brought into

correspondence; there will be locations where anatomical landmarks bear little relation to functional subdivisions, meaning structural similarity is not a sufficient condition for accurate registration; and there may be genuine topological reorganisations, whereby the standard registration approaches based on diffeomorphic warps could never succeed[1]. If these misalignments are not accounted for, then one expects the inferred mode time courses to be a farrago of contributions from the underlying 'true' set of modes.

In the temporal domain, functional connectomics is predicated on the idea that the interactions between modes encode interesting information about subjects, so we would like to capture that. However, the unconstrained nature of rfMRI data means that we can say very little about the time courses from a given run as there is no temporally consistent pattern of mode activation.

Therefore, we model subject and run variability as follows. Each run will have its own unique set of time courses, $\boldsymbol{A}^{(sr)}$, and a random realisation of the noise process, $\boldsymbol{\epsilon}^{(sr)}$. However, we assume that the spatial variability we observe across subjects, by virtue of it being driven primarily by cortical reorganisations, is consistent across all runs for a given subject. This gives a set of subject-specific spatial maps, $\boldsymbol{P}^{(s)}$, that will potentially be observed multiple times. This set of assumptions extends Equation 3.1 to give

$$\boldsymbol{D}^{(sr)} = \boldsymbol{P}^{(s)}\boldsymbol{A}^{(sr)} + \boldsymbol{\epsilon}^{(sr)} \ . \tag{3.3}$$

The last thing we need to specify is how to link the individual subject decompositions. Maintaining the formulation of this as a probabilistic model, we place priors on both the PFM spatial maps and time courses, as well as modelling

[1]It is somewhat contentious whether (structural) registration should be held responsible for the latter two processes. Our definition of registration is somewhat broader, as we hold it responsible for bringing subjects into structural and functional correspondence. While structural registration is unlikely to be sufficient here, this is nevertheless a reasonable aim for multi-modal registration approaches.

the contribution from the noise. Crucially, we introduce hierarchical priors to link subjects through a consistent set of group properties, and the forms of these distributions are explained in the following sections. For convenience we provide a graphical model in Figure 3.2.

3.1.1 Spatial model

As stated earlier, we aim to identify spatially extended modes, but this is not in direct competition with the idea of functional segregation. For example, note how even an extended system like the default mode is confined to several sizeable, but nevertheless well-defined regions of cortex. To do this we define a model that encodes sparsity in the spatial weights, and introduce a hierarchy over subjects to account for the types of variability we discussed earlier.

Subject level

At the subject level, we wish to capture this notion of functional segregation with a sparsity prior, as well as maintaining a set of subject-specific mode maps that are consistent with the rest of the group. To express this probabilistically, we formulate a delta-Gaussian mixture model, a natural extension of the spike-slab distribution [Mitchell and Beauchamp 1988; George and McCulloch 1993; Ishwaran and Rao 2005; Titsias and Lázaro-Gredilla 2011]. This contains a delta component at zero to capture the weights that are either not present or too weak to be observed, and a Gaussian component to model the observable weights and their variability over subjects. This is parameterised by the probability of a weight being present, π, as well as the mean, μ, and standard deviation, σ, of the Gaussian. Each spatial weight follows this distribution, so this parameterisation succinctly captures our beliefs about the answers to three very pertinent questions: does a voxel contribute to a given PFM? If so, how big is the contribution and how much does it vary from subject to subject?

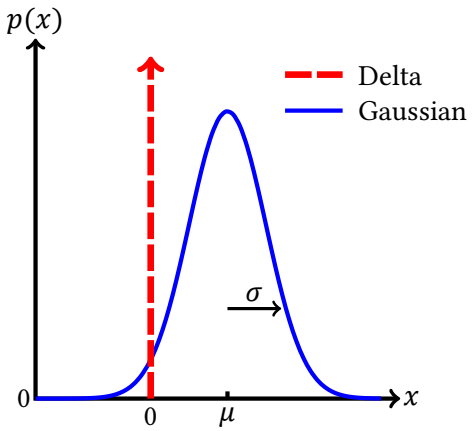


Figure 3.1: Diagrammatic representation of the probability density function for the delta-Gaussian mixture model. The relative contribution of each component is governed by the parameter π, while the Gaussian component is parameterised by its mean, μ, and standard deviation, σ.

To complete the formulation, we also need to include a binary indicator variable, q, to capture which component the weight is drawn from. Therefore, for voxel v in the spatial map of PFM m, this subject-level prior has the form given in Equation 3.4. This is also represented graphically in Figure 3.1.

$$\begin{aligned} \mathrm{p}\left(\boldsymbol{P}_{vm}^{(s)} \,\middle|\, q_{vm}^{(s)} = 1\right) &= \mathcal{N}\left(\boldsymbol{P}_{vm}^{(s)} \,\middle|\, \mu_{vm}, \sigma_{vm}^2\right) \\ \mathrm{p}\left(\boldsymbol{P}_{vm}^{(s)} \,\middle|\, q_{vm}^{(s)} = 0\right) &= \delta\left(\boldsymbol{P}_{vm}^{(s)}\right) \\ \mathrm{p}\left(q_{vm}^{(s)}\right) &= (\pi_{vm})^{q_{vm}^{(s)}} (1 - \pi_{vm})^{1 - q_{vm}^{(s)}} \end{aligned} \tag{3.4}$$

The fact that the non-zero weights are drawn from a Gaussian means that they can be either positive or negative, thereby allowing modes containing anti-correlated regions—if the data supports that inference. Similarly, we have placed no explicit prior on the relationship between the spatial distributions of different modes, so there is no explicit penalty on voxels being present in multiple modes. However, identifying several essentially identical modes is not a readily interpretable solution, so we do want some way to enforce that modes must be distinct.

In our case, this arises because there is an implicit penalty on model complexity via automatic relevance determination [Mackay 1995; 1994]. Intuitively, if two PFMs are almost identical then this is not a parsimonious way to represent the structured components present in the data, and one of the PFMs could be eliminated without reducing the ability of the model to represent the features of the data we are interested in.

Finally, note that there is a direct conceptual link between this approach and the traditional approach of significance testing within the general linear model framework for task fMRI data. For each spatial weight we want to assess whether, given the noise level and time course, the data supplies enough evidence to suggest that there is genuinely an effect present in that voxel. If there is insufficient evidence *a posteriori* for an effect we should just set the weight at that voxel to zero. However, if there is evidence for an effect, then we are interested in both its size and how it varies over subjects.

Group level

The subject-level model allows us to capture the notions of functional segregation and subject variability, but also enforces consistency across subjects. The parameters π, μ, and σ are shared between all subjects, so represent the description of the PFMs at the group level.

Note how much richer this description is than the single set of group-level means that most currently used techniques, like ICA, infer. For example, the σ parameters can capture the types of spatial non-uniformity in subject variability observed by Mueller et al. [2013]. Therefore, when inferring subject maps, the inference will automatically be informed by the data more than the group mean in regions inferred to exhibit high functional heterogeneity over subjects, and vice versa for regions with low subject-to-subject variability.

As is standard, we place a beta hyperprior on π and an inverse gamma hy-

perprior on σ^2. Finally, we place a spike-slab hyperprior on each mode's voxelwise means, again motivated by notions of functional segregation. This has precision γ and sparsity λ, as in Equation 3.5.

$$\begin{aligned} \mathrm{p}(\mu_{vm} \,|\, \rho_{vm} = 1) &= \mathcal{N}(\mu_{vm} \,|\, 0, \gamma^{-1}) \\ \mathrm{p}(\mu_{vm} \,|\, \rho_{vm} = 0) &= \delta(\mu_{vm}) \\ \mathrm{p}(\rho_{vm}) &= (\lambda)^{\rho_{vm}}(1-\lambda)^{1-\rho_{vm}} \end{aligned} \tag{3.5}$$

This is the same form as the prior used on $\boldsymbol{P}^{(s)}$ in Equation 3.4, with the exception that the mean of the Gaussian component is constrained to be zero.

The spike-slab hyperprior allows us to regularise the group spatial maps, if required, by altering the λ parameter. Intuitively, λ represents the proportion of voxels we expect to be non-zero *a priori* in each mode's group-level spatial map. In reality, as we expect modes to be spatially distributed and overlapping, we set $\lambda > M^{-1}$. This is simply another way of saying that we expect each voxel to be involved in more than one mode.

Clearly, there are several aspects of the spatial model that reflect trade-offs relating to computational concerns. The most obvious example of this is the independence over voxels in the spatial prior, thereby neglecting spatial structure, though the choice of spike-slab type spatial priors, which make the strong assumption of a delta function at zero, also relate to this. We know that modes tend to represent collections of spatially contiguous regions, and this is the type of structure that Varoquaux et al. [2011] try to capture with their model. For now, as fMRI data sets tend to contain large numbers of voxels, we neglect this in favour of a model that will result in a more computationally efficient inference procedure. However, imposing spatial structure—for example, it may be feasible to introduce interactions between spatially adjacent indicator variables in a manner analogous to an Ising model—is something that should much improve our model and warrants

future investigation.

Finally, note that when we present our results, the group maps we show are the marginal posterior means of the whole spatial distribution, $\mathrm{E}[\pi_{vm}\mu_{vm}|\mathscr{D}]$. We choose this quantity, rather than just the μ parameters, as this represents the expected value of $\boldsymbol{P}_{vm}^{(s)}$.

3.1.2 Temporal model

Primarily, we wish to formulate a model that captures hæmodynamic processes. However, we expect the inferred time courses to be corrupted by noise, even if we properly account for the noise process $\boldsymbol{\epsilon}^{(sr)}$. This is for two main reasons.

Firstly, as Bright and Murphy [2015] recently showed, even well-characterised modes can be identified from noise processes like subject motion. Conversely, this implies that even accurately identified functional modes may well correlate with non-neural processes that violate our hæmodynamic assumptions. Secondly, spatial smoothing is a standard pre-processing step for fMRI data. This ameliorates the problem of residual spatial mis-alignment after registration, but induces spatial correlations in the noise. While it would be possible to model this, estimating the true number of spatial degrees of freedom in the data is notoriously difficult [Worsley et al. 1996], and would be computationally expensive over a large number of voxels.

Therefore, we make the pragmatic decision to allow noisy time courses, and to reflect this our time course model contains two terms: the first represents the clean BOLD time courses, $\boldsymbol{B}^{(sr)}$, while the second represents the noise that corrupts these, $\boldsymbol{\xi}^{(sr)}$. This gives

$$\boldsymbol{A}^{(sr)} = \boldsymbol{B}^{(sr)} + \boldsymbol{\xi}^{(sr)} \,. \tag{3.6}$$

BOLD time courses

The spectral characteristics of the neurally-driven BOLD signal are dominated by the hæmodynamic response. In other words, we assume that any non-artefactual mode time courses we observe will be dominated by low frequencies, almost regardless of the frequency content of the underlying neuronal processes, because the HRF acts as such a strong low-pass filter.

Therefore, we formulate a temporal prior that captures the auto-correlation induced by the HRF. For a neuronal signal, $\mathrm{x}(t)$, and a linear HRF, $\mathrm{h}(t)$, the observed signal $\mathrm{y}(t)$ is a simple convolution of $\mathrm{x}(t)$ and $\mathrm{h}(t)$. If we assume that the neuronal process is white on the timescale of the fMRI acquisitions then it is straightforward to show that the auto-correlation induced in the observed signal is just the auto-correlation of the HRF, namely

$$\mathrm{E}\big[\mathrm{y}(t_1)\,\mathrm{y}(t_2)\big] = \sum_{\tau} \mathrm{h}(\tau)\,\mathrm{h}\big(\tau - (t_1 - t_2)\big)\ . \tag{3.7}$$

We assume a canonical double-gamma HRF [Friston et al. 2007] and use this correlation structure to construct a full covariance matrix, $\boldsymbol{K_B} \in \mathbb{R}^{T\times T}$, for all the time points in a given run.

We would also like to model the temporal interactions between modes, so we place a prior on the inter-modal precision matrix, $\boldsymbol{\alpha}^{(s)} \in \mathbb{R}^{M\times M}$. This is combined with the HRF-derived covariance structure in a matrix normal distribution. Therefore, the prior on the hæmodynamic time course for all the PFMs in given run becomes

$$\mathrm{p}\big(\boldsymbol{B}^{(sr)} \,\big|\, \boldsymbol{\alpha}^{(s)}\big) = \mathcal{MN}\big(\boldsymbol{B}^{(sr)} \,\big|\, \mathbf{0}, \boldsymbol{\alpha}^{(s)^{-1}}, \boldsymbol{K_B}\big)\ . \tag{3.8}$$

Of course, it is well known that the HRF is both highly variable and much more complex than the canonical linear HRF assumed here [Aguirre et al. 1998; Handwerker et al. 2004; Kriegeskorte et al. 2010]. However, this is a probabilistic

prior, rather than a hard constraint, so there is scope for the inferred time courses to match the temporal structure in the data. Similarly, this form of prior on the covariance structure is predicated on a white neuronal process. Some current theories, based on evidence from electrophysiology [Siegel et al. 2012] and biophysical models [Deco et al. 2011], predict that the neural basis of rfMRI activity is correlations between the envelopes of higher frequency signals, where these amplitude correlations occur at 'ultraslow' timescales (biased towards frequencies less than 0.1Hz). On the other hand, current evidence suggests that neuronal process are white over the range of frequencies estimable at typical fMRI sampling rates [Niazy et al. 2011] and that structure, consistent with the rfMRI literature, is present at surprisingly high frequencies [Baker et al. 2014].

Given the wide range of temporal models present in the literature, we deliberately choose a very simple, computationally efficient, linear model to capture the gross properties of the HRF. Implicitly, we are relying on the probabilistic inference procedure to be flexible enough to capture some of the more complex temporal properties. For example, this could relate to non-linearities in the hæmodynamic response or non-stationarities in the underlying neuronal activity.

Mode interactions

Our HRF-based prior on the time courses strongly predisposes the decomposition to identify modes that are neural in origin, rather than those which represent structured artefacts. However, the HRF covariance prior, considered in isolation, is rotationally invariant across modes. This limitation arises because of the assumption of a linear HRF operating on neural time courses that are independent and Gaussian. Intuitively, if the set of inferred time courses are all consistent with the frequency spectrum implied by our HRF and associated autocorrelation function, then any linear combination of the time courses would satisfy our temporal constraints equally well. In other words, satisfying the HRF constraint does not help

distinguish one mode from another.

However, as discussed earlier, we expect to observe temporal interactions between modes, and this will lend some structure to the mode time courses. We know that these interactions have been characterised as having a consistent structure across the group [Shehzad et al. 2009], so we introduce a hierarchical model to capture this. Subject variability will manifest itself as deviations from this set of group interactions.

Starting at the subject level, we estimate the subject-specific temporal precision matrix $\boldsymbol{\alpha}^{(s)}$ to keep track of the functional connectivity between modes. Our subject-level precision matrices follow a Wishart distribution, and we introduce a hyperparameter, $\boldsymbol{\beta} \in \mathbb{R}^{M \times M}$, to capture the interactions that are consistent across subjects. This takes the form of a hyperprior on the subject-specific scale matrices, and again this follows a Wishart distribution.

$$\mathrm{p}\left(\boldsymbol{\alpha}^{(s)} \,\middle|\, \boldsymbol{\beta}\right) = \mathcal{W}\left(\boldsymbol{\alpha}^{(s)} \,\middle|\, a_{\boldsymbol{\alpha}^{(s)}}, \boldsymbol{\beta}\right) \qquad \mathrm{p}\left(\boldsymbol{\beta}\right) = \mathcal{W}\left(\boldsymbol{\beta} \,\middle|\, a_{\boldsymbol{\beta}}, \boldsymbol{B}_{\boldsymbol{\beta}}\right) \tag{3.9}$$

As an aside, note that these interactions are calculated between the BOLD time courses after the auto-correlation induced by the HRF has been removed. It is straightforward to show that, dropping the subject and run indeces for simplicity, $\boldsymbol{\alpha} = \frac{1}{T}\,\mathrm{E}\left[\boldsymbol{B}\boldsymbol{K}_{\boldsymbol{B}}^{-1}\boldsymbol{B}\right]$. This is important as it is a weaker form of deconvolution. Our expectation is that if the BOLD time courses are inferred such that they are consistent with the temporal auto-correlation function that we use to construct $\boldsymbol{K}_{\boldsymbol{B}}$, then full deconvolution would be straightforward. This is a notoriously difficult process for rfMRI data in the presence of noise, but, as Gitelman et al. [2003] discuss, it can be important to take this into account when estimating neuronal interactions that vary over time[2].

[2]Of course, our model does not account for time-varying interactions. However, if we can infer time courses that can be deconvolved then this may increase the sensitivity of any post-hoc analyses that look for non-stationarities in the temporal domain.

Noise time courses

The noise time course of mode m at time t, $\xi_{mt}^{(sr)}$, is simply drawn from a Gaussian distribution with precision $\omega_m^{(sr)}$. This gives

$$\mathrm{p}\left(\xi_{mt}^{(sr)} \,\middle|\, \omega_m^{(sr)}\right) = \mathcal{N}\left(\xi_{mt}^{(sr)} \,\middle|\, 0, \omega_m^{(sr)^{-1}}\right) . \tag{3.10}$$

Each $\omega_m^{(sr)}$ takes a gamma hyperprior.

3.1.3 Noise model

The final part of the model to specify is the noise process in the outer product model in Equation 3.3, and hence the likelihood. This is simply zero-mean white Gaussian noise, with an overall precision for each run, $\psi^{(sr)}$, which takes a standard gamma hyperprior.

$$\begin{aligned} \mathrm{p}\left(\boldsymbol{\varepsilon}^{(sr)}\right) &= \mathcal{MN}\left(\boldsymbol{\varepsilon}^{(sr)} \,\middle|\, \mathbf{0}, (\psi^{(sr)})^{-1}\boldsymbol{I}_V, \boldsymbol{I}_T\right) \\ &= \mathrm{p}\left(\boldsymbol{D}^{(sr)} - \boldsymbol{P}^{(s)}\boldsymbol{A}^{(sr)}\right) \end{aligned} \tag{3.11}$$

We use this structure, rather than allowing voxel-specific precisions, as it allows us to exploit the properties of the matrix normal distribution, leading to very computationally efficient inference [Stegle et al. 2011]. In order to ensure that this is relatively accurate, we renormalise the data to ensure each voxel has unit variance before inferring the PFMs—a not uncommon pre-processing step for rfMRI data.

However, what is perhaps more problematic is that this model does not acknowledge the spatial smoothness of fMRI data. It is standard practice to spatially smooth fMRI data to reduce the impact of misalignments, but this means that the noise is not truly independent over voxels. It would be possible to model this, for example by inferring a full spatial covariance matrix for the noise that

acknowledged the dependencies between voxels that smoothing introduces. Again, we decide that the benefits of this more complex model are outweighed by the increased computational burden, and we discuss a way in which we can mitigate the effects of this model misspecification in the next chapter.

3.2 Equations

For completeness, we provide all the equations that define the model below.

Spatial model

$$\begin{aligned} \mathrm{p}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, q_{vm}^{(s)} = 1, \mu_{vm}, \sigma_{vm}\big) &= \mathcal{N}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, \mu_{vm}, \sigma_{vm}^2\big) \\ \mathrm{p}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, q_{vm}^{(s)} = 0\big) &= \delta\big(\boldsymbol{P}_{vm}^{(s)}\big) \\ \mathrm{p}\big(q_{vm}^{(s)}\big) &= (\pi_{vm})^{q_{vm}^{(s)}}(1-\pi_{vm})^{1-q_{vm}^{(s)}} \end{aligned} \tag{3.12}$$

$$\mathrm{p}\big(\pi_{vm}\big) = \beta\big(\pi_{vm} \,\big|\, a_{\pi_{vm}}, b_{\pi_{vm}}\big) \tag{3.13}$$

$$\mathrm{p}\big(\sigma_{vm}\big) = \Gamma\big(\sigma_{vm}^{-2} \,\big|\, a_{\sigma_{vm}}, b_{\sigma_{vm}}\big) \tag{3.14}$$

$$\begin{aligned} \mathrm{p}\big(\mu_{vm} \,\big|\, \rho_{vm} = 1\big) &= \mathcal{N}\big(\mu_{vm} \,\big|\, 0, \gamma^{-1}\big) \\ \mathrm{p}\big(\mu_{vm} \,\big|\, \rho_{vm} = 0\big) &= \delta\big(\mu_{vm}\big) \\ \mathrm{p}\big(\rho_{vm}\big) &= (\lambda)^{\rho_{vm}}(1-\lambda)^{1-\rho_{vm}} \end{aligned} \tag{3.15}$$

Temporal model

$$\boldsymbol{A}^{(sr)} = \boldsymbol{B}^{(sr)} + \boldsymbol{\xi}^{(sr)} \tag{3.16}$$

$$\mathrm{p}\big(\boldsymbol{B}^{(sr)} \,\big|\, \boldsymbol{\alpha}^{(s)}\big) = \mathcal{N}\big(\mathrm{vec}\big(\boldsymbol{B}^{(sr)}\big) \,\big|\, \boldsymbol{0}, {\boldsymbol{\alpha}^{(s)}}^{-1} \otimes \boldsymbol{K_B}\big) \tag{3.17}$$

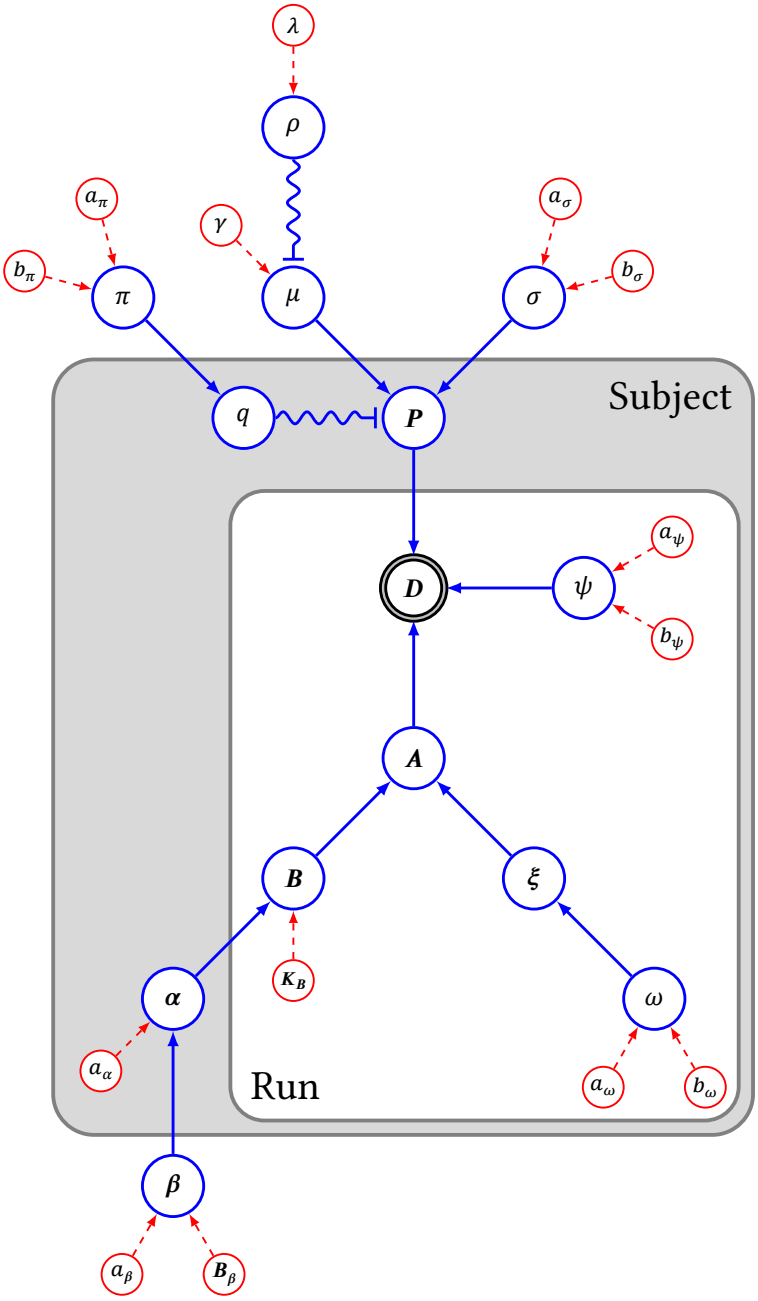


Figure 3.2: Graphical representation of the full model structure. Variables are shown in the blue circles and the dependencies are depicted as the arrows joining them, with switches conditioned on indicator variables shown as wiggly lines. Pre-specified prior parameters are shown in red. We have indicated the parameters that are inferred at the subject and run levels, though we have omitted the dimensions and numbers of variables for simplicity.

$$\mathrm{p}\left(\boldsymbol{\alpha}^{(s)} \,\middle|\, \boldsymbol{\beta}\right) = \mathcal{W}\left(\boldsymbol{\alpha}^{(s)} \,\middle|\, a_{\boldsymbol{\alpha}^{(s)}}, \boldsymbol{\beta}\right) \tag{3.18}$$

$$\mathrm{p}\left(\boldsymbol{\beta}\right) = \mathcal{W}\left(\boldsymbol{\beta} \,\middle|\, a_{\beta}, \boldsymbol{B}_{\beta}\right) \tag{3.19}$$

$$\mathrm{p}\left(\boldsymbol{\xi}_t^{(sr)} \,\middle|\, \boldsymbol{\omega}^{(sr)}\right) = \mathcal{N}\left(\boldsymbol{\xi}_t^{(sr)} \,\middle|\, \mathbf{0}, \boldsymbol{\omega}^{(sr)-1}\right) \tag{3.20}$$

$$\mathrm{p}\left(\omega_m^{(sr)}\right) = \Gamma\left(\omega_m^{(sr)} \,\middle|\, a_{\omega_m^{(sr)}}, b_{\omega_m^{(sr)}}\right) \tag{3.21}$$

Noise model

$$\mathrm{p}\left(\psi^{(sr)}\right) = \Gamma\left(\psi^{(sr)} \,\middle|\, a_{\psi^{(sr)}}, b_{\psi^{(sr)}}\right) \tag{3.22}$$

3.3 Independence and Mode Interactions

Within the fMRI community, ICA is often criticised for assuming statistical independence between modes [Varoquaux et al. 2010]. However, the huge success and widespread applicability of the algorithm, even on fMRI data, suggest this is not too injudicious an assumption in practice [Hyvärinen et al. 2001b; Hyvärinen 2013]. We would certainly hope not, as the PFM spatial prior is also independent over modes (Equations 3.4 and 3.5).

So why is this independence assumption such a concern amongst those working with fMRI data? In practice, when researchers talk about spatial interactions between modes they are referring to the inter-modal covariance matrix. If modes spatially interact this manifests itself as off-diagonal elements in this matrix, or

$$\mathrm{E}\left[\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\right] \neq \boldsymbol{I}_M \ . \tag{3.23}$$

ICA precludes this type of interaction[3], and this seems to be the major point of concern. However, it is actually relatively easy to come up with models that can capture this very simple notion of spatial dependence. For example, the toy model given below (Equation 3.24) is both independent over modes and can support an inter-modal correlation structure.

$$\begin{aligned} \mathrm{p}(\boldsymbol{P}) &= \mathcal{MN}(\boldsymbol{\mu}, \boldsymbol{I}_V, \boldsymbol{I}_M) \\ &= \prod_{m=1}^{M} \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{I}_V) \\ \mathrm{E}\left[\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\right] &= \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} + V\boldsymbol{I}_M \end{aligned} \tag{3.24}$$

What we show in Appendix A is that these spatial correlations can be supported by the PFMs model, despite its set of spatial independence assumptions. Intuitively, this is because the spatial maps are not identically distributed across voxels, and therefore, the inter-modal correlations are carried in the voxelwise means, while the distributions themselves remain independent. By way of contrast, ICA makes two independence assumptions: the first is the widely recognised assumption that the marginal distributions for each component are independent (and non-Gaussian); the second is less widely acknowledged, and states that each observation is independently and identically distributed—thereby precluding the type of mean structure described above. What this means is that sICA is blind to the spatial structure of the data, and while residual dependencies between ICA components do exist [Beckmann et al. 2005], one feels that being unable to capture these characteristics of modes at the model level is not necessarily a desirable property.

[3]Technically, strict orthogonalisation of components by pre-whitening—which is what is usually referred to in practice—is actually an algorithmic convenience, but the ICA generative model still does not support this type of correlation.

Chapter 4

Inference

Given a set of rfMRI data, we wish to infer a set of modes that is consistent with our probabilistic model. We aim to do this in a Bayesian framework, which will give a set of PFMs described by a posterior distribution. However, calculating the full posterior analytically is intractable and a sampling procedure for a model with this many parameters would be prohibitively slow. Therefore, we choose a computationally-efficient variational approach.

In this chapter, we give a brief overview of variational Bayesian methods, before describing our implementation of this for the PFM model.

4.1 Variational Bayesian Inference

4.1.1 Bayesian inference

While well established, Bayesian inference still seems to generate more than its fair share of polemical debate—'missionary zeal' and 'intemperate rhetoric' were the terms Gelman and Robert [2013] used. However, on the face of it, the procedure does not seem too contentious: one simply specifies a model $\mathscr{M}$ with associated parameters $\boldsymbol{\theta}$, taking care to write down what one knows about the parameters a

priori, and then collects some data $\mathscr{D}$. Bayes' rule [Bayes and Price 1763] does the rest.

$$\underbrace{\mathrm{p}(\boldsymbol{\theta} \mid \mathscr{D}, \mathscr{M})}_{\text{Posterior}} = \frac{\overbrace{\mathrm{p}(\mathscr{D} \mid \boldsymbol{\theta}, \mathscr{M})}^{\text{Likelihood}} \overbrace{\mathrm{p}(\boldsymbol{\theta} \mid \mathscr{M})}^{\text{Prior}}}{\underbrace{\mathrm{p}(\mathscr{D} \mid \mathscr{M})}_{\text{Evidence}}} \tag{4.1}$$

All the information about the model parameters is then encoded in the posterior distribution: it not only tells us what parameter values are consistent with the data, but also how certain we can be about those values.

Facetiousness aside, Bayesian methods are not the only way to infer such models, and working with the full posterior rather than point estimates is a frequently intractable computational burden. So should we do Bayesian inference at all? Proponents often point to its admissability; or the fact that as well as inferring the parameters, one simultaneously learns about the uncertainty associated with the parameter estimates; or that it allows principled model comparison, while critics often refer to the subjectivity of prior information. However, our views are borne from more prosaic considerations than those lofty ideals, and, as we use it, the Bayesian approach simply represents a principled framework for inference on probabilistic models[1].

That is not to say that we are not interested in such justifications; rather, that, for various reasons, we do not exploit the full scope of the Bayesian machinery. As we interact with it, the key property of Bayesian inference is the uncertainty over parameter estimates it encodes. Firstly, on an inferential level, it seems self-evident that keeping track of parameter uncertainty allows more reasonable conditional inferences to be made about the other parameters[2]. Secondly, the posterior distribution is an enormously rich description of the data. While this is challenging to visualise, we expect that this will contain genuinely interesting

[1] Realistically, this probably classifies us as 'pseudo-Bayesian' at best [Neal 1998].
[2] Especially within a framework where parameter estimates are updated iteratively.

information, and furthermore, this is information that could be incorporated into any subsequent analyses of our results, connectomic or otherwise.

4.1.2 Variational approximations

Rather than evaluating the exact posterior distribution, Variational Bayesian (VB) inference proceeds by optimising an approximation to this[3]. The trick is to choose an approximating distribution, $q(\boldsymbol{\Theta})$, that is simple enough to be tractable, but complex enough to capture the key features of the true posterior, $p(\boldsymbol{\Theta} \mid \mathscr{D})$.

There are two key decisions to be made: how to choose an approximation, and how to optimise it. We will deal with the latter first.

Optimisation by minimisation of the Kullback-Leiber divergence

The Kullback-Leibler (KL) divergence, D_{KL}, is a frequently-used measure of the difference between probability distributions [Kullback and Leibler 1951]. Therefore, if one minimises $D_{KL}\big(q(\boldsymbol{\Theta}) \,\big\|\, p(\mathscr{D} \mid \boldsymbol{\Theta})\big)$ then, in some sense, the resulting $q(\boldsymbol{\Theta})$ will be the 'best' approximation of $p(\mathscr{D} \mid \boldsymbol{\Theta})$. As we shall see, the KL divergence turns out to be a convenient choice as there exists a whole family of approximating distributions for which the necessary optimisations are relatively straightforward.

Before we continue, we reformulate the minimisation of the KL divergence as

[3] Many summaries of the VB approach are available. The brief overview we present here is based on a handful of these [Attias 2000; MacKay 2003; Winn et al. 2005; Chappell et al. 2008].

a maximisation problem.

$$\begin{aligned}
\mathrm{D}_{\mathrm{KL}}\Big(\mathrm{q}(\boldsymbol{\Theta}) \,\Big\|\, \mathrm{p}(\mathscr{D} \mid \boldsymbol{\Theta})\Big) &= \int \mathrm{q}(\boldsymbol{\Theta}) \ln\left(\frac{\mathrm{q}(\boldsymbol{\Theta})}{\mathrm{p}(\boldsymbol{\Theta} \mid \mathscr{D})}\right) \mathrm{d}\boldsymbol{\Theta} \\
&= \int \mathrm{q}(\boldsymbol{\Theta}) \ln\left(\frac{\mathrm{q}(\boldsymbol{\Theta})\,\mathrm{p}(\mathscr{D})}{\mathrm{p}(\boldsymbol{\Theta}, \mathscr{D})}\right) \mathrm{d}\boldsymbol{\Theta} \\
&= -\int \mathrm{q}(\boldsymbol{\Theta}) \ln\left(\frac{\mathrm{p}(\boldsymbol{\Theta}, \mathscr{D})}{\mathrm{q}(\boldsymbol{\Theta})}\right) \mathrm{d}\boldsymbol{\Theta} \\
&\quad + \int \mathrm{q}(\boldsymbol{\Theta}) \ln\Big(\mathrm{p}(\mathscr{D})\Big) \,\mathrm{d}\boldsymbol{\Theta} \\
&= -\mathscr{F}\Big(\mathrm{q}(\boldsymbol{\Theta})\Big) + \ln\Big(\mathrm{p}(\mathscr{D})\Big)
\end{aligned} \tag{4.2}$$

$$\ln\Big(\mathrm{p}(\mathscr{D})\Big) \geq \mathscr{F}\Big(\mathrm{q}(\boldsymbol{\Theta})\Big) \tag{4.3}$$

The term $\mathscr{F}\Big(\mathrm{q}(\boldsymbol{\Theta})\Big)$ is referred to as the free energy, and as the above equations show, this has the attractive property that it is a rigorous lower bound on the true log model evidence for any approximating distribution $\mathrm{q}(\boldsymbol{\Theta})$. As such, the free energy can also be used for principled model comparison.

Therefore, we can recast our minimisation of the KL divergence as a maximisation of the free energy. However, this has not made the problem any easier in and of itself: the challenge is now to choose an approximating distribution that means that the necessary integrations can be solved without resorting to numerical methods.

Factorised approximations

A straightforward approach is to assume $\mathrm{q}(\boldsymbol{\Theta})$ factorises over convenient groups of variables, namely

$$\mathrm{q}(\boldsymbol{\Theta}) = \prod_{\theta \in \boldsymbol{\Theta}} \mathrm{q}(\theta) \,. \tag{4.4}$$

Probabilistically, these factorisations represent groups of variables that are independent in $\mathrm{q}(\boldsymbol{\theta})$. This is a strong assumption, but one that allows us to proceed.

Using the calculus of variations, it is then possible to show that the form for each $\mathrm{q}(\theta)$ that maximises $\mathscr{F}$ is

$$\ln\Big(\mathrm{q}(\theta)\Big) = \Big\langle \ln\Big(\mathrm{p}(\boldsymbol{\theta}, \mathscr{D})\Big)\Big\rangle_{\mathrm{q}(\boldsymbol{\theta}\neq\theta)} + \text{const}\,, \tag{4.5}$$

where $\big\langle \bullet \big\rangle_{\mathrm{q}(\boldsymbol{\theta}\neq\theta)}$ represents the expectation taken with respect to the approximating distributions of all variables apart from θ.

This yields a very straightforward optimisation routine. We consider each independent group of variables in turn, and given the current approximating distributions of all the other variables, update the approximate posterior using Equation 4.5. This procedure is guaranteed to converge to a *locally* optimal solution [Attias 2000].

Of course, while this factorisation allows us to proceed, it introduces a deviation from the true posterior and the implications of this are very hard to quantify. We should of course evaluate our results critically, and our tests with simulations and real data, described later, lead us to believe that the effects of this are not too pernicious. Unfortunately, a post hoc examination of the results like this conflates two distinct issues: the inadequacy of our model, given the simplifying assumptions we need to make to ensure tractability, and the inaccuracy of the approximate posterior. It may be possible to investigate the latter systematically on a much reduced data set with an exhaustive sampling approach, and this would allow us to inspect where the approximation is not adequately capturing the form of the true posterior.

However, investigating the modelling assumptions is much harder. The concept of a mode is still inchoate, and as such, our stated aims—for example, about inducing sparsity in the spatial maps to represent functional segregation—are vague enough

to be satisfied by whole families of priors. In a sense this is a natural consequence of this type of exploratory analysis, and we would expect that we will be able to refine our model based on what we learn from our results. A different approach would be to try and use task data to verify some of the modelling assumptions. We could try and learn about the key statistics of the spatial patterns of activity induced by task manipulations, though it is not clear that there would be a direct mapping between these induced patterns and what we observe at rest.

Conjugate-exponential models

It is possible to make one final simplification to the derivations of the variational updates. If the distribution of a set of variables is conjugate with respect to the relevant hyperpriors—as Winn et al. [2005] describe, 'a [hyperprior] distribution $\mathrm{p}(X \mid Y)$ is said to be conjugate to the distribution of ... $\mathrm{p}(W \mid X)$ if $\mathrm{p}(X \mid Y)$ has the same functional form, with respect to X, as $\mathrm{p}(W \mid X)$'—and if the distributions are drawn from the exponential family, then the update rules are particularly straightforward [Attias 2000; Winn et al. 2005]. If the model has this conjugate-exponential property, then the approximate posterior for a set of variables can take the same functional form as the relevant hyperprior, and the expectations needed to update these distributions, as in Equation 4.5, are all straightforward to calculate.

In fact, if the model is conjugate-exponential, it is possible to build generic algorithms that derive and solve the update rules automatically [Winn et al. 2005; Kucukelbir et al. 2015].

4.2 Implementation of the PFM Model

As all the distributions in the PFM model are drawn from the conjugate-exponential family the variational update rules are somewhat routine; therefore, we provide

them in Appendix B rather than in the main body of the thesis. The rather more involved computational implementation of the update rules for the BOLD time courses can be found in Appendix C. The code itself, and some associated documentation, can be found at git.fmrib.ox.ac.uk/samh/profumo.

Algorithmically, the key observation is that 'the variational update equation for a node … depends only on expectations over variables in the Markov blanket of that node' [Winn et al. 2005]. This naturally suggests a modular structure to the code. If we define a node as a representation of some variables in the graphical model, each node only needs to be aware of a limited set of other nodes, be responsible for its own update rules, and provide expectations for other nodes as required. This nodal structure can be concisely expressed with a handful of classes in C++.

However, rather than implementing this as a completely generic message passing framework, we explicitly choose a handful of key parameter groupings, with expectations that can conveniently be expressed in matrix form, and base key nodes around them. This still results in a modular structure, but allows us to harness efficient linear algebra libraries to do the bulk of the computation [Sanderson 2010].

For example, note that the only expectations the time course and noise models require from the spatial maps are the two matrices $\langle \boldsymbol{P} \rangle$ and $\langle \boldsymbol{P}^{\mathrm{T}} \boldsymbol{P} \rangle$. Therefore, we can implement any spatial model as a node, and providing it returns those expectations, it can simply be dropped into the existing framework. In fact, we provide five models for the spatial maps (eleven accounting for different formulations of the hyperpriors), three time course models, four models for precision matrices, and many more besides.

So why provide so many models above and beyond those described in the Modelling Approach chapter? The reason is that this enables us to dynamically adapt the model structure during the inference procedure. Our experience is that

a model with as many parameters as we outlined earlier is relatively unstable when randomly initialised, but that the full PFM model converges as we would hope when initialised with a set of modes from a relatively simpler model. This is conceptually similar to multi-scale registration approaches, where a registration is performed at a low spatial resolution before being gradually refined as more spatial detail is introduced [Robinson et al. 2014]. Alternatively, this can be thought of as a heuristic approximation to a simulated annealing approach, where the introduction of a more complex model mirrors the reduction in temperature [MacKay 2003].

The main simplification we use is to define a non-hierarchical spatial model, where we define a set of group maps $\boldsymbol{P}^{(g)}$ and turn off the spatial variability by setting $\boldsymbol{P}^{(s)} = \boldsymbol{P}^{(g)}$. For consistency with the group hyperpriors, $\boldsymbol{P}^{(g)}$ follows the same spike-slab distribution as the matrix of μ parameters (Equation 3.5). We make a similar simplification in the temporal domain by setting a group-level between-mode precision matrix that does not vary from subject to subject.

Our inference then follows the following procedure. The simple group-level spatial and temporal models are initialised with a random set of spatial maps. Inference on this model identifies a set of group-level modes, which is a much better starting point for the hierarchical models than a random initialisation. Our code structure allows these more complex nodes to be introduced into the graphical model at run-time, so we slowly build up the model complexity, interspersed with applications of the update rules, until we reach the PFMs model as outlined earlier. We provide a more detailed description of this approach in Appendix D.

4.2.1 Spatial degrees of freedom correction

We make one final change to the probabilistic model, motivated by the spatial smoothing that is a standard pre-processing step for most fMRI data. This was something we discussed in relation to the noisy time course model (Section 3.1.2),

but in reality has somewhat broader implications. In essence, by not acknowledging the inherent spatial smoothness of the data in our specification of the noise process, the model assumes more independent spatial measurements than actually exist.

Fortunately, as Groves et al. [2011] discuss, there is a simple way to mitigate some of the effects of this within the VB framework. Intuitively, if we have smoothed the data then we should be able to downsample it. At some stage, this would result in the noise becoming genuinely spatially independent again. However, this presents several practical problems, so rather than actually downsample the data, we simply downweight the spatial information by a factor ν. This represents the proportion of voxels that would be retained if we were to optimally downsample. 'This is analogous to fixing that only a random fraction of the data points will be kept, but at each stage averaging over all possible choices of decimated voxels' [Groves et al. 2011].

The implementation of this at the code level is relatively straightforward. Whenever an update rule requires a sum over voxels, for example as is implied when regressing the spatial maps into the data to update the time courses, the result of this sum is simply multiplied by ν.

While this approach still does not explicitly acknowledge the relationship between noise in nearby voxels, it does counter most of the deleterious effects of this model misspecification, especially when combined with the noisy time course model. The main advantage of this approach, compared to a more formal model for smoothness, is that it remains particularly computationally efficient.

4.3 Data

The data we use for this thesis has been collected as part of the Human Connectome Project[4] (HCP): 'an ambitious 5-year effort to characterize brain connectivity and

[4] humanconnectome.org

function and their variability in healthy adults' [Van Essen et al. 2012b]. The project is collecting neuroimaging, behavioural and genetic data from 1,200 subjects, with a large technical focus on pioneering neuroimaging acquisitions that represent a revolutionary increase in data quality.

For each subject, high quality structural, task and resting-state functional, and diffusion MRI data are acquired, allowing a description of the 'macro-connectome' at an unprecedented level of detail [Van Essen et al. 2012b; 2013]. This information is augmented with data from a battery of behavioural tests and genotyping, and as many of the subjects are either twins or siblings, this should allow the investigation of both the environmental and genetic correlates of brain connectivity.

In this section, we give a brief overview of the rfMRI data the HCP collects, before describing the additional processing and data reduction strategies we use to make inference on such a large data set feasible.

4.3.1 Human Connectome Project rfMRI data

The HCP rfMRI data represents a step change in both the spatial and temporal resolution of fMRI data, as well as scanning each subject for between five and ten times as long as is currently the norm[5]—see the manuscript by Smith et al. [2013a] for a much more detailed overview. Using a highly optimised protocol relying heavily on multiband acceleration, the data is acquired in 2 mm isotropic voxels at a sub-second TR. Each subject is scanned four times, for fifteen minutes, resulting in 4,800 time points per subject.

Spatial pre-processing corrects for distortion and motion and then registers the data into a common space. For the HCP, functional data are registered onto a common set of 'grayordinates' [Glasser et al. 2013]. As the rfMRI signals of

[5]While it has been suggested that between 5 and 15 minutes of rfMRI data per subject is adequate [Van Dijk et al. 2010; Birn et al. 2013], several studies have noted that estimation is improved by much, much more data [Anderson et al. 2011; Hacker et al. 2013; Laumann et al. 2015].

interest reside in grey matter, registering data to the cortical surface is a much more parsimonious representation of the data than the traditional volumetric formats. However, due to their 3D geometry or fine-grained spatial structure, subcortical structures and the cerebellum are represented volumetrically. This yields a set of 60,000 surface vertices split across the two cortical hemispheres, and 30,000 voxels distributed across the other structures. Finally, the data is spatially smoothed with a 2 mm FWHM Gaussian spatial kernel, and a similarly conservative temporal high-pass filter is used to remove drifts.

The last stage of the pipeline is the automated removal of structured artefacts—a development which hugely improves the temporal-SNR of the data. sICA is used to decompose each fifteen minute run, at which stage the FIX algorithm classifies the resulting components as either signal or artefact, and the artefacts are regressed out of the data [Griffanti et al. 2014; Salimi-Khorshidi et al. 2014].

The result is an hour of data per subject, at very high spatio-temporal and resolution, that is essentially free from structured artefacts. Furthermore, this is all parsimoniously represented as a set of grayordinates. For this thesis, we use data from the 461 subjects released to date[6], with two minor modifications. On a run by run basis, we subtract the voxelwise means and then set the voxelwise variances to unity. These simply bring the data more closely into line with our assumption that the PFM noise process is zero-mean and isotropic across voxels.

4.3.2 Subject-specific data reduction

Even efficiently represented as a set of grayordinates, the four rfMRI runs still combine to 1.6 GB of data per subject [Glasser et al. 2013]. This will result in just shy of 2 TB for the whole HCP cohort, and represents a not insubstantial 750 GB for the current release that we are working with here. As such, it is necessary to

[6] 500 Subjects MR + MEG2 Release: humanconnectome.org/documentation/S500

employ some form of data reduction strategy in order to work with the whole group of subjects simultaneously.

The PFM model is explicitly formulated with terms that represent both the spatial and temporal characteristics of modes. Therefore, in an approach somewhat similar to Calhoun et al. [2001], we work with a set of subject-specific low rank approximations of the data, rather than the full data matrices.

We approximate each run with a low-rank singular value decomposition (SVD). However, as the PFM model assumes that subject-specific spatial maps are conserved across all runs for a given subject, we make further savings by only maintaining a single set of spatial singular vectors per subject.

To do this, we calculate the SVD of the matrix formed by temporally concatenating all data from a given subject. This combined data matrix, $\boldsymbol{D}^{(s)} \in \mathbb{R}^{V \times R_s T}$, is then represented by $\boldsymbol{U}^{(s)}$, $\boldsymbol{S}^{(s)}$ and $\boldsymbol{V}^{(s)}$. To approximate this with a low rank SVD, we simply only retain the singular vectors associated with the top N singular values. For example, assuming $V > R_s T$ and ignoring columns associated with singular values equal to zero, $\boldsymbol{U}^{(s)} \in \mathbb{R}^{V \times R_s T}$ is replaced by $\hat{\boldsymbol{U}}^{(s)} \in \mathbb{R}^{V \times N}$. Finally, we can partition the temporal singular vectors, according to the order the individual runs were concatenated, in order to reconstruct the data from each run individually, or in other words, $\hat{\boldsymbol{V}}^{(s)} \in \mathbb{R}^{R_s T \times N}$ is decomposed into a set of $\hat{\boldsymbol{V}}^{(sr)} \in \mathbb{R}^{T \times N}$. In summary, each data matrix, $\boldsymbol{D}^{(sr)}$, has three approximating matrices, namely $\hat{\boldsymbol{U}}^{(s)} \in \mathbb{R}^{V \times N}$, $\hat{\boldsymbol{S}}^{(s)} \in \mathbb{R}^{N \times N}$ and $\hat{\boldsymbol{V}}^{(sr)} \in \mathbb{R}^{T \times N}$.

