UNIVERSITY OF CALIFORNIA
SANTA CRUZ

BAYESIAN MODELING OF COMPLEX-VALUED FMRI SIGNALS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICS AND APPLIED MATHEMATICS

by

Cheng-Han Yu

June 2018

The Dissertation of Cheng-Han Yu
is approved:

Professor Raquel Prado, Chair

Professor Hernando Ombao

Professor Rajarshi Guhaniyogi

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Bayesian Modeling of Complex-valued FMRI signals

by

Cheng-Han Yu

Detecting which voxels or brain regions are activated by an external stimulus is a common objective in functional magnetic resonance (fMRI) studies, however, most studies use magnitude-only fMRI data and discard the phase data. We consider a set of statistical models for detecting brain activation at the voxel level that make use of the entire complex-valued fMRI time courses. We develop a regression model on the Cartesian representation of the complex fMRI time courses and use a complex normal spike-and-slab mixture prior on the parameters that determine brain activation at the voxel level. Our model also incorporates autoregressive components to capture temporal structure in the data. We then develop a general complex-valued expectation maximization algorithm (C-EMVS) that allows us to detect brain activation in a computationally efficient manner within this modeling framework.

To further improve detection performance, a computationally efficient Bayesian spatial model is developed to explicitly capture the spatial dependence across voxels through kernel convolution. This model encourages voxels to be activated in clusters and is able to eliminate isolated voxels that are incorrectly labeled as active in models that do not assume a spatial structure. The kernel-based method significantly reduces the computational burden compared to other spatial approaches, as it leads to dimension reduction.

We then generalize the spatial model mentioned above from a single-subject to a multi-subject model, and use the information from multiple subjects to infer brain connectivity. This model is general and practical due to its ability to infer brain activation and connectivity simultaneously in multi-subject studies.

We illustrate the performance of our statistical models and tools in extensive and physically realistic simulation studies and also in the analysis of human complex-valued fMRI data from a task-related study.

To my mom and Ya-Huei, who have loved and supported me every step of the way.

Acknowledgments

First and foremost I offer my sincerest gratitude to my advisor, Professor Raquel Prado, who has supported me throughout my thesis with her patience and knowledge whilst allowing me the room to work in my own way. She gives me insightful suggestions and warm encouragement when I get stuck with my research, and shows me how to conduct research and write a decent paper. I would also like to express my great appreciation to my committee members and collaborators, Professor Hernando Ombao, Professor Daniel B. Rowe and Dr. Rajarshi Guhaniyogi. Without their guidance and constructive comments, this thesis would not have been possible.

I cannot be an independent researcher without receiving thorough training in Bayesian modeling and computing. I would like to pay special thankfulness, warmth and appreciation to the professors below who made my research successful and assisted me at every point to cherish my goal: Dr. Bruno Sanso, Dr. Athanasios Kottas, Dr. Abel Rodriguaz, Dr. David Draper, Dr. Juhee Lee and Dr. Dongwook Lee. Moreover, I have greatly benefited from the Space-Time Meetings with UC Irvine. I thank Dr. Michele Guindani, Dr. Mark Fiecas and all participants for presenting various interesting space-time models with applications in neuroimaging.

In addition to professors in Santa Cruz, I wish to express my gratitude to professors who taught and advised me in Bloomington, Indiana. They cultivated my mathematics, statistics and computing background, and made my dream about pursuing a PhD come true. I must give my special thanks to Professor Gabriel Huerta and Richard C. Bradley. Dr. Huerta showed me the beauty of Bayesian statistics, and let

me know there is a wonderful place to do Bayesian research, UC Santa Cruz. Because of Professor Bradley's excellent teaching, probability becomes my favorite subject. I would also like to thank Professor Wen-Chieh Wu and Shuen-Zen Liu in Taiwan, who encouraged me to chase my dream and shared their experience about studying abroad in the United States.

I would not forget generous support from my friends and fellows in Bloomington and Santa Cruz. I am particularly grateful for the academic assistance given by Yu-Min, Hao-Wei, Yifei, Devin and Juan. I thank Ruei-Jr, Wei-Chieh, Chia-Te, Brenton and all other friends who helped me in life and made my life full of joy. I will always remember every single moment we shared.

Being raised in a single parent family since 11, I owe my deepest gratitude to my mother. She sent me to a better middle school far away from home and drove me to school everyday. She worked night shifts and 15 hours a day to support me financially through college. She always value my decisions and back me up. It is she who made all of these impossible possible. Finally, Ya-Huei, my best friend and the most important person in my days of studying abroad. Her companion, smile, encouragement and support made me stronger and courageous to face any challenges. I deeply appreciate everything she did for me. I cannot thank her more. Mom and Ya-Huei, I love you.

Chapter 1

Introduction

Humans are real and complex, so are data. In many science and engineering fields, complex-valued data are common and more useful than real-valued data in describing social and natural phenomena in the world. The main area of application that motivates the proposed statistical approaches and related computational methods detailed in this thesis is the analysis of brain images, in particular, functional magnetic resonance imaging (fMRI) data. We are interested in models that can describe brain activation and connectivity patterns from complex-valued fMRI (CV-fMRI) data recorded in task-related experiments.

FMRI data consist of a sequence of magnetic resonance images (MRI). In a typical fMRI experiment, three-dimensional (3D) MR images, each consisting of a large amount of equally spaced volume elements, or voxels, are acquired over time while the subject lies in the MRI scanner. The measurements of the MR signal can be represented by the so called k-space which is an array of numbers with spatial frequencies in the MR

image. As k-space data correspond to the Fourier transformation (FT) of the nuclear spin densities, the measurements are complex-valued. It is typically assumed that the real and imaginary components are measured with independent normally distributed error (Lindquist, 2008). As a result, the reconstructed voxelwise image data after inverse FT are also complex-valued with both parts following a Gaussian distribution.

However, most fMRI studies only use the magnitude portion of the signal, and phase information is discarded as it is usually assumed that the phase information does not contain any relevant signal information. Therefore, researchers often do not make use of all information contained in CV-fMRI data. Moreover, if the real and imaginary parts of the CV-fMRI data at a given time point and at given voxel are assumed to be two independent Gaussian random variables with the same variance, the magnitude data, in fact, would follow the so called Ricean distribution, which cannot be well-approximated by the Gaussian distribution when the signal-to-noise ratio (SNR) of the data is low, or the noise variance is large (Rowe & Logan, 2004). Hence, a model that takes both the real and imaginary parts into account is needed for better inference results.

Detecting which voxels or brain regions are activated by an external stimulus is one of the most fundamental objectives in fMRI studies. However, even though it is the first question fMRI researchers usually ask, it has not been fully answered yet, and there is still much room for improving the performance of currently available methods that detect activation. In fact, Eklund et al. (2016) pointed out that the most common software packages for fMRI analysis can lead to false positive rates of up to 70% when

detecting activation at the voxel level. In this work we develop novel statistical models that can utilize both real and imaginary parts of CV-fMRI data and then produce better activation detection results.

In the context of task-based fMRI experiments, brain activation at the voxel level can be modeled as a variable selection problem (Xia et al., 2009; Zhang et al., 2015). More specifically, using a regression model with explanatory variables associated to the stimuli, a voxel is labeled as active if and only if its corresponding regression coefficient in the model has a nonzero estimated value. Using a complex normal spike-and-slab mixture prior, we develop a complex-valued expectation-maximization variable selection (C-EMVS) algorithm that efficiently and accurately detects fMRI brain activation at the lowest voxel level. A corresponding complex-valued stochastic search Markov chain Monte Carlo (MCMC) algorithm is also developed to quantify the uncertainty of the activation strength at each voxel in the image. We show that our complex-valued model generates less false positives and less false negatives, leading to higher accuracy and precision rates, especially when the SNR is low in the data. Moreover, the complex-valued model can better estimate the activation strength and noise variance as well. The Bayesian nature of the model avoids multiple testing and p-value correction issues, and directly provides a probability map of activation. A model with an autoregressive noise structure is also developed to better capture the temporal structure in the data.

To further improve detection performance, a Bayesian spatial model is developed to explicitly capture the spatial dependence across voxels through kernel convolu-

tion. This model encourages voxels to be activated in clusters and is able to eliminate isolated voxels that are incorrectly labeled as active in models that do not assume a spatial structure, such as voxels outside the brain. The kernel-based model greatly reduces computational burden as it leads to significant dimension reduction in the parameter space. Furthermore, the model does not need users to pre-specify pre-specify the neighboring structure of a voxel as required in models that use Markov random fields (MRFs).

The proposed models described above have several advantages, as will be shown later in this thesis, but they can only handle data from a single subject. However, in practice, most fMRI studies, either resting-state or task-based, involve multiple participants. Therefore, instead of fitting a single-subject model independently many times, a more general approach that considers a multi-subject model to borrow information from all the subjects in the study is desired. Furthermore, in addition to localizing brain activations, another important topic is inferring brain connectivity, i.e., how brain regions interact with each other over the course of an experiment. Changes in brain connectivity are shown to be related to brain diseases such as stroke and neurodegenerative disease like Alzheimer's disease (Baldassarre et al., 2016; Iturria-Medina & Evans, 2015; Pievani et al., 2014). Therefore, we generalize our proposed kernel-based spatial model to the case of multi-subject data, and then use the information from the multiple subjects to infer brain connectivity by estimating partial correlation or conditional dependence structure of different brain regions. This estimation is done by examining the covariance matrix of subject-specific region-level spatial effects that

are associated with the probability of activation. The model is general and practical because it can not only do the multi-subject analysis, but also infer brain activation and connectivity simultaneously.

In Chapter 2 we give a general introduction to fMRI, including how a typical fMRI experiment is performed, the characteristics of fMRI data and also provide an introduction and literature review of fMRI activation and connectivity studies. Chapter 3 discusses the statistical techniques and methods on which our proposed models are based. Chapters 4, 5 and 6 present our proposed models and discuss their application to synthetic and human data. Chapter 7 concludes with a summary of our contributions and possible extensions to our statistical models and methods for CV-fMRI analysis.

Chapter 2

Functional magnetic resonance imaging

In this chapter, we give an overview about fMRI studies and fMRI data analysis, including a description of task-based fMRI experiments and the characteristics of fMRI data. We then provide a review of statistical models and methods for inferring brain activation and connectivity from fMRI data.

2.1 BOLD fMRI: Experiments and data

Neuronal activation occurs in milliseconds and cannot be easily observed. However, it is associated with localized changes in metabolism. As an imaging modality, fMRI is a noninvasive neuroimaging method that indirectly measures neuronal activity by detecting changes in the blood oxygen level dependent (BOLD) signals. When neuronal activity of a brain region increases in response to a task, the brain

area requires more local oxygen consumption, and therefore more oxygen-rich blood flows to the active brain area. As a result, due to the increased supply of oxygen in the activated brain regions, the oxyhemogloblin level is increased. As a proxy for neuronal activation, BOLD signals then measure the contrast between the blood oxyhemoglobin and deoxyhemoglobin levels arising from changes in local cerebral blood flow (Lazar, 2008; Lindquist, 2008; Zhang et al., 2015).

In comparison to high temporal resolution imaging modalities such as electroencephalography (EEG) and magnetoencephalography (MEG) that measure, respectively, electrical and magnetic activity in the brain, fMRI data are acquired with high spatial resolution but low temporal resolution due to the slower rate of brain hemodynamics.

In a typical single subject task-related fMRI experiment, a massive data set is generated by observing hemodynamic activity or BOLD responses over the entire brain volume every two to three seconds at several hundreds or thousands of time points, while a subject performs a series of tasks. In a single session, a full 3D whole brain fMRI data can be obtained by acquiring many adjacent 2D slices of fMRI data. The size of an entire 3D fMRI data is roughly $96 \times 96 \times 96$, i.e., 96 slices of 2D images each having size 96×96. In practice, many fMRI studies focus on analyzing one or several slices of 2D images. Similarly, our proposed models analyze 2D fMRI slices that usually contain 96×96 or 128×128 voxels, each having a time course of BOLD signals.

If the experiment is repeated several times for the same subject, or if multiple

subjects are involved in a fMRI study, a massive collection of data needs to be analyzed. The fact that fMRI data consist of measurements collected over time suggests a temporal dependence structure. Moreover, voxels are spatially dependent because a voxel behaves similarly to its neighbors. Hence, voxels tend to be activated in clusters or groups (Lu et al., 2003; Lazar, 2008). Another issue of fMRI data is that the signals are relatively weak in comparison to noises in the data. Therefore, a typical fMRI experiment results in a large dimensional noisy data set that exhibits a complicated spatio-temporal correlation structure (Caballero-Gaudes & Reynolds, 2017).

The BOLD fMRI signal does not rise or fall simultaneously with neural reaction when one performs a task. In general, the positive rise in signal due to the increase in the inflow of oxygenated blood to active regions of the brain gets started approximately 2 seconds after the onset of neural activity, and reaches its peak $5-8$ seconds after the neural activity has peaked. After reaching its peak, the BOLD signal decreases to a level lower than the baseline for about 10 seconds (Lindquist, 2008). This special hemodynamic response behavior is described by the so called hemodynamic response function (HRF). A fixed HRF can be used in a fMRI study, or the HRF can be assumed unknown and estimated by a statistical model. Various HRFs have been considered in fMRI studies. Figure 2.1 shows two default HRFs in two of the most commonly used software packages for fMRI analysis, the Statistical Parametric Mapping (SPM) and the FMRIB Software Library (FSL) (Pauli et al., 2016; Wager & Lindquist, 2015). SPM uses the so called canonical HRF (or double-gamma, gamma difference HRF) as the default HRF. This HRF is obtained as a linear com-

bination of two gamma functions, and it has a significant "undershoot" effect. The default HRF used in FSL is a gamma probability density function with mean 6 and standard deviation 3, or equivalently, a gamma density with shape parameter 4 and the rate parameter is 2/3. This HRF does not have the undershoot effect. In FSL, the double-gamma HRF option can be used as well.

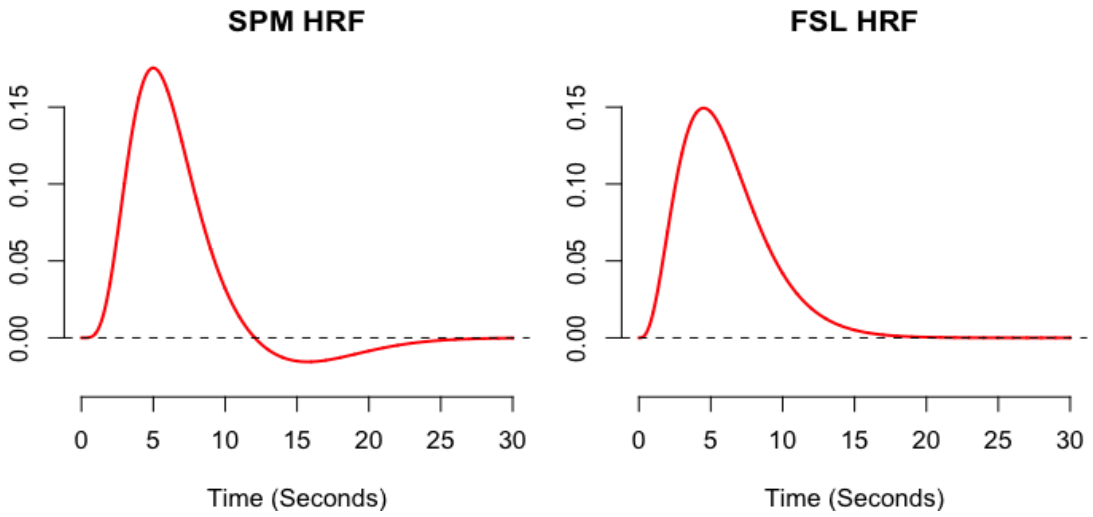


Figure 2.1: Default HRF in SPM (left) and FSL (right).

Regarding fMRI task-related studies, there are two major design classes of fMRI experiments: block designs and event-related designs. A block design separates experimental conditions into extended time intervals, or blocks. For example, a subject might perform a task, say finger tapping, for a period of time and then take a rest during another block of time. In general, block designs have high statistical power to detect activation and the detecting performance is less sensitive to the shape of the HRF. In an event-related design, several short discrete external stimuli occur randomly at different timings, offering a higher flexibility in experimental terms. For instance, a subject is

seeing pictures of faces or brief light flashes with randomized timing. Event-related designs can avoid the effects of fatigue, boredom, and systematic thought patterns that are not related to the task when block time intervals are long. Figure 2.2 gives examples of a block design and an event-related design.

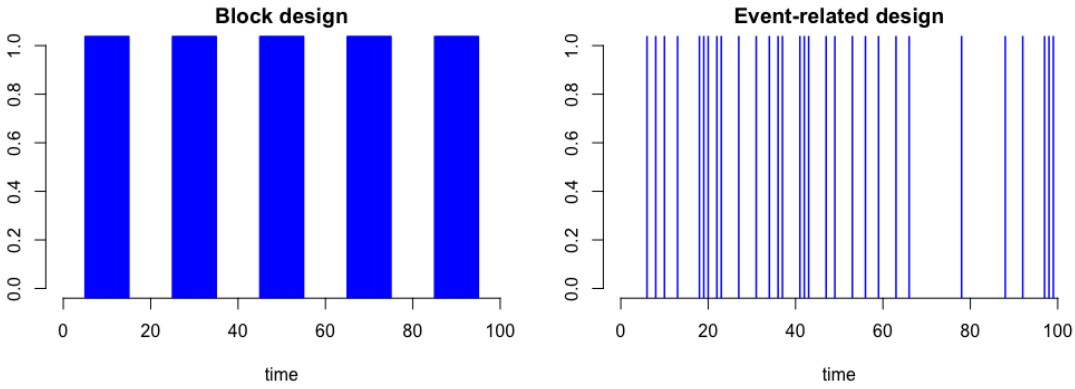


Figure 2.2: Block design and event-related design.

A widely used method for estimating BOLD signal at a given voxel and time, denoted as $x_v(t)$, is to model it as the convolution of the stimulus function $s(t)$ with the HRF $h_v(t)$ for voxel v (Poldrack et al., 2011; Lindquist, 2008; Wager & Lindquist, 2015), that is,

$$x_v(t) = (s * h_v)(t) = \int_0^t s(u)h_v(t-u)\,du.$$

Figure 2.3 shows the expected BOLD signals under a block design and an event-related design using the default canonical HRF in SPM.

To conduct activation and connectivity studies, researchers usually assume that $h_v(t)$ is known and fixed, and possibly further assume $h_v(t) = h(t)$ for all voxels v. However, the shape of HRF varies across brain voxels, within an individual and across individuals, and the true hemodynamic response is usually unknown. Recently,

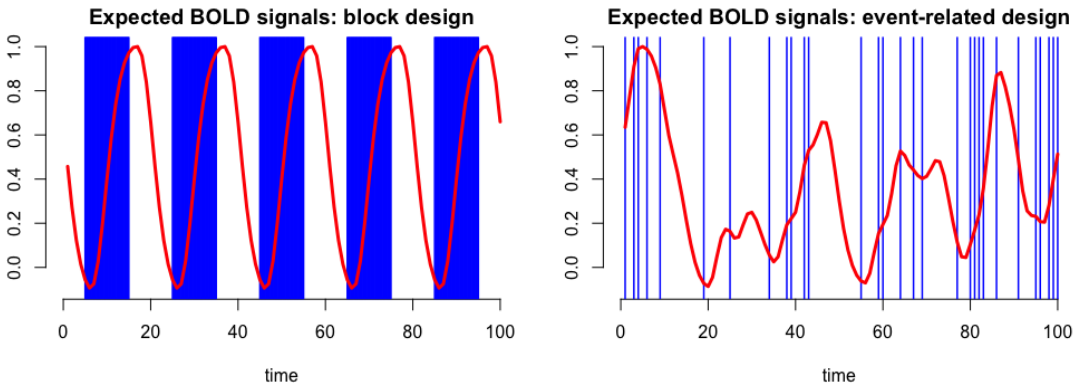


Figure 2.3: The expected BOLD signals (red) under block design and event-related design. The HRF is the canonical HRF used in SPM.

estimating HRFs has become an important research topic and several studies that use different approaches have been conducted to estimate the HRF (Goutte et al., 2000; Ciuciu et al., 2003; Lindquist et al., 2009; Pedregosa et al., 2015; Yu et al., 2016). In addition, some studies detect activation with voxel-dependent HRFs or do the detection and estimate the HRF simultaneously by a general Bayesian spatio-temporal model (Quiros et al., 2010; Woolrich et al., 2004; Xia et al., 2009; Zhang et al., 2014).

2.2 Complex-valued fMRI

As mentioned in Chapter 1, fMRI data are actually complex-valued, but most fMRI studies are based on real-valued models that use magnitude data only. In this thesis, we propose three statistical models and related inferential tools for the analysis of CV-fMRI data. In MRI and fMRI, images or voxel measurements are complex-valued due to phase imperfections after Fourier encoding and inverse Fourier image reconstruction from the k-space frequency domain to the image space and vice

versa. Each k-space point contains spatial frequency and phase information about every voxel in the final image. Thus in fMRI, voxel time course measurements consist of real and imaginary components (Bernstein et al., 1989; Haacke et al., 1999; Macovski, 1996) and these are generally converted to magnitude and phase voxel time courses. However, most fMRI brain activation studies discard the phase information and rely on magnitude-only image time courses. When this is done, the original complex-valued data are unrecoverable as operations that involve magnitude-only reconstruction are not unique.

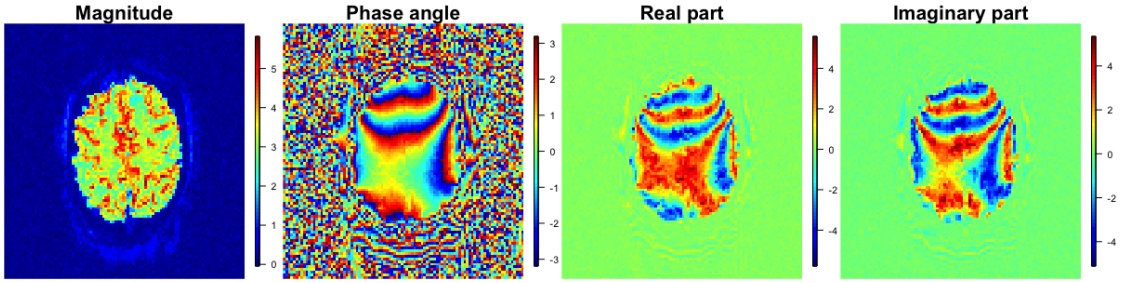


Figure 2.4: Images of CV-fMRI signals. From left to right: magnitude, phase angle, real part and imaginary part. The fMRI data set is generated from a finger-tapping experiment. The images are produced at time point 400.

Figure 2.4 shows the images of magnitude, phase angle, real part and imaginary part of CV-fMRI signals recorded in a subject who participated in a task-related experiment at time point 400. This is only a 96×96 slice of the full CV-fMRI data. The entire data set consists of 3D volumes recorded at 510 time points. The data set is generated from a finger tapping block design of experiment and is used for analysis in later chapters. Clearly, phase angles are not homogeneous across voxels and hence provide some other information that the magnitude image cannot capture. Figures 2.5 and 2.6 show time series of magnitude, phase, real and imaginary parts of signals

for a voxel labeled as activated and for a voxel labeled as non-activated, respectively. The four time series courses taken from non-activated voxels look closer to white noise, while the signals from voxels labeled as activated display some patterns that are not consistent with white noise that may be related to brain activation in such location.

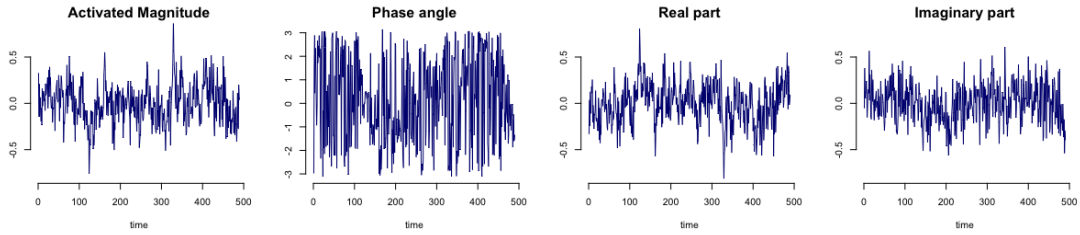


Figure 2.5: Time series of magnitude, phase, real and imaginary parts of signals for a voxel labeled as activated.

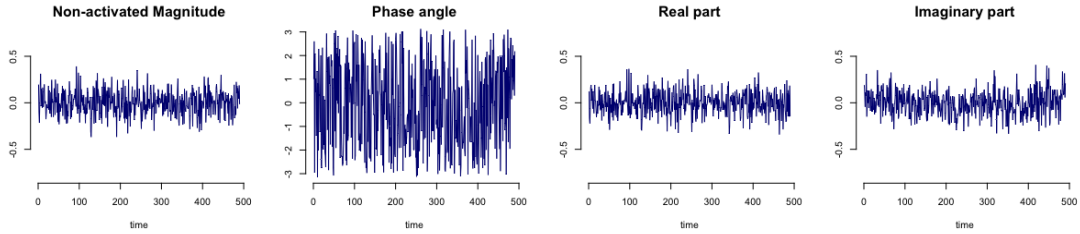


Figure 2.6: Time series of magnitude, phase, real and imaginary parts of signals for a voxel labeled as non-activated.

We note that CV-fMRI data that jointly consist of magnitude and phase images are not provided by the scanners as the default output, but they are usually readily available. For instance, GE scanners typically provide an output file that contains the raw complex-valued k-space data and other information, as well as the magnitude images. Magnitude and phase images, or real and imaginary images, can be easily obtained by simply changing a pre-set control variable in an input file, making CV-fMRI data available to neuroimaging researchers and practitioners.

2.3 fMRI activation and connectivity studies

One main objective of fMRI analysis is localizing which voxels or brain regions are activated by a task. Bandettini et al. (1992) demonstrated that voxel time courses can be used as effective tools for localizing brain function in humans. Early common model-based approaches to the analysis of magnitude fMRI data relied on the general linear model (GLM), as first proposed by Friston et al. (1994). In this model, the observed magnitude-only fMRI signal is modeled as the underlying expected BOLD response plus a noise component. In other words, for each voxel $v = 1, \ldots, V$, the voxel-wise GLM can be written as

$$\mathbf{y}^v = \mathbf{X}^v \boldsymbol{\gamma}^v + \boldsymbol{\epsilon}^v, \tag{2.1}$$

where $\mathbf{y}^v$ is the $T \times 1$ response vector of magnitude-only fMRI time course for voxel v, $\mathbf{X}^v$ is the $T \times q$ design matrix whose components include the expected BOLD responses for each of p experimental tasks or input stimuli and possibly other regressors such as trends (and so, $p \leq q$), $\boldsymbol{\gamma}^v$ is a $q \times 1$ vector of regression coefficients and $\boldsymbol{\epsilon}^v$ is a $T \times 1$ error vector, which captures random noises due to scanner artifacts and any additional subject-related physiological noise. In the absence of intercepts, trends, or any other covariates that are not task-specific, i.e., when $q = p$, each of the p BOLD responses in $\mathbf{X}^v$ is the discretized convolution of a stimulus on-and-off signal with the HRF that models the hemodynamic delay in the magnetic resonance signal (Friston et al., 2007). In addition, the HRF is often assumed to be the same across voxels,

resulting in $\mathbf{X}^v = \mathbf{X}$ for all v.

The GLM described above is for single-subject studies only. To analyze multi-subject fMRI data, several hierarchical or multi-level models have been providing a framework for performing mixed-effects analysis and dealing with individual subject level effect and group level effect in different layers of the model. Friston et al. (2002) reviews hierarchical models and establishes a connection between classical inference and parametric empirical Bayes thorough covariance component estimation of the error term $\boldsymbol{\epsilon}^v$ in the GLM (2.1). The covariance component estimation can deal with estimating nonsphericity or variance components in fMRI time series arising from serial correlations induced by multiple subjects. Beckmann et al. (2003) shows the equivalence between a two-level mixed-effects model and a single complete mixed-effects model if the covariance at the group level is set equal to the sum of the covariances in the single-level, using the best linear unbiased estimator with known covariances. The second-level analysis then can be carried out using only first-level results, without a need to revisit the fMRI time series data. Other sophisticated spatio-temporal models for detecting brain activation, particularly Bayesian models, will be reviewed in Chapter 3.

While constructing activation maps is a major interest in fMRI studies, recently, there has been an increased interest in inferring brain connectivity, i.e., studying how brain regions interact with each other in resting-state fMRI studies, and also in studying how brain connection structures react to certain experimental conditions. Functional connectivity and effective connectivity are two main types of connectivity

defined in Friston (1994). Functional connectivity refers to the undirected association between fMRI time series, or similar temporal patterns from brain regions that are spatially distant. Effective connectivity, on the other hand, estimates causal dependence or the directed influence of one brain region on another.

A simple approach to infer functional connectivity is based on estimating temporal correlations between regions of interests or between a "seed" region and other voxels throughout the brain. Dimension reduction methods such as Principal Component Analysis (PCA) in Andersen et al. (1999) and Independent Component Analysis (ICA) in McKeown et al. (1998), Calhoun et al. (2001) and Calhoun et al. (2002), are alternative methods to discover spatial patterns that account for most of the variability in the fMRI time series. Approaches that apply graphical Lasso (GLasso) to estimate sparse precision matrices of regions of interest in the image have also been proposed by Varoquaux et al. (2010).

Commonly used approaches for estimating effective connectivity include Structural Equation Modeling (SEM) by Mclntosh & Gonzalez-Lima (1994), Dynamic Causal Modeling (DCM) by Friston et al. (2003), vector autoregressive (VAR) models by Harrison et al. (2003), Granger causality by Goebel et al. (2003) and Bayesian networks by Zheng & Rajapakse (2006).

Some complex-valued based connectivity studies have also been conducted, including complex-valued time series correlation (Kociuba & Rowe, 2016), complex ICA (Du et al., 2014, 2016), and DCM for complex-valued data (Friston et al., 2012). However, the approaches mentioned in the previous paragraphs and most of other

studies are for magnitude fMRI data. Moreover, so far there are no approaches that detect activation and infer connectivity simultaneously in the complex-valued domain, and this motivates us to build the models described in Chapter 6.

.

Chapter 3

Statistical modeling background

This chapter provides some background on the statistical tools and methods that will be used in later chapters.

3.1 The complex normal distribution and complex-valued models for CV-fMRI

This section first introduces the complex normal distribution and discusses some of its properties. We then present the basic assumptions underlying complex-valued linear regression models in the context of fMRI data analysis as these are the basic building blocks of the Bayesian models for brain activation that we develop in subsequent chapters.

3.1.1 The complex normal distribution

A complex-valued random vector (r.v.) $\boldsymbol{Z} \in \mathbb{C}^n$ consists of a pair of real r.v.'s $\boldsymbol{X}$ and $\boldsymbol{Y}$ of $\mathbb{R}^n$ so that $\boldsymbol{Z} = \boldsymbol{X} + i\boldsymbol{Y}$, where $i^2 = -1$. The expectation of $\boldsymbol{Z}$ is defined as $\mathrm{E}(\boldsymbol{Z}) = \mathrm{E}(\boldsymbol{X}) + i\mathrm{E}(\boldsymbol{Y})$. Unlike real-valued Gaussian r.v.'s, the covariance matrix of $\boldsymbol{Z}$ is not enough to fully describe the second order statistical properties of $\boldsymbol{Z}$. Another matrix called the relation matrix is needed. Without loss of generality, the $n \times n$ covariance matrix $\boldsymbol{\Gamma}$ and the $n \times n$ relation matrix $\boldsymbol{C}$ are defined, with $\mathrm{E}(\boldsymbol{Z}) = 0$ as $\boldsymbol{\Gamma} := \mathrm{E}\left(\boldsymbol{Z}\boldsymbol{Z}^H\right)$ and $\boldsymbol{C} := \mathrm{E}(\boldsymbol{Z}\boldsymbol{Z}')$, where $\boldsymbol{Z}^H$ denotes the transpose conjugate of $\boldsymbol{Z}$ and $\boldsymbol{Z}'$ denotes its transpose. The covariance matrix $\boldsymbol{\Gamma}$ is complex-valued, positive (semi) definite and Hermitian, i.e., $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^H$. The relation matrix $\boldsymbol{C}$ is complex-valued and symmetric, i.e., $\boldsymbol{C} = \boldsymbol{C}'$ or $\boldsymbol{C}^* = \boldsymbol{C}^H$, where $\boldsymbol{C}^*$ denotes the conjugate of $\boldsymbol{C}$.

Definition 3.1.1 (Complex normal random variable (Picinbono, 1996))**.** A complex r.v. $\boldsymbol{Z} \in \mathbb{C}^n$ is complex normal if its real and imaginary parts $\boldsymbol{X}$ and $\boldsymbol{Y}$ are jointly (real) normal. That is, an n-dimensional r.v. $\boldsymbol{Z} = \boldsymbol{X} + i\boldsymbol{Y}$ is complex normal if and only if $\boldsymbol{Z}$ has real representation:

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \sim N_{2n}\left(\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}\right), \tag{3.1}$$

where $\Sigma_X = \mathrm{E}(\boldsymbol{X}\boldsymbol{X}')$; $\Sigma_{XY} = \mathrm{E}(\boldsymbol{X}\boldsymbol{Y}')$; $\Sigma_Y = \mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}')$.

From Definition 3.1.1, it is possible to deal with problems concerning complex r.v.'s in $\mathbb{C}^n$ by using a real r.v. of dimension $2n$, but it is usually more tedious than using the r.v. $\boldsymbol{Z}$ of $\mathbb{C}^n$ directly. Assuming $\boldsymbol{\mu}_x = \boldsymbol{\mu}_y = \boldsymbol{0}$, if the $2n \times 2n$ real covariance

matrix $\boldsymbol{\Sigma}$ in the real representation (3.1) is positive definite, the real r.v. $(\boldsymbol{X}, \boldsymbol{Y})'$ has density

$$f(\boldsymbol{x}, \boldsymbol{y}) = (2\pi)^{-n} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ \frac{-1}{2} (\boldsymbol{x}', \boldsymbol{y}') \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}, \boldsymbol{y})' \right\}.$$

The corresponding complex representation of a complex augmented r.v. $(\boldsymbol{Z}, \boldsymbol{Z}^*)'$, as shown in van den Bos (1995) and Picinbono (1996), has density

$$\begin{aligned} f(\boldsymbol{z}, \boldsymbol{z}') &= (\pi)^{-n} [|\boldsymbol{\Gamma}||\boldsymbol{P}|]^{-1/2} \exp\left\{ \frac{-1}{2} (\boldsymbol{z}^H, \boldsymbol{z}') \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{C} \\ \boldsymbol{C}^H & \boldsymbol{\Gamma}^* \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{z} \\ \boldsymbol{z}^* \end{pmatrix} \right\} \\ &= (\pi)^{-n} [|\boldsymbol{\Gamma}||\boldsymbol{P}|]^{-1/2} \exp\left\{ -\boldsymbol{z}^H \boldsymbol{P}^{-*} \boldsymbol{z} + \mathrm{Re}\left(\boldsymbol{z}' \mathbf{R}' \boldsymbol{P}^{-*} \boldsymbol{z} \right) \right\}, \end{aligned} \tag{3.2}$$

where $\boldsymbol{P} := \boldsymbol{\Gamma}^* - \boldsymbol{C}^H \boldsymbol{\Gamma}^{-1} \boldsymbol{C}$ which is Hermitian and positive definite, $\boldsymbol{P}^{-*} := \boldsymbol{\Gamma}^{-1} + \boldsymbol{\Gamma}^{-1} \boldsymbol{C} \boldsymbol{\Gamma}^{-1} \boldsymbol{C}^H \boldsymbol{\Gamma}^{-1}$ and $\mathbf{R} = \boldsymbol{C}^H \boldsymbol{\Gamma}^{-1}$. $\mathrm{Re}\left(\boldsymbol{z}' \mathbf{R}' \boldsymbol{P}^{-*} \boldsymbol{z} \right)$ denotes the real part of $\boldsymbol{z}' \mathbf{R}' \boldsymbol{P}^{-*} \boldsymbol{z}$. We write $\boldsymbol{Z} \sim CN_n(\mathbf{0}, \boldsymbol{\Gamma}, \boldsymbol{C})$. Moreover, real and complex representations are related to each other by the following equations

$$\begin{aligned} \boldsymbol{\Gamma} &= \Sigma_X + \Sigma_Y + i(\Sigma_{XY} - \Sigma_{YX}); \quad \boldsymbol{C} = \Sigma_X - \Sigma_Y + i(\Sigma_{XY} + \Sigma_{YX}), \\ \Sigma_X &= \frac{1}{2} \mathrm{Re}(\boldsymbol{\Gamma} + \boldsymbol{C}), \quad \Sigma_Y = \frac{1}{2} \mathrm{Re}(\boldsymbol{\Gamma} - \boldsymbol{C}), \\ \Sigma_{XY} &= \frac{1}{2} \mathrm{Im}(-\boldsymbol{\Gamma} + \boldsymbol{C}), \quad \Sigma_{YX} = \frac{1}{2} \mathrm{Im}(\boldsymbol{\Gamma} + \boldsymbol{C}). \end{aligned} \tag{3.3}$$

A special case of a complex normal r.v. $\boldsymbol{Z}$ appears when $\boldsymbol{C} = \mathbf{0}$, meaning that the second-order statistics of $\boldsymbol{Z}$ and its rotated variable $e^{i\alpha} \boldsymbol{Z}$ are identical for

any α. Such variable $\boldsymbol{Z}$ is called is proper or (second-order) circular when $\boldsymbol{Z}$ has mean zero (Adali et al., 2011). From (3.3), $\boldsymbol{C} = \mathbf{0}$ if and only if $\Sigma_X = \Sigma_Y$ and $\Sigma_{XY} = -\Sigma_{YX}$. This means that for a circular complex r.v., its real and imaginary parts have exactly the same covariance structure, and they do not have any correlation. With this property and the density shown in (3.2), we obtain that the density function of a circular normal r.v. $\boldsymbol{Z}$ is of the form

$$f(\boldsymbol{z}) = (\pi)^{-n} |\boldsymbol{\Gamma}|^{-1} \exp\left\{-\boldsymbol{z}^H \boldsymbol{\Gamma}^{-1} \boldsymbol{z}\right\}. \tag{3.4}$$

This density, therefore, does not contain $\boldsymbol{C}$ and $\boldsymbol{z}^*$, and the r.v.'s behavior is fully characterized by $\boldsymbol{\Gamma}$. Next, we define independence of two complex normal random variables.

Definition 3.1.2. Suppose $\boldsymbol{Z} = (Z_1, Z_2)' \sim CN_2(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{C})$. Z_1 and Z_2 are independent if and only if $\boldsymbol{\Gamma}$ and $\boldsymbol{C}$ are both diagonal.

By this definition, the following corollary is straightforward.

Corollary 3.1.1. *Suppose* $\boldsymbol{Z} = (\boldsymbol{Z}_1', \boldsymbol{Z}_2')' \sim CN_{n_1+n_2}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{C})$, *where* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')'$, $\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}_{12}^H & \boldsymbol{\Gamma}_2 \end{pmatrix}$ *and similarly for* $\boldsymbol{C}$. $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}_2$ *are independent if and only if* $\boldsymbol{\Gamma}_{12} = \boldsymbol{C}_{12} = \mathbf{0}$.

Notice that a circular complex r.v. $\boldsymbol{Z}$ can have non-independent elements as long as Σ_X, and hence $\boldsymbol{\Gamma}$, is not diagonal. A non-circular complex r.v. $\boldsymbol{Z}$ can also have independent elements as long as Σ_X, Σ_Y and Σ_{XY} are all diagonal. Hence, circularity does not imply independence, and vice versa.

Like real normality, complex normality is preserved under linear transformations. That is, $\mathbf{A}\boldsymbol{Z}+\mathbf{b} \sim CN_k\left(\mathbf{A}\boldsymbol{\mu}+\mathbf{b}, \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^H, \mathbf{A}\boldsymbol{C}\mathbf{A}'\right)$ for any $k \times n$ complex-valued matrix $\mathbf{A}$ and $k \times 1$ complex constant vector $\mathbf{b}$.

Now we summarize the properties of conditional complex normal distributions. From Picinbono (1996), if $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are two complex normal r.v.'s with each mean zero, then $\left(\boldsymbol{Z}_1', \boldsymbol{Z}_2'\right)'$ is jointly complex normal $CN\left((\mathbf{0}', \mathbf{0}')', \boldsymbol{\Gamma}, \boldsymbol{C}\right)$, where $\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_2 \end{pmatrix}$, $\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_2 \end{pmatrix}$ and $\boldsymbol{\Gamma}_j = \mathrm{E}\left(\boldsymbol{Z}_j \boldsymbol{Z}_j^H\right)$, $\boldsymbol{C}_j = \mathrm{E}\left(\boldsymbol{Z}_j \boldsymbol{Z}_j'\right), j = 1, 2,$ and $\boldsymbol{\Gamma}_{12} = \mathrm{E}\left(\boldsymbol{Z}_1 \boldsymbol{Z}_2^H\right)$, $\boldsymbol{C}_{12} = \mathrm{E}\left(\boldsymbol{Z}_1 \boldsymbol{Z}_2'\right)$. Moreover, the r.v. $\boldsymbol{Z}_1 \mid \boldsymbol{Z}_2$ also follows a complex normal distribution $CN\left(m\left(\boldsymbol{z}_2\right), \boldsymbol{E}, \boldsymbol{F}\right)$, where

$$
\begin{aligned}
m_1\left(\boldsymbol{z}_2\right) &= \mathbf{A}\boldsymbol{z}_2 + \mathbf{B}\boldsymbol{z}_2^*, \\
\boldsymbol{E} &= \boldsymbol{\Gamma}_1 - \mathbf{A}\boldsymbol{\Gamma}_{12}^H - \mathbf{B}\boldsymbol{C}_{12}^H; \quad \boldsymbol{F} = \boldsymbol{C}_1 - \mathbf{A}\boldsymbol{C}_{12}' - \mathbf{B}\boldsymbol{\Gamma}_{12}', \\
\mathbf{A} &= \left(\boldsymbol{\Gamma}_{12} - \boldsymbol{C}_{12}\boldsymbol{\Gamma}_2^{-*}\boldsymbol{C}_2^H\right)\mathbf{P}_2^{-*}; \quad \mathbf{B} = \left(\boldsymbol{C}_{12} - \boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{C}_2\right)\mathbf{P}_2^{-1}, \\
\mathbf{P}_2 &= \boldsymbol{\Gamma}_2^* - \boldsymbol{C}_2^H\boldsymbol{\Gamma}_2^{-1}\boldsymbol{C}_2,
\end{aligned}
\tag{3.5}
$$

and $\boldsymbol{\Gamma}_2^{-*} = \left(\boldsymbol{\Gamma}_2^{-1}\right)^*$. $\boldsymbol{Z}_2 \mid \boldsymbol{Z}_1$ has the same distribution structure as $\boldsymbol{Z}_1 \mid \boldsymbol{Z}_2$ with appropriate indices replaced.

3.1.2 Complex-valued models for CV-fMRI

Many magnitude-only approaches work under the assumption that errors in the model are normally distributed which may be problematic, resulting in incorrect standard errors that can produce inaccurate activation results. In fact, if both the real

and imaginary components of the CV-fMRI signals $y_t^v = y_{t,Re}^v + iy_{t,Im}^v$, have independent normally distributed errors with the same variance, the magnitude-only signals $y_{t,Mag}^v = \sqrt{(y_{t,Re}^v)^2 + (y_{t,Im}^v)^2}$ follow a Ricean (Rice or Rician) distribution. Here, a_{Re} and a_{Im} generically denote the real and imaginary parts of any complex-valued quantity $a = a_{Re} + ia_{Im}$. The argument or phase, $\phi_t^v = \arg(y_t^v)$, can be determined from the equation $\tan(\phi_t^v) = y_{t,Im}^v / y_{t,Re}^v$, and $\left(\phi_t^v \mid y_{t,Mag}^v\right)$ follows Tikhonov distribution (O'Donoughue & Moura, 2012).

The Ricean distribution is approximately normal only in the case of large SNRs (Rowe & Logan, 2004; Gudbjartsson & Patz, 1995; Rice, 1944). However, the SNRs may not be large enough in practice for this approximate normality to hold. This is increasingly true in cases with higher voxel resolutions and for voxels with a large degree of signal drop-out, i.e., those for which the signal is not available or has small SNR, such as voxels located near air/tissue boundaries. In particular, Adrian et al. (2013) show that with magnitude-only models, tests derived using Ricean modeling are superior to Gaussian-based activation tests for SNRs below 0.6. Rowe (2005b) also shows that Gaussian based activation parameter estimates were biased for SNRs under 10.

Complex-valued modeling has been widely used in several applied areas allowing full utilization of real and imaginary, or equivalently magnitude and phase, information in certain signals and images, providing a general framework for the analysis of several classes of processes (see, e.g., Mandic & Goh, 2009). The incorporation of phase information has proven key in communications and imaging (Oppenheim & Lim,

1981), as complex-valued modeling simultaneously handles the intensity and direction when dealing with radar, sonar and wind data.

A number of tools for CV-fMRI data analysis have been proposed in the literature, including non model-based exploratory ICA (Calhoun et al., 2002), as well as direct modeling of the complex activation data (Lai & Glover, 1997; Rowe & Logan, 2004, 2005; Rowe, 2005a; Lee et al., 2007; Rowe, 2009; Lee et al., 2009). Approaches such as those in Rowe & Logan (2004, 2005); Rowe (2005a) and Rowe (2005b) model the phase to directly estimate the phase angle using a polar coordinates representation, while the methods in Lee et al. (2007) and Lee et al. (2009) are based on Cartesian representations. More recently, complex-valued models with temporal correlations (including autoregressive structures) have also been developed (Adrian et al., 2017; Kociuba & Rowe, 2016). In particular, Rowe (2005a) specifies the following structure for the complex-valued image measurement at time t and voxel v, $y_t^v = y_{t,Re}^v + iy_{t,Im}^v \in \mathbb{C}$, as

$$y_t^v = \rho_t^v \cos(\phi_t^v) + i\rho_t^v \sin(\phi_t^v) + \eta_t^v, \tag{3.6}$$

where $\rho_t^v = \beta_0^v + \beta_1^v x_{1,t} + \ldots + \beta_{p_1}^v x_{p_1,t}$ is the magnitude of y_t^v with p_1 magnitude regressors, $\phi_t^v = \alpha_0^v + \alpha_1^v u_{1,t} + \ldots + \alpha_{p_2}^v u_{p_2,t}$ is the phase of y_t^v with p_2 regressors, and $i = \sqrt{-1}$. All the regression coefficients $\beta_0^v, \ldots, \beta_{p_1}^v$ and $\alpha_0^v, \ldots, \alpha_{p_2}^v$ are real-valued. The noise term η_t^v is also assumed to be complex-valued, i.e., $\eta_t^v = \eta_{t,Re}^v + i\eta_{t,Im}^v$. When $\alpha_0^v \neq 0$ and $\alpha_j^v = 0$ for all $j = 1, \ldots, p_2$ we have the Rowe-Logan constant phase model. Note that when no trends are included, the magnitude and phase regressors

could be chosen to be identical to the expected BOLD responses associated with the p experimental tasks, i.e., $p_1 = p_2 = p$ and $x_{j,t} = u_{j,t}$ for all $j = 1, \ldots, p$. Rowe (2005a) identifies active voxels using a generalized likelihood ratio test.

Lee et al. (2007) and Lee et al. (2009) propose a method based on a Cartesian model representation which has the following matrix form

$$\mathbf{y}^v = \mathbf{X}\boldsymbol{\gamma}^v + \boldsymbol{\eta}^v, \tag{3.7}$$

with $\mathbf{y}^v = (y_1^v, \ldots, y_T^v)'$, $\boldsymbol{\gamma}^v = \boldsymbol{\gamma}_{Re}^v + i\boldsymbol{\gamma}_{Im}^v$, $\boldsymbol{\gamma}_{Re}^v = \left(\gamma_{Re,1}^v, \ldots, \gamma_{Re,q}^v\right)'$, $\boldsymbol{\gamma}_{Im}^v = \left(\gamma_{Im,1}^v, \ldots, \gamma_{Im,q}^v\right)'$, with $q = p + 1$, $\mathbf{X} = (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_T')'$, where $\boldsymbol{x}_t = (1, x_{1,t}, \ldots, x_{p,t})'$, $t = 1, \ldots, T$, and complex-valued noise vector $\boldsymbol{\eta}^v = (\eta_1^v, \ldots, \eta_T^v)$. Lee et al. (2007) combines this general linear model representation in Cartesian coordinates with a Hotelling's T^2-test to detect active sites. Model (3.7) is equivalent to the Rowe-Logan constant phase complex-valued model (Rowe & Logan, 2004) if $p_1 = p$, $\boldsymbol{\gamma}_{Re}^v = \left(\beta_0^v, \ldots, \beta_p^v\right)' \cos\left(\alpha_0^v\right)$ and $\boldsymbol{\gamma}_{Im}^v = \left(\beta_0^v, \ldots, \beta_p^v\right)' \sin\left(\alpha_0^v\right)$. Model (3.7) is also equivalent to the complex-valued magnitude and phase activation model in Rowe & Logan (2005) when there is only a single regressor in both, magnitude and phase, corresponding to a 0/1 vector representing a boxcar block design.

The references cited above show that modeling the complete CV-fMRI data leads to superior power in detecting active voxels when compared to magnitude-only approaches, especially for situations in which the SNRs are relatively small. However, in spite of their advantages, currently available methods for CV-fMRI data rely on mechanisms that control some notion of error to correct for multiple testing, such

as Bonferroni corrections, and therefore involve two-step procedures. The first step provides estimates of the potentially active voxels according to some model, while the second step involves using one of the standard methods to correct for multiple testing. Furthermore, available methods for CV-fMRI data assume that the voxels are independent and do not offer a principled framework for parameter learning through borrowing information across voxels.

In the next chapters we propose a class of complex-valued Bayesian models for detecting activation at the voxel level. Our models are based on Bayesian variable selection methods that do not require additional adjustments for multiple testing. All the models utilize the real and imaginary information provided by the CV-fMRI data to obtain a more accurate detection of active voxels. Before providing a description of our statistical models and inferential tools we review some variable selection approaches that are widely used in the context of analyzing neuroimaging data.