The last thing we do is to combine these three matrices into two matrices. This simply saves some computation each time we need to calculate any expectations

involving the data. The final form for the approximate data is therefore

$$\begin{aligned} \boldsymbol{W}^{(s)} &= \big(\hat{\boldsymbol{U}}^{(s)}\big)\big(\hat{\boldsymbol{S}}^{(s)}\big)^{\frac{1}{2}} \\ \boldsymbol{X}^{(sr)} &= \big(\hat{\boldsymbol{S}}^{(s)}\big)^{\frac{1}{2}}\big(\hat{\boldsymbol{V}}^{(sr)}\big)^{*} \\ \boldsymbol{D}^{(sr)} &\approx \boldsymbol{W}^{(s)}\boldsymbol{X}^{(sr)} \end{aligned} \tag{4.6}$$

We can simply substitute this approximate expression for $\boldsymbol{D}^{(sr)}$ any time it appears in the VB update rules, and this has the added bonus of being computationally, as well as space, efficient. However, we explicitly calculate, and cache, $\mathrm{Tr}\big((\boldsymbol{D}^{(sr)})^{\mathsf{T}}\boldsymbol{D}^{(sr)}\big)$ from the full data, rather than ignoring the data variance in the subspace of discarded singular values. This means that the estimate for the noise precision, $\psi^{(sr)}$, will be comparable whether or not we choose to utilise this low-rank approximation, or indeed across different values of N.

We now have an explicit method for reducing the HCP data to a more manageable size. However, there is one final complication: computationally, calculating the SVD of every $\boldsymbol{D}^{(s)}$ actually turns out to be prohibitively expensive. In order to circumvent this, we utilise the fact that we are explicitly looking for a low-rank approximation and implement an extremely efficient randomised algorithm to directly calculate the truncated SVD. This approach is described in the excellent review by Halko et al. [2011].

Choice of dimensionality

We plot the singular values from an example combined data matrix, $\boldsymbol{D}^{(s)} \in \mathbb{R}^{V\times R_s T}$, in Figure 4.1. Our choice is to only retain the top 500 singular values from the random SVD, and, as can be seen from the figure, this lies well to the right of the 'elbow' in the plot. Even if we are particularly worried about the error bounds on the random implementation of the SVD, we can see that one could make a case for including only half as many singular values, giving us a huge buffer against this

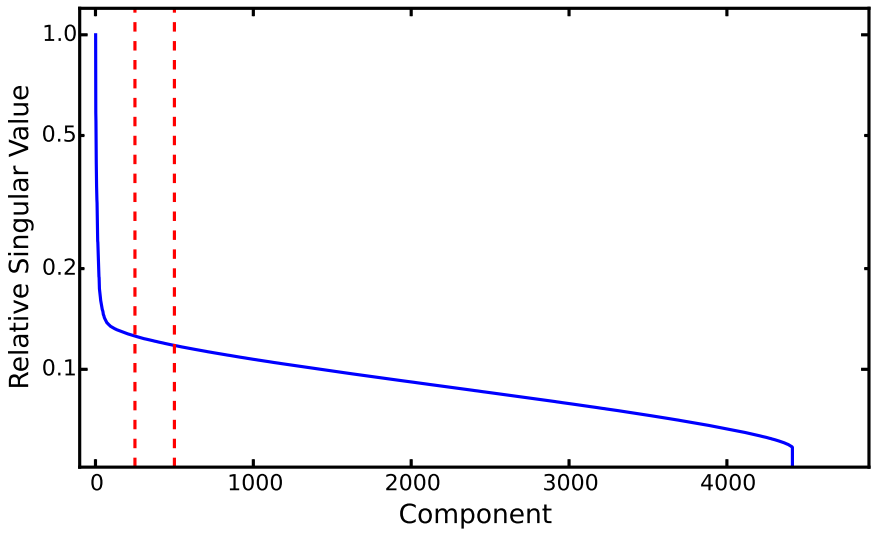


Figure 4.1: Singular values from four concatenated HCP runs for an example subject. For each run, the grayordinates were de-meaned and renormalised to unit variance. The runs were temporally concatenated and the SVD was calculated for this combined matrix. We plot the singular values of this matrix, normalised such that the largest singular value is equal to one. The removal of structured artefacts by FIX means that several hundred singular values are equal to zero. The vertical lines are plotted at the 250$^{\text{th}}$ and 500$^{\text{th}}$ singular values.

type of error[7].

On the data reduction front, this reduces the memory demands by nearly a factor of 10. The full data requires nearly 750 GB of data, but this is reduced to 83 GB if we only take the top 500 singular values. Therefore, we are satisfied that this data reduction approach achieves a reasonable balance between minimising the memory demands and retaining as much of the structure in the original data as possible.

[7] In fact, the claim made by Halko et al. [2011] is that, in practice, the theoretical error bounds on the random SVD tend to be rather pessimistic.

Chapter 5

Results

In this chapter, we give an overview of the properties of a set of PFMs as inferred from the HCP data. We compare the PFMs to modes that have been characterised using previous methods, and demonstrate the types of extra information that are captured by both the group- and subject-level aspects of the PFM model.

However, before we start, there is one pressing question to answer: how many modes do we expect to find? If we take the default mode as our canonical example, then we can see that that the regions commonly associated with this mode cover approximately ten percent of the cortex[1]. This would imply of the order of 10 modes, though in reality we expect several more. The default mode is relatively large compared to other previously characterised modes, and allowing spatial overlap will potentially increase the number of identifiable modes again. Furthermore, as we are using HCP data of a hitherto unattainable quality, we expect that this will allow us to identify a larger set of modes. Bearing the above in mind, the results in this section are based on an analysis where we aimed to

[1] This is hopefully corroborated by a quick visual inspection of Figure 1.2. Furthermore, note that the default mode is one of the 10 modes identified by Beckmann et al. [2005], and the regions classically associated with the default mode were grouped together when Yeo et al. [2011] divided the cortex into either 7 or 17 putative functional subsystems—both consistent with approximately 10 % cortical coverage. As an aside, parcellations lie at the other end of the spectrum to modes and current estimates suggest that there are a couple of hundred cortical areas [Van Essen et al. 2012a].

identify 50 PFMs. Further details of the analyses, including all parameters, can be found in Appendix D.

In theory, it would be possible to identify several sets of PFMs at different dimensionalities, and use the Bayesian model evidence, or in our case the free energy, to select the number of PFMs. Again, we are limited by computational practicalities. It takes around six hours to load the data and three or four days to run through the whole inference procedure, and as such, multiple runs of the algorithm were too expensive for us to pursue this approach. Note however, that the model can eliminate PFMs via automatic relevance determination if they are not supported by the data [Mackay 1995; 1994]. This should at least mitigate the impact of over-specification of the dimensionality.

For the rest of the chapter, we focus on a subset of 30 PFMs. Of the remainder, 14 PFMs were excluded because they had been eliminated from the model, 5 PFMs were too inconsistent over subjects to be interpretable and 1 PFM seemed to be artefactual in origin.

The group-level spatial maps for the 30 PFMs we will focus on for the rest of the chapter can be found in Appendix F. Finally, for visualisation purposes, we group these modes into six broad categories based on their spatial localisations. These are:

- *Global (Glo): PFM 1.* A PFM representing activity distributed across the whole cortex.
- *Visual (Vis): PFMs 2–8.* PFMs primarily confined to the primary visual cortex or the visual processing streams.
- *Motor (Mot): PFMs 9–12.* PFMs primarily confined to the somato-motor cortex.
- *Auditory (Aud): PFM 13.* A PFM lying within primary auditory cortex.

- *Cognitive (Cog): PFMs 14–29.* PFMs primarily confined to association cortex.
- *Subcortical (Sub): PFM 30.* A PFM primarily residing within subcortical structures and the cerebellum.

5.1 Group Maps

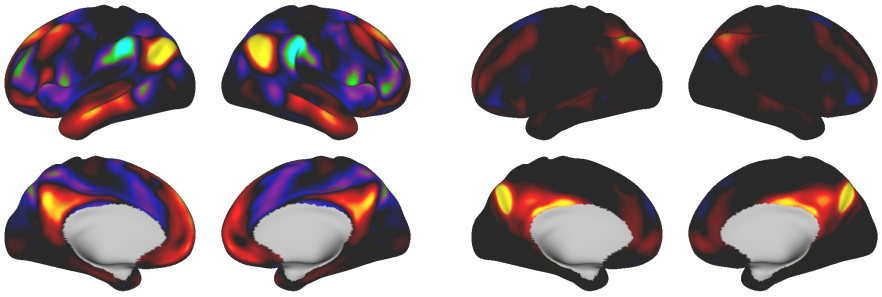
It is currently standard to simply identify modes based on the brain regions present in their group-level spatial maps. With that in mind, we show a selection of the cognitive PFMs in Figure 5.1.

There are several things that are apparent from a qualitative inspection of at these maps. Firstly, the PFMs are sparse, some contain anti-correlations and there are regions where several PFMs overlap. Therefore, the group-level spatial prior appears to be performing as intended. Secondly, at the group level, the PFMs are consistent with the existing literature. Five of the PFMs shown are recognisable from previous studies, including the seminal default mode (PFM 14, Figure 5.1(a)). The only PFM of the six shown here that has not been widely reported previously (PFM 15, Figure 5.1(b)) seems to match a set of regions identified from a memory retrieval meta-analysis [Power et al. 2011].

However, while the identification of established modes is ostensibly encouraging, it is unclear how we should interpret points of difference between the PFMs and the literature, or indeed, those PFMs that we do not recognise at all. In short, can we trust these results?

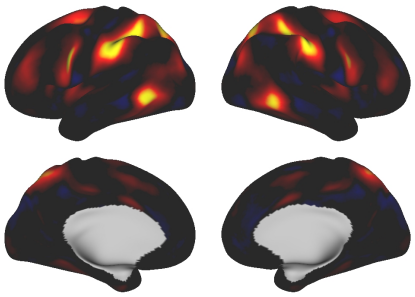
5.1.1 Reliability

How best to assess the reliability of mode decompositions is still a contentious issue. Tests on simulated data are widespread [Welvaert and Rosseel 2014], though these are almost invariably designed to test a specific method and, as such, tend

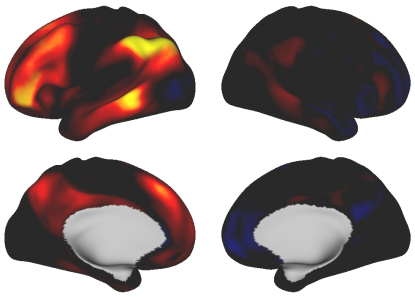


(a) Default mode (PFM 14).

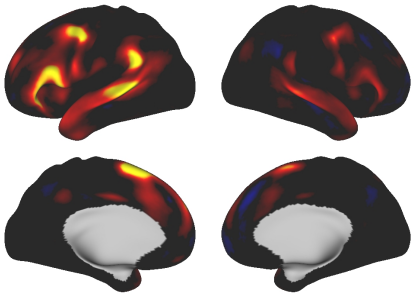
(b) Mode residing in medial parietal cortex and the intraparietal sulcus (PFM 15).



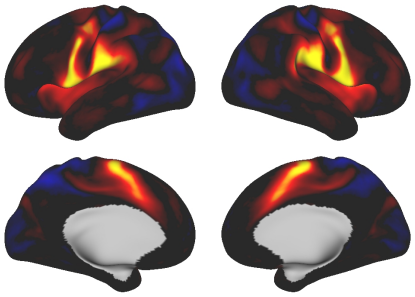
(c) Dorsal attention mode (PFM 18).



(d) Left-lateralised fronto-parietal control mode (PFM 17).



(e) Language mode (PFM 20).



(f) Cingulo-opercular mode (PFM 21).

Figure 5.1: Example group-level PFM spatial maps. All PFMs were classified as cognitive and, where possible, PFMs are referred to by their common names in the literature. For example, see Corbetta and Shulman [2002], Greicius et al. [2003], Beckmann et al. [2005], Dosenbach et al. [2006, 2007], Vincent et al. [2008], Smith et al. [2009], Spreng et al. [2010], Power et al. [2011], Yeo et al. [2011] and M. W. Cole et al. [2013]. Posterior mean spatial weights are displayed on the left and right cortical surfaces. Subcortical structures were also included in the analysis, but are not shown here. The colour scale for each map is symmetric around zero and is clipped at the 99.5 % level of the distribution of the spatial weight magnitudes. The default mode is clipped at 99 % as it is spatially extended relative to the other 5 PFMs shown here. Cortical surface views were generated using Connectome Workbench (humanconnectome.org/software/connectome-workbench.html).

to be either 'too simple or [too] specialized to be of general use' [Erhardt et al. 2012]. As such, the most common way of assessing reliability is to quantify how consistent the results are across multiple runs of the algorithm, and ideally on different data [Himberg et al. 2004; Groppe et al. 2009]. Here we use both simulations and measures of consistency on real data.

Firstly, we assess the reliability of PFMs using simulations. While there have been attempts to develop general rfMRI simulation frameworks [Smith et al. 2011; Erhardt et al. 2012], our aim was to better understand the implications of variability across subjects, so we resorted to designing another specialised test suite to explicitly capture this. The full details of these proof-of-principle simulations can be found in Appendix E. The results were encouraging for the PFM approach, with evidence that the model could genuinely capture both subject variability and modes that are simultaneously spatially and temporally correlated. The crucial result, however, was that the consistency with which a given PFM was identified over multiple runs of the algorithm tended to correlate well with ground-truth accuracy (Figure E.1). ICA, by way of contrast, seemed to be highly reliable but systematically biased. This gave us confidence that assessing the stability of PFMs over multiple runs and data is a genuinely useful test.

Secondly, therefore, we evaluate the split-half reliability of the PFMs on the HCP data. To do this, we split the HCP data into two sets of 230 subjects and infer a set of PFMs from each of the halves. In an approach inspired by Groppe et al. [2009], we independently pair the PFMs from each half to the PFMs we are presenting from the full data. Then, for each of the PFMs from the full data, we report the correlation between the two paired PFMs.

Encouragingly, 21 of the PFMs have split-half reproducibilities above 0.8. This is shown in Figure 5.2, and the group maps from each half are shown alongside the full PFMs in Appendix F. Furthermore, the modes achieving high scores are not

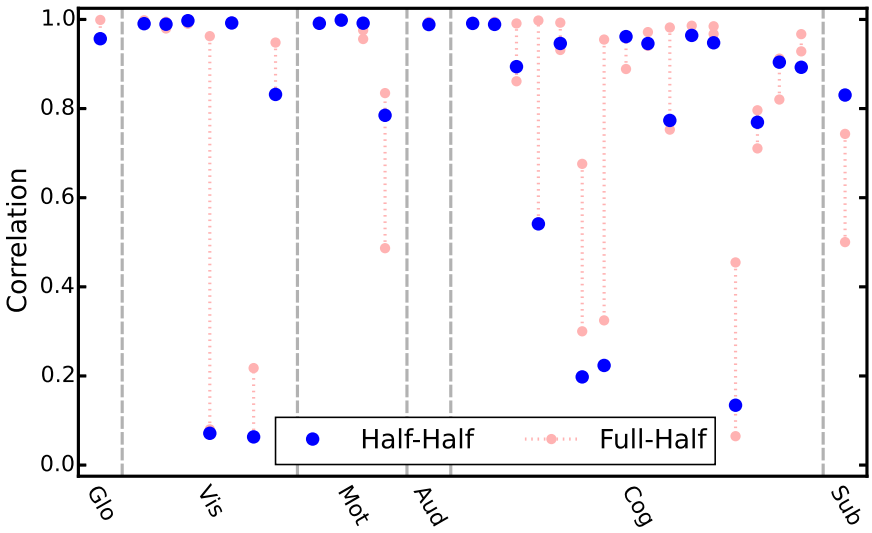


Figure 5.2: Split-half reproducibility of PFMs. As well as the previously reported PFMs from the full data, two sets of 50 PFMs are independently inferred after partitioning the data into halves. For each of 30 pre-selected PFMs from the full data set, we report the correlation between group-level spatial maps of the two best matched PFMs from the halved data (Half-Half). For completeness, we also show the correlation between each of the PFMs inferred from half the data and the full PFMs (Full-Half).

simply those which have been previously identified. For example, PFM 15, which we flagged as unknown in Figure 5.1(b), has a split-half reproducibility of 0.989. Therefore, this suggests that we can have some confidence in these PFMs.

However, it is unclear how we should interpret the lower-scoring PFMs. Interestingly, a closer inspection of Figure 5.2 reveals that of the nine PFMs scoring below 0.8, five of these are clearly present in one half of the data, but not both. The MT complex (PFM 5, Figure F.5) is an excellent example of this. Our interpretation of this is that some genuine PFMs are eliminated from the model because the random initialisation procedure we employ does not capture any of the signal they are responsible for. We touch on this issue again in our discussion of future lines of investigation for the PFM framework (Section 7.2).

In summary, a large proportion of PFMs are strongly reproducible, though this could of course be improved. However, there is one final interpretation of these

results we need to rule out. As we are using data of an unprecedented quality, it is possible that any new results are solely attributable to improvements in the acquisitions, rather than the PFM framework itself. In other words, does the PFM approach make any different predictions to existing techniques, when they are run on the HCP data?

5.1.2 Comparison with ICA

As the established technique for identifying modes, ICA is the obvious technique to compare with PFMs. Furthermore, as there is so much data available from the HCP, we can get reliable results from both spatial and temporal ICA [Smith et al. 2012; 2013, b].

Therefore, we identify a set of modes using both sICA and tICA and compare these to the PFMs. We display the similarity between the PFMs and the ICA-derived modes in Figure 5.3, and the ICA group-level spatial maps are displayed alongside the matched PFMs in Appendix F.

Clearly, the PFMs represent a characterisation of rfMRI data that is distinct from both sICA and tICA. There are obviously some similarities; indeed, that is what we expect given that several of the PFMs correspond to modes that have been identified by many different methods and from many different data sets. However, what this shows is that the differences between PFMs and modes previously identified by more traditional approaches cannot simply be explained away by the higher quality data.

5.1.3 Prediction of task contrasts

Finally, we use the task fMRI data acquired as part of the HCP to make one further assessment of the PFM group-level spatial maps. As part of the HCP nearly one hour of task fMRI data is acquired for each subject, with the battery of tasks

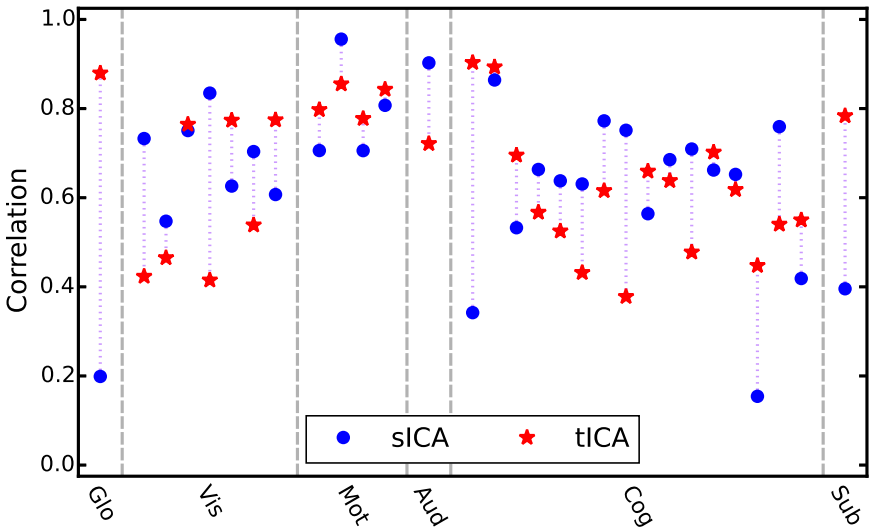


Figure 5.3: Comparison between PFMs and both spatial and temporal ICA. All techniques were made to identify 50 modes from the full set of available HCP data. Modes identified by ICA were paired with the PFMs, and the correlations between the ICA and PFM group-level spatial maps are displayed here for the subset of PFMs outlined earlier.

designed to cover several broad cognitive domains [Van Essen et al. 2012b]. As we know there should be a strong correspondence between the spatial patterns of activity during tasks and at rest [Smith et al. 2009], we can utilise these task results to validate the PFMs.

In Figure 5.4 we investigate how well the set of PFMs predicts the task observations, and again we compare this to the ICA modes. From this figure, it is clear that all the methods produce a set of modes that are consistent with knowledge from task data. Note, however, that this does not imply that the PFMs represent a more accurate separation of the underlying modes, but it does suggest that they are capturing the correct subspace of the rfMRI data. This consistency is therefore a necessary rather than sufficient condition for modes derived from rfMRI data, but it nevertheless gives us more confidence in the PFM approach.

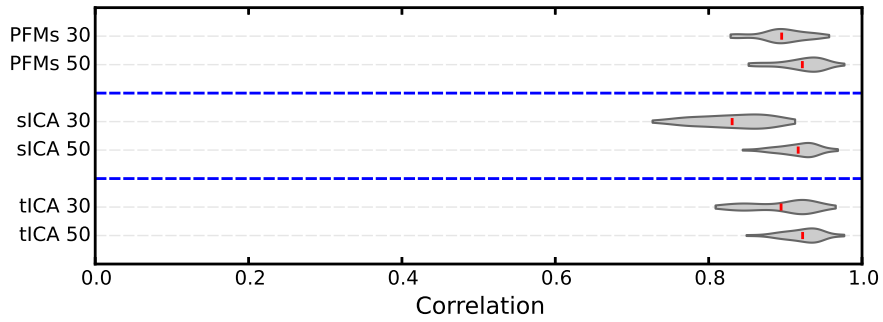


Figure 5.4: Prediction of group-level task maps by PFMs and ICA modes. For each task map in turn, we find the optimal approximation, in the least squares sense, using a weighted linear combination of the PFM, sICA or tICA spatial maps. We calculate the correlation between the true map and the approximation from the modes for each technique, and we do this for both the full set of 50 modes and the subset of 30 modes. The figure shows the distribution of correlations for each of the 86 HCP task contrasts. For the ICA approaches, we use the same 30 components as were paired to the PFMs for the comparison in Figure 5.3. Due to the different signal-to-noise ratios in cortical and subcortical grayordinates, these results are only calculated on the cortical surface.

Summary

In this section, we have shown that the set of PFMs we have identified possess a set of group-level spatial maps that are reproducible, are accordant with both the existing literature and task results, but do not simply represent a recapitulation of modes identified by existing techniques. These results, in and of themselves, suggest that PFMs represent an interesting decomposition that is genuinely novel. However, the PFM model is much richer than a simple group-level spatial representation, so in the following sections we outline some of the extra information that is associated with a set of PFMs.

5.2 Time Courses

In the temporal domain, one of the key aspects of the PFM model is the way the hæmodynamics are used to constrain the inferred time courses. In Figure 5.5 we plot the frequency content of the time courses from the default mode (PFM 14,

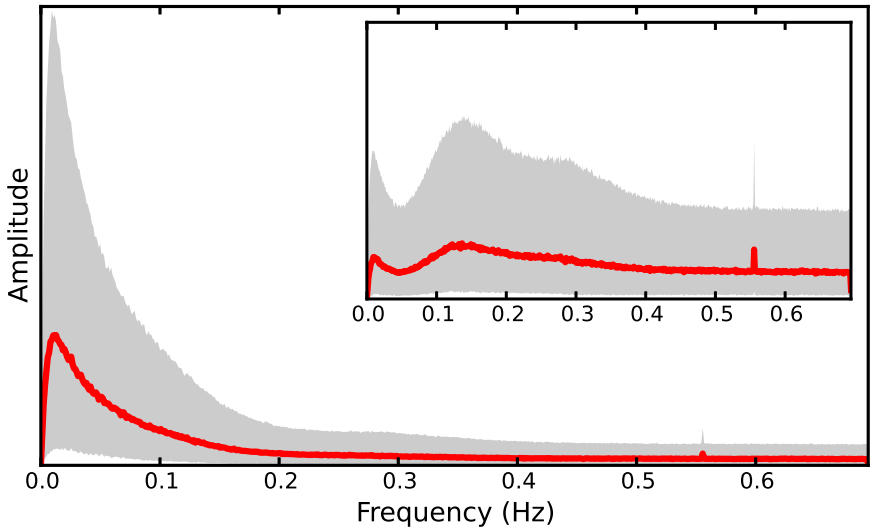
Figure 5.1(a)).

The combined time courses, $\boldsymbol{A}$, are shown in Figure 5.5(a). The vast majority of the power is concentrated at frequencies below 0.2 Hz in a manner that is completely consistent with the temporal blurring induced by the assumed HRF. However, there is a small amount of power present at frequencies all the way up to the Nyquist limit. The PFM model attributes this to the noise time courses, $\boldsymbol{\xi}$, and these have an amplitude spectrum which is nearly flat across frequencies, as can be seen from the inset panel.

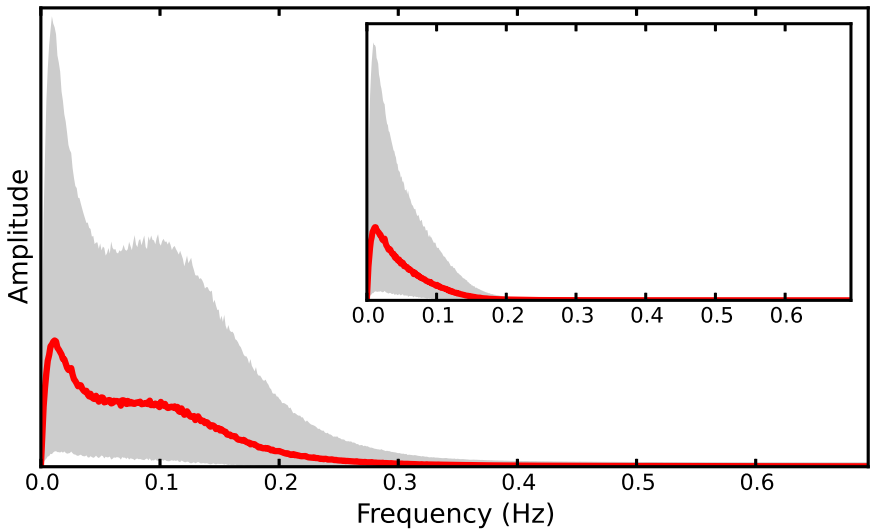
The frequency content of the BOLD time courses, $\boldsymbol{B}$, is shown in Figure 5.5(b). The inset panel shows that, once the noise processes have been accounted for, the BOLD time courses carry almost no power at frequencies above 0.2 Hz. The frequency spectrum of these time courses after correction for the autocorrelation induced by the HRF is shown in the main panel. The decorrelation with respect to the HRF does indeed boost power at higher frequencies, though the effect is slight—the frequency spectrum becomes flat between 0.05 Hz and 0.1 Hz, and the subsequent roll-off is then much slower, with power remaining up to 0.3 Hz. Importantly, this is stable[2]—hæmodynamic deconvolution is a non-trivial process in and of itself [Gaudes et al. 2011; Karahanoğlu et al. 2013].

There are two observations to be made from this. Firstly, there is essentially no power in the inferred BOLD signal above 0.3 Hz, despite the fact that the higher sampling rate of HCP data would appear to make this possible. In practice, the canonical linear HRF we have used admits such low amounts of signal at high frequencies that the model simply eliminates all power near the Nyquist limit, which is why the decorrelation is stable. It would be fascinating to see whether or not this would change if a more complex hæmodynamic model were used.

[2] In other words, the power at high frequencies does not 'explode' after decorrelation. This is normally caused by the amplification of high-frequency noise, as at these frequencies the inverted power spectral density is extremely large.



(a) Frequency content of the combined time courses, $\boldsymbol{A}$. **Inset:** Frequency content of the noise time courses, $\boldsymbol{\xi}$ (expanded vertical scale).



(b) Frequency content of the decorrelated BOLD time courses, $\boldsymbol{K_B}^{-\frac{1}{2}}\boldsymbol{B}$. **Inset:** Frequency content of the BOLD time courses, $\boldsymbol{B}$ (different vertical scales).

Figure 5.5: Frequency content of the time courses of the default mode (PFM 14, Figure 5.1(a)). The magnitude of the DFT coefficients are calculated for the specified time course from each of the 1,844 runs, and for each frequency bin we fit a gamma distribution to the histogram of observed magnitudes. The mode of this distribution is plotted in red, and the grey region represents the 95 % highest density interval [Kruschke 2014].

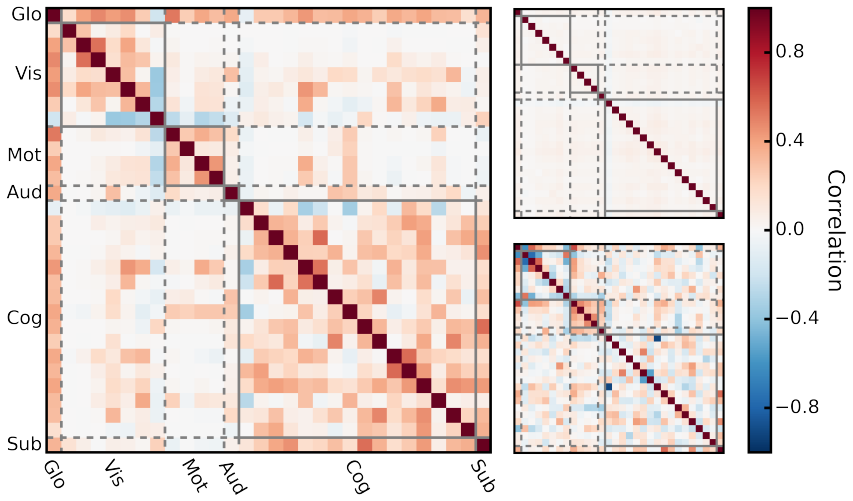
Secondly, there remains an excess of power below 0.05 Hz. While this could be explained as a misspecification of the HRF, this observation is tantalisingly similar to the predictions of some biophysical models, that posit that modes are driven by ultraslow (<0.1 Hz) fluctuations [Deco et al. 2009; 2011].

These results raise some intriguing questions about the nature of the HRF and intrinsic mode dynamics. However, the primary conclusion is that PFMs have time courses that are consistent with current models of the hæmodynamics that drive the BOLD signal, and furthermore, are stable under decorrelation. It is the latter, in particular, that is of particular interest next, as this gives us confidence that the interactions between PFM time courses more closely reflect genuinely neuronal processes [Gitelman et al. 2003]. We investigate those interactions in the next section.

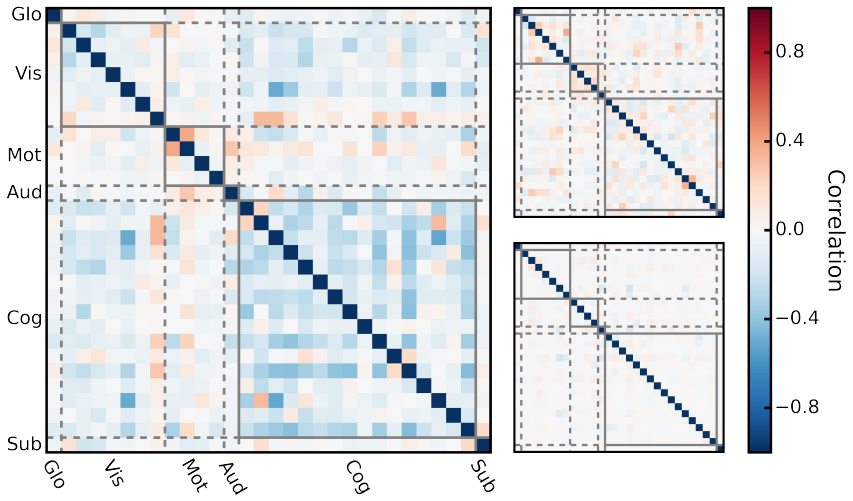
5.3 Interactions Between PFMs

One of the key motivations for the PFM model was to relax the assumptions ICA makes about the interactions between modes. To see if this is borne out by the results, we plot the spatial and temporal correlations between the inferred PFMs in Figure 5.6. For comparison, we also plot the correlations between the best matched sICA and tICA modes.

In the spatial domain, the PFMs tend to be strongly correlated, and especially so within each group. This confirms the intuition from the qualitative examination of the spatial maps in Figure 5.1, namely that those PFMs seemed to be spatially overlapping. The sICA modes are uncorrelated by construction, whereas the tICA modes do contain spatial correlations of a comparable magnitude to the PFMs. This ability to extract spatially overlapping PFMs is exactly the type of behaviour we had designed the PFM model to produce, and it is clear that the spatial sparsity prior is not detrimental in this regard.



(a) Correlations between group-level spatial maps. **Left:** PFMs; **top right:** sICA; **bottom right:** tICA.



(b) Partial correlations between time courses. **Left:** PFMs; **top right:** sICA; **bottom right:** tICA.

Figure 5.6: Spatial and temporal correlations between PFMs. The spatial correlations were calculated from the group-level spatial maps. The PFM temporal partial correlations were calculated from the group-level hyperprior on subject-specific temporal precision matrices, $\boldsymbol{\beta}$. For a comparable metric on the ICA time course correlations, we compute the partial correlations from the mean subject-specific temporal precision matrix. The ICA components were matched as in Figure 5.3.

However, the completely unexpected result is that the PFMs are predominantly anti-correlated with one another in the temporal domain. This is at odds with currently held ideas about mode interactions[3], and there is no evidence for this in either the sICA or tICA temporal correlations[4]. At the model level, inferring strong temporal correlations is very encouraging, but unlike the spatial correlations, there is no intuitive interpretation of this result.

To reiterate, this set of temporal correlations is a genuinely radical departure from previous results. While it is too early to speculate on interpretations of this, it is important to understand how it has arisen. What has enabled the PFM model to infer a set of mode interactions so completely different to those predicted by ICA?

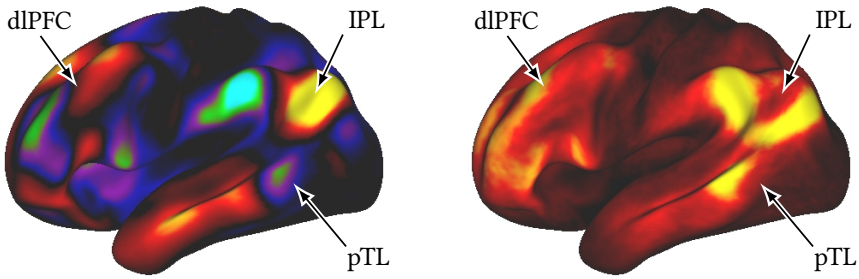
5.4 Subject Variability

The PFM model is built with the concept of subject variability at its heart. It is therefore natural to suspect that this may be driving the differences between the PFMs and modes inferred by ICA. To begin with, we simply plot the key group-level spatial parameters and some example subject-specific spatial maps for the default mode in Figure 5.7.

Several things are clear from a qualitative inspection of the subject-specific spatial maps. Firstly, despite high levels of variability, these maps are recognisable as variants of the group-level map. We can therefore be confident that the time courses, and any associated interactions, are all related to the default mode as a

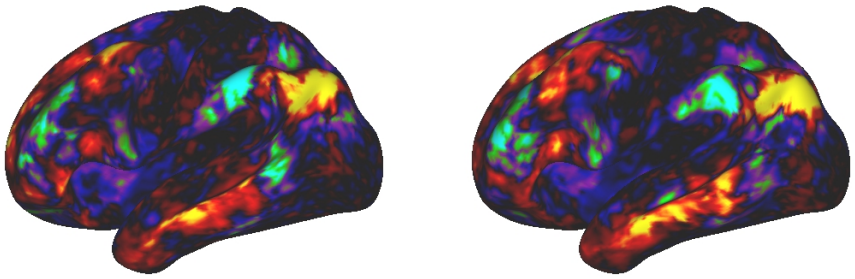
[3]Even accounting for the fact that the model infers a global PFM—as such, the calculation of the partial correlations can be thought of as including a weak version of global signal regression—one would expect the between PFM correlations to have zero mean, rather than being consistently anti-correlated [Murphy et al. 2009]. We discuss global signal regression in more detail in the next chapter (Section 6.2.2).

[4]By construction, the tICA time series are uncorrelated if they are concatenated over all subjects. However, this does not guarantee that the mean over subject-specific partial correlations will be zero.



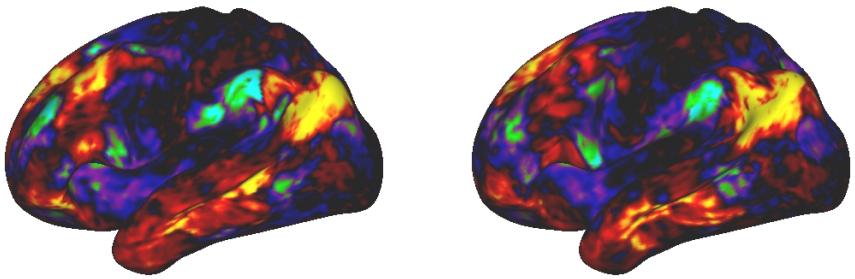
(a) Group-level spatial map.

(b) Group-level variance parameters, σ.



(c) Subject A spatial map.

(d) Subject B spatial map.



(e) Subject C spatial map.

(f) Subject D spatial map.

Figure 5.7: Subject variability of the default mode (PFM 14, Figure 5.1(a)). We display the group-level spatial map, the variance parameters encoded by the PFM model, and four exemplar subject-level spatial maps. Spatial weights are displayed on the left lateral cortical surface. The colour scale for each map is symmetric around zero and is clipped at the 99 % level of the distribution of the spatial weight magnitudes. Cortical surface views were generated using Connectome Workbench (humanconnectome.org/software/connectome-workbench.html).
Labels: dLPFC: dorsolateral prefrontal cortex; IPL: inferior parietal lobule; pTL: posterior temporal lobe.

functionally distinct entity from other PFMs. Secondly, there is enormous variation in the size, shape and location of clusters of spatial weights. For example, note the changing pattern of positive and negative weights in the dorsolateral prefrontal cortex, or the much stronger negative weights in the posterior temporal lobe of subject A. This variability appears to be highly structured, rather than representing a uniformally distributed set of deviations around the group maps.

We can investigate this more quantitatively, as the PFM model has an explicit set of parameters designed to capture spatial non-uniformity of subject variability. We plot these σ parameters in Figure 5.7(b). The most striking feature is the huge ring of cortex exhibiting particularly high variability that surrounds the set of positive weights in the inferior parietal lobule, though the aforementioned posterior temporal lobe and dorsolateral prefrontal cortex also show up strongly. What is interesting is that the variability is relatively low within the regions of the default mode. Rather, it is concentrated around the edges of regions, and as such, appears to be primarily driven by changes in the size, shape and position of cortical areas.

While this is a compelling story for the default mode, this only represents a small proportion of the cortex. To investigate the effects of subject variability more generally, we plot the similarity between the subject-specific maps and the group-level maps for each of the 30 PFMs in Figure 5.8.

What this shows is that the cognitive PFMs seem to be much more variable than those residing in primary visual cortex or the motor strip[5]— a finding that almost exactly matches the results Mueller et al. [2013] reported from their investigations into subject variability. Previous work has speculated about the potential evolutionary and developmental reasons for this heightened variability in associ-

[5]PFMs 2, 3 and 4 reside in primary visual cortex and are highly consistent, and likewise for PFMs 9 and 10 in the motor strip. The other visual PFMs tend to represent higher order visual processing regions, and PFMs 11 and 12 represent a somewhat ambiguous lateralisation of the hand and arm representations in the motor cortex.

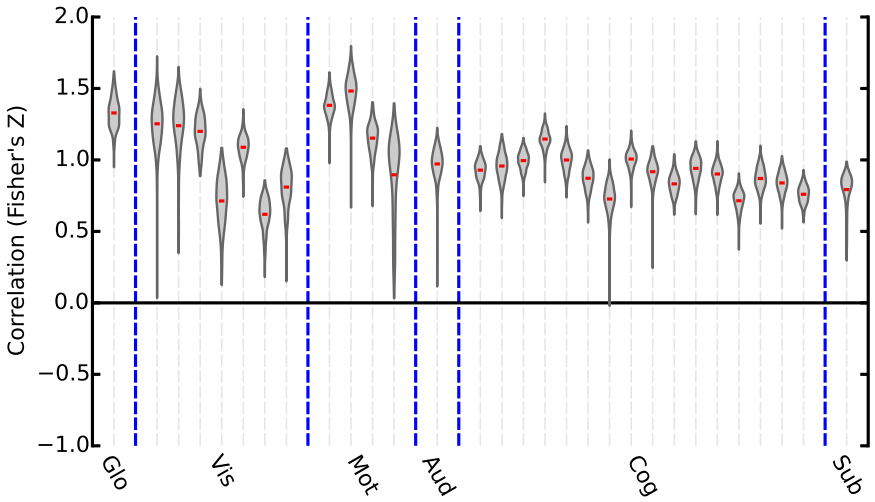


Figure 5.8: Similarity between the subject-specific PFM spatial maps and those at the group level. For each of the 30 PFMs, we plot the distribution of correlations between each of the 461 subject variants and the group-level map. We apply the Fisher transformation, $z = \operatorname{artanh}(\rho)$, to the correlation coefficients.

ation cortex relative to the primary sensory cortices [Buckner and Krienen 2013], and it will be fascinating to see what further insights the PFM approach yields.

In summary, the variability captured by the PFMs seems to reflect a specific type of cortical reorganisation, and is spatially distributed in a manner which is consistent with past findings. However, we can also use the information about the familial relationships between subjects that the HCP collects to validate these findings further, and we investigate the heritability of the subject-specific PFMs in the next section.

5.4.1 Heritability

In an attempt to capture some of the genetic and environmental correlates of neurological variability, the HCP subject cohort contains a large proportion of twins and siblings. While full genetic information will not be available until the end of the project, we can use the information about family structure to do a

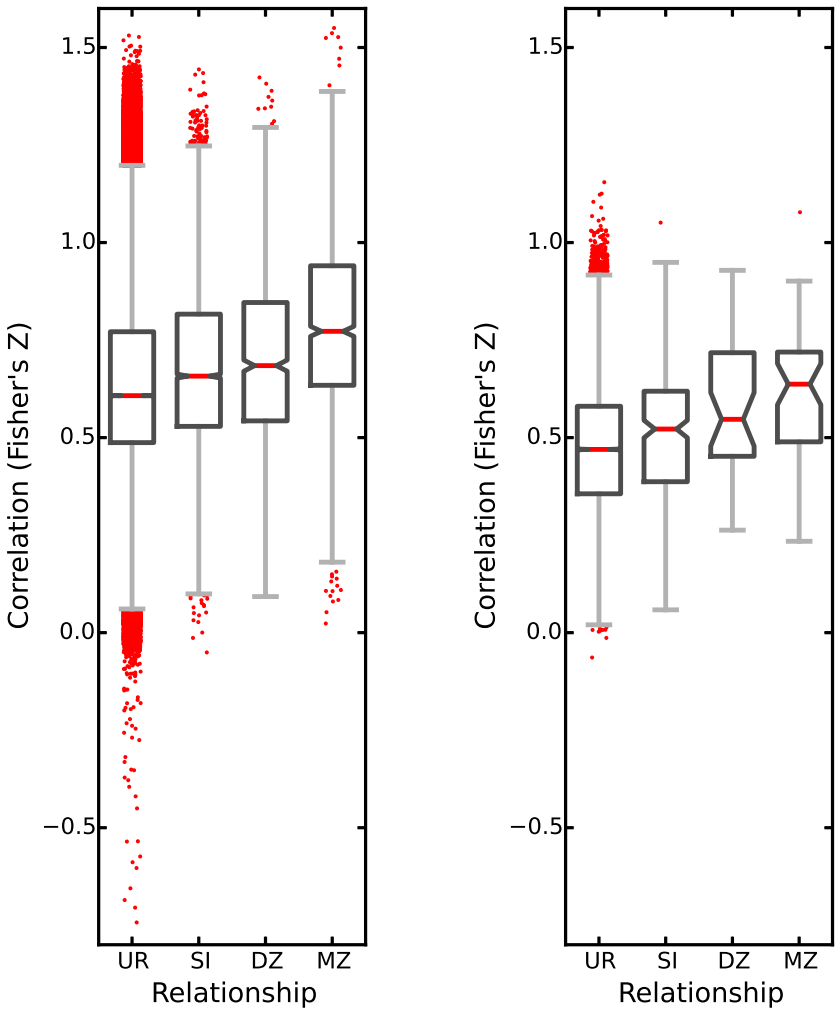
simplified analysis of the heritability of subject information. For each pair of subjects, we calculate the similarity in the subject-specific information encoded by PFMs, and separate this based on the type of familial relationship the subjects have. This is plotted in Figure 5.9.

The information about subjects that PFMs encode is clearly heritable, as closely related subjects have much more similar spatial maps and between-PFM temporal correlations. This is hugely important: it means that a large proportion of the variability we observe is underpinned by genetically and developmentally relevant changes, and does not simply reflect noise. This substantially increases our confidence in that we can extract meaningful information about individual subjects, and in particular that the unexpected temporal interaction structure is underpinned by subject-specific interactions that vary in a meaningful way.

5.4.2 Implications of subject variability

What we have hopefully demonstrated is that the PFMs capture a surprising degree of variability over subjects, but in a way that is consistent with past results and the extra information we have about HCP subjects.

One hypothesis is that this subject-specific modelling is key to reconciling the different predictions that PFMs and other methods make about the temporal interactions between modes. If subject variability in the spatial domain is primarily driven by spatial misalignments and cortical reorganisations, as the PFM results suggest, then this will seriously affect any methods that, like ICA, extract subject-specific mode time courses based on a fixed group template. Intuitively, any spatial shifts relative to the group would mean that the ICA time courses for a given subject would contain a mix of time courses from the 'true' subject-level modes. As the PFMs model also predicts that modes are spatially positively correlated, this type of spatial mixing would mean that the inferred time courses would tend to



(a) Heritability of subject-specific PFM spatial maps.

(b) Heritability of subject-specific PFM temporal partial correlation matrices.

Figure 5.9: Similarity between subject-specific PFM information, stratified by family structure. For each pair of subjects we calculate the similarity between the spatial maps and the temporal interactions and divide this based on the relationship between the two subjects. In the spatial domain, similarity is the correlation between subject-specific PFM spatial maps. In the temporal domain, we calculate the subject-specific partial correlation matrices from the precision matrices, $\boldsymbol{\alpha}^{(s)}$. The similarity is taken as the correlation between the elements in the upper triangles of these matrices, and we apply the Fisher transformation, $z = \operatorname{artanh}(\rho)$, to these correlation coefficients. For these box plots, the whisker extent is 1.5 times the inter-quartile range.
Labels: UR: unrelated; SI: siblings; DZ: dizygotic twins; MZ: monozygotic twins.

be an additive mix of the true time courses. Therefore, this would tend to induce spurious positive temporal correlations between any spatially adjacent modes derived from a group-level template.

While the above represents a conceptual framework within which our surprising results can be reconciled with ICA, there is much work still to do on verifying and comparing these results. However, it seems clear that proper modelling of subject variability will be absolutely crucial to any attempt to understand the complex set of spatio-temporal interactions between modes.

Summary

In this chapter, we have given an overview of the different types of information about modes that the PFM model captures. This approach yields PFMs that are reliable, can often be linked to modes characterised by other techniques, and are consistent with our knowledge about task data and hæmodynamics. However, by inferring modes that are simultaneously strongly spatially and temporally correlated, the PFM approach goes beyond what can be achieved with ICA-based techniques. Finally, it is the novel temporal structure that is particularly surprising, and this leads us to believe that the principled modelling of subject variability is absolutely crucial for any technique that tries to better characterise modes, and that this is therefore a key advantage of the PFM framework.

Chapter 6

Further Analyses

In this chapter we go beyond simply characterising the properties of PFMs, and briefly outline two examples of how we are using PFMs to investigate the properties of rfMRI data in more detail. To begin with, we examine the implications of different registration schemes, before describing a surprising set of observations that links the non-stationary temporal dynamics of PFMs with an ongoing issue of much debate in the rfMRI community, global signal regression.

6.1 Functional Registration

The striking feature of the PFM subject modelling in the last chapter was the extent to which this seemed to capture variability in spatial localisation. As we discussed in the model chapter, there are several reasons why registration driven by gross morphological features may fail to bring subjects into functional alignment. Therefore, it is natural to suspect that using information derived from functional data to drive the registration would improve alignment. This is the rationale behind several recently proposed multi-modal registration approaches [Conroy et al. 2013; Li and Fan 2014; Robinson et al. 2014]. However, if the PFMs are already able to account for the spatial variability in the data, then will PFMs inferred from more

accurately registered data make any different predictions?

In order to better understand the implications of registration accuracy, and the interactions between this and the subject-specific modelling within the PFM framework, we inferred two new sets of PFMs from one data set registered in two different ways. We use data from an earlier HCP release[1], consisting of 209 subjects. Crucially, this data was available with and without functional registration. The data pre-processing pipeline proceeds essentially as described earlier in Section 4.3.1, until the stage at which subject-specific cortical surfaces need to be registered to the group atlas. The standard approach uses FreeSurfer's surface-based registration algorithm informed by sulcal depth alone [Glasser et al. 2013], and we will refer to this as the *structurally registered data*. The alternative approach uses the Multimodal Surface Matching (MSM) algorithm introduced by Robinson et al. [2014], and this performs the registration based on sulcal depth, myelination [Glasser and Van Essen 2011] and a set of subject-specific modes identified from the rfMRI data by sICA and dual regression. We will refer to this as the *functionally registered data*, though it does of course use both structural and functional features. Finally, note that the newer data used in the previous chapter was also structurally registered, but it is not yet available with functional registration applied. However, there have been some tweaks to the pre-processing pipeline that mean that it would not be a fair comparison to use this as the structurally registered data here.

6.1.1 Impact of functional registration

To begin with, we simply try to characterise the effect that functional registration has had on alignment accuracy. Firstly, we plot the PFM that represents the default mode from each of the two analyses, along with the group-level variance parameters and an example subject map. This is shown in Figure 6.1. Several

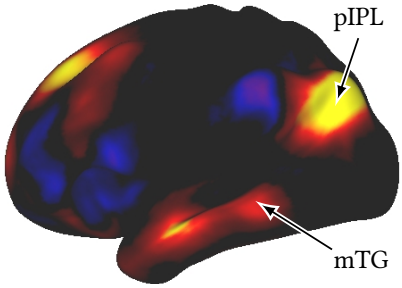
[1] Q3 Data Release: humanconnectome.org/documentation/Q3

things are apparent from a visual inspection of these maps. The first striking result is that the anti-correlated regions are not inferred from the structurally registered data. Secondly, the group-level parameters that capture the spatial non-uniformity of subject variability seem to represent altogether different types of variation. As with the results in the previous chapter, the PFMs inferred from the functionally registered data are highly variable around the edges of regions, as evidenced by the ring of highly variable cortex surrounding the positive spatial weights in the posterior inferior parietal lobule. However, for the structurally registered data the variability seems to be reside within the regions themselves, rather than being concentrated at the edges.

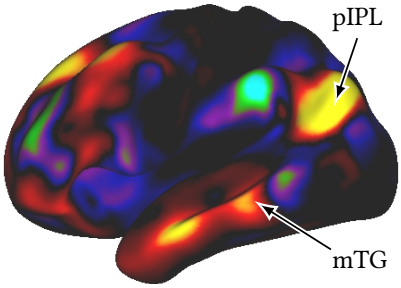
Given the results in the previous chapter, where the data was also structurally registered, it is a particular surprise to observe a hugely different pattern of variability in the structurally registered data here. Our hypothesis is that we will observe this characteristic ring of variability—or in other words, the variability at the centre of a region will be small relative to the variability at the edges—if the region is larger than the spatial shifts between subjects. If that is the case, then the subject-specific modes will all overlap at the centre of the region, despite the spatial shifts, and the variability there would be relatively low. Therefore, the fact that we do not observe this pattern after structural registration could be explained if the spatial shifts were much, much larger in this data.

This does seem to be borne out by the results of our next analysis. In Figure 6.2 we have plotted the similarity between the subject-specific spatial maps and the group-level maps for five cognitive PFMs, and there is a pronounced increase in consistency across subjects after functional registration.

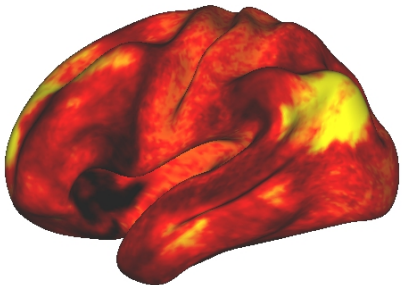
Therefore, it seems likely that this different pattern of variability arises because the spatial misalignments in the structurally registered are markedly larger than those in the functionally registered data. This result is not a surprise, as



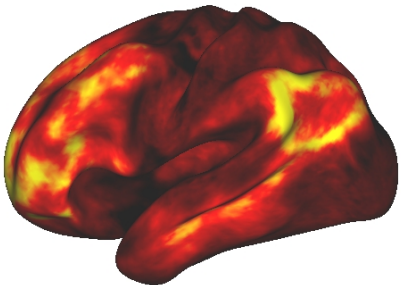
(a) Group-level spatial map, structurally registered data.



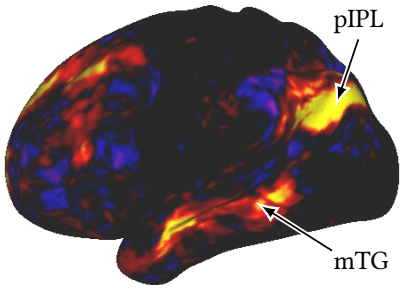
(b) Group-level spatial map, functionally registered data.



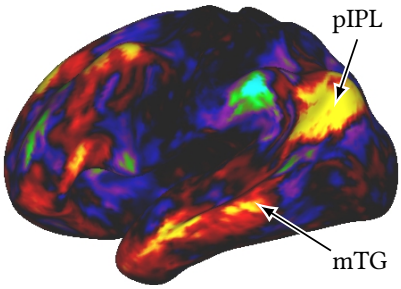
(c) Group-level variance parameters, structurally registered data.



(d) Group-level variance parameters, functionally registered data.



(e) Subject A spatial map, structurally registered data.



(f) Subject A spatial map, functionally registered data.

Figure 6.1: Subject variability of the default mode, as identified from the structural and functional data sets. Structural results are in the left column, with functional on the right. We display the group-level spatial map, the variance parameters encoded by the PFM model, and exemplar subject-level spatial maps from the same subject. Note that these are termed PFMs S1 and F1 respectively, and the full group-level spatial maps can be found in Appendix F. Spatial weights are displayed on the left lateral cortical surface. The colour scale for each map is symmetric around zero and is clipped at the 99 % level of the distribution of the spatial weight magnitudes. Cortical surface views were generated using Connectome Workbench (humanconnectome.org/software/connectome-workbench.html).
Labels: pIPL: posterior inferior parietal lobule; mTG: middle temporal gyrus.

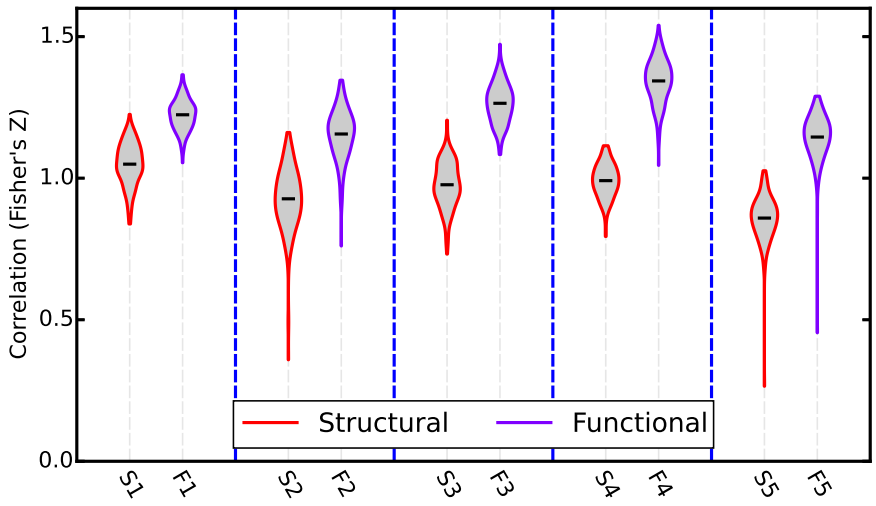


Figure 6.2: Similarity between the subject-specific PFM spatial maps and those at the group level for both the structurally and functionally registered data. This was calculated for five cognitive PFMs, for which the group-level spatial maps can be found in Appendix F. These PFMs correspond to the six PFMs that we highlighted by plotting the group-level spatial maps in the previous chapter (Figure 5.1), with the exception of the language PFM which was not inferred from the structurally registered data. For each of the 5 PFMs, we plot the distribution of correlations between each of the 209 subject-specific spatial maps and the group-level map. We apply the Fisher transformation, $z = \mathrm{artanh}(\rho)$, to the correlation coefficients.

we expected an improvement from functional registration. Rather, it is the size of the improvement that we did not expect. It appears that the misalignments in the structurally registered data are of a similar spatial scale to a region as big as the group of weights in the posterior inferior parietal lobule, whereas they are substantially diminished by functional registration. However, note that improvements from functional registration are not solely confined to the alignment of large-scale features. The increased strength of the group-level weights—and the improved localisation of the subject weights relative to these—in the area of the middle temporal gyrus we have highlighted imply that this is operating at a relatively fine spatial scale. Finally, our suspicion is that the heightened level of variability in the structurally registered data is responsible for the disappearance

of the anti-correlated regions[2].

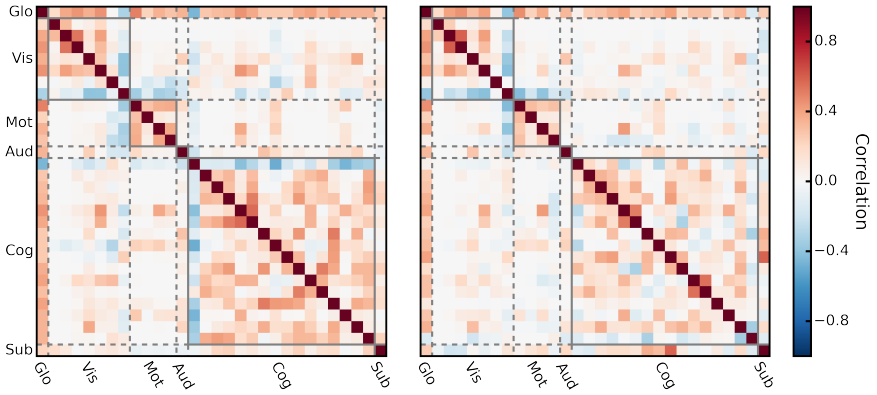
However, it is not our intention that this explanation for our observations be used to traduce structural registration. Rather, it should be emphasised that functional registration, as implemented within the multimodal surface matching framework, has yielded a remarkable improvement in inter-subject spatial alignment.

6.1.2 Interaction with PFM subject modelling

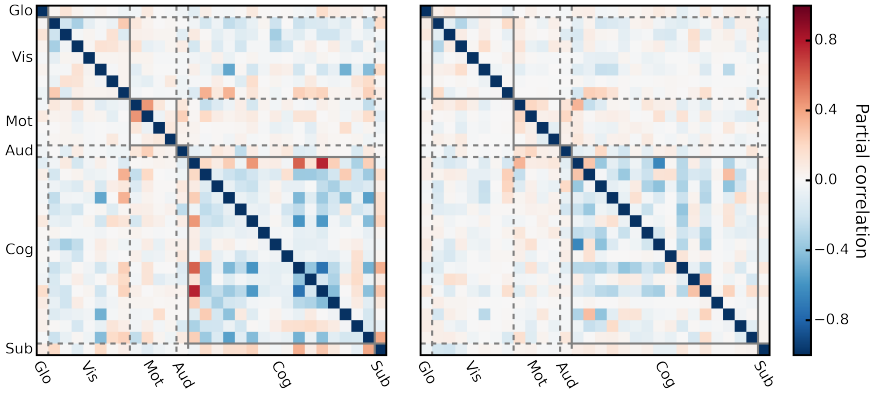
The major impact of functional registration on the PFMs is to reduce the uncertainty over the spatial localisation of regions. However, there are also a number of more subtle effects, like the disappearance of the anti-correlated regions in the default mode, that suggest that functional registration is changing the PFMs themselves, rather than simply their associated uncertainties. To investigate this further we plot the spatial and temporal interactions between the PFMs inferred from the structurally and functionally registered data in Figure 6.3. The equivalent matrices from the analysis in the previous chapter can be found in Figure 5.6.

The spatial interactions are similar to what we have seen before, namely that within the groups the PFMs are positively correlated with one another. The correlations look slightly weaker for the functionally registered data, and this is consistent with regions being better delineated at the group level if there is less variability in location at the subject level. However, in the temporal domain, the partial correlations between the PFMs inferred from the structurally registered data seem to be much less consistent than between the PFMs extracted from the functionally registered data. Anti-correlations are still predominant, but there is a greater range of strengths, as well as stronger between-group correlations, in the interactions inferred from the structurally registered data. Qualitatively,

[2]The reduced number of subjects, even relative to the split-half analysis in the previous chapter, will also have an effect.



(a) Correlations between group-level spatial maps. **Left:** structurally registered data; **Right:** functionally registered data.



(b) Partial correlations between time courses. **Left:** structurally registered data; **Right:** functionally registered data.

Figure 6.3: Spatial and temporal correlations between PFMs, as derived from the structurally and functionally registered data. These PFMs were matched to the set of 30 PFMs examined in the previous chapter, and were matched based on similarity between the group-level spatial maps. The spatial correlations were calculated from the group-level spatial maps. The PFM temporal partial correlations were calculated from the group-level hyperprior on subject-specific temporal precision matrices, $\boldsymbol{\beta}$.

the temporal correlation structure extracted from the functionally registered data seems to be a closer match to the results from the equivalent analysis in the previous chapter, where over twice as many subjects were used.

Therefore, it appears that while the gross properties of PFMs that we outlined in the previous chapter can be seen in both data sets, registration does induce some subtle changes to the PFMs. Our interpretation is that while PFMs can correct for subject variability to a large extent, there is only so much that they can do. Better quality data goes beyond simply reducing uncertainty in the PFM model, and seems to offer some improvements to the PFM decomposition generally.

6.1.3 Implications

In summary, in this section we have demonstrated that using functional information to drive the registration profoundly affects data quality by vastly reducing the observed between-subject variability. While the subject-specific modelling at the heart of the PFMs framework can ameliorate misalignments to a large extent, there are still some marked differences between the PFMs inferred from structurally and functionally registered data.

However, we believe that improved registration will have a much more profound effect than this on PFMs: it should allow us to identify more modes. While the resolution at which the HCP rfMRI data acquired is 2 mm, it is our view that the effective functional resolution—by which we mean the scale at which a distinct functional unit can be resolved across subjects[3]—is far larger than this, and therefore, our interpretation is that registration accuracy represents the current spatial limitation for group studies.

In their discussion of the work of Saygin et al. [2012], which predicted the spatial localisation of functional regions using structural connectivity, Jbabdi and

[3] In fact, for some parts of the cortex, we already have a simple way of estimating this: this is simply the width of the ring of the high variability region we observe around high spatial weights.

Behrens [2012] observed that '[b]y mapping onto a purely spatial template, we lose a great deal of detail that is present in individual responses, and we are left to interpret only the spatial peaks that are consistent across subjects.' A crucial advantage of the PFM framework is that if these spatially consistent peaks do occur, then it is also possible to reconstruct the individual responses. Intuitively though, this approach breaks down if the misalignments are larger than the spatial scale of the modes we want to identify. However, if functional registration reduces the scale of misalignments between subjects, then the expectation is that more spatially consistent peaks will appear, and that this will naturally lead to a larger set of functional modes being identified.

6.2 Temporal Non-Stationarities & the Global Signal

6.2.1 Non-stationary temporal dynamics

There has been much recent interest in the dynamic nature of functional connectivity [Hutchison et al. 2013]. Rather than simply investigating the interactions between regions as inferred from the entire scan duration, several methods have been proposed that try and characterise the time-varying properties of functional connectivity [Chang and Glover 2010; Cribben et al. 2012; E. A. Allen et al. 2014; Zalesky et al. 2014], as well as several attempts to explain how this property arises [Baker et al. 2014; Hansen et al. 2015]. But what can PFMs tell us about these non-stationarities?

Our interest in this was piqued by a surprisingly simple analysis of the PFM time courses. The aim was to investigate if there were any systematic changes in PFM activity over the duration of a typical rfMRI scan. To do this, we took each PFM in turn and, for each time point, collected the observed values of the BOLD time courses for all the runs analysed. We calculated the variance of this

distribution to give an estimate of the typical level of activity of PFM m at time point t, which we denote $\mathrm{Var}(\boldsymbol{B}_{mt})$, and this can be calculated as per the equation below.

$$\mathrm{Var}(\boldsymbol{B}_{mt}) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{R_s}\sum_{r=1}^{R_s}\left(\boldsymbol{B}_{mt}^{(sr)} - \overline{\boldsymbol{B}_{mt}}\right)^2 \tag{6.1}$$

This variance gives us a dynamic estimate of the activity of each PFM as a function of the time subjects had spent being scanned.

The remarkable result is that there seems to be a clear dissociation between these time-resolved variance estimates for the cognitive PFMs and those residing in sensory cortices. This is shown in Figure 6.4. The activity of the sensory PFMs increases markedly over the course of the scan, whereas the activity in the cognitive PFMs is essentially constant, apart from a sharp increase during the first minute.

That there should such a clear separation by such a simple measure of non-stationarity is a huge surprise. To investigate this further, we repeated the analysis but this time on the raw data, rather than the PFM time courses. We calculate a time-resolved variance estimate for each grayordinate, $\mathrm{Var}(\boldsymbol{D}_{vm})$, which can again be calculated from Equation 6.1. We form a matrix of these spatially and temporally resolved variances, before decomposing this via the SVD. The spatial map and time course of the first component of this—or in other words, the first columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ respectively—are plotted in Figures 6.5(a) and 6.5(c) respectively.

The first SVD component seems to independently verify the observations from the PFM time courses. The SVD time course shows a consistent increase in variance over the scan length, and the spatial map is confined to primary visual and primary motor cortex, with a weaker recruitment of primary auditory cortex. What is notable is how similar this spatial pattern is to the global PFM (Figure 6.5(b)). Similarly, if we look back to the PFM time course variances (Figure 6.4) then we can see that the global PFM has one of the strongest trends for increasing variance over the scan, and that the strength of this signal in sensory regions explains the

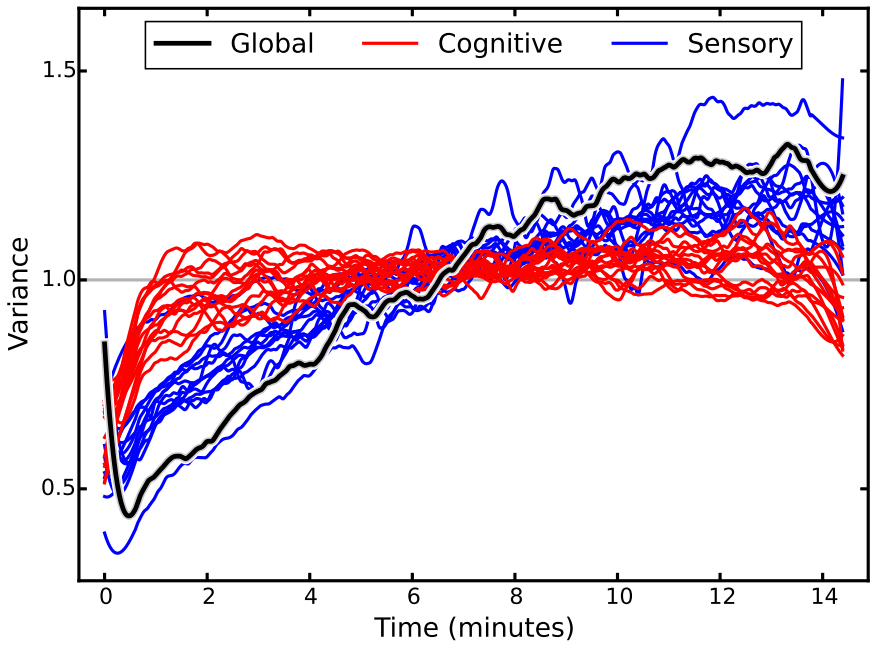
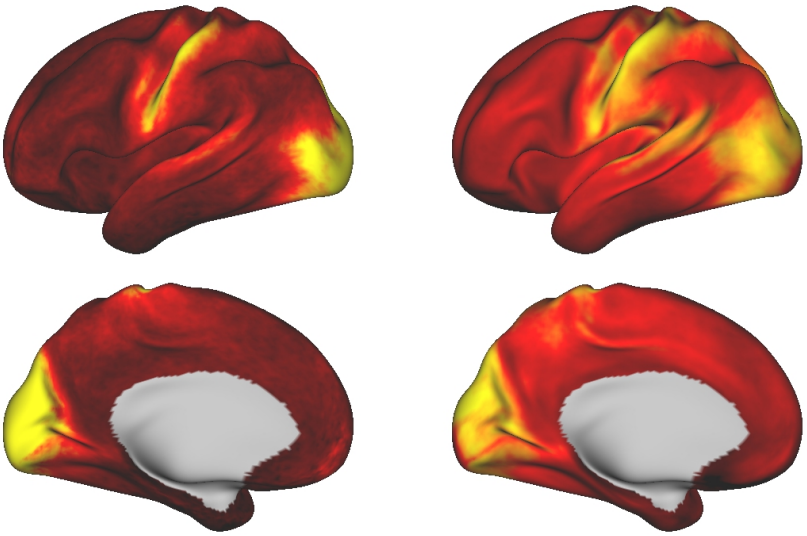


Figure 6.4: Variance changes over the course of an rfMRI scan for the PFM time courses. For each of the 30 PFMs analysed in Chapter 5, we calculate the time-resolved variance of the inferred BOLD time courses over all 1,844 runs. These variance time courses are rescaled and smoothed with a Savitzky-Golay filter (order=3, window length=101) before plotting [Savitzky and Golay 1964]. The sensory category includes all the PFMs from the visual, motor, auditory and subcortical groups.

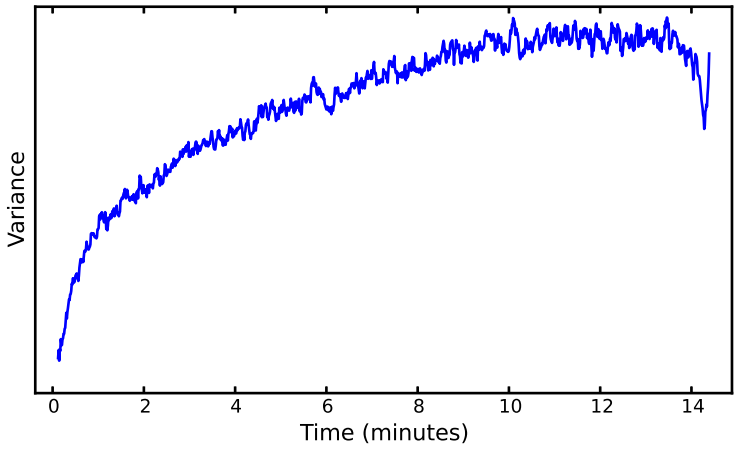
dissociation between cognitive and sensory PFMs.

This is an interesting result in and of itself. As far as we can tell, this spatial non-uniformity of temporal non-stationarities has not been reported before, and could have a profound effect on estimates of dynamic functional connectivity. This is because, as Friston [2011] notes, it is possible to induce '[a] difference in [functional connectivity] by changing the amplitude of neuronal fluctuations without changing the coupling'. However, the relationship with the global PFM is of particular interest, as at the moment the global signal is a particularly contentious topic within the fMRI community. We give a quick overview of the debates around the global signal in the next section, and in the final section we describe why we believe these non-stationarities are an important contribution to the discussion.



(a) Spatial map of the first component from the variance SVD.

(b) The global mode (PFM 1).



(c) Time course of the first component from the variance SVD.

Figure 6.5: Variance changes over the course of an rfMRI scan for the raw HCP data. For each grayordinate we calculate the time-resolved variance of the rfMRI signal over all 1,844 runs. The signal variance, pooled over time and runs, is set to unity in each grayordinate and these spatio-temporally resolved variance estimates are arranged into a matrix of size: number of grayordinates × time points. We calculate the SVD of this matrix after removing the mean, and we plot the spatial map and time course of the first SVD component. For comparison, we also show the spatial map of the global mode (PFM 1) as inferred from the same data. Spatial weights are displayed on the left cortical surface. The colour scale for each map is symmetric around zero and is clipped at the 97 % and 99.5 % levels of the respective distributions of the spatial weight magnitudes. Cortical surface views were generated using Connectome Workbench (humanconnectome.org/software/connectome-workbench.html).

Our ultimate aim is to use these observations, in conjunction with the existing literature, to better understand what the global signal might represent.

6.2.2 The fMRI global signal

If one takes the mean over all grey matter voxels in an fMRI data set, a low-frequency time course emerges—the global signal. While this signal is present across the entire grey matter, it is expressed most strongly in the primary sensory cortices [Fox et al. 2009], in a manner that is consistent with the spatial distribution of our global PFM. Empirically, regressing the global signal out of the data improves the spatial specificity of functional connectivity analyses [Fox et al. 2009; Mennes and Beckmann 2015]. This regression step is not necessarily as drastic as it sounds—for example, it is common practice to regress the mean white matter signal out of the data during the pre-processing pipeline.

However, there is good evidence that the global signal is neural in origin [Schölvinck et al. 2010]. That being the case, then at best global signal regression removes interesting signal, and at worst it fundamentally changes our interpretation of the data. The main controversy centres around whether or not any temporal anti-correlations observed after removing the global signal are a neuroscientifically meaningless consequence of this regression. By definition, if all voxels contain a portion of the same signal then they will be positively correlated, and regressing that common signal out will remove the mean correlation [Murphy et al. 2009]. There will now be anti-correlations present in the data, but are they interpretable? Proponents have argued that this simply enhances anti-correlations that were already present in the data [Fox et al. 2009], but from a mathematical standpoint, the interpretation of correlations after regressing out the mean signal is far from clear [D. M. Cole et al. 2010].

There is, however, a crucial distinction to be made here. If, as Mennes and

Beckmann [2015] recently proposed, the global signal simply represents the mean mode activity then the concerns that D. M. Cole et al. [2010] raised about the interpretability of correlations after the regression are valid. However, just because a signal can be identified from the mean over voxels, this does not imply that it is simply a mathematical construct. The other interpretation is that the global signal represents a neurophysiologically distinct process from the resting-state fluctuations in mode activity, but one that, by virtue of appearing in more of the cortex than any one mode, can be more reliably identified from a simple mean operation. In that case, regressing it out is perfectly valid, in the same way that it is valid to take the partial correlation between a set of modes, or regress out the white matter signal.

Finally, note that the our results are consistent with the latter interpretation: a global PFM, distinct from any other modes, is identified. This is in contrast to the prediction that the work of Mennes and Beckmann [2015] makes, which is that this activity would be subsumed by the other modes. Furthermore, because the global PFM is inferred alongside the other modes, rather than from the mean over voxels, it should not contain the mean activity from the other PFMs. This means that the concerns that D. M. Cole et al. [2010] raised, about regressing out the mean specifically, are less of a concern for the global PFM.

6.2.3 Speculation

We have identified a pattern of temporal non-stationarities that almost exactly matches the spatial distribution of the global PFM. But is there a plausible reason to believe that the global signal would be particularly non-stationary?

Two ECoG studies on monkeys provide the crucial piece of information. In the first study, Liu et al. [2015a] identified a consistent, cortically distributed sequence of neural activity that occurred sporadically during both wake and sleep, but that

this was less pronounced and occurred less regularly when the monkeys were awake. Furthermore, this pattern seemed to relate to the transition to sleep, as it could be induced by the administration of an anæsthetic. In the second study, Liu et al. [2015b] performed a similar set of experiments, but this time using a combined fMRI and ECoG paradigm, and demonstrated that this sequence of events coincided with peaks in the fMRI global signal. Taken together, these studies have claimed to identify a neural basis for the fMRI global signal, with the key prediction that this will be more prevalent during sleep.

Finally, we can link the global signal back to our observed non-stationarities. As we mentioned at the start of the introduction, Tagliazucchi and Laufs [2014] have demonstrated that during rfMRI scans subjects tend to fall asleep, and increasingly so as the scan progresses. Therefore, if the global signal does indeed increase as alertness decreases, then the steadily increasing activity over the course of the scan that we have observed may simply represent the increasing prevalence of subjects drifting in and out of light sleep[4].

This is somewhat counter-intuitive, as it might be expected that a transition to sleep would be associated with a reduction in the variance of cognitive regions. However, mode activity persists even while asleep [Horovitz et al. 2009; Sämann et al. 2011], and the HCP rfMRI acquisition is specifically broken into four runs to prevent subjects falling into a deep sleep. Rather, what we are seeing here is only the very early stages of sleep, and the implication is that this is primarily associated with a reduction in the coupling between sensory inputs and the activity in sensory regions[5]. In other words, a reduction in feedback in the early stages of sleep leads to increased activity in primary visual, motor and auditory cortices,

[4]Tagliazucchi and Laufs [2014] also observed an increase in BOLD signal variance in visual and motor cortices during N1 and N2 sleep, again consistent with our hypothesis.

[5]The implication from functional connectivity is that the major difference between wakefulness and the earliest stages of sleep is a reduction in cortico-thalamic connectivity [Spoormaker et al. 2010].

rather than a reduction of activity in cognitive regions.

In summary, we have identified a type of temporal non-stationarity that manifests itself as variance increases over the scan duration, and this is most strongly expressed in primary sensory cortices—a pattern of expression that is strikingly similar to the spatial distribution of the fMRI global signal. Our suspicion is that this relates to arousal changes. In the absence of an overt task, there is a strong tendency for subjects to fall asleep over the course of an rfMRI scan, and as previous work has indicated that the global signal is modulated by alertness, this suggests a natural mechanism by which global signal variance would increase over time. If true, this implies that the global signal represents a neurophysiologically distinct process from the fluctuations in mode activity, and this could potentially explain the empirical observation that global signal regression improves the specificity of functional connectivity analyses[6].

[6]Similarly, this would also explain why PFMs categorise the global signal as distinct from other modes.

Chapter 7

Conclusions & Future Work

In this chapter, we give a brief summary of the main characteristics of the PFM framework that we have developed, before outlining some key lines of enquiry that we believe will offer fecund ground for future investigations.

7.1 Conclusions

In this thesis, we have developed a new model, and inference framework, for characterising functional modes from resting-state fMRI data. Our stated aims were to identify a set of modes without making restrictive assumptions about their spatio-temporal relationships, and to do this in a way that allowed us to capture the way these modes varied across subjects.

In the Results chapter we explored the properties of a set of PFMs inferred from data collected as part of the Human Connectome Project, containing over 450 subjects. Our analyses demonstrated that, at the group-level, PFMs were consistent with current knowledge of the macro-scale functional organisation of the brain, and furthermore, this held for insights gained from both previous resting-state and task studies. However, the PFMs neither represent a direct replication of previous results nor reflect the new results from previous mode identification

techniques that are made possible by the unprecedented quality and quantity of the HCP data. In particular, they represent a genuinely different characterisation of functional modes from ICA, the technique that is currently dominant within the rfMRI community. In addition, despite the complexity of the hierarchical model that is inferred from the data, the PFMs can be reliably identified.

However, it is the modelling of between-mode interactions and subject variability that enables the PFMs to offer novel insights into the properties of the brain at rest. The scale of the variability between subjects that the PFMs capture, and the extent to which this seems to reflect variability in the spatial localisation of functional regions, is a surprise and has profound implications for attempts to characterise functional organisation at fine spatial scales using fMRI. That being said, the fact that aspects of this variability are heritable, and the manner in which it decreases after improvements to the registration process, give us confidence that the variability captured by the PFMs genuinely reflects—and, therefore, ameliorates the impact of—cortical reorganisations. This, combined with the ability to decouple noise from the temporal activity that arises as a result of hæmodynamic processes, leads us to believe that the temporal information that we extract from PFMs more closely reflects the activity of the underlying subject-specific modes than the time courses from any technique that makes its inferences from a set of group-derived modes. Therefore, it is the observation that the cognitive PFMs are almost exclusively temporal anti-correlated with one another that represents the most radical departure from our current understanding of functional modes and it will be fascinating to investigate the implications of this further.

Then, in Chapter 6, we explored some of the more subtle properties of rfMRI data. Using PFMs, we demonstrated that functional registration offers huge improvements in the quality of inter-subject alignment, and we suspect that this will be crucial to attempts to characterise more modes in the future. Finally, inspired by

a surprising observation of non-stationarity in the PFM time courses, we demonstrated that there is a striking relationship between the spatial distribution of the global signal and the regions that exhibit a consistent increase in the variance of the BOLD signal over the course of an rfMRI scan, and suggested a mechanism by which these two might be linked. Our hope is that these observations shed fresh light on what is an increasingly contentious debate.

7.2 Future Work

In the first instance, there is much work to be done on the results that have been presented here. The HCP includes a wealth of information about subjects beyond the neuroimaging data, and with the PFMs capturing so much subject variability, this is an obvious area of investigation. As Smith et al. [2015] elegantly demonstrated, resting-state functional connectivity can be related back to a multitude of external factors. It would be remiss of us not to investigate the behavioural correlates of the subject-level PFMs, or what different predictions this makes to the simpler dual regression based models of subject variability.

At an algorithmic level, we believe that the way in which we initialise the PFMs could be improved, and that this would lead to much more reliable decompositions. As with all VB approaches, our approach will only ever be guaranteed to converge to a local optimum, and as such, initialisation will inevitably have some bearing on the final set of PFMs. As we calculate the subject-specific SVDs to reduce the memory footprint of the data, we have already performed the first stage of many group ICA approaches [Calhoun et al. 2001; Varoquaux et al. 2010]. If we combine the subject-specific spatial singular vectors, then we can construct an initial estimate of the subspace in which we expect the PFM group-level spatial maps to reside—we could even run sICA if we were so inclined! Our expectation is that this would be a more effective way of initialising the PFMs than the random

procedure we currently use.

In the medium term, the HCP is set to release another tranche of data, increasing the number of subjects available to over 900, and crucially, this will all be functionally registered. As we discussed earlier, the combination of more and better data is likely to have a profound impact on the quality of the PFM decomposition. While this represents a significant computational challenge, the benefits are likely to be enormous.

On a related note, given that the PFMs capture spatial misalignments between subjects, it would be an obvious next step to drive the functional registration process using the subject-specific PFM spatial maps. Furthermore, it will be fascinating to see if PFMs can tell us about the limits of current registration techniques based on diffeomorphic warps, and the extent to which the currently observed spatial misalignments represent topological reorganisations.

Finally, the ultimate aim here is to better characterise modes. The PFMs are already making some new predictions about the structure of modes, their complex spatio-temporal interactions, and their non-stationary temporal dynamics. Our hope is that, if PFMs do indeed represent a more accurate description of modes than currently exists, they facilitate attempts to better understand the functional organisation of the brain.

Appendix A

Comparison of PFMs and ICA

The definitions in this appendix are all standard and can be found from many sources. However, the notational conventions are, for the most part, borrowed from *Independent Component Analysis* by Hyvärinen et al. [2001b] and *Information Theory, Inference and Learning Algorithms* by MacKay [2003].

A.1 Definitions

A.1.1 Independence

Random variables X and Y are independent if and only if the joint probability density factorises, so

$$\mathrm{p}_{X,Y}(x, y) = \mathrm{p}_X(x)\ \mathrm{p}_Y(y) \tag{A.1}$$

Therefore, for any absolutely integrable functions f and g, independent variables satisfy the following:

$$\mathrm{E}\big[\mathrm{f}(X)\,\mathrm{g}(Y)\big] = \mathrm{E}\big[\mathrm{f}(X)\big]\,\mathrm{E}\big[\mathrm{g}(Y)\big] \tag{A.2}$$

This can be extended to more than two variables, and random vectors, to give

$$\mathrm{p}_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\ldots}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\ldots) = \mathrm{p}_{\boldsymbol{X}}(\boldsymbol{x})\ \mathrm{p}_{\boldsymbol{Y}}(\boldsymbol{y})\ \mathrm{p}_{\boldsymbol{Z}}(\boldsymbol{z})\ldots \tag{A.3}$$

Clearly, 'the components of $\boldsymbol{x}$ can be mutually dependent, while they are independent with respect to the components of the other random vectors $\boldsymbol{y}$ and $\boldsymbol{z}$' [Hyvärinen et al. 2001b].

A.1.2 Similarity, correlation and covariance

In the scalar case, we will adopt the following definitions. Let the 'similarity' be the expectation of the product of the variables,

$$\mathrm{sim}(X,Y) = r_{X,Y} = \mathrm{E}[XY]\ , \tag{A.4}$$

and the covariance be the version of this after the variables have been demeaned,

$$\begin{aligned}\mathrm{cov}(X,Y) = c_{X,Y} &= \mathrm{E}\big[\big(X-\mathrm{E}[X]\big)\big(Y-\mathrm{E}[Y]\big)\big] \\ &= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] \\ &= r_{X,Y} - \mathrm{E}[X]\,\mathrm{E}[Y]\ .\end{aligned} \tag{A.5}$$

Correlation is simply the covariance normalised by the standard deviations of X and Y,

$$\mathrm{corr}(X,Y) = \rho_{X,Y} = \frac{c_{X,Y}}{\sqrt{c_{X,X}c_{Y,Y}}} \tag{A.6}$$

In the vector case, the definitions are analogous. For random vectors $\boldsymbol{X} \in \mathbb{R}^{K_X}$ and $\boldsymbol{Y} \in \mathbb{R}^{K_Y}$, with mean vectors $\boldsymbol{m}_{\boldsymbol{X}} = \mathrm{E}[\boldsymbol{X}]$ and $\boldsymbol{m}_{\boldsymbol{Y}} = \mathrm{E}[\boldsymbol{Y}]$ respectively, then

we can define the similarity and covariance as:

$$\boldsymbol{R}_{\boldsymbol{X},\boldsymbol{Y}} = \mathrm{E}\left[\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\right] \tag{A.7}$$

$$\begin{aligned} \boldsymbol{C}_{\boldsymbol{X},\boldsymbol{Y}} &= \mathrm{E}\left[\left(\boldsymbol{X} - \mathrm{E}[\boldsymbol{X}]\right)\left(\boldsymbol{Y} - \mathrm{E}[\boldsymbol{Y}]\right)^{\mathrm{T}}\right] \\ &= \boldsymbol{R}_{\boldsymbol{X},\boldsymbol{Y}} - \boldsymbol{m}_{\boldsymbol{X}}\boldsymbol{m}_{\boldsymbol{Y}}^{\mathrm{T}} \end{aligned} \tag{A.8}$$

Note that in general these will not be square matrices as $\boldsymbol{R}_{\boldsymbol{X},\boldsymbol{Y}}, \boldsymbol{C}_{\boldsymbol{X},\boldsymbol{Y}} \in \mathbb{R}^{K_X \times K_Y}$.

A.1.3 Results from finite samples

The above results require enough to be known about the distributions of the random variables for the various expectations to be calculated. An alternative scenario is where we observe multiple realisations of the random variables and wish to estimate the expectations directly from the observations. As this frequently uses the same notation for slightly different quantities, we will explicitly provide some results here.

Let $\tilde{\boldsymbol{X}} \in \mathbb{R}^N$ be a *row* vector consisting of n random realisations of X,

$$\tilde{\boldsymbol{X}} = (x_1, x_2, \ldots, x_N) \ . \tag{A.9}$$

We can average over these N samples to estimate the expectations, as

$$\mathrm{E}\left[\mathrm{f}(X)\right] \approx \frac{1}{N}\sum_{n=1}^{N} \mathrm{f}(x_n) \ . \tag{A.10}$$

For example, if we define $\tilde{\boldsymbol{Y}}$ similarly to $\tilde{\boldsymbol{X}}$ then we can compute

$$\begin{aligned} r_{X,Y} &= \mathrm{E}[XY] \\ \hat{r}_{X,Y} &= \frac{1}{N}\sum_{n=1}^{N} x_n y_n \\ &= \frac{1}{N}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Y}}^{\mathrm{T}} \end{aligned} \tag{A.11}$$

The extension to the case where we have random vectors requires that we now collect matrices of random samples. Let $\tilde{\boldsymbol{X}} \in \mathbb{R}^{K_X \times N}$ be a matrix consisting of N realisations of $\boldsymbol{X} \in \mathbb{R}^{K_X}$,

$$\tilde{\boldsymbol{X}} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) \ , \tag{A.12}$$

and let $\tilde{\boldsymbol{Y}} \in \mathbb{R}^{K_Y \times N}$ be similarly defined for $\boldsymbol{Y}$. Then, for example,

$$\begin{aligned} \boldsymbol{R}_{X,Y} &= \mathrm{E}\left[\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\right] \\ \hat{\boldsymbol{R}}_{X,Y} &= \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{y}_n^{\mathrm{T}} \\ &= \frac{1}{N}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Y}}^{\mathrm{T}} \ . \end{aligned} \tag{A.13}$$

Note that if two rows of $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{Y}}$ are orthogonal then the estimated similarity between them is 0.

A.2 ICA model

ICA assumes that an observation, $\boldsymbol{d}$, is a linear mixture of underlying sources, $\boldsymbol{x}$. The observation $\boldsymbol{d}$ is a realisation of a random variable $\boldsymbol{D} \in \mathbb{R}^{K_D}$, and similarly for $\boldsymbol{x}$ and $\boldsymbol{X} \in \mathbb{R}^{K_X}$. For simplicity, the dimensionality of $\boldsymbol{D}$ and $\boldsymbol{X}$ are often assumed to be the same i.e. $K_{\boldsymbol{D}} = K_{\boldsymbol{X}}$. The linear mixing is defined by the mixing matrix,

$\boldsymbol{G}$, so

$$\boldsymbol{d} = \boldsymbol{G}\,\boldsymbol{x}\ . \tag{A.14}$$

The aim is to recover $\boldsymbol{x}$, which will require the inversion of $\boldsymbol{G}$. This is possible with essentially only two assumptions: firstly, each of the elements of $\boldsymbol{X}$ are generated independently; secondly, these are drawn from non-Gaussian distributions.

$$\mathrm{p}(\boldsymbol{X}) = \prod_{k=1}^{K_X} \mathrm{p}_k(\boldsymbol{X}_k) \tag{A.15}$$

In reality, we only observe a finite number of observations, $\{\boldsymbol{d}_n\}_{n=1}^{N}$, and wish to recover the set of corresponding sources, $\{\boldsymbol{x}_n\}_{n=1}^{N}$. We can use the sample notation, introduced in Section A.1.3, to formulate the inference problem in this finite case as a matrix factorisation.

$$\tilde{\boldsymbol{D}} = \boldsymbol{G}\,\tilde{\boldsymbol{X}}\ . \tag{A.16}$$

A.2.1 Correlations

By construction, the generative model states that the covariance between different sources is zero. However, this is only a very weak statement; in reality, we observe a finite set of observations and it is the correlations between the sources that we observe, given the data, that are arguably of more interest in the case of fMRI, as we shall see later.

$$\begin{aligned} \hat{\boldsymbol{R}}_{\boldsymbol{X},\boldsymbol{X}} &= \frac{1}{N}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}} \\ &= \frac{1}{N}\boldsymbol{G}^{-1}\hat{\boldsymbol{R}}_{\boldsymbol{D},\boldsymbol{D}}\boldsymbol{G}^{-\mathrm{T}} \end{aligned} \tag{A.17}$$

We can actually say surprisingly little about $\hat{\boldsymbol{R}}_{\boldsymbol{X},\boldsymbol{X}}$ if we only make the two previously stated assumptions that underpin ICA.

A.2.2 Whitening

Whitening is a standard pre-processing technique for ICA algorithms, as it reduces the number of parameters to infer. A random variable is white if its covariance matrix is equal to the identity, and it is always possible to transform variables to this form. The following explains why this is useful and the implications for the sample correlations between components.

First, assume we have a whitened form of the data, $\boldsymbol{Z}$, such that $\boldsymbol{C}_{Z,Z} = \boldsymbol{I}$. This is achieved with a whitening matrix, $\boldsymbol{W}$, such that $\boldsymbol{z}_n = \boldsymbol{W}\boldsymbol{d}_n$. Equation A.14 then becomes

$$\begin{aligned} \boldsymbol{z}_n &= \boldsymbol{W}\,\boldsymbol{G}\,\boldsymbol{x}_n \\ &= \boldsymbol{G}'\,\boldsymbol{x}_n \ . \end{aligned} \tag{A.18}$$

To see why this is a useful transformation, we need to calculate the covariance of Equation A.18. This yields

$$\boldsymbol{C}_{Z,Z} = \boldsymbol{G}'\,\boldsymbol{C}_{X,X}\,\boldsymbol{G}'^{\mathrm{T}} \ . \tag{A.19}$$

Under the generative model $\boldsymbol{C}_{X,X} = \boldsymbol{I}$, and $\boldsymbol{C}_{Z,Z} = \boldsymbol{I}$ by definition, so this simplifies to

$$\boldsymbol{G}'\boldsymbol{G}'^{\mathrm{T}} = \boldsymbol{I} \ , \tag{A.20}$$

or in other words, $\boldsymbol{G}'$ is an orthogonal matrix.

A good rule of thumb is that an orthogonal matrix has only half the degrees of freedom of an arbitrary matrix, so restricting the search for $\boldsymbol{G}'$ to the space of orthogonal matrices will be much more efficient than inferring $\boldsymbol{G}$.

In practice, whitening is nearly always achieved by taking the SVD, $\tilde{\boldsymbol{D}} = \tilde{\boldsymbol{U}}\,\tilde{\boldsymbol{S}}\tilde{\boldsymbol{V}}^{\mathrm{T}}$, and setting $\tilde{\boldsymbol{Z}} = \tilde{\boldsymbol{V}}^{\mathrm{T}}$. Dimensionality reduction can be incorporated by

only selecting a subset of the SVD components. Technically, the SVD actually ensures that $\hat{\boldsymbol{C}}_{\boldsymbol{Z},\boldsymbol{Z}} = \boldsymbol{I}$, as we only observe N data points rather than the 'true' distribution for $\boldsymbol{Z}$. Given that we still restrict $\boldsymbol{G}'$ to be orthogonal, it is trivial to show that this now ensures $\hat{\boldsymbol{C}}_{\boldsymbol{X},\boldsymbol{X}} = \boldsymbol{I}$.

In summary, it makes algorithmic sense to whiten the data and infer an orthogonal mixing matrix, and this ensures that the set of sources are uncorrelated under the sample definition.

A.2.3 ICA and fMRI

When used for the analysis of fMRI data, ICA is normally interpreted in terms of the matrix factorisation formulation of Equation A.16. For example, spatial ICA would infer a matrix $\tilde{\boldsymbol{X}}$, of size components by voxels. These are interpreted as a set of independent spatial maps, and this is why the finite sample approximation to the source covariance, $\hat{\boldsymbol{C}}_{\boldsymbol{X},\boldsymbol{X}}$, is of particular interest—this now represents the covariance between the observed spatial maps. However, it is important to note that the model itself (Equation A.14) is blind to this interpretation and any spatial structure in the resulting maps essentially emerges by accident.