3.2 Variable selection methods

Detection of brain activation in task-related fMRI can be viewed as a variable selection problem (Xia et al., 2009; Zhang et al., 2015) in which a voxel is activated if and only if the corresponding regression coefficient related to the expected BOLD contrasts in the regression model is nonzero. Tibshirani (1996) provides the variable selection method "Lasso" (least absolute shrinkage and selection operator) that has been prevalently used because of its computational efficiency and easy interpretation. In an ordinary, real-valued, linear regression model of the form $y_i = \gamma_0 + \sum_{j=1}^{q} \gamma_j x_{i,j} + \epsilon_i$,

or $\mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$ in matrix form, the lasso estimate coefficient is

$$\hat{\boldsymbol{\gamma}}^{lasso} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \left\{ \|\mathbf{y} - \mathbf{X}\boldsymbol{\gamma}\|^2 + \lambda \sum_{j=1}^{q} |\gamma_j| \right\},$$

where q is the number of tasks and $\lambda \geq 0$ is a penalty or complexity parameter that controls the amount of shrinkage. The larger value of λ, the more coefficients shrink to zero. Lasso forces the coefficients to be equally penalized, leading to asymptotic inconsistency for variable selection under some scenarios. To solve this problem, Zou (2006) proposed a new version of the lasso, called the adaptive lasso, where adaptive weights λ_j are used for penalizing different coefficients in the penalty. That is,

$$\hat{\boldsymbol{\gamma}}^{adaptive} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \left\{ \|\mathbf{y} - \mathbf{X}\boldsymbol{\gamma}\|^2 + \sum_{j=1}^{q} \lambda_j |\gamma_j| \right\}.$$

Adaptive lasso is shown to have better asymptotic results than lasso, and it can be solved by the same efficient algorithm used for solving the lasso.

There are also several different kinds of approaches of Bayesian variable selection methods. O'Hara & Sillanpaa (2009) classifies the approaches to variable selection into four categories: indicator model selection, stochastic search variable selection (SSVS), adaptive shrinkage, and model space approach. In this thesis, we do variable selection by introducing indicator variables that determine whether a voxel is activated or not. In particular, we develop complex-valued version of SSVS, and add spatial dependence on the indicators to detect activation. Therefore, here we briefly review the

indicator model selection and SSVS approaches. For details about adaptive shrinkage methods, including the Bayesian lasso, and model space approaches that use reversible jump MCMC, please see the discussion in O'Hara & Sillanpaa (2009).

Let ψ_j be an auxiliary indicator variable where $\psi_j = 1$ indicates presence of covariate j in a linear regression model $y_i = \gamma_0 + \sum_{j=1}^{q} \gamma_j x_{i,j} + \epsilon_i$, Kuo & Mallick (1998) consider the expanded linear regression model

$$y_i = \gamma_0 + \sum_{j=1}^{q} \psi_j \gamma_j x_{i,j} + \epsilon_i,$$

and assume that indicators ψ_j and covariate effects γ_j are independent *a priori*, i.e., $\pi(\psi_j, \gamma_j) = \pi(\psi_j)\,\pi(\gamma_j)$. The posterior distribution of the indicators is obtained by means of the MCMC sample. The model does not require tuning parameters, but when $\psi_j = 0$, the updated value of γ_j is sampled from its prior distribution, which may lead to poor mixing if the prior distribution is too vague, as the sampled γ_j may not fall in the region in which γ_j has high posterior density value when $\psi_j = 1$.

SSVS (George & McCulloch, 1993, 1997) has been widely used in many fields and extended to various models such as generalized linear models, time series and factor analysis. SSVS assumes that $\pi(\psi_j, \gamma_j) = \pi(\gamma_j \mid \psi_j)\,\pi(\psi_j)$, and that the spike part is a narrow continuous distribution centered at zero. Specifically,

$$\gamma_j \mid \psi_j \sim (1 - \psi_j)\, N\left(0, \tau^2\right) + \psi_j N\left(0, c\tau^2\right),$$

where $c > 1$ is set large so that if $\psi_j = 1$, a non-zero estimate of γ_j should be

included in the model. In general, we can write the spike-and-slab prior with two tuning parameters v_0 and v_1 as

$$\gamma_j \mid \psi_j \sim (1 - \psi_j) N(0, v_0) + \psi_j N(0, v_1),$$

where the hyperparameters v_0 and v_1 are set small and large respectively. Continuity on both the spike and slab part makes the Gibbs sampler feasible and easy to implement.

George & McCulloch (1997) provides a "conjugate" version in which the error variance in the regression model is in the spike-and-slab mixture:

$$\gamma_j \mid \psi_j \sim (1 - \psi_j) N\left(0, v_0\sigma^2\right) + \psi_j N\left(0, v_1\sigma^2\right).$$

One advantage of using the conjugate prior is that $\boldsymbol{\gamma}$ and σ^2 can be analytically integrated out from the posterior. George & McCulloch (1997) shows that the non-conjugate prior generates iterations faster than the conjugate prior, especially when the number of covariates is large. However, when covariates are correlated or the multicollinearity exists, the conjugate prior is superior to the non-conjugate prior. In addition, if v_1/v_0 is too large, say above 2500, the non-conjugate prior performs poorly, having higher Monte Carlo standard errors, and very small transition probabilities for ψ_j from 0 to 1 and 1 to 0.

More recently Rockova & George (2014) proposes a Expectation-Maximization variable selection (EMVS) algorithm, a deterministic alternative to stochastic search based on an EM algorithm, that can quickly identify promising high posterior modes.

EMVS uses the conjugate version of the spike-and-slab normal mixture. Due to the continuity of the spike distribution, the EM algorithm has rapidly computable closed form expressions. For model evaluation, a point mass spike distribution is used to determine which of the discovered submodels is best supported by the data. One of our proposed algorithms, C-EMVS, uses this type of approach due to its computational feasibility for massive fMRI data.

Detecting activation can also be viewed as a multiple comparison problem. Müller et al. (2006) uses a decision theoretic perspective to discuss Bayesian approaches to multiple comparison problems. In fact, posterior inference adjusts for multiplicities, and no further adjustment is required if the model includes a positive prior probability of non-activation for each voxel and includes a hyperparameter that defines the prior probability mass for non-activation. Müller et al. (2004) showed that under several loss functions that combine false negative and false discovery counts and/or rates, the optimal decision rule δ_v^* is to declare all voxels with marginal probability beyond a threshold t as activated, i.e., $\delta_v^* = I\left(Pr(\psi^v = 1 \mid \mathbf{y}) > t\right)$, where $I(\cdot)$ is an indicator function. For example, if the loss function is of the type $L(\delta) = c\overline{FD} + \overline{FN}$, where $\overline{FD}$ and $\overline{FN}$ are posterior mean of false discoveries and false negatives, respectively, then the optimal threshold value is given by $t^* = \frac{c}{c+1}$. Therefore, when false discoveries and false negatives are weighted equally, the optimal threshold is 0.5.

3.3 Spatio-temporal modeling for fMRI data

The traditional GLM (2.1) performs fMRI data analysis in a univariate setting, assuming that neighboring voxels are independent and there is no time dependence. This is not reasonable since it ignores the fact that voxels tend to be activated in clusters or groups, and measurements recorded over the time course of the experiment may induce temporal correlation. Bayesian hierarchical models based on GLM (2.1) are flexible in modeling temporal and spatial dependence structures of fMRI data.

To capture temporal correlation in fMRI data, one of the most common approaches is to put an autoregressive structure of order k (AR(k)) on $\boldsymbol{\epsilon}^v$, and then assign prior distributions on the AR coefficients (Lee et al., 2014; Penny et al., 2003, 2005). Other time dependence sources include low-frequency drift due to slow fluctuations in the scanner hardware, and physiological noise due to subject motion, heart beat and respiratory rate. To account for these sources, researchers often perform high-pass filtering by adding covariates such as a level shift or a deterministic trend, for example, a pth order polynomial function (Lindquist, 2008).

More sophisticated Bayesian models, including spatial and spatio-temporal approaches, have been developed for magnitude-only fMRI data. Some approaches induce spatial dependency by imposing a spatial prior on the regression coefficients. For example, Gossl et al. (2001) and Quiros et al. (2010) used Gaussian Markov random field (GMRF) priors on the jth regression coefficient vector $\gamma_{(j)} = \left(\gamma_j^1, \ldots, \gamma_j^V\right)$, and Penny et al. (2005) considered a spatial prior on the regression coefficient vector using the Laplacian operator. Other spatial priors on the regression coefficients include the

diffusion-based spatial priors of Harrison et al. (2008), the conditional autoregressive (CAR) priors of Harrison & Green (2010) and the sparse spatial basis function (SSBF) priors of Flandin & Penny (2007).

Alternative spatial models view the detection activation as a variable selection problem and impose spatial dependence on the binary indicator variables ψ^v in the spike-and-slab mixture priors. Kalus et al. (2014) specified a spatial probit model such that $Pr(\psi^v = 1) = \Phi(\alpha_v)$, where Φ is the standard normal cdf and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_V)$ follows a GMRF. Smith & Fahrmeir (2007) propose an Ising prior to smooth spatially the indicator variables representing whether or not the voxel is activated. Bowman et al. (2008) consider a two-stage Bayesian hierarchical model with temporal correlations at the first stage and spatial correlations at the second stage. In Lee et al. (2014), temporal dependence is characterized via autoregressive models, Zellner's g-priors are assumed for the regression coefficients, and a binary spatial Ising prior is used to specify anatomical information and spatial interaction between voxels. In Zhang et al. (2014) a general error structure is used to capture general dependence, and a Markov random field prior is used to detect activations in a nonparametric way.

Notice that all the approaches referenced above are for magnitude-only data. To analyze CV-fMRI data, in this thesis, we develop complex-valued Bayesian variable selection models that utilize both real and imaginary signals for detecting activation. In order to keep the model structure as simple and parsimonious as possible, we avoid introducing spatial dependence on the complex-valued regression coefficients. Instead, we follow the approach of inducing a spatial structure through the binary indicator

variables. Note that in this way we also jointly use the information provided by the real and imaginary data to determine activation. Furthermore, in Chapter 5, we model spatial dependence through kernel convolution which leads to dimension reduction and hence computational efficiency. Moreover, unlike MRF models, our proposed spatial model does not require a specification of the voxel neighboring structure before the analysis.

Chapter 4

A Bayesian variable selection approach yields improved activation detection from CV-fMRI

The focus of this chapter is the development of a Bayesian variable selection approach that efficiently yields improved activation detection from CV-fMRI. As mentioned above, we develop a model that makes use of the complete magnitude and phase information provided by the CV-fMRI data. However, unlike previous approaches (Rowe & Logan, 2004, 2005; Rowe, 2005a,b, 2009; Lee et al., 2007) discussed in Chapter 2, we utilize a fully Bayesian framework for identifying active voxels via variable selection in the complex-valued domain.

We follow the Cartesian coordinates approach of Lee et al. (2007) given in (3.7) and further assume independent and identically distributed complex-normal error vectors, i.e.,

$$\mathbf{y}^v = \mathbf{X}\boldsymbol{\gamma}^v + \boldsymbol{\eta}^v, \quad \boldsymbol{\eta}^v \sim CN_T\left(\mathbf{0}, \boldsymbol{\Gamma}_v, \mathbf{C}_v\right), \tag{4.1}$$

with $CN_L\left(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C}\right)$ denoting a complex normal distribution of dimension L with mean $\boldsymbol{\mu}$, complex-valued, Hermitian and non-negative definite covariance matrix $\boldsymbol{\Gamma}$, and complex-valued symmetric relation matrix $\mathbf{C}$. As shown below, the linear structure in this representation is computationally relevant, as it leads to fast Bayesian posterior estimation of active sites. Note also that any complex-valued normal distribution of dimension L has a real-valued normal representation of dimension $2L$ (see Chapter 3 and Wooding, 1956; van den Bos, 1995; Picinbono, 1996). Thus, letting $\boldsymbol{\Sigma}^v_{Re,Re} = \frac{1}{2}Re\left(\boldsymbol{\Gamma}_v + \mathbf{C}_v\right)$, $\boldsymbol{\Sigma}^v_{Im,Im} = \frac{1}{2}Re\left(\boldsymbol{\Gamma}_v - \mathbf{C}_v\right)$, $\boldsymbol{\Sigma}^v_{Re,Im} = \frac{1}{2}Im\left(-\boldsymbol{\Gamma}_v + \mathbf{C}_v\right)$, and $\boldsymbol{\Sigma}^v_{Im,Re} = \frac{1}{2}Im\left(\boldsymbol{\Gamma}_v + \mathbf{C}_v\right)$, model (4.1) also has a real-valued representation as

$$\mathbf{y}^v_r = \mathbf{X}^r\boldsymbol{\gamma}^v_r + \boldsymbol{\eta}^v_r, \tag{4.2}$$

with $\mathbf{y}^v_r = \left((\mathbf{y}^v_{Re})', (\mathbf{y}^v_{Im})'\right)'$, $\mathbf{X}^r = \text{blockdiag}\left(\mathbf{X}, \mathbf{X}\right)$, $\boldsymbol{\gamma}^v_r = \left((\boldsymbol{\gamma}^v_{Re})', (\boldsymbol{\gamma}^v_{Im})'\right)'$, and $\boldsymbol{\eta}^v_r = \left((\boldsymbol{\eta}^v_{Re})', (\boldsymbol{\eta}^v_{Im})'\right)'$, where $\boldsymbol{\eta}^v_r \sim N_{2T}\left(\mathbf{0}, \boldsymbol{\Sigma}_v\right)$ with

$$\boldsymbol{\Sigma}_v = \begin{pmatrix} \boldsymbol{\Sigma}^v_{Re,Re} & \boldsymbol{\Sigma}^v_{Re,Im} \\ \boldsymbol{\Sigma}^v_{Im,Re} & \boldsymbol{\Sigma}^v_{Im,Im} \end{pmatrix}.$$

The simplest possible structure for $\boldsymbol{\eta}^v$ is that obtained by taking $\boldsymbol{\eta}^v \sim CN_T\left(\mathbf{0}, 2\sigma_v^2\mathbf{I}_T, \mathbf{0}\right)$ or equivalently, setting $\boldsymbol{\Sigma}_v = \sigma_v^2\mathbf{I}_{2T}$ in the real-valued Gaussian representation. This implies that there is no correlation within the real components and within the imaginary components of $\boldsymbol{\eta}^v$, and also that there is no correlation between the real and imaginary components of $\boldsymbol{\eta}^v$. These assumptions can be relaxed to include correlations within the real and imaginary components in order to capture temporal structure (as illustrated in some of the analysis of synthetic and human CV-fMRI data presented in Sections 4.3 and 4.4), or correlations between the real and imaginary components for more structured noise.

Below we describe the priors and the corresponding posterior inference for the simplest noise structure, focusing on complex-valued priors for $\boldsymbol{\gamma}^v$ that lead to posterior inference of activation in CV-fMRI at the voxel-specific level.

4.1 Priors

In the absence of any trends and intercepts, and without loss of generality, i.e., for the case in which $\mathbf{X}$ in (4.1) contains only the expected BOLD signals for each of p stimuli/tasks with no baselines or trends, activation can be viewed as a variable selection problem (Xia et al., 2009; Zhang et al., 2015). In other words, if $\gamma_j^v = \gamma_{Re,j}^v + i\gamma_{Im,j}^v \neq 0$ for voxel v and task j, such voxel is identified as active under task j. Note that complex-valued priors must be considered for γ_j^v. Here we develop a complex-valued analogue of the Bayesian variable selection methods of George & McCulloch (1993, 1997) and Rockova & George (2014). If trends and/or intercepts are needed

they can easily be included in the model along with priors on their corresponding parameters and integrated out, as done in the applications illustrated in Sections 4.3 and 4.4. Thus, we focus the discussion below to the case in which $\mathbf{X}$ only consists of the expected BOLD signals associated to each of the p experimental stimuli/tasks.

Our proposed complex-valued spike-and-slab priors for γ_j^v extend the widely used real-valued spike-and-slab priors by considering

$$\gamma_j^v \mid \psi_j^v \sim \left(1-\psi_j^v\right) g_0\left(\gamma_j^v\right) + \psi_j^v g\left(\gamma_j^v\right),$$

with $g_0(\cdot)$ and $g(\cdot)$ complex-valued distributions with mean zero, and $\psi_j^v \in \{0,1\}$, where $\psi_j^v = 1$ indicates that voxel v is active during task j. Therefore, this prior allows us to determine if a voxel is active by jointly considering the real and imaginary components of γ_j^v. In general, we consider priors with $g_0\left(\gamma_j^v\right) = CN_1\left(0, \sigma_v^2\omega_0, \sigma_v^2\lambda_0\right)$, and $g\left(\gamma_j^v\right) = CN_1\left(0, \sigma_v^2\omega_1, \sigma_v^2\lambda_1\right)$, and their corresponding vectorial representation given by

$$\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v \sim CN_p\left(\mathbf{0}, \sigma_v^2\boldsymbol{\Omega}_v, \sigma_v^2\boldsymbol{\Lambda}_v\right), \tag{4.3}$$

with $\boldsymbol{\Omega}_v = \text{diag}\left(\left(1-\psi_1^v\right)\omega_0 + \psi_1^v\omega_1, \ldots, \left(1-\psi_p^v\right)\omega_0 + \psi_p^v\omega_1\right)$,
$\boldsymbol{\Lambda}_v = \text{diag}\left(\left(1-\psi_1^v\right)\lambda_0 + \psi_1^v\lambda_1, \ldots, \left(1-\psi_p^v\right)\lambda_0 + \psi_p^v\lambda_1\right)$ and $\boldsymbol{\psi}^v = \left[\psi_1^v, \ldots, \psi_p^v\right]$.

In the data analyses presented below we take $g_0\left(\gamma_j^v\right) = CN_1\left(0, 2v_0\sigma_v^2, 0\right)$ and $g\left(\gamma_j^v\right) = CN_1\left(0, 2v_1\sigma_v^2, 0\right)$, with parameters $0 < v_0 < v_1$, and with smaller values of v_0 favoring the detection of even weakly activated voxels. As shown in Section 4.2,

this prior structure leads to a closed form complex-valued EMVS algorithm, referred to as C-EMVS here, that allows for fast identification of active voxels.

We complete the prior specification taking $\sigma_v^2 \sim IG\left(a_\sigma, b_\sigma\right)$ and $\psi_j^v \sim Bernoulli\left(\theta_j\right)$ with $\theta_j \sim Beta\left(a_\theta, b_\theta\right)$, for all $j = 1, \ldots, p$ and $a_\sigma, b_\sigma, a_\theta, b_\theta$ constants. In particular, as discussed in the examples, we consider $a_\sigma = b_\sigma = 1/2$ and values of v_0, v_1, a_θ and b_θ selected following guidelines similar to those provided in Rockova & George (2014) and Wang et al. (2015). This prior structure relates voxels through the common probability that the binary variables for a given task j are equal to one, i.e., $Pr\left(\psi_j^v = 1 \mid \theta_j\right) = \theta_j$, for all the voxels $v = 1, \ldots, V$.

4.2 Posterior inference

We summarize the algorithms for posterior inference below. We first describe a complex-valued EMVS algorithm, C-EMVS, that leads to fast detection of active sites under the Bayesian model. A similar EMVS algorithm can be derived for magnitude-only models. We then provide a Markov chain Monte Carlo (MCMC) scheme that allows us to obtain full posterior inference. The simulations and experimental data analyzed in Sections 4.3 and 4.4 focus on the performance of the complex-valued and magnitude-only EMVS algorithms, as full MCMC is usually not computationally efficient for the analysis of large-dimensional voxel-level fMRI and CV-fMRI.

4.2.1 A C-EMVS algorithm for fast posterior computations

Rockova & George (2014) proposed an expectation-maximization approach to Bayesian variable selection (EMVS) that takes advantage of the continuity of the spike distribution to produce rapidly computable closed form expressions. Here we develop an EMVS-based approach to posterior computation that combines the linear and complex-valued Gaussian structure in (4.1), the complex-valued spike-and-slab prior for $\boldsymbol{\gamma}^v$ in (4.3), and the priors for the remaining model parameters described above. More specifically, we now summarize the steps of the C-EMVS algorithm for the simplest model specification considered in the simulation studies presented in Section 4.3 (algorithms for general models are detailed in Appendix A). This model is given by

$$\begin{aligned}
\mathbf{y}^v &= \mathbf{X}\boldsymbol{\gamma}^v + \boldsymbol{\eta}^v, \;\; \boldsymbol{\eta}^v \sim CN_T\left(\mathbf{0}, 2\sigma_v^2\mathbf{I}, \mathbf{0}\right), \\
\gamma_j^v \mid \psi_j^v &\sim \left(1-\psi_j^v\right) CN_1\left(0, 2v_0\sigma_v^2, 0\right) + \psi_j^v CN_1\left(0, 2v_1\sigma_v^2, 0\right), \;\; j = 1, \ldots, p, \\
\sigma_v^2 &\sim IG\left(a_\sigma, b_\sigma\right), \quad \psi_j^v \mid \theta_j \sim Bernoulli\left(\theta_j\right), \quad \theta_j \sim Beta\left(a_\theta, b_\theta\right).
\end{aligned} \tag{4.4}$$

Note that, for each task j, model (4.4) relates voxels through the common probability that the binary variables that specify the activation at the voxel-level for such task are equal to one, i.e., $Pr\left(\psi_j^v = 1 \mid \theta_j\right) = \theta_j$ for all voxels $v = 1, \ldots, V$ and each task $j = 1, \ldots, p$. Letting $\boldsymbol{\gamma} = \left[\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^V\right]$, $\boldsymbol{\psi} = \left[\boldsymbol{\psi}^1, \ldots, \boldsymbol{\psi}^V\right]$, with $\boldsymbol{\psi}^v = \left(\psi_1^v, \ldots, \psi_p^v\right)'$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$, $\boldsymbol{\sigma}^2 = \left[\sigma_1^2, \ldots, \sigma_V^2\right]$, and $\mathbf{y} = \left[\mathbf{y}^1, \ldots, \mathbf{y}^V\right]$, we find that the full

posterior density is given by

$$\begin{aligned}\pi\left(\boldsymbol{\gamma},\boldsymbol{\psi},\boldsymbol{\theta},\boldsymbol{\sigma}^2 \mid \mathbf{y}\right) &\propto \prod_{v=1}^{V}\left[f\left(\mathbf{y}^v \mid \boldsymbol{\gamma}^v,\sigma_v^2\right)\pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v,\sigma_v^2\right)\pi\left(\boldsymbol{\psi}^v \mid \boldsymbol{\theta}\right)\pi\left(\sigma_v^2\right)\right]\pi\left(\boldsymbol{\theta}\right) \\ &\propto \prod_{v=1}^{V}\left[CN_T\left(\mathbf{y}^v \mid \mathbf{X}\boldsymbol{\gamma}^v, 2\sigma_v^2\mathbf{I},\mathbf{0}\right)CN_p\left(\boldsymbol{\gamma}^v \mid \mathbf{0},\sigma_v^2\boldsymbol{\Omega}_v,\mathbf{0}\right)\right] \\ &\quad\times\prod_{v=1}^{V}\left[\pi\left(\sigma_v^2\right)\prod_{j=1}^{p}Bernoulli\left(\psi_j^v \mid \theta_j\right)\right]\prod_{j=1}^{p}Beta\left(\theta_j \mid a_\theta,b_\theta\right),\end{aligned} \tag{4.5}$$

where $\boldsymbol{\Omega}_v = 2\times\operatorname{diag}\left(\left(1-\psi_1^v\right)v_0+\psi_1^v v_1,\ldots,\left(1-\psi_p^v\right)v_0+\psi_p^v v_1\right)$.

An EM algorithm for maximizing the full posterior $\pi\left(\boldsymbol{\gamma},\boldsymbol{\theta},\boldsymbol{\sigma}^2 \mid \mathbf{y}\right)$ for this complex-valued model, referred to as C-EMVS, is derived by iteratively maximizing the objective function

$$Q\left(\boldsymbol{\gamma},\boldsymbol{\theta},\boldsymbol{\sigma}^2 \mid \boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\boldsymbol{\sigma}^{2,(l)}\right) = \mathrm{E}_{\boldsymbol{\psi}|\cdot}\left[\log\pi\left(\boldsymbol{\gamma},\boldsymbol{\psi},\boldsymbol{\theta},\boldsymbol{\sigma}^2 \mid \mathbf{y}\right)\middle|\boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\boldsymbol{\sigma}^{2,(l)},\mathbf{y}\right],$$

at iteration $l+1$, where $\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(\cdot\right) = \mathrm{E}_{\boldsymbol{\psi}|\boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\boldsymbol{\sigma}^{2,(l)},\mathbf{y}}\left(\cdot\right)$. Note that at iteration $l+1$, the function $Q\left(\cdot\right)$ uses the maxima found at iteration l. Given the form of the log posterior in this case we can write

$$Q\left(\boldsymbol{\gamma},\boldsymbol{\theta},\boldsymbol{\sigma}^2 \mid \boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\boldsymbol{\sigma}^{2,(l)}\right) = Q_1\left(\boldsymbol{\gamma},\boldsymbol{\sigma}^2 \mid \boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\boldsymbol{\sigma}^{2,(l)}\right) + Q_2\left(\boldsymbol{\theta} \mid \boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\boldsymbol{\sigma}^{2,(l)}\right) + K_Q, \tag{4.6}$$

with $Q_1\left(\boldsymbol{\gamma},\boldsymbol{\sigma}^2 \mid \boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\boldsymbol{\sigma}^{2,(l)}\right) = \sum_{v=1}^{V}Q_1^v\left(\boldsymbol{\gamma}^v,\sigma_v^2 \mid \boldsymbol{\gamma}^{v,(l)},\boldsymbol{\theta}^{(l)},\sigma_v^{2,(l)}\right)$ and K_Q a constant. For the E-step, we compute the conditional expectations $\mathrm{E}_{\psi^v|\cdot}[\psi_j^v]$ and $\mathrm{E}_{\psi^v|\cdot}\left[\frac{1}{\left(1-\psi_j^v\right)v_0+\psi_j^v v_1}\right]$. The M-step solves for $\left(\boldsymbol{\gamma}^{(l+1)},\boldsymbol{\sigma}^{2,(l+1)}\right)$ and $\boldsymbol{\theta}^{(l+1)}$ by maximizing Q_1^v for $v=1,\ldots,V$ and Q_2 in (4.6). The complete details for this C-EMVS

algorithm, as well as those for algorithms under more general complex-valued priors (e.g., non-circular priors) can be found in Appendix A.

The C-EMVS algorithm is iterated until $\|\boldsymbol{\gamma}^{(l)} - \boldsymbol{\gamma}^{(l-1)}\| < \epsilon$, $\|\boldsymbol{\theta}^{(l)} - \boldsymbol{\theta}^{(l-1)}\| < \epsilon$ and $\|\boldsymbol{\sigma}^{2,(l)} - \boldsymbol{\sigma}^{2,(l-1)}\| < \epsilon$, with ϵ small. In the analyses of simulated and human experimental data presented in Sections 4.3 and 4.4 we use $\epsilon = 10^{-3}$. We assess convergence by monitoring that the log-posterior distribution increases at each step of the algorithm. Once the EM algorithm converges, we obtain estimated posterior modes $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\sigma}}^2$, and $\hat{\boldsymbol{\theta}}$. Then, for each voxel we compute $Pr\left(\psi_j^v = 1 \mid \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\sigma}}^2, \mathbf{y}\right)$, and we label a given voxel v active for task j if $Pr\left(\psi_j^v = 1 \mid \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\sigma}}^2, \mathbf{y}\right) > \delta$, where δ is a fixed threshold value. This is equivalent to saying that a voxel is active if its corresponding strength is greater than some real-valued threshold $\gamma_j^{*,v}$, i.e., $|\hat{\gamma}_j^v| > \gamma_j^{*,v}$. A common choice of δ is 0.5, which leads to a local version of the median probability model of Barbieri & Berger (2004). Some researchers in the fMRI community suggest using $\delta = 0.8722$ for magnitude-only models. Smith & Fahrmeir (2007) gives a clear description of the motivation for this threshold value in the context of a Bayesian spatial model. Given that our models do not explicitly incorporate a spatial structure we use $\delta = 0.5$ in the following analyses. A further alternative that could be considered within a Bayesian decision theoretic framework is to choose the threshold by minimizing a well defined loss function, or via Bayesian false discovery rates (see, e.g., Müller et al., 2006 and Sun et al., 2015).

Finally note that, if desired, the algorithm can also be implemented for the

real-valued version of the model in (4.4) given by

$$\begin{pmatrix} \mathbf{y}_{Re}^{v} \\ \mathbf{y}_{Im}^{v} \end{pmatrix} = \mathbf{X}^{r}\boldsymbol{\gamma}_{r}^{v} + \boldsymbol{\eta}_{r}^{v}, \quad \boldsymbol{\eta}_{r}^{v} \sim N_{2T}\left(\mathbf{0}, \sigma_{v}^{2}\mathbf{I}_{2T}\right),$$

$$\boldsymbol{\gamma}_{r}^{v} = \begin{pmatrix} \boldsymbol{\gamma}_{Re}^{v} \\ \boldsymbol{\gamma}_{Im}^{v} \end{pmatrix} \sim N_{2p}\left(\mathbf{0}, \sigma_{v}^{2}\begin{pmatrix} \boldsymbol{\Sigma}_{Re,Re}\left(\boldsymbol{\psi}^{v}\right) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{Im,Im}\left(\boldsymbol{\psi}^{v}\right) \end{pmatrix}\right),$$

and the same priors on σ_v^2, ψ_j^v and θ_j specified above.

4.2.2 Posterior inference via Markov chain Monte Carlo

Full posterior inference can be obtained via MCMC. Similar to the C-EMVS case described above, we generalize the Stochastic Search Variable Selection algorithm (SSVS) proposed by George & McCulloch (1993) to the complex-valued domain. Suppose we have a simplified complex-valued model such as (4.4) except that we now use a "non-conjugate" version of the spike-and-slab prior on $\boldsymbol{\gamma}^v$, i.e., $\gamma_j^v \mid \psi_j^v \sim \left(1-\psi_j^v\right) CN_1\left(0, 2v_0, 0\right) + \psi_j^v CN_1\left(0, 2v_1, 0\right)$, $j = 1, \ldots, p$. The general vectorized form of this prior can be written as $\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v \sim CN_p\left(\mathbf{0}, \boldsymbol{\Omega}_v, \mathbf{0}\right)$, with $\boldsymbol{\Omega}_v = 2 \times \text{diag}\left[\left(1-\psi_1^v\right)v_0 + \psi_1^v v_1, \ldots, \left(1-\psi_p^v\right)v_0 + \psi_p^v v_1\right]$. Then, the posterior full conditional distributions for a Gibbs sampling scheme can be derived as follows:

- For each v, $v = 1, \ldots, V$, $\boldsymbol{\gamma}^v \mid \mathbf{y}^v, \sigma_v^2, \boldsymbol{\psi}^v \sim CN_p\left(\boldsymbol{\mu}_{\boldsymbol{\gamma}}^v, \boldsymbol{\Omega}_{pos}^v, \mathbf{0}\right)$, with $\boldsymbol{\Omega}_{pos}^v = \left(2^{-1}\sigma_v^{-2}\mathbf{X}'\mathbf{X} + \boldsymbol{\Omega}_v^{-1}\right)^{-1}$, and $\boldsymbol{\mu}_{\boldsymbol{\gamma}}^v = \boldsymbol{\Omega}_{pos}^v \mathbf{X}'\mathbf{y}^v/\sigma_v^2$.

- $\sigma_v^2 \mid \mathbf{y}^v, \boldsymbol{\gamma}^v \sim IG\left(a_\sigma^{v,pos}, b_\sigma^{v,pos}\right)$, with $a_\sigma^{v,pos} = T + a_\sigma$ and $b_\sigma^{v,pos} = \|\mathbf{y}^v - \mathbf{X}\boldsymbol{\gamma}^v\|^2/2 + b_\sigma$.

- $Pr\left(\psi_j^v = 1 \mid \mathbf{y}^v, \boldsymbol{\gamma}^v, \sigma^2, \theta, \boldsymbol{\psi}_{-j}^v\right) = \frac{c_j^v}{c_j^v + e_j^v}$, with

  $$c_j^v = \pi\left(\boldsymbol{\gamma}_j^v \mid \mathbf{y}^v, \psi_j^v = 1, \boldsymbol{\psi}_{-j}^v\right) \times \theta_j \quad \text{and} \quad e_j^v = \pi\left(\boldsymbol{\gamma}_j^v \mid \mathbf{y}^v, \psi_j^v = 0, \boldsymbol{\psi}_{-j}^v\right) \times (1 - \theta_j).$$

  Here $\pi\left(\boldsymbol{\gamma}_j^v \mid \mathbf{y}^v, \psi_j^v = 1, \boldsymbol{\psi}_{-j}^v\right)$ and $\pi\left(\boldsymbol{\gamma}_j^v \mid \mathbf{y}^v, \psi_j^v = 0, \boldsymbol{\psi}_{-j}^v\right)$ are complex-normal densities (see Appendix A for details).

- For each $j = 1, \ldots, p$, $\theta_j \mid \mathbf{y}, \boldsymbol{\psi}^v \sim Beta\left(\sum_{v=1}^V \psi_j^v + a_\theta, V - \sum_{v=1}^V \psi_j^v + b_\theta\right)$.

In order to decide whether a voxel v is active or not after MCMC convergence is achieved, we look at the posterior probability of $\psi_j^v = 1$, for each task-related BOLD signal $j = 1, \ldots, p$. A detailed derivation of general complex-valued SSVS algorithm and the corresponding full conditional distributions above can be found in Appendix A.

4.3 Simulation studies

We show the performance of the proposed complex-valued variable selection methods for detecting activation in two simulation studies. The first study compares the C-EMVS algorithm to computationally fast alternatives that are often used in practice, such as lasso and adaptive lasso (Tibshirani, 1996; Zou, 2006). We also compare the results obtained by the proposed complex-valued model and priors via the C-EMVS algorithm with those obtained using a magnitude-only Bayesian model with the real-valued priors in Rockova & George (2014). The magnitude-only voxel time series courses are obtained by taking the moduli of the CV-fMRI signals at each

voxel. The second study considers a physically realistic simulated CV-fMRI dataset.

4.3.1 Simulation study I

We simulated 20 datasets consisting of 48×48 CV-fMRI slices with a constant baseline signal and a single expected BOLD signal (i.e., $p = 1$) resulting from the convolution of a stimulus indicator function and the canonical HRF. Three activation regions were simulated using the function `specifyregion` in the `R` package `neuRosim` (Welvaert et al., 2011). More specifically, for $v = 1, \ldots, 48 \times 48$, and $t = 1, \ldots, 200$, the time series for each voxel v were simulated as follows:

$$\begin{aligned} y^v_{t,Re} &= (\beta_0 + \beta_1 f_v z_t) \cos(\alpha_0) + \eta^v_{t,Re}, \quad \eta^v_{t,Re} \sim N\left(0, \sigma^2\right) \\ y^v_{t,Im} &= (\beta_0 + \beta_1 f_v z_t) \sin(\alpha_0) + \eta^v_{t,Im}, \quad \eta^v_{t,Im} \sim N\left(0, \sigma^2\right) \end{aligned} \tag{4.7}$$

where f_v is the BOLD signal strength or intensity rate of voxel v, with $f_v = 0$ if voxel v is non-active and $f_v \neq 0$ if voxel v is in an active region. The values of f_v for active voxels were specified using the argument `fading` in the function `specifyregion` in `neurosim`. Here, the fading of the expected BOLD signal decays exponentially depending on the distance of the active voxel v with coordinates (i, j), to the center of the active region with coordinates (i', j'), i.e., the fading for voxel v is given by

$$f_v(i, j) = \frac{1}{4}\left\{2 \cdot \exp\left[-\left(\left(i - i'\right)^2 + (j - j')^2\right) \cdot \varrho\right] + 2\right\},$$

where ϱ is the decay rate in $[0, 1]$ with 0 and 1 corresponding, respectively, to no decay and to the strongest decay. z_t in (4.7) is the BOLD signal given by the convolution

of the canonical HRF, denoted as h_t, and the stimulus indicator function s_t, i.e., $z_t = h_t \otimes s_t$.

We used $\alpha_0 = \pi/4$ and different values of β_0 and β_1 to examine the performance of the proposed complex-valued models using the C-EMVS algorithm for posterior computations. These were chosen to set specific values of the SNR and the contrast-to-noise ratio (CNR) as defined in Rowe & Logan (2004), with SNR $= \beta_0/\sigma$ and CNR $= \beta_1/\sigma$. Note that active voxels have different CNRs given by $\text{CNR}_v = (\beta_1 f_v)/\sigma$, with $\text{CNR}_v \leq \text{CNR}$ for all v, as $f_v \in [0, 1]$. Hence, the largest CNR for active voxels is β_1/σ, computed using no fading, while the smallest CNR is $\beta_1 f_{min}/\sigma$, where $f_{min} = \min_{\{v \in A\}} f_v$ and A is the set of active voxels. The average CNR is $\sum_{\{v \in A\}} \beta_1 f_v/|A|$. In this simulation we used $f_{min} \approx 0.50$ and $\sum_{\{v \in A\}} f_v/|A| \approx 0.71$, with $|A| = 103$ active voxels, which accounts for 4.47% of all voxels.

The top left plot in Figure 4.1 shows the experimental block design, with $s_t = 1$ if the stimulus is on and $s_t = 0$ otherwise. It consists of five epochs of 20 s on and 20 s off with an observation interval of 1. The resulting BOLD signal z_t is shown in the bottom left plot. The right plot displays the active regions with the corresponding f_v values. The three active regions are centered at the coordinates $(20, 20)$, $(30, 30)$ and $(40, 10)$, with radius arguments 3, 2, 1, and fading arguments 0.5, 0.01, and 0.3, respectively for each region. The bottom-right region is a square and the other two are circles.

Four different SNRs, 0.5, 1, 5, and 10, and three different CNRs, 0.5, 1, and 1.5, were considered, resulting in 12 different SNR-CNR data types. These are num-

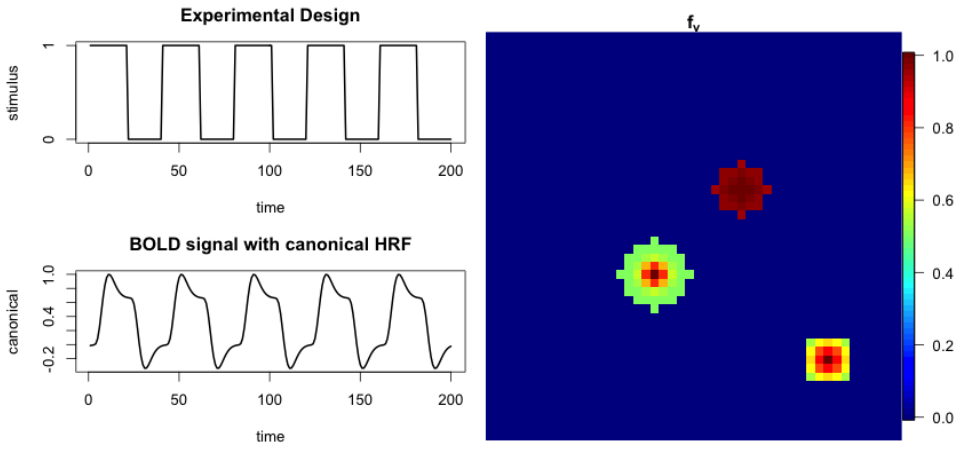


Figure 4.1: Left: Block experimental design (top); expected BOLD signal obtained from convolving the stimulus indicator signal with the canonical hemodynamic function (bottom). Right: activation regions and f_v values for active voxels.

SNR	0.5			1			5			10		
CNR	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
Data Type	1	2	3	4	5	6	7	8	9	10	11	12

Table 4.1: Twelve data types and thier corresponding SNR and CNR.

bered as shown in Table 4.1. We generated 20 simulated datasets for each SNR-CNR data type and computed classification performance measures (sensitivity, specificity, precision and accuracy) to examine how well our algorithm and other methods perform in the different scenarios.

Four methods are compared in this simulation study, the proposed Bayesian complex-valued model using the C-EMVS algorithm for posterior computations, (referred to as CV in the results below), the Bayesian magnitude-only model with the EMVS algorithm (MO), and the lasso (LA) and adaptive lasso (ALA), both for magnitude-only data.

The Bayesian complex-valued model used here has the form

$$y_t^v = \gamma_1^* + \gamma_2^{*,v} x_t + \eta_t^v, \;\; \eta_t^v \sim CN_1\left(0, 2\sigma^2, 0\right),$$

with γ_1^* a baseline parameter and $\gamma_2^{*,v}$ the complex-valued activation parameters for each voxel and $x_t = z_t$. For the baseline parameter we use a prior of the form $\pi\left(\gamma_1^*\right) \propto 1$. For the activation parameters and the remaining model parameters we used the following priors:

$$\begin{aligned} \gamma_2^{*,v} \mid \psi^v &\sim \left(1-\psi^v\right) CN_1\left(0, 2v_0\sigma^2, 0\right) + \psi^v CN_1\left(0, 2v_1\sigma^2, 0\right), \\ \sigma^2 &\sim IG\left(1/2, 1/2\right), \quad \psi^v \mid \theta \sim Bernoulli\left(\theta\right), \quad \theta \sim Beta\left(1,1\right). \end{aligned} \tag{4.8}$$

The baseline parameter was integrated out before proceeding with the C-EMVS or MCMC algorithms for posterior inference and detection of active sites, so we used the algorithms outlined in Section 4.2 and detailed in Appendix A.

We also consider a Bayesian model for the magnitude-only data. The magnitude-only time courses are obtained as $y_{t,Mag}^v = \sqrt{\left(y_{t,Re}^v\right)^2 + \left(y_{t,Im}^v\right)^2}$. The MO model used to analyze these data is essentially the same as the CV model used for the complex-valued data, except that the linear model is now real-valued and the priors on the

regression coefficients are real-valued Gaussian spike-and-slab priors. This is

$$y^v_{t,Mag} = \gamma^*_{Mag,1} + \gamma^{*,v}_{Mag,2} x_t + \eta^v_t, \;\; \eta^v_t \sim N\left(0, \sigma^2\right),$$

$$\gamma^{*,v}_{Mag,2} \mid \psi^v_{Mag} \sim \left(1 - \psi^v_{Mag}\right) N_1\left(0, v_0\sigma^2\right) + \psi^v_{Mag} N_1\left(0, v_1\sigma^2\right),$$

$$\sigma^2 \sim IG\left(1/2, 1/2\right), \;\; \psi^v_{Mag} \mid \theta \sim Bernoulli\left(\theta\right), \;\; \theta \sim Beta\left(1, 1\right),$$

and $\pi\left(\gamma^*_{Mag,1}\right) \propto 1$.

The tuning parameters in the Bayesian CV and MO models above, v_0 and v_1, are chosen as suggested in Rockova & George (2014) and Wang et al. (2015). More specifically, we fix v_1, taking $v_1 = 1$ and choose the optimal v_0 in each case, denoted as v_0^{CV} and v_0^{MO}, for the CV and MO models respectively, by maximizing the marginal posterior $\pi_0\left(\boldsymbol{\psi} \mid \mathbf{y}\right)$ that evaluates $\boldsymbol{\psi}$ according to the submodel that contains only those variables for which $\psi^v_j = 1$. This marginal can be derived in closed form up to a normalizing constant. From our experience with the real and simulated datasets analyzed here, the optimal v_0 takes values around $1/\sqrt{100Tp}$ and usually lies in the interval $\left(1/\sqrt{1000Tp}, 1/\sqrt{10Tp}\right)$, where p is the number of tasks. In this simulation we only have one task so $p = 1$.

Finally, we also applied the lasso (LA) and adaptive lasso (ALA) methods (Tibshirani, 1996; Zou, 2006) to the magnitude-only data. Both LA and ALA use a regularization parameter and ALA uses additional weights to allow for different penalizations in the regression coefficients (the $\gamma^{*,v}_{Mag,2}$ parameters in our case). The regularization parameter was chosen using a five fold cross validation approach and the weights in the ALA were set to $1/|\hat{\gamma}^{*,v}_{Mag,2}|$, where $\hat{\gamma}^{*,v}_{Mag,2}$ is the ordinary least square

estimator of $\gamma^{*,v}_{Mag,2}$. LA and ALA were implemented using the `R` package `glmnet` (Friedman et al., 2010).

The resulting average performance measures over the 20 simulated datasets for the four different methods are summarized in Figure 4.2. Note that this simulation contains 2201 non-active voxels out of a total 2304 voxels, so any model can achieve 95.53% accuracy by simply classifying all voxels as non-active. Hence the accuracy subfigure is plotted from 0.95 to 1 for clear comparison. Similarly, the specificity is plotted from 0.997 to 1.

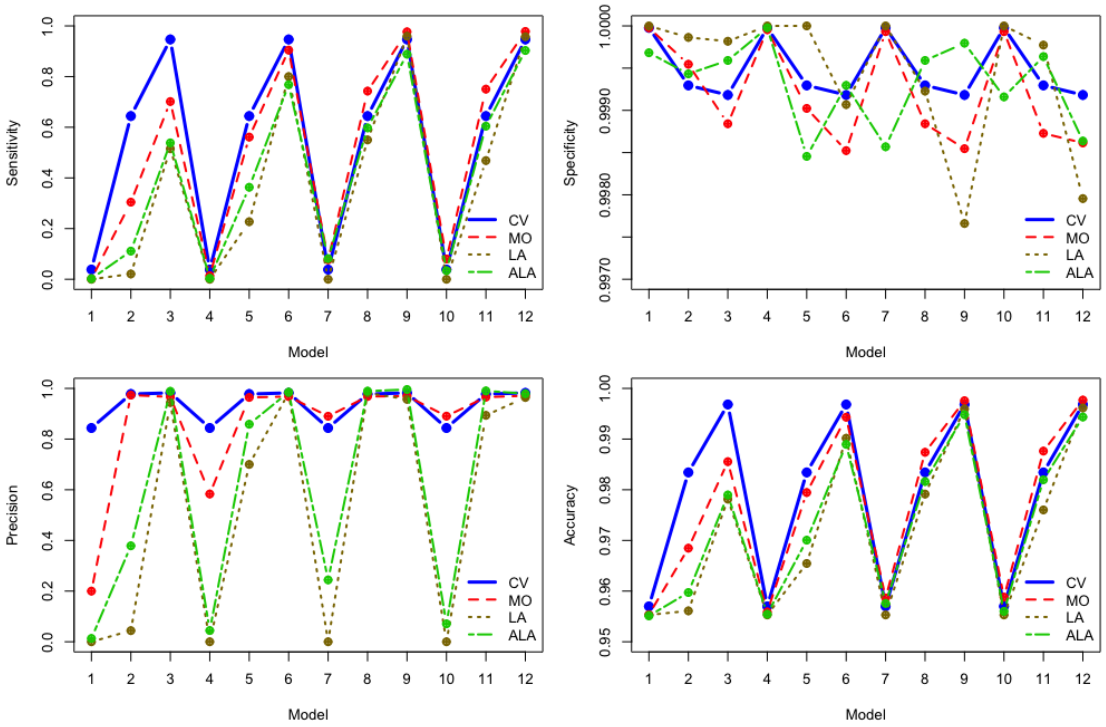
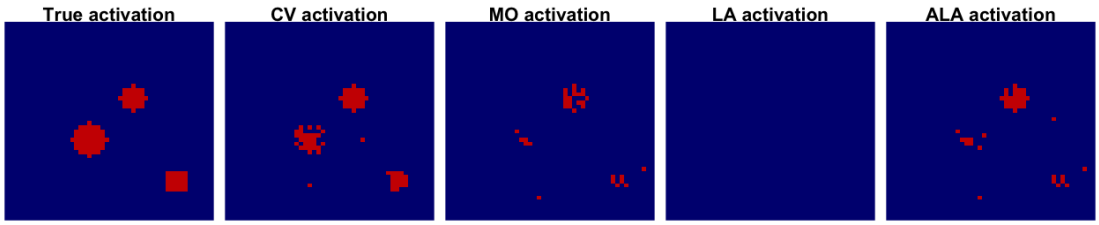


Figure 4.2: Sensitivity (top-left), specificity (top-right), precision (bottom-left) and accuracy (bottom-right) for four models: complex-valued EM (CV; blue, solid), magnitude-only EM (MO; red, dash), Lasso (LA; brown, dotted), and Adaptive Lasso (ALA; green, dash-dotted).

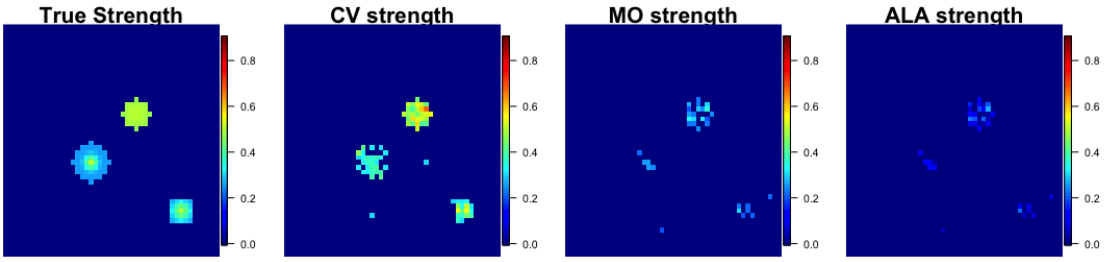
First, we see that both Bayesian variable selection approaches, the one for

the CV-fMRI and the one for magnitude-only data (MO), dominate the traditional lasso (LA) and adaptive lasso (ALA) for magnitude-only data in terms of sensitivity (power), precision and accuracy. The Bayesian approaches are able to eliminate most of the false positives by borrowing strength across voxels via the common probability of activation parameter θ. The Bayesian CV and MO methods are comparable to lasso and adaptive lasso in terms of specificity, while the first provide a more complete inferential analysis. The main advantage of the Bayesian CV model with respect to the Bayesian MO model is that the CV model significantly detects more true positives than the MO when the SNR is small, which leads to higher sensitivity, precision and accuracy. When the SNR is fairly large, using the information provided only by the magnitude leads to good activation results in these simulated scenarios. In fact, the MO model even has a slightly larger sensitivity than the CV model when the SNR is 5 or 10. On the other hand, the CV model leads to higher specificity and precision than the MO model even when the SNR is 5 or 10. Moreover, the performance of the CV model is very consistent across different SNRs. Hence, when the CV-fMRI data are recorded under small SNRs or when researchers are uncertain about the magnitude of the SNR in their data, the CV model stands out as the best option among the models considered here. Given that improved MRI technology allows for improved spatial resolution and therefore reduces SNR, we would expect that complex-valued models will become an essential tool for detecting active sites in CV-fMRI data.