Finally, it is well known that residual dependencies between sources often remain after ICA has converged—in fact, there are ICA models that explicitly utilise these dependencies [Hyvärinen and Hoyer 2000; Hyvärinen et al. 2001a]. As we discussed in the Literature Review, Beckmann et al. [2005] demonstrated that, in the case of fMRI, the restriction to uncorrelated sources may not hinder the inference of spatial maps that are highly overlapping—that is, correlated after thresholding—if the data are noisy 'enough'.

A.3 PFM model

The PFM model is explicitly based round a matrix factorisation of fMRI data, and specifies a prior over subject specific maps:

$$\begin{aligned} \mathrm{p}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, q_{vm}^{(s)} = 1, \mu_{vm}, \sigma_{vm}\big) &= \mathcal{N}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, \mu_{vm}, \sigma_{vm}^2\big) \\ \mathrm{p}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, q_{vm}^{(s)} = 0\big) &= \delta\big(\boldsymbol{P}_{vm}^{(s)}\big) \\ \mathrm{p}\big(q_{vm}^{(s)}\big) &= (\pi_{vm})^{q_{vm}^{(s)}}(1-\pi_{vm})^{1-q_{vm}^{(s)}} \end{aligned} \tag{A.21}$$

The prior over the mean parameters, that can be thought of as being analogous to the ICA group maps, takes a similar form:

$$\begin{aligned} \mathrm{p}\big(\mu_{vm} \,\big|\, \rho_{vm} = 1\big) &= \mathcal{N}\big(\mu_{vm} \,\big|\, 0, \gamma^{-1}\big) \\ \mathrm{p}\big(\mu_{vm} \,\big|\, \rho_{vm} = 0\big) &= \delta\big(\mu_{vm}\big) \\ \mathrm{p}\big(\rho_{vm}\big) &= (\lambda)^{\rho_{vm}}(1-\lambda)^{1-\rho_{vm}} \end{aligned} \tag{A.22}$$

Clearly, both these and the factorised posterior distributions are independent, and in fact this independence is over both modes and voxels. However, PFMs are fundamentally different to ICA in that they define distributions over matrices, rather than assuming the matrices are collections of random realisations.

As explained in Section A.2.3, $\hat{\boldsymbol{C}}_{\boldsymbol{X},\boldsymbol{X}}$ is of particular interest for sICA, as it captures the dependencies between the maps we actually infer. However, this sampling approximation of the between-source covariance no longer makes sense in the PFM framework. If we let $\boldsymbol{P}(v,m) = \mu_{vm}\rho_{vm}$ and consider this as analogous to the ICA group maps, then we can explicitly calculate a conceptually similar quantity to $\hat{\boldsymbol{R}}_{\boldsymbol{X},\boldsymbol{X}}$,

$$\boldsymbol{R}^*_{\boldsymbol{P},\boldsymbol{P}} = \frac{1}{V}\sum_{v=1}^{V} \mathrm{E}\big[\boldsymbol{P}_v^{\mathrm{T}}\boldsymbol{P}_v\big] \,. \tag{A.23}$$

The approximate VB posterior for these parameters (see Appendix B) is

$$\begin{aligned} \mathrm{q}(\mu_{vm} \,|\, \rho_{vm} = 1) &= \mathcal{N}(\mu_{vm} \,|\, \hat{\mu}_{\mu_{vm}}, \hat{\sigma}^2_{\mu_{vm}}) \\ \mathrm{q}(\mu_{vm} \,|\, \rho_{vm} = 0) &= \delta(\mu_{vm}) \\ \mathrm{q}(\rho_{vm}) &= (\hat{\phi}_{\rho_{vm}})^{\rho_{vm}} (1 - \hat{\phi}_{\rho_{vm}})^{1-\rho_{vm}} \end{aligned} \tag{A.24}$$

Taking expectations with respect to this yields

$$\boldsymbol{R}^*_{\boldsymbol{P},\boldsymbol{P}}(i,j) = \begin{cases} \frac{1}{V} \sum_{v=1}^{V} \hat{\phi}_{\rho_{vi}} (\hat{\mu}^2_{\mu_{vi}} + \hat{\sigma}^2_{\mu_{vi}}) & \text{if } i = j \\ \frac{1}{V} \sum_{v=1}^{V} (\hat{\phi}_{\rho_{vi}} \hat{\mu}_{\mu_{vi}})(\hat{\phi}_{\rho_{vj}} \hat{\mu}_{\mu_{vj}}) & \text{otherwise} \end{cases} \tag{A.25}$$

Therefore, while the PFM model can clearly support stable correlations between the maps as a whole. The crux of the issue is that for PFMs, in contrast with ICA, each voxel in a spatial map is independently drawn from a *different* distribution. The finite sample approximation we use for ICA pools over voxels, under the assumption that they are identically distributed. If this is not the case, as with PFMs, then this metric will suggest the maps are correlated, even though they are independently distributed. Intuitively, the correlation between the PFM maps is carried by the mean parameters that vary across space. ICA assumes that each voxel has the same mean and is therefore unable to infer correlated maps via this mechanism.

Appendix B

VB Update Rules

B.1 Notation

Prior distributions are denoted $\mathrm{p}(\boldsymbol{A})$ while the approximate posteriors will be designated $\mathrm{q}(\boldsymbol{A})$. For convenience we will use the following notation to denote the expectation of a variable with respect to its approximate posterior:

$$\langle \boldsymbol{A} \rangle = \int \boldsymbol{A}\ \mathrm{q}(\boldsymbol{A})\ \mathrm{d}\boldsymbol{A}$$

We will use subscripts to index elements in matrices, so $\boldsymbol{A}_{mt}$ is the activity of mode m at time t. If there is only one subscript for a matrix then this represents the extraction of a particular row or column from the matrix. For example, $\boldsymbol{A}_m$ extracts the full time course from mode m. Whether a row or column is to be extracted should be obvious from the context.

We will use a tilde to denote extraction of all elements except the current index. Therefore, $\boldsymbol{A}_{\widetilde{m}}$ represents the matrix formed from the time courses of all modes except mode m.

Finally, we use a circumflex to denote parameters of the approximate posterior

distributions, with the variable in question denoted by a subscript. For example, a_β and $\hat{a}_\beta$ are the parameters of the prior and approximate posterior distributions of $\boldsymbol{\beta}$. This is particularly useful as, as we discussed earlier, the conjugate exponential nature of our model means that the approximate posteriors simply require modifications of the prior parameters. Clearly, most of our priors are predicated on variables as well as parameters, but we try and keep this convention as far as is reasonable.

B.2 Example derivation

In this section we will derive the update rules for $\boldsymbol{\alpha}^{(s)}$ from first principles. All other update rules can be derived in a similar fashion.

Many overviews of VB are available, but the notation and approach demonstrated here is particularly informed by Attias [2000], MacKay [2003], Winn et al. [2005] and Chappell et al. [2008].

The crucial equation, as given in Section 4.1.2, is

$$\ln\Big(\mathrm{q}\big(\theta\big)\Big) = \Big\langle \ln\Big(\mathrm{p}\big(\boldsymbol{\Theta}, \mathscr{D}\big)\Big)\Big\rangle_{\mathrm{q}\left(\boldsymbol{\Theta}\neq\theta\right)} + \text{const}\ , \tag{B.1}$$

To begin with, we need to collect all terms that involve $\boldsymbol{\alpha}^{(s)}$.

$$\begin{aligned} \mathrm{p}\big(\boldsymbol{\alpha}^{(s)} \,\big|\, \boldsymbol{\beta}\big) &= \mathcal{W}\big(\boldsymbol{\alpha}^{(s)} \,\big|\, a_{\boldsymbol{\alpha}^{(s)}}, \boldsymbol{\beta}\big) \\ \ln\Big(\mathrm{p}\big(\boldsymbol{\alpha}^{(s)} \,\big|\, \boldsymbol{\beta}\big)\Big) &= \frac{a_{\boldsymbol{\alpha}^{(s)}} - M - 1}{2} \ln\big|\boldsymbol{\alpha}^{(s)}\big| \\ &\quad - \frac{1}{2}\,\mathrm{Tr}\big(\boldsymbol{\beta}\boldsymbol{\alpha}^{(s)}\big) + \text{const}\big(\boldsymbol{\alpha}^{(s)}\big) \end{aligned} \tag{B.2}$$

$$\begin{aligned} \mathrm{p}\left(\boldsymbol{B}^{(sr)} \,\middle|\, \boldsymbol{\alpha}^{(s)}\right) &= \mathcal{MN}\left(\boldsymbol{B}^{(sr)} \,\middle|\, \mathbf{0}, \boldsymbol{\alpha}^{(s)^{-1}}, \boldsymbol{K}_{\boldsymbol{B}}\right) \\ \ln\left(\mathrm{p}\left(\boldsymbol{B}^{(sr)} \,\middle|\, \boldsymbol{\alpha}^{(s)}\right)\right) &= \frac{T}{2}\ln\left|\boldsymbol{\alpha}^{(s)}\right| - \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{B}^{(sr)}\boldsymbol{K}_{\boldsymbol{B}}^{-1}(\boldsymbol{B}^{(sr)})^{\mathrm{T}}\boldsymbol{\alpha}^{(s)}\right) \\ &\quad + \mathrm{const}\left(\boldsymbol{\alpha}^{(s)}\right) \end{aligned} \tag{B.3}$$

We can now plug these into Equation B.1:

$$\begin{aligned} \ln\left(\mathrm{q}\left(\boldsymbol{\alpha}^{(s)}\right)\right) &= \left\langle \ln\left(\mathrm{p}\left(\boldsymbol{\Theta}, \mathscr{D}\right)\right)\right\rangle_{\mathrm{q}\left(\boldsymbol{\Theta}\neq\boldsymbol{\alpha}^{(s)}\right)} \\ &= \frac{a_{\boldsymbol{\alpha}^{(s)}} - M - 1 + R(s)T}{2}\ln\left|\boldsymbol{\alpha}^{(s)}\right| \\ &\quad - \frac{1}{2}\mathrm{Tr}\left(\langle\boldsymbol{\beta}\rangle\boldsymbol{\alpha}^{(s)} + \sum_{r=1}^{R(s)}\left\langle\boldsymbol{B}^{(sr)}\boldsymbol{K}_{\boldsymbol{B}}^{-1}(\boldsymbol{B}^{(sr)})^{\mathrm{T}}\right\rangle\boldsymbol{\alpha}^{(s)}\right) \\ &\quad + \mathrm{const}\left(\boldsymbol{\alpha}^{(s)}\right) \end{aligned} \tag{B.4}$$

Crucially, this has the same functional form as the hyperprior $\mathrm{p}\left(\boldsymbol{\alpha}^{(s)} \,\middle|\, \boldsymbol{\beta}\right)$: it has one term in $\ln\left|\boldsymbol{\alpha}^{(s)}\right|$ and one term in $\mathrm{Tr}\left(\boldsymbol{\alpha}^{(s)}\right)$. Therefore, we can satisfy Equation B.1 by setting $\mathrm{q}\left(\boldsymbol{\alpha}^{(s)}\right)$ to be a Wishart distribution and matching terms. If we let:

$$\begin{aligned} \mathrm{q}\left(\boldsymbol{\alpha}^{(s)}\right) &= \mathcal{W}\left(\boldsymbol{\alpha}^{(s)} \,\middle|\, \hat{a}_{\boldsymbol{\alpha}^{(s)}}, \hat{\boldsymbol{B}}_{\boldsymbol{\alpha}^{(s)}}\right) \\ \ln\left(\mathrm{q}\left(\boldsymbol{\alpha}^{(s)}\right)\right) &= \frac{\hat{a}_{\boldsymbol{\alpha}^{(s)}} - M - 1}{2}\ln\left|\boldsymbol{\alpha}^{(s)}\right| \\ &\quad - \frac{1}{2}\mathrm{Tr}\left(\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}^{(s)}}\boldsymbol{\alpha}^{(s)}\right) + \mathrm{const}\left(\boldsymbol{\alpha}^{(s)}\right) \end{aligned} \tag{B.5}$$

An inspection of Equation B.4 and Equation B.5 tells us that we simply need to set the parameters of $\mathrm{q}\left(\boldsymbol{\alpha}^{(s)}\right)$ to the following values:

$$\hat{a}_{\boldsymbol{\alpha}^{(s)}} = a_{\boldsymbol{\alpha}^{(s)}} + R(s)T \tag{B.6}$$

$$\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}^{(s)}} = \langle \boldsymbol{\beta} \rangle + \sum_{r=1}^{R(s)} \langle \boldsymbol{B}^{(sr)} \boldsymbol{K}_{\boldsymbol{B}}^{-1} (\boldsymbol{B}^{(sr)})^{\mathrm{T}} \rangle \tag{B.7}$$

B.3 Spatial Model

B.3.1 Subject Maps

Prior:

$$\begin{aligned} \mathrm{p}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, q_{vm}^{(s)} = 1, \mu_{vm}, \sigma_{vm}\big) &= \mathcal{N}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, \mu_{vm}, \sigma_{vm}^2\big) \\ \mathrm{p}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, q_{vm}^{(s)} = 0\big) &= \delta\big(\boldsymbol{P}_{vm}^{(s)}\big) \\ \mathrm{p}\big(q_{vm}^{(s)}\big) &= (\pi_{vm})^{q_{vm}^{(s)}} (1 - \pi_{vm})^{1 - q_{vm}^{(s)}} \end{aligned} \tag{B.8}$$

Posterior:

$$\begin{aligned} \mathrm{q}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, q_{vm}^{(s)} = 1\big) &= \mathcal{N}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, \hat{\mu}_{\boldsymbol{P}_{vm}^{(s)}}, \hat{\sigma}^2_{\boldsymbol{P}_{vm}^{(s)}}\big) \\ \mathrm{q}\big(\boldsymbol{P}_{vm}^{(s)} \,\big|\, q_{vm}^{(s)} = 0\big) &= \delta\big(\boldsymbol{P}_{vm}^{(s)}\big) \\ \mathrm{q}\big(q_{vm}^{(s)}\big) &= (\hat{\phi}_{q_{vm}^{(s)}})^{q_{vm}^{(s)}} (1 - \hat{\phi}_{q_{vm}^{(s)}})^{1 - q_{vm}^{(s)}} \end{aligned} \tag{B.9}$$

Parameters:

$$\hat{\sigma}^2_{\boldsymbol{P}_{vm}^{(s)}} = \left(\langle \beta_{vm} \rangle + \sum_{r=1}^{R(s)} \langle \psi^{(sr)} \rangle \langle \boldsymbol{A}_m^{(sr)} (\boldsymbol{A}_m^{(sr)})^{\mathrm{T}} \rangle \right)^{-1} \tag{B.10}$$

$$\begin{aligned} \hat{\mu}_{\boldsymbol{P}_{vm}^{(s)}} = \hat{\sigma}^2_{\boldsymbol{P}_{vm}^{(s)}} \Bigg(& \sum_{r=1}^{R(s)} \langle \psi^{(sr)} \rangle \boldsymbol{D}_v^{(sr)} \langle \boldsymbol{A}_m^{(sr)} \rangle^{\mathrm{T}} \\ & - \sum_{r=1}^{R(s)} \langle \psi^{(sr)} \rangle \langle \boldsymbol{P}_{\widetilde{m}}^{(s)} \rangle \langle \boldsymbol{A}_{\widetilde{m}}^{(sr)} (\boldsymbol{A}^{(sr)})^{\mathrm{T}} \rangle \\ & + \langle \beta_{vm} \rangle \langle \mu_{vm} \rangle \Bigg) \end{aligned} \tag{B.11}$$

$$
\begin{aligned}
\ln\big(\hat{\phi}^{(0)}_{q^{(s)}_{vm}}\big) &= \big\langle \ln\big(1-\pi_{vm}\big)\big\rangle \\
\ln\big(\hat{\phi}^{(1)}_{q^{(s)}_{vm}}\big) &= \big\langle \ln\big(\pi_{vm}\big)\big\rangle + \frac{1}{2}\big\langle \ln\big(\beta_{vm}\big)\big\rangle - \frac{1}{2}\big\langle \beta_{vm}\big\rangle\big\langle \mu^2_{vm}\big\rangle \\
&\quad - \frac{1}{2}\ln\big(\hat{\sigma}^{-2}_{\boldsymbol{P}^{(s)}_{vm}}\big) + \frac{1}{2}\hat{\sigma}^{-2}_{\boldsymbol{P}^{(s)}_{vm}}\hat{\mu}^2_{\boldsymbol{P}^{(s)}_{vm}} \\
\hat{\phi}_{q^{(s)}_{vm}} &= \frac{\hat{\phi}^{(1)}_{q^{(s)}_{vm}}}{\hat{\phi}^{(0)}_{q^{(s)}_{vm}} + \hat{\phi}^{(1)}_{q^{(s)}_{vm}}}
\end{aligned}
\tag{B.12}
$$

B.3.2 Group Mixture Weights

Prior:

$$
\mathrm{p}\big(\pi_{vm}\big) = \beta\big(\pi_{vm} \,\big|\, a_{\pi_{vm}}, b_{\pi_{vm}}\big) \tag{B.13}
$$

Posterior:

$$
\mathrm{q}\big(\pi_{vm}\big) = \beta\big(\pi_{vm} \,\big|\, \hat{a}_{\pi_{vm}}, \hat{b}_{\pi_{vm}}\big) \tag{B.14}
$$

Parameters:

$$
\hat{a}_{\pi_{vm}} = a_{\pi_{vm}} + \sum_{s=1}^{S}\big\langle q^{(s)}_{vm}\big\rangle \tag{B.15}
$$

$$
\hat{b}_{\pi_{vm}} = b_{\pi_{vm}} + S - \sum_{s=1}^{S}\big\langle q^{(s)}_{vm}\big\rangle \tag{B.16}
$$

B.3.3 Group Standard Deviations

Prior:

$$
\mathrm{p}\big(\sigma_{vm}\big) = \Gamma\big(\sigma^{-2}_{vm} \,\big|\, a_{\sigma_{vm}}, b_{\sigma_{vm}}\big) \tag{B.17}
$$

Posterior:

$$
\mathrm{q}\big(\sigma_{vm}\big) = \Gamma\big(\sigma^{-2}_{vm} \,\big|\, \hat{a}_{\sigma_{vm}}, \hat{b}_{\sigma_{vm}}\big) \tag{B.18}
$$

Parameters:

$$\hat{a}_{\sigma_{vm}} = a_{\sigma_{vm}} + \frac{1}{2}\sum_{s=1}^{S}\langle q_{vm}^{(s)}\rangle \tag{B.19}$$

$$\hat{b}_{\sigma_{vm}} = b_{\sigma_{vm}} + \frac{1}{2}\sum_{s=1}^{S}\langle q_{vm}^{(s)}\rangle\Big(\langle (\boldsymbol{P}_{vm}^{(s)})^2 | q_{vm}^{(s)} = 1\rangle + \langle (\mu_{vm})^2\rangle - 2\langle \boldsymbol{P}_{vm}^{(s)} | q_{vm}^{(s)} = 1\rangle\langle \mu_{vm}\rangle\Big) \tag{B.20}$$

B.3.4 Group Means

Prior:

$$\begin{aligned} \mathrm{p}(\mu_{vm} \,|\, \rho_{vm} = 1) &= \mathcal{N}(\mu_{vm} \,|\, 0, \gamma^{-1}) \\ \mathrm{p}(\mu_{vm} \,|\, \rho_{vm} = 0) &= \delta(\mu_{vm}) \\ \mathrm{p}(\rho_{vm}) &= (\lambda)^{\rho_{vm}}(1-\lambda)^{1-\rho_{vm}} \end{aligned} \tag{B.21}$$

Posterior:

$$\begin{aligned} \mathrm{q}(\mu_{vm} \,|\, \rho_{vm} = 1) &= \mathcal{N}(\mu_{vm} \,|\, \hat{\mu}_{\mu_{vm}}, \hat{\sigma}^2_{\mu_{vm}}) \\ \mathrm{q}(\mu_{vm} \,|\, \rho_{vm} = 0) &= \delta(\mu_{vm}) \\ \mathrm{q}(\rho_{vm}) &= (\hat{\phi}_{\rho_{vm}})^{\rho_{vm}}(1-\hat{\phi}_{\rho_{vm}})^{1-\rho_{vm}} \end{aligned} \tag{B.22}$$

Parameters:

$$\hat{\sigma}^2_{\mu_{vm}} = \Big(\gamma + \langle \sigma_{vm}^{-2}\rangle \sum_{s=1}^{S}\langle q_{vm}^{(s)}\rangle\Big)^{-1} \tag{B.23}$$

$$\hat{\mu}_{\mu_{vm}} = \hat{\sigma}^2_{\mu_{vm}}\langle \sigma_{vm}^{-2}\rangle \sum_{s=1}^{S}\langle \boldsymbol{P}_{vm}^{(s)} | q_{vm}^{(s)} = 1\rangle\langle q_{vm}^{(s)}\rangle \tag{B.24}$$

$$\begin{aligned}
\ln\left(\hat{\phi}^{(0)}_{\rho_{vm}}\right) &= \ln(1-\lambda) \\
\ln\left(\hat{\phi}^{(1)}_{\rho_{vm}}\right) &= \ln(\lambda) + \frac{1}{2}\ln(\gamma) - \frac{1}{2}\ln\left(\hat{\sigma}^{-2}_{\mu_{vm}}\right) + \frac{1}{2}\hat{\sigma}^{-2}_{\mu_{vm}}\hat{\mu}^{2}_{\mu_{vm}} \\
\hat{\phi}_{\rho_{vm}} &= \frac{\hat{\phi}^{(1)}_{\rho_{vm}}}{\hat{\phi}^{(0)}_{\rho_{vm}} + \hat{\phi}^{(1)}_{\rho_{vm}}}
\end{aligned} \tag{B.25}$$

B.4 Temporal Model

B.4.1 BOLD Time Courses

Prior:

$$\mathrm{p}\left(\boldsymbol{B}^{(sr)} \,\middle|\, \boldsymbol{\alpha}^{(s)}\right) = \mathcal{N}\left(\mathrm{vec}\left(\boldsymbol{B}^{(sr)}\right) \,\middle|\, \mathbf{0}, \boldsymbol{\alpha}^{(s)^{-1}} \otimes \boldsymbol{K}_{\boldsymbol{B}}\right) \tag{B.26}$$

Posterior:

$$\mathrm{q}\left(\boldsymbol{B}^{(sr)}\right) = \mathcal{N}\left(\mathrm{vec}\left(\boldsymbol{B}^{(sr)}\right) \,\middle|\, \mathrm{vec}\left(\hat{\boldsymbol{M}}_{\boldsymbol{B}^{(sr)}}\right), \hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}^{(sr)}}\right) \tag{B.27}$$

Parameters:

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}^{(sr)}} = \left(\left\langle \boldsymbol{\alpha}^{(s)} \right\rangle \otimes \boldsymbol{K}_{\boldsymbol{B}}^{-1} + \left\langle \psi^{(sr)} \right\rangle \left\langle \left(\boldsymbol{P}^{(s)}\right)^{\mathrm{T}} \boldsymbol{P}^{(s)} \right\rangle \otimes \boldsymbol{I}_{T}\right)^{-1} \tag{B.28}$$

$$\mathrm{vec}\left(\hat{\boldsymbol{M}}_{\boldsymbol{B}^{(sr)}}\right) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}^{(sr)}} \mathrm{vec}\left(\left\langle \psi^{(sr)} \right\rangle \left\langle \boldsymbol{P}^{(s)} \right\rangle^{\mathrm{T}} \boldsymbol{D}^{(sr)} - \left\langle \boldsymbol{\xi}^{(sr)} \right\rangle\right) \tag{B.29}$$

B.4.2 BOLD Time Course Subject-Level Precisions

Prior:

$$\mathrm{p}\left(\boldsymbol{\alpha}^{(s)} \,\middle|\, \boldsymbol{\beta}\right) = \mathcal{W}\left(\boldsymbol{\alpha}^{(s)} \,\middle|\, a_{\boldsymbol{\alpha}^{(s)}}, \boldsymbol{\beta}\right) \tag{B.30}$$

Posterior:

$$\mathrm{q}\left(\boldsymbol{\alpha}^{(s)}\right) = \mathcal{W}\left(\boldsymbol{\alpha}^{(s)} \,\middle|\, \hat{a}_{\boldsymbol{\alpha}^{(s)}}, \hat{\boldsymbol{B}}_{\boldsymbol{\alpha}^{(s)}}\right) \tag{B.31}$$

Parameters:

$$\hat{a}_{\boldsymbol{\alpha}^{(s)}} = a_{\boldsymbol{\alpha}^{(s)}} + R(s)T \tag{B.32}$$

$$\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}^{(s)}} = \langle \boldsymbol{\beta} \rangle + \sum_{r=1}^{R(s)} \left\langle \boldsymbol{B}^{(sr)} \boldsymbol{K}_{\boldsymbol{B}}^{-1} (\boldsymbol{B}^{(sr)})^{\mathrm{T}} \right\rangle \tag{B.33}$$

B.4.3 BOLD Time Course Group-Level Precisions

Prior:

$$\mathrm{p}(\boldsymbol{\beta}) = \mathcal{W}\left(\boldsymbol{\beta} \,\middle|\, a_{\beta}, \boldsymbol{B}_{\beta}\right) \tag{B.34}$$

Posterior:

$$\mathrm{q}(\boldsymbol{\beta}) = \mathcal{W}\left(\boldsymbol{\beta} \,\middle|\, \hat{a}_{\beta}, \hat{\boldsymbol{B}}_{\beta}\right) \tag{B.35}$$

Parameters:

$$\hat{a}_{\beta} = a_{\beta} + \sum_{s=1}^{S} a_{\boldsymbol{\alpha}^{(s)}} \tag{B.36}$$

$$\hat{\boldsymbol{B}}_{\beta} = \boldsymbol{B}_{\beta} + \sum_{s=1}^{S} \left\langle \boldsymbol{\alpha}^{(s)} \right\rangle \tag{B.37}$$

B.4.4 Noise Time Courses

Let $\boldsymbol{\omega}^{(sr)} \in \mathbb{R}^{M \times M}$ represent the diagonal matrix formed by placing each $\omega_m^{(sr)}$ along the diagonal, or in other words,

$$\boldsymbol{\omega}_{ij}^{(sr)} = \begin{cases} \omega_i^{(sr)} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{B.38}$$

Prior:

$$\mathrm{p}\left(\xi_t^{(sr)} \,\middle|\, \boldsymbol{\omega}^{(sr)}\right) = \mathcal{N}\left(\xi_t^{(sr)} \,\middle|\, \mathbf{0}, \boldsymbol{\omega}^{(sr)^{-1}}\right) \tag{B.39}$$

Posterior:

$$\mathrm{q}\left(\xi_t^{(sr)}\right) = \mathcal{N}\left(\xi_t^{(sr)} \,\middle|\, \hat{\boldsymbol{M}}_{\xi_t^{(sr)}}, \hat{\boldsymbol{\Sigma}}_{\xi_t^{(sr)}}\right) \tag{B.40}$$

Parameters:

$$\hat{\boldsymbol{\Sigma}}_{\xi_t^{(sr)}} = \left(\boldsymbol{\omega}^{(sr)} + \left\langle \psi^{(sr)} \right\rangle \left\langle \left(\boldsymbol{P}^{(s)}\right)^{\mathrm{T}} \boldsymbol{P}^{(s)} \right\rangle\right)^{-1} \tag{B.41}$$

$$\mathrm{vec}(\hat{\boldsymbol{M}}_{\xi_t^{(sr)}}) = \hat{\boldsymbol{\Sigma}}_{\xi_t^{(sr)}} \left(\left\langle \psi^{(sr)} \right\rangle \left\langle \boldsymbol{P}^{(s)} \right\rangle^{\mathrm{T}} \boldsymbol{D}^{(sr)} - \left\langle \boldsymbol{B}^{(sr)} \right\rangle \right) \tag{B.42}$$

B.4.5 Noise Time Course Precisions

Prior:

$$\mathrm{p}\left(\omega_m^{(sr)}\right) = \Gamma\left(\omega_m^{(sr)} \,\middle|\, a_{\omega_m^{(sr)}}, b_{\omega_m^{(sr)}}\right) \tag{B.43}$$

Posterior:

$$\mathrm{q}\left(\omega_m^{(sr)}\right) = \Gamma\left(\omega_m^{(sr)} \,\middle|\, \hat{a}_{\omega_m^{(sr)}}, \hat{b}_{\omega_m^{(sr)}}\right) \tag{B.44}$$

Parameters:

$$\hat{a}_{\omega_m^{(sr)}} = a_{\omega_m^{(sr)}} + \frac{T}{2} \tag{B.45}$$

$$\hat{b}_{\omega_m^{(sr)}} = b_{\omega_m^{(sr)}} + \frac{1}{2} \sum_{t=1}^{T} \left\langle \xi_t^{(sr)} \left(\xi_t^{(sr)}\right)^{\mathrm{T}} \right\rangle \tag{B.46}$$

B.5 Noise Model

B.5.1 Precision

Prior:

$$\mathrm{p}\left(\psi^{(sr)}\right) = \Gamma\left(\psi^{(sr)} \,\middle|\, a_{\psi^{(sr)}}, b_{\psi^{(sr)}}\right) \tag{B.47}$$

Posterior:

$$\mathrm{q}\left(\psi^{(sr)}\right) = \Gamma\left(\psi^{(sr)} \,\middle|\, \hat{a}_{\psi^{(sr)}}, \hat{b}_{\psi^{(sr)}}\right) \tag{B.48}$$

Parameters:

$$\hat{a}_{\psi^{(sr)}} = a_{\psi^{(sr)}} + \frac{VT}{2} \tag{B.49}$$

$$\begin{aligned} \hat{b}_{\psi^{(sr)}} = b_{\psi^{(sr)}} &+ \frac{1}{2}\,\mathrm{Tr}\Big(\big(\boldsymbol{D}^{(sr)}\big)^{\mathrm{T}}\boldsymbol{D}^{(sr)}\Big) - \mathrm{Tr}\Big(\big(\boldsymbol{D}^{(sr)}\big)^{\mathrm{T}}\big\langle\boldsymbol{P}^{(s)}\big\rangle\big\langle\boldsymbol{A}^{(sr)}\big\rangle\Big) \\ &+ \frac{1}{2}\,\mathrm{Tr}\Big(\big\langle\big(\boldsymbol{P}^{(s)}\big)^{\mathrm{T}}\boldsymbol{P}^{(s)}\big\rangle\big\langle\boldsymbol{A}^{(sr)}\big(\boldsymbol{A}^{(sr)}\big)^{\mathrm{T}}\big\rangle\Big) \end{aligned} \tag{B.50}$$

Appendix C

Computation of the BOLD Time Course Updates

C.1 Distributions

C.1.1 Matrix normal distribution

The matrix normal distribution is a generalisation of the multivariate normal distribution to matrices, for certain covariance structures. It is used when the covariance structure for the whole matrix can be expressed as the Kronecker product of covariance matrices over the rows and columns of the matrix [Dawid 1981].

If the random matrix $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ follows a matrix normal distribution, with mean $\boldsymbol{M} \in \mathbb{R}^{m \times n}$, column covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$ and row covariance $\boldsymbol{\Omega} \in \mathbb{R}^{n \times n}$, then the following hold:

$$\boldsymbol{X} \sim \mathcal{MN}\left(\boldsymbol{M}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}\right)$$

$$\mathrm{vec}\left(\boldsymbol{X}\right) \sim \mathcal{N}\left(\mathrm{vec}\left(\boldsymbol{M}\right), \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}\right)$$

$$\mathrm{p}\big(\boldsymbol{X} \,\big|\, \boldsymbol{M}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}\big) = \frac{\exp\Big(-\frac{1}{2}\mathrm{Tr}\big(\boldsymbol{\Omega}^{-1}(\boldsymbol{X}-\boldsymbol{M})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\big)\Big)}{(2\pi)^{\frac{mn}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}|\boldsymbol{\Omega}|^{\frac{m}{2}}}$$

$$\big\langle \boldsymbol{X} \big\rangle = \boldsymbol{M}$$

$$\big\langle \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} \big\rangle = \boldsymbol{M}^T\boldsymbol{A}\boldsymbol{M} + \boldsymbol{\Omega}\,\mathrm{Tr}(\boldsymbol{\Sigma}\boldsymbol{A})$$

$$\big\langle \boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^T \big\rangle = \boldsymbol{M}\boldsymbol{B}\boldsymbol{M}^T + \boldsymbol{\Sigma}\,\mathrm{Tr}(\boldsymbol{\Omega}\boldsymbol{B})$$

$$\big\langle \mathrm{Tr}\big(\boldsymbol{A}\boldsymbol{X}^T\boldsymbol{Z}\boldsymbol{X}\big) \big\rangle = \mathrm{Tr}\big((\boldsymbol{B}\big\langle \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} \big\rangle\big) = \mathrm{Tr}\big(\big\langle \boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^T \big\rangle \boldsymbol{A}\big)$$

C.2 $\boldsymbol{B}$ updates

For computational reasons, we define the prior over $\boldsymbol{B}^{\mathrm{T}}$ rather than over $\boldsymbol{B}$ directly. This does not change the form of any of the distributions, but gives the posterior covariance matrix a more convenient block structure. With this formulation, the blocks of the matrix are defined by full time courses, rather than the set of all mode activities at one time point—intuitively, $\mathrm{vec}(\boldsymbol{B}^{\mathrm{T}})$ concatenates time courses rather than time points.

The update rules are only sketched in outline to try and keep the notation relatively simple: we drop the run and subject indices and ignore the noise time courses, $\boldsymbol{\xi}$, as these can simply be subtracted from $\big\langle \psi \big\rangle \boldsymbol{D}^{\mathrm{T}} \big\langle \boldsymbol{P} \big\rangle$. However, in general, we may wish to infer $\boldsymbol{B}$ from a range of observations[1], in which case there may be multiple sets of ψ, $\boldsymbol{P}$ and $\boldsymbol{D}$. Rather than writing the sum over these sets of variables out in full each time, we will keep these implicit. Note that this is why ψ does not always commute, even though it is a scalar—in general, it is paired with a

[1] The obvious example is multiple subjects doing the same task and therefore having equivalent time courses.

set of $\boldsymbol{P}$ and $\boldsymbol{D}$.

$$\begin{aligned} \mathrm{p}\big(\boldsymbol{B} \,\big|\, \boldsymbol{\alpha}\big) &= \mathcal{MN}\big(\boldsymbol{B} \,\big|\, \boldsymbol{0}, \boldsymbol{\alpha}^{-1}, \boldsymbol{K}_{\boldsymbol{B}}\big) \\ &= \mathcal{N}\big(\mathrm{vec}\big(\boldsymbol{B}^{\mathrm{T}}\big) \,\big|\, \boldsymbol{0}, \boldsymbol{\alpha}^{-1} \otimes \boldsymbol{K}_{\boldsymbol{B}}\big) \end{aligned} \tag{C.1}$$

$$\mathrm{q}\big(\boldsymbol{B}\big) = \mathcal{N}\big(\mathrm{vec}\big(\boldsymbol{B}^{\mathrm{T}}\big) \,\big|\, \mathrm{vec}\big(\hat{\boldsymbol{M}}_{\boldsymbol{B}}^{\mathrm{T}}\big), \hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}\big) \tag{C.2}$$

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}} = \Big(\big\langle \boldsymbol{\alpha} \big\rangle \otimes \boldsymbol{K}_{\boldsymbol{B}}^{-1} + \big\langle \boldsymbol{P}^{\mathrm{T}} \boldsymbol{P} \big\rangle \otimes \big\langle \psi \big\rangle \boldsymbol{I}_{T}\Big)^{-1} \tag{C.3}$$

$$\mathrm{vec}\big(\hat{\boldsymbol{M}}_{\boldsymbol{B}}^{\mathrm{T}}\big) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}\, \mathrm{vec}\Big(\big\langle \psi \big\rangle \boldsymbol{D}^{\mathrm{T}} \big\langle \boldsymbol{P} \big\rangle\Big) \tag{C.4}$$

$$\begin{aligned} \mathrm{D}_{\mathrm{KL}}\Big(\mathrm{q}\big(\boldsymbol{B}\big) \,\Big\|\, \mathrm{p}\big(\boldsymbol{B} \,\big|\, \boldsymbol{\alpha}\big)\Big) &= \frac{1}{2}\left\langle \ln\left(\frac{|\boldsymbol{\alpha}^{-1} \otimes \boldsymbol{K}_{\boldsymbol{B}}|}{|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}|}\right)\right\rangle + \frac{1}{2}\left\langle \mathrm{Tr}\Big(\big(\boldsymbol{\alpha}^{-1} \otimes \boldsymbol{K}_{\boldsymbol{B}}\big)^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}\Big)\right\rangle \\ &\quad + \frac{1}{2}\, \mathrm{vec}\big(\hat{\boldsymbol{M}}_{\boldsymbol{B}}^{\mathrm{T}}\big)^{\mathrm{T}} \left\langle \big(\boldsymbol{\alpha}^{-1} \otimes \boldsymbol{K}_{\boldsymbol{B}}\big)^{-1} \right\rangle \mathrm{vec}\big(\hat{\boldsymbol{M}}_{\boldsymbol{B}}^{\mathrm{T}}\big) - \frac{TM}{2} \\ &= \frac{M}{2} \ln\big(|\boldsymbol{K}_{\boldsymbol{B}}|\big) - \frac{T}{2}\big\langle \ln\big(|\boldsymbol{\alpha}|\big)\big\rangle - \frac{1}{2} \ln\big(|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}|\big) \\ &\quad + \frac{1}{2}\, \mathrm{Tr}\Big(\big(\big\langle \boldsymbol{\alpha} \big\rangle \otimes \boldsymbol{K}_{\boldsymbol{B}}^{-1}\big) \hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}\Big) \\ &\quad + \frac{1}{2}\, \mathrm{Tr}\big(\big\langle \boldsymbol{\alpha} \big\rangle \hat{\boldsymbol{M}}_{\boldsymbol{B}} \boldsymbol{K}_{\boldsymbol{B}}^{-1} \hat{\boldsymbol{M}}_{\boldsymbol{B}}^{\mathrm{T}}\big) - \frac{TM}{2} \\ &= \frac{M}{2} \ln\big(|\boldsymbol{K}_{\boldsymbol{B}}|\big) - \frac{T}{2}\big\langle \ln\big(|\boldsymbol{\alpha}|\big)\big\rangle - \frac{1}{2} \ln\big(|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}|\big) \\ &\quad + \frac{1}{2}\, \mathrm{Tr}\Big(\big\langle \boldsymbol{\alpha} \big\rangle \big\langle \boldsymbol{B} \boldsymbol{K}_{\boldsymbol{B}}^{-1} \boldsymbol{B}^{\mathrm{T}} \big\rangle\Big) - \frac{TM}{2} \end{aligned} \tag{C.5}$$

C.3 $\boldsymbol{\alpha}$ updates

This gives the loose form of the updates necessary to infer a posterior covariance across modes. The precise way the information from different time courses is pooled to infer the covariance is arbitrary—the updates are simply provided to illustrate the expectations we need to calculate.

Prior:

$$\mathrm{p}(\boldsymbol{\alpha} \mid \boldsymbol{\beta}) = \mathcal{W}(\boldsymbol{\alpha} \mid a_{\boldsymbol{\alpha}}, \boldsymbol{\beta}) \tag{C.6}$$

Posterior:

$$\mathrm{q}(\boldsymbol{\alpha}) = \mathcal{W}(\boldsymbol{\alpha} \mid \hat{a}_{\boldsymbol{\alpha}}, \hat{\boldsymbol{B}}_{\boldsymbol{\alpha}}) \tag{C.7}$$

Parameters:

$$\hat{a}_{\boldsymbol{\alpha}} = a_{\boldsymbol{\alpha}} + RT \tag{C.8}$$

$$\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}} = \langle \boldsymbol{\beta} \rangle + \sum_{r=1}^{R} \langle \boldsymbol{B} \boldsymbol{K}_{\boldsymbol{B}}^{-1} \boldsymbol{B}^{\mathrm{T}} \rangle \tag{C.9}$$

C.4 Computational approach

The above demonstrate that we need to be able to manipulate $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$, the posterior covariance matrix over all elements in $\boldsymbol{B}$. However, this is a square matrix of side MT. For even relatively modest values, say $M = 100$ and $T = 1000$, this contains over five billion unique entries, far too large for a naïve computational approach.

In this section we detail how it is possible to utilise the structure of this matrix to extract all the necessary expectations, without ever forming the full matrix.

We need to calculate four separate quantities using $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$. Firstly, it is used in Equation C.4 to calculate the posterior mean of $\boldsymbol{B}$. Secondly, we need it to calculate $\langle \boldsymbol{B} \boldsymbol{B}^{\mathrm{T}} \rangle$, which is needed by the other parts of the matrix factorisation

model. Thirdly, the update rules for $\boldsymbol{\alpha}$ and the free energy require $\left\langle \boldsymbol{B}\boldsymbol{K}_{\boldsymbol{B}}^{-1}\boldsymbol{B}^{\mathrm{T}} \right\rangle$. Finally, we need its determinant, again for the free energy.

C.4.1 Eigendecomposition of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$ using the Kronecker structure

For two real, symmetric matrices with accompanying eigendecompositions, $\boldsymbol{X} = \boldsymbol{U}_{\boldsymbol{X}}\boldsymbol{\Lambda}_{\boldsymbol{X}}\boldsymbol{U}_{\boldsymbol{X}}^{\mathrm{T}}$ and $\boldsymbol{Y} = \boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{\Lambda}_{\boldsymbol{Y}}\boldsymbol{U}_{\boldsymbol{Y}}^{\mathrm{T}}$, it is straightforward to show that the Kronecker product takes the form

$$\boldsymbol{X} \otimes \boldsymbol{Y} = \left(\boldsymbol{U}_{\boldsymbol{X}} \otimes \boldsymbol{U}_{\boldsymbol{Y}}\right)\left(\boldsymbol{\Lambda}_{\boldsymbol{X}} \otimes \boldsymbol{\Lambda}_{\boldsymbol{Y}}\right)\left(\boldsymbol{U}_{\boldsymbol{X}}^{\mathrm{T}} \otimes \boldsymbol{U}_{\boldsymbol{Y}}^{\mathrm{T}}\right)$$

The other property we require is a general one of eigendecompositions:

$$\boldsymbol{X} + c\boldsymbol{I} = \boldsymbol{U}_{\boldsymbol{X}}\left(\boldsymbol{\Lambda}_{\boldsymbol{X}} + c\boldsymbol{I}\right)\boldsymbol{U}_{\boldsymbol{X}}^{\mathrm{T}}$$

As Stegle et al. [2011] noted, it is possible to utilise these two properties for very efficient solutions of problems involving matrix normal distributions. The combination of the two yields the following equation, which we can use to invert $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$.

$$\left(\boldsymbol{X} \otimes \boldsymbol{Y} + c\boldsymbol{I}\right)^{-1} = \left(\boldsymbol{U}_{\boldsymbol{X}} \otimes \boldsymbol{U}_{\boldsymbol{Y}}\right)\left(\boldsymbol{\Lambda}_{\boldsymbol{X}} \otimes \boldsymbol{\Lambda}_{\boldsymbol{Y}} + c\boldsymbol{I}\right)^{-1}\left(\boldsymbol{U}_{\boldsymbol{X}}^{\mathrm{T}} \otimes \boldsymbol{U}_{\boldsymbol{Y}}^{\mathrm{T}}\right)$$

Once we have expressed $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$ in terms of its eigendecomposition, it is possible to simplify the formulation of the necessary expectations. To do this, we will also make use of the following property of the matrix inverse:

$$\left(\boldsymbol{X} + \boldsymbol{Y}\right)^{-1} = \boldsymbol{X}^{-1}\left(\boldsymbol{I} + \boldsymbol{Y}\boldsymbol{X}^{-1}\right)^{-1}$$

The form of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$ we require is:

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}} &= \left(\langle\boldsymbol{\alpha}\rangle \otimes \boldsymbol{K}_{\boldsymbol{B}}^{-1} + \langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle \otimes \langle\psi\rangle\boldsymbol{I}_T\right)^{-1} \\ &= \left(\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle \otimes \langle\psi\rangle\boldsymbol{I}_T\right)^{-1}\left(\left(\langle\boldsymbol{\alpha}\rangle \otimes \boldsymbol{K}_{\boldsymbol{B}}^{-1}\right)\left(\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle \otimes \langle\psi\rangle\boldsymbol{I}_T\right)^{-1} + \boldsymbol{I}_{MT}\right)^{-1} \\ &= \left(\left(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\right)^{-1} \otimes \boldsymbol{I}_T\right)\left(\left(\langle\boldsymbol{\alpha}\rangle\left(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\right)^{-1}\right) \otimes \boldsymbol{K}_{\boldsymbol{B}}^{-1} + \boldsymbol{I}_{MT}\right)^{-1}\end{aligned} \tag{C.10}$$

The relevant eigendecompositions are denoted $\boldsymbol{K}_{\boldsymbol{B}} = \boldsymbol{U}_{\boldsymbol{K}}\boldsymbol{\Lambda}_{\boldsymbol{K}}\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}$, $\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle = \boldsymbol{U}_{\boldsymbol{P}}\boldsymbol{\Lambda}_{\boldsymbol{P}}\boldsymbol{U}_{\boldsymbol{P}}^{\mathrm{T}}$, $\langle\boldsymbol{\alpha}\rangle = \boldsymbol{U}_{\boldsymbol{\alpha}}\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}\boldsymbol{U}_{\boldsymbol{\alpha}}^{\mathrm{T}}$ and $\langle\boldsymbol{\alpha}\rangle\left(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\right)^{-1} = \boldsymbol{U}_{\boldsymbol{\alpha}\boldsymbol{P}}\boldsymbol{\Lambda}_{\boldsymbol{\alpha}\boldsymbol{P}}\boldsymbol{U}_{\boldsymbol{\alpha}\boldsymbol{P}}^{-1}$. Combining these with Equation C.10 gives:

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}} = \left(\left(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\right)^{-1} \otimes \boldsymbol{I}_T\right)\left(\boldsymbol{U}_{\boldsymbol{\alpha}\boldsymbol{P}} \otimes \boldsymbol{U}_{\boldsymbol{K}}\right)\left(\boldsymbol{\Lambda}_{\boldsymbol{\alpha}\boldsymbol{P}} \otimes \boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1} + \boldsymbol{I}_{MT}\right)^{-1}\left(\boldsymbol{U}_{\boldsymbol{\alpha}\boldsymbol{P}}^{-1} \otimes \boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\right) \tag{C.11}$$

Importantly, it is possible to precompute the eigendecomposition of $\boldsymbol{K}_{\boldsymbol{B}}$, and the decompositions of the various $M \times M$ matrices should be relatively inexpensive.

Note that in general $\langle\boldsymbol{\alpha}\rangle\left(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\right)^{-1}$ is not symmetric. However, it is possible to show that the eigenvalues will still be real, though the eigenvectors are not guaranteed to be orthogonal. This is useful as it means that it is not necessary to consider the general case of complex matrices when doing the various computations.

To show this, we will use the concept of matrix similarity. Two matrices, $\boldsymbol{X}$ and $\boldsymbol{Y}$, are considered similar if there exists an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{Y} = \boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{P}$. If this holds, then the matrices share a number of properties, including properties of the characteristic polynomial (and hence eigenvalues).

Consider two real, positive semidefinite matrices, $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let $\boldsymbol{Z} = \boldsymbol{P}^{-1}\boldsymbol{X}\boldsymbol{Y}\boldsymbol{P}$, and take $\boldsymbol{P} = \boldsymbol{Y}^{-\frac{1}{2}}$ (this exists, and is positive semidefinite, because $\boldsymbol{Y}$ is positive semidefinite). Rearranging gives that $\boldsymbol{Z} = \boldsymbol{Y}^{\frac{1}{2}}\boldsymbol{X}\boldsymbol{Y}^{\frac{1}{2}}$, which is clearly symmetric.

In other words, the product of two real, positive semidefinite matrices is similar to a real, symmetric matrix, so will have real eigenvalues.

C.4.2 Calculation of expectations

Vector multiplication

The posterior mean requires $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$ to be multiplied by a vector. To do this we will require the following:

$$\left(\boldsymbol{X} \otimes \boldsymbol{Y}\right) \mathrm{vec}\left(\boldsymbol{Z}\right) = \mathrm{vec}\left(\boldsymbol{Y}\boldsymbol{Z}\boldsymbol{X}^{\mathrm{T}}\right)$$

Using Equation C.11, the simplified vector multiplication operation, for $\boldsymbol{Z} \in \mathbb{R}^{T\times M}$, is

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}\,\mathrm{vec}\left(\boldsymbol{Z}\right) = \left(\langle\psi\rangle^{-1}\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle^{-1} \otimes \boldsymbol{I}_{T}\right)\left(\boldsymbol{U}_{\alpha\boldsymbol{P}} \otimes \boldsymbol{U}_{\boldsymbol{K}}\right)\ \ldots \\ \ldots\ \left(\boldsymbol{\Lambda}_{\alpha\boldsymbol{P}} \otimes \boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1} + \boldsymbol{I}_{MT}\right)^{-1} \mathrm{vec}\left(\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-\mathrm{T}}\right)\end{aligned} \tag{C.12}$$

As the eigenvalue matrix is diagonal, we can simplify this further. For notational purposes, let $\boldsymbol{L} \in \mathbb{R}^{M\times T}$ be a rearrangement of this eigenvector matrix, such that $\boldsymbol{L}(m,t) = \left(\boldsymbol{\Lambda}_{\alpha\boldsymbol{P}}(m)\boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1}(t) + 1\right)^{-1}$. Using $\circ$ to represent the Hadamard product, we can continue simplifying Equation C.12 as

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}\,\mathrm{vec}\left(\boldsymbol{Z}\right) &= \left(\left(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\right)^{-1} \otimes \boldsymbol{I}_{T}\right)\left(\boldsymbol{U}_{\alpha\boldsymbol{P}} \otimes \boldsymbol{U}_{\boldsymbol{K}}\right)\mathrm{vec}\left(\boldsymbol{L}^{\mathrm{T}} \circ \left(\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-\mathrm{T}}\right)\right) \\ &= \left(\left(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\right)^{-1} \otimes \boldsymbol{I}_{T}\right)\mathrm{vec}\Big(\boldsymbol{U}_{\boldsymbol{K}}\left(\boldsymbol{L}^{\mathrm{T}} \circ \left(\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-\mathrm{T}}\right)\right)\boldsymbol{U}_{\alpha\boldsymbol{P}}^{\mathrm{T}}\Big) \\ &= \mathrm{vec}\Big(\boldsymbol{U}_{\boldsymbol{K}}\left(\boldsymbol{L}^{\mathrm{T}} \circ \left(\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-\mathrm{T}}\right)\right)\boldsymbol{U}_{\alpha\boldsymbol{P}}^{\mathrm{T}}\left(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\right)^{-1}\Big)\end{aligned} \tag{C.13}$$

Finally, combining Equation C.4 and Equation C.13 gives the final expression

for calculating the posterior mean.

$$\begin{aligned}\mathrm{vec}\big(\hat{\boldsymbol{M}}_{\boldsymbol{B}}^{\mathrm{T}}\big) &= \mathrm{vec}\Big(\boldsymbol{U}_{\boldsymbol{K}}\big(\boldsymbol{L}^{\mathrm{T}} \circ (\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\langle \boldsymbol{P}\rangle\langle\psi\rangle\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-\mathrm{T}})\big)\boldsymbol{U}_{\alpha\boldsymbol{P}}^{\mathrm{T}}\big(\langle\psi\rangle\langle \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\big)^{-1}\Big) \\ \hat{\boldsymbol{M}}_{\boldsymbol{B}} &= \big(\langle\psi\rangle\langle \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\big)^{-1}\boldsymbol{U}_{\alpha\boldsymbol{P}}\big(\boldsymbol{L} \circ (\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-1}\langle\psi\rangle\langle \boldsymbol{P}^{\mathrm{T}}\rangle\boldsymbol{D}\boldsymbol{U}_{\boldsymbol{K}})\big)\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\end{aligned} \tag{C.14}$$

Blockwise trace

The contribution of the posterior covariance to $\langle \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\rangle$ can be formulated as the trace of different blocks of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$, each of size $T\times T$. The expression for the expectation of the ith mode's time course when multiplied by that of the jth mode is:

$$\langle \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\rangle(i,j) = \sum_{t=1}^{T}\hat{\boldsymbol{M}}_{\boldsymbol{B}}(i,t)\hat{\boldsymbol{M}}_{\boldsymbol{B}}(j,t) + \sum_{t=1}^{T}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}\big((i-1)T+t,(j-1)T+t\big) \tag{C.15}$$

Firstly, we will derive an expression for the i,jth block of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$, which we will denote $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}^{[i,j]} \in \mathbb{R}^{T\times T}$. Note that the expression for blockwise matrix multiplication has a similar form to the standard expression in terms of indices, so

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{Y} \qquad \boldsymbol{Z}(i,j) = \sum_{k}\boldsymbol{X}(i,k)\boldsymbol{Y}(k,j) \qquad \boldsymbol{Z}^{[i,j]} = \sum_{k}\boldsymbol{X}^{[i,k]}\boldsymbol{Y}^{[k,j]} \tag{C.16}$$

provided the blocks are sized appropriately. Using Equation C.11,

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}^{[i,j]} &= \sum_{k1}\sum_{k2}\sum_{k3}\big(\big(\langle\psi\rangle\langle \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\big)^{-1}\otimes \boldsymbol{I}_{T}\big)^{[i,k1]}\big(\boldsymbol{U}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{U}_{\boldsymbol{K}}\big)^{[k1,k2]} \ \dots \\ &\qquad \dots\ \Big(\big(\boldsymbol{\Lambda}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1}+\boldsymbol{I}_{MT}\big)^{-1}\Big)^{[k2,k3]}\big(\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-1}\otimes\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\big)^{[k3,j]} \\ &= \sum_{k1}\sum_{k2}\big(\big(\langle\psi\rangle\langle \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\big)^{-1}\otimes \boldsymbol{I}_{T}\big)^{[i,k1]}\big(\boldsymbol{U}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{U}_{\boldsymbol{K}}\big)^{[k1,k2]} \ \dots \\ &\qquad \dots\ \Big(\big(\boldsymbol{\Lambda}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1}+\boldsymbol{I}_{MT}\big)^{-1}\Big)^{[k2,k2]}\big(\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-1}\otimes\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\big)^{[k2,j]} \\ &= \sum_{k1}\sum_{k2}\Big(\big(\langle\psi\rangle\langle \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\big)^{-1}(i,k1)\Big)\Big(\boldsymbol{U}_{\alpha\boldsymbol{P}}(k1,k2)\Big)\Big(\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-1}(k2,j)\Big) \ \dots \\ &\qquad \dots\ \boldsymbol{U}_{\boldsymbol{K}}\Big(\big(\boldsymbol{\Lambda}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1}+\boldsymbol{I}_{MT}\big)^{-1}\Big)^{[k2,k2]}\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\end{aligned} \tag{C.17}$$

Taking the trace over the blocks of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$ allows us to simplify the expression in Equation C.17. This gives:

$$\begin{aligned}\mathrm{Tr}\big(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}^{[i,j]}\big) &= \sum_{k1}\sum_{k2}\Big(\big(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\big)^{-1}(i,k1)\Big)\Big(\boldsymbol{U}_{\alpha\boldsymbol{P}}(k1,k2)\Big)\Big(\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-1}(k2,j)\Big)\ \ldots \\ &\qquad \ldots\ \mathrm{Tr}\Big(\boldsymbol{U}_{\boldsymbol{K}}\Big(\big(\boldsymbol{\Lambda}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1}+\boldsymbol{I}_{MT}\big)^{-1}\Big)^{[k2,k2]}\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\Big) \\ &= \sum_{k1}\sum_{k2}\Big(\big(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\big)^{-1}(i,k1)\Big)\Big(\boldsymbol{U}_{\alpha\boldsymbol{P}}(k1,k2)\Big)\Big(\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-1}(k2,j)\Big)\ \ldots \\ &\qquad \ldots\ \mathrm{Tr}\Big(\Big(\big(\boldsymbol{\Lambda}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1}+\boldsymbol{I}_{MT}\big)^{-1}\Big)^{[k2,k2]}\Big)\end{aligned} \tag{C.18}$$

At this point we can use this form to calculate each of the elements of $\langle\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\rangle$. However, we can turn the above back into a matrix multiplication, and calculate all of $\langle\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\rangle$ in one go. To do this, form the diagonal matrix $\widehat{\boldsymbol{L}} \in \mathbb{R}^{M\times M}$, where $\widehat{\boldsymbol{L}}(m,m) = \sum_{t=1}^{T}\boldsymbol{L}(m,t)$. Using this, we can combine Equation C.15 and Equation C.18 to give:

$$\langle\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\rangle = \hat{\boldsymbol{M}}_{\boldsymbol{B}}\hat{\boldsymbol{M}}_{\boldsymbol{B}}^{\mathrm{T}} + \big(\langle\psi\rangle\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\rangle\big)^{-1}\boldsymbol{U}_{\alpha\boldsymbol{P}}\,\widehat{\boldsymbol{L}}\,\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-1} \tag{C.19}$$

Extension to blockwise trace with matrix multiplication

It is relatively straightforward to extend the above result to $\langle\boldsymbol{B}\boldsymbol{K}_{\boldsymbol{B}}^{-1}\boldsymbol{B}^{\mathrm{T}}\rangle$. It can be shown that the terms that involve $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$ that we require take the form $\mathrm{Tr}\big(\boldsymbol{K}_{\boldsymbol{B}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}^{[i,j]}\big)$. Note that if we multiply the expression for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}^{[i,j]}$ in Equation C.17 by $\boldsymbol{K}_{\boldsymbol{B}}^{-1}$ then the eigenvectors cancel and we are simply left with a modified diagonal matrix of eigenvectors.

Therefore, define $\boldsymbol{L}_{\boldsymbol{K}} \in \mathbb{R}^{M\times T}$ be a modification of $\boldsymbol{L}$, such that $\boldsymbol{L}_{\boldsymbol{K}}(m,t) = \boldsymbol{\Lambda}_{\boldsymbol{K}}(t,t)^{-1}\boldsymbol{L}(m,t)$. Then, defining $\widehat{\boldsymbol{L}}_{\boldsymbol{K}}$ in an analogous manner to $\widehat{\boldsymbol{L}}$, the extension

to Equation C.19 is trivial, yielding:

$$\left\langle \boldsymbol{B}\boldsymbol{K}_{\boldsymbol{B}}^{-1}\boldsymbol{B}^{\mathrm{T}}\right\rangle = \hat{\boldsymbol{M}}_{\boldsymbol{B}}\boldsymbol{K}_{\boldsymbol{B}}^{-1}\hat{\boldsymbol{M}}_{\boldsymbol{B}}^{\mathrm{T}} + \left(\left\langle\psi\right\rangle\left\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\right\rangle\right)^{-1}\boldsymbol{U}_{\alpha\boldsymbol{P}}\,\hat{\boldsymbol{L}}_{\boldsymbol{K}}\,\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-1} \tag{C.20}$$

Finally, rather than calculating $\hat{\boldsymbol{M}}_{\boldsymbol{B}}\boldsymbol{K}_{\boldsymbol{B}}^{-1}\hat{\boldsymbol{M}}_{\boldsymbol{B}}^{\mathrm{T}}$ directly, note that Equation C.14 gives an expression for $\hat{\boldsymbol{M}}_{\boldsymbol{B}}$ involving terms from the eigendecomposition of $\boldsymbol{K}_{\boldsymbol{B}}$. Some of these will cancel, allowing calculations to proceed without explicitly forming $\boldsymbol{K}_{\boldsymbol{B}}^{-1}$.

Determinant

This is straightforward to calculate using the expression for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}$ in Equation C.11. There are a few properties of the determinant we need in order to proceed. Firstly, $|\boldsymbol{X}\boldsymbol{Y}| = |\boldsymbol{X}||\boldsymbol{Y}|$. Secondly, $|\boldsymbol{X}^{-1}| = |\boldsymbol{X}|^{-1}$. Thirdly, for $\boldsymbol{X} \in \mathbb{R}^{M\times M}$ and $\boldsymbol{Y} \in \mathbb{R}^{N\times N}$, $|\boldsymbol{X}\otimes\boldsymbol{Y}| = |\boldsymbol{X}|^{N}|\boldsymbol{Y}|^{M}$. Combining the above gives:

$$\begin{aligned}\left|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}\right| &= \left|\left(\left\langle\psi\right\rangle\left\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\right\rangle\right)^{-1}\otimes\boldsymbol{I}_{T}\right|\left|\boldsymbol{U}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{U}_{\boldsymbol{K}}\right|\left|\boldsymbol{\Lambda}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1}+\boldsymbol{I}_{MT}\right|^{-1}\left|\boldsymbol{U}_{\alpha\boldsymbol{P}}^{-1}\otimes\boldsymbol{U}_{\boldsymbol{K}}^{\mathrm{T}}\right| \\ &= \left|\left\langle\psi\right\rangle\left\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\right\rangle\right|^{-T}\left|\boldsymbol{\Lambda}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1}+\boldsymbol{I}_{MT}\right|^{-1}\end{aligned} \tag{C.21}$$

The free energy requires the log of the determinant, which follows straightforwardly from Equation C.21. Note that we can reuse the matrix of eigenvalues, $\boldsymbol{L}$, from previous calculations.

$$\begin{aligned}\ln\left(\left|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{B}}\right|\right) &= -T\ln\left(\left|\left\langle\psi\right\rangle\left\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\right\rangle\right|\right) + \ln\left(\left|\boldsymbol{\Lambda}_{\alpha\boldsymbol{P}}\otimes\boldsymbol{\Lambda}_{\boldsymbol{K}}^{-1}+\boldsymbol{I}_{MT}\right|^{-1}\right) \\ &= -T\ln\left(\left|\left\langle\psi\right\rangle\left\langle\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\right\rangle\right|\right) + \sum_{t=1}^{T}\sum_{m=1}^{M}\ln\left(\boldsymbol{L}(m,t)\right)\end{aligned} \tag{C.22}$$

Appendix D

Parameters

D.1 Pre-processing

We apply two further pre-processing steps to the HCP data. For each run independently, we demean the time series from each grayordinate and then renormlaise such that the time series has unit variance. The data is then reduced via the random SVD, from which we retain 500 components.

D.2 Alternative Models

We use several other models other than those outlined in Chapter 3. In the spatial domain, we turn off subject variability by defining a set of group maps, $\boldsymbol{P}^{(g)}$, and setting $\boldsymbol{P}^{(s)} = \boldsymbol{P}^{(g)}$ for all subjects s. We use two models for $\boldsymbol{P}^{(g)}$, both based on the hyperprior for the group-level μ parameters. In fact, the λ and γ parameters take the same values as in the full model.

We define the group-level binary mixture model prior as:

$$\begin{aligned} \mathrm{p}\big(\boldsymbol{P}_{vm}^{(g)} \,\big|\, \rho_{vm} = 1\big) &= \delta\big(\boldsymbol{P}_{vm}^{(g)} - 1\big) \\ \mathrm{p}\big(\boldsymbol{P}_{vm}^{(g)} \,\big|\, \rho_{vm} = 0\big) &= \delta\big(\boldsymbol{P}_{vm}^{(g)}\big) \\ \mathrm{p}\big(\rho_{vm}\big) &= (\lambda)^{\rho_{vm}}(1-\lambda)^{1-\rho_{vm}} \end{aligned} \tag{D.1}$$

We also define the group-level delta-Gaussian mixture model prior as:

$$\begin{aligned} \mathrm{p}\big(\boldsymbol{P}_{vm}^{(g)} \,\big|\, \rho_{vm} = 1\big) &= \mathcal{N}\big(\boldsymbol{P}_{vm}^{(g)} \,\big|\, 0, \gamma^{-1}\big) \\ \mathrm{p}\big(\boldsymbol{P}_{vm}^{(g)} \,\big|\, \rho_{vm} = 0\big) &= \delta\big(\boldsymbol{P}_{vm}^{(g)}\big) \\ \mathrm{p}\big(\rho_{vm}\big) &= (\lambda)^{\rho_{vm}}(1-\lambda)^{1-\rho_{vm}} \end{aligned} \tag{D.2}$$

We employ a specific approach in the temporal domain, and turn off the hierarchical model over precision matrices by setting $\boldsymbol{\alpha}^{(s)} = \boldsymbol{\beta}$ for all subjects s. In this case we use a group-level diagonal prior for $\boldsymbol{\beta}$, in a manner similar to the noise covariance matrices. This takes the form:

$$\mathrm{p}\big(\boldsymbol{\beta}_{ij}\big) = \begin{cases} \Gamma\big(\boldsymbol{\beta}_i \,\big|\, c_{\beta_i}, d_{\beta_i}\big) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{D.3}$$

D.3 Inference Procedure

We start with the group-level binary mixture model in the spatial domain, and similarly adopt the group-level prior for $\boldsymbol{\beta}$. The spatial maps are initialised with a random draw from this prior, except for one mode where $\{\rho_{vm} = 1\}_{v=1}^{V}$. We infer on this model for 100 iterations of the update rules.

Next, we switch to the group-level spatial delta-Gaussian mixture model, holding all other approximate distributions constant, and infer on that new model for 100 iterations.

The next change is to switch to the full spatial model, and infer on that for 150 iterations.

Next, the subject-variability in the temporal domain is switched on and a further 50 iterations of the VB update rules are computed.

At this stage, in order to ensure that the subject-level modes do not diverge too far from the group-level means, we turn off spatial variability and switch back to the group-level delta-Gaussian mixture model for 100 iterations.

We switch back to the full spatial model, infer on this for 500 iterations, re-initialise the temporal model, and finally compute the VB update rules a further 1,000 times.

It is the output from the full model, after the combined 2,000 iterations of the VB update rules, that we report in the main body of the thesis.

D.4 Prior Parameters

The full set of prior parameters can be found in Table D.1.

Parameter	Value
λ	0.05
γ	1
$a_{\sigma_{vm}}$	$0.2\,S$
$b_{\sigma_{vm}}$	$0.02\,S$
$a_{\pi_{vm}}$	1
$b_{\pi_{vm}}$	4
$a_{\boldsymbol{\alpha}^{(s)}}$	M
a_{β}	M
$\boldsymbol{B}_{\beta}$	$M\,\boldsymbol{I}_M$
c_{β}	50
d_{β}	
$a_{\omega_m^{(sr)}}$	50
$b_{\omega_m^{(sr)}}$	10
a_{ψ}	1
b_{ψ}	
η	0.1

Table D.1: Prior parameters used when inferring on the PFM model.

Appendix E

Simulations

In this appendix, we give a brief overview of the simulations used to validate the PFM model and inference approach. A more comprehensive description can be found in Harrison et al. [2015], and full details can be found in the supplementary material associated with that paper.

E.1 Data Generation

The aim was to simulate a scaled-down version of the HCP data, though for computational reasons we reduced the number of grayordinates and subjects to 12,500 and 30 respectively. However, the number of runs per subject, time points per run, and TR remained constant.

In the spatial domain, we simulated a set of parcels and then generated modes by taking weighted combinations of these. The parcels represented a set of 200 spatially contiguous, binary regions, with smooth spatial warps applied to generate subject-specific parcellations with misalignments. A set of 25 modes was generated by taking weighted combinations of the parcels, and these modes included negative weights to simulate anti-correlations. There was a tendency for modes to cover a greater continuous spatial extent than parcels, and for modes to be correlated with

one another. Sparse noise was added to the group-level mode weights to generate subject-specific variants. The subject-level modes in the space of the full set of grayordinate could be constructed by combining the subject-specific parcellations and the subject-specific mode weights.

In the temporal domain, 'neural' time courses were simulated at a sampling frequency of 10 Hz, and these time courses were both sparse, correlated with one another, and contained an excess of power below 0.1 Hz.

The BOLD signal was constructed by convolving the time courses with a random HRF (generated from the FLOBS basis set [Woolrich et al. 2004]), down-sampling, and projecting into the full grayordinate space via the subject-specific mode spatial maps. A sigmoid function was then applied to these voxelwise BOLD signal time series to mimic the nonlinear saturations of a more complex hæmodynamic model.

Finally, noise was added to the BOLD signal to achieve an overall signal-to-noise ratio of −10 dB.

A complete description of this procedure, including all the parameters used, can be found in the supplementary material of the aforementioned paper. This also includes several images of the simulated data itself.

E.2 Models Tested

We compared the PFM approach to PCA and several flavours of ICA: sICA, tICA and 'stICA'. The last of those, stICA, is an approach where sICA is used to identify the 200 parcels, and tICA is run in that data subspace. All of the ICA methods are also presented after a dual regression step to extract subject-specific information.

The PFMs were based on a slightly different model to the one outlined in Chapter 3, with a simpler temporal model. The noise modelling at the time course level was absent, which is equivalent to setting $\boldsymbol{A}^{(sr)} = \boldsymbol{B}^{(sr)}$, and we used the

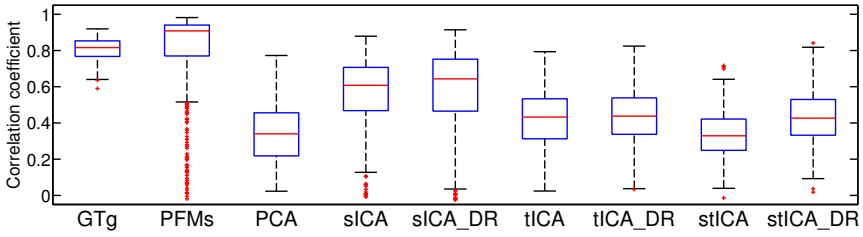
group-level diagonal prior for $\boldsymbol{\beta}$ as outlined in Appendix D.

E.3 Results

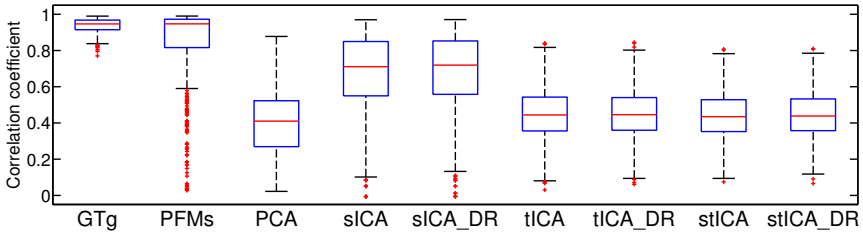
We show the accuracy, relative to the ground truth, in figure E.1, and PFMs perform considerably better than any of the other approaches.

We also show the relationship between ground truth accuracy and test-retest reliability (a very similar test to split-half reproducibility) in figure E.2. There is a consistent relationship between these two quantities for the PFMs, whereas the ICA methods display a consistent bias and ground truth accuracy is essentially independent of test-retest reliability.

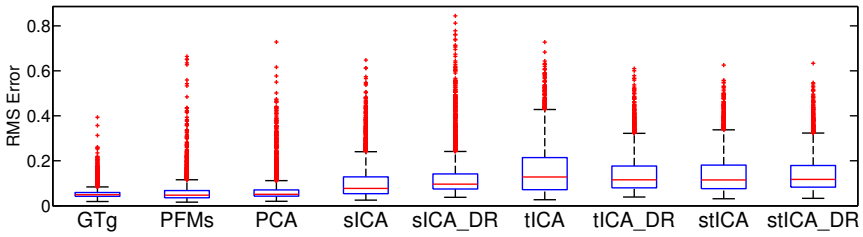
Several more plots, including a demonstration that the inferred set of PFMs is relatively robust to mis-specification of the true number of modes, can be found in the supplementary material of the paper.



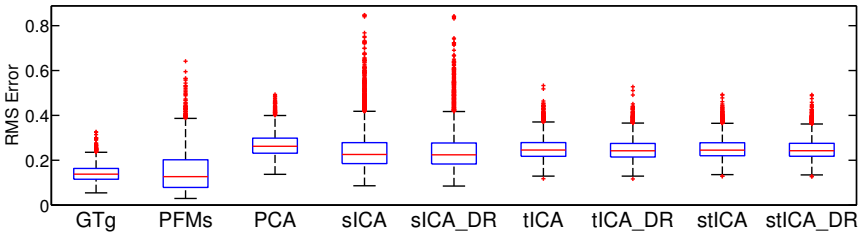
(a) Spatial maps.



(b) Time courses.

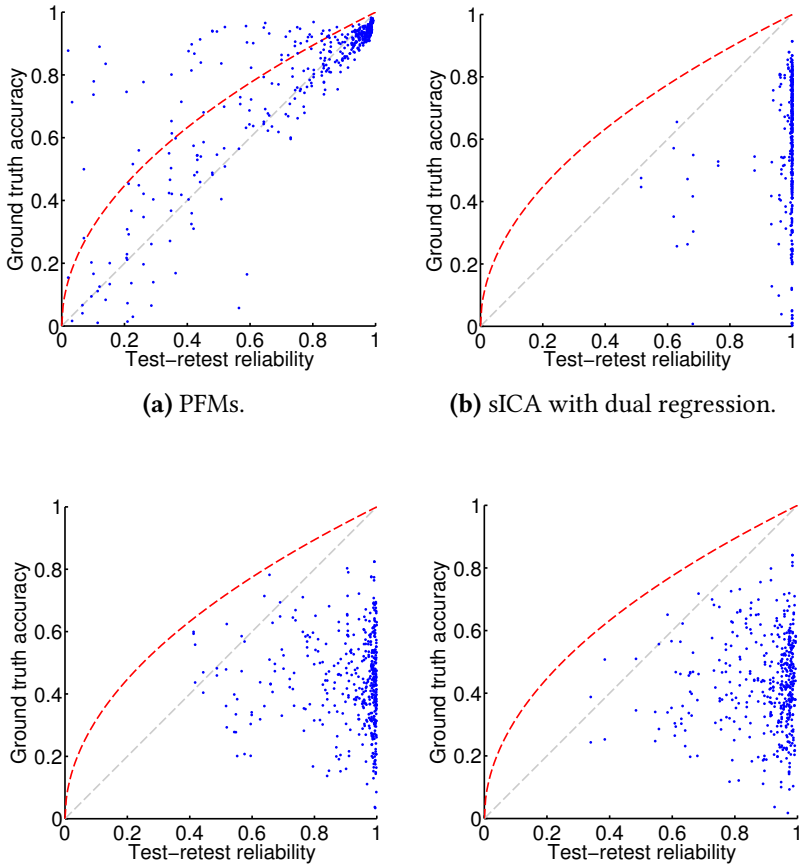


(c) Spatial correlation coefficients.



(d) Temporal correlation coefficients.

Figure E.1: Accuracy in recovery of the ground truth on simulated data. Multiple data sets were simulated, and the accuracy with which the above data characteristics are inferred, for each of the 25 modes, is shown for each method. Dual regression is indicated by the suffix DR.
GTg illustrates the scores that are achieved if the subject maps are just set to the mean of the ground truth subject maps; as such, it both illustrates the amount of subject variability in the data and is also a useful benchmark for those methods which do not model individual subjects.



(a) PFMs.

(b) sICA with dual regression.

(c) tICA with dual regression.

(d) stICA with dual regression.

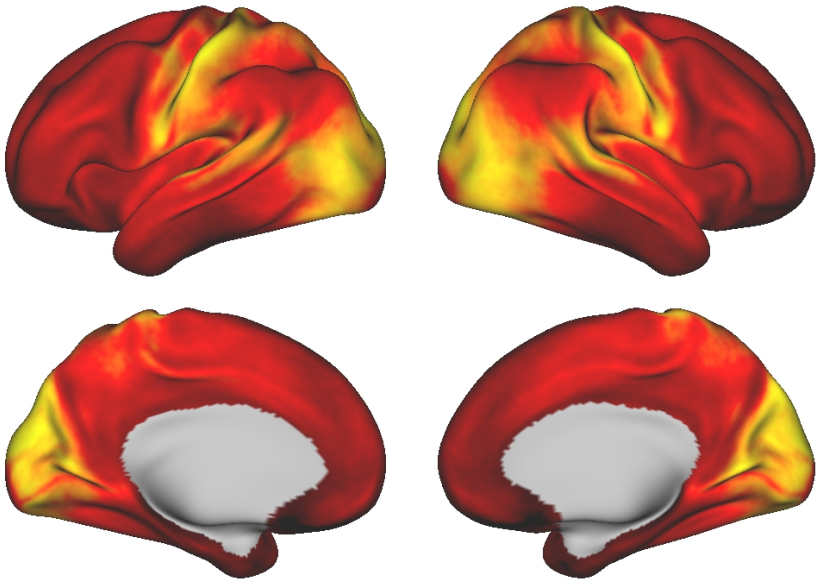
Figure E.2: Accuracy in recovery of ground truth subject-specific spatial maps plotted against test-retest reliability for four methods tested on simulated data. Each method was run twice on the same data set; both the accuracy scores, as plotted in figure E.1(a), and the test-retest reliability, scored using the same correlation metric, were calculated for each mode.
The grey line indicates equality between the two scores, whereas the red line indicates the range of scores possible if the inferred maps are just the ground truth maps with independent additive noise.

Appendix F

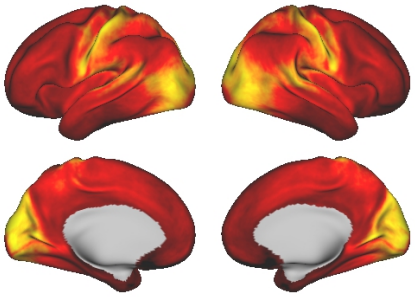
Group-Level Spatial Maps

In this appendix, we show the spatial maps for all 30 PFMs highlighted in the main body of the thesis. Furthermore, we show the PFMs used to generate the split-half reproducibilities and the ICA maps used for that comparison.

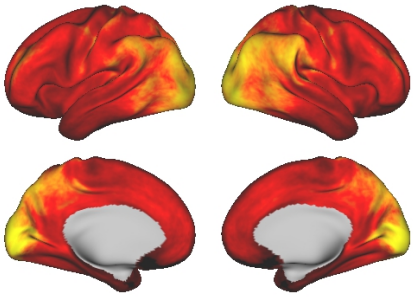
Only the cortex is shown, though the PFMs also contain subcortical and cerebellar regions. The colour scale for each map is clipped at the 99.5 % level of the distribution of the spatial weight magnitudes, and is symmetric around zero. Cortical surface views were generated using Connectome Workbench (humanconnectome.org/software/connectome-workbench.html).



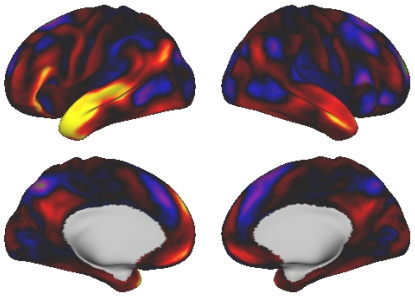
(a) PFM, all data.



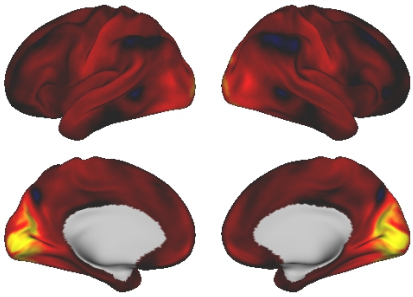
(b) PFM, first half of data.



(c) PFM, second half of data.

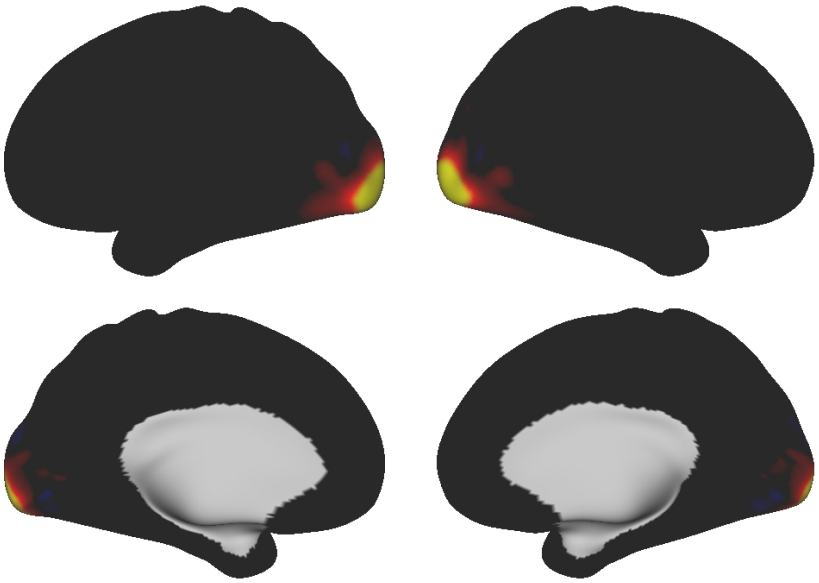


(d) sICA, all data.

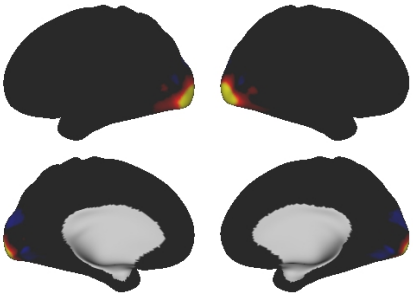


(e) tICA, all data.

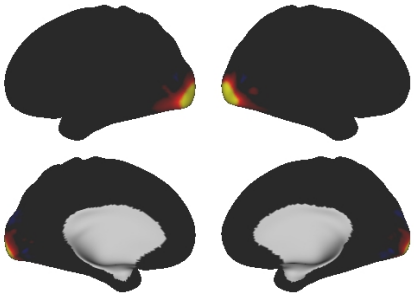
Figure F.1: PFM 1



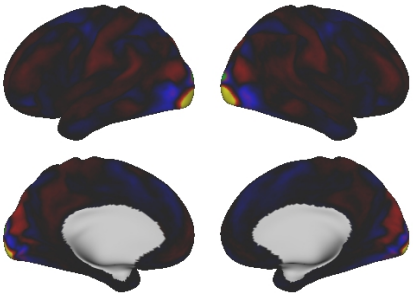
(a) PFM, all data.



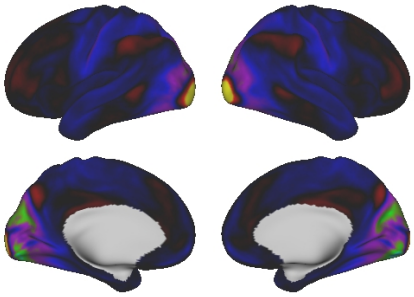
(b) PFM, first half of data.



(c) PFM, second half of data.

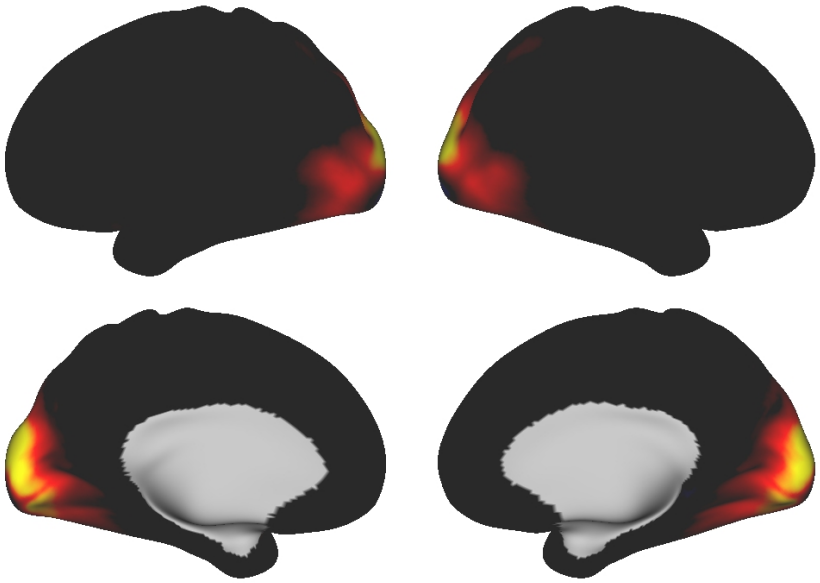


(d) sICA, all data.

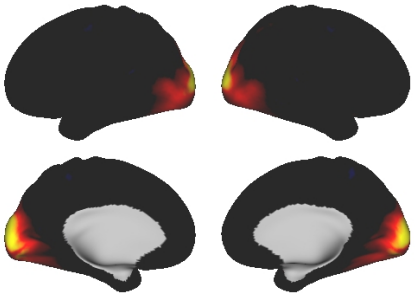


(e) tICA, all data.

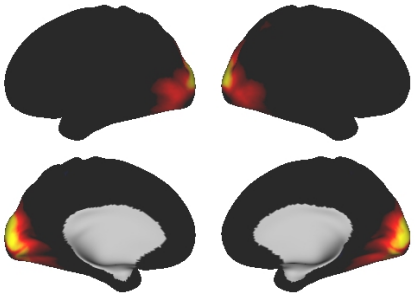
Figure F.2: PFM 2



(a) PFM, all data.

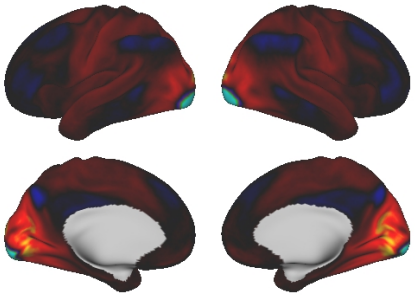


(b) PFM, first half of data.



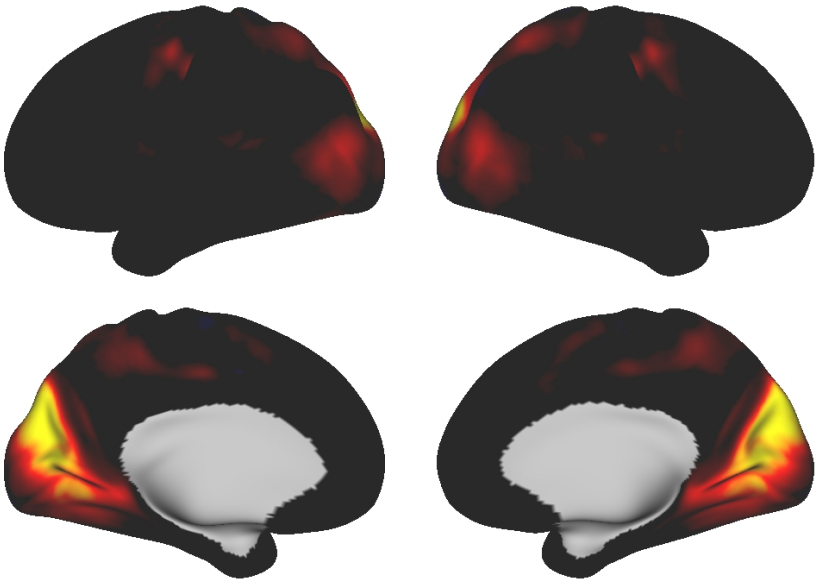
(c) PFM, second half of data.

(d) sICA, all data.

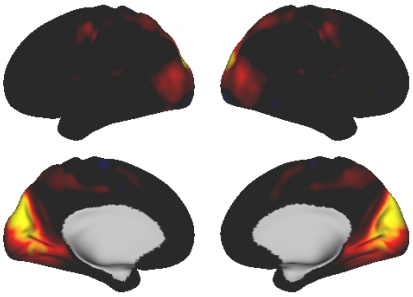


(e) tICA, all data.

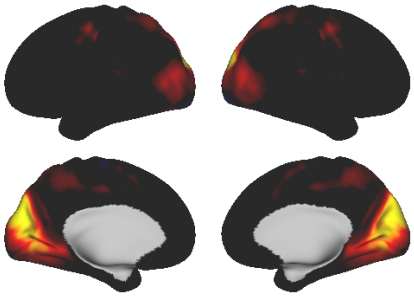
Figure F.3: PFM 3



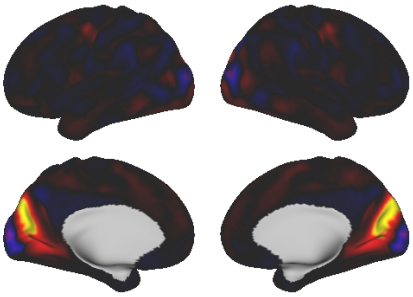
(a) PFM, all data.



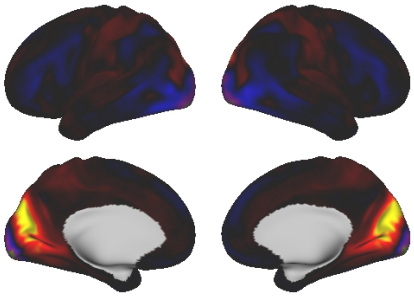
(b) PFM, first half of data.



(c) PFM, second half of data.

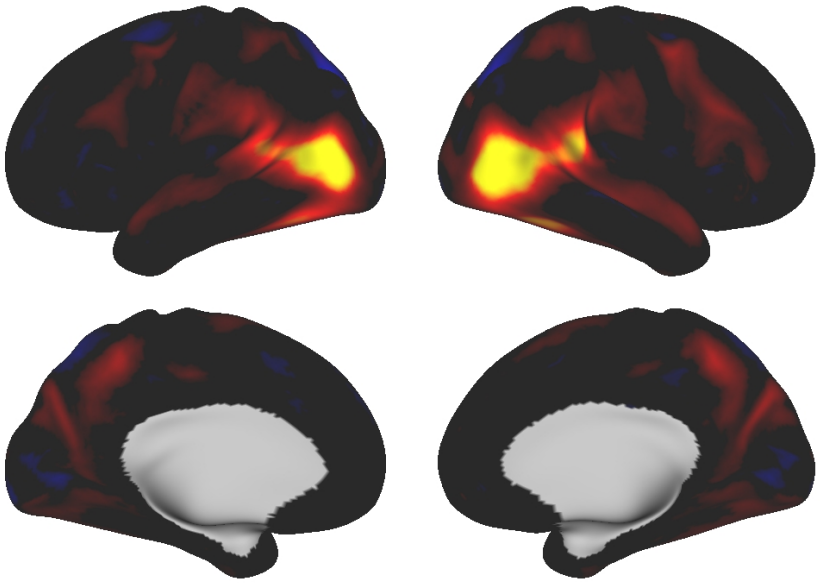


(d) sICA, all data.

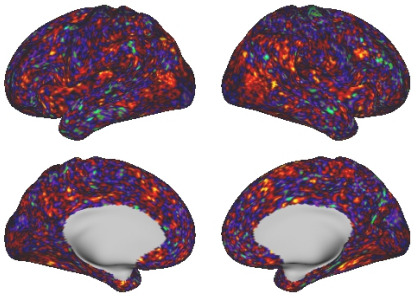


(e) tICA, all data.

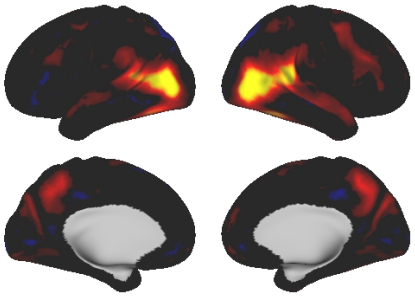
Figure F.4: PFM 4



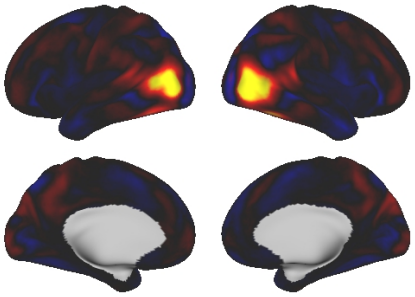
(a) PFM, all data.



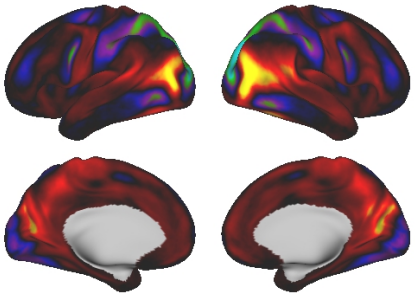
(b) PFM, first half of data.



(c) PFM, second half of data.

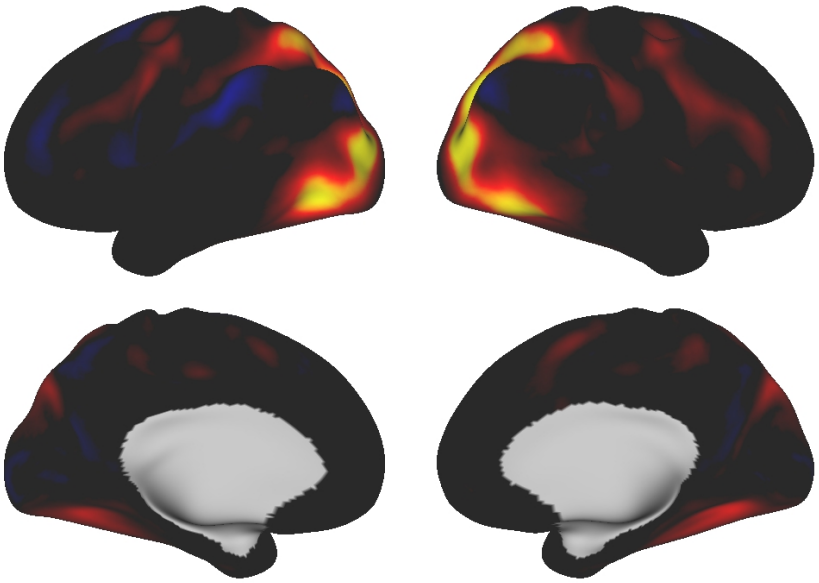


(d) sICA, all data.

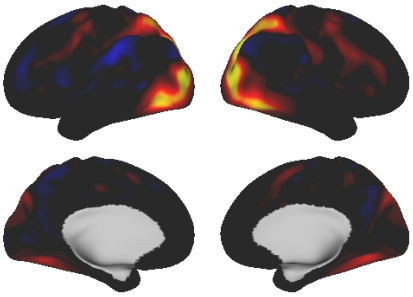


(e) tICA, all data.

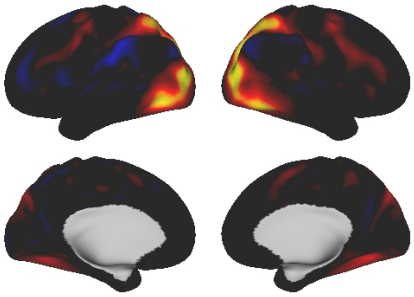
Figure F.5: PFM 5



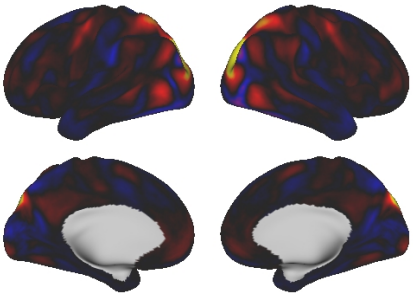
(a) PFM, all data.



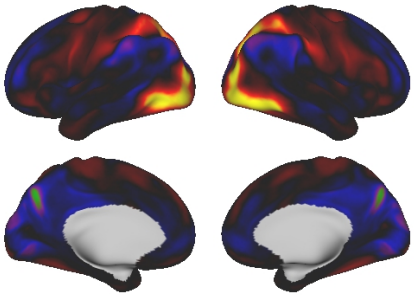
(b) PFM, first half of data.



(c) PFM, second half of data.