Figure 4.3 shows the true activation and strength maps for one of the 20 simulated datasets with SNR = 0.5 and CNR = 1 along with the estimated activation



(a) Activation maps



(b) Strength maps

Figure 4.3: Activation and strength maps for a simulated dataset with SNR = 0.5 and CNR = 1. (a) Activation maps showing the true active sites, and the activation results obtained from C-EMVS, MO-EMVS, Lasso and Adaptive Lasso. Activated sites are colored in red. (b) Strength maps: true strength and estimated strengths from C-EMVS, MO-EMVS, Lasso and Adaptive Lasso.

and strength maps (only for sites labeled as active) obtained from the C-EMVS (CV), the magnitude-only EMVS (MO), and adaptive lasso (ALA). The strength maps for lasso are not shown, as lasso detected no active sites. Both activation maps for the complex-valued and magnitude-only EMVS display activation levels that result from setting $v_1 = 1$ and choosing the optimal values of v_0 for each method as discussed above. For this dataset and with our prior distribution settings, we found that the optimal values were $v_0^{CV} = 0.0071$ and $v_0^{MO} = 0.0056$. The C-EMVS approach clearly outperforms all the other approaches: it has higher power for detecting active voxels while simultaneously controlling for false positives, and also leads to more accurate estimation of the activation strength (note that MO and ALA clearly underestimate the strength). In relation to this point, we computed the mean squared errors (MSEs)

for this simulated dataset under the C-EMVS, MO and ALA approaches for voxels that are labeled as active for at least one of the 3 methods and found that the MSEs values were, respectively, 0.0080, 0.0084, and 0.1162. The complex-valued model also leads to more accurate inference for σ. Magnitude-only models underestimate σ when the SNR is small as a consequence of the fact that the MO error distribution is truly Ricean at low SNRs. This can lead to an increase of false positives when detecting activation (in fact, we can see that the specificity values obtained with the complex-valued model are generally higher than those obtained with magnitude-only model as shown in Figure 4.2). For example, for a dataset generated under a true value of $\sigma = 0.5$, when SNR $= 0.5$, we found $\hat{\sigma}_{CV} = 0.497$, while $\hat{\sigma}_{MO} = 0.346$. In order to obtain better estimates of σ with MO models we need to considerably increase the SNR. For instance, for a simulated dataset with SNR $= 10$, we obtained $\hat{\sigma}_{MO} = 0.495$ which is closer to the true value 0.5. These results are consistent with the findings of Rowe (2005b).

Finally, we also implemented the MCMC sampling approach outlined in Section 4.2 and detailed in Appendix A to achieve full posterior inference for the complex-valued models. We obtained similar results to those from the C-EMVS algorithm in terms of the number of active sites and the strength of those sites, but we highlight that, in addition, the MCMC approach allows us to compute uncertainty measures related to activation strength and any other functions of the model parameters. For instance, Figure 4.4 shows posterior mean strength maps and 95% posterior credibility strength maps for a single dataset obtained from the complex-valued model. As seen in this figure, the posterior mean estimates for the strength are similar to those

obtained via the C-EMVS algorithm but the MCMC-based posterior credibility maps provide additional information about the strength maps. We see that, in general, there is less uncertainty about activation strength for voxels located in region centered at (30,30) than for voxels located in the region centered at (40,10). This makes sense given the true strength maps used to generate the simulated data (see Figure 4.3). In cases where this Gibbs sampling scheme is not computationally feasible (e.g., when several large-dimensional images for multiple subjects need to be analyzed) one could consider a hybrid approach that, say, uses the C-EMVS method to determine which sites are active and then uses the Gibbs sampling scheme only on regions of the brain that present active sites to obtain posterior uncertainty measures on strength maps and/or activation maps for those regions only. Alternative methods based on obtaining approximate inference via variational Bayes could also be considered (see, e.g., Yu et al., 2016).

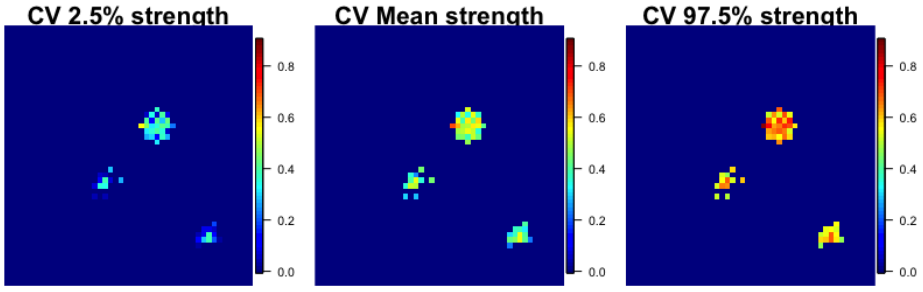


Figure 4.4: Strength maps for a simulated dataset with SNR = 0.5 and CNR = 1 obtained from a complex-valued model via MCMC. Left: 2.5% quantile map; Middle: Posterior mean map; Right: 97.5% quantile map.

4.3.2 Additional structure: Temporal correlation

We also analyzed synthetic CV-fMRI data simulated under (4.7) but with errors following an autoregressive structure of order one, i.e., $\eta^v_{t,Re} = \varphi\eta^v_{t-1,Re} + \zeta^v_{t,Re}$, with $\zeta^v_{t,Re}$ independent Gaussian for all t, $\zeta^v_{t,Re} \sim N\left(0, \sigma^2\right)$, and $\eta^v_{t,Im} = \varphi\eta^v_{t-1,Im} + \zeta^v_{t,Im}$, with $\zeta^v_{t,Im}$ also independent Gaussian for all t, $\zeta^v_{t,Im} \sim N\left(0, \sigma^2\right)$ and $\varphi \in [0, 1)$ the AR coefficient. We considered values of φ ranging from 0.1 to 0.9, and the same 12 SNR-CNR scenarios described in the previous simulation, with $\sigma^2 = 0.25$. We analyzed these data using two versions of the model $y^v_t = \gamma^*_1 + \gamma^{*,v}_2 x_t + \eta^v_t$: one version with η^v_t i.i.d. complex normal, and another version with η^v_t following a complex-valued AR(1) structure in $\eta^v_t = \eta^v_{t,Re} + i\eta^v_{t,Im}$ as described above. Figure 4.5 displays the sensitivity, specificity, precision and accuracy for the two versions of the CV model (independent and autoregressive errors) and 2 types of data (AR errors with $\varphi = 0.5$ and $\varphi = 0.9$). Overall we find that the larger the value of φ the harder it is to detect active sites, particularly for small SNR and CNR. This makes sense, as AR(1) errors with φ close to 1 may add a temporal structure that locally resembles a linear trend and can easily hide/mask the temporal behavior that characterizes active sites due to increased variability in the observed time series. We also see that while the CV model with independent errors has higher sensitivity, it also leads to a larger number of false positives (we only have about 77% specificity for the model with independent errors while we obtain 100% specificity for the model with AR errors when $\varphi = 0.9$). Therefore, the CV model with AR errors is overall a better option in terms of specificity, precision and accuracy, particularly when φ is large.

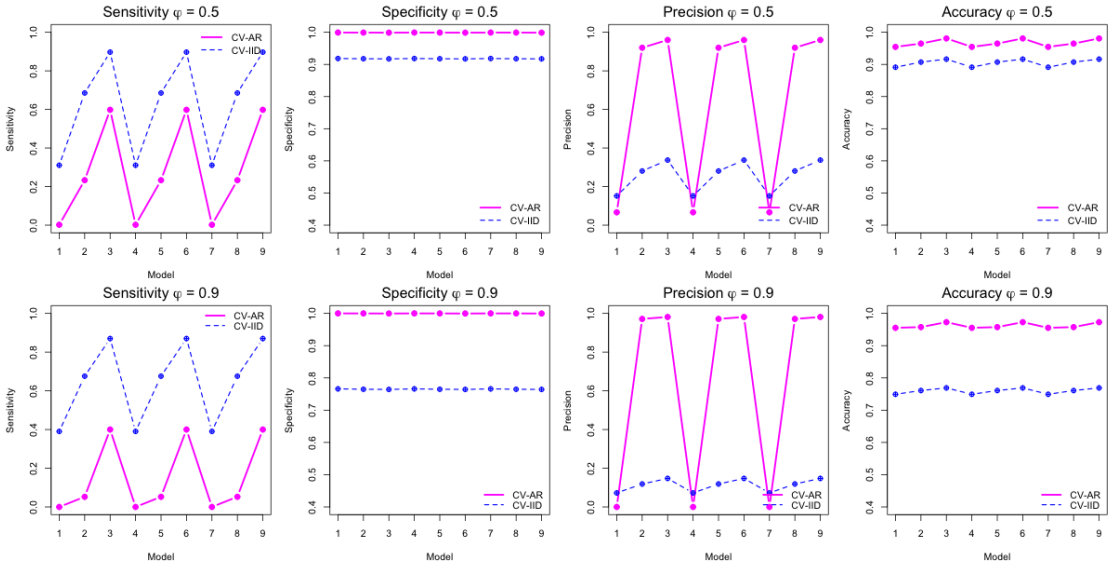


Figure 4.5: Sensitivity, specificity, precision, and accuracy plots for synthetic AR(1) CV-fMRI data with AR coefficients 0.5 (top plots) and 0.9 (bottom plots). The plots are based on results obtained from analyzing 20 datasets using models that assumed independent errors (dotted lines) and AR(1) errors (solid lines).

4.3.3 Additional structure: HRF effect and prior sensitivity analyses

We also studied the effects of the HRF choice and the prior distributions. Regarding the HRFs, we analyzed the simulated and human data presented in Sections 4.3 and 4.4 with 3 different classes of HRF functions, namely, canonical, gamma and boxcar with different choices for the parameters that define each particular class. For a given HRF we can select the optimal v_0 and then choose the HRF and corresponding v_0 that leads to the smallest mean squared error (MSE) for a particular dataset. Overall we found that the MSEs for the optimal HRFs within each class were comparable. Furthermore, the results in terms of the number and locations of the sites labeled as active were also similar across the optimal HRFs within each class.

We studied the sensitivity of our posterior results with respect to the prior

distributions. In particular, as mentioned above, we generally assume $\theta \sim Beta\,(1,1)$. In cases where a sparser structure is desired a priori, i.e., when it makes sense biologically to assume that the number of active sites is just a very small percentage of the total number of sites, priors of the form $\theta \sim Beta\,(1,b)$ with b large can be used. In this simulation study we found that the activation results were essentially the same for any prior with $b \leq 1000$. Priors with values of $b > 1000$ lead to sparser results (i.e., less active sites) in the simulated data. For the human data presented in Section 4.4 we found that we are able to detect similar numbers and locations of active sites for priors with values of $b \in [1, 100000]$. Note that choosing $b = 1000$ leads to a fairly informative prior, with about 0.09% of active sites expected a priori and rarely above 0.4% of active sites expected a priori.

Finally, we assessed the effect of using non-circular complex-normal priors on $\boldsymbol{\gamma}^v$, i.e., priors of the form $\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v \sim CN_p\left(\mathbf{0}, \sigma_v^2\boldsymbol{\Omega}_v, \sigma_v^2\boldsymbol{\Lambda}_v\right)$, with $\boldsymbol{\Lambda}_v \neq \mathbf{0}$, so that there is a non-zero correlation between the real and imaginary components of $\boldsymbol{\gamma}^v$. As expected, allowing for a correlation structure between the real and imaginary components of $\boldsymbol{\gamma}^v$ leads to improved results when such underlying structure is present in the data, i.e., having a more flexible prior that accounts for this correlation leads to higher power for detecting activation and reduces the number of false positives. On the other hand, such priors also lead to models that are more computationally costly and may potentially lead to biases in the posterior results. Therefore, we recommend the use of non-circular priors only when there is a strong indication that there is a significant correlation between the real and complex components of $\boldsymbol{\gamma}^v$, and that such

correlation structure is similar for active and non-active voxels. Alternative priors will be developed and investigated in the future but are out of the scope of this work. We now illustrate the use of non-circular priors in the analysis of simulated a dataset with high correlation among the real and imaginary components for both types of voxels, active and non-active. The data was simulated following

$$\begin{aligned} y^v_{t,Re} &= \left(\beta_0 + \beta^v_{1,Re} z_t\right) \cos\left(\alpha_0\right) + \eta^v_{t,Re}, \quad \eta^v_{t,Re} \sim N\left(0, \sigma^2\right), \\ y^v_{t,Im} &= \left(\beta_0 + \beta^v_{1,Im} z_t\right) \sin\left(\alpha_0\right) + \eta^v_{t,Im}, \quad \eta^v_{t,Im} \sim N\left(0, \sigma^2\right), \end{aligned}$$

with $\alpha_0 = \pi/4$, $\sigma^2 = 0.1$, SNR $= 0.4$, $\beta_0 = 0.8$, and the same z_t used in the previous simulation study. In addition, the parameters $\beta^v_{1,Re}$ and $\beta^v_{1,Im}$ were obtained from complex-normal distributions as follows. For active voxels we sampled $\beta^v_{1,Re} + i\beta^v_{1,Im}$ from a complex non-circular normal with mean 0.7 and covariance and relation values that lead to a correlation of 0.9 between $\beta^v_{1,Re}$ and $\beta^v_{1,Im}$. For non-active voxels we sampled $\beta^v_{1,Re} + i\beta^v_{1,Im}$ from a complex non-circular normal with mean 0 and covariance and relation values that lead to a correlation of 0.9 between $\beta^v_{1,Re}$ and $\beta^v_{1,Im}$. Note that $\eta^v_{t,Re}$ and $\eta^v_{t,Im}$ are assumed independent for all the voxels and also across time. The location of the active voxels was determined using the same activation map used in the previous simulation and displayed in the left plot of Figure 4.3 (a).

Figure 4.6 shows the results obtained from a model that uses a non-circular prior on $\boldsymbol{\gamma}^v$ that captures the induced correlation structure in these coefficients (left plot) and also shows the results obtained using a circular prior that assumes no correlation structure. Clearly, the model with a non-circular prior leads to much better

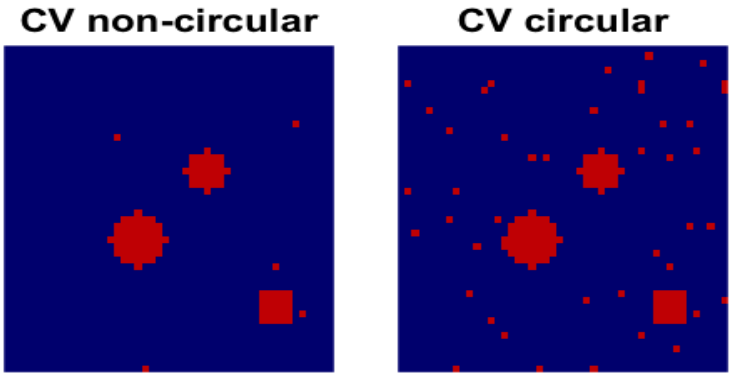


Figure 4.6: Left: Activation results obtained from a model with a non-circular prior. Right: Activation results obtained with a circular prior. The data were simulated so that the real and complex components of the activation coefficients are highly correlated.

results as it adequately identifies the active regions and leads to a much smaller number of false positives than those obtained under the model with the circular prior. The model with the non-circular prior also leads to better results in terms of estimation of activation strength and reduced MSE.

4.3.4 Simulation study II: Physically realistic simulated data

A more realistic simulated dataset was generated using a discrete version of the magnetic resonance (MR) signal equation after steady state magnetization (Karaman et al., 2015). This equation is given by

$$s(k_x, k_y \mid t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x,y) e^{-t/T_2^*(x,y)} \left(1 - e^{-TR/T_1(x,y)}\right) e^{i\Gamma_H \Delta B(x,y) t} e^{-i2\pi(k_x x + k_y y)} \mathrm{d}x\mathrm{d}y, \tag{4.9}$$

where $s(k_x, k_y \mid t)$ is the k-space location at intra slice time t, $\rho(x,y)$ is the proton spin density (PSD), $T_2^*(x,y)$ is the transverse relaxation rate (TRR), $T_1(x,y)$ is the longitudinal relaxation rate (LRR), $\Delta B(x,y)$ is the magnetic field inhomogeneity (MFI),

and Γ_H is the proton gyromagnetic ratio (Haacke et al., 1999). The k-space points in (4.9) are defined by the temporal integral of the magnetic field gradients $G_x(\cdot)$ and $G_y(\cdot)$:

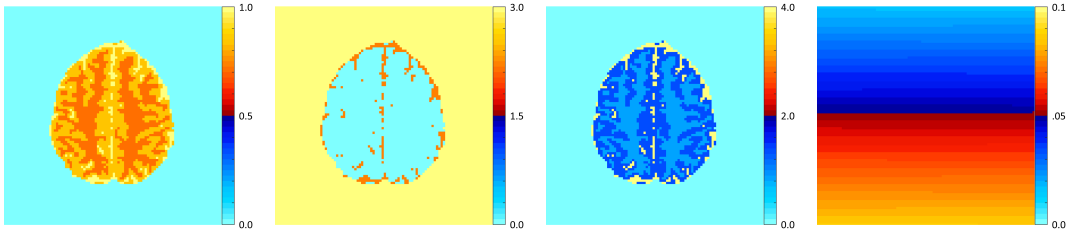
$$k_x = \frac{\Gamma_H}{2\pi}\int_0^t G_x(t')\mathrm{d}t', \qquad \text{and} \quad k_y = \frac{\Gamma_H}{2\pi}\int_0^t G_y(t')\mathrm{d}t'.$$

As input to this data generation process, 3.0 T tissue specific physical parameters (Peters et al., 2006) for the brain slice as given in Table 4.2 and displayed in Figure 4.7 were utilized. The units of measurement for T_2^* and T_1 are ms. In generating these data, a simplified version of (4.9) was used where the MFI $\Delta B(x, y)$ was not included. Without the inclusion of $\Delta B(x, y)$ MFI, the k-space array after being reconstructed yielded a real-valued image with a maximum around one. The average value in grey matter of this image was computed, and the entire image multiplied by a magnitude SNR $\mathrm{SNR_M} \times \sigma/\overline{GM}$, where $\sigma = 1$ is the standard deviation of noise added to the simulation, $\mathrm{SNR_M} = 25$ is the SNR for the simulation, and $\overline{GM} = 0.3545$ is the average grey matter value before scaling. This scaled real-valued image was utilized as the magnitude of the true images. In order to have nonzero mean phase, baseline phase as shown in Figure 4.8b was added to each tissue type according to $\alpha_{0,OB} = 0$ for outside brain, $\alpha_{0,WM} = \pi/12$ for white matter, $\alpha_{0,GM} = \pi/6$ for grey matter, and $\alpha_{0,CSF} = \pi/4$ for cerebrospinal fluid. The remaining imaging parameters were selected to mirror those of an experimental dataset with field of view $(FOV) = 240$ mm, time to echo $(TE) = 50$ ms, flip angle $(FA) = 90°$, effective echo spacing $(EESP) = 720$ μs, and bandwidth $(BW) = 125$ kHz.

The simulated data have slices of dimension 96×96 over $T = 490$ time points.

	Grey Matter	White Matter	Cerebrospinal Fluid	Outside Brain
PSD	0.83	0.71	1.00	10^{-9}
T_2^*	59.7	54.6	2200	10^{10}
T_1	1331	832	4000	10^{-6}

Table 4.2: Tissue physical parameter values.

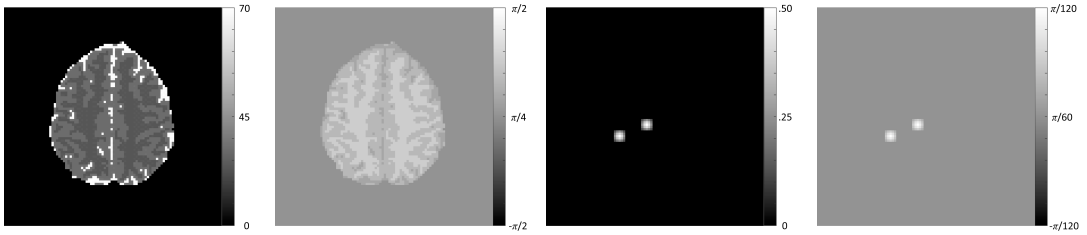


(a) PSD, $\rho(x, y)$ (b) TRR, $T_2^*(x, y)$ (c) LRR, $T_1(x, y)$ (d) Time map, $t(x, y)$

Figure 4.7: Slice physical parameters.

The true activation regions are the two 5×5 red squares shown in the left panel of Figure 4.9. Each active voxel has different intensity and the voxels near the center of the region have stronger intensities than the ones around the edges of the region. At each time point, the magnitude contrast (β_1) in Figure 4.8c was multiplied by a task response waveform and then added to the magnitude baseline (β_0) in Figure 4.8a to form the image magnitude. At each time point, the phase contrast (α_1) in Figure 4.8d was multiplied by a task response waveform then added to the phase baseline (α_0) in Figure 4.8b to form the image phase. Independent zero mean and unit variance normal noise was also added to the real and imaginary parts at each time point. In this simulation, the maximum magnitude CNR in the center of each ROI was set to $\text{CNR}_\text{M} = \beta_1/\sigma = 0.5/1$ and the maximum phase CNR in each ROI was set to $\text{CNR}_\text{P} = \alpha_1/\text{SNR}_\text{M} = (\pi/120)/25$. The contrast values ($\beta_1$ and α_1) in each ROI were then multiplied by an unnormalized Gaussian kernel with full-width-at-half-max

$(FWHM) = 4$ voxels.



(a) Magnitude baseline (b) Phase baseline (c) Magnitude contrast (d) Phase contrast

Figure 4.8: True simulated image values.

We fitted a Bayesian complex-valued (CV) model given by

$$y_t^v = \gamma_1^{*,v} + \gamma_2^{*,v} t/T + \gamma_3^{*,v} x_t + \eta_t^v, \quad \eta_t^v \sim CN_1\left(0, 2\sigma^2, 0\right), \tag{4.10}$$

with $\gamma_1^{*,v}$ and $\gamma_2^{*,v}$ baseline and trend parameters, and $\gamma_3^{*,v}$ the activation parameters for $v = 1, \ldots, V$. Here we set x_t to be the BOLD response obtained from the convolution of the experimental block design and the canonical HRF. We used priors of the form $\pi\left(\gamma_1^{*,v}\right) \propto 1$ and $\pi\left(\gamma_2^{*,v}\right) \propto 1$ for the baseline and trend parameters. For the activation parameters $\gamma_3^{*,v}$ for $v = 1, \ldots, V$ we used priors of the form $\gamma_3^{*,v} \mid \psi^v \sim (1 - \psi^v) CN_1\left(0, 2v_0\sigma^2, 0\right) + \psi^v CN_1\left(0, 2v_1\sigma^2, 0\right)$, and for σ^2, ψ^v and θ we used the priors given in (4.8). As done in the simulation study I, we also fitted a magnitude-only version of this model (MO). The posterior results for both models summarized here were obtained after integrating out the baseline and trend parameters. In both models, we chose the optimal values of v_0 by maximizing the marginal posterior $\pi_0\left(\boldsymbol{\psi} \mid \mathbf{y}\right)$ as described in the previous simulation study. In this case the optimal v_0 values were found to be $v_0^{MO} = v_0^{CV} = 0.006$.

The posterior activation maps for each model are shown in Figure 4.9. First note that both Bayesian models, the complex-valued model (CV) and the magnitude-only one (MO), perform reasonably well in terms of detecting active sites, particularly considering that the CV-fMRI data were not generated from these models and instead followed a much more complicated physically realistic model, and also considering the low SNR and CNR in this setting. The main advantage of the Bayesian CV and MO models is that their linear structure allows us to obtain posterior estimates in a computationally feasible manner that scales well with the large dimension of the images. Regarding the comparison between the complex-valued and magnitude-only models we see that, once again, the CV model shows a better performance than the MO model. The MO model produces a larger number of false positives without detecting more true positives than the CV model. Table 4.3 shows the performance measures for both models. We also see that the mean squared errors are smaller for the CV model.

Model	True Positives (50 active)	False Positives (9166 non-active)	Precision	Accuracy	MSE (all)	MSE (active)
CV	24	0	1.000	0.9972	0.0046	0.0147
MO	23	5	0.821	0.9965	0.1582	0.0915

Table 4.3: Performance measures obtained from the complex-valued EMVS and magnitude-only EMVS in simulation study II.

In terms of the strength, the CV model also leads to more accurate results. Figure 4.10 shows the estimated strengths obtained from the C-EMVS and MO EMVS approaches with their corresponding optimal v_0 values. The magnitude-only model overestimates the strengths for the true active sites and does not appropriately capture

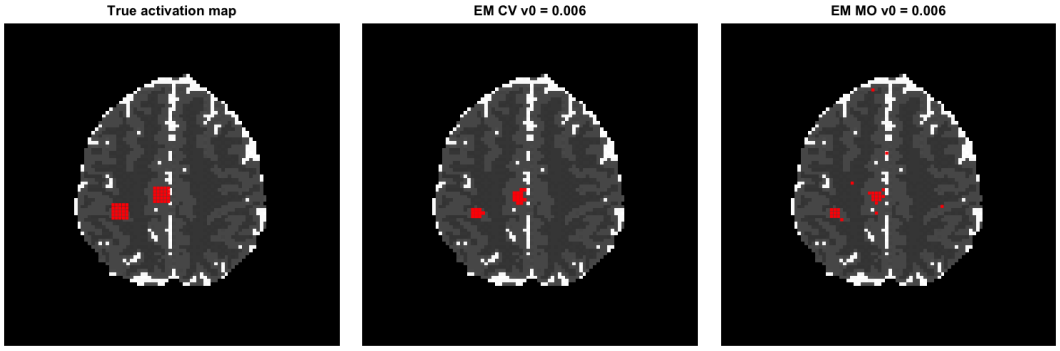


Figure 4.9: Left: True activation map. Middle: Activation map from C-EMVS at optimal v_0. Right: Activation map from magnitude-only EMVS at optimal v_0.

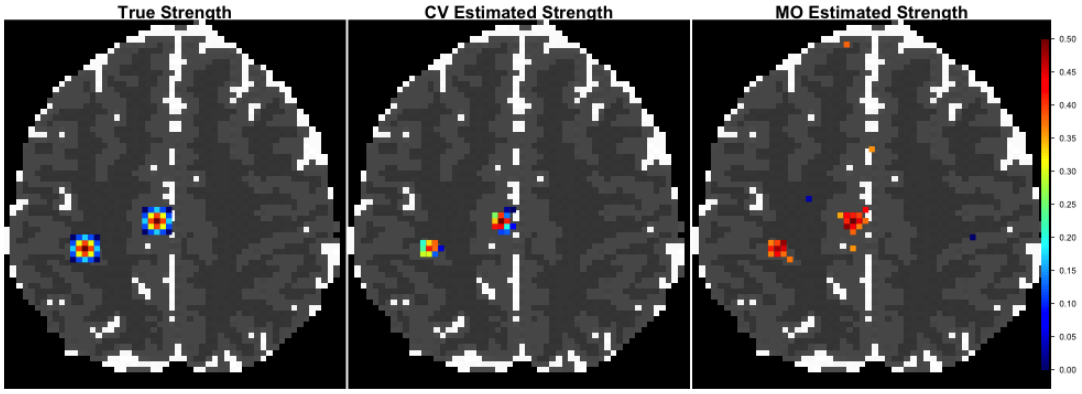


Figure 4.10: Left: True strength map. Middle: Strength map from C-EMVS at optimal v_0. Right: Strength map from magnitude-only EMVS at optimal v_0.

the fading effect observed in the true strength map. Finally, we note that full posterior results obtained via MCMC (not shown) were similar to those obtained through the EM approaches for both models.

4.4 Analysis of human CV-fMRI data

We analyzed human data recorded during an fMRI experiment performed on a single subject on a 3.0-T General Electric SIGNA LX magnetic resonance imager. The experiment consisted of a unilateral finger-tapping task performed with a visual

cue indicating whether to tap or rest. A block designed experiment with an initial 20 s of rest followed by 16 epochs of 15 s on and 15 s off was used. The full dataset is composed of seven 2.5 mm thick axial slices of dimension 96×96. A single slice was used for the analysis presented here. Further details about the experiment, the dataset and previous analyses of these data are found in Karaman et al. (2014). The original time series at each voxel had 510 TRs, however, following the approach of Karaman et al. (2014), we discarded the first 20 observations for the analysis with the C-EMVS approach. Activation from this finger-tapping experiment is well-studied. However, the methods that have been used so far could have limitations in detecting activation – as suggested by the simulation studies. Our goal here is to demonstrate that our novel Bayesian complex-valued method is able to simultaneously produce activation results that are consistent with previous results and additionally lead to a reduction of spurious results, such as activation outside of the brain or in regions that are not implicated in the finger tapping task.

Karaman et al. (2014) analyzed these data with 3 different models: a complex-valued constant phase activation model that linearly describes the temporally varying magnitude (we refer to this model as KBR14-CV), a similar magnitude-only activation model (KBR14-MO), and a non-linear model referred to as DeTeCT-ING that incorporates tissue and imaging parameters T_1 and T_2^* into physical magnetization equation to model magnetic resonance (MR) magnetization. More specifically, the DeTeCT-ING model considers the physical nonlinear signal equation to model MR magnetization, uses the first scans of the CV-fMRI data to estimate the parameters T_1 and T_2^*, and

incorporates these GM (gray matter) parameter values to detect active voxels. Further details about these models and related activation maps obtained by Karaman et al. (2014) are included in Appendix B.

We applied the C-EMVS approach to these human CV-fMRI data using models with baseline, trend and activation parameters and considered different noise structures. As in the previous examples we used reference priors on the baseline and trend parameters and the proposed complex-valued spike-and-slab prior on activation parameters. We also used the canonical HRF to obtain the BOLD signals for all the voxels. Other classes of HRFs were considered, as explained in Section 4.3, resulting in similar activation results to those presented here for the canonical HRF. Regarding the noise structure, we considered independent noise and noise with a temporal correlation modeled in terms of an autoregressive process of order one or AR(1). The model with AR(1) noise was specified as follows,

$$\begin{aligned} y_t^v &= \gamma_1^{v,*} + \gamma_2^{v,*} t/T + \gamma_3^{v,*} x_t + \eta_t^v, \\ \eta_t^v &= \varphi_v \eta_{t-1}^v + \zeta_t^v, \quad \zeta_t^v \overset{\text{iid}}{\sim} CN_1\left(0, 2\sigma_v^2, 0\right), \end{aligned}$$

where φ_v is the AR(1) coefficient for voxel v. For this model we considered a prior structure with $\pi\left(\gamma_1^{*,v}\right) \propto 1$, $\pi\left(\gamma_2^{*,v}\right) \propto 1$, $\gamma_3^{*,v} \mid \psi^v \sim (1-\psi^v) CN_1\left(0, 2v_0\sigma_v^2, 0\right) + \psi^v CN_1\left(0, 2v_1\sigma_v^2, 0\right)$, $\sigma_v^2 \sim IG\left(1/2, 1/2\right)$, $\psi^v \mid \theta \sim Bernoulli\left(\theta\right)$, $\theta \sim Beta\left(1,1\right)$, and $\varphi_v \sim U\left(-1,1\right)$. In addition, we also considered models with common variance for all voxels, i.e., $\sigma_v^2 = \sigma^2$ for all v and $\sigma^2 \sim IG\left(1/2, 1/2\right)$ and models with common AR coefficient for all voxels, i.e., $\varphi_v = \varphi$ with $\varphi \sim U\left(-1,1\right)$. All the different models that

were considered are summarized in Table 4.4.

Error structure	common variance	voxel-specific variance
independent	model (i): σ^2	model (iv): σ_v^2
AR(1), common AR parameter	model (ii): σ^2, φ	model (v): σ_v^2, φ
AR(1), voxel-specific AR parameter	model (iii): σ^2, φ_v	model (vi): σ_v^2, φ_v

Table 4.4: CV models considered for the human CV-fMRI data.

Here we only present the results for models (iii) and (vi). We found that these two models led to better activation maps (i.e., smaller numbers of clear false positives in areas outside the brain) than the other models considered. The left and center plots in Figure 4.11 show the estimated values of φ_v for models (iii) and (vi). These pictures demonstrate that there is a large variability in the estimated AR coefficients at the voxel level. The voxels outside the brain essentially show no temporal correlation since the estimated AR coefficient values are close to zero. We also see that some voxels have relatively large temporal correlation with $\hat{\varphi}_v$ around 0.6, however these voxels do not lie in the activation areas. Figure 4.11 also shows the estimated values of σ_v^2 from model (vi) (right plot). It is clear from this plot that the estimated σ_v^2 are larger for those voxels inside the brain than for those outside. These estimated values are also able to differentiate gray matter from the rest and are consistent with results in Karaman et al. (2015). In particular, the right plot in Figure 4.11 shares similarities with the estimated T_1 map in Karaman et al. (2015). This is an important result given that our proposed models are able to capture a relatively sophisticated brain structure without incorporating non-linear physically-based components that would make posterior computations extremely challenging for these large dimensional data.

Figure 4.12 shows the strength maps obtained from models (iii) (left plot)

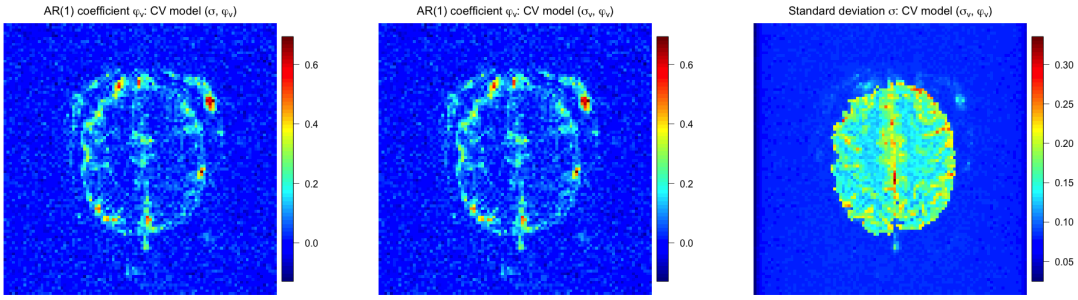


Figure 4.11: Human data: Estimated values of φ_v^2 for the CV model with a single σ^2 (model (iii), left plot) and the CV model with voxel-specific σ_v^2 (model (vi), center plot); estimated values of the φ_vs in model (vi).

and (vi) (right plot). These maps are fairly similar for both models. Comparing these C-EMVS results with those results obtained from the model of Karaman et al. (2014) we observe that the C-EMVS models have higher power of detecting active sites than the magnitude-only KBR14-MO model, and also show a better performance than the complex-valued KBR14-CV model (KBR14 maps provided in Appendix B). The maps obtained from models (iii) and (vi) show either no false positives outside of the brain and also no false positives in the upper left side of the brain close to the no signal area (model (iii)), or a much more reduced number of false positives (model (vi)) when compared to the activation map obtained from model KBR14-CV. In addition, The KBR14-MO and KBR14-CV models both use a FWE of 5% and are therefore procedures that require two steps. The Bayesian C-EMVS approach is a one-step procedure and does not require additional adjustments for multiple comparisons. The C-EMVS approach also compares favorably with the more physically realistic non-linear DeTeCT-ING model. Models (iii) and (iv) identify most of the sites in the left and supplementary motor region and produce none or a reduced number of the false

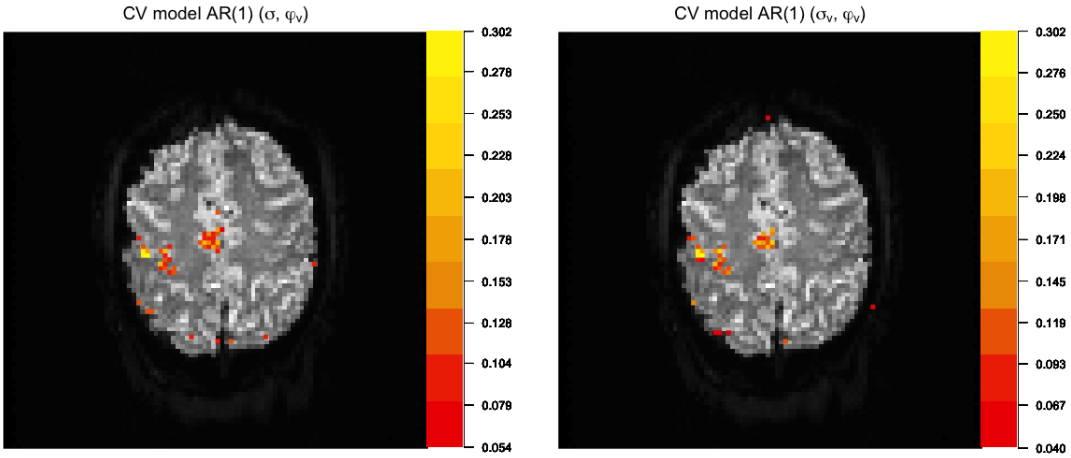


Figure 4.12: Human data: Strength maps obtained from models (iii) (left) and (vi) right.

positive sites in the anterior left side of the brain than those that were obtained by the DeTeCT-ING model. These findings are not trivial especially given that, unlike the DeTeCT-ING model, the proposed Bayesian C-EMVS approach does not incorporate any physical aspects to model MR magnetization. This example shows that methods with improved power for detecting activations, such as the one developed here, are essential to increase the understanding of human brain function, particularly in scenarios that involve CV-fMRI images with low SNR.

4.5 Discussion

We develop a new Bayesian variable selection approach for detection of brain activation from single or multi-task complex-valued fMRI signals at the voxel-specific level. Although we focused on circular complex-valued priors and the methods were only illustrated in the context of CV-fMRI data, the models and algorithms proposed

here are general, and can be applied to the general case of non-circular complex-valued priors and to other types of data.

Our simulation studies show that by considering both, real and imaginary information the Bayesian complex-valued variable selection methods are able to detect more true positives and less false positives than magnitude-only models, especially when the SNR is small. We also found that both, the Bayesian complex-valued and magnitude-only EMVS approaches performed better than lasso and adaptive lasso and were computationally fast, with run times comparable to those needed by lasso or adaptive lasso. Finally, we demonstrated that the activation results in the finger-tapping experiment obtained from the C-EMVS approach compared favorably to those results obtained from more sophisticated non-linear models that are physically realistic as they incorporate tissue and imaging parameters. The computational efficiency and the performance obtained in the analysis of experimental and simulated CV-fMRI data presented here make the C-EMVS approach a very useful tool for detecting brain activation.

We note that the new Bayesian complex-valued models considered here do not use any sophisticated spatio-temporal structure that can more appropriately describe fMRI data (we only considered an AR(1) temporal structure). Adding spatio-temporal structure that can better describe the data could potentially lead to further improved results, but would also lead to more computationally-intensive models that may be not be feasible for detecting brain activation at the voxel-specific level. Chapter 5 explores computationally efficient Bayesian complex-valued spatial models and Chapter

6 discusses multi-subject models for CV-fMRI data. The C-EMVS methods presented here serve as as a highly useful starting point, especially for analyzing high-dimensional CV-fMRI data.

Chapter 5

Bayesian spatial models for brain activation in CV-fMRI

In this chapter, we consider the Cartesian complex-valued model representation of Lee et al. (2007, 2009) and Yu et al. (2018) and propose a flexible and computationally feasible Bayesian variable selection approach that uses a low-rank spatial model through kernel convolution (KC) for detecting activation in CV-fMRI at the voxel level. More specifically, our complex-valued Bayesian variable selection model assumes a complex normal noise and uses complex-valued g-priors on the regression coefficients. We then add a spatial structure on the binary indicator variables that determine if a given voxel is activated. Following Bezener et al. (2017), we first parcellate the image into several low-rank second layer spatial regions. However, unlike Bezener et al. (2017), we describe the spatial structure through a kernel convolution model rather than a Gaussian process (GP). We are then able to recover the entire

underlying spatial process at the voxel level by convolving a set of region-specific spatial random effects with voxel and region-dependent kernel functions. The parcellation allows us to reduce the resolution of the image and consequently leads to faster computation, while kernel convolution provides increased spatial flexibility as it allows us to relate a given voxel with neighboring voxels located in the same brain region, and also with voxels located in other regions that are farther away in the brain. Our approach uses the CV-fMRI data to infer the parameters that define the spatial kernel functions, providing a flexible framework for determining the strength and range of the spatial dependency across voxels in the context of brain activation.

We first present our Bayesian variable selection method with kernel-based spatial structure for detecting activation at the voxel level in CV-fMRI data. We refer to this method as the Complex-Valued Kernel Convolution (CV-KC) method. We also describe a complex-valued Bayesian variable selection method with spatial Gaussian processes, referred to as CV-GP. This model extends the approach of Bezener et al. (2017) to the complex-valued domain. We compare both methods in terms of their power and accuracy for detecting activation in different simulation scenarios. We also examine how the spatial models improve activation detection relative to the non-spatial complex-valued approach of our non-spatial complex-valued approach summarized in Chapter 4 and in Yu et al. (2018). We show that our kernel-based approach significantly increases the detection power when activation strength is weak, i.e., in cases of low CNRs, for both, complex-valued and magnitude-only models. Including a spatial structure in the model allows us to eliminate or significantly reduce the number of iso-

lated voxels incorrectly labeled as active voxels. Therefore, this approach encourages clusters of brain activation without leading to increased false-positives in contrast to other magnitude-only spatial approaches implemented in common software packages for fMRI analysis (Eklund et al., 2016). Finally, we also compare the complex-valued approaches with magnitude-only approches in several simulation studies and in the analysis of human fMRI data. Section 5.1 describes the Bayesian spatial CV-KC and CV-GP models. Section 5.2 discusses the MCMC algorithms for full posterior inference under these models. Section 5.3 illustrates the performance of the proposed methods in an extensive simulation study. The performance of the complex-valued spatial approaches are compared to complex-valued approaches and magnitude only approaches that do not incorporate spatial structure. Section 5.4 discusses the results obtained from analyzing a human CV-fMRI dataset and Section 5.5 presents a discussion and future extensions.

5.1 Bayesian spatial models for brain activation in complex-valued fMRI

We consider two different Bayesian spatial hierarchical models for detecting brain activation at the voxel level in CV-fMRI data. We first extend the model proposed in Bezener et al. (2017) to the complex-valued domain. This model uses a standard Gaussian process to describe the spatial random effects that determine the probability of activation detection for each voxel in the image. We then propose a novel model that uses kernel convolution to describe the spatial random effects. More

specifically, the spatial structure on the activation probabilities is induced by writing the spatial random effects as a convolution of a given kernel function, e.g., a Gaussian kernel, and Gaussian distributed random variables. As illustrated later via simulation studies, and also in the analysis of human CV-fMRI data, the proposed kernel-based model induces dimension reduction as the voxel-level spatial effects are written as a weighted average of a smaller number of region-specific effects and kernel values that are voxel and region-specific. Through this smoothing process, each voxel borrows information about activation from its neighboring voxels. Therefore, the kernel-based method leads to more accurate activation results than the model that assumes a Gaussian process directly on the random effects.

We begin by providing some background on kernel-convolution processes and by describing the complex-valued regression model for CV-fMRI data. We then discuss in detail the spatial structure in the two complex-valued models mentioned above.

5.1.1 Some background on kernel convolution methods

Kernel convolution, or process convolution, has been widely used for spatial modeling in environmental sciences and geostatistics (Higdon, 1998, 2002; Lee et al., 2005). It provides a convenient way to construct a Gaussian process over a general spatial region domain denoted as $\mathcal{S}$, such as a two-dimensional real field. In order to define a process convolution, consider a kernel function $k(s;\phi)$, where $s \in \mathcal{S}$, and ϕ is the associated parameter (possibly a vector) that defines the kernel function, and let $w(s)$ be a white noise process, i.e., $\mathrm{E}(w(s)) = 0$, $\mathrm{Var}(w(s)) = \sigma^2$, and

$\text{Cov}\left(w\left(s\right), w\left(s'\right)\right) = 0$. Then, the process

$$x\left(s\right) = \int_{\mathcal{S}} k\left(s-u;\phi\right) w\left(u\right) du \tag{5.1}$$

is a stationary Gaussian process that has mean zero and

$$\begin{aligned} \text{Var}\left(x\left(s\right)\right) &= \sigma^2 \int_{\mathcal{S}} k^2\left(s-u;\phi\right) du \\ c\left(d\right) = \text{Cov}\left(x\left(s\right), x\left(s'\right)\right) &= \sigma^2 \int_{\mathcal{S}} k\left(t;\phi\right) k\left(t-d;\phi\right) dt, \quad d = s - s'. \end{aligned}$$

If $\mathcal{S}$ is $\mathbb{R}^m$ and $k\left(s;\phi\right)$ is isotropic[1], there is a one-to-one correspondence between the kernel and the covariance function $c\left(d\right)$ of $x\left(s\right)$, provided that the first two moments of $k\left(s;\phi\right)$ exist or $c\left(d\right)$ is integrable and positive definite. Note however that if the process is not isotropic multiple kernels can map to the same covariance function. In addition, based on the convolution theorem for Fourier transformations, the Fourier transform of the covariance of $x\left(s\right)$ is the square of the Fourier transform of $k\left(s;\phi\right)$.

In practice, due to the complexities involved in doing numerical integration, a finite sum approximation to (5.1) is applied (Higdon, 1998), i.e., for a grid of sites $u_1, \ldots, u_G$ in $\mathcal{S}$, the process is defined as

$$x\left(s\right) = \sum_{g=1}^{G} k\left(s-u_g;\phi\right) w\left(u_g\right).$$

A result of this discrete approximation is that the process is no longer stationary

[1] A kernel function $k\left(s-s';\phi\right)$ is isotropic if its value depends on distance alone, i.e., $k\left(s-s';\phi\right) = k\left(\tau;\phi\right)$, where $\tau = \|s-s'\|$. An isotropic kernel function has a corresponding covariance function $c\left(\tau\right)$ that is also isotropic.

since the covariance of $x(s)$ is $\mathrm{Cov}(x(s), x(s')) = \sigma^2 \sum_{g=1}^{G} k(s - u_g; \phi) k(s' - u_g; \phi)$, which is not a function of the distance $\|s - s'\|$. However, such finite sum approximation provides dimension reduction and allows us to simplify the computational complexity of the algorithms for posterior inference. A relatively small number G of parameters $w(u_1), \ldots, w(u_G)$ will effectively control the entire spatial process $x(s)$.

Instead of using MRFs or defining a specific local neighboring structure for each voxel, we use a convolution-based spatial prior on the parameters that specify the voxel-specific probabilities of activation. This structure allows us to uncover the underlying full probability map of brain activation from CV-fMRI data. Regarding the kernel function, $k(s; \phi)$, several choices are available and have been used in the literature. A popular choice is the Gaussian kernel because of its analytical properties and computational appeal. However, the Gaussian kernel leads to the Gaussian covariance function, and its corresponding random field is infinitely smooth. This assumption is not very realistic for fMRI data. Here we use another popular kernel function, the Bezier kernel, whose 2-D version is given by

$$k(z_v - u_g; \phi) = \left(1 - \frac{\|z_v - u_g\|^2}{\phi^2}\right)^{\nu}, \quad \|z_v - u_g\| < \phi, \tag{5.2}$$

with $\nu > 0$ a smoothing parameter, which will be assumed fixed, and range parameter $\phi > 0$. In the later fMRI analysis, z_v is the location of voxel v, and u_g can be seen as the site of the spatial region g in the image. This kernel can be seen as a general version of several commonly used kernel functions such as the parabolic ($\nu = 1$), quartic (biweight) ($\nu = 2$) or the triweight kernel ($\nu = 3$) up to some constant. One

of the key features of the Benzier kernel is that it has a compact support. This avoids unrealistically relating any two voxels in the image and does not require us to pre-specify a particular neighboring structure. Instead, the model learns the neighboring structure from the data by inferring the range parameter ϕ. In the simulation and human data studies presented below we consider different values of ν, and compare the Bezier kernel and Gaussian kernel. As an illustration, Figure 5.1 shows the effect of ϕ on the Bezier kernel when $\nu = 2$. As ϕ increases, the range affected by the kernel becomes larger.

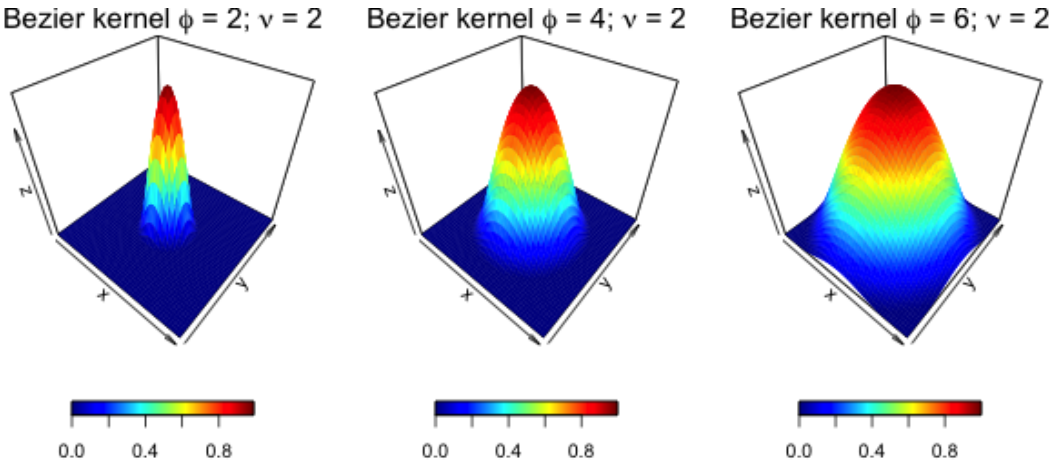


Figure 5.1: Effect of ϕ on the Bezier kernel when $\nu = 2$. Both x and y are from -5 to 5. The color bar below shows kernel values z.

5.1.2 A complex-valued linear model for task-related CV-fMRI

As in the previous chapter, we continue to use the Cartesian coordinates representation of Lee et al. (2007), Lee et al. (2009) and Rowe (2009), or the complex-valued linear model (3.7). Without loss of generality we assume that the fMRI signals have been centered and detrended, and so, in our setting each $x_{j,t}$ corresponds to the

discretized convolution of the on-and-off signal of the stimulus sequence for task j with a given HRF. Again, we further assume that the HRF is known for all t and v. This assumption can be relaxed by considering models that also infer the voxel-specific HRF (Rowe, 2001), but this significantly increases the computational cost of posterior inference, especially because our model has a voxel-level resolution.

We add a spatial structure in $\boldsymbol{\gamma}^v$ as follows. Let ψ_j^v be an indicator variable such that $\gamma_j^v = 0$, with $\gamma_j^v = \gamma_{Re,j}^v + i\gamma_{Im,j}^v$, if $\psi_j^v = 0$ and $\gamma_j^v \neq 0$ if $\psi_j^v = 1$. Then, model (3.7) can be rewritten to make the dependence on the indicator variables explicit, i.e.,

$$\mathbf{y}^v = \mathbf{X}\left(\boldsymbol{\psi}^v\right)\boldsymbol{\gamma}^v\left(\boldsymbol{\psi}^v\right) + \boldsymbol{\eta}^v, \quad \boldsymbol{\eta}^v \overset{\text{iid}}{\sim} CN_T\left(\mathbf{0}, 2\sigma_v^2\mathbf{I}_T, \mathbf{0}\right), \tag{5.3}$$

where $\boldsymbol{\psi}^v = \left(\psi_1^v, \ldots, \psi_p^v\right)'$. Thus, detecting activation at the voxel-specific level can be viewed as a variable selection problem where a voxel v is labeled as active under task j if and only if $\gamma_j^v \neq 0$, which corresponds to $\psi_j^v = 1$, and labeled as non-activated otherwise.

The model in (5.3) defines the likelihood so we now proceed to discuss the prior specification for the model parameters. In particular, we focus on the prior for the indicator variables $\boldsymbol{\psi}^v$, as spatial dependence is introduced through these binary variables. For the remaining parameters, we choose their prior distributions so that posterior inference can be computationally efficient.

Let $\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \ldots, \sigma_V^2\right)$. We assume $\pi\left(\boldsymbol{\sigma}^2\right) = \prod_{v=1}^V \pi\left(\sigma_v^2\right)$, with $\sigma_v^2 \overset{\text{iid}}{\sim} IG\left(a_\sigma, b_\sigma\right)$. In practice we typically set $a_\sigma = b_\sigma = 1/2$, which leads to a relatively non-informative prior $IG\left(1/2, 1/2\right)$. The invariant prior $\pi\left(\sigma_v^2\right) \propto \sigma_v^{-2}$ can also be con-

sidered. In order to make comparisons with the approach of Bezener et al. (2017), we use a complex circular normal g-prior on $\boldsymbol{\gamma}^v(\boldsymbol{\psi}^v)$, which is given by

$$\boldsymbol{\gamma}^v(\boldsymbol{\psi}^v) \mid \boldsymbol{\gamma}^v, \sigma_v^2 \overset{\text{ind}}{\sim} CN_p\left(\hat{\boldsymbol{\gamma}}^v(\boldsymbol{\psi}^v), 2T\sigma_v^2\left(\mathbf{X}'(\boldsymbol{\gamma}^v)\mathbf{X}_v(\boldsymbol{\gamma}^v)\right)^{-1}, \mathbf{0}\right),$$

with $\hat{\boldsymbol{\gamma}}^v(\boldsymbol{\psi}^v) = \left(\mathbf{X}'(\boldsymbol{\gamma}^v)\mathbf{X}(\boldsymbol{\gamma}^v)\right)^{-1}\mathbf{X}(\boldsymbol{\gamma}^v)\mathbf{y}^v$. As it is the case in the real-valued domain, this circular complex normal prior allows $\boldsymbol{\gamma}^v(\boldsymbol{\psi}^v)$ to be easily integrated out and hence makes the computation more efficient. Note that, similar to Bezener et al. (2017), the parameter g in the g-prior is fixed at T, the length of the time course of the CV-fMRI signal, for both real and imaginary parts, which yields a unit information prior that produces similar results as those derived from BIC (Raftery, 1998). An empirical Bayes approach can also be used to select the value of g.

Due to the large number of voxels in the images, it is computationally challenging to do spatial modeling at the voxel-specific level. Therefore, we consider a two-step procedure following Bezener et al. (2017), i.e., we first parcellate the images into G spatial regions of voxels, and then construct a latent spatial random process based on the G spatial regions. The regions must be non-overlapping, so that each voxel belongs to only one spatial region. In addition, since the centroids of the regions will determine the location of the regions and will be used to compute how strongly spatially correlated the different regions are, the shapes of the regions should be such that their centroid lies in their region. Hence, a ring-shaped region or any other non-convex set may not be valid. Anatomical information or brain atlases can be used to define the different regions or to determine regions of interests (ROIs), however, in the

absence of any prior anatomical information, using equally-spaced grids to parcellate the image is a natural starting point. In general, choosing a larger number of spatial regions G will lead to better activation detection performance, but it will also imply a heavier computational burden. Because of this it is important to choose a spatial prior that is not too sensitive to the choice of G. As mentioned before, we consider two classes of spatial priors that lead to two spatial models in the complex-valued domain. One of these models extends the Gaussian process prior approach of Bezener et al. (2017) to the complex-valued domain and will be referred to as the CV-GP model. The alternative model uses a kernel-convolution approach to induce the spatial structure and will be referred to as the CV-KC model. We show via simulation studies that the performance of the CV-GP model is sensitive to G and that its computing time grows significantly with G. The performance of the CV-KS approach, on the contrary, is not very sensitive to G and its computing time grows much slower as G increases. We describe the spatial priors below.

5.1.3 Hierarchical Gaussian process priors on ψ^v for the GP model

Let $\boldsymbol{\psi}_{(j)}$ be the set of indicator variables for task j and all the voxels v with $v = 1, \ldots, V$, i.e., $\boldsymbol{\psi}_{(j)} = \left(\psi_j^1, \ldots, \psi_j^V\right)$. Define $\boldsymbol{S}_{(j)} = \left(S_j^1, \ldots, S_j^G\right)$ as a collection of region-specific spatial random effects for task j for a given parcellation. Let $\mathcal{R}_g$ denote the set of all voxels in region g, for $g = 1, \ldots, G$. The spatial prior for this model assumes that the indicator variables are conditionally independent given the spatial random effects, and that these random effects are generated from a Gaussian process

as follows:

$$\begin{aligned}\pi\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{S}_{(j)}\right) &= \prod_{g=1}^{G}\prod_{v\in\mathcal{R}_g}\pi\left(\psi_j^v \mid S_j^g\right), \\ \psi_j^v \mid S_j^g &\overset{\text{ind}}{\sim} Bernoulli\left(\frac{1}{1+e^{-\left(a_j^g+S_j^g\right)}}\right).\end{aligned}$$

Here a_j^g allows us to incorporate any prior external anatomical information into the model. Finally, for this model it is assumed that the spatial random effects $\boldsymbol{S}_{(j)}$ are generated from a Gaussian process, i.e.,

$$\boldsymbol{S}_{(j)} \mid \delta_j^2, r_j \overset{\text{ind}}{\sim} N\left(0, \delta_j^2\Gamma_j\right),$$

where Γ_j is an exponential covariance function whose (i,k)-th element is given by $\Gamma_j(i,k) = \exp\left(-\frac{\|s_i - s_k\|}{r_j}\right)$, with s_i and s_k denoting the centroid coordinates of regions i and k, respectively, and $\|\cdot\|$ denoting the Euclidean distance. Here δ_j^2 is a parameter that controls the smoothness of $\boldsymbol{S}_{(j)}$, and therefore the smoothness of $\boldsymbol{\psi}_{(j)}$, while r_j is a range parameter that determines the strength of the spatial dependence between neighboring regions under task j. Now, letting $\boldsymbol{\delta}^2 = \left(\delta_1^2, \ldots, \delta_p^2\right)$, $\boldsymbol{r} = (r_1, \ldots, r_p)$, and $\boldsymbol{S} = \left(\boldsymbol{S}_{(1)}, \ldots, \boldsymbol{S}_{(p)}\right)$, we assume

$$\begin{aligned}\pi\left(\boldsymbol{S}, \boldsymbol{\delta}, \boldsymbol{r}\right) &= \pi\left(\boldsymbol{S} \mid \boldsymbol{\delta}, \boldsymbol{r}\right)\pi\left(\boldsymbol{\delta}\right)\pi\left(\boldsymbol{r}\right), \\ \pi\left(\boldsymbol{S} \mid \boldsymbol{\delta}, \boldsymbol{r}\right) &= \prod_{j=1}^{p}\pi\left(\boldsymbol{S}_{(j)} \mid \delta_j, r_j\right), \quad \pi\left(\boldsymbol{\delta}^2\right) = \prod_{j=1}^{p}\pi\left(\delta_j^2\right), \quad \pi\left(\boldsymbol{r}\right) = \prod_{j=1}^{p}\pi\left(r_j\right),\end{aligned}$$

with $\pi\left(\delta_j^2\right) \propto \delta_j^{-2}$ and χ^2 priors for r_j, for $j = 1, \ldots, p$. Notice that through this spatial prior setting, voxels in the same spatial region g, i.e., all $v \in \mathcal{R}_g$, share the same region-level spatial random effect S_j^g and the same a_j^g, and hence, the same probability of activation $\frac{1}{1+e^{-\left(a_j^g + S_j^g\right)}}$. On the other hand, voxels in different regions will be related through the underlying Gaussian structure of the spatial random effects, which may be stronger or weaker depending on the scale and range parameters that will be learned from the data.

5.1.4 Hierarchical spatial priors on ψ^v for the KC model

As it was the case with the CV-GP spatial priors described above, our proposed kernel convolution model model assumes that $\boldsymbol{\psi}_{(j)}$ is conditionally independent given the spatial random effects $\boldsymbol{S}_{(j)}$. However, instead of assuming region-specific spatial random effects, we now assume that $\boldsymbol{S}_{(j)}$ contains voxel-specific spatial random effects, and so $\boldsymbol{S}_{(j)} = \left(S_j^1, \ldots, S_j^V\right)$. Then, letting w_j^g, be the spatial random effect for task j and region g, $z_1, \ldots, z_V$, denote the locations of the voxels, and $k\left(z_v - s_g; \phi\right)$ a kernel function, we have

$$S_j^v = \sum_{g=1}^{G} k\left(z_v - s_g; \phi\right) w_j^g,$$

where s_g is the centroid of region g. The region-specific spatial effects w_j^g are assumed to be Gaussian distributed such that $w_j^g \mid \tau_j^2 \overset{\text{ind}}{\sim} N\left(0, \tau_j^2\right)$, and $\tau_j^2 \overset{\text{iid}}{\sim} IG\left(a_\tau, b_\tau\right)$, or equivalently, $\boldsymbol{S}_{(j)} = \mathbf{K}(\phi)\mathbf{w}_{(j)}$ follows a V-dimensional Gaussian distribution $N\left(\mathbf{0}, \tau_j^2 \mathbf{K}(\phi)\mathbf{K}(\phi)'\right)$, where $\mathbf{K}(\phi) = \left[k\left(z_v - s_g; \phi\right)\right]$ is the $V \times G$ matrix of kernel values

whose (v, g)-th element is $k(z_v - s_g; \phi)$ and $\mathbf{w}_{(j)} = \left(w_j^1, \ldots, w_j^G\right)$. The parameter ϕ in the kernel function could be a positive scalar common for all the regions or a region-specific vector, i.e., $\phi = \left(\phi^1, \ldots, \phi^G\right)$. This parameter controls the shape of the kernels, and hence the smoothness of the activation maps. Different priors can be assigned depending on the scalar or multivariate nature of ϕ. A common gamma or inverse gamma prior on a scalar ϕ or on independent priors on each ϕ^g can be used, for example, $\phi^g \overset{\text{iid}}{\sim} Ga(a_\phi, b_\phi)$. Region-specific a_{ϕ^g} and b_{ϕ^g} can also be considered if one has any prior information about how the spatial effect w^g in region g can contribute to the spatial structure in other regions.

Note that instead of having V independent $S_j^1, \ldots, S_j^V$, spatial dependence is achieved by relating $\psi_j^1, \ldots, \psi_j^V$ using kernel convolution. In addition, this kernel convolution approach leads to dimension reduction, as only a relatively small number of region-specific effects may be needed to describe the entire underlying spatial process in the image at the voxel level. The voxel-level spatial effect S_j^v is a weighted average of the region-specific spatial effects w_j^g and the kernel values that are voxel and region specific. Note also that $\{k(z_v - s_g; \phi)\}_{g=1}^G$ can be viewed as a set of basis functions. Hence, a voxel v in region g can relate not only to voxels in its own region, but to voxels in other regions. Voxels in the same region under the KC model can have less similarities in their probabilities of activation than those implied by the GP model because in the KC approach each voxel is allowed to borrow information from different regions through different values in the kernel convolution. This property allows us to obtain higher resolution latent spatial effects and more precise posterior probability activation maps,

as it is illustrated in the simulation studies and in the human data example presented in Sections 5.3 and 5.4. As it is shown later, the posterior probability activation maps obtained from the GP approach usually display a probability inflation effect, i.e., when a given spatial region contains a non-negligible number of activated voxels, other non-activated voxels in the region will also have relatively high probabilities of activation. On the other hand, two voxels located next to each other may also have significantly different probabilities of activation in the GP model due to their different spatial region assignment. This effect usually occurs when the correlation between spatial regions decays very fast, and it highly depends on the resolution of the grid that defines the spatial regions, and on how similar is the behavior of voxels that belong to the same region. The KC model does not have this issue because of the "spillover effect" of convolution methods. Kernel functions are more flexible in terms of determining how a given voxel is related to voxels in other spatial region through the kernel structure and the parameter ϕ. The strength of the relationship between voxels is explained by the kernel values. As a result, the KC approach is less sensitive to the parcellation than the GP approach. Finally, note that although voxels in the same region have their own unique spatial random effect, they can also share common region-specific but external information through the constant a_j^g that describes features that are only specific to this particular region for the different tasks.

5.2 Posterior Inference

In this section we focus on posterior inference for the CV-KC model via MCMC. The posterior distribution and sampling scheme for the magnitude-only GP model can be found in Bezener et al. (2017). The MCMC algorithm for posterior sampling in the complex-valued case is very similar to the algorithm of Bezener et al. (2017), so we do not describe it here. Instead, we focus on the algorithm for posterior sampling under the CV-KC model.

Let $\mathbf{y} = \left(\mathbf{y}^1, \ldots, \mathbf{y}^V\right)$ be the observed complex-valued image, $\boldsymbol{\gamma}\left(\boldsymbol{\psi}\right) = \left(\boldsymbol{\gamma}^1\left(\boldsymbol{\psi}^1\right), \ldots, \boldsymbol{\gamma}^V\left(\boldsymbol{\psi}^V\right)\right)$, $\boldsymbol{\psi} = \left(\boldsymbol{\psi}_{(1)}, \ldots, \boldsymbol{\psi}_{(p)}\right)$, $\mathbf{w} = \left(\mathbf{w}_{(1)}, \ldots, \mathbf{w}_{(p)}\right)$, $\mathbf{w}_{(j)} = \left(w_j^1, \ldots, w_j^G\right)$, $\boldsymbol{\tau}^2 = \left(\tau_1^2, \ldots, \tau_p^2\right)$, the model parameters and $\boldsymbol{S}_{(j)} = \mathbf{K}(\phi)\mathbf{w}_{(j)}$. Without loss of generality, we assume that ϕ is a scalar with $\phi > 0$. Based on the model and prior setup given in the previous section, the posterior density of the CV-KC model is given by

$$\begin{aligned} q\left(\boldsymbol{\gamma}\left(\boldsymbol{\psi}\right), \boldsymbol{\psi}, \mathbf{w}, \boldsymbol{\tau}^2, \boldsymbol{\sigma}^2 \mid \mathbf{y}\right) \quad \propto \quad & p\left(\mathbf{y} \mid \boldsymbol{\gamma}\left(\boldsymbol{\psi}\right), \boldsymbol{\psi}, \boldsymbol{\sigma}^2\right) \pi\left(\boldsymbol{\gamma}\left(\boldsymbol{\psi}\right) \mid \boldsymbol{\psi}, \boldsymbol{\sigma}^2\right) \pi\left(\boldsymbol{\sigma}^2\right) \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi\right) \\ & \times \pi\left(\mathbf{w} \mid \boldsymbol{\tau}^2\right) \pi\left(\boldsymbol{\tau}^2\right) \pi\left(\phi\right). \end{aligned}$$

Note that this model is high-dimensional with $V(2p+1)+Gp+p+1$ parameters, even for single-subject activation analysis. Therefore, to further improve computational efficiency, we take advantage of the conjugate priors and integrate out the parameters

$\boldsymbol{\gamma}(\boldsymbol{\psi}), \boldsymbol{\sigma}^2$ and $\boldsymbol{\tau}^2$ to get the marginal posterior

$$q(\boldsymbol{\psi}, \mathbf{w}, \phi \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\psi}) \pi(\boldsymbol{\psi} \mid \mathbf{w}, \phi) \pi(\mathbf{w}) \pi(\phi).$$

Let $\tilde{\mathbf{y}}^v = \left(y^v_{1,Re}, \ldots, y^v_{T,Re}, y^v_{1,Im}, \ldots, y^v_{T,Im}\right)'$ and $\tilde{\mathbf{X}}_v = \tilde{\mathbf{X}}(\boldsymbol{\psi}^v) =$ blockdiag $(\mathbf{X}(\boldsymbol{\psi}^v), \mathbf{X}(\boldsymbol{\psi}^v))$. In the circular normal case, after integrating $\boldsymbol{\gamma}(\boldsymbol{\psi})$ and $\boldsymbol{\sigma}^2$ out, the real representation of the marginal $p(\tilde{\mathbf{y}} \mid \boldsymbol{\psi})$ in the complex-valued model model is given by

$$\begin{aligned} p(\tilde{\mathbf{y}} \mid \boldsymbol{\psi}) &= \int p\left(\tilde{\mathbf{y}} \mid \tilde{\boldsymbol{\gamma}}(\boldsymbol{\psi}), \boldsymbol{\psi}, \boldsymbol{\sigma}^2\right) \pi\left(\tilde{\boldsymbol{\gamma}}(\boldsymbol{\psi}) \mid \boldsymbol{\psi}, \boldsymbol{\sigma}^2\right) \pi\left(\boldsymbol{\sigma}^2\right) d\boldsymbol{\gamma} d\boldsymbol{\sigma}^2 \\ &= \prod_{v=1}^{V} (1+T)^{-q_v} \left((\tilde{\mathbf{y}}^v)' \tilde{\mathbf{y}}^v - (\tilde{\mathbf{y}}^v)' \tilde{\mathbf{X}}_v \left(\tilde{\mathbf{X}}_v' \tilde{\mathbf{X}}_v \right)^{-1} \tilde{\mathbf{X}}_v' \tilde{\mathbf{y}}^v + 2b_\sigma \right)^{-(T+a_\sigma)} \\ &= \prod_{v=1}^{V} (1+T)^{-q_v} M(\boldsymbol{\psi}^v)^{-(T+a_\sigma)}, \end{aligned}$$

where $q_v = \sum_{j=1}^{p} \psi_j^v$ and $M(\boldsymbol{\psi}^v) = \left((\tilde{\mathbf{y}}^v)' \tilde{\mathbf{y}}^v - (\tilde{\mathbf{y}}^v)' \tilde{\mathbf{X}}_v \left(\tilde{\mathbf{X}}_v' \tilde{\mathbf{X}}_v \right)^{-1} \tilde{\mathbf{X}}_v' \tilde{\mathbf{y}}^v + 2b_\sigma \right)$. Then we have the marginal posterior

$$q\left(\boldsymbol{\psi}, \mathbf{w}, \boldsymbol{\tau}^2, \phi \mid \mathbf{y}\right) \propto p(\tilde{\mathbf{y}} \mid \boldsymbol{\psi}) \pi(\boldsymbol{\psi} \mid \mathbf{w}, \phi) \pi\left(\mathbf{w} \mid \boldsymbol{\tau}^2\right) \pi\left(\boldsymbol{\tau}^2\right) \pi(\phi).$$

Each τ_j^2 can further be integrated out to get the marginal distribution of $\mathbf{w}_{(j)}$, and so,

for each $j = 1, \ldots, p$,

$$\begin{aligned}\pi\left(\mathbf{w}_{(j)}\right) &= \int \pi\left(\mathbf{w}_{(j)} \mid \tau_j^2\right) \pi\left(\tau_j^2 \mid a_\tau, b_\tau\right) d\tau_j^2 = \int \prod_{g=1}^{G} N\left(w_j^g \mid \tau_j^2\right) IG\left(\tau_j^2 \mid a_\tau, b_\tau\right) d\tau_j^2 \\ &= \frac{b_\tau^{a_\tau}}{\Gamma\left(a_\tau\right)} \times \frac{\Gamma\left(G/2 + a_\tau\right)}{\left(\frac{1}{2}\sum_{g=1}^{G}\left(w_j^g\right)^2 + b_\tau\right)^{(G/2+a_\tau)}}.\end{aligned}$$

As a result, we now have $q\left(\boldsymbol{\psi}, \mathbf{w}, \phi \mid \mathbf{y}\right) \propto p\left(\tilde{\mathbf{y}} \mid \boldsymbol{\psi}\right) \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi\right) \pi\left(\mathbf{w}\right) \pi\left(\phi\right)$ and can propose a MCMC algorithm to sample from this distribution.

First, note that the conditional distribution for $\boldsymbol{\psi}$ is

$$\begin{aligned}q\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi\right) p\left(\tilde{\mathbf{y}} \mid \boldsymbol{\psi}\right), \\ &\propto \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi\right) \prod_{v=1}^{V} \left(1+T\right)^{-q_v} M\left(\boldsymbol{\psi}^v\right)^{-(T+a_\sigma)},\end{aligned}$$

where

$$\pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi\right) = \prod_{j=1}^{p} \pi\left(\boldsymbol{\psi}_{(j)} \mid \phi, \mathbf{w}_{(j)}\right) = \prod_{j=1}^{p}\prod_{v=1}^{V} \pi\left(\psi_j^v \mid \mathbf{K}_v'(\phi)\mathbf{w}_{(j)}\right),$$

where $\mathbf{K}_v(\phi)$ is the v-th row vector of $\mathbf{K}$. Similarly, we obtain the conditional distributions for $\mathbf{w}$ and ϕ from $q\left(\mathbf{w} \mid \boldsymbol{\psi}, \phi, \mathbf{y}\right) \propto \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi\right) \pi\left(\mathbf{w}\right)$, and $q\left(\phi \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y}\right) \propto \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi\right) \pi\left(\phi\right)$.

The marginal posterior $q\left(\boldsymbol{\psi}, \mathbf{w}, \phi \mid \mathbf{y}\right)$ has dimension $p(V+G)+1$. We generate posterior samples from this distribution sequentially by sampling $\mathbf{w}$, $\boldsymbol{\psi}$, and ϕ. We sample $\mathbf{w}$ and ϕ using Metropolis-Hastings steps, and then sample $\boldsymbol{\psi}$ using a Gibbs step as follows:

- Sampling $\mathbf{w}$: this involves pG steps. For $j = 1, \ldots, p$ and for each g with $g = 1, \ldots, G$ we have that the conditional density of w_j^g is given by

$$\begin{aligned} q\left(w_j^g \;\middle|\; \mathbf{w}_{(-j)}^{(-g)}, \boldsymbol{\psi}, \phi, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi\right) \pi\left(\mathbf{w}_{(j)}\right) \\ &\propto \pi\left(w_j^g\right) \prod_{v=1}^{V} \pi\left(\psi_j^v \;\middle|\; S_j^v\left(w_j^g; \mathbf{K}_v(\phi), \mathbf{w}_{(-j)}^{(-g)}\right)\right), \end{aligned}$$

where $\mathbf{w}_{(-j)}^{(-g)}$ are all region level spatial random effects w_l^h such that $l \neq j$, $h \neq g$. . Here, $S_j^v\left(w_j^g; \mathbf{K}_v(\phi), \mathbf{w}_{(-j)}^{(-g)}\right)$ denotes that we view S_j^v as a function of w_j^g leaving the remaining parameters in $K_v(\phi)$ and $\mathbf{w}_{(-j)}^{(-g)}$ fixed. A random walk proposal $\left(w_j^g\right)^* \sim N\left(\left(w_j^g\right)^{(m)}, \sigma^2_{w_j^g}\right)$ is used, where $\left(w_j^g\right)^{(m)}$ is the current state and $\sigma^2_{w_j^g}$ is a tuning parameter. Hence, the Metropolis-Hastings ratio is

$$\begin{aligned} \frac{q\left(\left(w_j^g\right)^* \;\middle|\; w_{-j}^{-g}, \boldsymbol{\psi}, \phi, \mathbf{y}\right)}{q\left(\left(w_j^g\right)^{(m)} \;\middle|\; w_{-j}^{-g}, \boldsymbol{\psi}, \phi, \mathbf{y}\right)} &= \frac{\prod_{v=1}^{V} \pi\left(\psi_j^v \;\middle|\; S_j^v\left(\left(w_j^g\right)^*; \mathbf{K}_v(\phi), \mathbf{w}_{(-j)}^{(-g)}\right)\right)}{\prod_{v=1}^{V} \pi\left(\psi_j^v \;\middle|\; S_j^v\left(\left(w_j^g\right)^{(m)}; \mathbf{K}_v(\phi), \mathbf{w}_{(-j)}^{(-g)}\right)\right)} \\ &\quad \times \frac{\pi\left(\left(w_j^g\right)^*\right)}{\pi\left(\left(w_j^g\right)^{(m)}\right)}. \end{aligned}$$

- Sampling $\boldsymbol{\psi}$: this involves pV updates. For $j = 1, \ldots, p$ and $v = 1, \ldots, V$ we have that $Pr\left(\psi_j^v = 1 \;\middle|\; \boldsymbol{\psi}_{(-j)}^{(-v)}, \mathbf{w}_{(j)}, \phi, \mathbf{y}\right) \propto \theta\left(\psi_j^v = 1\right)$, with $\theta\left(\psi_j^v = 1\right) = \pi\left(\psi_j^v = 1 \;\middle|\; S_j^v = \mathbf{K}_v'(\phi)\mathbf{w}_{(j)}\right)(1+T)^{-\psi_j^v} M\left(\boldsymbol{\psi}^v\right)^{-(T+a_\sigma)}$, and $\boldsymbol{\psi}_{(-j)}^{(-v)}$

are all binary indicator variables ψ_l^u, $l \neq j$, $u \neq v$. Therefore,

$$\left(\psi_j^v \;\middle|\; \boldsymbol{\psi}_{(-j)}^{(-v)}, \mathbf{w}_{(j)}, \phi, \mathbf{y}\right) \sim Bernoulli\left(\frac{\theta\left(\psi_j^v = 1\right)}{\theta\left(\psi_j^v = 1\right) + \theta\left(\psi_j^v = 0\right)}\right).$$

- Sampling ϕ: Let $\xi = \log(\phi)$, then, the full conditional of ξ is

$$\begin{aligned} q\left(\xi \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y}\right) &= q\left(\phi(\xi) \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y}\right) e^{\xi} \\ &= Ga\left(e^{\xi} \mid a_\phi, b_\phi\right) \cdot e^{\xi} \cdot \prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_j^v \;\middle|\; S_j^v\left(\mathbf{K}_v(e^{\xi}); \mathbf{w}_{(j)}\right)\right). \end{aligned}$$

We use a normal random walk proposal on ξ, i.e., $\xi^* \sim N\left(\xi^{(m)}, \sigma_\xi^2\right)$, where $\xi^{(m)}$ is current state value and σ_ξ^2 is a tuning parameter. Then, the Metropolis-Hastings ratio is given by

$$\begin{aligned} \frac{q\left(\xi^* \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y}\right)}{q\left(\xi^{(m)} \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y}\right)} &= \frac{\prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_j^v \mid S_j^v\left(\mathbf{K}_v(e^{\xi^*}); \mathbf{w}_{(j)}\right)\right)}{\prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_j^v \mid S_j^v\left(\mathbf{K}_v(e^{\xi^{(m)}}); \mathbf{w}_{(j)}\right)\right)} \\ &\quad \times \frac{Ga\left(e^{\xi^*} \mid a_\phi, b_\phi\right) \cdot e^{\xi^*}}{Ga\left(e^{\xi^{(m)}} \mid a_\phi, b_\phi\right) \cdot e^{\xi^{(m)}}}. \end{aligned}$$

For the general case that $\phi = \left(\phi^1, \ldots, \phi^G\right)$, each ϕ^g can be updated using a scheme similar to the one described above.

Although this algorithm works, we improve the mixing considerably by sampling the parameters in blocks and also use the empirical covariance matrix from posterior

samples in previous iterations to generate proposal candidates. Once M posterior samples are obtained, $\left\{\left(\boldsymbol{\psi}^{(m)}, \mathbf{w}^{(m)}, \phi^{(m)}\right)\right\}_{m=1}^{M}$, we can estimate the posterior probabilities of activation for each voxel and each task, $Pr\left(\psi_j^v = 1 \,\middle|\, \mathbf{y}\right)$, by computing $\hat{p}_j^v = \frac{1}{M}\sum_{m=1}^{M}\left(\psi_j^v\right)^{(m)}$.

5.3 Simulation studies

We show the performance of the proposed complex-valued spatial methods for detecting activation in two simulation studies. We compare three methods for CV-fMRI data, namely, the two proposed spatial models referred to as the CV-KC and CV-GP models, and the non-spatial C-EMVS approach proposed by Yu et al. (2018). The results obtained using exclusively magnitude fMRI and magnitude-only spatial models, referred to as MO-KC and MO-GP[2], and the non-spatial MO-EMVS are also compared. In fact, the synthetic CV-fMRI data generated in the simulation study below was obtained using a complex-valued version of the data generating process used in a simulation study presented in Bezener et al. (2017).

5.3.1 Simulated data

We simulated 10 datasets consisting of 20×20 CV-fMRI slices with a constant baseline signal and a single expected BOLD signal (i.e., $p = 1$) resulting from the convolution of a stimulus indicator function and the canonical HRF. Three activation regions were simulated using the function `specifyregion` in the `R` package `neuRosim` (Welvaert et al., 2011). Specifically, for $v = 1, \ldots, 20 \times 20$, and $t = 1, \ldots, 50$, the time

[2]MO-GP is in fact the model proposed by Bezener et al. (2017).

series for each voxel v were simulated as follows:

$$
\begin{aligned}
y^v_{t,Re} &= (\beta_{v,0} + \beta_{v,1} x_t) \cos(\alpha_0) + \eta^v_{t,Re}, \quad \eta^v_{t,Re} \sim N\left(0, \sigma^2\right), \\
y^v_{t,Im} &= (\beta_{v,0} + \beta_{v,1} x_t) \sin(\alpha_0) + \eta^v_{t,Im}, \quad \eta^v_{t,Im} \sim N\left(0, \sigma^2\right),
\end{aligned}
\tag{5.4}
$$

where x_t is the expected BOLD signal, which is obtained as a convolution of a given HRF and a stimulus indicator function (Lindquist, 2008). We used $\alpha_0 = \pi/4$ and different values of β_0 and β_1 to examine the performance of the proposed CV-KC and CV-GP models. These were chosen to set specific values of the SNR and the CNR as defined in Rowe & Logan (2004), with SNR $= \beta_0/\sigma$ and CNR $= \beta_1/\sigma$.

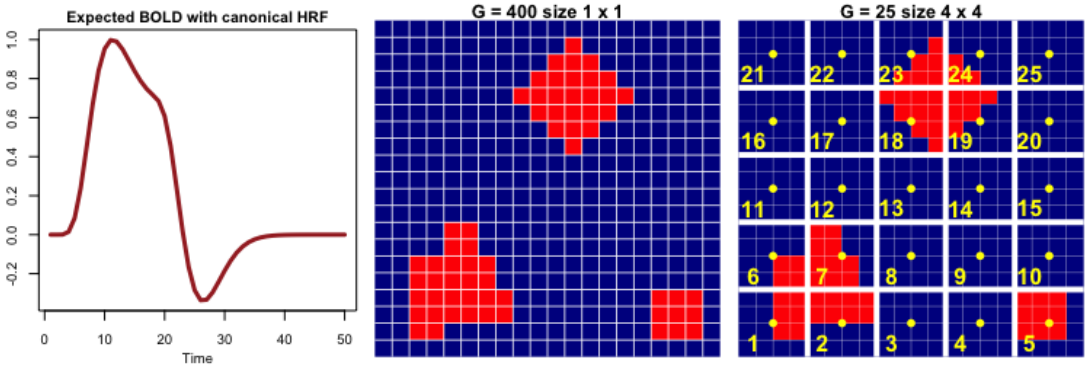


Figure 5.2: Left: Design matrix of the expected BOLD signal with the canonical HRF used in the simulation. Middle: Activation map of a 20×20 image. Right: Activation map with white spatial region grid lines. The number of regions is $G = 25$. Each region is of size 4×4 containing 16 voxels. Yellow dots represent the centroids of regions. Yellow numbers are region numbers.

The expected BOLD signal, true activation map, and and a parcellation of the image with $G = 25$ regions are shown in Figure 5.2.[3] The expected BOLD signal is the convolution of the canonical HRF with a 0-1 stimulus function that is on for $1 \leq t \leq 16$ and off for $17 < t \leq 50$. Here, we consider four different combinations of $(\beta_{v,0}, \beta_{v,1})$ that

[3]The simulated image is the same as the one in the benchmark example in Bezener et al. (2017).

correspond to high or low SNRs or CNRs, namely, **case LL:** $(\beta_{v,0}, \beta_{v,1}) = (1, 2)$ - low SNR and low CNR, **case LH**: $(\beta_{v,0}, \beta_{v,1}) = (1, 5)$ - low SNR and high CNR; **case HL**: $(\beta_{v,0}, \beta_{v,1}) = (300, 2)$ - high SNR and low CNR; and **case HH**: $(\beta_{v,0}, \beta_{v,1}) = (300, 5)$ - high SNR and high CNR. Case **HH** is in fact one of the coefficient settings in Bezener et al. (2017). However, their model does not include the phase angle $\alpha_0 = \pi/4$ as shown in Eq. (5.4) and considers only one real-valued time course of signals. That is, the simulated data are generated from

$$y_t^v = \beta_{v,0} + \beta_{v,1} x_t + \epsilon_t^v, \quad \epsilon_t^v \sim N\left(0, \sigma^2\right).$$

Also, in order to compare our results to those in Bezener et al. (2017) we also set $\sigma^2 = 3$.

5.3.2 Fitted complex-valued models

We fit the CV-KC and CV-GP spatial models proposed in Section 5.1, and the non-spatial C-EMVS model proposed in Yu et al. (2018) to the simulated data. We measured the performance of these models for detecting brain activation at voxel level by measuring the sensitivity, specificity, precision and accuracy in the different simulation settings. Specifically, for $v = 1, \ldots, V$, with $V = 400$, $T = 50$, $G = 25$ and

$p = 1$, we considered a CV-KC model of the form

$$\begin{aligned}
\mathbf{y}^v &= \mathbf{x}\left(\psi^v\right)\gamma^v\left(\psi^v\right) + \boldsymbol{\eta}^v, \quad \boldsymbol{\eta}^v \sim CN_T\left(\mathbf{0}, 2\sigma_v^2\mathbf{I}_T, \mathbf{0}\right), \\
\gamma^v\left(\psi^v\right) \mid \psi^v, \sigma_v^2 &\overset{\text{ind}}{\sim} CN_1\left(\hat{\gamma}_v\left(\psi^v\right), 2T\sigma_v^2\left(\mathbf{x}'\left(\psi^v\right)\mathbf{x}\left(\psi^v\right)^{-1}\right), 0\right), \\
\psi^v \mid S^v &\overset{\text{ind}}{\sim} Bernoulli\left(\frac{1}{1+e^{-S^v}}\right), \quad S^v = \sum_{g=1}^{G} k\left(z_v - s_g; \phi\right) w^g, \\
w^g \mid \tau^2 &\overset{\text{ind}}{\sim} N\left(0, \tau^2\right), \quad \tau^2 \sim IG\left(1/2, 1/2\right), \\
\sigma_v^2 &\overset{\text{iid}}{\sim} IG\left(1/2, 1/2\right), \quad \phi \sim IG\left(1/2, 1/2\right),
\end{aligned}$$

where $\hat{\gamma}^v\left(\psi^v\right) = \left(\mathbf{x}'\left(\psi^v\right)\mathbf{x}\left(\psi^v\right)\right)^{-1}\mathbf{x}\left(\psi^v\right)\mathbf{y}^v$, z_v denotes the coordinates of voxel v and s_g denotes the centroid coordinates of region g. The centroids for the regions in the parcellation are shown as a yellow dots in Figure 5.2. Without any prior information, we assume that $a^g = 0$ for all g. Note that in this simulation study there is only one task ($p = 1$), so we removed the subindex j from the notation.

The CV-GP model has the same likelihood and the same priors on γ^v and σ_v^2. The rest of the model setup is given by

$$\psi^v \mid S^g \overset{\text{ind}}{\sim} Bernoulli\left(\frac{1}{1+e^{-S^g}}\right), \quad \left(\boldsymbol{S} \mid \delta^2, r\right) \overset{\text{ind}}{\sim} N\left(0, \delta^2\Gamma\right), \quad \pi\left(\delta^2\right) \propto \delta^{-2}, \quad r \sim \chi_8^2,$$

where Γ is the exponential covariance function described in the previous section. The degrees of freedom in the χ^2 distribution are set to 8 for comparison with Bezener et al. (2017).

The C-EMVS model has the same likelihood as the CV-KC and CV-GP

models. The rest of the model specification is given by

$$\gamma^v \mid \psi^v \sim (1-\psi^v)\, CN_1\left(0, 2v_0\sigma_v^2, 0\right) + \psi^v CN_1\left(0, 2\sigma_v^2, 0\right),$$

$$\sigma_v^2 \sim IG\left(1/2, 1/2\right), \quad \psi^v \mid \theta \sim Bernoulli\left(\theta\right), \quad \theta \sim Beta\left(1,1\right).$$

Note that although there is no explicit spatial structure in this model, activation is inferred by borrowing information across voxels through a common probability of activation for all voxels in the Bernoulli prior. The tuning parameter v_0 is chosen from a grid of values so that the optimal v_0 maximizes $p\left(\boldsymbol{\psi} \mid \mathbf{y}\right)$ that evaluates $\boldsymbol{\psi}$ containing those variables for which $\psi^v = 1$. That is, the optimal v_0 leads to the most likely activation pattern given the observed data.

For the MCMC algorithms of the KC and GP models, 51,000 samples were drawn. After burning the first 1,000 samples and thinning the sequences by keeping every 25th draw, 2,000 draws were used for analysis. For the EM algorithm, we iterated the expectation and maximization steps until the Euclidean distance between the current and new parameter values was less than 0.001, that is, $\|\boldsymbol{\theta}^{(l)} - \boldsymbol{\theta}^{(l-1)}\| < 10^{-3}$, where $\boldsymbol{\theta}^{(l)}$ denotes the full parameter vector at the lth iteration.

5.3.3 Results of the analysis of the simulated data

In order to examine the performance of the different methods, we consider two threshold values for detecting activation, namely, 0.8722 and 0.5. A threshold value of 0.5 corresponds to the so called "median probability model", i.e., the model that labels the sites as active if they have posterior probabilities of activation higher than 0.5. This

model has been shown to be the optimal predictive model under certain assumptions for the squared error loss (Barbieri & Berger, 2004). A threshold value of 0.8722 has been used in some Bayesian spatial approaches for magnitude-only fMRI data (Smith & Fahrmeir, 2007; Lee et al., 2014). This threshold corresponds to a p-value 0.05 for the likelihood ratio statistic $-2\log\left(\left(1-Pr\left(\psi^v=1\mid\mathbf{y}\right)\right)/Pr\left(\psi^v=1\mid\mathbf{y}\right)\right)$ (Raftery, 1996). Alternative threshold values can be used. For example, a threshold value can be determined by a Bayesian decision theoretic approach. Müller et al. (2004) and Müller et al. (2006) show that when the loss function being considered is of the form $c\mathrm{E}\left(FD\mid\mathbf{y}\right)+\mathrm{E}\left(FN\mid\mathbf{y}\right)$, where FD is false discoveries and FN is false negatives, the optimal threshold value that minimizes the loss function is $\frac{c}{1+c}$. Note that for $c=1$, the optimal threshold is 0.5. In simulation settings, the performance with different threshold values for detecting activation can be examined by looking at the resulting receiver operating characteristic (ROC) curves as illustrated below.

Figures 5.3 and 5.4 show the activation performance measures for the complex-valued models we considered (the spatial CV-KC and CV-GP models, and the non-spatial C-EMVS) using, respectively, threshold values of 0.8722 and 0.5 for the spatial models. A threshold value of 0.5 was used for the non-spatial C-EMVS approach. When a threshold value of 0.8722 is used for detecting activation in the spatial models and a threshold value of 0.5 is used for the non-spatial C-EMVS, the C-EMVS has larger sensitivity values than those obtained with the spatial models in the cases with low CNR. However, both spatial models have higher precision and specificity than the the non-spatial model, with the CV-KC dominating both, CV-GP and C-EMVS. On

the other hand, when a threshold value for activation of 0.5 is used for all the models, the proposed CV-KC method clearly dominates the other two approaches, especially in terms of sensitivity and accuracy, as shown in Figure 5.4.

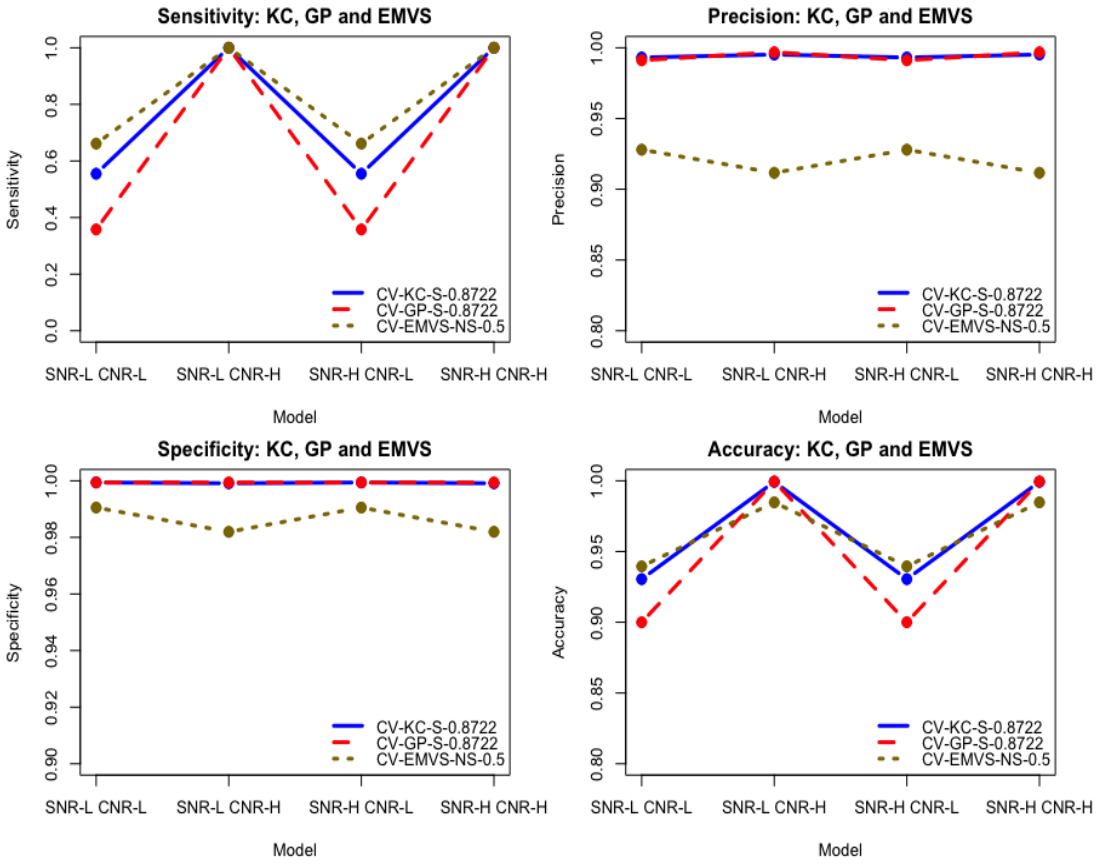


Figure 5.3: Performance measures of CV models under the spatial KC (blue-solid, CV-KC) priors, the spatial GP (red-dashed, CV-GP) priors, and the non-spatial EMVS (brown-dotted, CV-EMVS). Top-left: Sensitivity. Top-right: Precision. Bottom-left: Specificity. Bottom-right: Accuracy. A threshold value of 0.8722 is used for detecting active voxels in CV-KC and CV-GP, and a threshold value of 0.5 is used for the non-spatial C-EMVS.

As mentioned above, a threshold value of 0.8722 is commonly used for spatial models, however, it is not clear that this is the best threshold for complex-valued models in terms of balancing out false positives and false negatives. We also note that

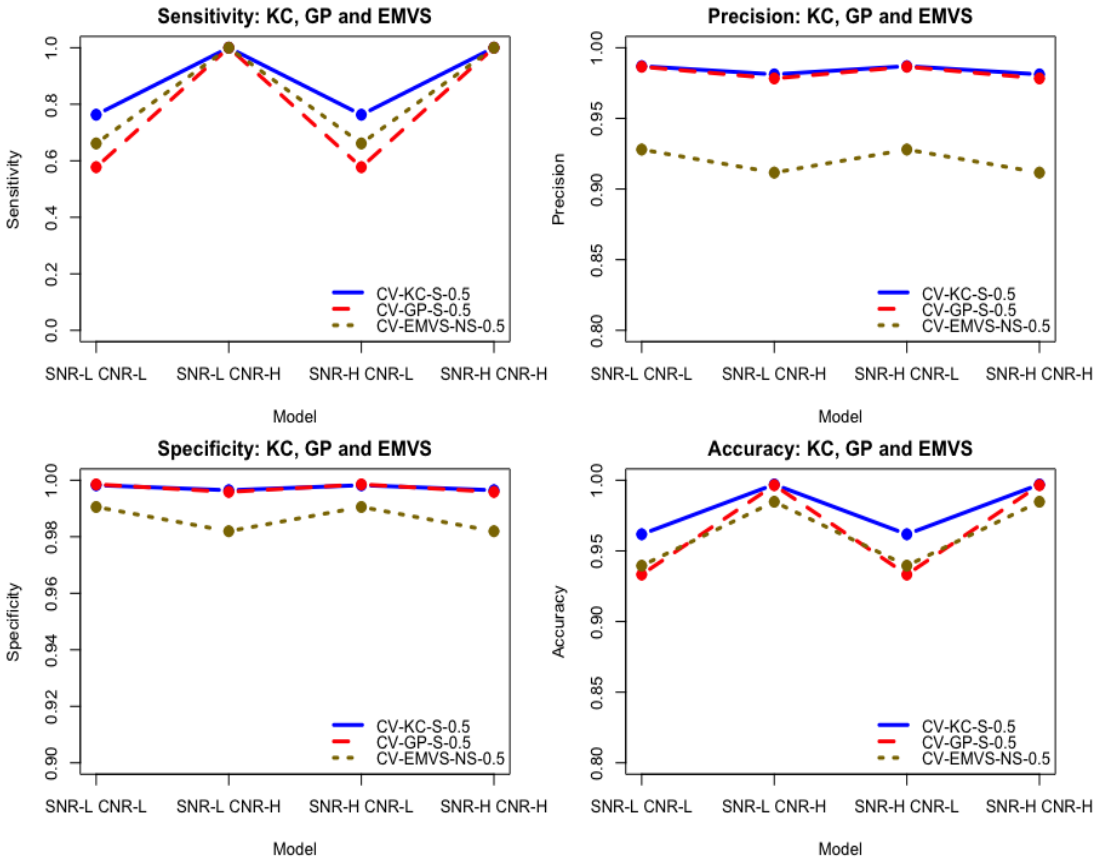


Figure 5.4: Performance measures of CV models under the spatial KC (blue-solid, CV-KC) priors, the spatial GP (red-dashed, CV-GP) priors, and the non-spatial EMVS (brown-dotted, CV-EMVS). Top-left: Sensitivity. Top-right: Precision. Bottom-left: Specificity. Bottom-right: Accuracy. A threshold value of 0.5 is used for all three models.

under the threshold value of 0.5 the GP model does not really improve in terms of the performance measures, showing values similar to those obtained with the 0.8722 threshold. Figure 5.5 shows the ROC curves for the three complex-valued models under the low CNR scenario for 2 different SNR scenarios. We see that the CV-KC method has the largest area under the ROC curve, showing that the CV-KC model leads to the best results in terms of detecting activation at the voxel level.

Figure 5.6 shows the estimated weights of the region-specific spatial effects

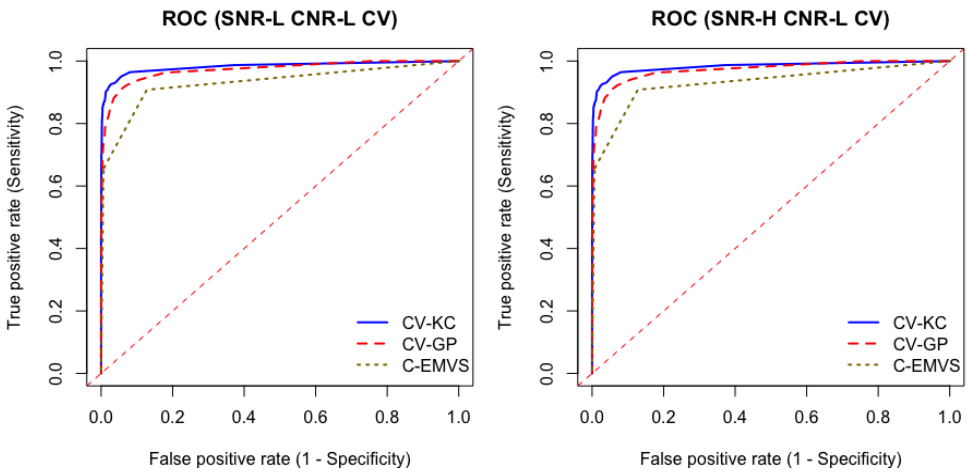


Figure 5.5: ROC curve under case LL (left) and case HL (right) for the CV models: CV-KC (blue-solid), CV-GP (red-dash) and C-EMVS (brown-dot).

w^1, ..., w^{25} on each voxel in the CV-KC model. Specifically, in this example with 25 regions, the voxel-level spatial effect is given by

$$S^v = \sum_{g=1}^{25} k\left(z_v - s_g; \phi\right) w^g = k\left(z_v - s_1; \phi\right) w^1 + \cdots + k\left(z_v - s_{25}; \phi\right) w^{25},$$

and hence the weights of the effects $w^1, \ldots, w^{25}$ are column vectors $\{\mathbf{K}_{v,1}\}_{v=1}^{400}, \ldots,$ $\{\mathbf{K}_{v,25}\}_{v=1}^{400}$ of the 400×25 kernel matrix $\mathbf{K}$. Figure 5.6 shows each of these columns in an image. Since a common ϕ is used for all spatial regions, the radii of the kernel effects are the same. The effects displayed in the plot are obtained by evaluating the Bezier kernel function $k\left(z_v - s_g; \tilde{\phi}\right)$ at the posterior mean of ϕ, $\tilde{\phi}$. Then, this figure shows how a given voxel is related to other voxels in the image. In general, voxels in a given region g are mostly influenced by the spatial effect of the same region, as indicated in the picture, but they are also affected by the spatial effects of other regions with the kernel values shown in the picture acting as weights for the different effects.