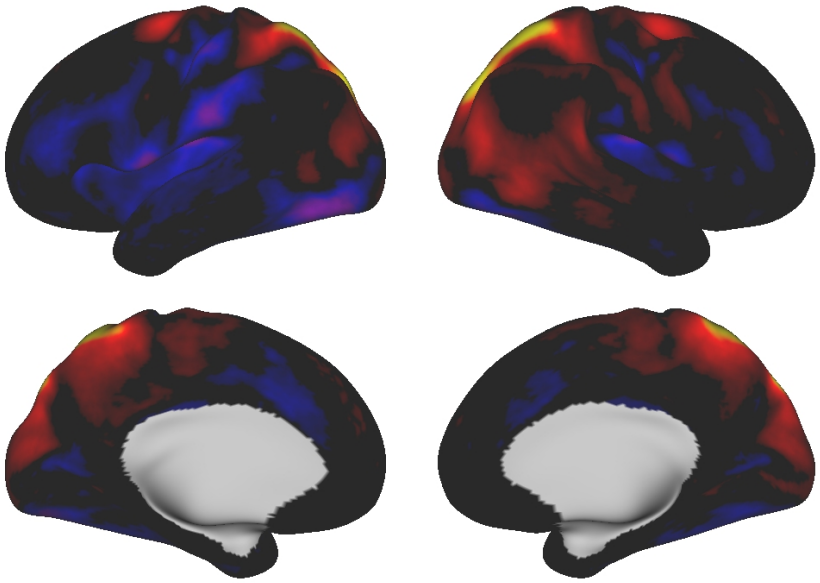


(d) sICA, all data.

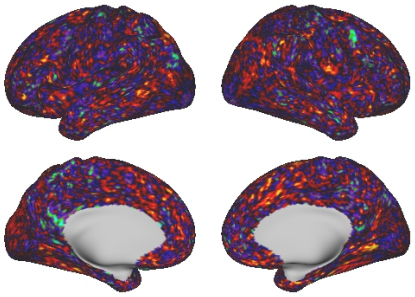


(e) tICA, all data.

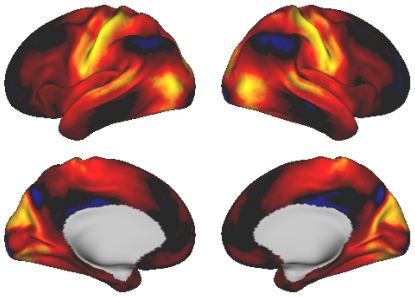
Figure F.6: PFM 6



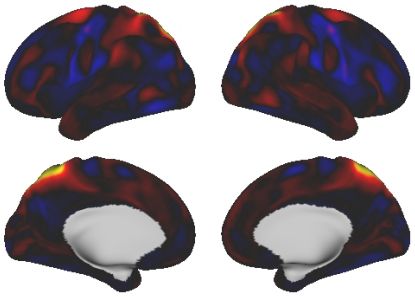
(a) PFM, all data.



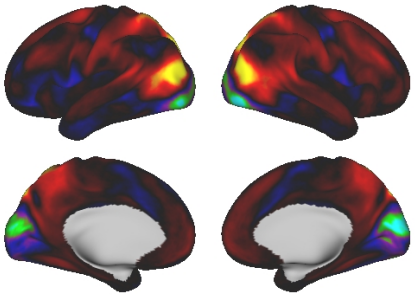
(b) PFM, first half of data.



(c) PFM, second half of data.

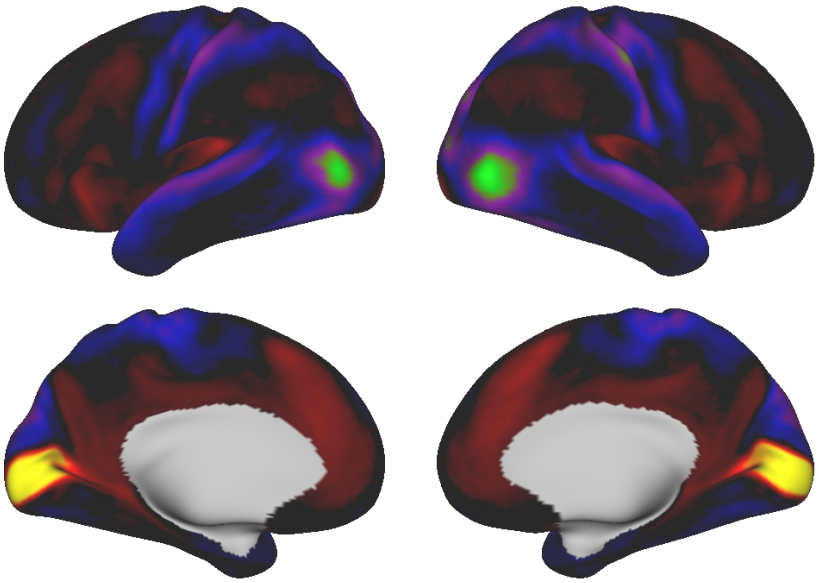


(d) sICA, all data.

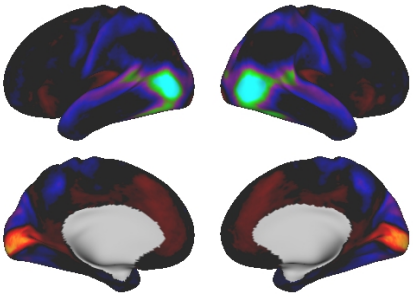


(e) tICA, all data.

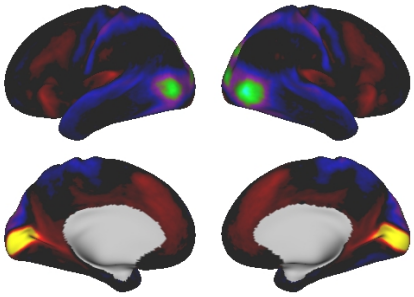
Figure F.7: PFM 7



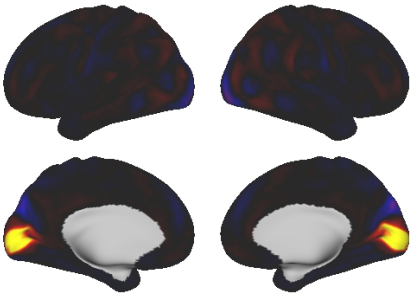
(a) PFM, all data.



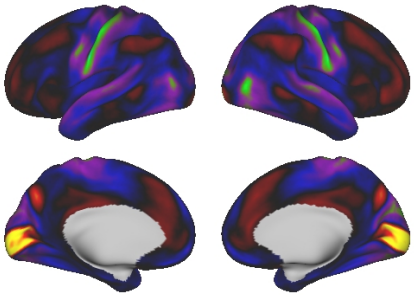
(b) PFM, first half of data.



(c) PFM, second half of data.

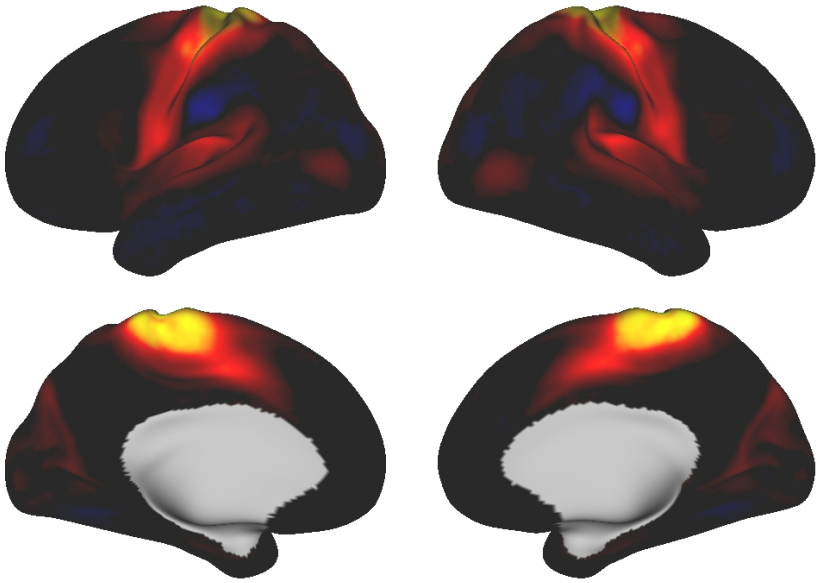


(d) sICA, all data.

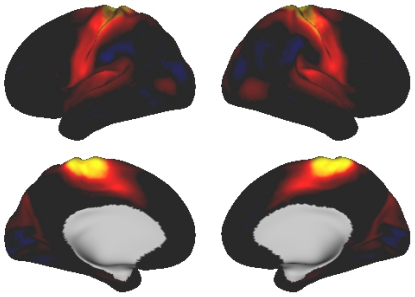


(e) tICA, all data.

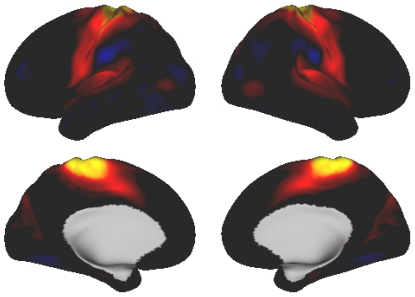
Figure F.8: PFM 8



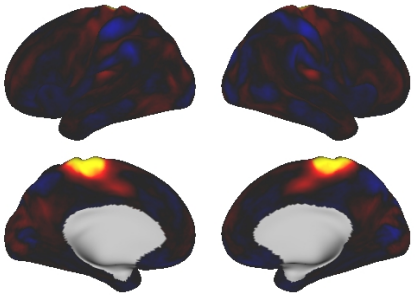
(a) PFM, all data.



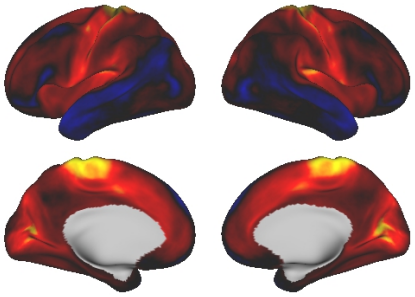
(b) PFM, first half of data.



(c) PFM, second half of data.

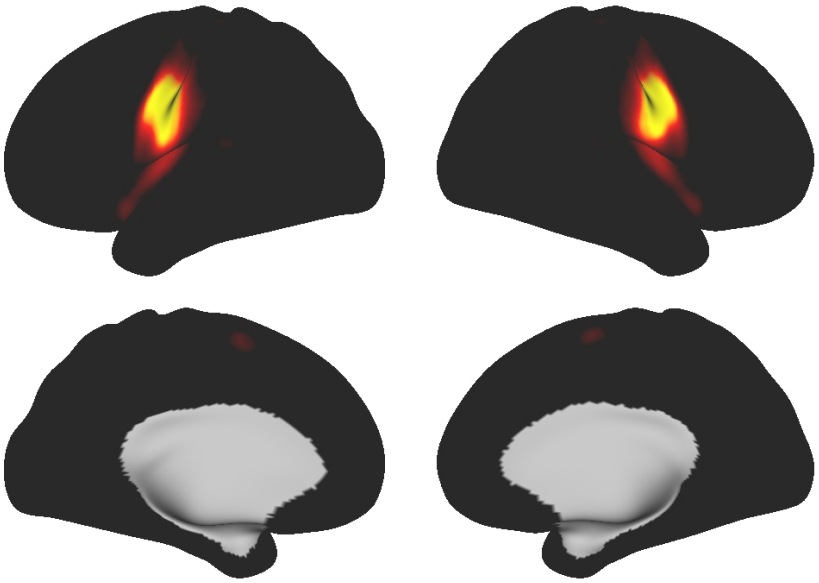


(d) sICA, all data.

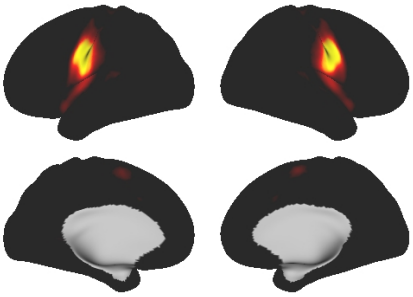


(e) tICA, all data.

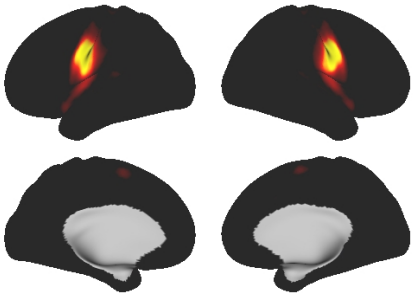
Figure F.9: PFM 9



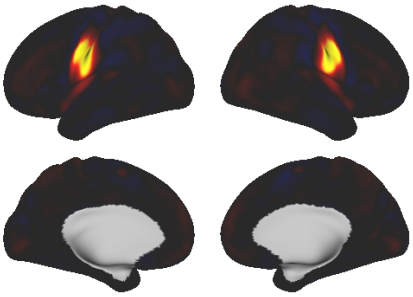
(a) PFM, all data.



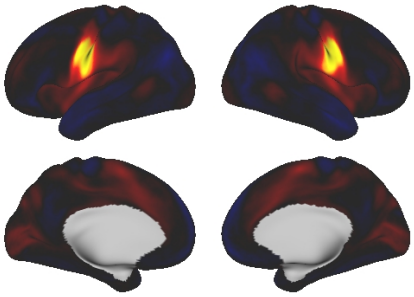
(b) PFM, first half of data.



(c) PFM, second half of data.

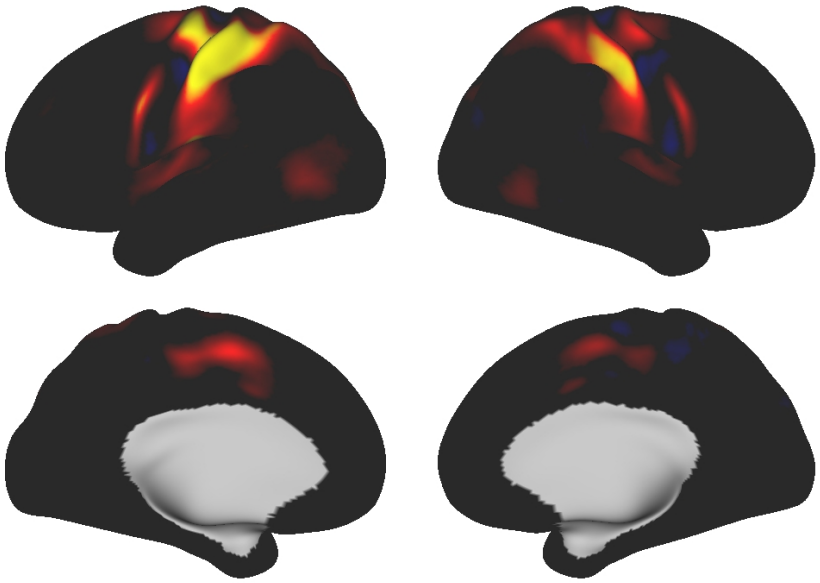


(d) sICA, all data.

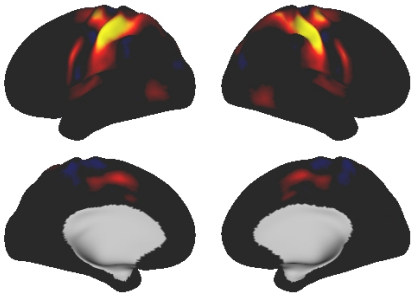


(e) tICA, all data.

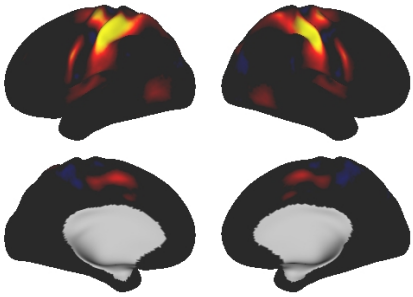
Figure F.10: PFM 10



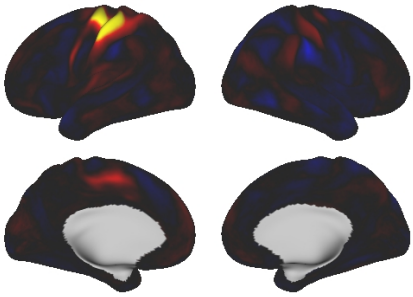
(a) PFM, all data.



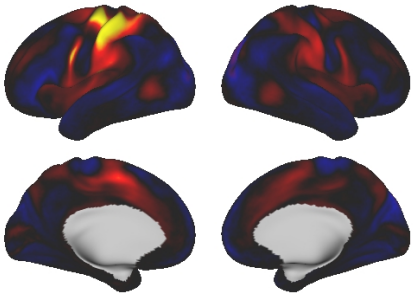
(b) PFM, first half of data.



(c) PFM, second half of data.

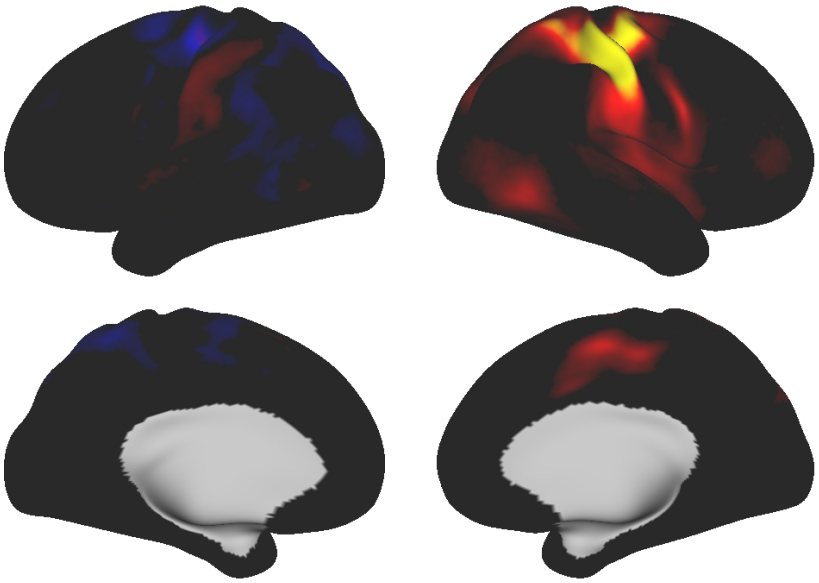


(d) sICA, all data.

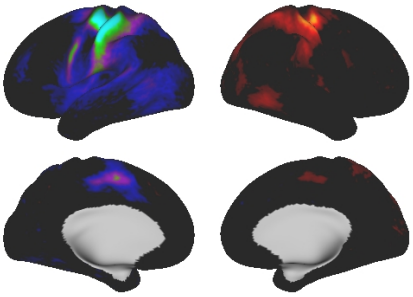


(e) tICA, all data.

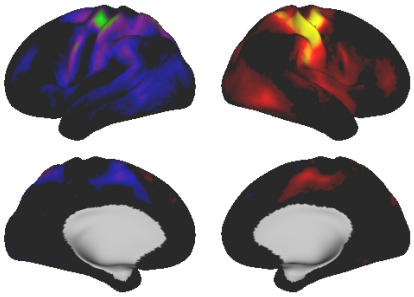
Figure F.11: PFM 11



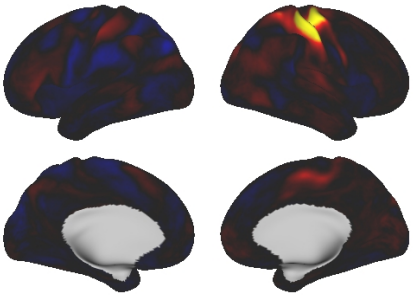
(a) PFM, all data.



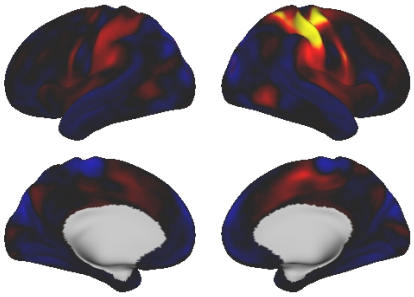
(b) PFM, first half of data.



(c) PFM, second half of data.

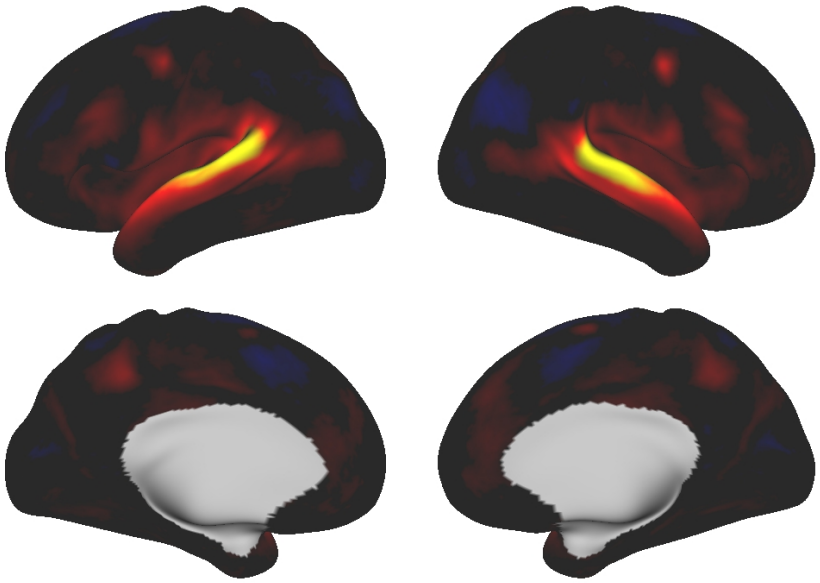


(d) sICA, all data.

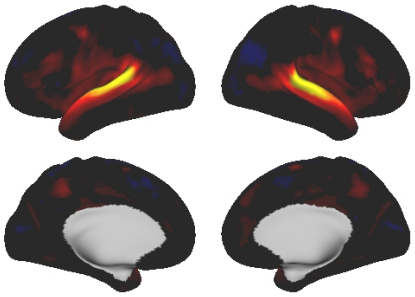


(e) tICA, all data.

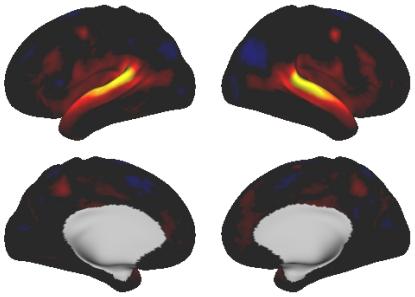
Figure F.12: PFM 12



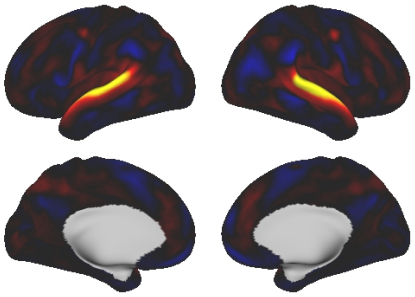
(a) PFM, all data.



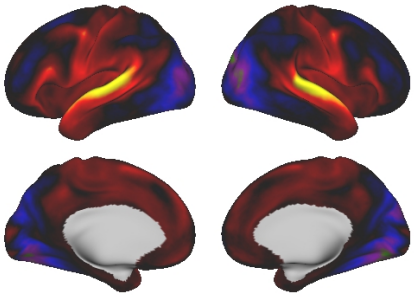
(b) PFM, first half of data.



(c) PFM, second half of data.

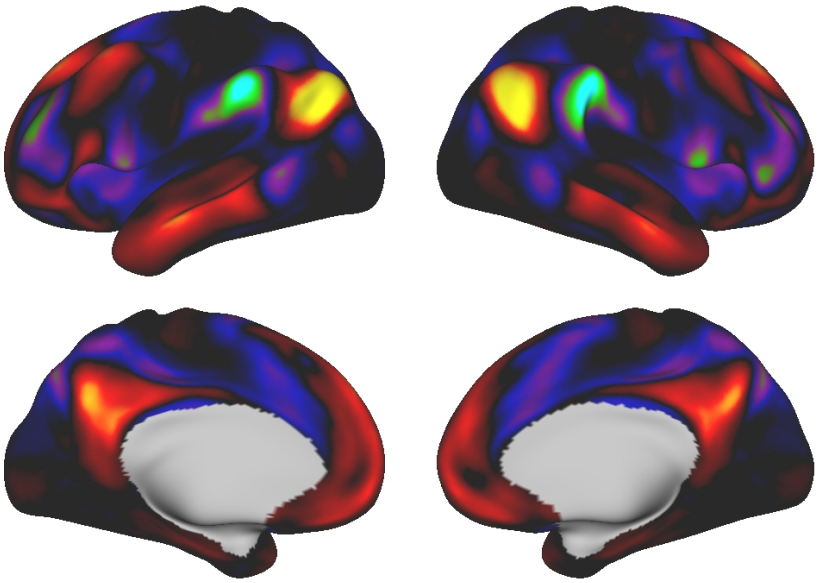


(d) sICA, all data.

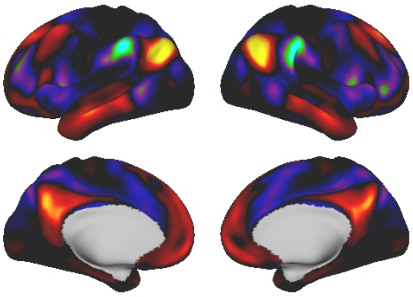


(e) tICA, all data.

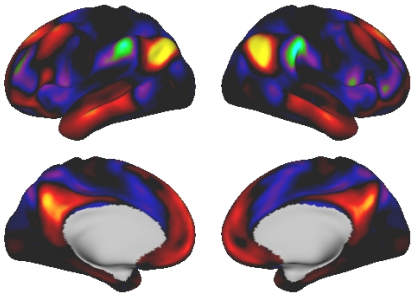
Figure F.13: PFM 13



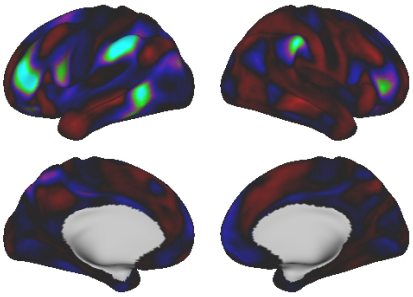
(a) PFM, all data.



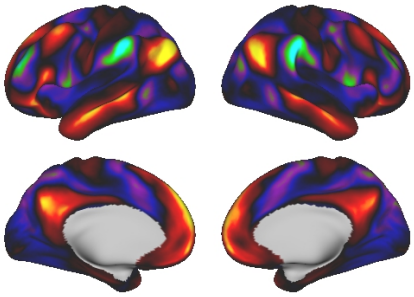
(b) PFM, first half of data.



(c) PFM, second half of data.

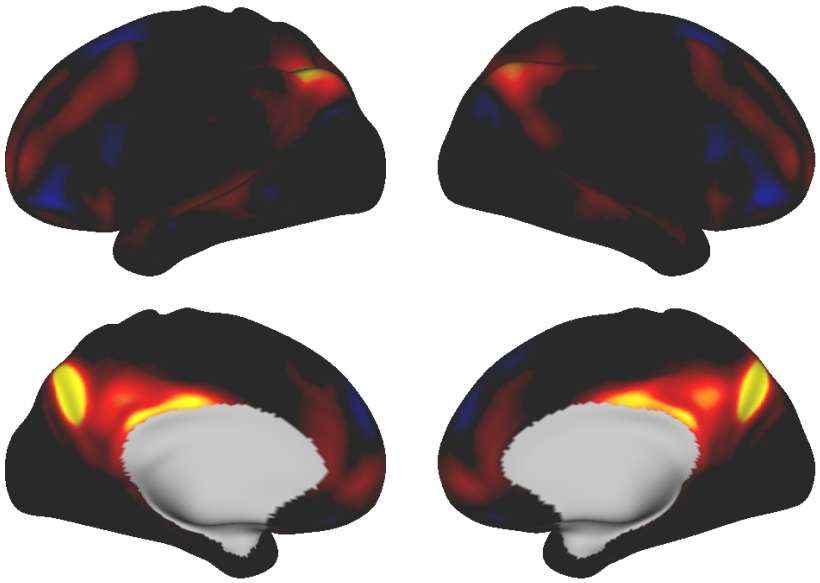


(d) sICA, all data.

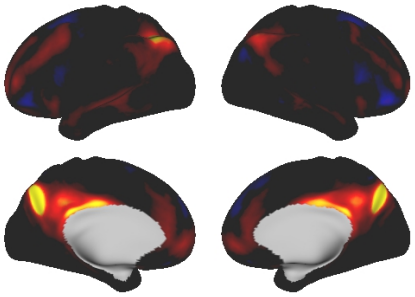


(e) tICA, all data.

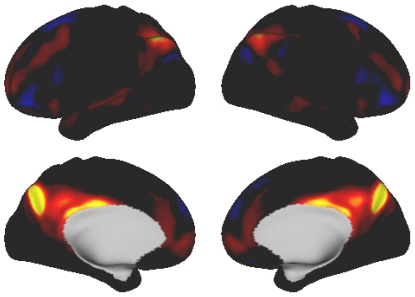
Figure F.14: PFM 14



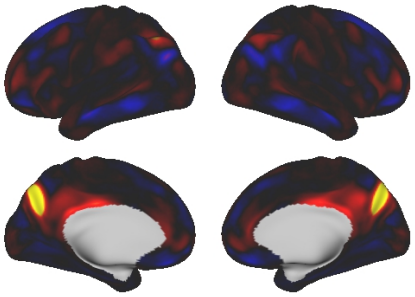
(a) PFM, all data.



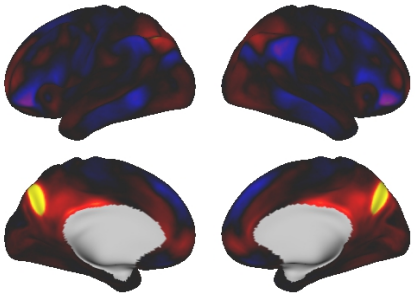
(b) PFM, first half of data.



(c) PFM, second half of data.

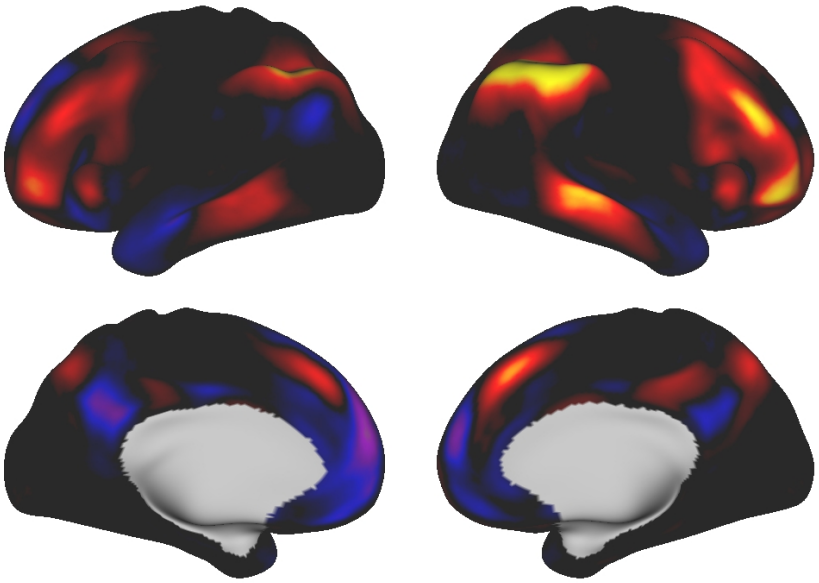


(d) sICA, all data.

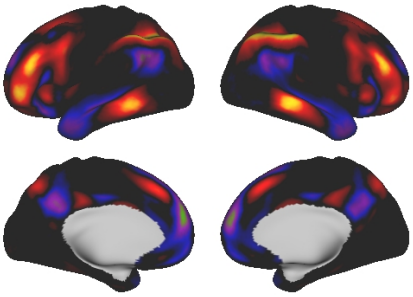


(e) tICA, all data.

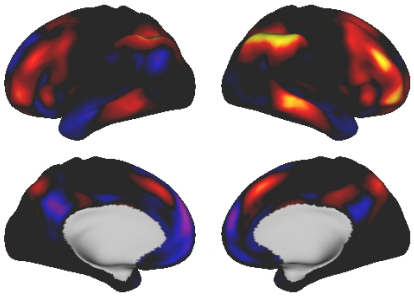
Figure F.15: PFM 15



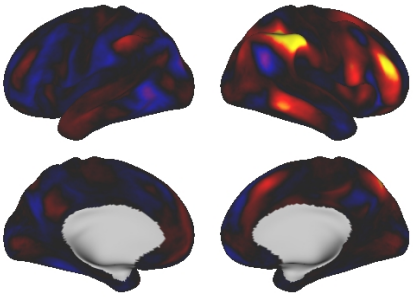
(a) PFM, all data.



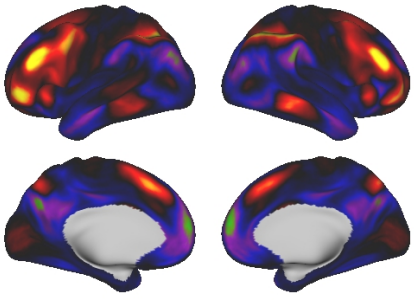
(b) PFM, first half of data.



(c) PFM, second half of data.

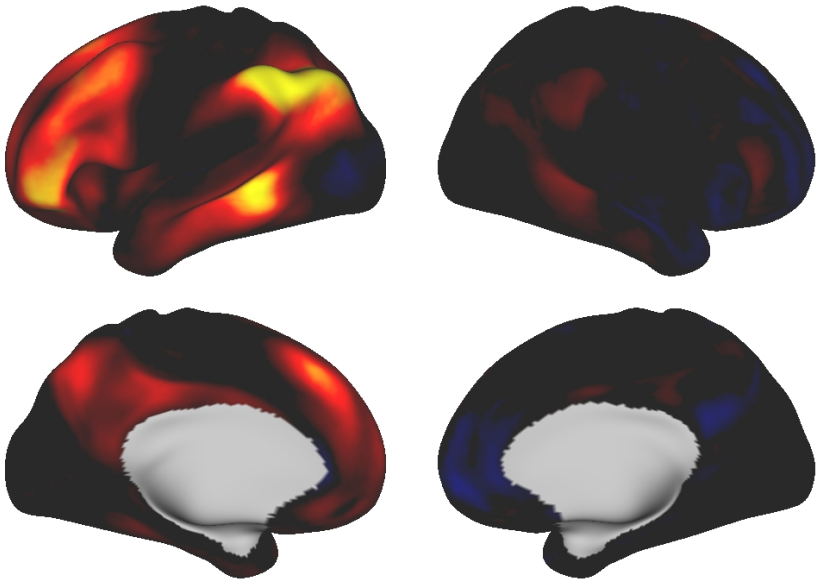


(d) sICA, all data.

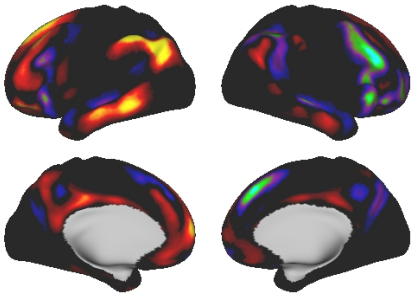


(e) tICA, all data.

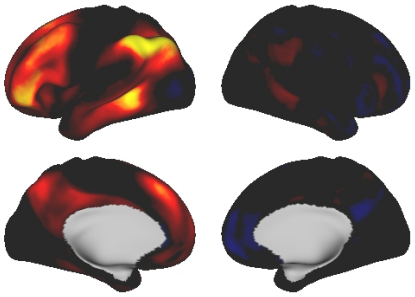
Figure F.16: PFM 16



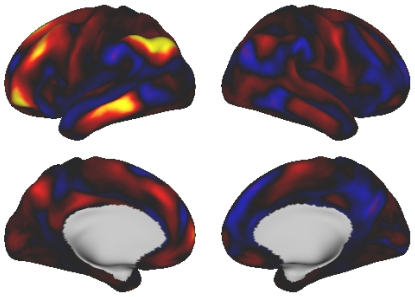
(a) PFM, all data.



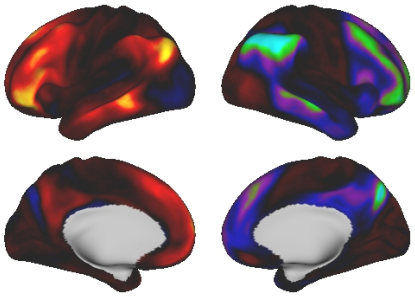
(b) PFM, first half of data.



(c) PFM, second half of data.

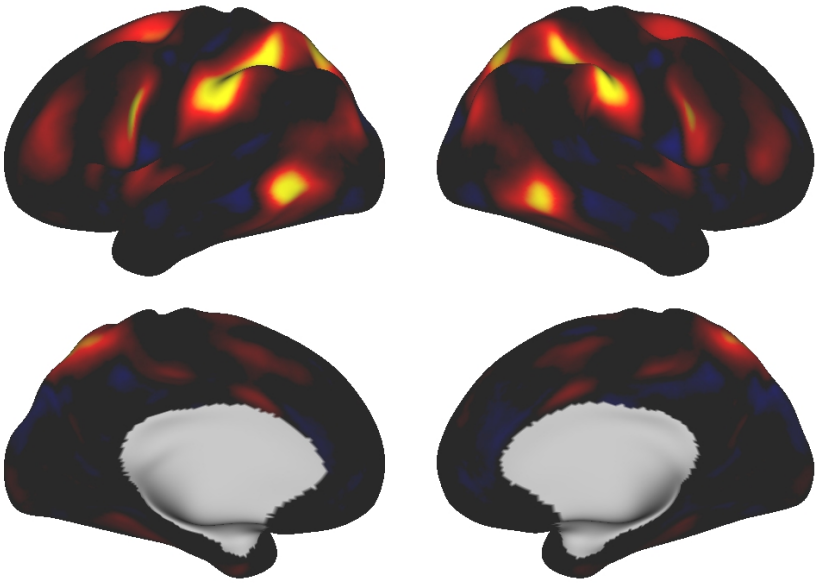


(d) sICA, all data.

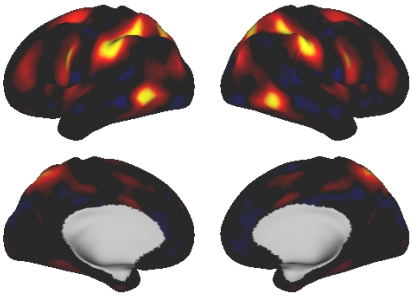


(e) tICA, all data.

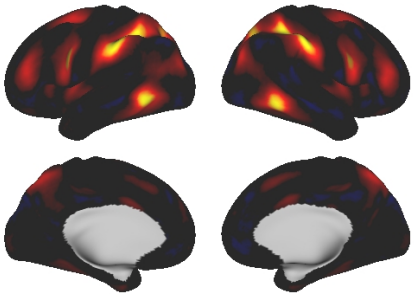
Figure F.17: PFM 17



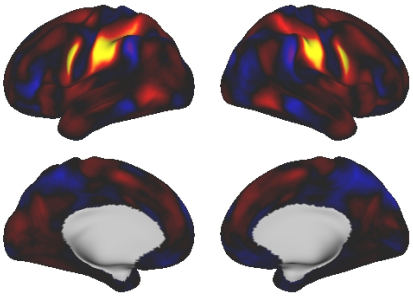
(a) PFM, all data.



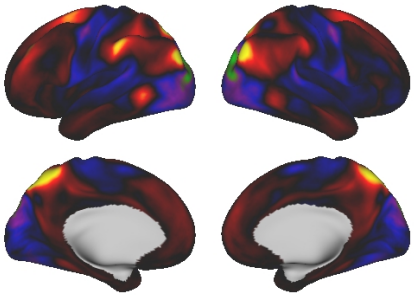
(b) PFM, first half of data.



(c) PFM, second half of data.

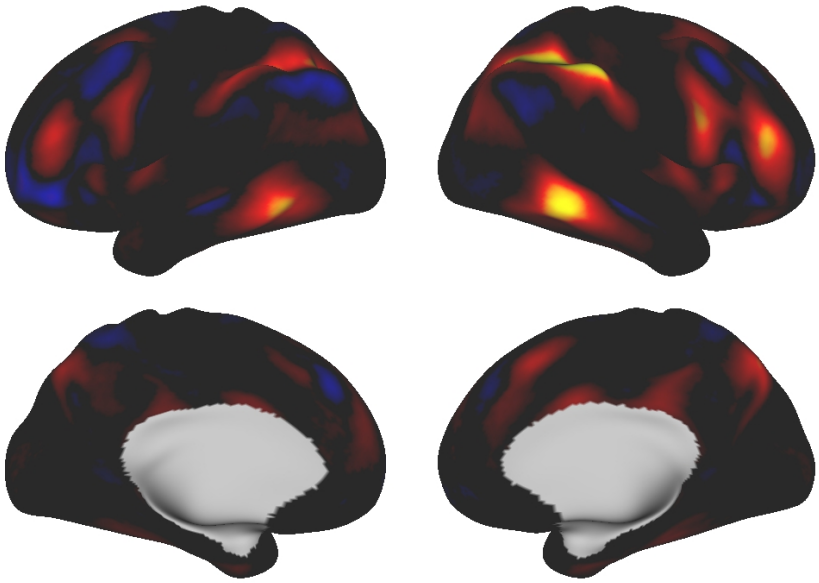


(d) sICA, all data.

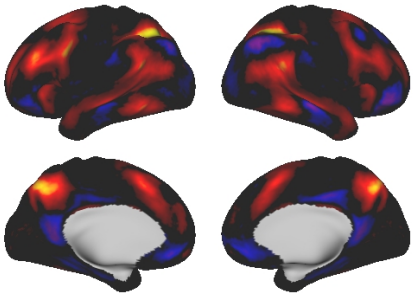


(e) tICA, all data.

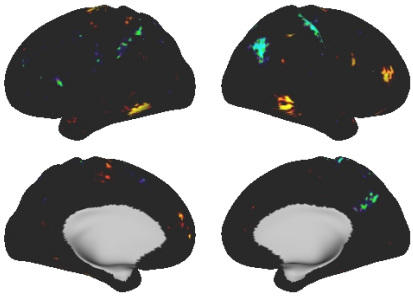
Figure F.18: PFM 18



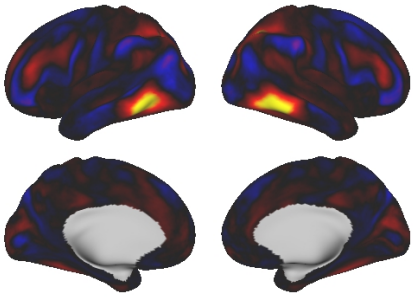
(a) PFM, all data.



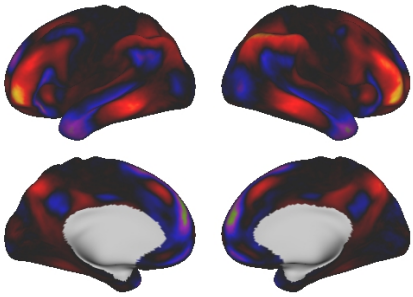
(b) PFM, first half of data.



(c) PFM, second half of data.

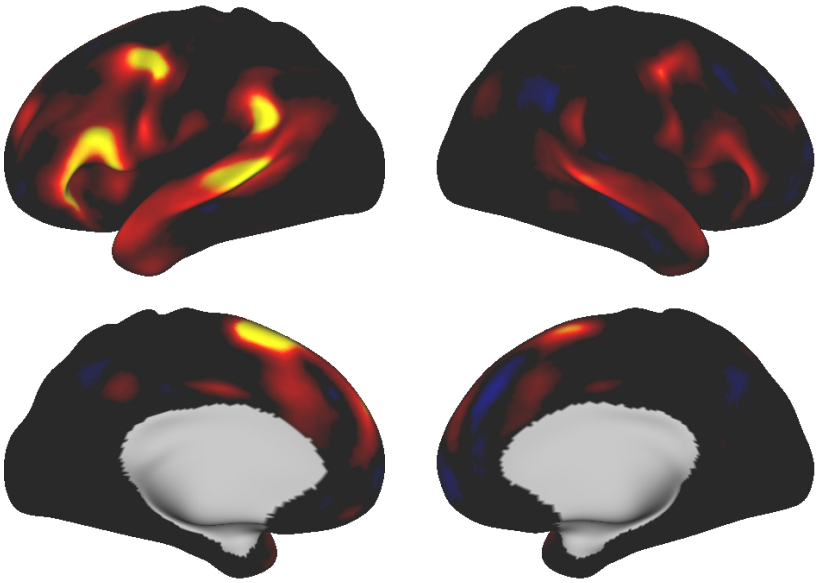


(d) sICA, all data.

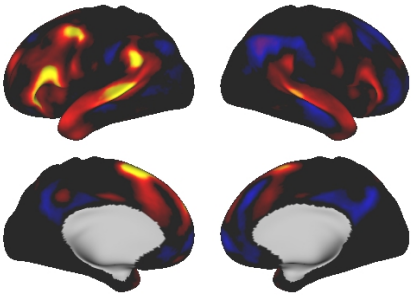


(e) tICA, all data.

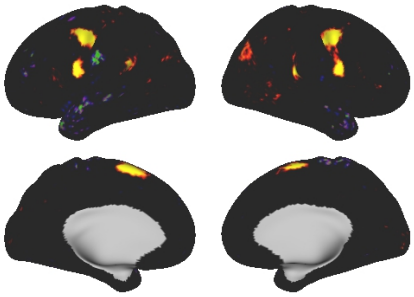
Figure F.19: PFM 19



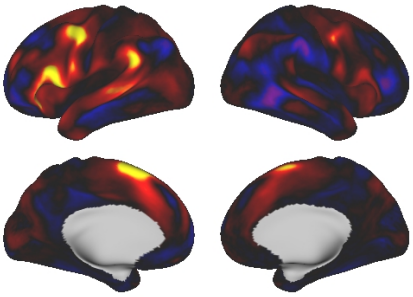
(a) PFM, all data.



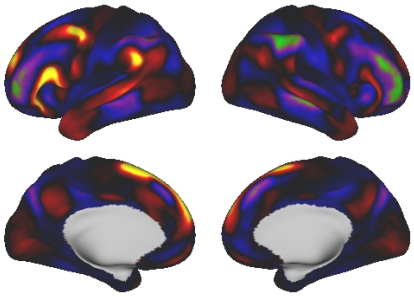
(b) PFM, first half of data.



(c) PFM, second half of data.

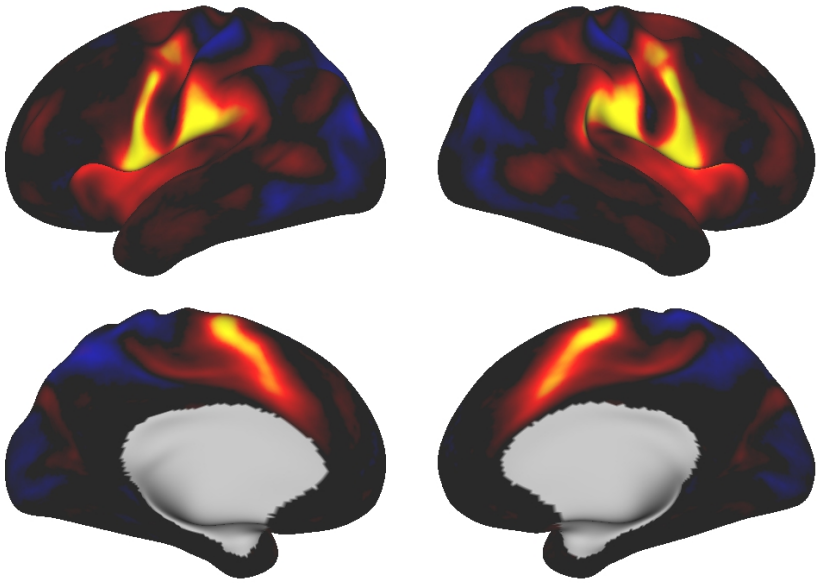


(d) sICA, all data.

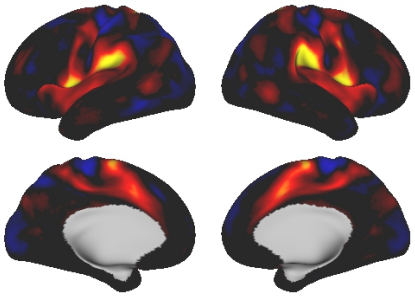


(e) tICA, all data.