For example, voxels in region 2 (see Figure 5.2 for region numbers) are heavily affected by w^2 with the largest weights given by $k\left(z_v - s_2; \tilde{\phi}\right), v \in \mathcal{R}_2$. This is shown in the image corresponding to Column 2 of K in Figure 5.6. However, voxels in region 2 are also impacted by the spatial effects from regions 1, 3, 6, 7, and 8 with smaller weights $k\left(z_v - s_g; \tilde{\phi}\right), v \in \mathcal{R}_2$, for $g = 1, 3, 6, 7, 8$.

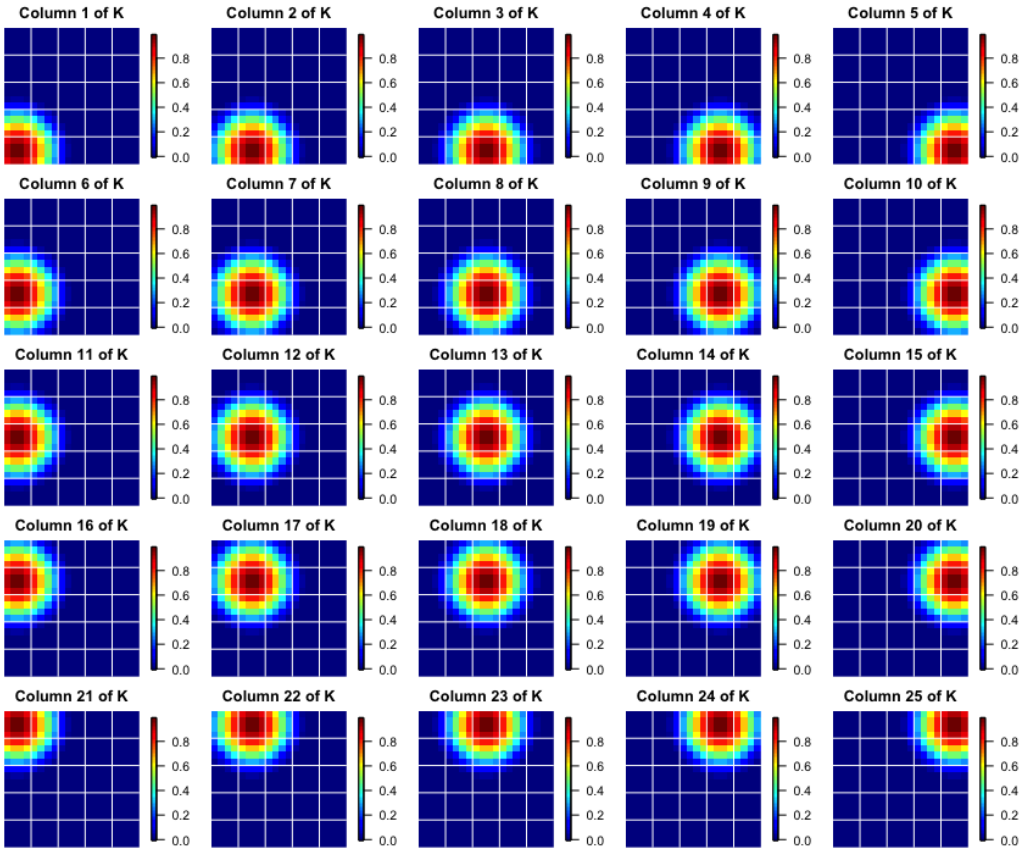
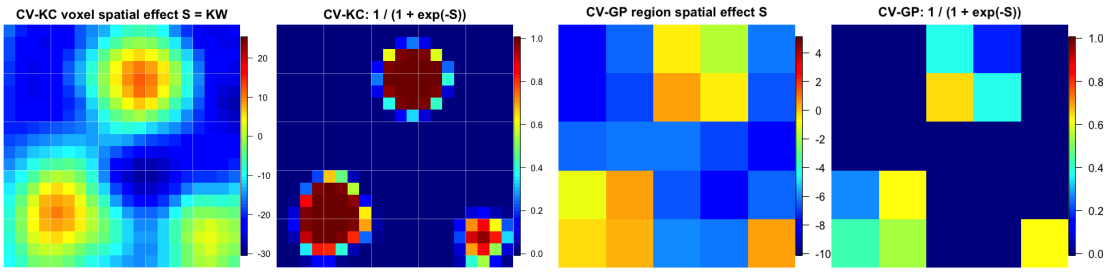


Figure 5.6: Kernel effect with the kernel function $k\left(z_v - s_g; \tilde{\phi}\right)$ evaluated at the posterior mean of ϕ, $\tilde{\phi}$. Each image shows the effect of spatial region g.

The CV-KC model produces a higher resolution latent spatial process S^v at the voxel level than that obtained by the GP model, as shown in Figure 5.7. The posterior voxel-level spatial effect $\boldsymbol{S}$ is computed via $\mathbf{K}\left(\tilde{\phi}\right)\tilde{\mathbf{w}}$, where $\tilde{\phi}$ and $\tilde{\mathbf{w}}$ are the

posterior means of ϕ and $\mathbf{w}$, respectively. The spatial process map correctly identifies three activation areas. The probabilities of activation are obtained from the posterior means of the spatial effects via the logistic transformation. We see that these estimated probabilities of activation are close to one for voxels that are truly activated. On the other hand, the GP model does not allow us to infer activation at the voxel level by simply examining its resulting spatial process. As shown in Figure 5.7 (b), voxels in the same region share the same spatial effect. This is not realistic unless there is a clever way to parcellate the image so that voxels in the same region behave similarly and hence are all active or non-active. In fact, these region-specific probabilities can be seen as the overall mean probability of activation in a region g. Take region 18 for example, the probability of activation for this region is estimated at the posterior mean of S^{18}, $\tilde{S}^{18}$, as $1/(1+e^{-\tilde{S}^{18}}) \approx 0.626$. This number is similar to the proportion of activated voxels in the region 18, $10/16 = 0.625$. Hence, based on the region-specific spatial effects, the GP model can tell us that there are approximately 10 voxels out of 16 being activated in region 18, but it cannot further specify which voxels are more likely to be activated as there is no voxel-level information for this model except the posterior probabilities $\hat{p}^v$. Later, we will show that if magnitude-only data are used, the GP model tends to inflate its region-level spatial effects, and potentially lead to more false positives.

Given the fact that the CV-GP model forces voxels in the same region to share the same spatial effect, its detection performance is very sensitive to the number of spatial regions and to how the image is parcellated. In general, the larger G is,



(a) CV-KC voxel-level spatial effect (b) CV-GP region-level spatial effect

Figure 5.7: Posterior mean of spatial random effects and corresponding logistic transformation for the CV-KC and CV-GP under the case HH.

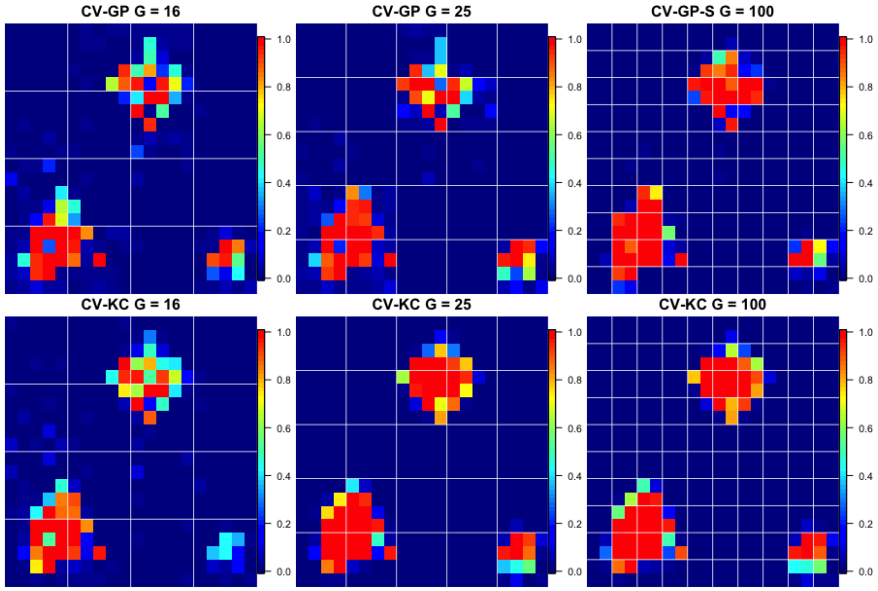


Figure 5.8: Effect of G under the case LL. Top: posterior probability maps of CV-GP. Bottom: posterior probability maps of CV-KC. From left to right are maps with $G = 16, 25, 100$, respectively.

the better detection we can have. Figure 5.8 shows how activation probability maps change with the number of spatial regions G. Notice that the CV-KC model is much less affected by G. In fact, the CV-KC model with $G = 25$ can lead to the same or even better activation detection than the CV-GP model with $G = 100$. This matters because more spatial regions results in longer computing time. To examine how the

G	4	16	25	100
CV-GP	0.469	0.587	0.950	9.809
CV-KC	0.607	0.908	1.198	3.510

Table 5.1: Average computing time of running the CV-GP and CV-KC MCMC algorithms 10 times. The time unit is seconds per 1000 MCMC iterations.

number of spatial regions G affects the computation efficiency of CV-GP and CV-KC, we run their MCMC algorithm 10 times each, and the their average computing time on the simulated data with $V = 400$, $T = 50$ and $p = 1$ under case LL is shown in Table 5.1. The time unit used in the table is seconds per 1000 MCMC iterations. The algorithms are coded in `R` version 3.3.1 environment with `C++` code embedded using the `Rcpp` package. The algorithms were run on the operating system OS X Yosemite 10.10.5 with the 1.3 GHz Intel Core i5 processor. Notice that the computing time of the CV-GP model grows faster than the computing time of the CV-KC model as G gets larger. For these data, when $G = 100$, we see that the CV-GP model essentially takes almost 2.80 times longer to obtain the same type of activation results obtained by the CV-KC model.

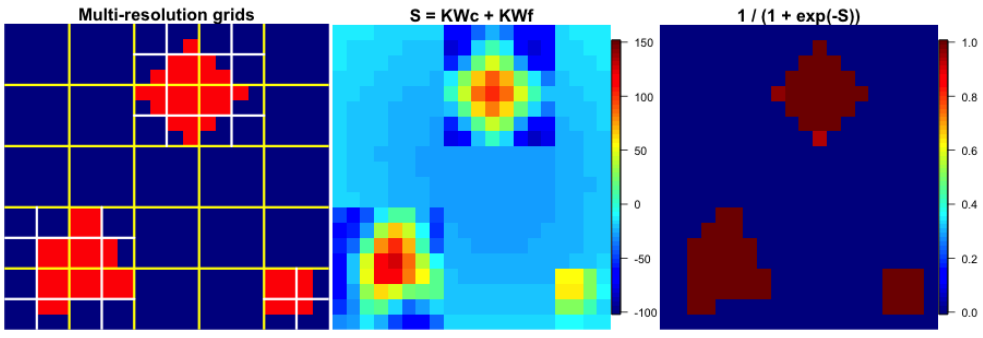


Figure 5.9: Multi-resolution spatial effects of CV-KS. From left to right: (1) Two resolution grids. Coarse grids are colored yellow and fine grids are colored white. (2) The posterior mean of overall spatial effects, i.e., the sum of spatial effects from coarse grids and spatial effects from fine grids. (3) The logistic transformation of the overall spatial effects.

Another advantage of the kernel method is that it allows for multi-resolution approaches. For example, one can use finer grids on areas that are more likely to be activated. Also, kernels used in coarse and fine grids could be different to capture different activation patterns. The mathematical formulation is straightforward. Suppose that G regions are created using a coarse grid and H regions are created using a fine grid. Denote the "coarse" kernel functions by $k_c\left(z_v - s_g, \phi_c\right)$ and the "fine" kernel functions by $k_f\left(z_v - d_h, \phi_f\right)$, respectively, where s_g denotes the centroid of each of the regions in the the coarse grid, and d_h denotes the centroid of each region in the fine grid. Then the voxel-level overall spatial effect is

$$S_j^v = \sum_{g=1}^{G} k_c\left(z_v - s_g; \phi_c\right) w_j^g + \sum_{h=1}^{H} k_f\left(z_v - d_h; \phi_f\right) u_j^h = S_{c,j}^v + S_{f,j}^v,$$

where u_j^h is another Gaussian variable independent of w_j^g. The posterior sampling scheme for this multi-resolution model is essentially the same as the scheme used for the model with only one resolution described in the previous section. Figure 5.9 shows how multi-resolution works. In this example, we put a finer grid on the regions that include the true activated voxels. This is not uncommon, as in many practical scenarios neuroscientists have some knowledge about the areas of the brain where they expect to see active voxels in a particular experimental setting. The middle picture shows the posterior mean of the overall voxel-level spatial effects. Unlike the spatial effects in Figure 5.7 (b), the spatial effects from the coarse grid become a very smooth and flat layer because all possible ups and downs are captured by the spatial effects from the fine grid. The estimated probability activation map obtained by computing the

Kernel type	Sensitivity	Specificity	Precision	Accuracy
Bezier $\nu = 1$	0.7387	0.9985	0.9893	0.9582
Bezier $\nu = 2$	0.7629	0.9982	0.9870	0.9618
Bezier $\nu = 3$	0.7790	0.9979	0.9858	0.9640
Bezier $\nu = 5$	0.7677	0.9982	0.9877	0.9625
Gaussian	0.7516	0.9982	0.9875	0.9600

Table 5.2: Performance measures of CV-KC under the Bezier kernel with $\nu = 1, 2, 3$ and 5 and the Gaussian kernel. The threshold value is 0.5.

logistic transformation of the overall spatial effects shows values near one for activated voxels and essentially zero for non-activated voxels. This shows that properly using a multi-resolution scheme can further improve activation detection without dramatically increasing the computation time.

5.3.4 Effect of smooth parameter ν and comparison to a convolution model with Gaussian kernels

We examine the effect of the smooth parameter ν in the Bezier kernel function by detecting activation performance and recovering the latent spatial process when $\nu = 1, 2, 3$ and 5. With a fixed value of parameter ϕ, smaller value of ν leads to larger degree of smoothness. Table 5.2 shows that the kernel method is not so sensitive to what value of ν is used in the model. This is because the KC model can learn the value ϕ, balance the effect of ν and ϕ, and hence produce a kernel effect that is best for detecting activation.

When a Gaussian kernel $k(s_v - u_g; \phi) \propto \exp\{-\|s_v - u_g\|^2/2\phi\}$ is used, the latent spatial process will be smoother than the spatial process generated by the Bezier kernel, but its performance measures do not change much, as shown in Table 5.2.

5.3.5 Comparison across magnitude-only models

Finally, we assess the effect of including a spatial structure in models for detecting brain activation from magnitude-only data. For this purpose, we obtained the magnitude-data from the complex-valued data simulated as described above and fitted three magnitude-only models, namely, a kernel-based spatial model (MO-KC), a spatial Gaussian process model (MO-GP) and the magnitude-only EMVS model of Yu et al. (2018) with no spatial structure (MO-EMVS). Note that by taking the magnitude data from the complex-valued data simulated above we are altering the SNR and CNR values, and so, the results obtained from magnitude-only models are not directly comparable to the results obtained under complex-valued models discussed above. However, we can still make comparisons across magnitude-only models as the magnitude-data are the same.

Figure 5.10 shows four different performance measures for detecting brain activation, namely, sensitivity, precision, specificity, and accuracy in the three magnitude-only models that we considered, i.e., MO-KC, MO-GP and MO-EMVS. Figure 5.11 shows the ROC curves for the magnitude-only models under two simulated scenarios. Once again the kernel-based spatial model dominates the other two models in terms of the ROC curves. The kernel-based spatial model improves sensitivity and accuracy when SNR is high and CNR is low. The higher specificity and precision for both spatial models indicate that introducing a spatial structure when analyzing magnitude-only fMRI reduces false positives. Note, however, that the spatial GP model has lower power for detecting activation than our non-spatial variable selection model when CNR is low,

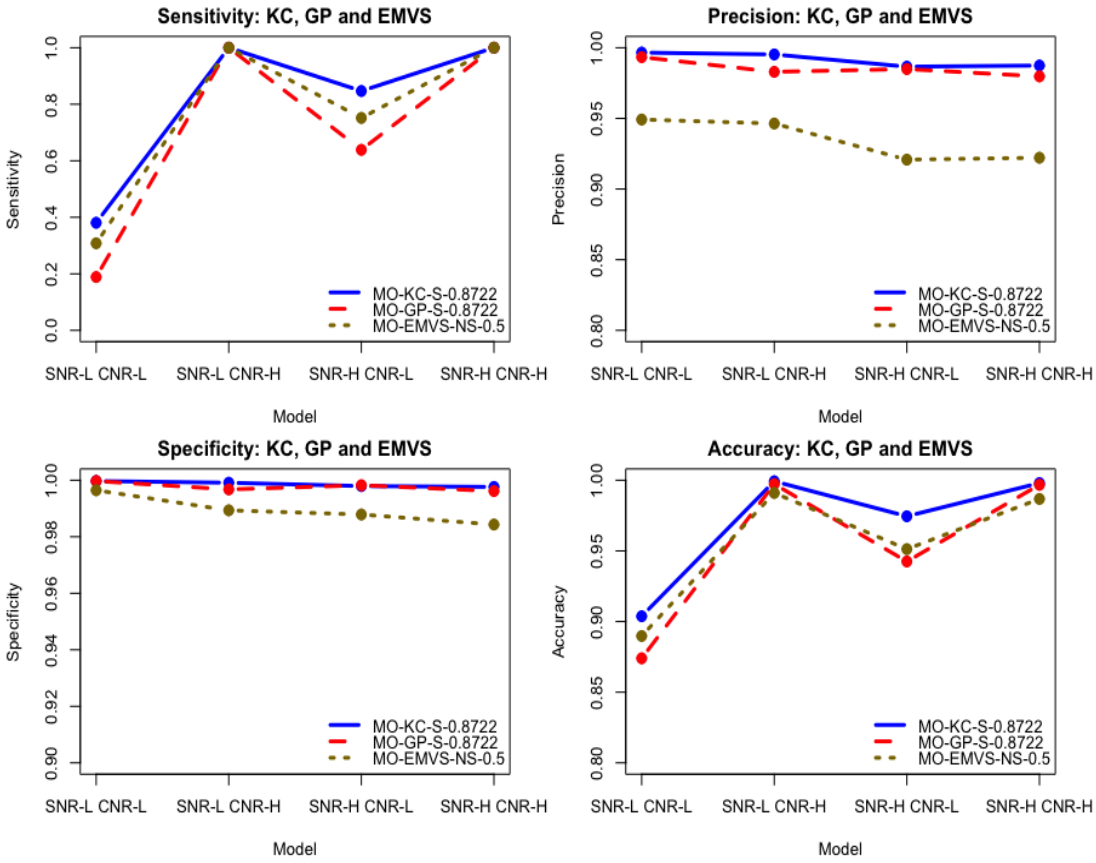


Figure 5.10: Performance measures of MO models under KC (blue-solid), GP (red-dash) and EMVS (brown-dot). Top-left: Sensitivity. Top-right: Precision. Bottom-left: Specificity. Bottom-right: Accuracy. y-axis is adjusted to magnify the differences between models. Threshold 0.8722 is used for the KC and GP models, and 0.5 for the non-spatial EMVS model.

i.e., when activation strength is weak. This is because the MO-GP model induces a probability inflation effect, i.e., if a spatial region contains true activated voxels, other non-activated voxels in the region will also have a relatively high probability of activation. One can clearly see this inflation effect in the estimated posterior probability map shown in Figure 5.12 (a). This effect increases the chance of false positives around true activated voxels, and so, the detection of active voxels for the spatial GP model is sensitive to the choice of the threshold value for detecting activation. Even though

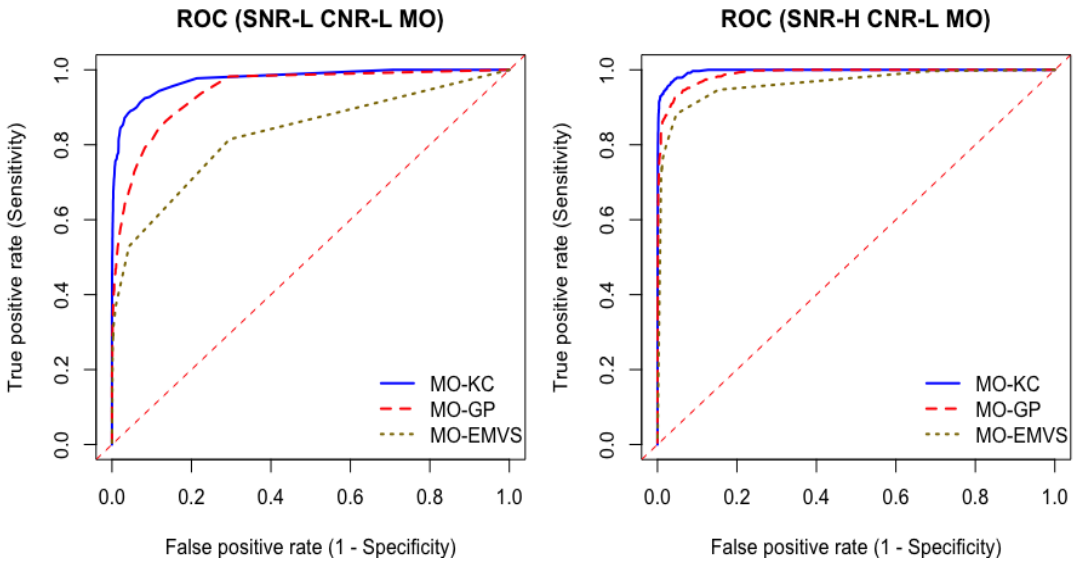
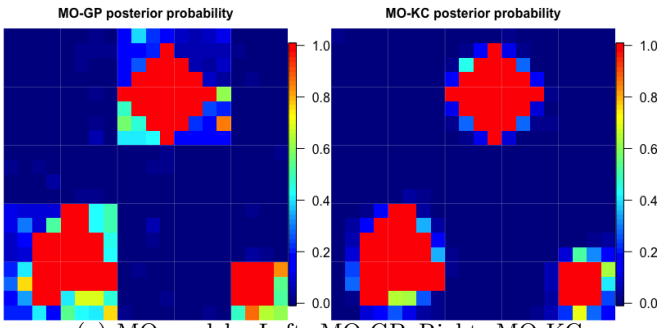


Figure 5.11: ROC curve under case LL (left) and case HL (right) for the MO models: MO-KC (blue-solid), MO-GP (red-dash) and MO-EMVS (brown-dot).

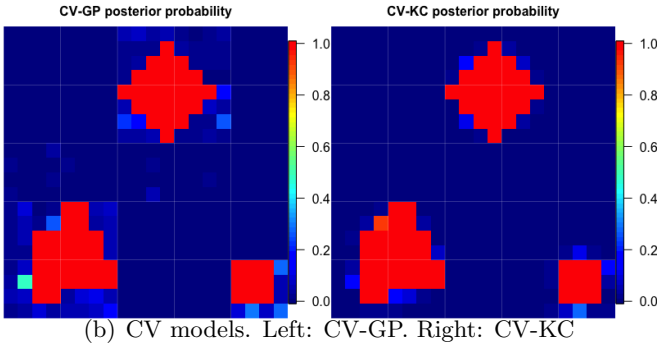
those voxels are not misclassified as activated in this example, this kind of map is not reasonable because two neighboring voxels that behave similarly may have significantly different probabilities of activation due to their spatial region assignment. On the contrary, the KC spatial model does not have this unfavorable effect and produces a more reasonable probability map. Note also that the probability inflation effect for the GP model is much smaller when complex-valued data and models are used, as one can see in Figure 5.12 (b).

5.4 Analysis of human CV-fMRI data

We also analyzed human data recorded during an fMRI experiment performed on a single subject on a 3.0-T General Electric Signa LX magnetic resonance image scanner. The experiment consisted of a unilateral finger-tapping task performed with



(a) MO models. Left: MO-GP. Right: MO-KC



(b) CV models. Left: CV-GP. Right: CV-KC

Figure 5.12: Posterior probability maps for GP and KC models. The maps are generated from a high SNR and high CNR simulated data set.

a visual cue indicating whether to tap or rest. A block designed experiment with an initial 20 s of rest followed by 16 epochs of 15 s on and 15 s off was used. The full dataset is composed of seven 2.5 mm thick axial slices of dimension 96×96. A single slice was used for the analysis presented here. Further details about the experiment, the data, and previous analyses of these data are found in Karaman et al. (2014) and Yu et al. (2018). The original time series at each voxel had 510 TRs, however, following the approach of Karaman et al. (2014), we discarded the first 20 observations for the analysis. Activation from this finger-tapping experiment is well-studied. Our goal here is to demonstrate that our novel Bayesian complex-valued spatial method is able to simultaneously produce activation results that are consistent

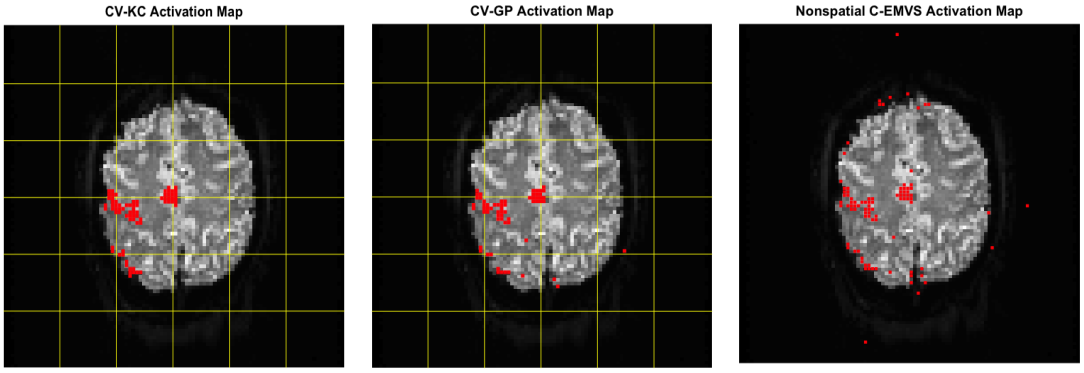
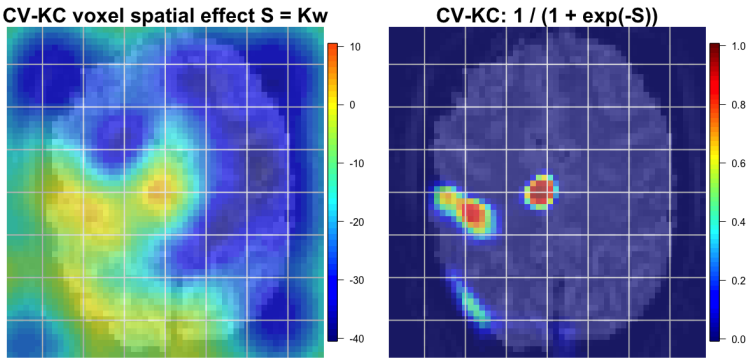


Figure 5.13: Activation of human subject CV-fMRI. Left: Activation map from KC-S with threshold 0.8722. Middle: Activation map from CV-GP with threshold 0.5. Right: Activation map from non spatial EM with threshold 0.5.

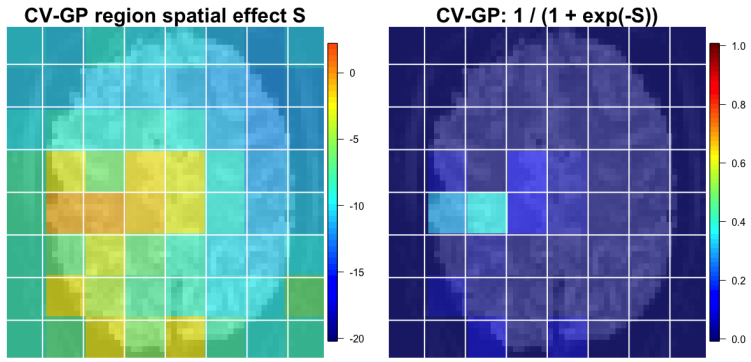
with previous analyses with the additional benefit of reducing of spurious activations, such as detecting activations outside the brain or in regions that are not implicated in the finger tapping task.

Yu et al. (2018) have shown that their C-EMVS approach removes many false positives outside the brain area and improves detection of activation in the expected motor-related brain regions with respect to results obtained with a complex-valued model proposed in Karaman et al. (2014). Here, we further improve activation results by using our proposed spatial models. We begin by parcellating the image into 36 equal-sized squared spatial regions containing 36 voxels each, and then fit the CV-KC and CV-GP models to the complex-valued human subject data set. Figure 5.13 shows the activation maps derived from the CV-KC, CV-GP and C-EMVS models under a threshold value of 0.5 for detecting activation. Clearly, spatial models eliminate isolated voxels around the brain margin incorrectly labeled as active by the C-EMVS approach. In addition, spatial modeling encourages activation in groups or clusters

of voxels. Both, the CV-KC and CV-GP models detect more activated voxels in the expected motor-related brain regions located on the left of the brain. These results are biologically plausible because execution of complex cognitive tasks usually involve populations of neurons spanning across many voxels rather than a single voxel.



(a) CV-KC voxel-level spatial effect



(b) CV-GP region-level spatial effect

Figure 5.14: Posterior mean of the spatial random effects and their corresponding logistic transformations for the complex-valued human data: (a) results from the CV-KC model; (b) results from the CV-GP model.

In order to focus on a smaller portion of the image that reduces areas outside the brain, an image of size 56×56 was furthered examined. This image was parcellated into 64 equally-sized regions, each containing 49 voxels. Figure 5.14 shows the voxel-level spatial effects inferred by the CV-KC model and the region-specific spatial effects

inferred by the CV-KC and the CV-GP models for this reduced image. The CV-KC model generates smooth and localized maps of the estimated spatial effects that show the strength of the spatial effects adequately concentrated around the active regions. On the other hand, the maps of estimated spatial effects are neither smooth nor localized.

Figure 5.15 shows the posterior probability maps obtained from the posterior samples of $\boldsymbol{\psi}$. We note that for this particular data set, the activation results are not so different for the CV-KC and CV-GP approaches when threshold values of 0.5 or 0.8722 are used for detecting activation. However, once again, we see that the CV-KC model leads to probability maps that are less uncertain than the maps obtained from the CV-GP model, e.g., there are less voxels with probabilities different from zero around the central activated region for the CV-KC maps. This is a clear advantage of the CV-KC approach, as more precise posterior probability maps for activation lead to final activation results are less sensitive to the choice of the threshold value used for detecting activation.

5.5 Discussion

We propose a new Bayesian spatial model via kernel convolution for detecting brain activation from CV-fMRI signals at the voxel-specific level. Our simulation studies show that the Bayesian complex-valued spatial methods lead to more true positives and less false positives than models that do not include a spatial structure, especially when the CNR is small. This is due to the fact that complex-valued models,

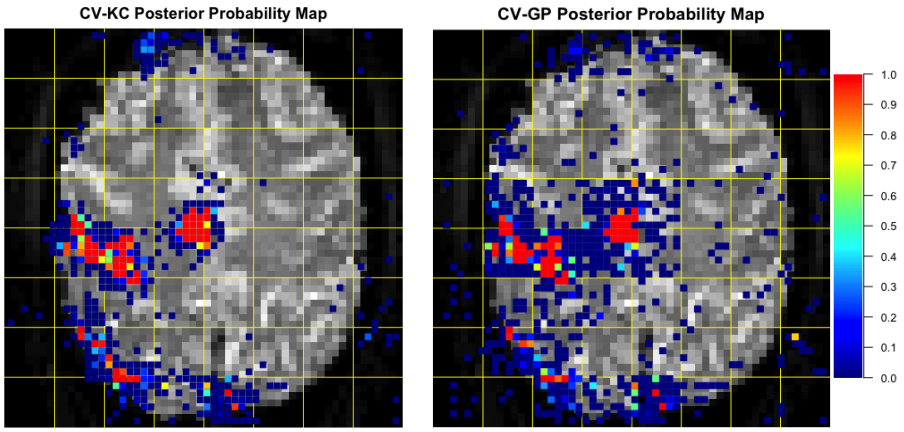


Figure 5.15: Posterior probability maps of human subject CV-fMRI. Left: CV-KC. Right: CV-GP. Uncolored voxels have posterior probability of activation numerically zero.

even those with no explicit spatial structure, use the additional information provided by the phase data to produce more accurate activation results, specially when the SNR is small.

We also see that our kernel-based complex-valued model outperforms the competing GP model by cleverly borrowing spatial information from neighboring voxels. Therefore, the CV-KC approach produces more precise posterior probability maps than the CV-GP model. Moreover, the CV-KC model is not very sensitive to the number of spatial regions, leading to dimension reduction and consequently decreasing the computational time for obtaining posterior inference. Future extensions of the models presented here will consider modeling functional connectivity through the a_j^g to capture long distance spatial dependencies.

Once again, we strongly recommend the use of CV-fMRI data and related models for detecting activation at the voxel level in fMRI task-related studies. Adding a spatial structure will definitely improve the performance of the complex-valued model

for detecting brain activation at the voxel level, but is computationally more costly than using a complex-valued model with no spatial structure. Having this in mind, we strongly advocate for the use of complex-valued models with probabilities of activation modeled using a spatial kernel-convolution approach for the following reasons. First, as shown in our simulation studies, these models lead to more flexible spatial modeling and more accurate estimated probability activation maps, and so, they are less sensitive to the choice of the threshold value that determines activation at the voxel level. Second, the complex-valued kernel-convolution models lead to dimension reduction and allow us to easily incorporate multi-resolution components to obtain more precise results without significantly increasing the computational efficiency.

Chapter 6

Multi-subject Bayesian spatial model for task-related CV-fMRI activation and connectivity

In this chapter, we generalize the spatial model proposed in the previous chapter into a multi-subject model. The model allows us not only to detect activation at the voxel level in multi-subject CV-fMRI, but also to infer connectivity by finding which areas of the brain are related in terms of activation. Therefore both, local spatial dependence and long distance connections between brain regions, can be captured by the multi-subject model.

6.1 Multi-subject model for CV-fMRI data

Given voxel-wise data, it is computationally challenging to estimate connectivity at the voxel level. Thus, common connectivity analyses focus on inferring the relationship between the fMRI time series from a given set of regions of interests or ROIs. In the previous chapter, the image is parcellated into several spatial regions and a local spatial structure across neighboring voxels is induced through these regions using Gaussian processes or via kernel convolutions.

In this chapter we propose multi-subjects models for detecting activation and connectivity that use two types of brain regions. One set of regions is defined through a relatively fine grid, as illustrated in the previous chapter, with the purpose of inducing local spatial structure across neighboring voxels. Then, another coarser set of regions, usually referred to as regions of interest or ROIs, is defined and our model uses the information provided by the multi-subject data to estimate brain connectivity by inferring co-activation patterns across these regions. The ROIs are typically provided by the neuroscientists, defined using a brain atlas or via prior information provided by other imaging modalities such as structural MRI of diffusion tensor imaging (DTI).

6.1.1 Multi-subject hierarchical model

We generalize the spatial models proposed in the previous chapter, i.e., CV-KC and CV-GP into multi-subject models. Specifically, for subject $s = 1, \ldots, n$, time $t = 1, \ldots, T$, voxel $v = 1, \ldots, V$, local spatial region $g = 1, \ldots, G$, ROI $d = 1, \ldots, D$

and task $j = 1, \ldots, p$, our full complex-valued multi-subject KC model[1] is

$$
\begin{aligned}
\mathbf{y}^v_{(s)} &= \mathbf{X}\left(\boldsymbol{\psi}^v_{(s)}\right)\boldsymbol{\gamma}^v_{(s)}\left(\boldsymbol{\psi}^v_{(s)}\right) + \boldsymbol{\eta}^v_{(s)}, \quad \boldsymbol{\eta}^v_{(s)} \sim CN\left(0, 2\sigma_v^2\mathbf{I}, \mathbf{0}\right), \\
\boldsymbol{\gamma}^v_{(s)}\left(\boldsymbol{\psi}^v_{(s)}\right) \mid \boldsymbol{\psi}^v_{(s)}, \sigma_v^2 &\overset{\text{ind}}{\sim} CN_p\left(\hat{\boldsymbol{\gamma}}^v_{(s)}\left(\boldsymbol{\psi}^v_{(s)}\right), 2T\sigma_v^2\left(\mathbf{X}'\left(\boldsymbol{\psi}^v_{(s)}\right)\mathbf{X}\left(\boldsymbol{\psi}^v_{(s)}\right)^{-1}\right), \mathbf{0}\right), \\
\hat{\boldsymbol{\gamma}}^v_{(s)}\left(\boldsymbol{\psi}^v_{(s)}\right) &= \left(\mathbf{X}'\left(\boldsymbol{\psi}^v_{(s)}\right)\mathbf{X}\left(\boldsymbol{\psi}^v_{(s)}\right)^{-1}\right)\mathbf{X}'\left(\boldsymbol{\psi}^v_{(s)}\right)\mathbf{y}^v_{(s)}, \\
\sigma_v^2 \mid a_\sigma, b_\sigma &\overset{\text{iid}}{\sim} IG\left(a_\sigma, b_\sigma\right), \\
\psi^v_{(s),j} \mid a^d_{(s)}, S^v_j &\sim Bernoulli\left(\frac{1}{1+e^{-\left(a^d_{(s)}+S^v_j\right)}}\right), \\
S^v_j \mid \phi, \mathbf{w}_{(j)} &= \sum_{g=1}^{G} k\left(z_v - s_g; \phi\right) w^g_j \qquad (6.1) \\
w^g_j \mid \tau^2 &\overset{\text{ind}}{\sim} N\left(0, \tau^2_j\right), \\
\tau^2_j \mid a_\tau, b_\tau &\overset{\text{iid}}{\sim} Ga\left(a_\tau, b_\tau\right), \\
\phi \mid a_\phi, b_\phi &\sim Ga\left(a_\phi, b_\phi\right), \\
\boldsymbol{a}_{(s)} = \left(a^1_{(s)}, \ldots, a^D_{(s)}\right)' &\overset{\text{iid}}{\sim} N_D\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}\right), \\
\boldsymbol{\mu} = \left(\mu_1, \ldots, \mu_D\right)' &\sim N_D\left(\mathbf{0}, \mathbf{I}\right), \\
\boldsymbol{\Omega} \mid r_0, \boldsymbol{\Omega}_0 &\sim Wishart_D\left(r_0, \boldsymbol{\Omega}_0\right).
\end{aligned}
$$

The connectivity is estimated through the correlation structure of the parameter $\boldsymbol{a} = \left(\boldsymbol{a}_{(1)}, \ldots, \boldsymbol{a}_{(n)}\right)$ with $\boldsymbol{a}_{(s)} = \left(a^1_{(s)}, \ldots, a^D_{(s)}\right)$ whose components appear in the Bernoulli prior

$Bernoulli\left(1/\left(1+e^{-\left(a^d_{(s)}+S^v_j\right)}\right)\right)$. By doing so, the model can potentially utilize connectivity information to further improve activation detection (Ng et al., 2011; Yoldemir

[1]The multi-subject model can also be generalized from the GP model in the previous chapter.

et al., 2013). Instead of assuming that $a^1_{(s)}, \ldots, a^D_{(s)}$ are constants representing prior external anatomical information as in the spatial GP or KC models in the previous chapter, here $\boldsymbol{a}_{(s)}$ is assumed normally distributed with a mean vector $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{\Omega}$. Therefore, the connectivity structure is captured by the $D \times D$ precision matrix $\boldsymbol{\Omega}$ or equivalently, the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}$ that shows the conditional dependence or correlation between the D connectivity regions or ROIs.

To keep the model and its MCMC algorithm computationally efficient, a simple and natural choice of prior for $\boldsymbol{\Omega}$ is the conjugate Wishart distribution, i.e., $\boldsymbol{\Omega} \sim Wishart\,(r_0, \Omega_0)$. Also, using a Wishart prior always leads to a positive definite posterior estimate for $\boldsymbol{\Omega}$. When no prior information about the connectivity across ROIs is available, one can choose the Wishart prior whose mean is the identity matrix, that is, $\boldsymbol{\Omega}_0 = \frac{1}{r_0}\mathbf{I}$. The degrees of freedom r_0 in the prior can be seen as the "prior sample size". The larger r_0 is, the more the posterior estimate will shrink to the identity matrix. The least informative but proper Wishart prior is obtained by setting $r_0 = D$. It should be noticed that our connectivity is not directly estimated by the correlation between time series of ROIs. Instead, the connectivity estimation is mainly based on n replicates of correlated connectivity parameters $\left(\boldsymbol{a}_{(1)}, \ldots, \boldsymbol{a}_{(n)}\right)$ which is usually much smaller than T. As a result, when there are many ROIs and relatively few subjects, D can be large and even larger than n. In this situation, even using the least informative proper prior may result in huge shrinkage to the identity matrix because information from the data is weighted less comparing to the prior information due to a small sample size n. Therefore, in order to obtain accurate connectivity results, the

number of subjects has to be larger than the number of ROIs or, alternatively, some prior information about connectivity needs to be included through the prior on $\mathbf{\Omega}$. If only a small number of subjects have participated in the study, a smaller number of ROIs should be considered. In addition, note that the ROIs do not need to cover the entire brain or all the voxels in the image. For those voxels that do not belong to any ROI, their parameter $a^d_{(s)}$ in the Bernoulli prior can be set to zero. An improper Wishart prior, i.e., a prior with $r_0 < D$, could also be considered if the posterior distribution remains proper. However, if an improper prior is used when n is small, a noisy estimation is expected.

Note that in our proposed model, the spatial effects S^v_j are voxel-specific but not subject-specific. While subjects have their own individual characteristics, they also share common features. Therefore, common spatial effects S^v_j across subjects can be seen as a group-level effect borrowing information about local dependence from all subjects. Subject-specific $a^d_{(s)}$, on the other hand, is a subject-level effect that account for individual differences.

6.1.2 Posterior inference

Let $\boldsymbol{\psi} = \left(\boldsymbol{\psi}_{(1)}, \ldots, \boldsymbol{\psi}_{(n)}\right)$ be the collection of indicator variables for identifying activation in all voxels and subjects, and $\boldsymbol{\psi}_{(s)} = \left(\boldsymbol{\psi}^1_{(s)}, \ldots, \boldsymbol{\psi}^V_{(s)}\right)$, $\boldsymbol{\psi}^v_{(s)} = \left(\psi^v_{(s),1}, \ldots, \psi^v_{(s),p}\right)$. Let $\boldsymbol{\gamma} = \left(\boldsymbol{\gamma}_{(1)}\left(\boldsymbol{\psi}_{(1)}\right), \ldots, \boldsymbol{\gamma}_{(n)}\left(\boldsymbol{\psi}_{(n)}\right)\right)$ be the collection of regression coefficients corresponding to expected BOLD signals in all voxels and subjects, and

$\boldsymbol{\gamma}_{(s)}\left(\boldsymbol{\psi}_{(s)}\right) = \left(\boldsymbol{\gamma}^1_{(s)}\left(\boldsymbol{\psi}^1_{(s)}\right), \ldots, \boldsymbol{\gamma}^V_{(s)}\left(\boldsymbol{\psi}^V_{(s)}\right)\right)$. $\mathbf{w} = \left(\mathbf{w}_{(1)}, \ldots, \mathbf{w}_{(p)}\right)$, $\mathbf{w}_{(j)} = \left(w^1_j, \ldots, w^G_j\right)$, $\boldsymbol{\tau}^2 = \left(\tau^2_1, \ldots, \tau^2_p\right)$, $\boldsymbol{a} = \left(\boldsymbol{a}_{(1)}, \ldots, \boldsymbol{a}_{(n)}\right)$, $\boldsymbol{\sigma}^2 = \left(\sigma^2_1, \ldots, \sigma^2_V\right)$ and $\boldsymbol{S}_{(j)} = \mathbf{K}\mathbf{w}_{(j)}$ where $\mathbf{K} = \mathbf{K}(\phi)$ is the matrix with elements $k\left(z_v - u_g; \phi\right)$. Without loss of generality, we assume ϕ is a scalar greater than zero. Based on the model and prior setup provided above in the previous section, the posterior distribution of the full multi-subject model is given by

$$\begin{aligned} q\left(\boldsymbol{\gamma}\left(\boldsymbol{\psi}\right), \boldsymbol{\psi}, \mathbf{w}, \boldsymbol{\tau}^2, \boldsymbol{\sigma}^2, \boldsymbol{a}, \boldsymbol{\mu}, \boldsymbol{\Omega} \mid \mathbf{y}\right) \propto\ & p\left(\mathbf{y} \mid \boldsymbol{\gamma}\left(\boldsymbol{\psi}\right), \boldsymbol{\psi}, \boldsymbol{\sigma}^2\right) \pi\left(\boldsymbol{\gamma}\left(\boldsymbol{\psi}\right) \mid \boldsymbol{\psi}, \boldsymbol{\sigma}^2\right) \pi\left(\boldsymbol{\sigma}^2\right) \\ & \times \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi, \boldsymbol{a}\right) \pi\left(\mathbf{w} \mid \boldsymbol{\tau}^2\right) \pi\left(\boldsymbol{\tau}^2\right) \pi\left(\phi\right) \\ & \times \pi\left(\boldsymbol{a} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}\right) \pi\left(\boldsymbol{\mu}\right) \pi\left(\boldsymbol{\Omega}\right). \end{aligned}$$

Integrating $\boldsymbol{\gamma}$, $\boldsymbol{\sigma}^2$ and $\boldsymbol{\tau}^2$ out, we obtain the density of $\mathbf{y}$ given $\boldsymbol{\psi}$:

$$p\left(\mathbf{y} \mid \boldsymbol{\psi}\right) = \prod_{v=1}^{V} (1+T)^{-\sum_{s=1}^{n} q^v_{(s)}} \left(\frac{1}{2}\sum_{s=1}^{n} M^v_{(s)} + b_\sigma\right)^{(nT + a_\sigma)},$$

where $M^v_{(s)} = \left(\mathbf{y}^v_{(s)}\right)' \mathbf{y}^v_{(s)} - \left(\mathbf{y}^v_{(s)}\right)' \mathbf{X}\left(\boldsymbol{\psi}\right)\left(\mathbf{X}\left(\boldsymbol{\psi}\right)' \mathbf{X}\left(\boldsymbol{\psi}\right)\right)^{-1} \mathbf{X}\left(\boldsymbol{\psi}\right)' \mathbf{y}^v_{(s)}$, and $q^v_{(s)}$ is the number of tasks for which voxel v of subject s is activated, i.e., $q^v_{(s)} = \sum_{j=1}^{p} \psi^v_{(s),j}$.

Then, the marginal posterior is

$$q\left(\boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{a}, \boldsymbol{\Omega} \mid \mathbf{y}\right) \propto p\left(\mathbf{y} \mid \boldsymbol{\psi}\right) \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi, \boldsymbol{a}\right) \pi\left(\mathbf{w}\right) \pi\left(\phi\right) \pi\left(\boldsymbol{a} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}\right) \pi\left(\boldsymbol{\mu}\right) \pi\left(\boldsymbol{\Omega}\right).$$

We then infer the parameters $\boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{a}$ and $\boldsymbol{\Omega}$ via a MCMC algorithm that allows us to obtain posterior samples from this marginal posterior using Metropolis-Hastings

steps and/or Gibbs steps as follows:

- Sampling $\psi^v_{(s),j}$: For $j = 1, \ldots, p$, $v = 1, \ldots, V$ and $s = 1, \ldots, n$,

$$q\left(\psi^v_{(s),j} \mid \psi^{-v}_{(-s),-j}, \psi^{-v}_{(-s),-j}, \mathbf{w}_{(j)}, \boldsymbol{a}, \phi, \mathbf{y}\right) \propto p\left(\mathbf{y} \mid \boldsymbol{\psi}\right) \pi\left(\psi^v_{(s),j} \mid \mathbf{w}_{(j)}, \phi, \boldsymbol{a}_{(s)}\right),$$

and $\pi\left(\psi^v_{(s),j} \mid \mathbf{w}_{(j)}, \phi, \boldsymbol{a}_{(s)}\right) = Bernoulli\left(\frac{1}{1+e^{-\left(a^d_{(s)}+\mathbf{K}'_{v(\phi)}\mathbf{w}_{(j)}\right)}}\right)$. Therefore, each $\psi^v_{(s),j}$ can be drawn from a Bernoulli distribution with the probability of activation proportional to $p\left(\mathbf{y} \mid \boldsymbol{\psi}\right) Bernoulli\left(\frac{1}{1+\exp\left\{-\left(a^d_{(s)}+\mathbf{K}'_{v}(\phi)\mathbf{w}_{(j)}\right)\right\}}\right)$.

- Sampling w^g_j: For $j = 1, \ldots, p$ and $g = 1, \ldots, G$, we have that the conditional density of w^g_j is given by

$$\begin{aligned} q\left(w^g_j \mid \mathbf{w}^{(-g)}_{(-j)}, \boldsymbol{\psi}, \phi, \boldsymbol{a}, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{\psi}_{(j)} \mid \mathbf{w}_{(j)}, \phi, \boldsymbol{a}\right) \pi\left(w^g_j\right) \\ &= \left(\prod_{s=1}^{n}\prod_{v=1}^{V} \pi\left(\psi^v_{(s),j} \mid \mathbf{w}_{(j)}, \phi, \boldsymbol{a}\right)\right) \pi\left(w^g_j\right). \end{aligned}$$

A Metropolis-Hastings step with a normal random walk proposal $\left(w^g_j\right)^* \sim N\left(\left(w^g_j\right)^{(m)}, \sigma^2_{w^g_j}\right)$ is used to sample w^g_j, where $\left(w^g_j\right)^{(m)}$ is the current state and $\sigma^2_{w^g_j}$ is a tuning parameter.