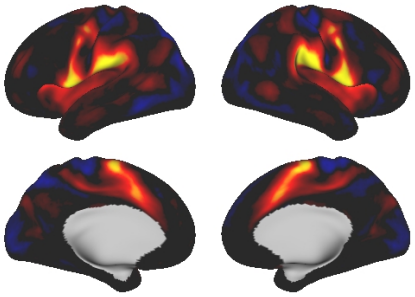
Figure F.20: PFM 20



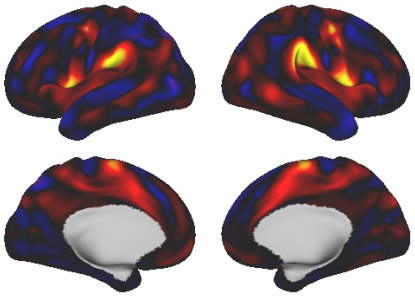
(a) PFM, all data.



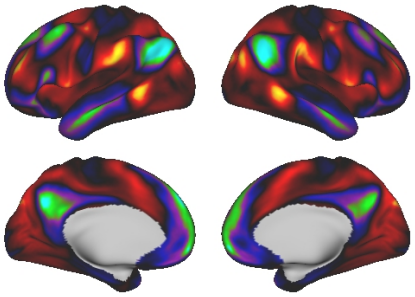
(b) PFM, first half of data.



(c) PFM, second half of data.

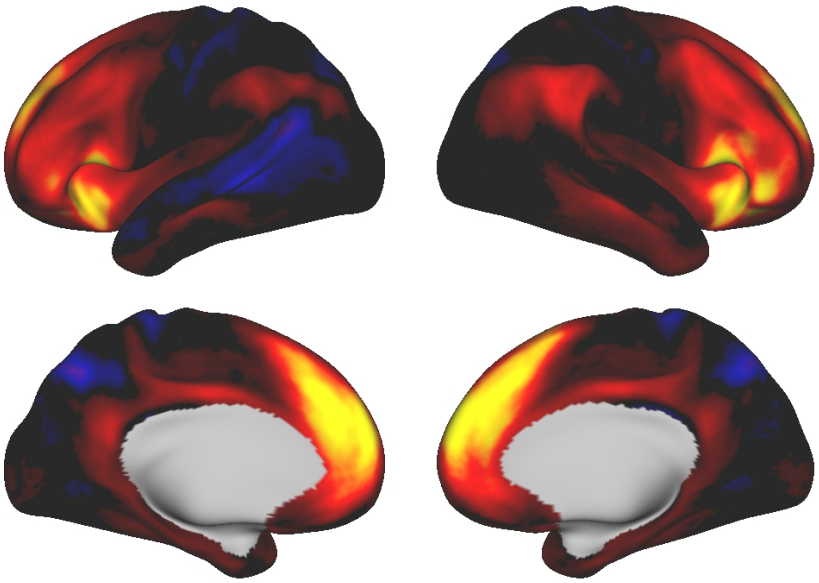


(d) sICA, all data.

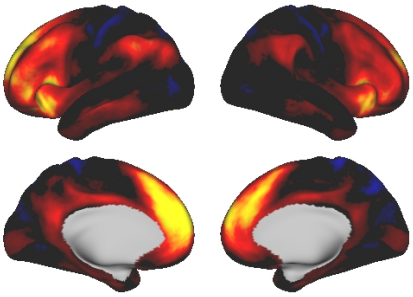


(e) tICA, all data.

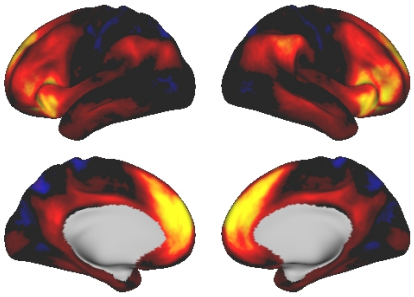
Figure F.21: PFM 21



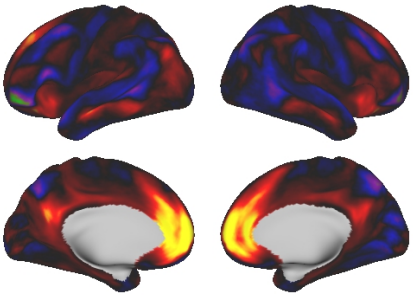
(a) PFM, all data.



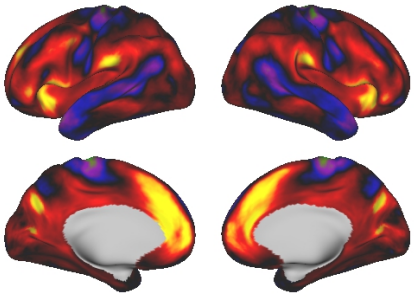
(b) PFM, first half of data.



(c) PFM, second half of data.

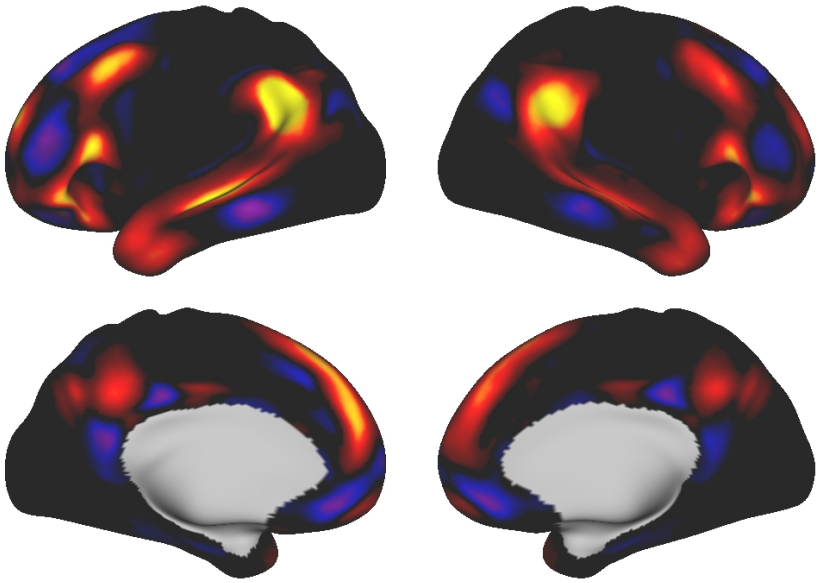


(d) sICA, all data.

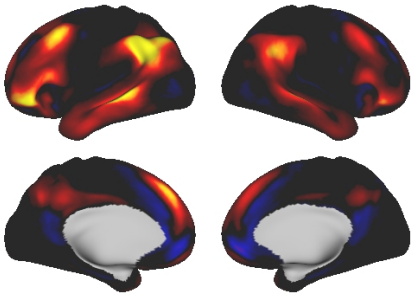


(e) tICA, all data.

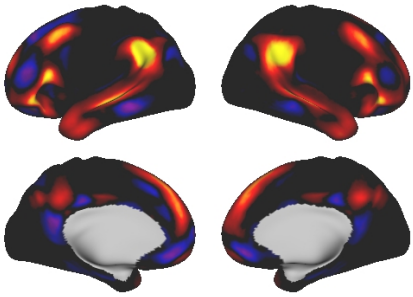
Figure F.22: PFM 22



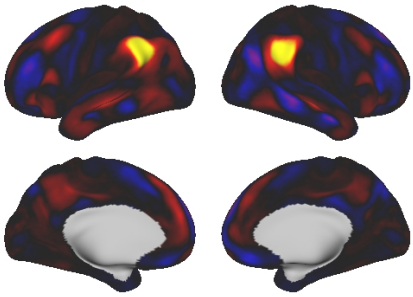
(a) PFM, all data.



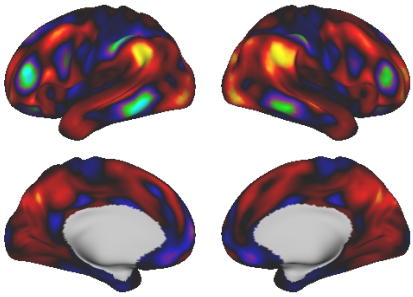
(b) PFM, first half of data.



(c) PFM, second half of data.

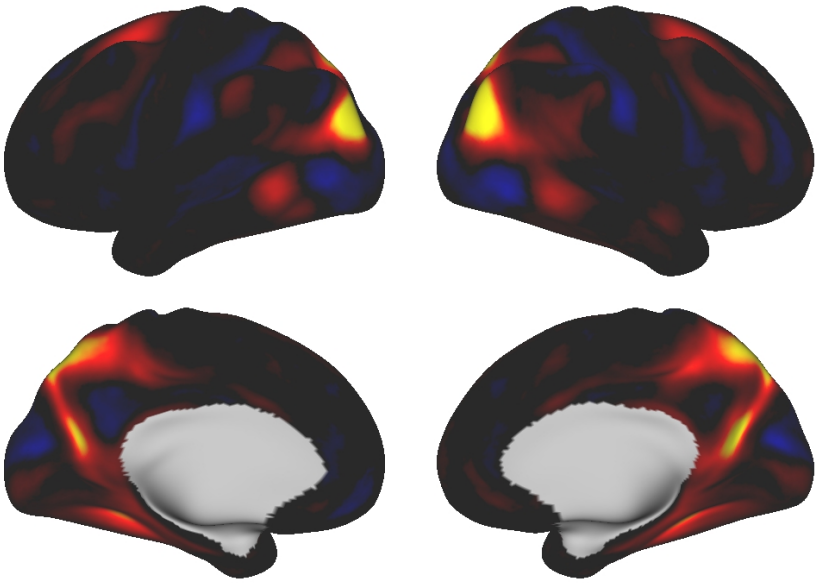


(d) sICA, all data.

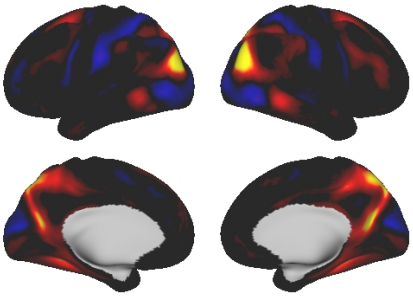


(e) tICA, all data.

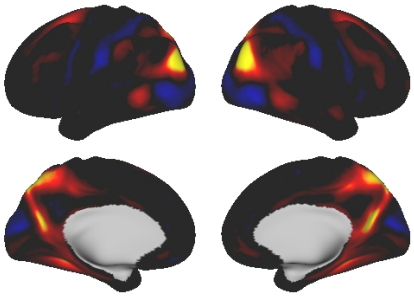
Figure F.23: PFM 23



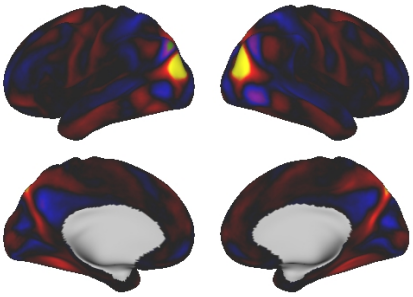
(a) PFM, all data.



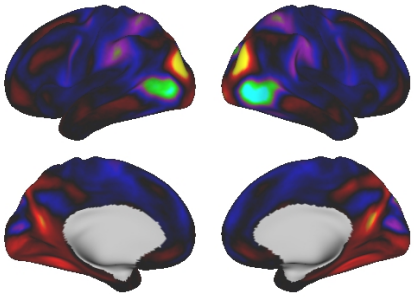
(b) PFM, first half of data.



(c) PFM, second half of data.

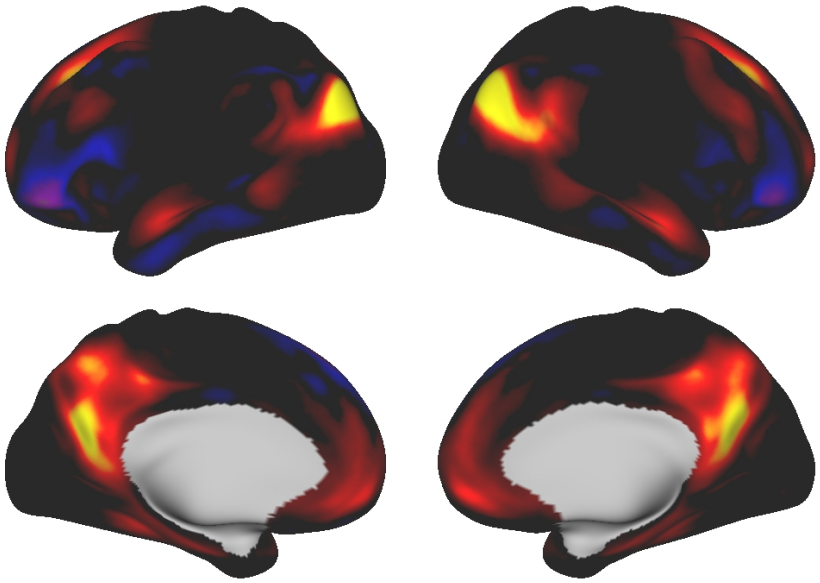


(d) sICA, all data.

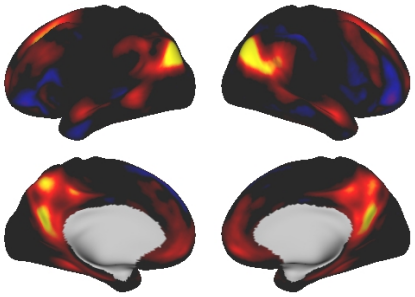


(e) tICA, all data.

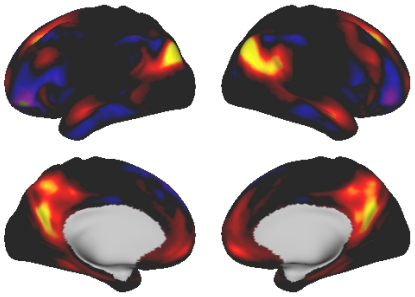
Figure F.24: PFM 24



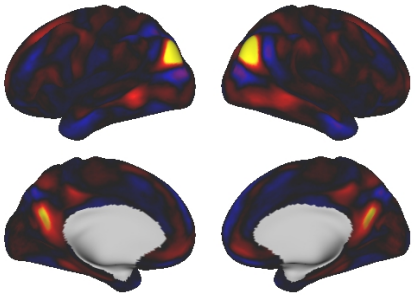
(a) PFM, all data.



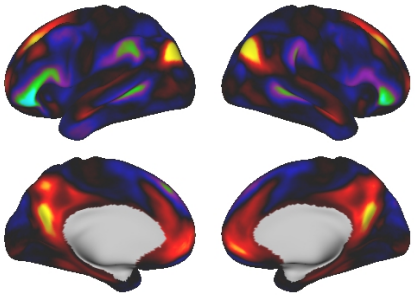
(b) PFM, first half of data.



(c) PFM, second half of data.

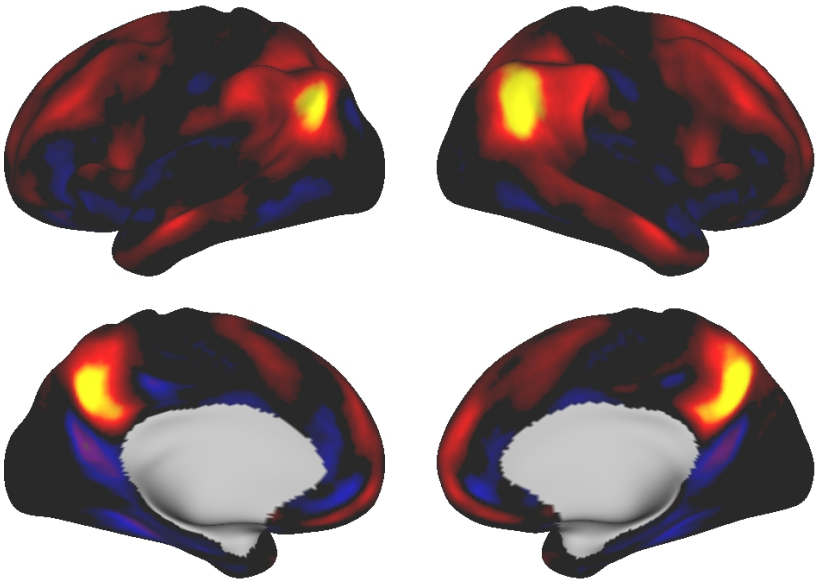


(d) sICA, all data.

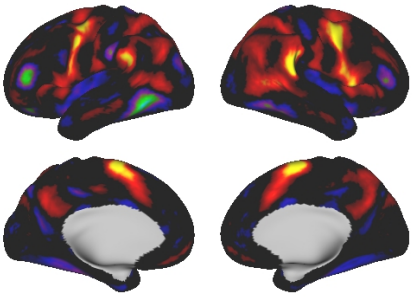


(e) tICA, all data.

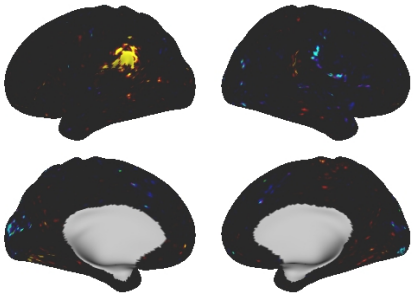
Figure F.25: PFM 25



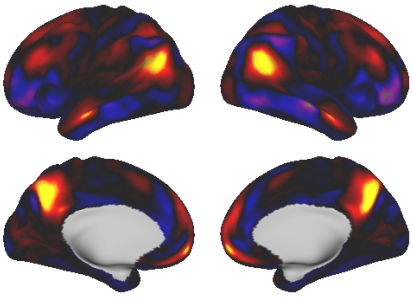
(a) PFM, all data.



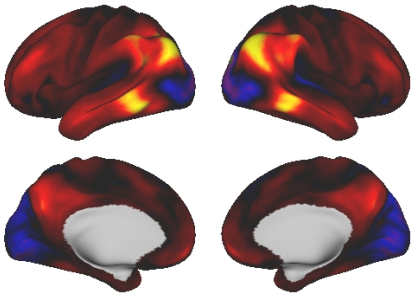
(b) PFM, first half of data.



(c) PFM, second half of data.

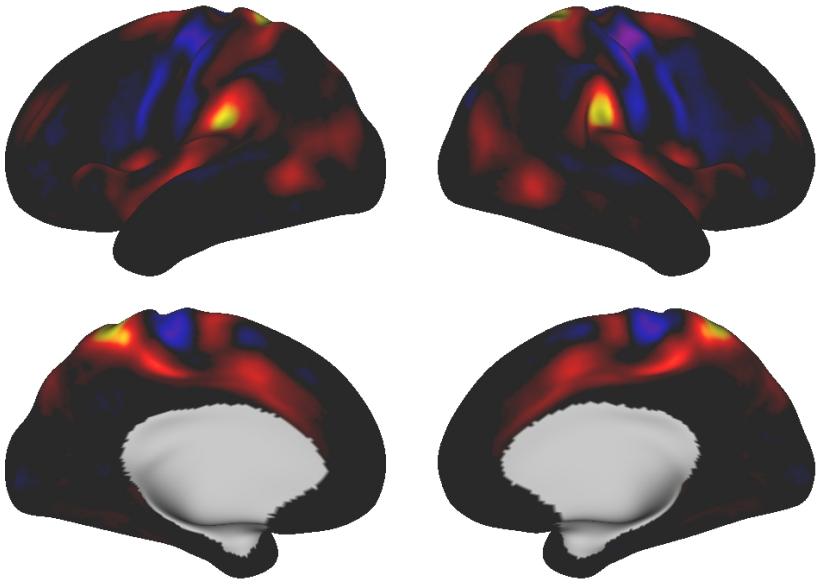


(d) sICA, all data.

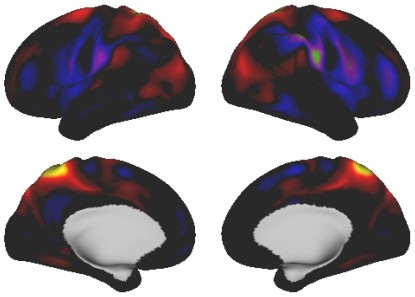


(e) tICA, all data.

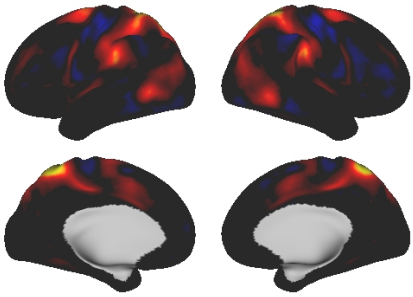
Figure F.26: PFM 26



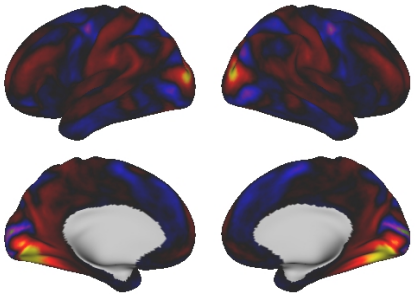
(a) PFM, all data.



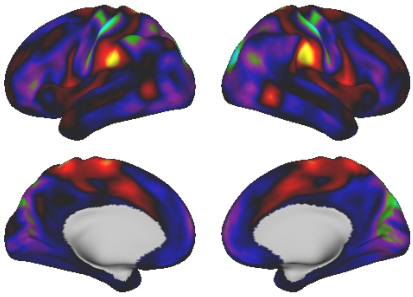
(b) PFM, first half of data.



(c) PFM, second half of data.

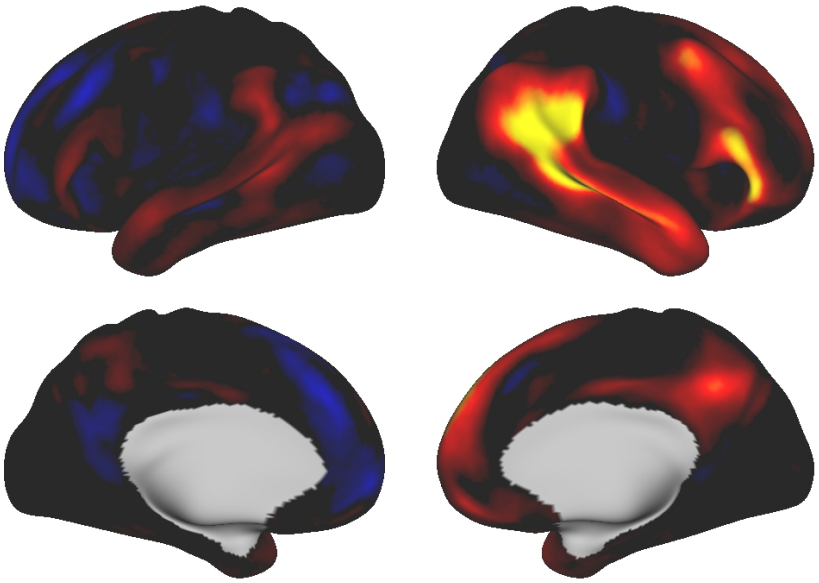


(d) sICA, all data.

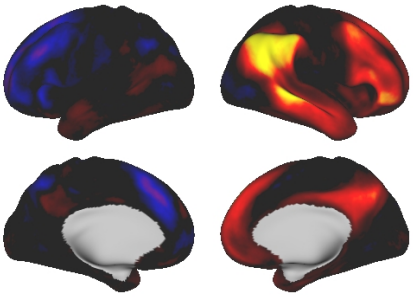


(e) tICA, all data.

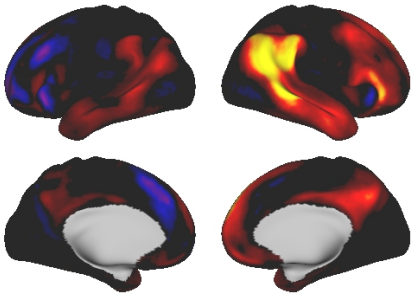
Figure F.27: PFM 27



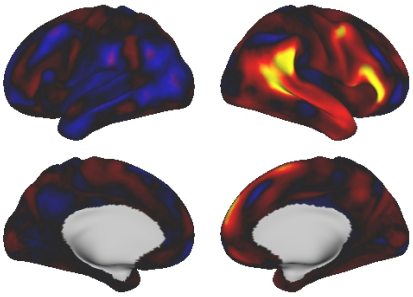
(a) PFM, all data.



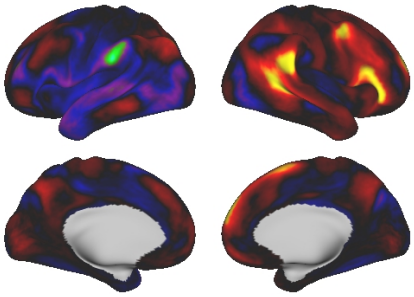
(b) PFM, first half of data.



(c) PFM, second half of data.

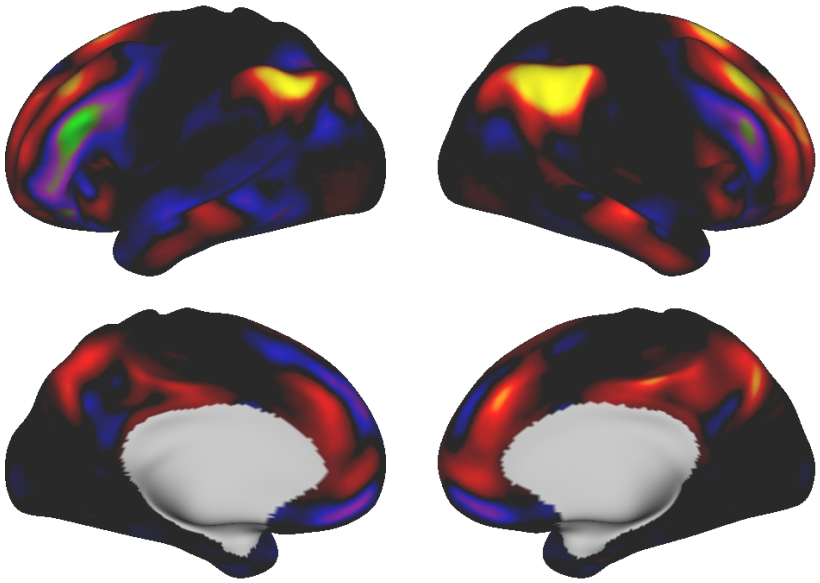


(d) sICA, all data.

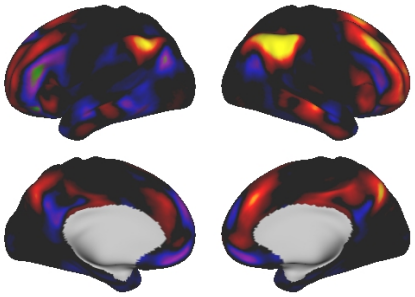


(e) tICA, all data.

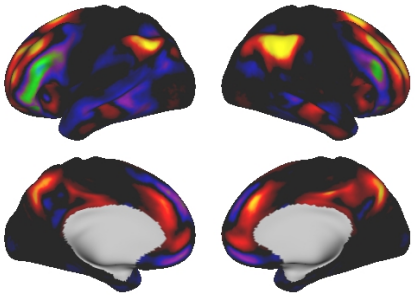
Figure F.28: PFM 28



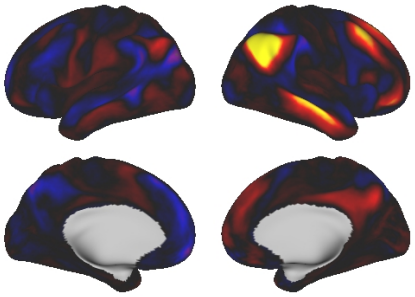
(a) PFM, all data.



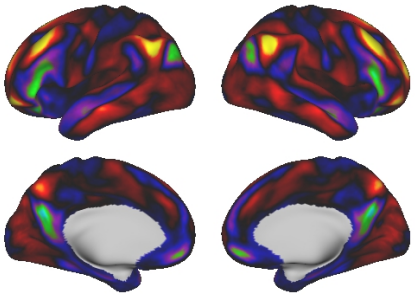
(b) PFM, first half of data.



(c) PFM, second half of data.

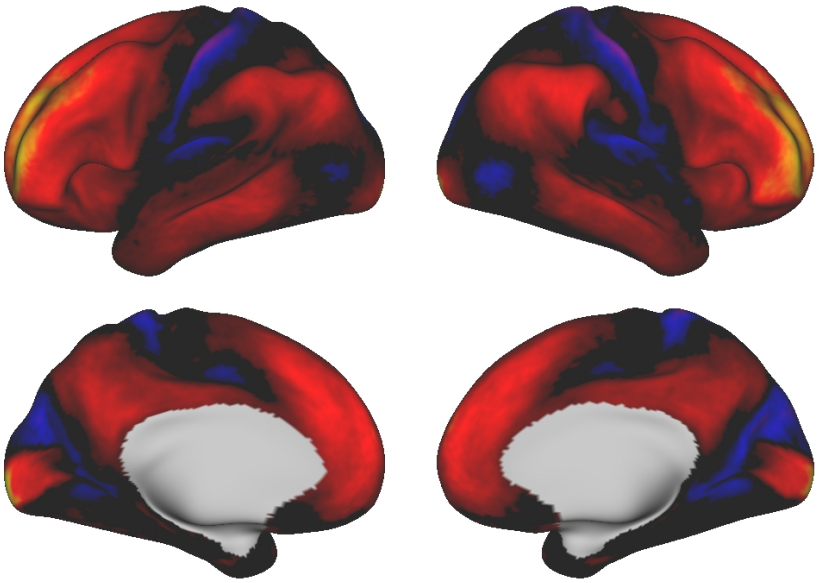


(d) sICA, all data.

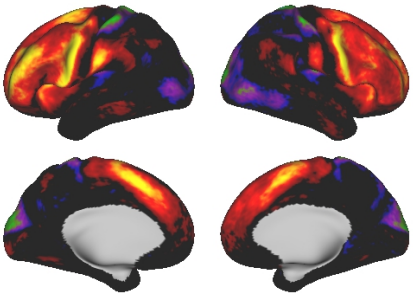


(e) tICA, all data.

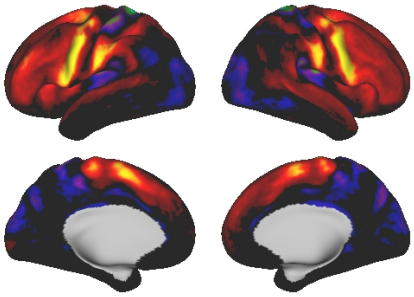
Figure F.29: PFM 29



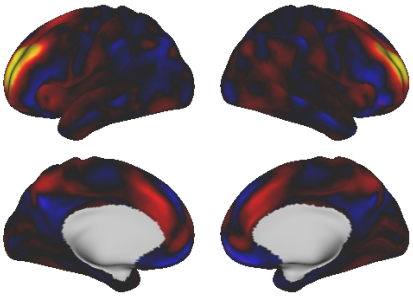
(a) PFM, all data.



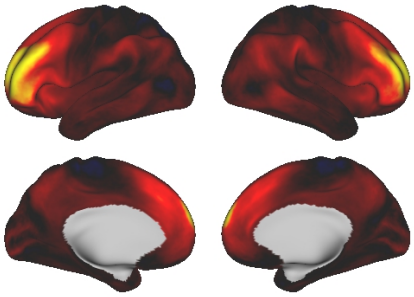
(b) PFM, first half of data.



(c) PFM, second half of data.

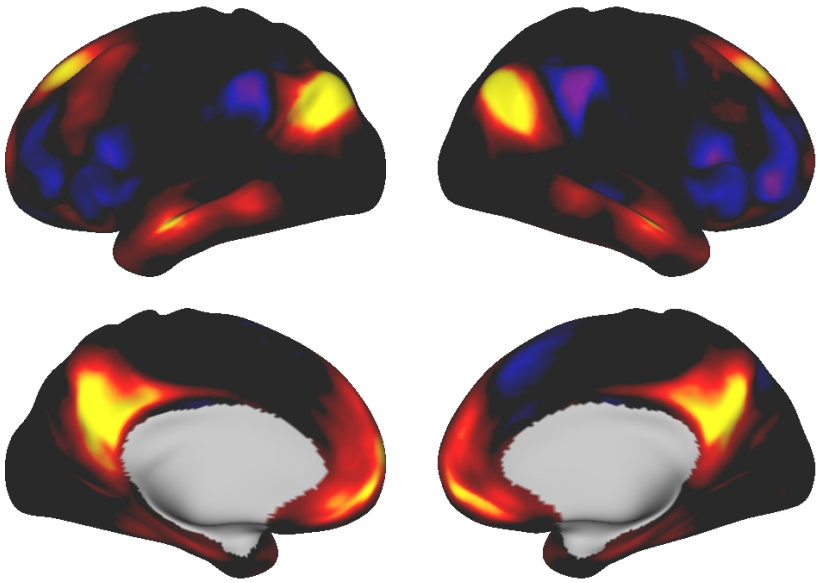


(d) sICA, all data.

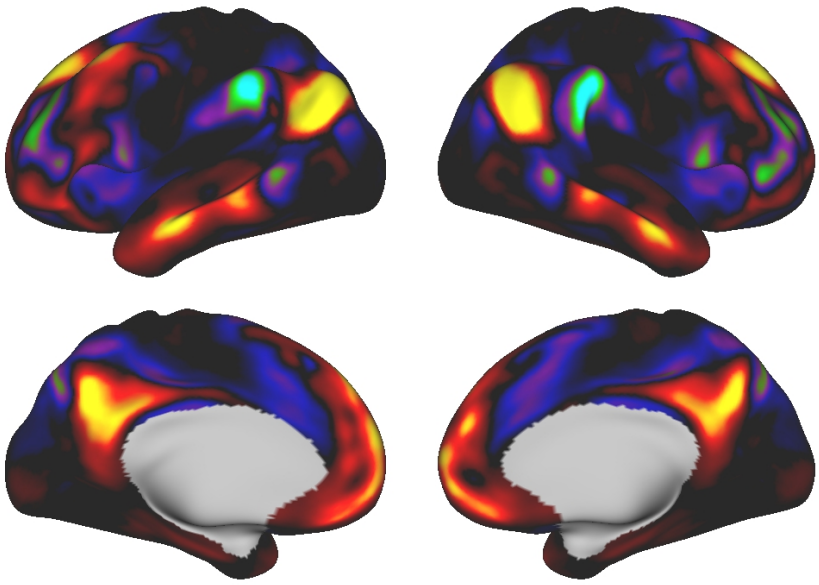


(e) tICA, all data.

Figure F.30: PFM 30

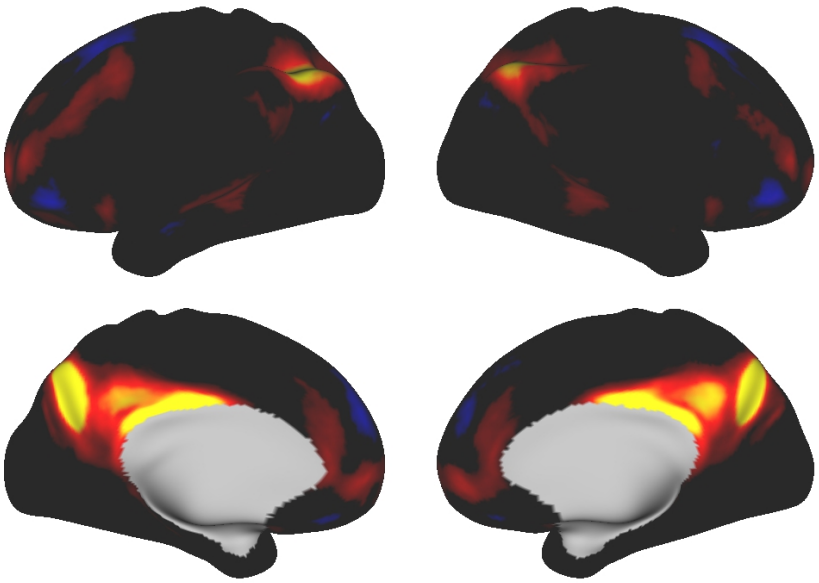


(a) PFM S1

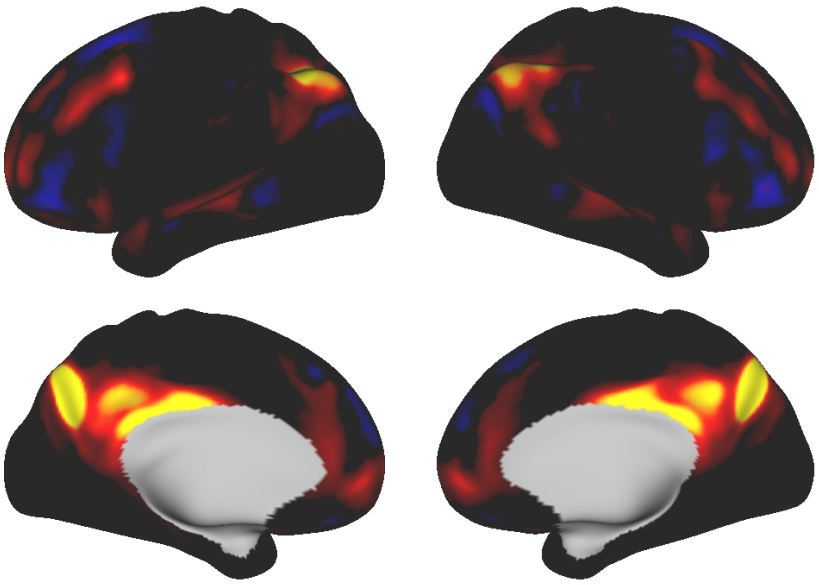


(b) PFM F1

Figure F.31: PFMs inferred under two different registration schemes. PFM S1 was inferred from data where the registration was driven by structural features, whereas PFM F1 was inferred from data that was registered using structural and functional features.

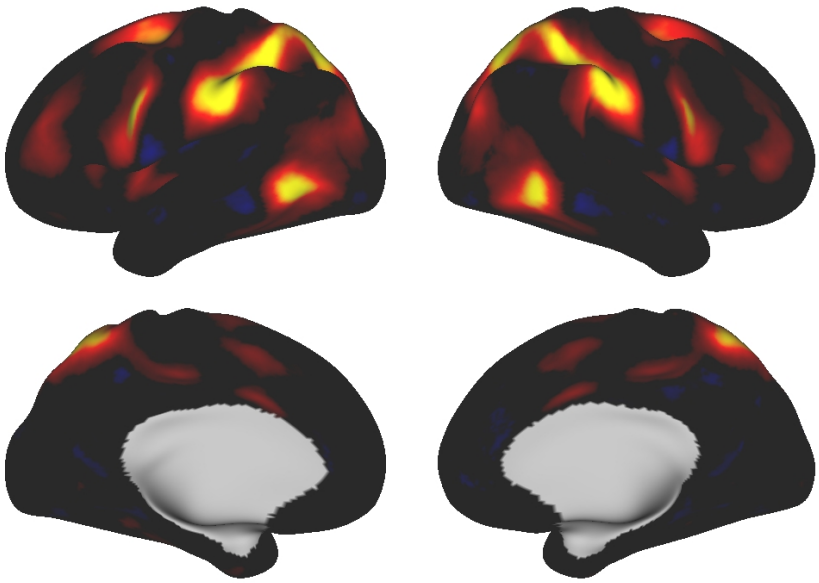


(a) PFM S2

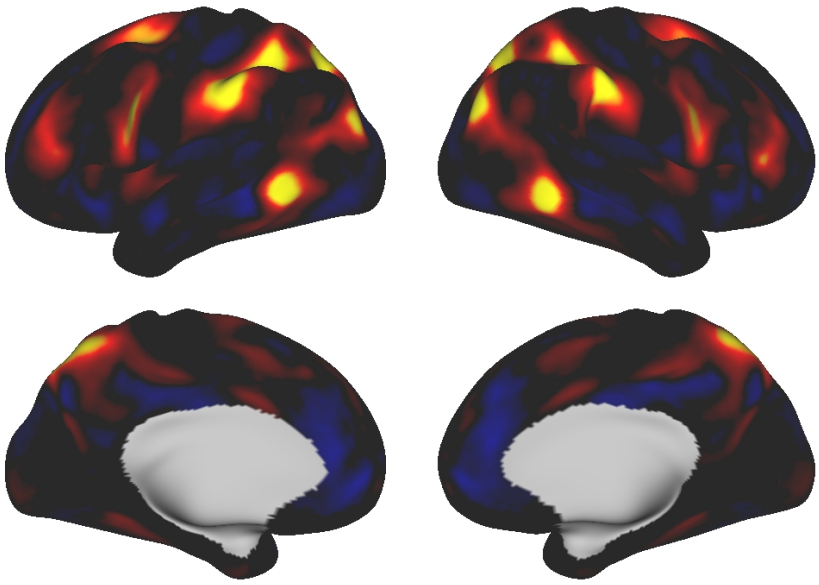


(b) PFM F2

Figure F.32: PFMs inferred under two different registration schemes. PFM S2 was inferred from data where the registration was driven by structural features, whereas PFM F2 was inferred from data that was registered using structural and functional features.

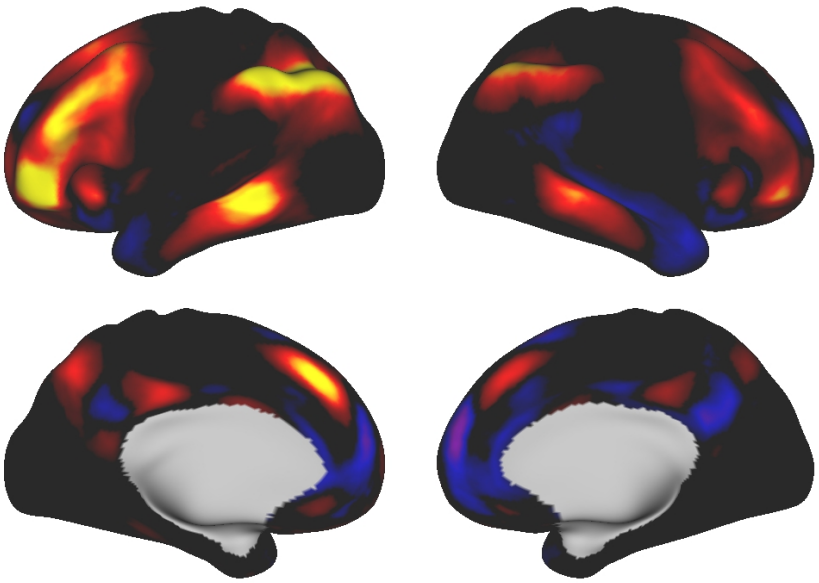


(a) PFM S3

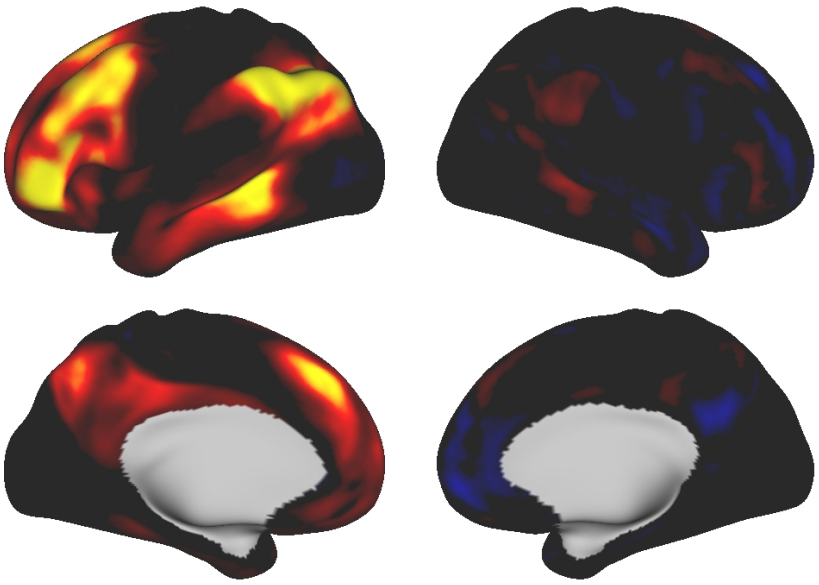


(b) PFM F3

Figure F.33: PFMs inferred under two different registration schemes. PFM S3 was inferred from data where the registration was driven by structural features, whereas PFM F3 was inferred from data that was registered using structural and functional features.

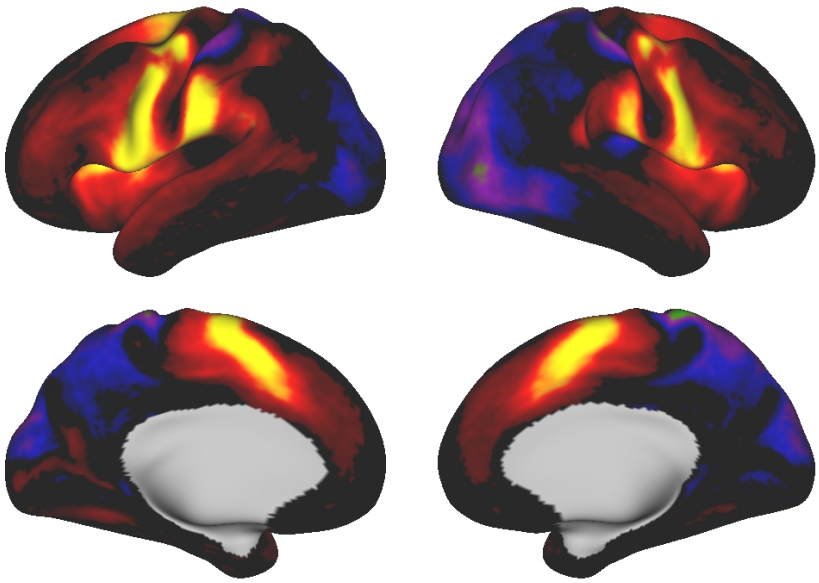


(a) PFM S4

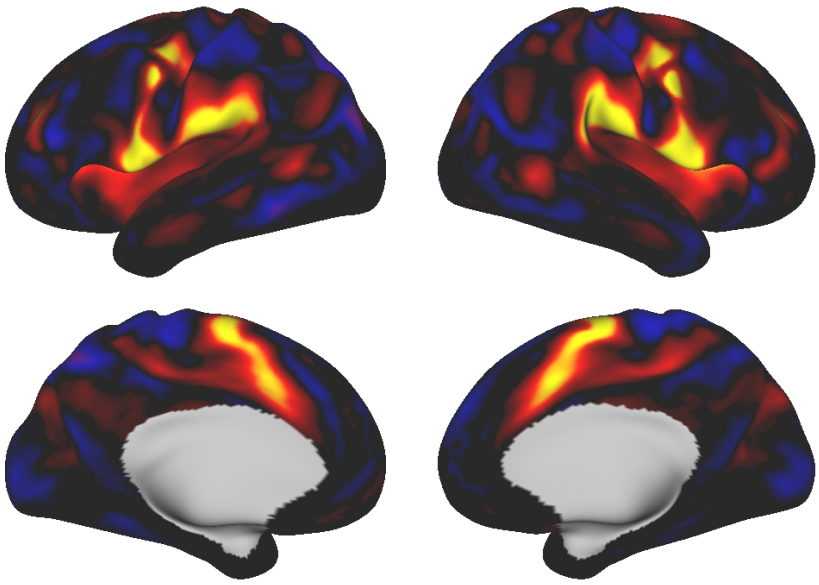


(b) PFM F4

Figure F.34: PFMs inferred under two different registration schemes. PFM S4 was inferred from data where the registration was driven by structural features, whereas PFM F4 was inferred from data that was registered using structural and functional features.



(a) PFM S5



(b) PFM F5

Figure F.35: PFMs inferred under two different registration schemes. PFM S5 was inferred from data where the registration was driven by structural features, whereas PFM F5 was inferred from data that was registered using structural and functional features.

Bibliography

Abraham, A., Dohmatob, E., Thirion, B., Samaras, D. and Varoquaux, G. (2013). *Extracting Brain Regions from Rest fMRI with Total-Variation Constrained Dictionary Learning*. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*. Ed. by K. Mori, I. Sakuma, Y. Sato, C. Barillot and N. Navab. Vol. 8150. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 607–615.

Aguirre, G., Zarahn, E. and D'Esposito, M. (1998). *The Variability of Human, BOLD Hemodynamic Responses*. In: *NeuroImage* 8.4, pp. 360–369.

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T. et al. (2014). *Tracking Whole-Brain Connectivity Dynamics in the Resting State*. In: *Cerebral Cortex* 24.3, pp. 663–676.