- Sampling ϕ: The full conditional of ϕ is

$$\begin{aligned} q\left(\phi \mid \boldsymbol{\psi}, \mathbf{w}, \boldsymbol{a}, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi, \boldsymbol{a}\right) \pi\left(\phi\right) \\ &= \left(\prod_{s=1}^{n}\prod_{j=1}^{p}\prod_{v=1}^{V} \pi\left(\psi^v_{(s),j} \mid \mathbf{w}_{(j)}, \phi, \boldsymbol{a}_{(s)}\right)\right) \pi\left(\phi\right). \end{aligned}$$

A Metropolis-Hastings step with a normal random walk proposal for $\xi = \log(\phi)$,

$\xi^* \sim N\left(\xi^{(m)}, \sigma_\xi^2\right)$ is used to sample ξ and hence ϕ, where $\xi^{(m)}$ is current state value and σ_ξ^2 is a tuning parameter.

- Sampling $\boldsymbol{a}_{(s)}$: For $s = 1, \ldots, n$,

$$\begin{aligned} q\left(\boldsymbol{a}_{(s)} \mid \boldsymbol{a}_{-(s)}, \boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi, \boldsymbol{a}\right) \pi\left(\boldsymbol{a} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}\right) \\ &= \left(\prod_{s=1}^{n} \prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_{(s),j}^{v} \mid \mathbf{w}_{(j)}, \phi, \boldsymbol{a}_{(s)}\right)\right) \\ &\quad \times N\left(\boldsymbol{a}_{(s)} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}\right). \end{aligned}$$

Block move update for $\boldsymbol{a}_{(s)}$ and a random walk proposal $\boldsymbol{a}_{(s)}^* \sim N_D\left(\boldsymbol{a}_{(s)}^{(m)}, \boldsymbol{\Sigma}_{\boldsymbol{a}_{(s)}}\right)$ is used, where $\boldsymbol{a}_{(s)}^{(m)}$ is the current state and $\boldsymbol{\Sigma}_{\boldsymbol{a}_{(s)}}$ is the tuning covariance matrix.

- Sampling $\boldsymbol{\mu}$: The full conditional of $\boldsymbol{\mu}$ is

$$\begin{aligned} q\left(\boldsymbol{\mu} \mid \boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{a}, \boldsymbol{\Omega}, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{a} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}\right) \pi\left(\boldsymbol{\mu}\right) = N_D\left(\boldsymbol{a}_{(s)} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}\right) N_D\left(\boldsymbol{\mu} \mid \mathbf{0}, \mathbf{I}\right) \\ &\sim N_D\left(\left(\mathbf{I} + n\boldsymbol{\Omega}\right)^{-1} \boldsymbol{\Omega} n \overline{\boldsymbol{a}}, \left(\mathbf{I} + n\boldsymbol{\Omega}\right)^{-1}\right), \end{aligned}$$

where $\overline{\boldsymbol{a}} = \frac{1}{n} \sum_{s=1}^{n} \boldsymbol{a}_{(s)}$. Then $\boldsymbol{\mu}$ is directly sampled from the multivariate normal distribution above.

- Sampling $\boldsymbol{\Omega}$: The full conditional of $\boldsymbol{\Omega}$ is

$$\begin{aligned} q\left(\boldsymbol{\Omega} \mid \boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{a}, \boldsymbol{\mu}, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{a} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}\right) \pi\left(\boldsymbol{\Omega}\right) \\ &= \left(\prod_{s=1}^{n} N_D\left(\boldsymbol{a}_{(s)} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}\right)\right) \times Wishart\left(\boldsymbol{\Omega} \mid r_0, \boldsymbol{\Omega}_0\right) \\ &\propto |\boldsymbol{\Omega}|^{-\frac{n+r_0-D-1}{2}} \exp\left(-\frac{1}{2} tr\left(\sum_{s=1}^{n} \boldsymbol{a}_{(s)}\left(\boldsymbol{a}_{(s)}\right)' + \boldsymbol{\Omega}_0^{-1}\right)\boldsymbol{\Omega}\right) \\ &\sim Wishart_D\left(n + r_0, \left(\sum_{s=1}^{n}\left(\boldsymbol{a}_{(s)} - \boldsymbol{\mu}\right)\left(\boldsymbol{a}_{(s)} - \boldsymbol{\mu}\right)' + \boldsymbol{\Omega}_0^{-1}\right)^{-1}\right). \end{aligned}$$

  $\boldsymbol{\Omega}$ is directly sampled from the Wishart distribution above.

Details of the algorithm can be found in Appendix A.6.

6.2 Simulation study I

In this study we consider a data set and a model without spatial random effects, and focus on obtaining posterior inference for the parameters $\boldsymbol{a}$ and $\boldsymbol{\Sigma}$.

The simulated image is of size 12×12 containing $V = 144$ voxels. In the study, the ROIs are $D = 9$ equally-sized grids each containing 16 voxels. The left panel of Figure 6.1 shows the 9 brain ROIs. Two regions are correlated if an edge is connected between the nodes of the regions. Therefore, for the 9 ROIs, the following region pairs are correlated: $(1,3)$, $(1,7)$, $(1,9)$, $(3,7)$, $(3,7)$, and $(7,9)$. Dark green edges indicate negative correlation, while red edges indicate positive correlation. The middle and right subplots of Figure 6.1 show the true covariance $\boldsymbol{\Sigma}$ and precision matrix $\boldsymbol{\Omega}$ used in the simulated data. The covariance matrix has all diagonal values being equal to one and hence it is also a correlation matrix. The positive and negative correlation

values between two regions are 0.7 and -0.7, respectively.

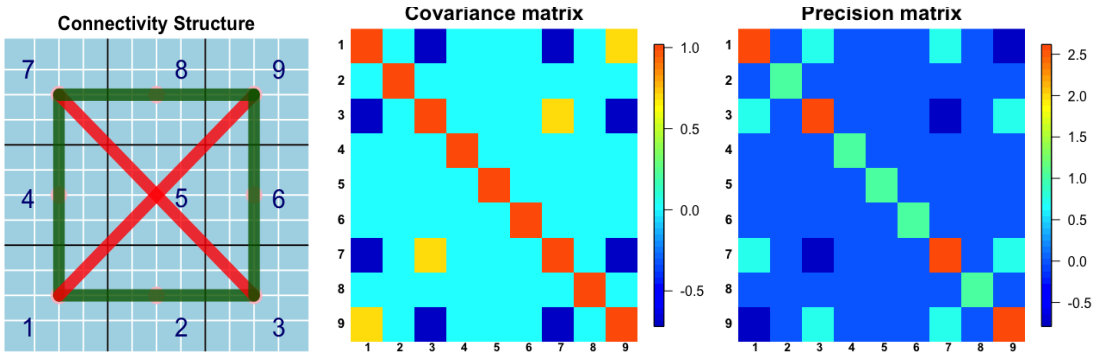


Figure 6.1: Left: 9 equally-sized connectivity brain regions. Two regions are correlated if an edge is connected between the nodes of the regions. Dark green edges indicate negative correlation while white edges indicate positive correlation. Middle: Covariance (correlation) matrix. Right: precision matrix.

Given the precision matrix $\boldsymbol{\Omega}$, the connectivity parameters $\boldsymbol{a}_{(s)}$, $s = 1, \ldots, 30$ are then generated from a $D = 9$ dimensional multivariate normal distribution $N_9\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}\right)$, where $\boldsymbol{\mu}$ is assumed to be a vector $(0, -3, 0, -3, -3, -3, 0, -3, 0)$. Drawn from $N_9\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, the vectors $\boldsymbol{a}_{(1)}, \ldots, \boldsymbol{a}_{(30)}$ have mean $(-0.07, -2.74, 0.09, -2.96, -3.00, -3.13, 0.03, -3.23, 0.15)$. With each $\boldsymbol{a}_{(s)}$, the true activation pattern for each subject is then generated by $\psi^v_{(s)} \mid a^d_{(s)} \sim Bernoulli\left(\frac{1}{1+\exp\left\{-a^d_{(s)}\right\}}\right)$. Notice that we expect regions 1, 3, 7 and 9 to have more activated voxels than other regions because about half of voxels in regions 1, 3, 7 and 9 are activated, and very few voxels in the remaining regions, indexed as 2, 4, 5, 6 and 8, are activated due to their small values of $a^d_{(s)}$, $d \in 2, 4, 5, 6, 8$. In addition, voxels in regions 1 and 9 tend to be activated together as the two regions are positively correlated. Voxels in regions 3 and 7 behave similarly. Moreover, when regions 1 and 9 yield more activated voxels, there are less activated voxels in regions 3 and 7 and vice versa due to the negative correlation between those

regions.

We consider 30 subjects who perform one task that produces similar activation maps. Specifically, in this simulation, there are two types of subjects. Some subjects show co-activation of regions 1 and 9, while some show co-activation of regions 3 and 7. Figure 6.2 shows the true activation maps of subjects 1, 10, 26 and 30. For subjects 1 and 10, they have more activated voxels in regions 1 and 9, while subjects 26 and 30 have more activated voxels in regions 3 and 7. Note that subjects do not have the same activation map because whether a voxel in a subject's image is activated or not is determined by its associated indicator variable that is randomly drawn from a Bernoulli distribution.

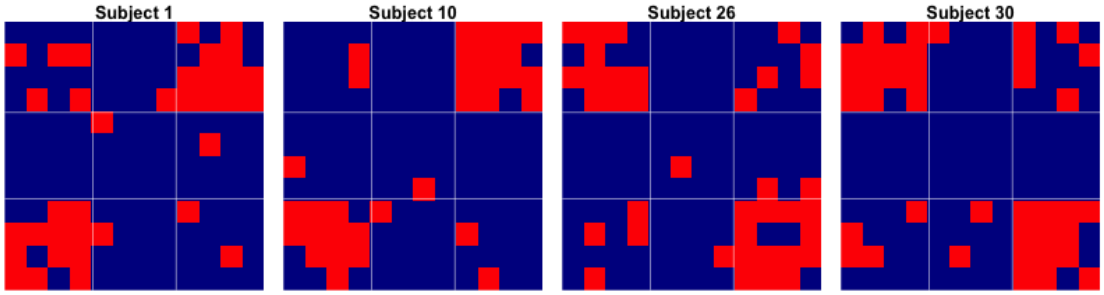


Figure 6.2: Activation maps of subject 1, 10, 26 and 30.

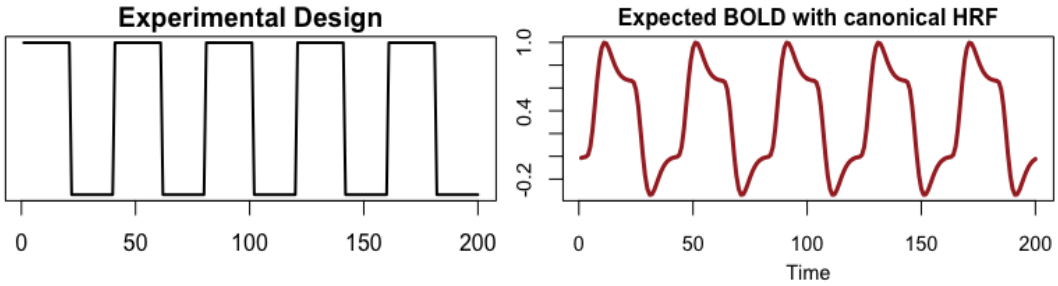


Figure 6.3: Experimental design and expected BOLD signal with canonical HRF.

Finally, for $v = 1, \ldots, 12 \times 12$, $t = 1, \ldots, 200$, and subject $s = 1, \ldots, 30$, the

real and imaginary time series components were simulated as follows:

$$\begin{aligned} y^v_{t,Re,(s)} &= \left(\beta_{v,0} + \beta_{v,1}\left(\psi^v_{(s)}\right) x_t\right)\cos\left(\alpha_0\right) + \eta^v_{t,Re,(s)}, \quad \eta^v_{t,Re,(s)} \sim N\left(0, \sigma^2\right), \\ y^v_{t,Im,(s)} &= \left(\beta_{v,0} + \beta_{v,1}\left(\psi^v_{(s)}\right) x_t\right)\sin\left(\alpha_0\right) + \eta^v_{t,Im,(s)}, \quad \eta^v_{t,Im,(s)} \sim N\left(0, \sigma^2\right), \end{aligned} \tag{6.2}$$

where again x_t is the expected BOLD signal, and $\alpha_0 = \pi/4$ as used in Chapters 4 and 5. The experimental design and expected BOLD signal, x_t, obtained using the canonical HRF are shown Figure 6.3. For $\beta_{v,0}$ and $\beta_{v,1}$, we use the case LL $(\beta_{v,0}, \beta_{v,1}) = (1, 2)$ and set $\sigma^2 = 3$ as in Chapter 5.

We then fit the following model to the simulated data set:

$$\begin{aligned} \mathbf{y}^v_{(s)} &= \mathbf{X}\left(\boldsymbol{\psi}^v_{(s)}\right)\boldsymbol{\gamma}^v_{(s)}\left(\boldsymbol{\psi}^v_{(s)}\right) + \boldsymbol{\eta}^v_{(s)}, \quad \boldsymbol{\eta}^v_{(s)} \sim CN\left(0, 2\sigma_v^2\mathbf{I}, \mathbf{0}\right), \\ \boldsymbol{\gamma}^v_{(s)}\left(\boldsymbol{\psi}^v_{(s)}\right) \mid \boldsymbol{\psi}^v_{(s)}, \sigma_v^2 &\overset{\text{ind}}{\sim} CN_p\left(\hat{\boldsymbol{\gamma}}^v_{(s)}\left(\boldsymbol{\psi}^v_{(s)}\right), 2T\sigma_v^2\left(\mathbf{X}'\left(\boldsymbol{\psi}^v_{(s)}\right)\mathbf{X}\left(\boldsymbol{\psi}^v_{(s)}\right)^{-1}\right), \mathbf{0}\right), \\ \hat{\boldsymbol{\gamma}}^v_{(s)}\left(\boldsymbol{\psi}^v_{(s)}\right) &= \left(\mathbf{X}'\left(\boldsymbol{\psi}^v_{(s)}\right)\mathbf{X}\left(\boldsymbol{\psi}^v_{(s)}\right)^{-1}\right)\mathbf{X}'\left(\boldsymbol{\psi}^v_{(s)}\right)\mathbf{y}^v_{(s)}, \\ \sigma_v^2 \mid a_\sigma, b_\sigma &\overset{\text{iid}}{\sim} IG\left(1/2, 1/2\right), \quad \psi^v_{(s)} \mid a^d_{(s)} \sim Bernoulli\left(\frac{1}{1+e^{-a^d_{(s)}}}\right), \\ \boldsymbol{a}_{(s)} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} &\overset{\text{iid}}{\sim} N_D\left(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}\right), \quad \boldsymbol{\mu} \sim N_D\left(\mathbf{0}, \mathbf{I}\right), \\ \boldsymbol{\Omega} \mid r_0 &\sim Wishart_D\left(r_0, (1/r_0)\mathbf{I}\right). \end{aligned}$$

For the parameter r_0, we set $r_0 = D = 9$, leading to the least informative proper Wishart prior. Later, we will discuss the effect of r_0 on the posterior connectivity structure.

We first examine the convergence behavior of the parameters $\boldsymbol{a}$ and $\boldsymbol{\mu}$. Figure 6.4 shows the traces, histograms, and ergodic means of posterior samples of $a^2_{(1)}$, $a^2_{(10)}$,

$a^2_{(26)}$, $a^2_{(30)}$ and μ_2, that is, the connectivity comes from $\boldsymbol{\Sigma}$ for the second region and for subject 1, 10, 26 and 30. μ_2 represents the overall mean of the second connectivity region effect averaged over $a^2_{(s)}$, $s = 1, \ldots, 30$. The trace plots show that mixing of the chain is good; the 2.5% and 97.5% quantiles and posterior means go to a steady state showing a healthy convergent behavior.

Figure 6.5 shows the posterior means of the overall means of the parameter $\boldsymbol{a}$ and its corresponding true mean. Specifically, the posterior values in the images are obtained as $\frac{1}{M}\sum_{m=1}^{M}\frac{1}{30}\sum_{s=1}^{30}\left(a^d_{(s)}\right)^{(m)}$, $d = 1, \ldots, 9$, where $\left(a^d_{(s)}\right)^{(m)}$ is the posterior sample of $a^d_{(s)}$ of the m-th MCMC iteration. The true values in the image are given by $\left(\frac{1}{30}\sum_{s=1}^{30} a^d_{(s)}\right)$, for $d = 1, \ldots, 9$. The posterior estimates are close to the true values of $\boldsymbol{a}$, and correctly identify regions 1, 3, 7 and 9 that have more activated voxels.

Figure 6.6 shows the posterior mean covariance matrices with $n = 30$ under the (improper) reference prior $r_0 = 0$, improper Wishart prior with $r_0 = 5$, Wishart prior with $r_0 = D = 9$ and the true covariance matrix. The posterior distribution of $\boldsymbol{\Omega}$ is proper as long as $n + r_0 \geq D$. Note that our prior mean covariance matrix is the identity matrix. Conditional on $\mathbf{y}$ and other parameters, at each MCMC iteration m, the conditional posterior mean covariance matrix can be broadly viewed as a weighted average of the covariance of samples of $\boldsymbol{a}^{(m)}_{(s)}$ and the prior mean covariance Ω_0, weighted by the sample size n and prior sample size r_0, providing a ridge-type estimation. Precisely speaking, the eigenvalues of the conditional posterior mean covariance matrix of $\boldsymbol{a}_{(s)}$, $\hat{\lambda}_d$, are $\frac{n}{n+r_0}\lambda_d + \frac{r_0}{n+r_0}1$, where λ_d are the eigenvalues of the covariance matrix computed by $\boldsymbol{a}^{(m)}_{(s)}$. Therefore, any correlation between region d and region d', $d \neq d'$

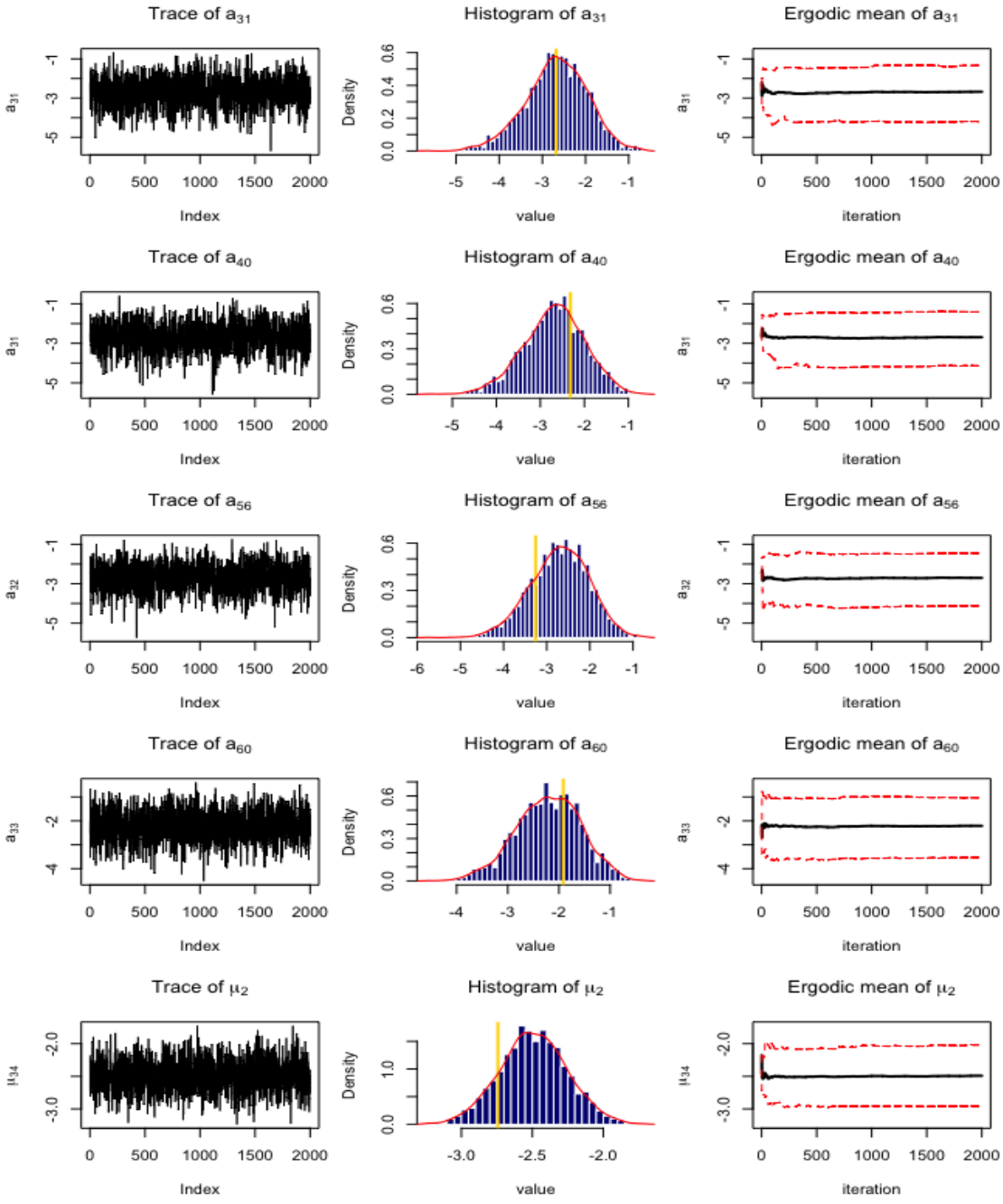


Figure 6.4: Traces, histograms, and ergodic means of posterior samples of $a^2_{(1)}$, $a^2_{(10)}$, $a^2_{(26)}$, $a^2_{(30)}$ and μ_2. Gold vertical lines in the histograms denote the true parameter values. Red dashed lines indicate 2.5% and 97.5% quantiles.

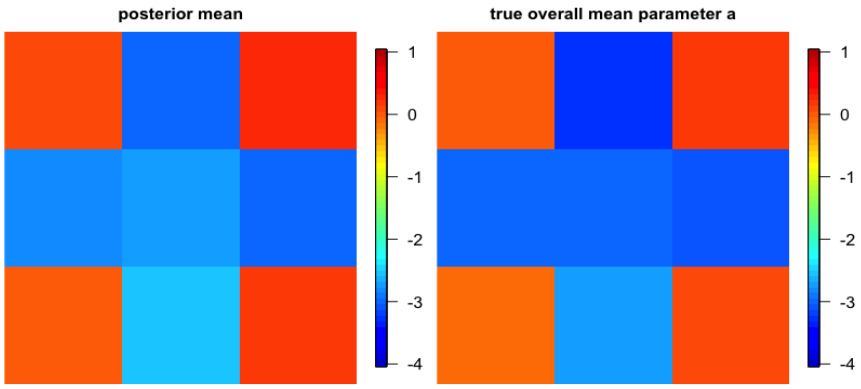


Figure 6.5: Images of posterior mean of overall mean of connectivity parameter $\boldsymbol{a}$ (left) and its corresponding true means (right).

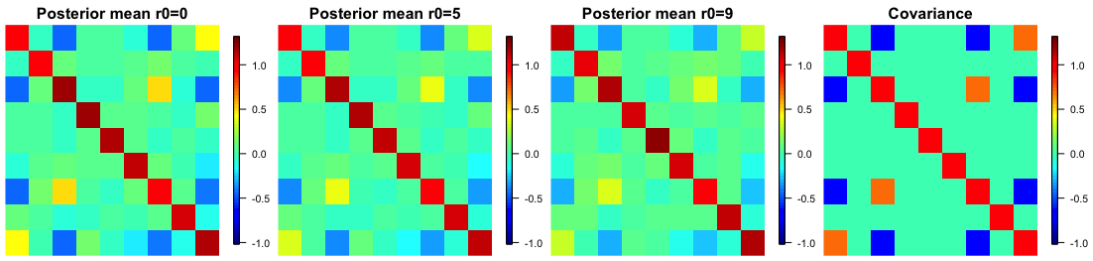


Figure 6.6: Left to right: posterior mean covariance matrices when $n = 30$ under the (improper) reference prior $r_0 = 0$, Wishart prior with $r_0 = 5$, Wishart prior with $r_0 = D = 9$ and the true covariance matrix.

will shrink more to zero as r_0 increases. This type of shrinkage effect also happens on the posterior distribution of $\boldsymbol{\Sigma}$. This can be seen in Figure 6.6, as the absolute values of off-diagonal correlations under $r_0 = 9$ are smaller than the values when the reference prior is used.

On the other hand, the Wishart prior makes the conditional posterior covariance matrix more well-conditioned as the prior increases sample eigenvalues λ_d smaller than one and shrinks λ_d greater than one. When the number of subjects is much less than the number of unknown elements to be estimated in the precision matrix, i.e., $n \ll \frac{1}{2}D(D+1)$, the eigenstructure of the sample covariance matrix computed

Threshold	Sensitivity	Specificity	Precision	Accuracy
0.5	1	0.998	0.985	0.999
0.8722	1	1	1	1

Table 6.1: Mean of sensitivity, specificity, precision and accuracy over 30 subjects.

by MCMC samples of $\boldsymbol{a}_{(s)}$ is ill-conditioned, and leads to a large estimation error. In this case, a moderate r_0 helps us get a more stable posterior covariance matrix. The posterior covariance of $\boldsymbol{\Omega}$ behaves similarly as the conditional posterior covariance generated from Wishart density.

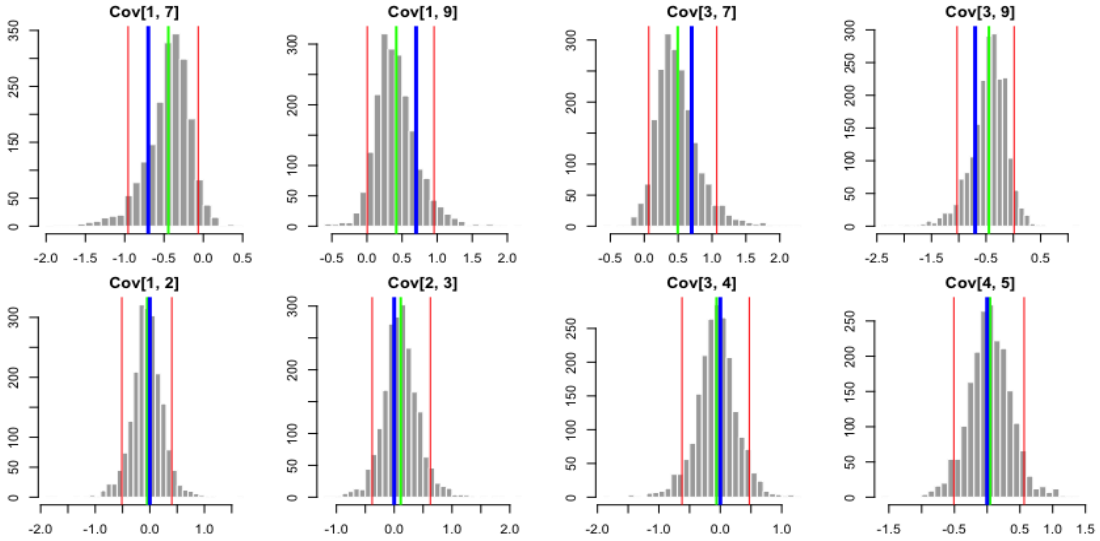


Figure 6.7: Posterior distribution of some covariance elements under the reference prior. Top: non-zero elements Σ_{17}, Σ_{19}, Σ_{37}, Σ_{39}. Bottom: zero elements Σ_{12}, Σ_{23}, Σ_{34}, Σ_{35}. Blue lines indicate the true value and green lines show the posterior mean. Red lines are 5% and 95% quantiles of the distribution.

Figure 6.7 shows the full posterior for some elements of the covariance matrix. In general, the distributions are either right or left-skewed depending on whether the true value is negative or positive. When the true correlation is zero, the posterior distribution is more symmetric.

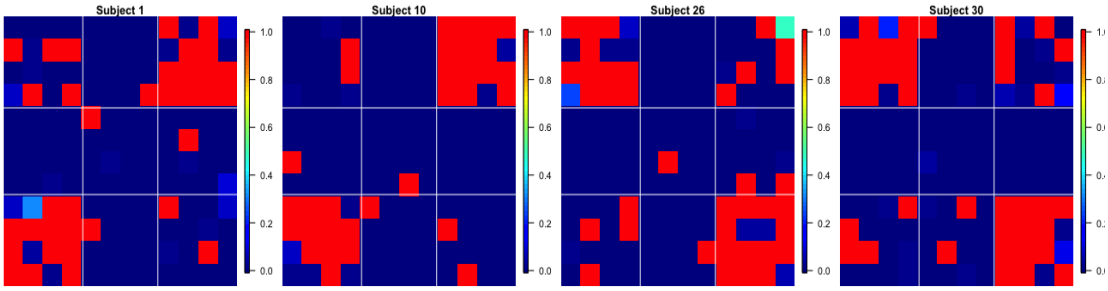


Figure 6.8: Posterior probability maps of subject 1, 10, 26 and 30.

Figure 6.8 and Table 6.1 show how well the model recovers the simulated activation. There will be only one false positive for subject 1, 10, 26 and 30 when 0.5 is used as the threshold. The mean posterior probabilities across 30 subjects are close to either $\frac{1}{1+\exp\{-0\}}$ or $\frac{1}{1+\exp\{3\}}$ depending on the mean of the connectivity parameter $\boldsymbol{a}$ or $\boldsymbol{\mu}$ at different regions. Therefore, our model adequately captures the underlying co-activation and voxel-level activation for each of the 30 subjects in the simulated data.

6.3 Simulation study II

In this section we generate a multi-subject simulated data set that has the same image size and activation areas as the simulated data set generated in Chapter 5 Figure 5.2. We then fit three different models, independent C-EMVS models on each subject, the multi-subject model proposed in this chapter with and without the spatial effect. Here, we model the spatial effect through kernel convolution $\boldsymbol{S} = \mathbf{K}\mathbf{w}$. We compare the different models in terms of their inferred activation and connectivity structures.

We consider 30 subjects who perform one task that produces three brain activation areas as in Figure 5.2 in Chapter 5. To account for individual differences, each subject has different activation strengths in the three areas which are generated from a multivariate normal distribution centered at the values used in the case LL, i.e., $(\beta_{v,0}, \beta_{v,1}) = (1, 2)$ in the simulation study in Chapter 5. Specifically, the simulated image is of size 20×20 containing $V = 400$ voxels. Six ROIs are created as shown in the left panel of Figure 6.9. Regions 1, 3 and 5 are assumed to be connected, and the correlation matrix $\mathbf{\Sigma}$ of the six regions is shown in the right panel in Figure 6.9. The nonzero correlations are fixed at value 0.7.

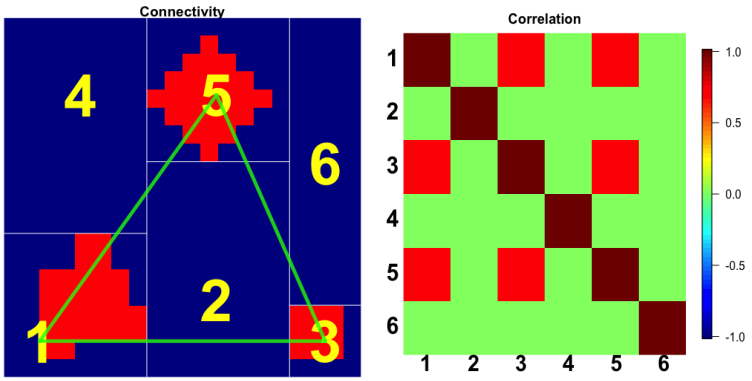


Figure 6.9: Left: connectivity regions of the simulated data. Right: correlation matrix of the connection regions.

We generate the coefficients $\beta^v_{(s),1}$ as follows. For each subject $s = 1, \ldots, 30$, active voxels in the same region d have the same value of coefficient $\beta^{(d)}_{(s),1}$. That is, $\beta^v_{(s),1} = \beta^{(d)}_{(s),1}$ for all $v \in d$. Similarly, non-active voxels in the same region have the same value of the coefficient. Notice that the proposed model describes the connectivity structure through the correlation between the ROIs in the Bernoulli prior for the indicator variables $\boldsymbol{\psi}$. Unlike Simulation study I that generates the cor-

related indicators from the model, in this simulation, the connectivity structure is described by the correlation between the coefficients of the associated regions. Denote the coefficients in the ROIs 1, 3, and 5 by $\boldsymbol{\beta}^{act}_{(s),1} = \left(\beta^{(1),act}_{(s),1}, \beta^{(3),act}_{(s),1}, \beta^{(5),act}_{(s),1}\right)$ and $\boldsymbol{\beta}^{non}_{(s),1} = \left(\beta^{(1),non}_{(s),1}, \beta^{(3),non}_{(s),1}, \beta^{(5),non}_{(s),1}\right)$ for active and non-active voxels respectively. Then $\boldsymbol{\beta}^{act}_{(s),1}$ is generated by $\boldsymbol{\beta}^{act}_{(s),1} \sim N_3\left(2 \times \mathbf{1}_3, 0.05\boldsymbol{\Sigma}_3\right)$ for activated voxels, and $\boldsymbol{\beta}^{non}_{(s),1}$ is generated by $\boldsymbol{\beta}^{non}_{(s),1} \sim N_3\left(\mathbf{0}_3, 0.05\boldsymbol{\Sigma}_3\right)$ for non-active voxels, where $\mathbf{1}_3 = (1,1,1)'$ and $\boldsymbol{\Sigma}_3$ is the sub-correlation matrix of the connectivity regions 1, 3 and 5. By doing so, individual differences are induced as each subject s would have different coefficient sizes, centered around either 0 or 2. For ROIs 2, 4, 6, their simulated coefficients are generated independently by $N(0, 0.05)$.

Figure 6.10 shows the simulated coefficients $\beta^{v}_{(1),1}$, $\beta^{v}_{(2),1}$ and $\beta^{v}_{(3),1}$, for $v = 1, \ldots, 20 \times 20$. Clearly, each subject has different coefficient values, i.e., different activation strengths, and will have different activation estimation results, but the true activation maps are the same for the 3 subjects.

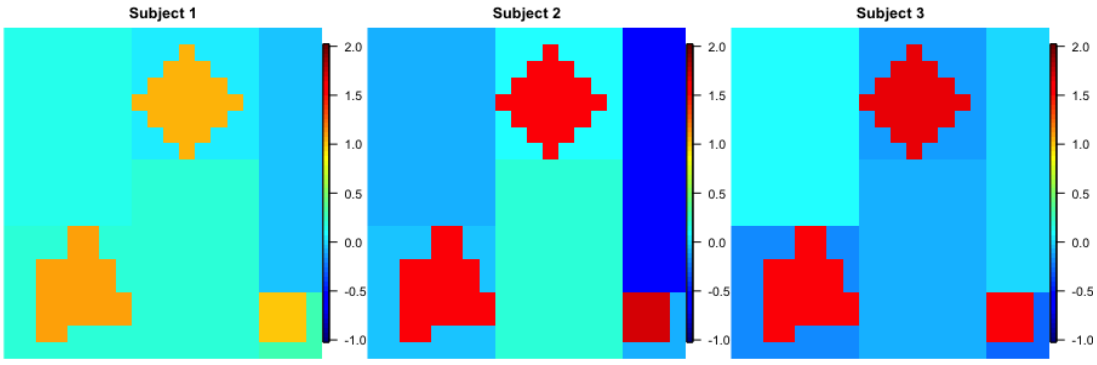


Figure 6.10: Simulated coefficients $\beta^{v}_{(1),1}$, $\beta^{v}_{(2),1}$ and $\beta^{v}_{(3),1}$, $v = 1, \ldots, 20 \times 20$.

Finally, for each subject $s = 1, \ldots, 30$, at voxel $v = 1, \ldots, 20 \times 20$ and time

$t = 1, \ldots, 50$, the real and imaginary parts of time courses are simulated as follows:

$$\begin{aligned} y^v_{t,Re,(s)} &= \left(\beta_{v,0} + \beta^v_{(s),1} x_t\right) \cos(\alpha_0) + \eta^v_{t,Re,(s)}, \quad \eta^v_{t,Re,(s)} \sim N\left(0, \sigma^2\right), \\ y^v_{t,Im,(s)} &= \left(\beta_{v,0} + \beta^v_{(s),1} x_t\right) \sin(\alpha_0) + \eta^v_{t,Im,(s)}, \quad \eta^v_{t,Im,(s)} \sim N\left(0, \sigma^2\right), \end{aligned} \tag{6.3}$$

where again x_t is the expected BOLD signal. $\beta_{v,0} = 1$, $\alpha_0 = \pi/4$ and $\sigma^2 = 3$ are used as in the previous simulation.

We consider 3 different approaches, labeled as A, B and C, to analyze the simulated data, where the first method uses the single-subject C-EMVS model and the other two are multi-subject models with and without spatial effects, i.e.,

A. *Independent single-subject C-EMVS on each subject* $1, \ldots, 30$. The independent single-subject C-EMVS is the model (4.8) in Chapter 4, that is,

$$\begin{aligned} \mathbf{y}^v &= \mathbf{X}\boldsymbol{\gamma}^v + \boldsymbol{\eta}^v, \;\; \boldsymbol{\eta}^v \sim CN_T\left(\mathbf{0}, 2\sigma_v^2\mathbf{I}, \mathbf{0}\right), \\ \gamma^v \mid \psi^v &\sim (1-\psi^v)\, CN_1\left(0, 2v_0\sigma^2, 0\right) + \psi^v CN_1\left(0, 2v_1\sigma^2, 0\right), \\ \sigma^2 &\sim IG(1/2, 1/2), \quad \psi^v \mid \theta \sim Bernoulli(\theta), \quad \theta \sim Beta(1,1). \end{aligned}$$

The parameter settings are the same as the setting in the first simulation study discussed in Chapter 4.

B. *Multi-subject model without the spatial effects* $\boldsymbol{S}$. The model is

$$\begin{aligned}
\mathbf{y}_{(s)}^{v} &= \mathbf{X}\left(\boldsymbol{\psi}_{(s)}^{v}\right)\boldsymbol{\gamma}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right)+\boldsymbol{\eta}_{(s)}^{v}, \quad \boldsymbol{\eta}_{(s)}^{v} \sim CN\left(0, 2\sigma_{v}^{2}\mathbf{I}, \mathbf{0}\right), \\
\boldsymbol{\gamma}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right) \mid \boldsymbol{\psi}_{(s)}^{v}, \sigma_{v}^{2} &\overset{\text{ind}}{\sim} CN_{p}\left(\hat{\boldsymbol{\gamma}}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right), 2T\sigma_{v}^{2}\left(\mathbf{X}'\left(\boldsymbol{\psi}_{(s)}^{v}\right)\mathbf{X}\left(\boldsymbol{\psi}_{(s)}^{v}\right)^{-1}\right), \mathbf{0}\right), \\
\hat{\boldsymbol{\gamma}}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right) &= \left(\mathbf{X}'\left(\boldsymbol{\psi}_{(s)}^{v}\right)\mathbf{X}\left(\boldsymbol{\psi}_{(s)}^{v}\right)^{-1}\right)\mathbf{X}'\left(\boldsymbol{\psi}_{(s)}^{v}\right)\mathbf{y}_{(s)}^{v}, \\
\sigma_{v}^{2} \mid a_{\sigma}, b_{\sigma} &\overset{\text{iid}}{\sim} IG\left(1/2, 1/2\right), \quad \psi_{(s)}^{v} \mid a_{(s)}^{d} \sim Bernoulli\left(\frac{1}{1+e^{-a_{(s)}^{d}}}\right), \\
\boldsymbol{a}_{(s)} \mid \boldsymbol{\mu}, \boldsymbol{\Omega} &\overset{\text{iid}}{\sim} N_{D}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}\right), \quad \boldsymbol{\mu} \sim N_{D}\left(\mathbf{0}, \mathbf{I}\right), \\
\boldsymbol{\Omega} \mid r_{0} &\sim Wishart_{D}\left(r_{0}, (1/r_{0})\mathbf{I}\right).
\end{aligned}$$

C. *Multi-subject model with the spatial effects using kernel convolution.* The fitted full multi-subject model is

$$\begin{aligned}
\mathbf{y}_{(s)}^{v} &= \mathbf{X}\left(\boldsymbol{\psi}_{(s)}^{v}\right)\boldsymbol{\gamma}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right)+\boldsymbol{\eta}_{(s)}^{v}, \quad \boldsymbol{\eta}_{(s)}^{v} \sim CN\left(0, 2\sigma_{v}^{2}\mathbf{I}, \mathbf{0}\right), \\
\boldsymbol{\gamma}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right) \mid \boldsymbol{\psi}_{(s)}^{v}, \sigma_{v}^{2} &\overset{\text{ind}}{\sim} CN_{p}\left(\hat{\boldsymbol{\gamma}}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right), 2T\sigma_{v}^{2}\left(\mathbf{X}'\left(\boldsymbol{\psi}_{(s)}^{v}\right)\mathbf{X}\left(\boldsymbol{\psi}_{(s)}^{v}\right)^{-1}\right), \mathbf{0}\right), \\
\hat{\boldsymbol{\gamma}}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right) &= \left(\mathbf{X}'\left(\boldsymbol{\psi}_{(s)}^{v}\right)\mathbf{X}\left(\boldsymbol{\psi}_{(s)}^{v}\right)^{-1}\right)\mathbf{X}'\left(\boldsymbol{\psi}_{(s)}^{v}\right)\mathbf{y}_{(s)}^{v}, \\
\psi_{(s)}^{v} \mid a_{(s)}^{d}, S^{v} &\sim Bernoulli\left(\frac{1}{1+e^{-\left(a_{(s)}^{d}+S^{v}\right)}}\right), \quad \sigma_{v}^{2} \overset{\text{iid}}{\sim} IG\left(1/2, 1/2\right), \\
S^{v} \mid \phi, \mathbf{w} &= \sum_{g=1}^{G} k\left(z_{v}-s_{g};\phi\right)w^{g}, \quad w^{g} \mid \tau^{2} \overset{\text{ind}}{\sim} N\left(0, \tau^{2}\right), \\
\tau^{2} &\overset{\text{iid}}{\sim} IG\left(1/2, 1/2\right), \quad \phi \mid a_{\phi}, b_{\phi} \sim Ga\left(a_{\phi}, b_{\phi}\right), \\
\boldsymbol{a}_{(s)} &\overset{\text{iid}}{\sim} N_{D}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}\right), \quad \boldsymbol{\mu} \sim N_{D}\left(\mathbf{0}, \mathbf{I}\right), \\
\boldsymbol{\Omega} \mid r_{0} &\sim Wishart_{D}\left(r_{0}, (1/r_{0})\mathbf{I}\right).
\end{aligned}$$

Here we use the reference prior for $\boldsymbol{\Sigma}^{-1}$ for both methods B and C.

Figure 6.11 shows the correlation estimation at the 5% quantile, 95% quantile and posterior mean derived from method B. Although the nonzero correlations are underestimated, the model can adequately capture the connectivity structure, even though these data were not simulated from this model. The estimation from method C is similar to the estimation from method B, and method A is not able to learn the connectivity structure as the C-EMVS treats each subject independently.

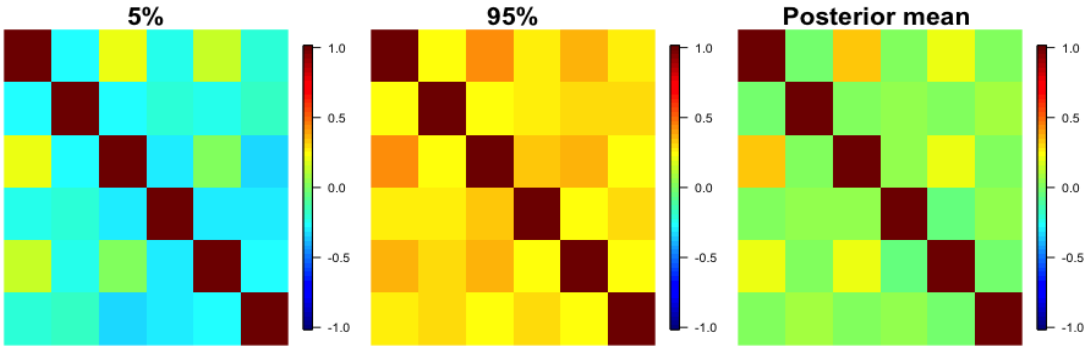


Figure 6.11: 5% quantile, 95% quantile and posterior mean of the correlation matrix of $\boldsymbol{a}_{(s)}$.

For activations, we first examine how the detection activation performance can be improved when a multi-subject model is used, and then check how spatial effects from kernel convolutions further improve the detection performance.

Figure 6.12 shows the estimate activation maps for subjects 1, 2 and 3 with threshold value of 0.5 to detect activation from the three approaches. Since subject 1 has smaller regression coefficients, i.e., weaker activation strengths, as shown in Figure 6.10, we expect to see less activated voxels for subject 1. One can see that the single-subject C-EMVS leads to more false positives than the multi-subject models. This is

reasonable because the multi-subject models can borrow information across subjects and learn that voxels in ROIs 2, 4 and 6 tend to be non-activated. Furthermore, when the spatial effects are included in the model, the local dependence between neighboring voxels is more adequately captured by kernel convolution. Therefore, the full multi-subject spatial model detects more true positives that are missed when the local dependence is not characterized in the model. Table 6.2 shows sensitivity, specificity, precision and accuracy measures for the three models. Simply using the multi-subject model without spatial effects can increase precision, comparing to single-subject C-EMVS analysis. When the latent spatial effects are in the model, the power, or true positive rate, also increases significantly.

To examine the activation pattern at the population level, the overall activation map can be derived from averaging activation maps over all subjects. One advantage of MCMC is that uncertainty can be quantified. Hence, we could also average posterior probabilities from the MCMC samples over all subjects. Figure 6.13 shows the proportion of voxels being activated in 30 subjects. Single-subject C-EMVS produces some false positives in the individual images, but it can capture the overall activation pattern as the multi-subject models. Once again we see that when the spatial effects are included, more voxels are classified as activated in the three true activation areas, showing high detection power. Figure 6.14 shows uncertainty measures for the posterior probability maps obtained from the multi-subject models. Clearly, the full multi-subject model with the kernel convolution spatial effect appropriately increases the posterior probabilities of activation for the true activation areas. Finally,

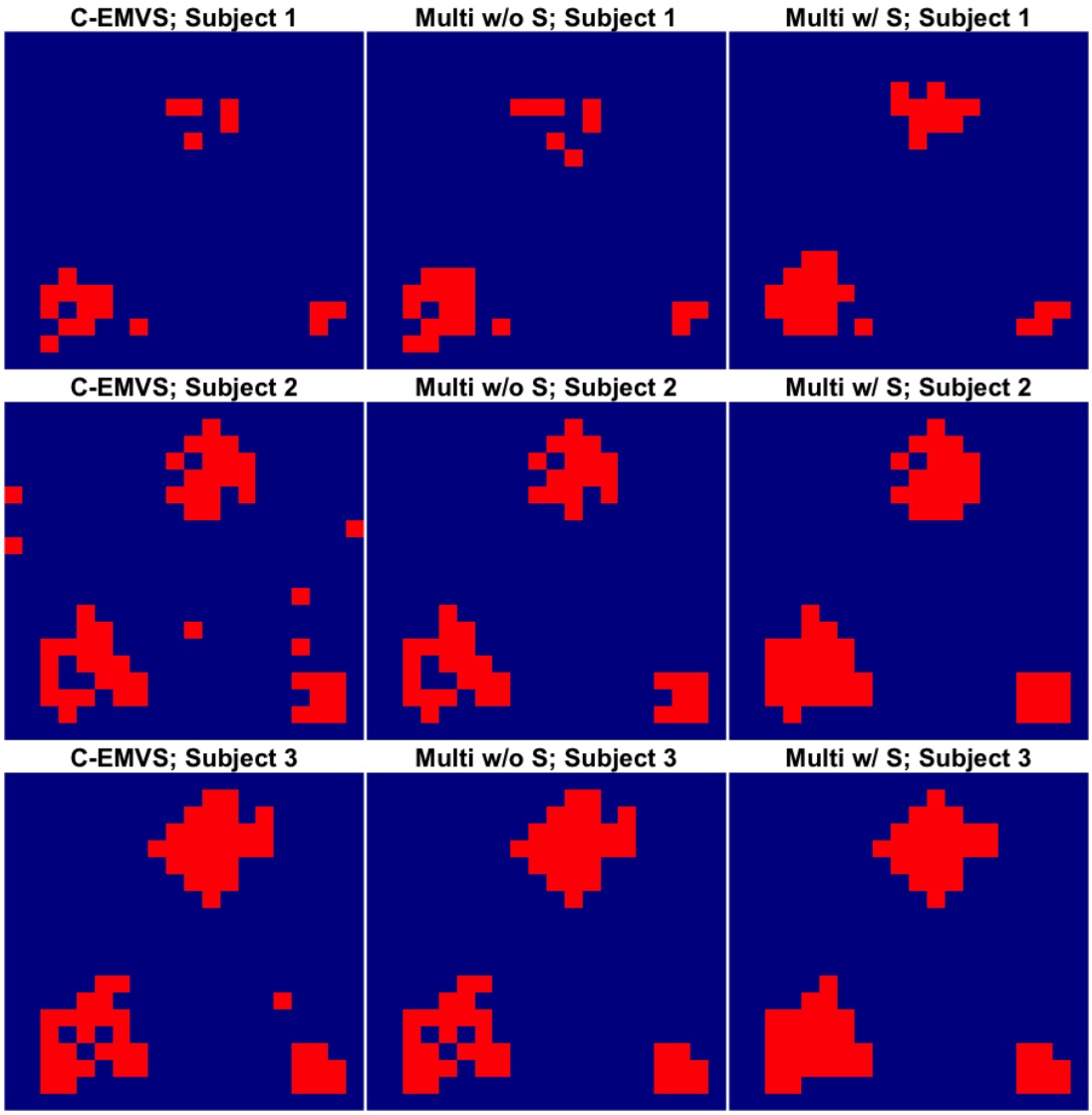


Figure 6.12: Activation maps for subjects 1, 2 and 3 (top, center and bottom) with threshold value 0.5 from the three fitted models, C-EMVS (left), multi-subject model without spatial effects (middle), and multi-subject model with spatial effects (right).

the posterior mean of population-level spatial effects, and its corresponding logistic transformation, from the full multi-subject model are shown in Figure 6.15. Superimposing this common population-level effect onto each subject, the multi-subject model with spatial effect included not only eliminates false positives but also increase true

Model	Sensitivity	Specificity	Precision	Accuracy
A. C-EMVS	0.647	0.984	0.892	0.932
B. Multi w/o S	0.651	0.993	0.956	0.941
C. Multi w/ S	**0.810**	**0.998**	**0.985**	**0.969**

Table 6.2: Sensitivity, specificity, precision and accuracy of single-subject C-EMVS, multi-subject without and with spatial effects.

positives.

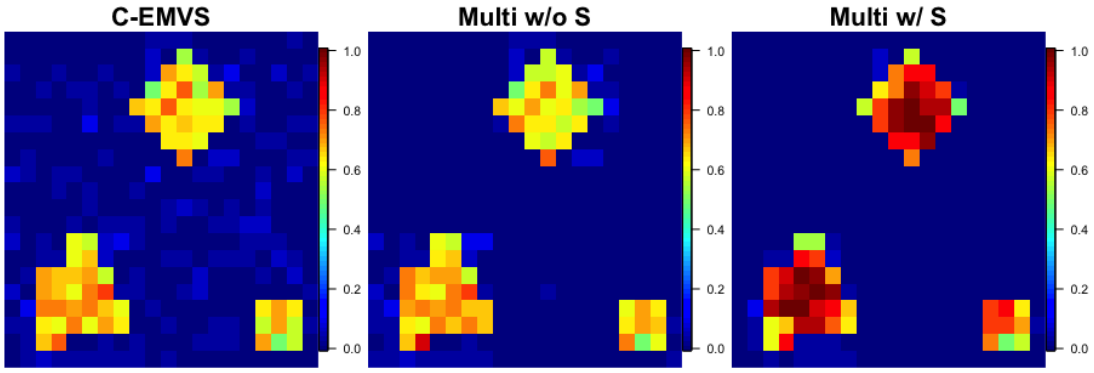
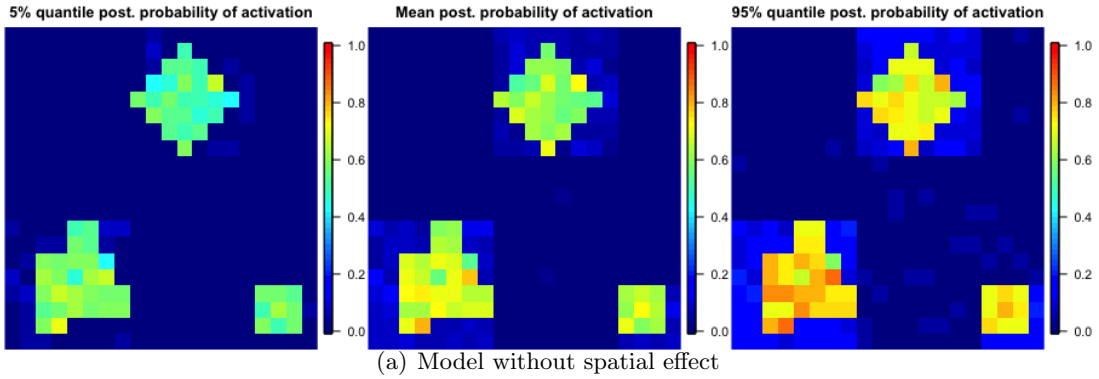


Figure 6.13: Proportions of activated voxels from approaches A, B and C.

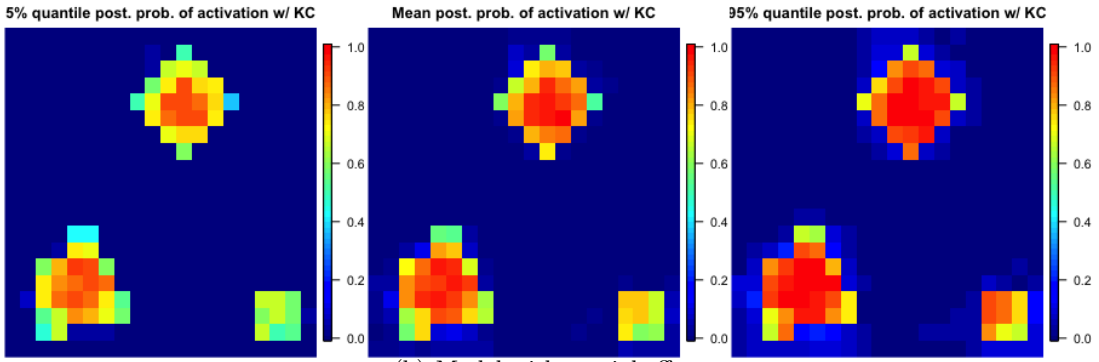
In sum, we find that the multi-subject models eliminate isolated false positives that are generated from fitting single-subject C-EMVS independently. When the local spatial structure is included in the model, more true positives are discovered and voxels tend to be activated in clusters.

6.4 Discussion

This chapter discusses a multi-subject model that can borrow information from all subjects and improve activation detection at the subject level. The model can also learn the brain connectivity structure via the correlation between the connectivity regions. Even without spatial effects that describe the local dependence, the



(a) Model without spatial effect



(b) Model with spatial effect

Figure 6.14: Average posterior probability map over 30 subjects. Left to right: 2.5% quantile, posterior mean, 95% quantile.

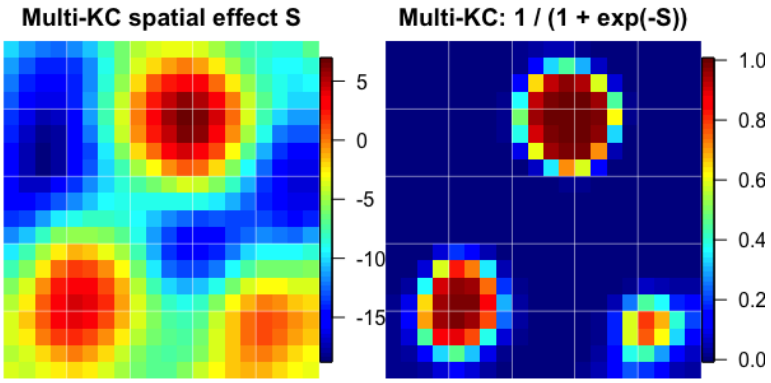


Figure 6.15: Posterior mean spatial effect and its logistic transformation from the full multi-subject model.

multi-subject model eliminates false positives shown in the images from C-EMVS. The common population spatial effect $\boldsymbol{S}$ introduces local dependence and further enhances detection power.

Chapter 7

Conclusions

Several application areas, particularly in engineering, require the analysis of complex-valued datasets. For instance, some radar, sonar, and imaging data are generated after Fourier transformation (FT) and hence are complex-valued. Neuroimaging modalities such as fMRI also employ FT and/or inverse FT, and therefore lead to real and imaginary data, or equivalently, magnitude and phase fMRI. However, mainstream MR scanners do not output complex-valued raw data by default, but only produce the magnitude part of complex-valued signals for analysis. Such default setting discard the phase data. As a result, researchers do not use the complete complex-valued data. In this thesis we show that statistical models using the entire complex-valued data improve fMRI brain activation. Therefore, we recommend neuroimaging researchers to change the MRI scanner setting and make the CV-fMRI data available for analysis.

We first build a complex-valued Bayesian variable selection model to detect which voxels are activated in task-related CV-fMRI data. A novel complex normal

spike-and-slab mixture prior is used for variable selection. The proposed computationally efficient C-EMVS model described in Chapter 4 is shown to produce less false positives and more true positives than the magnitude-only models, including magnitude-only EMVS as well as popular lasso and adaptive lasso, especially when SNR is low, which occurs more frequently in practice due to higher voxel resolutions of advanced MRI technology. We show that the C-EMVS leads to better activation strength and noise variance estimates. We also consider a version of the complex-valued variable selection model that includes a temporal autoregressive process for improved activation detection. In addition, a complex-valued version of MCMC algorithm is developed to quantify the uncertainty of activation strength.

In order to explicitly describe spatial dependence of voxels in the image, we then build a complex-valued Bayesian spatial model via kernel convolution that further improves detection performance. The kernel-based model takes advantage of dimension reduction and leads to computational efficiency. We show that this spatial model encourages local spatial dependence and activation in clusters, which is more biologically plausible and further increases detection power. The model is also able to remove isolated false positives and improves specificity. Furthermore, unlike models using MRFs, the proposed model does not need us to pre-specify the neighboring structure of a voxel. Instead, the neighboring structure is learned from the data.

Finally, we develop a multi-subject model that can infer brain activation at the voxel level and connectivity or co-activation across brain regions by borrowing information across multiple participants. The conjugacy of normal and Wishart priors

in the model keeps the model simple and computationally feasible. We show that this multi-subject model can capture how regions of interests are correlated and leads to less voxels incorrectly labeled as active (false positives) than fitting single-subject models on each subject separately.

Notice that all of our proposed models are built in Cartesian coordinates and based on complex-valued linear regression. This linearity makes the models and their associated algorithms computationally efficient. More importantly, as shown in Chapter 4, the performance of our proposed models can be as good as some nonlinear models that use the MR signal equation, for example, the DeTeCT-ING model proposed by Karaman et al. (2014).

In terms of future work, in addition to the complex normal distribution, we plan to consider using other distributions for complex random variables. We will also work on alternative approaches for inferring connectivity, especially on how to utilize the correlation of the real and imaginary parts of the signals. For this purpose, we expect to collect more real human CV-fMRI data and apply the multi-subject models to human multi-subject CV-fMRI. Finally, as MR technology requires many preprocessing steps to get the raw data, a key question is how these preprocessing steps affect complex-valued data and the resulting activation and connectivity inference. Complex-valued and magnitude data may behave differently depending on the pre-processing steps. For example, magnitude data may be more sensitive to the spatial-smoothing steps used in the preprocessing pipeline than complex-valued data, as complex-valued data contain some additional spatial information.

In sum, we develop complex-valued Bayesian statistical models and computationally efficient algorithms for detecting fMRI brain activation and connectivity. The advantages of using the complex-valued data and our proposed models for inferring activation and connectivity will hopefully encourage more neuroimaging researchers to utilize the complete CV-fMRI data. While many sophisticated real-valued spatial-temporal models have been proposed to improve activation detection performance, complex-valued modeling provides an opportunity to further improve activation detection.

Finally, although our proposed models are developed for CV-fMRI applications, these methods are general, and we believe that they can also be applied to complex-valued data from different fields.

Bibliography

Adali, T., Schreier, P. & Scharf, L. L. (2011). Complex-valued signal processing: The proper way to deal with impropriety. *IEEE Transactions on Signal Processing* **59**, 5101–5125.

Adrian, D. W., Maitra, R. & Rowe, D. B. (2013). Ricean over Gaussian modelling in magnitude fMRI analysis–added complexity with negligible practical benefits. *Stat* **2**, 303–316.

Adrian, D. W., Maitra, R. & Rowe, D. B. (2017). Complex-valued time-series modeling for improved accuracy of activation detection in fMRI studies. *Annals of Applied Statistics* .

Andersen, A., Gash, D. & Avison, M. (1999). Principal component analysis of the dynamic response measured by fMRI: A generalized linear systems framework. *Magnetic Resonance Imaging* **17**, 795–815.

Baldassarre, A., Ramsey, L., Siegel, J., Shulman, G. & Corbetta, M. (2016). Brain connectivity and neurological disorders after stroke. *Current Opinion in Neurology* **29**, 706–713.

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S. & Hyde, J. S. (1992). Time course EPI of human brain function during task activation. *Magnetic Resonance in Medicine* **25**, 390–97.

Barbieri, M. M. & Berger, J. O. (2004). Optimal predictive model selection. *Annals of Statistics* **32**, 870–897.

Beckmann, C. F., Jenkinson, M. & M.Smith, S. (2003). General multilevel linear modeling for group analysis in fMRI. *NeuroImage* **20**, 1052–1063.

Bernstein, M. A., Thomasson, D. M. & Perman, W. H. (1989). Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction. *Medical Physics* **16**, 813–817.

Bezener, M., Hughes, J. & Jones, G. (2017). Bayesian spatiotemporal modeling using hierarchical spatial priors, with applications to functional magnetic resonance imaging. Technical report, University of Minnesota.

van den Bos, A. (1995). The multivariate complex normal distribution – a generalization. *IEEE Transactions on Information Theory* **41**, 537–539.

Bowman, D. F., Caffo, B., Bassett, S. S. & Kilts, C. (2008). A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* **39**, 146–156.

Brandwood, D. H. (1983). A complex gradient operator and its application in adaptive array theory. *Communications, Radar and Signal Processing, IEE Proceedings F* **130**, 11–16.

Caballero-Gaudes, C. & Reynolds, R. C. (2017). Methods for cleaning the BOLD fMRI signal. *NeuroImage* **154**, 128–149.

Calhoun, V., Adali, T., Pearlson, G. & Pekar, J. (2001). A method for making group differences from function MRI data using independent component analysis. *Human Brain Mapping* **14**, 140–151.

Calhoun, V. D., Adali, T., Pearlson, G. D., van Zijl, P. C. M. & Pekar, J. J. (2002). Independent component analysis of fMRI data in the complex domain. *Magnetic Resonance in Medicine* **48**, 180–192.

Ciuciu, P., Poline, J.-B., Marrelec, G., Idier, J., Pallier, C. & Benali, H. (2003). Unsupervised robust nonparametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Transactions on Medical Imaging* **22**, 1235–1251.

Du, W., Fu, G.-S., Calhoun, V. D. & Adah, T. (2014). Performance of complex-valued ICA algorithms for fMRI analysis: Importance of taking full diversity into account. *2014 IEEE International Conference on Image Processing (ICIP)* .

Du, W., Levin-Schwartz, Y., Fu, G.-S., Ma, S., Calhoun, V. D. & Adah, T. (2016). The role of diversity in complex ICA algorithms for fMRI analysis. *Journal of Neuroscience Methods* **264**, 129–135.

Eklund, A., Nichols, T. & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *PNAS* **113**, 7900–7905.

Flandin, G. & Penny, W. (2007). Bayesian fMRI data analysis with sparse spatial basis function priors. *NeuroImage* **34**, 1108–1125.

Friedman, J., Hastie, T. & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**, 1–22.

Friston, K., Bastos, A., Litvak, V., Stephan, K., Fries, P. & Moran, R. (2012). DCM for complex-valued data: Cross-spectra, coherence and phase-delays. *NeuroImage* **59**, 439–455.

Friston, K., Harrison, L. & Penny, W. (2003). Dynamic causal modelling. *NeuroImage* **19**, 1273–1302.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping* **2**, 56–78.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E. & Penny, W. D. (2007). *Statistical parametric mapping: The analysis of functional brain images*. London: Academic Press.

Friston, K. J., Jezzard, P. & Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping* **1**, 153–171.

Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G. & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage* **16**, 465–483.

George, E. I. & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.

George, E. I. & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339–373.

Goebel, R., Roebroeck, A., Kim, D. & Formisano, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic Resonance Imaging* **21**, 1251–1261.

Gossl, C., Auer, D. & Fahrmeir, L. (2001). Bayesian spatio-temporal inference in functional magnetic resonance imaging. *Biometrics* **57**, 554–562.

Goutte, C., Nielsen, F. & Hansen, K. (2000). Modeling the hemodynamic response in fMRI using smooth FIR filters. *IEEE Transactions on Medical Imaging* **19**, 1188–1201.

Gudbjartsson, H. & Patz, S. (1995). The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine* **34**, 910–914.

Haacke, E., Brown, R., Thompson, M. & Venkatesan, R. (1999). *Magnetic Resonance Imaging: Physical Principles and Sequence Design.* New York: John Wiley and Sons.

Harrison, L. & Green, G. (2010). A Bayesian spatiotemporal model for very large data sets. *NeuroImage* **50**, 1126–1141.

Harrison, L., Penny, W., Daunizeau, J. & Friston, K. (2008). Diffusion-based spatial priors for functional magnetic resonance images. *NeuroImage* **41**, 408–423.

Harrison, L., Penny, W. & Friston, K. (2003). Multivariate autoregressive modeling of fMRI time series. *NeuroImage* **19**, 1477–1491.

Higdon, D. (1998). A process-convolution approach to modeling temperatures in the north Atlantic ocean. *Environmental and Ecological Statistics* **5**, 173–190.

Higdon, D. (2002). Space and space-time modeling using process convolutions. In *Quantitative methods for current environmental issues*, Eds. C. W. A. By A. H. El-Shaarawi, V. Barnett & P. C. Chatwin. Oxford: Springer London.

Hjørungnes, A. & Gesbert, D. (2007). Complex-valued matrix differentiation: Techniques and key results. *IEEE Transactions on Signal Processing* **55**, 2740–2746.

Iturria-Medina, Y. & Evans, A. C. (2015). On the central role of brain connectivity in neurodegenerative disease progression. *Frontiers in Aging Neuroscience* **7:90**, –.

Kalus, S., Samann, P. G. & Fahrmeir, L. (2014). Classification of brain activation via spatial Bayesian variable selection in fMRI regression. *Advanced Data Analysis Classification* .

Karaman, M., Bruce, I. P. & Rowe, D. B. (2015). Incorporating relaxivities to more accurately reconstruct MR images. *Magnetic Resonance Imaging* **33**, 374–384.

Karaman, M. M., Bruce, I. P. & Rowe, D. B. (2014). A statistical fMRI model

for differential T2* contrast incorporating T1 and T2 of gray matter. *Magnetic Resonance Imaging* **32**, 9–27.

Kociuba, M. C. & Rowe, D. B. (2016). Complex-valued time-series correlation increases sensitivity in fMRI analysis. *Magnetic Resonance Imaging* **34**, 765–770.

Kuo, L. & Mallick, B. (1998). Variable selection for regression models. *Sankhya: The Indian Journal of Statistics, Series B* **60**, 65–81.

Lai, S. & Glover, G. H. (1997). Detection of BOLD fMRI signals using complex data. *Proc ISMRM* p. 1671.

Lazar, N. A. (2008). *The Statistical Analysis of Functional MRI Data*. Springer.

Lee, H., Higdon, D., Calder, K. & Holloman, C. (2005). Efficient models for correlated data via convolutions of intrinsic processes. *Statistical Modelling* **5**, 53–74.

Lee, J., Shahram, M. & Pauly, J. M. (2009). Combining complex signal change in functional MRI. *Magnetic Resonance in Medicine* **62**, 1358–1360.

Lee, J., Shahram, M., Schwartzman, A. & Pauly, J. M. (2007). Complex data analysis in high-resolution SSFP fMRI. *Magnetic Resonance in Medicine* **57**, 905:917.

Lee, K. J., Jones, G. L., Caffo, B. S. & Bassett, S. S. (2014). Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. *Bayesian Analysis* **9**, 699–732.

Lindquist, M. (2008). The statistical analysis of fMRI data. *Statistical Science* **23**, 439–464.

Lindquist, M. A., Loh, J. M., Atlas, L. Y. & Wager, T. D. (2009). Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage* **45**, 187–198.

Lu, Y., Jiang, T. & Zang, Y. (2003). Region growing method for the analysis of functional MRI data. *NeuroImage* **20**, 455–465.

Macovski, A. (1996). Noise in MRI. *Magnetic Resonance in Medicine* **36**, 494–497.

Mandic, P. D. & Goh, V. S. L. (2009). *Complex-valued Nonlinear Adaptive Filters*. Great Britain: John Wiley & Sons.

McKeown, M., Makeig, S., Brown, G., Jung, T., Kindermann, S. & abd T. Sejnowski, A. B. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping* **6**, 160–188.

Mclntosh, A. & Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping* **2**, 2–22.

Müller, P., Parmigiani, G. & Rice, K. (2006). FDR and Bayesian multiple comparisons rule. In *Bayesian Statistics 8 (J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West, eds.)*. Oxford University Press.

Müller, P., Parmigiani, G., Robert, C. & Rouseau, J. (2004). Optimal sample size for multiple testing: the case of gene expression microarrays. *Journal of American Statistical Association* **99**, 990–1001.

Ng, B., Abugharbieh, R., Varoquaux, G., Poline, J. B. & Thirion, B. (2011). Connectivity-informed fMRI activation detection. *Medical Image Computing and Computer-Assisted Intervention MICCAI 2011* pp. 285–292.

O'Donoughue, N. & Moura, J. M. F. (2012). On the product of independent complex Gaussians. *IEEE Transactions on Signal Processing* **60**, 1050–1063.

O'Hara, R. & Sillanpaa, M. J. (2009). A review of Bayesian variable selection methods: What, how and which. *Bayesian Analysis* **4**, 85–118.

Oppenheim, A. V. & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE* **69**, 529–541.

Pauli, R., Bowring, A., Reynolds, R., Chen, G., Nichols, T. E. & Maume, C. (2016). Exploring fMRI results space: 31 variants of an fMRI analysis in AFNI, FSL, and SPM. *Frontiers in Neuroinformatics* **10**.

Pedregosa, F., Eickenberg, M., Ciuciu, P., Thirion, B. & Gramfort, A. (2015). Data-driven HRF estimation for encoding and decoding models. *NeuroImage* **104**, 209–220.

Penny, W., Kiebel, S. & Friston, K. (2003). Variational Bayesian inference for fMRI time series. *NeuroImage* **19**, 727–741.

Penny, W., Trujillo-Barreto, N. & Friston, K. (2005). Bayesian fMRI time series analysis with spatial priors. *NeuroImage* **24**, 350–362.

Peters, A. M., Broooks, M. J., Hoogenrad, F. G., Gowland, P. A., Francis, S. T., Morris, P. G. & Bowtell, R. W. (2006). Comparison of T2* measurements in human brain at 1.5, 3 and 7 T. In *Proc. Intl. Soc. Magn. Reson. Med.*, volume 14, p. 926. Seattle, WA, USA: John Wiley & Sons.

Picinbono, B. (1996). Second-order complex random vectors and normal distributions. *IEEE Transactions on Signal Processing* **44**, 2637–2640.

Pievani, M., Filippini, N., van den Heuvel, M. P., Cappa, S. F. & Frisoni, G. B. (2014). Brain connectivity in neurodegenerative diseasesfrom phenotype to proteinopathy. *Nature Reviews Neurology* **10**, 620–633.

Poldrack, R. A., Mumford, J. A. & Nichols, T. E. (2011). *Handbook of Functional MRI Data Analysis*. Cambridge.

Quiros, A., Diez, R. & Gamerman, D. (2010). Bayesian spatiotemporal model of fMRI data. *NeuroImage* **49**, 442–456.

Raftery, A. (1998). Bayes factors and BIC: Comment on Weakliem. Technical report, University of Washington.

Raftery, A. E. (1996). Hypothesis testing and model selection. *Markov Chain Monte Carlo in Practice, Chapman and Hall, London* .

Rice, S. O. (1944). Mathematical analysis of random noise. *The Bell System Technical Journal* **23**, 282–332.

Rockova, V. & George, E. I. (2014). EMVS: The EM approach to Bayesian variable selection. *Journal of the American Statistical Association* **109**, 828–846.

Rowe, D. B. (2001). A Bayesian source separation for reference function determination in fMRI. *Magnetic Resonance in Medicine* **46**, 374–378.

Rowe, D. B. (2005a). Modeling both the magnitude and phase of complex-valued fMRI data. *NeuroImage* **25**, 1310–1324.

Rowe, D. B. (2005b). Parameter estimation in the magnitude-only and complex-valued fMRI data models. *NeuroImage* **25**, 1124–1132.

Rowe, D. B. (2009). Magnitude and phase signal detection in complex-valued fMRI data. *Magnetic Resonance in Medicine* **62**, 1356–1357.

Rowe, D. B. & Logan, B. R. (2004). A complex way to compute fMRI activation. *NeuroImage* **23**, 1078–1092.

Rowe, D. B. & Logan, B. R. (2005). Complex fMRI analysis with unrestricted phase is equivalent to a magnitude-only model. *NeuroImage* **24**, 603–606.

Smith, M. & Fahrmeir, L. (2007). Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *Journal of the American Statistical Association* **102**, 417–431.

Sun, W., Reich, B., Cai, T., Guindani, M. & Schwartzman, A. (2015). False discovery control in large-scale spatial multiple testing. *Journal of the Royal Statistical Association, Series B* **77**, 59–83.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* **58**, 267–288.

Varoquaux, G., Gramfort, A., Poline, J.-B. & Thirion, B. (2010). Brain covariance selection: Better individual functional connectivity models using population prior. *Advances in Neural Information Processing Systems* pp. 2334–2342.

Wager, T. D. & Lindquist, M. A. (2015). *Principles of fMRI*. New York: Leanpub.

Wang, J., Liang, F., Li, Y. & Zhu, Y. (2015). A scalable algorithm for Bayesian variable selection (SAR) with application to miRNA-mRNA regulation in cancer. Technical report, Department of Statistics, University of Illinois at Urbana-Champaign.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G. & Rosseel, Y. (2011). neurosim: An R package for generating fMRI data. *Journal of Statistical Software* **44**, 1–18.

Wooding, R. A. (1956). The multivariate distribution of complex normal variables. *Biometrika* **43**, 212–215.

Woolrich, M., Jenkinson, M., Brady, J. & Smith, S. (2004). Fully Bayesian spatiotemporal modeling of fMRI data. *IEEE Transactions on Medical Imaging* **23**, 213–231.

Xia, J., Liang, F. & Wang, Y. (2009). FMRI analysis through Bayesian variable selection with a spatial prior. In *Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 714–717. Boston, MA, USA.

Yoldemir, B., Ng, B., Woodward, T. S. & Abugharbieh, R. (2013). Fiber connectivity integrated brain activation detection. *Information Processing in Medical Imaging* pp. 135–146.

Yu, C., Prado, R., Ombao, H. & Rowe, D. B. (2018). A Bayesian variable selection approach yields improved detection of brain activation from complex-valued fMRI. *Journal of American Statistical Association: Application and Case Studies Accepted* .

Yu, Z., Prado, R., Quinlan, E. B., Cramer, S. C. & Ombao, H. (2016). Understanding the impact of stroke on brain motor function: A hierarchical Bayesian approach. *Journal of American Statistical Association: Application and Case Studies* **111**, 549–563.

Zhang, L., Guindani, M. & Vannucci, M. (2015). Bayesian models for fMRI data analysis. *WIREs Computational Statistics* **7**, 21–41.

Zhang, L., Guindani, M., Versace, F. & Vannucci, M. (2014). A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. *NeuroImage* **95**, 162–175.

Zheng, X. & Rajapakse, J. (2006). Learning functional structure from fMRI images. *NeuroImage* **31**, 1601–1613.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429.

Appendix A

Derivation of the C-EMVS and C-SSVS algorithms

A.1 Derivation of the C-EMVS algorithm

Assume that $\mathbf{y}^v \sim CN_T\left(\mathbf{X}\boldsymbol{\gamma}^v, 2\sigma_v^2\mathbf{I}, \mathbf{0}\right)$ with a prior for $\boldsymbol{\gamma}^v$ of the form (4.3), i.e., $\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v \sim CN_p\left(\mathbf{0}, \sigma_v^2\boldsymbol{\Omega}_v, \sigma_v^2\boldsymbol{\Lambda}_v\right)$ with $\boldsymbol{\Omega}_v = \text{diag}\left(\kappa_1^v, \ldots, \kappa_p^v\right)$, and $\boldsymbol{\Lambda}_v = \text{diag}\left(\tau_1^v, \ldots, \tau_p^v\right)$, where $\kappa_j^v = \left(1 - \psi_j^v\right)\omega_0 + \psi_j^v\omega_1$ and $\tau_j^v = \left(1 - \psi_j^v\right)\lambda_0 + \psi_j^v\lambda_1$ for $j = 1, \ldots, p$. In addition, assume $\sigma_v^2 \overset{\text{iid}}{\sim} IG\left(a_\sigma, b_\sigma\right)$, $\psi_j^v \overset{\text{iid}}{\sim} Bernoulli\left(\theta_j\right)$, and $\theta_j \overset{\text{iid}}{\sim} Beta\left(a_\theta, b_\theta\right)$.

Now, using the notation introduced in Section 4.2, let $\boldsymbol{\gamma} = \left[\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^V\right]$, with $\boldsymbol{\psi} = \left[\boldsymbol{\psi}^1, \ldots, \boldsymbol{\psi}^V\right]$, $\boldsymbol{\psi}^v = \left(\psi_1^v, \ldots, \psi_p^v\right)'$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$, $\boldsymbol{\sigma}^2 = \left[\sigma_1^2, \ldots, \sigma_V^2\right]$, and $\mathbf{y} = \left[\mathbf{y}^1, \ldots, \mathbf{y}^V\right]$. Given the values of the model parameters at the l-th step of the EM

algorithm, denoted as $\boldsymbol{\gamma}^{(l)}$, $\boldsymbol{\theta}^{(l)}$ and $\boldsymbol{\sigma}^{2,(l)}$, the objective function is given by

$$Q\left(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2 \mid \boldsymbol{\gamma}^{(l)}, \boldsymbol{\theta}^{(l)}, \boldsymbol{\sigma}^{2,(l)}\right) = \mathrm{E}_{\boldsymbol{\psi}|\cdot}\left[\log \pi\left(\boldsymbol{\gamma}, \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2 \mid \mathbf{y}\right) \,\middle|\, \boldsymbol{\gamma}^{(l)}, \boldsymbol{\theta}^{(l)}, \boldsymbol{\sigma}^{2,(l)}, \mathbf{y}\right],$$

where $\mathrm{E}_{\boldsymbol{\psi}|\cdot}$ denotes the conditional expectation of $\boldsymbol{\psi}$ given $\boldsymbol{\gamma}^{(l)}, \boldsymbol{\theta}^{(l)}, \boldsymbol{\sigma}^{2,(l)}$ and $\mathbf{y}$, i.e., $\mathrm{E}_{\boldsymbol{\psi}|\boldsymbol{\gamma}^{(l)}, \boldsymbol{\theta}^{(l)}, \boldsymbol{\sigma}^{2,(l)}, \mathbf{y}}$.

Using the complex-valued normal density (Wooding, 1956; van den Bos, 1995; Picinbono, 1996), we write the posterior distribution as

$$\pi\left(\boldsymbol{\gamma}, \theta, \boldsymbol{\sigma}^2, \boldsymbol{\psi} \mid \mathbf{y}\right) \propto \prod_{v=1}^{V}\left[f\left(\mathbf{y}^v \mid \boldsymbol{\gamma}^v, \sigma_v^2\right) \pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v, \sigma_v^2\right) \pi\left(\boldsymbol{\psi}^v \mid \boldsymbol{\theta}\right) \pi\left(\sigma_v^2\right)\right] \pi\left(\boldsymbol{\theta}\right),$$

where for each voxel v, we have

$$f\left(\mathbf{y}^v \mid \boldsymbol{\gamma}^v, \sigma_v^2\right) \propto \left|2\sigma_v^2 \mathbf{I}\right|^{-1} \exp\{-(\mathbf{y}^v - \mathbf{X}\boldsymbol{\gamma}^v)^H \left(2\sigma_v^2 \mathbf{I}\right)^{-1} (\mathbf{y}^v - \mathbf{X}\boldsymbol{\gamma}^v)\},$$

$$\begin{aligned}\pi\left(\boldsymbol{\gamma}^v \mid \sigma_v^2, \boldsymbol{\psi}\right) &\propto \left[\left|\sigma_v^2 \boldsymbol{\Omega}_v\right| \left|\boldsymbol{P}_v\right|\right]^{-1/2} \\ &\quad \times \exp\left\{-\left(\boldsymbol{\gamma}^v\right)^H \boldsymbol{P}_v^{-*} \boldsymbol{\gamma}^v + \frac{1}{2}\left[\left(\boldsymbol{\gamma}^v\right)' \boldsymbol{R}_v' \boldsymbol{P}_v^{-*} \boldsymbol{\gamma}^v + \left(\boldsymbol{\gamma}^v\right)^H \boldsymbol{R}_v^H \boldsymbol{P}_v^{-1} \left(\boldsymbol{\gamma}^v\right)^*\right]\right\},\end{aligned}$$

$\pi\left(\sigma_v^2\right) \propto \left(\sigma_v^2\right)^{-(a_\sigma+1)} \exp\{-b_\sigma/\sigma_v^2\}$, $\pi\left(\boldsymbol{\psi}^v \mid \boldsymbol{\theta}\right) \propto \prod_{j=1}^{p} \theta_j^{\psi_j^v} \left(1-\theta_j\right)^{1-\psi_j^v}$, and $\pi\left(\boldsymbol{\theta}\right) \propto \prod_{j=1}^{p} \theta_j^{a_\theta - 1} \left(1-\theta_j\right)^{b_\theta - 1}$, with

$$\boldsymbol{P}_v = \left(\sigma_v^2 \boldsymbol{\Omega}_v\right)^* - \left(\sigma_v^2 \boldsymbol{\Lambda}_v\right)^H \left(\sigma_v^2 \boldsymbol{\Omega}_v\right)^{-1} \left(\sigma_v^2 \boldsymbol{\Lambda}_v\right) = \sigma_v^2\left[\boldsymbol{\Omega}_v^* - \boldsymbol{\Lambda}_v^H \boldsymbol{\Omega}_v^{-1} \boldsymbol{\Lambda}_v\right],$$

and $\boldsymbol{R}_v = \left(\sigma_v^2 \boldsymbol{\Lambda}_v\right)^H \left(\sigma_v^2 \boldsymbol{\Omega}_v\right)^{-1} = \boldsymbol{\Lambda}_v^H \boldsymbol{\Omega}_v^{-1}$. Here $\boldsymbol{Z}^*$ denotes the complex conjugate of a matrix $\boldsymbol{Z}$ and $\boldsymbol{Z}^H$ denotes its conjugate transpose, i.e., $\boldsymbol{Z}^H = (\boldsymbol{Z}^*)'$.

Hence, the log-posterior distribution is

$$\begin{aligned}\log \pi\left(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\psi} \mid \mathbf{y}\right) &= \sum_{v=1}^{V}\Bigg\{ -\frac{1}{2\sigma_v^2}\left(\mathbf{y}^v - \mathbf{X}\boldsymbol{\gamma}^v\right)^H\left(\mathbf{y}^v - \mathbf{X}\boldsymbol{\gamma}^v\right) - (T+p+a_\sigma+1)\log\left(\sigma_v^2\right) \\ &\quad -\frac{b_\sigma}{\sigma_v^2} - \left(\boldsymbol{\gamma}^v\right)^H \boldsymbol{P}_v^{-*}\boldsymbol{\gamma}^v + \frac{1}{2}\left[\left(\boldsymbol{\gamma}^v\right)' \boldsymbol{R}_v' \boldsymbol{P}_v^{-*}\boldsymbol{\gamma}^v + \left(\boldsymbol{\gamma}^v\right)^H \boldsymbol{R}_v^H \boldsymbol{P}_v^{-1}\left(\boldsymbol{\gamma}^v\right)^*\right] \\ &\quad + \sum_{j=1}^{p}\psi_j^v \log\left(\theta_j\right) + \sum_{j=1}^{p}\left(1-\psi_j^v\right)\log\left(1-\theta_j\right)\Bigg\} \\ &\quad + (a_\theta - 1)\sum_{j=1}^{p}\log\left(\theta_j\right) + (b_\theta - 1)\sum_{j=1}^{p}\log\left(1-\theta_j\right) + K_\pi,\end{aligned}$$

for some constant K_π.

The log-posterior above can be further simplified using some complex-valued linear algebra. First, note that $\boldsymbol{\Lambda}_v$ is a symmetric matrix, which implies that $\boldsymbol{\Lambda}_v' = \boldsymbol{\Lambda}_v$ and $\boldsymbol{\Lambda}_v^H = \boldsymbol{\Lambda}_v^*$, and hence $\boldsymbol{P}_v^* = \sigma_v^2\left[\boldsymbol{\Omega}_v - \boldsymbol{\Lambda}_v\boldsymbol{\Omega}_v^{-*}\boldsymbol{\Lambda}_v^*\right]$. Using the Sherman-Morrison-Woodbury formula we have

$$\boldsymbol{P}_v^{-*} = \left(\sigma_v^2\right)^{-1}\left[\boldsymbol{\Omega}_v^{-1} + \boldsymbol{\Omega}_v^{-1}\boldsymbol{\Lambda}_v\boldsymbol{P}_v^{-1}\boldsymbol{\Lambda}_v^H\boldsymbol{\Omega}_v^{-1}\right],$$

which, given the structure assumed for $\boldsymbol{\Omega}_v$ and $\boldsymbol{\Lambda}_v$, can be written as

$$\begin{aligned}\boldsymbol{P}_v^{-*} &= \left(\sigma_v^2\right)^{-1}\operatorname{diag}\left(\frac{1}{\kappa_1^v} + \frac{\left(\tau_1^v\right)^*\tau_1^v}{\kappa_1^v\left(\left(\kappa_1^v\right)^*\kappa_1^v - \left(\tau_1^v\right)^*\tau_1^v\right)}, \ldots, \frac{1}{\kappa_p^v} + \frac{\left(\tau_p^v\right)^*\tau_p^v}{\kappa_p^v\left(\left(\kappa_p^v\right)^*\kappa_p^v - \left(\tau_p^v\right)^*\tau_p^v\right)}\right), \\ &= \left(\sigma_v^2\right)^{-1}\operatorname{diag}\left(\frac{1}{\kappa_1^v}\left(1+b_1^v\right), \ldots, \frac{1}{\kappa_p^v}\left(1+b_p^v\right)\right),\end{aligned}$$

with $b_j^v = \frac{|\tau_j^v|^2}{|\kappa_j^v|^2 - |\tau_j^v|^2} = \frac{\left(\tau_j^v\right)^*\tau_j^v}{\left(\kappa_j^v\right)^*\kappa_j^v - \left(\tau_j^v\right)^*\tau_j^v}$, for $j = 1, \ldots, p$, where b_j^v is real-valued for all j and all v.

Similarly,

$$\begin{aligned}\boldsymbol{R}_v'\boldsymbol{P}_v^{-*} &= \left(\sigma_v^2\right)^{-1}\left(\boldsymbol{\Lambda}_v^H\boldsymbol{\Omega}_v^{-1}\right)'\left(\boldsymbol{\Omega}_v^{-1}+\boldsymbol{\Omega}_v^{-1}\boldsymbol{\Lambda}_v\boldsymbol{P}_v^{-1}\boldsymbol{\Lambda}_v^H\boldsymbol{\Omega}_v^{-1}\right),\\ &= \left(\sigma_v^2\right)^{-1}\left(\boldsymbol{\Omega}_v^{-1}\boldsymbol{\Lambda}_v^*\boldsymbol{\Omega}_v^{-1}+\boldsymbol{\Omega}_v^{-1}\boldsymbol{\Lambda}_v^*\boldsymbol{\Omega}_v^{-1}\boldsymbol{\Lambda}_v\boldsymbol{P}_v^{-1}\boldsymbol{\Lambda}_v^H\boldsymbol{\Omega}_v^{-1}\right),\end{aligned}$$

which can be simplified further by writing

$\boldsymbol{\Omega}_v^{-1}\boldsymbol{\Lambda}_v^*\boldsymbol{\Omega}_v^{-1} = \text{diag}\left(\left(\tau_1^v\right)^*/\left(\kappa_1^v\right)^2,\ldots,\left(\tau_p^v\right)^*/\left(\kappa_p^v\right)^2\right)$, and

$$\boldsymbol{\Omega}_v^{-1}\boldsymbol{\Lambda}_v^*\boldsymbol{\Omega}_v^{-1}\boldsymbol{\Lambda}_v\boldsymbol{P}_v^{-1}\boldsymbol{\Lambda}_v^H\boldsymbol{\Omega}_v^{-1} = \text{diag}\left(b_1^v\frac{\left(\tau_1^v\right)^*}{\left(\kappa_1^v\right)^2},\ldots,b_p^v\frac{\left(\tau_p^v\right)^*}{\left(\kappa_p^v\right)^2}\right).$$

We now proceed with defining some quantities and computing some expectations that are required in the E-step of the EM algorithm. Let $b^{(i)}$ be the value of b_j^v if $\psi_j^v = i$, $i = 0, 1$. That is, $b^{(0)} = \frac{|\lambda_0|^2}{|\omega_0|^2-|\lambda_0|^2}$, and $b^{(1)} = \frac{|\lambda_1|^2}{|\omega_1|^2-|\lambda_1|^2}$. Then, for $j = 1,\ldots,p$, we have expectations

$$\begin{aligned}\text{E}_{\boldsymbol{\psi}|\cdot}\left[\frac{1}{\kappa_j^v}\left(1+b_j^v\right)\right] &= \frac{1}{\omega_0}\left(1+b^{(0)}\right)Pr\left(\psi_j^v=0\mid\cdots\right)+\frac{1}{\omega_1}\left(1+b^{(1)}\right)Pr\left(\psi_j^v=1\mid\cdots\right),\\ &= \frac{1}{\omega_0}\left(1+b^{(0)}\right)\text{E}_{\boldsymbol{\psi}|\cdot}\left(1-\psi_j^v\right)+\frac{1}{\omega_1}\left(1+b^{(1)}\right)\text{E}_{\boldsymbol{\psi}|\cdot}\left(\psi_j^v\right),\end{aligned}$$

$$\begin{aligned}\text{E}_{\boldsymbol{\psi}|\cdot}\left[\frac{\left(\tau_j^v\right)^*}{\left(\kappa_j^v\right)^2}\right] &= \frac{\lambda_0^*}{\omega_0^2}Pr\left(\psi_j^v=0\mid\cdots\right)+\frac{\lambda_1^*}{\omega_1^2}Pr\left(\psi_j^v=1\mid\cdots\right)\\ &= \frac{\lambda_0^*}{\omega_0^2}\text{E}_{\boldsymbol{\psi}|\cdot}\left(1-\psi_j^v\right)+\frac{\lambda_1^*}{\omega_1^2}\text{E}_{\boldsymbol{\psi}|\cdot}\left(\psi_j^v\right),\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left[b_j^v\frac{\tau_j^*}{(\kappa_j^v)^2}\right] &= b^{(0)}\frac{\lambda_0^*}{\omega_0^2}Pr(\psi_j^v=0\mid\cdots)+b^{(1)}\frac{\lambda_1^*}{\omega_1^2}Pr(\psi_j^v=1\mid\cdots)\\ &= b^{(0)}\frac{\lambda_0^*}{\omega_0^2}\mathrm{E}_{\boldsymbol{\psi}|\cdot}(1-\psi_j^v)+b^{(1)}\frac{\lambda_1^*}{\omega_1^2}\mathrm{E}_{\boldsymbol{\psi}|\cdot}(\psi_j^v).\end{aligned}$$

Therefore, the objective function can be written as

$$\begin{aligned}Q\left(\boldsymbol{\gamma},\boldsymbol{\theta},\boldsymbol{\sigma}^2\mid\boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\boldsymbol{\sigma}^{2,(l)}\right)=&\sum_{v=1}^{V}Q_1^v\left(\boldsymbol{\gamma}^v,\sigma_v^2\mid\boldsymbol{\gamma}^{v,(l)},\boldsymbol{\theta}^{(l)},\sigma_v^{2,(l)}\right)\\ &+Q_2\left(\boldsymbol{\theta}\mid\boldsymbol{\gamma}^{(l)},\boldsymbol{\theta}^{(l)},\boldsymbol{\sigma}^{2,(l)}\right)+K_Q,\end{aligned}$$

for some constant K_Q, where

$$\begin{aligned}&Q_1^v\left(\boldsymbol{\gamma}^v,\sigma_v^2\mid\boldsymbol{\gamma}^{v,(l)},\boldsymbol{\theta}^{(l)},\sigma_v^{2,(l)}\right)\\ &\quad=-\frac{1}{2\sigma_v^2}\left(\mathbf{y}^v-\mathbf{X}\boldsymbol{\gamma}^v\right)^H\left(\mathbf{y}^v-\mathbf{X}\boldsymbol{\gamma}^v\right)-\left(T+p+a_\sigma+1\right)\log\left(\sigma_v^2\right)-\frac{b_\sigma}{\sigma_v^2}\\ &\qquad-\frac{1}{\sigma_v^2}\sum_{j=1}^{p}\left(\gamma_j^v\right)^*\left(\gamma_j^v\right)\left[\frac{1}{\omega_0}\left(1+b^{(0)}\right)\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(1-\psi_j^v\right)+\frac{1}{\omega_1}\left(1+b^{(1)}\right)\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(\psi_j^v\right)\right]\\ &\qquad-\frac{1}{\sigma_v^2}\mathrm{Re}\left(\sum_{j=1}^{p}\left(\gamma_j^v\right)^2\left[\frac{\lambda_0^*}{\omega_0^2}\left(1+b^{(0)}\right)\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(1-\psi_j^v\right)+\frac{\lambda_1^*}{\omega_1^2}\left(1+b^{(1)}\right)\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(\psi_j^v\right)\right]\right),\end{aligned}$$

and

$$\begin{aligned}Q_2\left(\boldsymbol{\theta}\mid\boldsymbol{\gamma}^{(l)},\theta^{(l)},\boldsymbol{\sigma}^{2,(l)}\right) &= \sum_{v=1}^{V}\left\{\sum_{j=1}^{p}\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(\psi_j^v\right)\log\left(\theta_j\right)+\sum_{j=1}^{p}\left(1-\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(\psi_j^v\right)\right)\log\left(1-\theta_j\right)\right\}\\ &\quad+\left(a_\theta-1\right)\sum_{j=1}^{p}\log\left(\theta_j\right)+\left(b_\theta-1\right)\sum_{j=1}^{p}\log\left(1-\theta_j\right).\end{aligned}$$

Finally, $p_j^v = \mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(\psi_j^v\right) = Pr\left(\psi_j^v = 1 \mid \boldsymbol{\gamma}^{(l)}, \boldsymbol{\theta}^{(l)}, \boldsymbol{\sigma}^{2,(l)}\right)$ can be computed as $p_j^v = \dfrac{p^1}{p^0 + p^1}$, where $p^0 = \pi\left(\gamma_j^{v,(l)} \mid \sigma_v^{2,(l)}, \psi_j^v = 0, \psi_{-j}^v\right) Pr\left(\psi_j^v = 0 \mid \boldsymbol{\theta}^{(l)}, \psi_{-j}^v\right)$ and $p^1 = \pi\left(\gamma_j^{v,(l)} \mid \sigma_v^{2,(l)}, \psi_j^v = 1, \psi_{-j}^v\right) Pr\left(\psi_j^v = 1 \mid \boldsymbol{\theta}^{(l)}, \psi_{-j}^v\right)$. This completes the E-step of the algorithm.

We now continue with the maximization step of the EM algorithm. For each voxel v, define

$$d_j^v = \left[\frac{1}{\omega_0}\left(1 + b^{(0)}\right)\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(1 - \psi_j^v\right) + \frac{1}{\omega_1}\left(1 + b^{(1)}\right)\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(\psi_j^v\right)\right]$$

and

$$g_j^v = \left[\frac{\lambda_0^*}{\omega_0^2}\left(1 + b^{(0)}\right)\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(1 - \psi_j^v\right) + \frac{\lambda_1^*}{\omega_1^2}\left(1 + b^{(1)}\right)\mathrm{E}_{\boldsymbol{\psi}|\cdot}\left(\psi_j^v\right)\right].$$

We begin with maximization of Q_1^v. Regardless of $\sigma_v^{2,(l+1)}$, we solve the following problem for $\boldsymbol{\gamma}^v$:

$$\left(\boldsymbol{\gamma}^v\right)^{(l+1)} = \arg\min_{\boldsymbol{\gamma}^v \in \mathbb{C}^p} h\left(\boldsymbol{\gamma}^v\right),$$

where $h\left(\boldsymbol{\gamma}^v\right) : \mathbb{C}^p \to \mathbb{R}$ is a real-valued function of complex-valued vectors given by

$$\begin{aligned} h\left(\boldsymbol{\gamma}^v\right) =& \frac{1}{2}\|\mathbf{y}^v - \mathbf{X}\boldsymbol{\gamma}^v\|^2 + \left(\boldsymbol{\gamma}^v\right)^H \mathbf{D}_v \boldsymbol{\gamma}^v - \frac{1}{2}\left[\left(\boldsymbol{\gamma}^v\right)' \mathbf{G}_v \boldsymbol{\gamma}^v + \left(\boldsymbol{\gamma}^v\right)^H \mathbf{G}_v^* \left(\boldsymbol{\gamma}^v\right)^*\right] \\ =& \frac{1}{2}\left[\left(\boldsymbol{\gamma}^v\right)^H \mathbf{X}^H \mathbf{X}\boldsymbol{\gamma}^v - \left(\boldsymbol{\gamma}^v\right)^H \mathbf{X}^H \mathbf{y}^v - \left(\mathbf{y}^v\right)^H \mathbf{X}\boldsymbol{\gamma}^v\right] + \left(\boldsymbol{\gamma}^v\right)^H \mathbf{D}_v \boldsymbol{\gamma}^v + \\ & - \frac{1}{2}\left[\left(\boldsymbol{\gamma}^v\right)' \mathbf{G}_v \boldsymbol{\gamma}^v + \left(\boldsymbol{\gamma}^v\right)^H \mathbf{G}_v^* \left(\boldsymbol{\gamma}^v\right)^*\right] + K_h \end{aligned}$$

with K_h a constant and with $\mathbf{D}_v = diag\left(d_1^v, \ldots, d_p^v\right) = \mathrm{E}_{\boldsymbol{\psi}|\cdot}[\boldsymbol{P}_v^{-*}]$ and

$\mathbf{G}_v = diag\left(g_1^v, \ldots, g_p^v\right) = \mathrm{E}_{\boldsymbol{\psi}|\cdot}\left[\boldsymbol{R}_v' \boldsymbol{P}_v^{-*}\right]$ and $\mathbf{G}_v^* = \mathrm{E}_{\boldsymbol{\psi}|\cdot}\left[\boldsymbol{R}_v^H \boldsymbol{P}_v^{-1}\right]$.