Allen, G. I., Grosenick, L. and Taylor, J. (2014). *A Generalized Least-Square Matrix Decomposition*. In: *Journal of the American Statistical Association* 109.505, pp. 145–159.

Amedi, A., Raz, N., Pianka, P., Malach, R. and Zohary, E. (2003). *Early 'visual' cortex activation correlates with superior verbal memory performance in the blind*. In: *Nature Neuroscience* 6.7 (July 2003), pp. 758–766.

Amunts, K. and Zilles, K. (2012). *Architecture and organizational principles of Broca's region*. In: *Trends in Cognitive Sciences* 16.8 (Aug. 2012), pp. 418–426.

Anderson, J. S., Ferguson, M. A., Lopez-Larson, M. and Yurgelun-Todd, D. (2011). *Reproducibility of Single-Subject Functional Connectivity Measurements*. In: *American Journal of Neuroradiology* 32.3, pp. 548–555.

Andrews, T. J., Halpern, S. D. and Purves, D. (1997). *Correlated Size Variations in Human Visual Cortex, Lateral Geniculate Nucleus, and Optic Tract*. In: *The Journal of Neuroscience* 17.8 (Apr. 1997), pp. 2859–2868.

Attias, H. (2000). *A Variational Bayesian Framework for Graphical Models*. In: *Advances in Neural Information Processing Systems 12*. Ed. by S. A. Solla, T. K. Leen and K. Müller. MIT Press, pp. 209–215.

Aydin, Ü., Vorwerk, J., Dümpelmann, M., Küpper, P., Kugel, H. et al. (2015). *Combined EEG/MEG Can Outperform Single Modality EEG or MEG Source Reconstruction in Presurgical Epilepsy Diagnosis*. In: *PLoS ONE* 10.3 (Mar. 2015), e0118753.

Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T. E. J. et al. (2014). *Fast transient networks in spontaneous human brain activity*. In: *eLife* 3, e01867.

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S. and Hyde, J. S. (1992). *Time course EPI of human brain function during task activation*. In: *Magnetic Resonance in Medicine* 25.2, pp. 390–397.

Barron, H. C., Dolan, R. J. and Behrens, T. E. J. (2013). *Online evaluation of novel choices by simultaneous representation of multiple memories*. In: *Nature Neuroscience* 16.10 (Oct. 2013), pp. 1492–1498.

Bartley, A. J., Jones, D. W. and Weinberger, D. R. (1997). *Genetic variability of human brain size and cortical gyral patterns*. In: *Brain* 120.2, pp. 257–269.

Bayes, T. and Price, R. (1763). *An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a Letter to John Canton, A.M.F.R.S.* In: *Philosophical Transactions (1683-1775)* 53, pp. 370–418.

Beckmann, C. F. (2012). *Modelling with independent components*. In: *NeuroImage* 62.2 (Aug. 2012), pp. 891–901.

Beckmann, C. F., DeLuca, M., Devlin, J. T. and Smith, S. M. (2005). *Investigations into resting-state connectivity using independent component analysis*. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 360.1457, pp. 1001–1013.

Beckmann, C. F., Mackay, C., Filippini, N. and Smith, S. M. (2009). *Group comparison of resting-state FMRI data using multi-subject ICA and dual regression*. In: *Organization for Human Brain Mapping*.

Beckmann, C. F. and Smith, S. M. (2002). *Probabilistic Independent Component Analysis for FMRI*. Tech. rep. TR02CB1. FMRIB.

Beckmann, C. F. and Smith, S. M. (2004). *Probabilistic independent component analysis for functional magnetic resonance imaging*. In: *Medical Imaging, IEEE Transactions on* 23.2 (Feb. 2004), pp. 137–152.

Bell, A. J. and Sejnowski, T. J. (1995). *An Information-Maximization Approach to Blind Separation and Blind Deconvolution*. In: *Neural Computation* 7.6 (Nov. 1995), pp. 1129–1159.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M. et al. (1999). *Conceptual Processing during the Conscious Resting State: A Functional MRI Study*. In: *Journal of Cognitive Neuroscience* 11.1 (Jan. 1999), pp. 80–93.

Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B. et al. (2013). *The effect of scan length on the reliability of resting-state fMRI connectivity estimates*. In: *NeuroImage* 83 (Dec. 2013), pp. 550–558.

Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C. et al. (2010). *Toward discovery science of human brain function*. In: *Proceedings of the National Academy of Sciences* 107.10 (Mar. 2010), pp. 4734–4739.

Biswal, B. B., Yetkin, F. Z., Haughton, V. M. and Hyde, J. S. (1995). *Functional connectivity in the motor cortex of resting human brain using echo-planar MRI*. In: *Magnetic Resonance in Medicine* 34.4, pp. 537–541.

Blaimer, M., Breuer, F., Mueller, M., Heidemann, R. M., Griswold, M. A. et al. (2004). *SMASH, SENSE, PILS, GRAPPA: How to Choose the Optimal Method*. In: *Topics in Magnetic Resonance Imaging* 15.4 (Aug. 2004), pp. 223–236.

Blumensath, T., Jbabdi, S., Glasser, M. F., Van Essen, D. C., Uğurbil, K. et al. (2013). *Spatially constrained hierarchical parcellation of the brain with resting-state fMRI*. In: *NeuroImage* 76 (Aug. 2013), pp. 313–324.

Bright, M. G. and Murphy, K. (2015). *Is fMRI "noise" really noise? Resting state nuisance regressors remove variance with network structure*. In: *NeuroImage* 114 (July 2015), pp. 158–169.

Broca, P. P. (1861). *Remarques sur la siège de la faculté de la langage articulé, suivie d'une observation d'aphémie (perte de la parole)*. In: *Bulletins de la Société Anatomique* 6, pp. 330–357.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde*. Leipzig: Barth.

Brookes, M. J., Woolrich, M. W., Luckhoo, H., Price, D., Hale, J. R. et al. (2011). *Investigating the electrophysiological basis of resting state networks using magnetoencephalography*. In: *Proceedings of the National Academy of Sciences* 108.40, pp. 16783–16788.

Buckner, R. L., Andrews-Hanna, J. R. and Schacter, D. L. (2008). *The Brain's Default Network: Anatomy, Function, and Relevance to Disease*. In: *Annals of the New York Academy of Sciences* 1124.1, pp. 1–38.

Buckner, R. L. and Krienen, F. M. (2013). *The evolution of distributed association networks in the human brain*. In: *Trends in Cognitive Sciences* 17.12 (Dec. 2013), pp. 648–665.

Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C. and Yeo, B. T. T. (2011). *The organization of the human cerebellum estimated by intrinsic functional connectivity*. In: *Journal of Neurophysiology* 106.5 (Nov. 2011), pp. 2322–2345.

Buckner, R. L. and Yeo, B. T. T. (2014). *Borders, map clusters, and supra-areal organization in visual cortex*. In: *NeuroImage* 93, Part 2 (June 2014), pp. 292–297.

Bullmore, E. and Sporns, O. (2009). *Complex brain networks: graph theoretical analysis of structural and functional systems*. In: *Nature Reviews Neuroscience* 10.3 (Mar. 2009), pp. 186–198.

Buxton, R. B., Wong, E. C. and Frank, L. R. (1998). *Dynamics of blood flow and oxygenation changes during brain activation: The balloon model*. In: *Magnetic Resonance in Medicine* 39.6 (June 1998), pp. 855–864.

Calhoun, V. D., Adalı, T., Hansen, L. K., Larsen, J. and Pekar, J. J. (2003). *ICA of Functional MRI Data: An Overview*. In: *4th International Workshop on Independent Component Analysis and Blind Signal Separation*, pp. 281–288.

Calhoun, V. D., Adalı, T., Pearlson, G. D. and Pekar, J. J. (2001). *A method for making group inferences from functional MRI data using independent component analysis*. In: *Human Brain Mapping* 14.3 (Nov. 2001), pp. 140–151.

Calhoun, V. D., Pekar, J. J., McGinty, V. B., Adalı, T., Watson, T. D. et al. (2002). *Different activation dynamics in multiple neural systems during simulated driving*. In: *Human Brain Mapping* 16.3 (July 2002), pp. 158–167.

Callard, F. and Margulies, D. S. (2011). *The Subject at Rest: Novel conceptualizations of self and brain from cognitive neuroscience's study of the 'resting state'*. In: *Subjectivity* 4.3 (Sept. 2011), pp. 227–257.

Canolty, R. T. and Knight, R. T. (2010). *The functional role of cross-frequency coupling*. In: *Trends in Cognitive Sciences* 14.11, pp. 506–515.

Castellanos, F. X., Di Martino, A., Craddock, R. C., Mehta, A. D. and Milham, M. P. (2013). *Clinical applications of the functional connectome*. In: *NeuroImage* 80 (Oct. 2013), pp. 527–540.

Chang, C. and Glover, G. H. (2010). *Time-frequency dynamics of resting-state brain connectivity measured with fMRI*. In: *NeuroImage* 50.1, pp. 81–98.

Chappell, M. A., Groves, A. and Woolrich, M. W. (2008). *The FMRIB Variational Bayes Tutorial: Variational Bayesian inference for a non-linear forward model*. Tech. rep. FMRIB.

Cohen, A. L., Fair, D. A., Dosenbach, N. U. F., Miezin, F. M., Dierker, D. et al. (2008). *Defining functional areas in individual human brains using resting functional connectivity MRI*. In: *NeuroImage* 41.1, pp. 45–57.

Cole, D. M., Smith, S. M. and Beckmann, C. F. (2010). *Advances and pitfalls in the analysis and interpretation of resting-state FMRI data*. In: *Frontiers in Systems Neuroscience* 4.8 (Apr. 2010).

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A. et al. (2013). *Multi-task connectivity reveals flexible hubs for adaptive task control*. In: *Nature Neuroscience* 16.9 (Sept. 2013), pp. 1348–1355.

Comon, P. (1994). *Independent component analysis, A new concept?* In: *Signal Processing* 36.3 (Apr. 1994), pp. 287–314.

Comon, P., Jutten, C. and Herault, J. (1991). *Blind separation of sources, part II: Problems statement*. In: *Signal Processing* 24.1 (July 1991), pp. 11–20.

Conroy, B. R., Singer, B. D., Guntupalli, J. S., Ramadge, P. J. and Haxby, J. V. (2013). *Inter-subject alignment of human cortical anatomy using functional connectivity*. In: *NeuroImage* 81 (Nov. 2013), pp. 400–411.

Corbetta, M. and Shulman, G. L. (2002). *Control of goal-directed and stimulus-driven attention in the brain*. In: *Nature Reviews Neuroscience* 3.3 (Mar. 2002), pp. 201–215.

Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A. et al. (2001). *Frequencies Contributing to Functional Connectivity in the Cerebral Cortex in "Resting-state" Data*. In: *American Journal of Neuroradiology* 22.7, pp. 1326–1333.

Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A. et al. (2000). *Mapping Functionally Related Regions of Brain with Functional Connectivity MR Imaging*. In: *American Journal of Neuroradiology* 21.9, pp. 1636–1644.

Correa, N., Adalı, T. and Calhoun, V. D. (2007). *Performance of blind source separation algorithms for fMRI analysis using a group ICA method*. In: *Magnetic Resonance Imaging* 25.5, pp. 684–694.

Craddock, R. C., James, G., Holtzheimer, P. E., Hu, X. P. and Mayberg, H. S. (2012). *A whole brain fMRI atlas generated via spatially constrained spectral clustering*. In: *Human Brain Mapping* 33.8, pp. 1914–1928.

Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D. and Lindquist, M. A. (2012). *Dynamic connectivity regression: Determining state-related changes in brain connectivity*. In: *NeuroImage* 61.4, pp. 907–920.

Critchley, M. and Critchley, E. A. (1998). *John Hughlings Jackson—Father Of English Neurology*. Oxford University Press.

Damadian, R., Goldsmith, M. and Minkoff, L. (1977). *NMR in cancer: XVI. FONAR image of the live human body.* In: *Physiological Chemistry and Physics* 9.1, pp. 97–100, 108.

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J. et al. (2006). *Consistent resting-state networks across healthy subjects*. In: *Proceedings of the National Academy of Sciences* 103.37 (Sept. 2006), pp. 13848–13853.

Dawid, A. P. (1981). *Some matrix-variate distribution theory: Notational considerations and a Bayesian application*. In: *Biometrika* 68.1, pp. 265–274.

De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M. and Smith, S. M. (2006). *fMRI resting state networks define distinct modes of long-distance interactions in the human brain*. In: *NeuroImage* 29.4 (Feb. 2006), pp. 1359–1367.

Deco, G., Jirsa, V. K. and McIntosh, A. R. (2011). *Emerging concepts for the dynamical organization of resting-state activity in the brain*. In: *Nature Reviews Neuroscience* 12.1 (Jan. 2011), pp. 43–56.

Deco, G., Jirsa, V., McIntosh, A. R., Sporns, O. and Kötter, R. (2009). *Key role of coupling, delay, and noise in resting brain fluctuations*. In: *Proceedings of the National Academy of Sciences* 106.25, pp. 10302–10307.

Deco, G. and Kringelbach, M. L. (2014). *Great Expectations: Using Whole-Brain Computational Connectomics for Understanding Neuropsychiatric Disorders*. In: *Neuron* 84.5 (Dec. 2014), pp. 892–905.

Dhillon, P. S., Wolk, D. A., Das, S. R., Ungar, L. H., Gee, J. C. et al. (2014). *Subject-specific functional parcellation via Prior Based Eigenanatomy*. In: *NeuroImage* 99 (Oct. 2014), pp. 14–27.

Doeller, C. F., Barry, C. and Burgess, N. (2010). *Evidence for grid cells in a human memory network*. In: *Nature* 463.7281 (Feb. 2010), pp. 657–661.

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K. et al. (2007). *Distinct brain networks for adaptive and stable task control in humans*. In: *Proceedings of the National Academy of Sciences* 104.26 (June 2007), pp. 11073–11078.

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K. et al. (2006). *A Core System for the Implementation of Task Sets*. In: *Neuron* 50.5 (June 2006), pp. 799–812.

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J. et al. (2003). *Visual field representations and locations of visual areas V1/2/3 in human visual cortex*. In: *Journal of Vision* 3.10 (Oct. 2003), pp. 586–598.

Du, Y. and Fan, Y. (2013). *Group information guided ICA for fMRI data analysis*. In: *NeuroImage* 69 (Apr. 2013), pp. 157–197.

Eavani, H., Satterthwaite, T. D., Gur, R. E., Gur, R. C. and Davatzikos, C. (2013). *Unsupervised Learning of Functional Network Dynamics in Resting State fMRI*. In: *Information Processing in Medical Imaging*. Springer, pp. 426–437.

Eickhoff, S. B., Schleicher, A., Zilles, K. and Amunts, K. (2006). *The Human Parietal Operculum. I. Cytoarchitectonic Mapping of Subdivisions*. In: *Cerebral Cortex* 16.2 (Feb. 2006), pp. 254–267.

Eickhoff, S. B., Walters, N. B., Schleicher, A., Kril, J., Egan, G. F. et al. (2005). *High-resolution MRI reflects myeloarchitecture and cytoarchitecture of human cerebral cortex*. In: *Human Brain Mapping* 24.3 (Mar. 2005), pp. 206–215.

Eling, P., ed. (1994). *Reader in the history of aphasia: From Franz Gall to Norman Geschwind*. Vol. 4. John Benjamins Publishing.

Erhardt, E. B., Allen, E. A., Wei, Y., Eichele, T. and Calhoun, V. D. (2012). *SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability*. In: *NeuroImage* 59.4, pp. 4160–4167.

Erhardt, E. B., Rachakonda, S., Bedrick, E. J., Allen, E. A., Adalı, T. et al. (2011). *Comparison of multi-subject ICA methods for analysis of fMRI data*. In: *Human Brain Mapping* 32.12, pp. 2075–2095.

Esposito, F., Scarabino, T., Hyvärinen, A., Himberg, J., Formisano, E. et al. (2005). *Independent component analysis of fMRI group studies by self-organizing clustering*. In: *NeuroImage* 25.1 (Mar. 2005), pp. 193–205.

Felleman, D. J. and Van Essen, D. C. (1991). *Distributed Hierarchical Processing in the Primate Cerebral Cortex*. In: *Cerebral Cortex* 1.1, pp. 1–47.

Filippini, N., MacIntosh, B. J., Hough, M. G., Goodwin, G. M., Frisoni, G. B. et al. (2009). *Distinct patterns of brain activity in young carriers of the APOE-ϵ4 allele*. In: *Proceedings of the National Academy of Sciences* 106.17, pp. 7209–7214.

Fornito, A., Zalesky, A. and Breakspear, M. (2013). *Graph analysis of the human connectome: Promise, progress, and pitfalls*. In: *NeuroImage* 80 (Oct. 2013), pp. 426–444.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C. et al. (2005). *The human brain is intrinsically organized into dynamic, anticorrelated functional networks*. In: *Proceedings of the National Academy of Sciences* 102.27, pp. 9673–9678.

Fox, M. D., Zhang, D., Snyder, A. Z. and Raichle, M. E. (2009). *The Global Signal and Observed Anticorrelated Resting State Brain Networks*. In: *Journal of Neurophysiology* 101.6, pp. 3270–3283.

Fransson, P. (2005). *Spontaneous low-frequency BOLD signal fluctuations: An fMRI investigation of the resting-state default mode of brain function hypothesis*. In: *Human Brain Mapping* 26.1, pp. 15–29.

Friston, K. J. (1998). *Modes or models: a critique on independent component analysis for fMRI*. In: *Trends in Cognitive Sciences* 2.10 (Nov. 1998), pp. 373–375.

Friston, K. J. (2011). *Functional and Effective Connectivity: A Review*. In: *Brain Connectivity* 1.1 (June 2011), pp. 13–36.

Friston, K. J., Ashburner, J., Kiebel, S., Nichols, T. and Penny, W., eds. (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.

Gaudes, C. C., Petridou, N., Dryden, I. L., Bai, L., Francis, S. T. et al. (2011). *Detection and characterization of single-trial fMRI bold responses: Paradigm free mapping*. In: *Human Brain Mapping* 32.9 (Sept. 2011), pp. 1400–1418.

Gelman, A. and Robert, C. P. (2013). *"Not Only Defended But Also Applied": The Perceived Absurdity of Bayesian Inference*. In: *The American Statistician* 67.1 (Feb. 2013), pp. 1–5.

George, E. I. and McCulloch, R. E. (1993). *Variable Selection via Gibbs Sampling*. In: *Journal of the American Statistical Association* 88.423, pp. 881–889.

Gitelman, D. R., Penny, W. D., Ashburner, J. and Friston, K. J. (2003). *Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution*. In: *NeuroImage* 19.1, pp. 200–207.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B. et al. (2013). *The minimal preprocessing pipelines for the Human Connectome Project*. In: *NeuroImage* 80 (Oct. 2013), pp. 105–124.

Glasser, M. F. and Van Essen, D. C. (2011). *Mapping Human Cortical Areas In Vivo Based on Myelin Content as Revealed by T1- and T2-Weighted MRI*. In: *The Journal of Neuroscience* 31.32 (Aug. 2011), pp. 11597–11616.

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M. et al. (2014). *Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations*. In: *Cerebral Cortex*.

Goyal, M. S., Hansen, P. J. and Blakemore, C. B. (2006). *Tactile perception recruits functionally related visual areas in the late-blind*. In: *NeuroReport* 17.13 (Sept. 2006), pp. 1381–1384.

Greicius, M. D., Krasnow, B., Reiss, A. L. and Menon, V. (2003). *Functional connectivity in the resting brain: A network analysis of the default mode hypothesis*. In: *Proceedings of the National Academy of Sciences* 100.1 (Jan. 2003), pp. 253–258.

Greicius, M. D., Srivastava, G., Reiss, A. L. and Menon, V. (2004). *Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI*. In: *Proceedings of the National Academy of Sciences* 101.13, pp. 4637–4642.

Greicius, M. D., Supekar, K., Menon, V. and Dougherty, R. F. (2009). *Resting-State Functional Connectivity Reflects Structural Connectivity in the Default Mode Network*. In: *Cerebral Cortex* 19.1, pp. 72–78.

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G. et al. (2014). *ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging*. In: *NeuroImage* 95 (July 2014), pp. 232–247.

Groppe, D. M., Makeig, S. and Kutas, M. (2009). *Identifying reliable independent components via split-half comparisons*. In: *NeuroImage* 45.4, pp. 1199–1211.

Groves, A. R., Beckmann, C. F., Smith, S. M. and Woolrich, M. W. (2011). *Linked independent component analysis for multimodal data fusion*. In: *NeuroImage* 54.3, pp. 2198–2217.

Guo, Y. and Pagnoni, G. (2008). *A unified framework for group independent component analysis for multi-subject fMRI data*. In: *NeuroImage* 42.3 (Sept. 2008), pp. 1078–1093.

Gusnard, D. A., Akbudak, E., Shulman, G. L. and Raichle, M. E. (2001). *Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function*. In: *Proceedings of the National Academy of Sciences* 98.7 (Mar. 2001), pp. 4259–4264.

Gusnard, D. A. and Raichle, M. E. (2001). *Searching for a baseline: Functional imaging and the resting human brain*. In: *Nature Reviews Neuroscience* 2.10 (Oct. 2001), pp. 685–694.

Hacker, C. D., Bundy, D., Pahwa, M., Snyder, A. Z., Szrama, N. P. et al. (2015). *Frequency-Specific ECoG Correlations Underlie fMRI Resting State Networks*. In: *Organization for Human Brain Mapping*.

Hacker, C. D., Laumann, T. O., Szrama, N. P., Baldassarre, A., Snyder, A. Z. et al. (2013). *Resting state network estimation in individual subjects*. In: *NeuroImage* 82 (Nov. 2013), pp. 616–633.

Halko, N., Martinsson, P. G. and Tropp, J. A. (2011). *Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions*. In: *SIAM Review* 53.2, pp. 217–288.

Handwerker, D. A., Ollinger, J. M. and D'Esposito, M. (2004). *Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses*. In: *NeuroImage* 21.4, pp. 1639–1651.

Hansen, E. C. A., Battaglia, D., Spiegler, A., Deco, G. and Jirsa, V. K. (2015). *Functional connectivity dynamics: Modeling the switching behavior of the resting state*. In: *NeuroImage* 105 (Jan. 2015), pp. 525–535.

Harrison, S. J., Woolrich, M. W., Robinson, E. C., Glasser, M. F., Beckmann, C. F. et al. (2015). *Large-scale Probabilistic Functional Modes from resting state fMRI*. In: *NeuroImage* 109 (Apr. 2015), pp. 217–231.

Hasan, A., McIntosh, A. M., Droese, U.-A., Schneider-Axmann, T., Lawrie, S. M. et al. (2011). *Prefrontal cortex gyrification index in twins: an MRI study*. English. In: *European Archives of Psychiatry and Clinical Neuroscience* 261.7, pp. 459–465.

Heinzel, G., Rüdiger, A. and Schilling, R. (2002). *Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new at-top windows*. Tech. rep. Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Teilinstitut Hannover.

Himberg, J., Hyvärinen, A. and Esposito, F. (2004). *Validating the independent components of neuroimaging time series via clustering and visualization*. In: *NeuroImage* 22.3 (July 2004), pp. 1214–1222.

Hinton, G. E. and Salakhutdinov, R. R. (2006). *Reducing the Dimensionality of Data with Neural Networks*. In: *Science* 313.5786, pp. 504–507.

Hjelm, R. D., Calhoun, V. D., Salakhutdinov, R., Allen, E. A., Adalı, T. et al. (2014). *Restricted Boltzmann machines for neuroimaging: An application in identifying intrinsic networks*. In: *NeuroImage* 96 (Aug. 2014), pp. 245–260.

Hoffman, E. J., Phelps, M. E., Mullani, N. A., Higgins, C. S. and Ter-Pogossian, M. M. (1976). *Design and Performance Characteristics of a Whole-Body Positron Transaxial Tomograph*. In: *Journal of Nuclear Medicine* 17.6, pp. 493–502.

Højen-Sørensen, P., Hansen, L. and Rasmussen, C. (1999). *Bayesian modelling of fMRI time series*. In: *Advances in Neural Information Processing Systems 12*. Ed. by S. Solla, T. Leen and K. Müller, pp. 754–760.

Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P. et al. (2009). *Predicting human resting-state functional connectivity from structural connectivity*. In: *Proceedings of the National Academy of Sciences* 106.6, pp. 2035–2040.

Horovitz, S. G., Braun, A. R., Carr, W. S., Picchioni, D., Balkin, T. J. et al. (2009). *Decoupling of the brain's default mode network during deep sleep*. In: *Proceedings of the National Academy of Sciences* 106.27, pp. 11376–11381.

Hsu, C.-T., Jacobs, A. M., Altmann, U. and Conrad, M. (2015). *The Magical Activation of Left Amygdala when Reading Harry Potter: An fMRI Study on How Descriptions of Supra-Natural Events Entertain and Enchant*. In: *PLoS ONE* 10.2 (Feb. 2015), e0118179.

Huettel, S. A., Song, A. W. and McCarthy, G. (2004). *Functional Magnetic Resonance Imaging*. Sinauer Associates, Apr. 2004.

Hughlings Jackson, J. (1879). *On Affectations of Speech from Disease of the Brain*. In: *Brain* 1, pp. 304–330.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D. et al. (2013). *Dynamic functional connectivity: Promise, issues, and interpretations*. In: *NeuroImage* 80 (Oct. 2013), pp. 360–378.

Hyvärinen, A. (1999). *Fast and robust fixed-point algorithms for independent component analysis*. In: *Neural Networks, IEEE Transactions on* 10.3, pp. 626–634.

Hyvärinen, A. (2013). *Independent component analysis: recent advances*. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.1984.

Hyvärinen, A. and Hoyer, P. O. (2000). *Emergence of Phase- and Shift-Invariant Features by Decomposition of Natural Images into Independent Feature Subspaces*. In: *Neural Computation* 12.7 (July 2000), pp. 1705–1720.

Hyvärinen, A., Hoyer, P. O. and Inki, M. (2001a). *Topographic Independent Component Analysis*. In: *Neural Computation* 13.7 (July 2001), pp. 1527–1558.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001b). *Independent Component Analysis*. John Wiley & Sons, Inc.

Hyvärinen, A. and Oja, E. (2000). *Independent component analysis: algorithms and applications*. In: *Neural Networks* 13.4–5 (June 2000), pp. 411–430.

Ishwaran, H. and Rao, J. S. (2005). *Spike and Slab Variable Selection: Frequentist and Bayesian Strategies*. In: *The Annals of Statistics* 33.2, pp. 730–773.

Jbabdi, S. and Behrens, T. E. J. (2012). *Specialization: the connections have it*. In: *Nature Neuroscience* 15.2 (Feb. 2012), pp. 171–172.

Jutten, C. and Herault, J. (1991). *Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture*. In: *Signal Processing* 24.1 (July 1991), pp. 1–10.

Karahanoğlu, F. I., Caballero-Gaudes, C., Lazeyras, F. and Van De Ville, D. (2013). *Total activation: fMRI deconvolution through spatio-temporal regularization*. In: *NeuroImage* 73, pp. 121–134.

Kety, S. S. (1957). *The General Metabolism of the Brain in vivo*. In: *Metabolism of the Nervous System*. Ed. by D. Richter. Pergamon, pp. 221–237.

Kiviniemi, V., Kantola, J.-H., Jauhiainen, J., Hyvärinen, A. and Tervonen, O. (2003). *Independent component analysis of nondeterministic fMRI signal sources*. In: *NeuroImage* 19.2, pp. 253–260.

Kiviniemi, V., Starck, T., Remes, J., Long, X., Nikkinen, J. et al. (2009). *Functional segmentation of the brain cortex using high model order group PICA*. In: *Human Brain Mapping* 30.12, pp. 3865–3886.

Kriegeskorte, N., Cusack, R. and Bandettini, P. A. (2010). *How does an fMRI voxel sample the neuronal activity pattern: Compact-kernel or complex spatiotemporal filter?* In: *NeuroImage* 49.3, pp. 1965–1976.

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. 2nd ed. Academic Press.

Kucukelbir, A., Ranganath, R., Gelman, A. and Blei, D. M. (2015). *Automatic Variational Inference in Stan*. In: *ArXiv e-prints* (June 2015).

Kullback, S. and Leibler, R. A. (1951). *On Information and Sufficiency*. In: *The Annals of Mathematical Statistics* 22.1 (Mar. 1951), pp. 79–86.

Kuzawa, C. W., Chugani, H. T., Grossman, L. I., Lipovich, L., Muzik, O. et al. (2014). *Metabolic costs and evolutionary implications of human brain development*. In: *Proceedings of the National Academy of Sciences* 111.36 (Sept. 2014), pp. 13010–13015.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M. et al. (1992). *Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation.* In: *Proceedings of the National Academy of Sciences* 89.12, pp. 5675–5679.

Larkman, D. J. and Nunes, R. G. (2007). *Parallel magnetic resonance imaging*. In: *Physics in Medicine and Biology* 52.7, R15.

Lashley, K. S. and Clark, G. (1946). *The cytoarchitecture of the cerebral cortex of ateles: A critical examination of architectonic studies*. In: *The Journal of Comparative Neurology* 85.2 (Oct. 1946), pp. 223–305.

Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J. et al. (2015). *Functional System and Areal Organization of a Highly Sampled Individual Human Brain*. In: *Neuron* 87.3 (Aug. 2015), pp. 657–670.

Lee, J.-H., Hashimoto, R., Wible, C. G. and Yoo, S.-S. (2011). *Investigation of spectrally coherent resting-state networks using non-negative matrix factorization for*

functional MRI data. In: *International Journal of Imaging Systems and Technology* 21.2, pp. 211–222.

Li, H. and Fan, Y. (2014). *Spatial alignment of human cortex by matching hierarchical patterns of functional connectivity*. In: *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*. Apr. 2014, pp. 329–332.

Liu, X., Yanagawa, T., Leopold, D. A., Chang, C., Ishida, H. et al. (2015a). *Arousal transitions in sleep, wakefulness, and anesthesia are characterized by an orderly sequence of cortical events*. In: *NeuroImage* 116 (Aug. 2015), pp. 222–231.

Liu, X., Yanagawa, T., Leopold, D., Schölvinck, M., Chang, C. et al. (2015b). *A Spontaneous Neurophysiological Event Underlying Spontaneous fMRI Signal Changes*. In: *Organization for Human Brain Mapping*.

Lukic, A. S., Wernick, M. N., Hansen, L. K. and Strother, S. C. (2002). *An ICA algorithm for analyzing multiple data sets*. In: *Proceedings of the International Conference on Image Processing*. Vol. 2, pp. 821–824.

Mackay, D. J. C. (1994). *Bayesian Methods for Backpropagation Networks*. In: *Models of Neural Networks III*. Ed. by E. Domany, J. L. van Hemmen and K. Schulten. New York: Springer-Verlag. Chap. 6, pp. 211–254.

MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press.

Mackay, D. J. C. (1995). *Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks*. In: *Network: Computation in Neural Systems* 6.3, pp. 469–505.

Markov, N. T., Ercsey-Ravasz, M.-M., Gariel, M.-A., Dehay, C., Knoblauch, A. et al. (2011). *Cerebral Plasticity*. Ed. by L. M. Chalupa, N. Berardi, M. Caleo, L. Galli-Resta and T. Pizzorusso. MIT Press. Chap. The tribal networks of the cerebral cortex, pp. 275–290.

Marsh, H. (2014). *Do No Harm*. Phoenix.

McKeown, M. J., Jung, T.-P., Makeig, S., Brown, G., Kindermann, S. S. et al. (1998a). *Spatially independent activity patterns in functional MRI data during the Stroop color-naming task*. In: *Proceedings of the National Academy of Sciences* 95.3 (Feb. 1998), pp. 803–810.

McKeown, M. J., Makeig, S., Brown, G. G., Jung, T.-P., Kindermann, S. S. et al. (1998b). *Analysis of fMRI Data by Blind Separation Into Independent Spatial Components*. In: *Human Brain Mapping* 6.3, pp. 160–188.

McKeown, M. J. and Sejnowski, T. J. (1998). *Independent component analysis of fMRI data: Examining the assumptions*. In: *Human Brain Mapping* 6.5-6, pp. 368–372.

Mennes, M. and Beckmann, C. F. (2015). *The Good, the Bad and the Ugly truth about global signal regression*. In: *Organization for Human Brain Mapping*.

Mitchell, T. J. and Beauchamp, J. J. (1988). *Bayesian Variable Selection in Linear Regression*. In: *Journal of the American Statistical Association* 83.404, pp. 1023–1032.

Mitra, A., Snyder, A. Z., Blazey, T. and Raichle, M. E. (2015). *Lag threads organize the brain's intrinsic activity*. In: *Proceedings of the National Academy of Sciences* 112.17, E2235–E2244.

Moeller, J. R. and Strother, S. C. (1991). *A Regional Covariance Approach to the Analysis of Functional Patterns in Positron Emission Tomographic Data*. In: *Journal of Cerebral Blood Flow and Metabolism* 11.S1 (Mar. 1991), A121–A135.

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T. T., Sepulcre, J. et al. (2013). *Individual Variability in Functional Connectivity Architecture of the Human Brain*. In: *Neuron* 77.3 (Feb. 2013), pp. 586–595.

Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B. and Bandettini, P. A. (2009). *The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced?* In: *NeuroImage* 44.3, pp. 893–905.

Neal, R. M. (1998). *Philosophy of Bayesian Inference*. Online. Jan. 1998.

Niazy, R. K., Xie, J., Miller, K. L., Beckmann, C. F. and Smith, S. M. (2011). *Spectral characteristics of resting state networks*. In: *Progress In Brain Research: Slow Brain Oscillations of Sleep, Resting State and Vigilance*. Ed. by E. J. W. Van Someren, Y. D. Van Der Werf, P. R. Roelfsema, H. D. Mansvelder and F. H. Lopes da Silva. Elsevier. Chap. 17, pp. 259–276.

Nutt, R. (2002). *The History of Positron Emission Tomography*. In: *Molecular Imaging & Biology* 4.1, pp. 11–26.

Ogawa, S. and Lee, T.-M. (1990). *Magnetic resonance imaging of blood vessels at high fields: In vivo and in vitro measurements and image simulation*. In: *Magnetic Resonance in Medicine* 16.1, pp. 9–18.

Ogawa, S., Lee, T.-M., Kay, A. R. and Tank, D. W. (1990a). *Brain magnetic resonance imaging with contrast dependent on blood oxygenation*. In: *Proceedings of the National Academy of Sciences* 87.24, pp. 9868–9872.

Ogawa, S., Lee, T.-M., Nayak, A. S. and Glynn, P. (1990b). *Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields*. In: *Magnetic Resonance in Medicine* 14.1, pp. 68–78.

Ogawa, S., Menon, R. S., Kim, S.-G. and Uğurbil, K. (1998). *On the Characteristics of Functional Magnetic Resonance Imaging of the Brain*. In: *Annual Review of Biophysics and Biomolecular Structure* 27.1, pp. 447–474.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G. et al. (1992). *Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging*. In: *Proceedings of the National Academy of Sciences* 89.13, pp. 5951–5955.

O'Reilly, J. X., Beckmann, C. F., Tomassini, V., Ramnani, N. and Johansen-Berg, H. (2010). *Distinct and Overlapping Functional Zones in the Cerebellum Defined by Resting State Functional Connectivity*. In: *Cerebral Cortex* 20.4, pp. 953–965.

Pauling, L. and Coryell, C. D. (1936). *The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin*. In: *Proceedings of the National Academy of Sciences* 22.4, pp. 210–216.

Pitzalis, S., Sereno, M. I., Committeri, G., Fattori, P., Galati, G. et al. (2010). *Human V6: The Medial Motion Area*. In: *Cerebral Cortex* 20.2, pp. 411–424.

Plis, S. M., Hjelm, D., Salakhutdinov, R., Allen, E. A., Bockholt, H. J. et al. (2014). *Deep learning for neuroimaging: a validation study*. In: *Frontiers in Neuroscience* 8.229 (Aug. 2014).

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A. et al. (2011). *Functional Network Organization of the Human Brain*. In: *Neuron* 72.4 (Nov. 2011), pp. 665–678.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A. et al. (2001). *A default mode of brain function*. In: *Proceedings of the National Academy of Sciences* 98.2 (Jan. 2001), pp. 676–682.

Robinson, E. C., Jbabdi, S., Glasser, M. F., Andersson, J., Burgess, G. C. et al. (2014). *MSM: A new flexible framework for Multimodal Surface Matching*. In: *NeuroImage* 100, pp. 414–426.

Rolfe, D. F. S. and Brown, G. C. (1997). *Cellular energy utilization and molecular origin of standard metabolic rate in mammals*. In: *Physiological Reviews* 77.3 (July 1997), pp. 731–758.

Rubinov, M. and Sporns, O. (2010). *Complex network measures of brain connectivity: Uses and interpretations*. In: *NeuroImage* 52.3, pp. 1059–1069.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L. et al. (2014). *Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers*. In: *NeuroImage* 90 (Apr. 2014), pp. 449–468.

Sämann, P. G., Wehrle, R., Hoehn, D., Spoormaker, V. I., Peters, H. et al. (2011). *Development of the Brain's Default Mode Network from Wakefulness to Slow Wave Sleep*. In: *Cerebral Cortex* 21.9, pp. 2082–2093.

Sanderson, C. (2010). *Armadillo: An Open Source C++ Linear Algebra Library for Fast Prototyping and Computationally Intensive Experiments*. Tech. rep. NICTA.

Savitzky, A. and Golay, M. J. E. (1964). *Smoothing and Differentiation of Data by Simplified Least Squares Procedures*. In: *Analytical Chemistry* 36.8 (July 1964), pp. 1627–1639.

Saygin, Z. M., Osher, D. E., Koldewyn, K., Reynolds, G., Gabrieli, J. D. E. et al. (2012). *Anatomical connectivity patterns predict face selectivity in the fusiform gyrus*. In: *Nature Neuroscience* 15.2 (Feb. 2012), pp. 321–327.

Scheperjans, F., Hermann, K., Eickhoff, S. B., Amunts, K., Schleicher, A. et al. (2008). *Observer-Independent Cytoarchitectonic Mapping of the Human Superior Parietal Cortex*. In: *Cerebral Cortex* 18.4, pp. 846–867.

Schleicher, A., Palomero-Gallagher, N., Morosan, P., Eickhoff, S. B., Kowalski, T. et al. (2005). *Quantitative architectural analysis: a new approach to cortical mapping*. English. In: *Anatomy and Embryology* 210.5-6 (Dec. 2005), pp. 373–386.

Schmithorst, V. J. and Holland, S. K. (2004). *Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data*. In: *Journal of Magnetic Resonance Imaging* 19.3 (Mar. 2004), pp. 365–368.

Schölvinck, M. L., Maier, A., Ye, F. Q., Duyn, J. H. and Leopold, D. A. (2010). *Neural basis of global resting-state fMRI activity*. In: *Proceedings of the National Academy of Sciences* 107.22 (June 2010), pp. 10238–10243.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W. et al. (1995). *Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging*. In: *Science* 268.5212 (May 1995), pp. 889–893.

Sereno, M. I., Lutti, A., Weiskopf, N. and Dick, F. (2013). *Mapping the Human Cortical Surface by Combining Quantitative T1 with Retinotopy*. In: *Cerebral Cortex* 23.9 (Sept. 2013), pp. 2261–2268.

Seung, S. (2012). *Connectome: How the Brain's Wiring Makes Us Who We Are*. Houghton Mifflin Harcourt.

Shehzad, Z., Kelly, A. M. C., Reiss, P. T., Gee, D. G., Gotimer, K. et al. (2009). *The Resting Brain: Unconstrained yet Reliable*. In: *Cerebral Cortex* 19.10 (Oct. 2009), pp. 2209–2229.

Shulman, G. L., Corbetta, M., Buckner, R. L., Fiez, J. A., Miezin, F. M. et al. (1997a). *Common Blood Flow Changes across Visual Tasks: I. Increases in Subcortical Structures and Cerebellum but Not in Nonvisual Cortex*. In: *Journal of Cognitive Neuroscience* 9.5 (Oct. 1997), pp. 624–647.

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M. et al. (1997b). *Common Blood Flow Changes across Visual Tasks: II. Decreases in Cerebral Cortex*. In: *Journal of Cognitive Neuroscience* 9.5 (Oct. 1997), pp. 648–663.

Siegel, M., Donner, T. H. and Engel, A. K. (2012). *Spectral fingerprints of large-scale neuronal interactions*. In: *Nature Reviews Neuroscience* 13.2 (Feb. 2012), pp. 121–134.

Smith, S. M., Feinberg, D. A., Griffanti, L., Harms, M. P., Kelly, M. et al. (2013a). *Resting-state fMRI in the Human Connectome Project*. In: *NeuroImage* 80 (Oct. 2013), pp. 144–168.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M. et al. (2009). *Correspondence of the brain's functional architecture during activation and rest*. In: *Proceedings of the National Academy of Sciences* 106.31, pp. 13040–13045.

Smith, S. M., Hyvärinen, A., Varoquaux, G., Miller, K. L. and Beckmann, C. F. (2014). *Group-PCA for very large fMRI datasets*. In: *NeuroImage* 101 (Nov. 2014), pp. 738–749.

Smith, S. M., Miller, K. L., Moeller, S., Xu, J., Auerbach, E. J. et al. (2012). *Temporally-independent functional modes of spontaneous brain activity*. In: *Proceedings of the National Academy of Sciences* 109.8, pp. 3131–3136.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F. et al. (2011). *Network modelling methods for FMRI*. In: *NeuroImage* 54.2, pp. 875–891.

Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E. J. et al. (2015). *A positive-negative mode of population covariation links brain connectivity, demographics and behavior*. In: *Nature Neuroscience* Advance online publication (Sept. 2015).

Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M. et al. (2013b). *Functional connectomics from resting-state fMRI*. In: *Trends in Cognitive Sciences* 17.12 (Dec. 2013), pp. 666–682.

Sokoloff, L., Mangold, R., Wechsler, R. L., Kennedy, C. and Kety, S. S. (1955). *The Effect of Mental Arithmetic on Cerebral Circulation and Metabolism*. In: *The Journal of Clinical Investigation* 34.7 (July 1955), pp. 1101–1108.

Spoormaker, V. I., Schröter, M. S., Gleiser, P. M., Andrade, K. C., Dresler, M. et al. (2010). *Development of a Large-Scale Functional Brain Network during Human Non-Rapid Eye Movement Sleep*. In: *The Journal of Neuroscience* 30.34, pp. 11379–11387.

Sporns, O., Tononi, G. and Kötter, R. (2005). *The Human Connectome: A Structural Description of the Human Brain*. In: *PLoS Computational Biology* 1.4 (Sept. 2005), e42.

Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W. and Schacter, D. L. (2010). *Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition*. In: *NeuroImage* 53.1, pp. 303–317.

Sprenger, C., Finsterbusch, J. and Büchel, C. (2015). *Spinal Cord–Midbrain Functional Connectivity Is Related to Perceived Pain Intensity: A Combined Spino-Cortical fMRI Study*. In: *The Journal of Neuroscience* 35.10 (Mar. 2015), pp. 4248–4257.

Stegle, O., Lippert, C., Mooij, J., Lawrence, N. and Borgwardt, K. (2011). *Efficient inference in matrix-variate Gaussian models with iid observation noise*. In: *Advances in Neural Information Processing Systems 24*. Ed. by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira and K. Weinberger, pp. 630–638.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P. et al. (2015). *UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age*. In: *PLOS Medicine* 12.3 (Mar. 2015), e1001779. published.

Svensén, M., Kruggel, F. and Benali, H. (2002). *ICA of fMRI Group Study Data*. In: *NeuroImage* 16.3, Part A (July 2002), pp. 551–563.

Tagliazucchi, E. and Laufs, H. (2014). *Decoding Wakefulness Levels from Typical fMRI Resting-State Data Reveals Reliable Drifts between Wakefulness and Sleep*. In: *Neuron* 82.3, pp. 695–708.

Tanner, J. M. (1962). *Growth at Adolescence*. 2nd ed. Blackwell Scientific Publications Ltd, Oxford.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A. et al. (2014). *The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data*. In: *Brain Imaging and Behavior* 8.2 (June 2014), pp. 153–182.

Titsias, M. and Lázaro-Gredilla, M. (2011). *Spike and Slab Variational Inference for Multi-Task and Multiple Kernel Learning*. In: *Advances in Neural Information*

Processing Systems 24. Ed. by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira and K. Weinberger, pp. 2339–2347.

Van Dijk, K. R. A., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W. et al. (2010). *Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties, and Optimization*. In: *Journal of Neurophysiology* 103.1 (Jan. 2010), pp. 297–321.

Van Essen, D. C. and Glasser, M. F. (2014). *In vivo architectonics: A cortico-centric perspective*. In: *NeuroImage* 93, Part 2 (June 2014), pp. 157–164.

Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J. and Coalson, T. (2012a). *Parcellations and Hemispheric Asymmetries of Human Cerebral Cortex Analyzed on Surface-Based Atlases*. In: *Cerebral Cortex* 22.10 (Oct. 2012), pp. 2241–2262.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E. et al. (2013). *The WU-Minn Human Connectome Project: An overview*. In: *NeuroImage* 80 (Oct. 2013), pp. 62–79.

Van Essen, D. C., Uğurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J. et al. (2012b). *The Human Connectome Project: A data acquisition perspective*. In: *NeuroImage* 62.4, pp. 2222–2231.

Van Horn, J. D. and Poldrack, R. A. (2009). *Functional MRI at the crossroads*. In: *International Journal of Psychophysiology* 73.1, pp. 3–9.

Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V. and Thirion, B. (2011). *Multi-subject Dictionary Learning to Segment an Atlas of Brain Spontaneous Activity*. In: *Information Processing in Medical Imaging*. Ed. by G. Székely and H. K. Hahn. Vol. 6801. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 562–573.

Varoquaux, G., Sadaghiani, S., Pinel, P., Kleinschmidt, A., Poline, J. et al. (2010). *A group model for stable multi-subject ICA on fMRI datasets*. In: *NeuroImage* 51.1, pp. 288–299.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E. and Buckner, R. L. (2008). *Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity*. In: *Journal of Neurophysiology* 100.6, pp. 3328–3342.

Walker, R., Hill, K., Burger, O. and Hurtado, A. M. (2006). *Life in the slow lane revisited: Ontogenetic separation between Chimpanzees and humans*. In: *American Journal of Physical Anthropology* 129.4, pp. 577–583.

Welvaert, M. and Rosseel, Y. (2014). *A Review of fMRI Simulation Studies*. In: *PLoS ONE* 9.7 (July 2014), e101953.

Wernicke, C. (1874). *Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis*. PhD thesis. Breslau.

Wig, G. S., Schlaggar, B. L. and Petersen, S. E. (2011). *Concepts and principles in the analysis of brain networks*. In: *Annals of the New York Academy of Sciences* 1224.1 (Apr. 2011), pp. 126–146.

Winn, J., Bishop, C. M. and Jaakkola (Ed.), T. (2005). *Variational message passing*. In: *Journal of Machine Learning Research* 6, pp. 661–694.

Woolrich, M. W., Behrens, T. E. J. and Smith, S. M. (2004). *Constrained linear basis sets for HRF modelling using Variational Bayes*. In: *NeuroImage* 21.4, pp. 1748–1761.

Worsley, K. J., Evans, A. C., Marrett, S. and Neelin, P. (1996). *A Three-Dimensional Statistical Analysis for CBF Activation Studies in Human Brain*. In: *Journal of Cerebral Blood Flow & Metabolism* 12.6 (Nov. 1996), pp. 900–918.

Yeo, T. B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D. et al. (2011). *The organization of the human cerebral cortex estimated by intrinsic functional connectivity*. In: *Journal of Neurophysiology* 106.3, pp. 1125–1165.

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L. L. and Breakspear, M. (2014). *Time-resolved resting-state brain networks*. In: *Proceedings of the National Academy of Sciences* 111.28 (July 2014), pp. 10341–10346.

Zuo, X.-N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X. et al. (2010). *Reliable intrinsic connectivity networks: Test-retest evaluation using ICA and dual regression approach*. In: *NeuroImage* 49.3 (Feb. 2010), pp. 2163–2177.