In general, we can view $h(\boldsymbol{\gamma})$ as a function of two variables, $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}^*$. If h is analytic in $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}^*$ independently, i.e., if h is partially differentiable with respect to $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}^*$, then, following Brandwood (1983), either condition $\partial h/\partial \boldsymbol{\gamma}$ or $\partial h/\partial \boldsymbol{\gamma}^*$ will allow us to find a stationary point that minimizes h. Note that in our case, given the Gaussian structure of our priors and noise terms the analytic condition of h is satisfied. Now, using Table IV in Hjørungnes & Gesbert (2007) (or Table 4.2 in Hjørungnes (2011)), we have

$$\begin{aligned}
\partial\left(\boldsymbol{\gamma}^H \mathbf{X}'\mathbf{X}\boldsymbol{\gamma}\right)/\partial\boldsymbol{\gamma} &= \boldsymbol{\gamma}^H \mathbf{X}'\mathbf{X}, \quad \partial\left(\boldsymbol{\gamma}^H \mathbf{D}\boldsymbol{\gamma}\right)/\partial\boldsymbol{\gamma} = \boldsymbol{\gamma}^H \mathbf{D}, \quad \partial\left(\boldsymbol{\gamma}^H \mathbf{X}'\mathbf{y}\right)/\partial\boldsymbol{\gamma} = 0, \\
\partial\left(\mathbf{y}^H \mathbf{X}\boldsymbol{\gamma}\right)/\partial\boldsymbol{\gamma} &= \mathbf{y}^H \mathbf{X}, \quad \partial\left(\boldsymbol{\gamma}'\mathbf{G}\boldsymbol{\gamma}\right)/\partial\boldsymbol{\gamma} = \gamma'\left(\mathbf{G} + \mathbf{G}'\right), \quad \partial\left(\boldsymbol{\gamma}^H \mathbf{G}^*\boldsymbol{\gamma}^*\right)/\partial\boldsymbol{\gamma} = 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\partial\left(\boldsymbol{\gamma}^H \mathbf{X}'\mathbf{X}\boldsymbol{\gamma}\right)/\partial\boldsymbol{\gamma}^* &= \boldsymbol{\gamma}'\left(\mathbf{X}'\mathbf{X}\right)', \quad \partial\left(\boldsymbol{\gamma}^H \mathbf{D}\boldsymbol{\gamma}\right)/\partial\boldsymbol{\gamma}^* = \boldsymbol{\gamma}'\mathbf{D}', \quad \partial\left(\boldsymbol{\gamma}^H \mathbf{X}'\mathbf{y}\right)/\partial\boldsymbol{\gamma}^* = (\mathbf{X}'\mathbf{y})', \\
\partial\left(\mathbf{y}^H \mathbf{X}\boldsymbol{\gamma}\right)/\partial\boldsymbol{\gamma}^* &= 0, \quad \partial\left(\boldsymbol{\gamma}'\mathbf{G}\boldsymbol{\gamma}\right)/\partial\boldsymbol{\gamma}^* = 0, \quad \partial\left(\boldsymbol{\gamma}^H \mathbf{G}^*\boldsymbol{\gamma}^*\right)/\partial\boldsymbol{\gamma}^* = \gamma^H\left(\mathbf{G}^* + \mathbf{G}^H\right).
\end{aligned}$$

Then

$$\begin{aligned}
\frac{\partial h}{\partial \boldsymbol{\gamma}^v} &= \frac{1}{2}\left(\boldsymbol{\gamma}^v\right)^H \mathbf{X}^H\mathbf{X} - \frac{1}{2}\left(\mathbf{y}^v\right)^H \mathbf{X} + \left(\boldsymbol{\gamma}^v\right)^H \mathbf{D}_v - \frac{1}{2}\left[\left(\boldsymbol{\gamma}^v\right)'\left(\mathbf{G}_v + \mathbf{G}_v'\right)\right] \\
&= \frac{1}{2}\left(\boldsymbol{\gamma}^v\right)^H \mathbf{X}^H\mathbf{X} - \frac{1}{2}\left(\mathbf{y}^v\right)^H \mathbf{X} + \left(\boldsymbol{\gamma}^v\right)^H \mathbf{D}_v - \left(\boldsymbol{\gamma}^v\right)' \mathbf{G}_v,
\end{aligned}$$

and the $\boldsymbol{\gamma}^v$ that minimizes h satisfies the equation

$$(\boldsymbol{\gamma}^v)^H \left(\mathbf{X}^H\mathbf{X} + 2\mathbf{D}_v\right) - 2\,(\boldsymbol{\gamma}^v)'\,\mathbf{G}_v = (\mathbf{y}^v)^H\,\mathbf{X}. \tag{A.1}$$

We also have

$$\begin{aligned}\frac{\partial h}{\partial\,(\boldsymbol{\gamma}^v)^*} &= \frac{1}{2}\,(\boldsymbol{\gamma}^v)'\left(\mathbf{X}^H\mathbf{X}\right)' - \frac{1}{2}\left(\mathbf{X}^H\mathbf{y}^v\right)' + (\boldsymbol{\gamma}^v)'\,\mathbf{D}_v' - \frac{1}{2}\left[(\boldsymbol{\gamma}^v)^H\left(\mathbf{G}_v^* + \mathbf{G}_v^H\right)\right]. \\ &= \frac{1}{2}\,(\boldsymbol{\gamma}^v)'\,\mathbf{X}^H\mathbf{X} - \frac{1}{2}\,(\mathbf{y}^v)'\,\mathbf{X} + (\boldsymbol{\gamma}^v)'\,\mathbf{D}_v - (\boldsymbol{\gamma}^v)^H\,\mathbf{G}_v^*,\end{aligned}$$

and so, the $\boldsymbol{\gamma}^v$ that minimizes h also satisfies the equation

$$(\boldsymbol{\gamma}^v)'\left(\mathbf{X}^H\mathbf{X} + 2\mathbf{D}_v\right) - 2\,(\boldsymbol{\gamma}^v)^H\,\mathbf{G}_v^* = (\mathbf{y}^v)'\,\mathbf{X}. \tag{A.2}$$

Note that (A.1) is the conjugate of (A.2) and both equations lead to the same solution of $\boldsymbol{\gamma}^v$. Transposing (A.2), we have

$$\left(\mathbf{X}^H\mathbf{X} + 2\mathbf{D}_v\right)\boldsymbol{\gamma}^v - 2\mathbf{G}_v^H\,(\boldsymbol{\gamma}^v)^* = \mathbf{X}'\mathbf{y}^v$$

Hence,

$$\left(\mathbf{X}^H\mathbf{X} + 2\mathbf{D}_v\right)(\boldsymbol{\gamma}_{Re}^v + \boldsymbol{\gamma}_{Im}^v i) - (2\mathbf{G}_{v,Re} - 2\mathbf{G}_{v,Im}i)\,(\boldsymbol{\gamma}_{Re}^v - \boldsymbol{\gamma}_{Im}^v i) = \mathbf{X}'\mathbf{y}^v,$$

and

$$\left[\left(\mathbf{X}^H\mathbf{X} + 2\mathbf{D}_v - 2\mathbf{G}_{v,Re}\right)\boldsymbol{\gamma}_{Re}^v + 2\mathbf{G}_{v,Im}\boldsymbol{\gamma}_{Im}^v\right] = \left(\mathbf{X}'\mathbf{y}^v\right)_{Re},$$

$$\left[\left(\mathbf{X}^H\mathbf{X} + 2\mathbf{D}_v + 2\mathbf{G}_{v,Re}\right)\boldsymbol{\gamma}_{Im}^v + 2\mathbf{G}_{v,Im}\boldsymbol{\gamma}_{Re}^v\right] = \left(\mathbf{X}'\mathbf{y}^v\right)_{Im}.$$

Therefore, taking $\mathbf{F}_v = \mathbf{X}^H\mathbf{X} + 2\mathbf{D}_v$, at iteration $(l+1)$ we have

$$\boldsymbol{\gamma}_{Re}^{v,(l+1)} = \left[\mathbf{F}_v^2 - \left(2\mathbf{G}_{v,Re}\right)^2 - \left(2\mathbf{G}_{v,Im}\right)^2\right]^{-1}\left[\left(\mathbf{F}_v + 2\mathbf{G}_{v,Re}\right)\left(\mathbf{X}'\mathbf{y}^v\right)_{Re} - 2\mathbf{G}_{v,Im}\left(\mathbf{X}'\mathbf{y}^v\right)_{Im}\right],$$
$$\boldsymbol{\gamma}_{Im}^{v,(l+1)} = \left[\left(2\mathbf{G}_{v,Im}\right)^2 - \mathbf{F}_v^2 + \left(2\mathbf{G}_{v,Re}\right)^2\right]^{-1}\left[2\mathbf{G}_{v,Im}\left(\mathbf{X}'\mathbf{y}^v\right)_{Re} - \left(\mathbf{F}_v - 2\mathbf{G}_{v,Re}\right)\left(\mathbf{X}'\mathbf{y}^v\right)_{Im}\right].$$

The update for each σ_v^2 is derived from $\partial Q_1^v/\partial\sigma_v^2 = 0$, which leads to

$$\sigma_v^{2,(l+1)} = \frac{SS_\sigma}{T + p + a_\sigma + 1},$$

where

$$SS_\sigma = \left[\left(\mathbf{y}^v - \mathbf{X}(\boldsymbol{\gamma}^v)^{(l+1)}\right)^H\left(\mathbf{y}^v - \mathbf{X}(\boldsymbol{\gamma}^v)^{(l+1)}\right)\right]/2 + (\boldsymbol{\gamma}^v)^{(l+1)H}\mathbf{D}_v(\boldsymbol{\gamma}^v)^{(l+1)}$$
$$- \mathrm{Re}\left((\boldsymbol{\gamma}^v)^{(l+1)'}\mathbf{G}_v(\boldsymbol{\gamma}^v)^{(l+1)}\right) + b_\sigma.$$

Finally, simply taking derivatives of Q_2 with respect to θ_j, for $j = 1, \ldots, p$, we update each θ_j as follows

$$\theta_j^{(l+1)} = \frac{\sum_{v=1}^V p_j^v + a_\theta - 1}{a_\theta + b_\theta + V - 2}.$$

This completes the derivation of the C-EMVS algorithm.

A.2 EM derivation for model (4.4)

For model (4.4) we have parameter values values $\lambda_0 = \lambda_1 = 0$, $\omega_0 = 2v_0$ and $\omega_1 = 2v_1$. Here we also assume that, $a_\sigma = b_\sigma = 1/2$, which determine the prior distribution for σ_v^2 that was used in the analyses of simulated and human data. Therefore, the prior for $\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v$, $v = 1, \ldots, V$ has the form $\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v \sim CN_p\left(\mathbf{0}, 2\sigma_v^2\boldsymbol{\Omega}_v, \boldsymbol{\Lambda}_v\right)$, where $\boldsymbol{\Lambda}_v = \mathbf{0}$ and $\boldsymbol{\Omega}_v = \text{diag}\left(\left(1-\psi_1^v\right)v_0 + \psi_1^v v_1, \ldots, \left(1-\psi_p^v\right)v_0 + \psi_p^v v_1\right)$.

Since $\boldsymbol{\Lambda}_v = \mathbf{0}$ and $\boldsymbol{\Omega}_v$ is real-valued, one has $\boldsymbol{P}_v = 2\sigma_v^2\boldsymbol{\Omega}_v$ and $\boldsymbol{R}_v = \mathbf{0}$. The E-step of the algorithm is a bit simpler as all the terms that involve $\lambda_0 = \lambda_1 = 0$ cancel. With real-valued $\mathbf{X}$ and $\mathbf{G}_{v,Re} = \mathbf{G}_{v,Im} = \mathbf{0}$, the M-step leads to

$$\left(\boldsymbol{\gamma}_{Re}^v\right)^{(l+1)} = \left(\mathbf{X}'\mathbf{X} + 2\mathbf{D}_v\right)^{-1}\left(\mathbf{X}'\mathbf{y}^v\right)_{Re},$$
$$\left(\boldsymbol{\gamma}_{Im}^v\right)^{(l+1)} = \left(\mathbf{X}'\mathbf{X} + 2\mathbf{D}_v\right)^{-1}\left(\mathbf{X}'\mathbf{y}^v\right)_{Im},$$

with $\mathbf{D}_v$ as previously defined. In addition,

$$\sigma_v^{2,(l+1)} = \frac{\|\mathbf{y}_r^v - \mathbf{X}\left(\boldsymbol{\gamma}_r^v\right)^{(l+1)}\|^2 + 2\|\left(\mathbf{D}_v\right)^{1/2}\left(\boldsymbol{\gamma}_r^v\right)^{(l+1)}\|^2 + 1}{2T + 2p + 3},$$

and

$$\theta_j^{(l+1)} = \frac{\sum_{v=1}^V p_j^v + a_\theta - 1}{V + a_\theta + b_\theta - 2},$$

for $j = 1, \ldots, p$.

A.3 Derivation of the C-EMVS algorithm for models with autoregressive noises

The analyses of the human CV-fMRI data presented in Section 4.4 with autoregressive errors consider four different models based on voxel-specific and/or common variance and/or AR coefficients. These models are labeled as: model (ii): (σ^2, φ);, model (iii): (σ^2, φ_v); model (v): (σ_v^2, φ); and model (vi): (σ_v^2, φ_v). The following derives the EM algorithm for the most general case of voxel-specific variances and AR coefficients, i.e., model (vi) with (σ_v^2, φ_v). Derivations for the other three cases are similar.

The model setting for the general case is as follows. For each voxel $v = 1, \ldots, V$, and each $j = 1, \ldots, p$, we have

$$\mathbf{y}^v = \mathbf{X}\boldsymbol{\gamma}^v + \boldsymbol{\eta}^v, \quad \eta_t^v = \varphi_v \eta_{t-1}^v + \zeta_t^v, \quad \zeta_t^v \overset{\text{iid}}{\sim} CN_1\left(0, 2\sigma_v^2, 0\right),$$

$$\boldsymbol{\gamma}^v | \boldsymbol{\psi}^v \sim CN_p\left(\mathbf{0}, \sigma_v^2 \boldsymbol{\Omega}_v, \sigma_v^2 \boldsymbol{\Lambda}_v\right), \quad \sigma_v^2 | a_\sigma, b_\sigma \overset{\text{iid}}{\sim} IG\left(a_\sigma, b_\sigma\right),$$

$$\psi_j^v | \theta_j \overset{\text{iid}}{\sim} Bernoulli\left(\theta_j\right), \quad \theta_j | a_\theta, b_\theta \overset{\text{iid}}{\sim} Beta\left(a_\theta, b_\theta\right), \quad \varphi_v \overset{\text{iid}}{\sim} Uniform\left(-1, 1\right).$$

First we apply the Prais-Winsten transformation on $\mathbf{y}^v$ and $\mathbf{X}$ so that

$$\tilde{\mathbf{y}}^v = \begin{bmatrix} \tilde{y}_1^v \\ \vdots \\ \tilde{y}_T^v \end{bmatrix} = \begin{bmatrix} \sqrt{1-\varphi_v^2} y_1^v \\ y_2^v - \varphi_v y_1^v \\ \vdots \\ y_T^v - \varphi_v y_{T-1}^v \end{bmatrix}, \quad \tilde{\mathbf{X}}^v = \begin{bmatrix} \tilde{x}_1^v \\ \vdots \\ \tilde{x}_T^v \end{bmatrix} = \begin{bmatrix} \sqrt{1-\varphi_v^2} x_1 \\ x_2 - \varphi_v x_1 \\ \vdots \\ x_T - \varphi_v x_{T-1} \end{bmatrix},$$

where x_t is the t-th row vector of $\mathbf{X}$. Note that now both $\tilde{\mathbf{y}}^v$ and $\tilde{\mathbf{X}}^v$ are functions of φ_v. We then consider the likelihood conditional on the first observation, y_1^v, instead of the full likelihood and obtain

$$f\left(\tilde{\mathbf{y}}_{-1}^v \mid \boldsymbol{\gamma}^v, \sigma_v^2, \varphi_v, y_1^v\right) \propto |2\sigma_v^2\mathbf{I}|^{-1} \exp\{-\left(\tilde{\mathbf{y}}_{-1}^v - \tilde{\mathbf{X}}_{-1}^v\boldsymbol{\gamma}^v\right)^H \left(2\sigma_v^2\mathbf{I}\right)^{-1} \left(\tilde{\mathbf{y}}_{-1}^v - \tilde{\mathbf{X}}_{-1}^v\boldsymbol{\gamma}^v\right)\},$$

where $\tilde{\mathbf{y}}_{-1}^v$ is $\tilde{\mathbf{y}}^v$ without $\tilde{y}_1^v$ and $\tilde{\mathbf{X}}_{-1}^v$ is $\tilde{\mathbf{X}}^v$ without the first row $\tilde{x}_1^v$. This leads to the following log-posterior density

$$\begin{aligned}
&\log \pi\left(\boldsymbol{\gamma}, \theta, \boldsymbol{\sigma}^2, \boldsymbol{\psi}, \boldsymbol{\varphi} \mid \mathbf{y}\right) = \\
&\sum_{v=1}^{V} \Bigg\{ -\frac{1}{2\sigma_v^2}\left(\tilde{\mathbf{y}}_{-1}^v - \tilde{\mathbf{X}}_{-1}^v\boldsymbol{\gamma}^v\right)^H \left(\tilde{\mathbf{y}}_{-1}^v - \tilde{\mathbf{X}}_{-1}^v\boldsymbol{\gamma}^v\right) - (T-1+p+a_\sigma+1)\log\left(\sigma_v^2\right) - \frac{b_\sigma}{\sigma_v^2} \\
&- (\boldsymbol{\gamma}^v)^H \boldsymbol{P}_v^{-*}\boldsymbol{\gamma}^v + \frac{1}{2}\left[(\boldsymbol{\gamma}^v)'\boldsymbol{R}_v'\boldsymbol{P}_v^{-*}\boldsymbol{\gamma}^v + (\boldsymbol{\gamma}^v)^H\mathbf{R}_v^H\mathbf{P}_v^{-1}(\boldsymbol{\gamma}^v)^*\right] + \sum_{j=1}^{p}\psi_j^v\log\left(\theta_j\right) + \\
&\sum_{j=1}^{p}\left(1-\psi_j^v\right)\log\left(1-\theta_j\right)\Bigg\} + (a_\theta-1)\sum_{j=1}^{p}\log\left(\theta_j\right) + (b_\theta-1)\sum_{j=1}^{p}\log\left(1-\theta_j\right) + \tilde{K}_\pi,
\end{aligned}$$

where $\boldsymbol{P}_v^{-*}$ and $\boldsymbol{R}_v$ are as defined as above and $\tilde{K}_\pi$ is a constant. The E-steps are essentially the same as the ones described above, but we now work with the transformed values $\tilde{\mathbf{y}}_{-1}^v$ and $\tilde{\mathbf{X}}_{-1}^v$. Similarly, taking $\tilde{\mathbf{F}}_v = \left(\tilde{\mathbf{X}}_{-1}^v\right)^H \tilde{\mathbf{X}}_{-1}^v + 2\mathbf{D}_v$, the M-step leads to the following updating rules for $\boldsymbol{\gamma}^v$ and σ_v^2, for each v, we obtain $(\boldsymbol{\gamma}^v)^{(l+1)} = \boldsymbol{\gamma}_{Re}^{v,(l+1)} +$

$i\boldsymbol{\gamma}_{Im}^{v,(l+1)}$, with

$$\begin{aligned}
\boldsymbol{\gamma}_{Re}^{v,(l+1)} &= [\tilde{\mathbf{F}}_v^2 - (2\mathbf{G}_{v,Re})^2 - (2\mathbf{G}_{v,Im})^2]^{-1} \\
&\quad \times \left[\left(\tilde{\mathbf{F}}_v + 2\mathbf{G}_{v,Re} \right) \left(\left(\tilde{\mathbf{X}}_{-1}^v \right)' \tilde{\mathbf{y}}_{-1}^v \right)_{Re} - 2\mathbf{G}_{v,Im} \left(\tilde{\mathbf{X}}_{-1}' \tilde{\mathbf{y}}_{-1}^v \right)_{Im} \right], \\
\boldsymbol{\gamma}_{Im}^{v,(l+1)} &= \left[(2\mathbf{G}_{v,Im})^2 - \tilde{\mathbf{F}}_v^2 + (2\mathbf{G}_{v,Re})^2 \right]^{-1} \\
&\quad \times \left[2\mathbf{G}_{v,Im} \left(\left(\tilde{\mathbf{X}}_{-1}^v \right)' \tilde{\mathbf{y}}_{-1}^v \right)_{Re} - \left(\tilde{\mathbf{F}}_v^2 - 2\mathbf{G}_{v,Re} \right) \left(\left(\tilde{\mathbf{X}}_{-1}^v \right)' \tilde{\mathbf{y}}_{-1}^v \right)_{Im} \right],
\end{aligned}$$

with $\mathbf{D}_v$ and $\mathbf{G}_v$ as defined above,

$$\sigma_v^{2,(l+1)} = \frac{(SS_\sigma^v)^{(l+1)}}{T + p + a_\sigma},$$

with

$$\begin{aligned}
SS_\sigma^{v,(l+1)} = & \frac{\left(\tilde{\mathbf{y}}_{-1}^v - \tilde{\mathbf{X}}_{-1}^v (\boldsymbol{\gamma}^v)^{(l+1)} \right)^H \left(\tilde{\mathbf{y}}_{-1}^v - \tilde{\mathbf{X}}_{-1}^v (\boldsymbol{\gamma}^v)^{(l+1)} \right)}{2} + \boldsymbol{\gamma}^{(l+1)H} \mathbf{D}_v \boldsymbol{\gamma}^{(l+1)} \\
& - \mathrm{Re} \left(\boldsymbol{\gamma}^{(l+1)'} \mathbf{G}_v \boldsymbol{\gamma}^{(l+1)} \right) + b_\sigma,
\end{aligned}$$

and the following updating rule for $\theta_j^{(l+1)}$,

$$\theta_j^{(l+1)} = \frac{\sum_{v=1}^V p_j^v + a_\theta - 1}{a_\theta + b_\theta + V - 2}.$$

Finally, to obtain the updating rule for the AR coefficients we proceed as follows. First note that, for $t = 2, \ldots, T$,

$$\tilde{y}_t^v - \tilde{x}_t^v \left(\boldsymbol{\gamma}_r^v \right)^{(l+1)} = \left(y_t^v - \varphi_v y_{t-1}^v \right) - \left(x_t - \varphi_v x_{t-1} \right) \left(\boldsymbol{\gamma}_r^v \right)^{(l+1)} = \left(\eta_t^v \right)^{(l+1)} - \varphi_v \left(\eta_{t-1}^v \right)^{(l+1)}.$$

We derive the M-step for each φ_v by setting

$$\frac{\partial \left[\sum_{t=2}^{T} \left((\eta_t^v)^* - \varphi_v \left(\eta_{t-1}^v\right)^*\right) \left(\eta_t^v - \varphi_v \eta_{t-1}^v\right)\right]}{\partial \varphi_v} = 0.$$

This leads to

$$\varphi_v^{(l+1)} = \frac{\sum_{t=2}^{T} \mathrm{Re}\left((\eta_t^v)^{*(l+1)} \left(\eta_{t-1}^v\right)^{(l+1)}\right)}{\sum_{t=2}^{T} |\left(\eta_{t-1}^v\right)^{(l+1)}|^2}.$$

If we consider common σ^2 instead of voxel-specific σ_v^2, its updating rule is

$$\sigma^{2,(l+1)} = \frac{\sum_{v=1}^{V} SS_\sigma^{v,(l+1)}}{V[(T-1)+p] + a_\sigma + 1}.$$

Similarly, the updating rule for a model with common φ across voxels is

$$\varphi^{(l+1)} = \frac{\sum_{v=1}^{V} \sum_{t=2}^{T} \mathrm{Re}\left((\eta_t^v)^{*(l+1)} \left(\eta_{t-1}^v\right)^{(l+1)}\right)}{\sum_{v=1}^{V} \sum_{t=2}^{T} |\left(\eta_{t-1}^v\right)^{(l+1)}|^2}.$$

For general AR(q) noises, we can transform variables from $\mathbf{y}^v$ and $\mathbf{X}$ to

$$\tilde{\mathbf{y}}^v = \begin{bmatrix} \tilde{y}_{q+1}^v \\ \vdots \\ \tilde{y}_T^v \end{bmatrix} = \begin{bmatrix} \varphi_v(L) y_{q+1} \\ \vdots \\ \varphi_v(L) y_T \end{bmatrix}, \quad \tilde{\mathbf{X}}^v = \begin{bmatrix} \tilde{x}_{q+1}^v \\ \vdots \\ \tilde{x}_T^v \end{bmatrix} = \begin{bmatrix} \varphi_v(L) x_{q+1} \\ \vdots \\ \varphi_v(L) x_T \end{bmatrix},$$

where L is the lag operator such that $\varphi_v(L) = 1 - \varphi_{1,v} L - \cdots - \varphi_{q,v} L^q$. Accordingly, the E-step and M-step for the parameters can be derived as the AR(1) case above. Specifically, the updating rules for $\boldsymbol{\gamma}^v$, σ_v^2 and $\boldsymbol{\theta}$ are similar to those with the AR(1)

structure, and the updating rules for $\varphi_{1,v}, \ldots, \varphi_{q,v}$ are given by

$$\varphi_{k,v}^{(l+1)} = \frac{A_k^v - B_k^v}{C_k^v}, \quad k = 1, \ldots, q,$$

where

$$\begin{aligned} A_k^v &= \sum_{t=q+1}^{T} \left[(\eta_t^v)^{*(l+1)} \left(\eta_{t-k}^v\right)^{(l+1)} + \left(\eta_{t-k}^v\right)^{*(l+1)} (\eta_t^v)^{(l+1)} \right] \\ &= \sum_{t=q+1}^{T} 2\mathrm{Re}\left((\eta_t^v)^{*(l+1)} \left(\eta_{t-k}^v\right)^{(l+1)} \right) \\ B_k &= \sum_{t=q+1}^{T} \sum_{n=1, n\neq k}^{q} \left[\varphi_{n,v}^{(l)} \left(\eta_{t-n}^v\right)^{*(l+1)} \left(\eta_{t-k}^v\right)^{(l+1)} + \varphi_{n,v}^{(l)} \left(\eta_{t-k}^v\right)^{*(l+1)} \left(\eta_{t-n}^v\right)^{(l+1)} \right] \quad \text{and} \\ &= \sum_{n=1, n\neq k}^{q} \varphi_{n,v}^{(l)} \sum_{t=q+1}^{T} 2\mathrm{Re}\left(\left(\eta_{t-n}^v\right)^{*(l+1)} \left(\eta_{t-k}^v\right)^{(l+1)} \right), \end{aligned}$$

and

$$C_k = \sum_{t=q+1}^{T} 2|(\eta_{r,t-k}^v)^{(l+1)}|^2.$$

A.4 Derivation of the C-SSVS algorithm

Once again consider the model specified by

$$\mathbf{y}^v = \mathbf{X}\boldsymbol{\gamma}^v + \boldsymbol{\eta}^v, \quad \boldsymbol{\eta}^v \sim CN_T\left(\mathbf{0}, 2\sigma_v^2\mathbf{I}, \mathbf{0}\right),$$

$$\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v \sim CN_p\left(\mathbf{0}, \boldsymbol{\Omega}_v, \boldsymbol{\Lambda}_v\right), \quad \sigma_v^2 \overset{\text{iid}}{\sim} IG\left(a_\sigma, b_\sigma\right),$$

$$\psi_j^v \mid \theta_j \overset{\text{iid}}{\sim} Bernoulli\left(\theta_j\right) \quad \text{and} \quad \theta_j \mid a_\theta, b_\theta \overset{\text{iid}}{\sim} Beta\left(a_\theta, b_\theta\right).$$

As in the derivation of the C-EMVS algorithm, we use the complex-valued normal density (Wooding, 1956; van den Bos, 1995; Picinbono, 1996) to find the full conditionals for $\boldsymbol{\gamma}$, $\boldsymbol{\sigma}^2$, $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ and obtain the steps of a Gibbs sampling algorithm for full posterior inference. The posterior distribution for $\boldsymbol{\gamma}$, $\boldsymbol{\sigma}^2$, $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ is

$$\begin{aligned}
\pi\left(\boldsymbol{\gamma}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}, \boldsymbol{\theta} \mid \mathbf{y}\right) &\propto \prod_{v=1}^{V}\left[f\left(\mathbf{y}^v \mid \boldsymbol{\gamma}^v, \sigma_v^2\right)\pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v, \sigma_v^2\right)\pi\left(\sigma_v^2 \mid \boldsymbol{\psi}^v\right)\pi\left(\boldsymbol{\psi}^v \mid \boldsymbol{\theta}\right)\right]\pi\left(\boldsymbol{\theta}\right) \\
&\propto \prod_{v=1}^{V}\Bigg[\left|2\sigma_v^2\mathbf{I}\right|^{-1}\exp\left\{-\left(\mathbf{y}^v-\mathbf{X}\boldsymbol{\gamma}^v\right)^H\left(2\sigma_v^2\mathbf{I}\right)^{-1}\left(\mathbf{y}^v-\mathbf{X}\boldsymbol{\gamma}^v\right)\right\}\left(|\boldsymbol{\Omega}_v||\boldsymbol{P}_v|\right)^{-1/2} \\
&\quad\times\exp\left\{-\left(\boldsymbol{\gamma}^v\right)^H\boldsymbol{P}_v^{-*}\boldsymbol{\gamma}^v+\frac{1}{2}\left[\left(\boldsymbol{\gamma}^v\right)'\boldsymbol{R}_v'\boldsymbol{P}_v^{-*}\boldsymbol{\gamma}^v+\left(\boldsymbol{\gamma}^v\right)^H\boldsymbol{R}_v^H\boldsymbol{P}_v^{-1}\left(\boldsymbol{\gamma}^v\right)^*\right]\right\} \\
&\quad\times\left(\sigma_v^2\right)^{-(a_\sigma+1)}\exp\left\{-b_\sigma/\sigma_v^2\right\}\prod_{j=1}^{p}\theta_j^{\psi_j^v}\left(1-\theta_j\right)^{1-\psi_j^v}\Bigg]\prod_{j=1}^{p}\theta_j^{a_\theta-1}\left(1-\theta_j\right)^{b_\theta-1},
\end{aligned}$$

where $\boldsymbol{P}_v = \boldsymbol{\Omega}_v^* - \boldsymbol{\Lambda}_v^H\boldsymbol{\Omega}_v^{-1}\boldsymbol{\Lambda}_v$ Hermitian and positive definite, and $\boldsymbol{R}_v = \boldsymbol{\Lambda}_v^H\boldsymbol{\Omega}_v^{-1}$.

Now assume that for each v, $\boldsymbol{\gamma}^v$ and $\boldsymbol{\eta}^v$ are uncorrelated and "unrelated", i.e., the covariance matrix is zero, $\mathrm{Cov}\left(\boldsymbol{\eta}^v, \boldsymbol{\gamma}^v\right) = \mathbf{0}$, and the relation matrix is also zero, $\mathrm{Rel}\left(\boldsymbol{\eta}^v, \boldsymbol{\gamma}^v\right) = \mathbf{0}$, Then $\left(\mathbf{y}^v, \boldsymbol{\gamma}^v\right)'$ are jointly complex normal, i.e., $\left(\mathbf{y}^v, \boldsymbol{\gamma}^v\right)' \sim CN\left((\mathbf{0}, \mathbf{0})', \bar{\boldsymbol{\Omega}}_v, \bar{\boldsymbol{\Lambda}}_v\right)$, where

$$\bar{\boldsymbol{\Omega}}_v = \begin{pmatrix} \boldsymbol{\Omega}_{\mathbf{y}}^v & \boldsymbol{\Omega}_{\mathbf{y},\boldsymbol{\gamma}}^v \\ \left(\boldsymbol{\Omega}_{\mathbf{y},\boldsymbol{\gamma}}^v\right)^H & \boldsymbol{\Omega}_{\boldsymbol{\gamma}}^v \end{pmatrix} = \begin{pmatrix} \mathbf{X}\boldsymbol{\Omega}_v\mathbf{X}' + 2\sigma_v^2\mathbf{I} & \mathbf{X}\boldsymbol{\Omega}_v \\ \left(\mathbf{X}\boldsymbol{\Omega}_v\right)^H & \boldsymbol{\Omega}_v \end{pmatrix},$$

$$\bar{\boldsymbol{\Lambda}}_v = \begin{pmatrix} \boldsymbol{\Lambda}_{\mathbf{y}}^v & \boldsymbol{\Lambda}_{\mathbf{y},\boldsymbol{\gamma}}^v \\ \left(\boldsymbol{\Lambda}_{\mathbf{y},\boldsymbol{\gamma}}^v\right)^H & \boldsymbol{\Lambda}_{\boldsymbol{\gamma}}^v \end{pmatrix} = \begin{pmatrix} \mathbf{X}\boldsymbol{\Lambda}_v\mathbf{X}' & \mathbf{X}\boldsymbol{\Lambda}_v \\ \left(\mathbf{X}\boldsymbol{\Lambda}_v\right)' & \boldsymbol{\Lambda}_v \end{pmatrix},$$

and

$$\begin{aligned}
\boldsymbol{\Omega}_{\mathbf{y}}^{v} &= \mathrm{Cov}\left(\mathbf{y}^{v}\right) = \mathrm{Cov}\left(\mathbf{X}\boldsymbol{\gamma}^{v} + \boldsymbol{\eta}^{v}\right) = \mathbf{X}\boldsymbol{\Omega}_{v}\mathbf{X}' + 2\sigma_{v}^{2}\mathbf{I}, \\
\boldsymbol{\Lambda}_{\mathbf{y}}^{v} &= \mathrm{Rel}\left(\mathbf{y}^{v}\right) = \mathrm{Rel}\left(\mathbf{X}\boldsymbol{\gamma}^{v} + \boldsymbol{\eta}^{v}\right) = \mathbf{X}\boldsymbol{\Lambda}_{v}\mathbf{X}', \\
\boldsymbol{\Omega}_{\mathbf{y},\boldsymbol{\gamma}}^{v} &= \mathrm{Cov}\left(\mathbf{X}\boldsymbol{\gamma}^{v} + \boldsymbol{\eta}^{v}, \boldsymbol{\gamma}^{v}\right) = \mathbf{X}\boldsymbol{\Omega}_{v}, \\
\boldsymbol{\Lambda}_{\mathbf{y},\boldsymbol{\gamma}}^{v} &= \mathrm{Rel}\left(\mathbf{X}\boldsymbol{\gamma}^{v} + \boldsymbol{\eta}^{v}, \boldsymbol{\gamma}^{v}\right) = \mathbf{X}\boldsymbol{\Lambda}_{v}.
\end{aligned}$$

The full conditional posterior of $\boldsymbol{\gamma}^{v}$ is then $\boldsymbol{\gamma}^{v} \mid \boldsymbol{\gamma}^{-v}, \boldsymbol{\sigma}^{2}, \boldsymbol{\psi}, \boldsymbol{\theta}, \mathbf{y} \sim$ $CN_{p}\left(\boldsymbol{\mu}_{\boldsymbol{\gamma}^{v}}\left(\mathbf{y}^{v}\right), \boldsymbol{\Omega}_{pos}^{v}, \boldsymbol{\Lambda}_{pos}^{v}\right)$, where

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{\gamma}^{v}}\left(\mathbf{y}^{v}\right) &= \mathbf{A}_{v}\mathbf{y}^{v} + \mathbf{B}_{v}\left(\mathbf{y}^{v}\right)^{*}, \\
\boldsymbol{\Omega}_{pos}^{v} &= \boldsymbol{\Omega}_{v} - \mathbf{A}_{v}\left(\mathbf{X}\boldsymbol{\Omega}_{v}\right) - \mathbf{B}_{v}\left(\left(\mathbf{X}\boldsymbol{\Lambda}_{v}\right)'\right)^{H}, \\
\boldsymbol{\Lambda}_{pos}^{v} &= \boldsymbol{\Lambda}_{v} - \mathbf{A}_{v}\left(\mathbf{X}\boldsymbol{\Lambda}_{v}\right) - \mathbf{B}_{v}\left(\left(\mathbf{X}\boldsymbol{\Omega}_{v}\right)^{H}\right)', \\
\mathbf{A}_{v} &= \left(\left(\left(\mathbf{X}\boldsymbol{\Omega}_{v}\right)^{H}\right) - \left(\mathbf{X}\boldsymbol{\Lambda}_{v}\right)'\left(\mathbf{X}\boldsymbol{\Omega}_{v}\mathbf{X}' + 2\sigma_{v}^{2}\mathbf{I}\right)^{-*}\left(\mathbf{X}\boldsymbol{\Lambda}_{v}\mathbf{X}'\right)^{H}\right)\mathbf{W}_{v}^{-*}, \\
\mathbf{B}_{v} &= \left(\left(\left(\mathbf{X}\boldsymbol{\Lambda}_{v}\right)'\right) - \left(\mathbf{X}\boldsymbol{\Omega}_{v}\right)^{H}\left(\mathbf{X}\boldsymbol{\Omega}_{v}\mathbf{X}' + 2\sigma_{v}^{2}\mathbf{I}\right)^{-1}\left(\mathbf{X}\boldsymbol{\Lambda}_{v}\mathbf{X}'\right)\right)\mathbf{W}_{v}^{-1}, \\
\mathbf{W}_{v} &= \left(\mathbf{X}\boldsymbol{\Omega}_{v}\mathbf{X}^{H} + 2\sigma_{v}^{2}\mathbf{I}\right)^{*} - \left(\mathbf{X}\boldsymbol{\Lambda}_{v}\mathbf{X}'\right)^{H}\left(\mathbf{X}\boldsymbol{\Omega}_{v}\mathbf{X}^{H} + 2\sigma_{v}^{2}\mathbf{I}\right)^{-1}\left(\mathbf{X}\boldsymbol{\Lambda}_{v}\mathbf{X}'\right).
\end{aligned}$$

The full conditional posterior for σ_{v}^{2} is $\sigma_{v}^{2} \mid \boldsymbol{\gamma}, \boldsymbol{\sigma}_{-v}^{2}, \boldsymbol{\psi}, \boldsymbol{\theta}, \mathbf{y} \sim IG\left(a_{\sigma}^{v,pos}, b_{\sigma}^{v,pos}\right)$, where $a_{\sigma}^{v,pos} = T + a_{\sigma}$, and $b_{\sigma}^{v,pos} = \|\mathbf{y}^{v} - \mathbf{X}\boldsymbol{\gamma}^{v}\|^{2}/2 + b_{\sigma}$. For the full conditional posterior for $\boldsymbol{\psi}^{v}$, notice that $Pr\left(\psi_{j}^{v} \mid \mathbf{y}, \boldsymbol{\gamma}, \boldsymbol{\sigma}^{2}, \boldsymbol{\psi}_{-j}^{v}, \boldsymbol{\theta}\right) = Pr\left(\psi_{j}^{v} \mid \boldsymbol{\gamma}^{v}, \sigma_{v}^{2}, \boldsymbol{\psi}_{-j}^{v}, \boldsymbol{\theta}\right)$ since $\boldsymbol{\psi}^{v}$ affect

$\mathbf{y}^v$ only through $\boldsymbol{\gamma}^v$. So $Pr\left(\psi_j^v = 1 \mid \boldsymbol{\gamma}^v, \sigma_v^2, \boldsymbol{\psi}_{-j}^v, \boldsymbol{\theta}\right) = \dfrac{c_j^v}{c_j^v + e_j^v}$, where

$$c_j^v = \pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}_{-j}^v, \psi_j^v = 1\right) \pi\left(\sigma_v^2 \mid \boldsymbol{\psi}_{-j}^v, \psi_j^v = 1\right) Pr\left(\psi_j^v = 1 \mid \boldsymbol{\psi}_{-j}^v, \boldsymbol{\theta}\right),$$

$$e_j^v = \pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}_{-j}^v, \psi_j^v = 0\right) \pi\left(\sigma_v^2 \mid \boldsymbol{\psi}_{-j}^v, \psi_j^v = 1\right) Pr\left(\psi_j^v = 0 \mid \boldsymbol{\psi}_{-j}^v, \boldsymbol{\theta}\right),$$

and $\pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}_{-j}^v, \psi_j^v = 1\right)$ and $\pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}_{-j}^v, \psi_j^v = 0\right)$ are complex normal densities $CN_p\left(\mathbf{0}, \boldsymbol{\Omega}_v\left(\boldsymbol{\psi}_{-j}^v, \psi_j^v = 1\right), \boldsymbol{\Lambda}_v\left(\boldsymbol{\psi}_{-j}^v, \psi_j^v = 1\right)\right)$ and $CN_p\left(\mathbf{0}, \boldsymbol{\Omega}_v\left(\boldsymbol{\psi}_{-j}^v, \psi_j^v = 0\right), \boldsymbol{\Lambda}_v\left(\boldsymbol{\psi}_{-j}^v, \psi_j^v = 0\right)\right)$, respectively. Since ψ_j^v are independent and identically distributed $Pr\left(\psi_j^v = 1 \mid \boldsymbol{\psi}_{-j}^v, \boldsymbol{\theta}\right) = \theta_j$, and $Pr\left(\psi_j^v = 0 \mid \boldsymbol{\psi}_{-j}^v, \boldsymbol{\theta}\right) = 1 - \theta_j$. Furthermore, for a_σ and b_σ constant, $c_j^v = \pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}_{-j}^v, \psi_j^v = 1\right)\theta_j$ and $e_j^v = \pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}_{-j}^v, \psi_j^v = 0\right)(1 - \theta_j)$.

Finally, the full conditional for each θ_j is Beta distributed $Beta\left(\sum_{v=1}^V \psi_j^v + a_\theta, V - \sum_{v=1}^V \psi_j + b_\theta\right)$. This completes the derivation of the complex-valued SSVS algorithm.

A.5 C-SSVS algorithm for model (4.4)

Here we derive the full conditionals for the model in Section 4.2.2 with $\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v \sim CN_p\left(\mathbf{0}, \boldsymbol{\Omega}_v, \mathbf{0}\right)$, where $\boldsymbol{\Omega}_v = 2\times\text{diag}\left(\left(1 - \psi_1^v\right)v_0 + \psi_1^v v_1, \ldots, \left(1 - \psi_p^v\right)v_0 + \psi_p^v v_1\right)$. Using the equations above, we have

$$\begin{aligned} \boldsymbol{\Omega}_{pos}^v &= \boldsymbol{\Omega}_v - \mathbf{A}_v\left(\mathbf{X}\boldsymbol{\Omega}_v\right), \quad \boldsymbol{\Lambda}_{pos}^v = \mathbf{0}, \\ \mathbf{A}_v &= \left(\mathbf{X}\boldsymbol{\Omega}_v\right)^H \mathbf{W}_v^{-*}, \quad \mathbf{B}_v = \mathbf{0}, \quad \mathbf{W}_v = \left(\mathbf{X}\boldsymbol{\Omega}_v\mathbf{X}^H + 2\sigma_v^2\mathbf{I}\right)^*. \end{aligned}$$

Since both $\mathbf{X}$ and $\mathbf{\Omega}_v$ are real-valued and $\mathbf{\Omega}_v$ is symmetric, the full conditional for $\boldsymbol{\gamma}^v$ is $CN_p\left(\boldsymbol{\mu}_{\boldsymbol{\gamma}^v}, \mathbf{\Omega}^v_{pos}, \mathbf{0}\right)$, where

$$\mathbf{\Omega}^v_{pos} = \mathbf{\Omega}_v - (\mathbf{X}\mathbf{\Omega}_v)^H \left(2\mathbf{X}\mathbf{\Omega}_v\mathbf{X}^H + 2\sigma_v^2\mathbf{I}\right)^{-1} (\mathbf{X}\mathbf{\Omega}_v) = \left(2^{-1}\sigma_v^{-2}\mathbf{X}'\mathbf{X} + \mathbf{\Omega}_v^{-1}\right)^{-1},$$

and

$$\boldsymbol{\mu}_{\boldsymbol{\gamma}^v}(\mathbf{y}^v) = \mathbf{A}_v\mathbf{y}^v = (\mathbf{X}\mathbf{\Omega}_v)^H \left(\mathbf{X}\mathbf{\Omega}_v\mathbf{X}^H + 2\sigma_v^2\mathbf{I}\right)^{-1}\mathbf{y}^v = \mathbf{\Omega}^v_{pos}\mathbf{X}'\mathbf{y}^v/\sigma_v^2.$$

The full conditional for σ_v^2 is $IG\left(a_\sigma^{v,pos}, b_\sigma^{v,pos}\right)$, where $a_\sigma^{v,pos} = T + a_\sigma$, $b_\sigma^{v,pos} = \|\mathbf{y}^v - \mathbf{X}\boldsymbol{\gamma}^v\|^2/2 + b_\sigma$. In addition, the full conditional for ψ_j^v is $Pr\left(\psi_j^v = 1 \,\middle|\, \boldsymbol{\gamma}^v, \sigma_v^2, \boldsymbol{\psi}^v_{-j}\right) = \dfrac{c_j^v}{c_j^v + e_j^v}$, where $c_j^v = \pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v_{-j}, \psi_j^v = 1\right)\theta_j$, and $e_j^v = \pi\left(\boldsymbol{\gamma}^v \mid \boldsymbol{\psi}^v_{-j}, \psi_j^v = 0\right)(1-\theta_j)$. Note that $\mathbf{\Omega}_v$ is diagonal, and so $\pi\left(\gamma_j^v \mid \boldsymbol{\psi}^v_{-j}, \psi_j^v = 1\right) = \pi\left(\gamma_j^v \mid \psi_j^v = 1\right) = CN(0, 2v_1, 0)$ and $\pi\left(\gamma_j^v \mid \boldsymbol{\psi}^v_{-j}, \psi_j^v = 0\right) = \pi\left(\gamma_j^v \mid \psi_j^v = 0\right) = CN(0, 2v_0, 0)$. Finally, the full conditional for θ_j is the same as in the general complex-valued case.

A.6 Derivation of MCMC algorithm of the multi-subject spatial model

Let $\tilde{\boldsymbol{\gamma}}$, $\tilde{\mathbf{X}}$, and $\tilde{\mathbf{y}}$ be the real representation of $\boldsymbol{\gamma}$, $\mathbf{X}$ and $\mathbf{y}$, respectively. We first integrate out $\tilde{\boldsymbol{\gamma}}(\boldsymbol{\psi})$ and $\boldsymbol{\sigma}^2$. For simplicity, $\tilde{\mathbf{X}}(\boldsymbol{\psi}^v)$ is replaced by $\tilde{\mathbf{X}}$.[1] For each

[1]Note that $\tilde{\mathbf{X}}$ has dimension $2T \times p$, but $\tilde{\mathbf{X}}(\psi^v)$ has dimension $2T \times q_v$.

voxel v and subject s,

$$p\left(\tilde{\mathbf{y}}_{(s)}^{v} \mid \tilde{\boldsymbol{\gamma}}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}\right), \boldsymbol{\psi}_{(s)}^{v}, \sigma_{v}^{2}\right) \propto \left(\sigma_{v}^{2}\right)^{-T} \exp\left(-\frac{1}{2\sigma_{v}^{2}}\left(\tilde{\mathbf{y}}_{(s)}^{v}\right)' \tilde{\mathbf{y}}_{(s)}^{v}\right),$$
$$\times \exp\left(-\frac{1}{2\sigma_{v}^{2}}\left(-2\left(\tilde{\boldsymbol{\gamma}}_{(s)}^{v}\right)' \tilde{\mathbf{X}}' \tilde{\mathbf{y}}_{(s)}^{v} + \left(\tilde{\boldsymbol{\gamma}}_{(s)}^{v}\right)' \tilde{\mathbf{X}}' \tilde{\mathbf{X}} \tilde{\boldsymbol{\gamma}}_{(s)}^{v}\right)\right);$$
$$\pi\left(\tilde{\boldsymbol{\gamma}}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right) \mid \boldsymbol{\psi}_{(s)}^{v}, \sigma_{v}^{2}\right) \propto \left(\sigma_{v}^{2}\right)^{-q_{(s)}^{v}} |T\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}|^{-1/2} \exp\left(B_{\boldsymbol{\gamma}}\right),$$

where

$$B_{\boldsymbol{\gamma}} = -\frac{1}{2\sigma_{v}^{2}}\left(\left(\tilde{\boldsymbol{\gamma}}_{(s)}^{v}\right)' T^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \tilde{\boldsymbol{\gamma}}_{(s)}^{v} - 2\left(\tilde{\boldsymbol{\gamma}}_{(s)}^{v}\right)' T^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \hat{\tilde{\boldsymbol{\gamma}}}_{(s)}^{v} + \left(\hat{\tilde{\boldsymbol{\gamma}}}_{(s)}^{v}\right)' T^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \hat{\tilde{\boldsymbol{\gamma}}}_{(s)}^{v}\right).$$

Note that

$$(\tilde{\boldsymbol{\gamma}}^{v})' T^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \hat{\tilde{\boldsymbol{\gamma}}}_{v} = (\tilde{\boldsymbol{\gamma}}^{v})' T^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{y}}_{v},$$
$$\hat{\tilde{\boldsymbol{\gamma}}}_{v}' T^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{X}} \hat{\tilde{\boldsymbol{\gamma}}}_{v} = \tilde{\mathbf{y}}_{v}' \tilde{\mathbf{X}} T^{-1} \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{y}}_{v}.$$

Hence,

$$p\left(\tilde{\mathbf{y}}_{(s)}^{v} \mid \tilde{\boldsymbol{\gamma}}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right), \boldsymbol{\psi}_{(s)}^{v}, \sigma_{v}^{2}\right) \pi\left(\tilde{\boldsymbol{\gamma}}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right) \mid \boldsymbol{\psi}_{(s)}^{v}, \sigma_{v}^{2}\right)$$
$$\propto \left(\sigma_{v}^{2}\right)^{-(T+q_{(s)}^{v})} |T\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}|^{-1/2} \exp\left(-\frac{1}{2\sigma_{v}^{2}}\left(\tilde{\mathbf{y}}_{(s)}^{v}\right)' \tilde{\mathbf{y}}_{(s)}^{v}\right)$$
$$\times \exp\left(-\frac{1}{2\sigma_{v}^{2}}\left(-2\left(\tilde{\boldsymbol{\gamma}}^{v}\right)'\left(1+T^{-1}\right)\tilde{\mathbf{X}}'\tilde{\mathbf{y}}_{v} + \left(\tilde{\boldsymbol{\gamma}}^{v}\right)'\left(1+T^{-1}\right)\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\tilde{\boldsymbol{\gamma}}^{v}\right)\right)$$
$$\times \exp\left(-\frac{1}{2\sigma_{v}^{2}}\left(\tilde{\mathbf{y}}_{(s)}^{v}\right)' \tilde{\mathbf{X}} T^{-1}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{y}}_{(s)}^{v}\right).$$

Therefore,

$$\begin{aligned}
&\int p\left(\tilde{\mathbf{y}}_{(s)}^{v} \mid \tilde{\boldsymbol{\gamma}}_{(s)}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right), \psi_{(s)}^{v}, \sigma_{v}^{2}\right) \pi\left(\tilde{\boldsymbol{\gamma}}^{v}\left(\boldsymbol{\psi}_{(s)}^{v}\right) \mid \boldsymbol{\psi}_{(s)}^{v}, \sigma_{v}^{2}\right) d\boldsymbol{\gamma}_{(s)}^{v} \\
&\quad = \left(\sigma_{v}^{2}\right)^{-(T)} \left|T\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\right|^{-1/2} \exp\left(-\frac{1}{2\sigma_{v}^{2}}\left(\tilde{\mathbf{y}}_{(s)}^{v}\right)'\tilde{\mathbf{y}}_{(s)}^{v}\right) \\
&\qquad \times \exp\left(-\frac{1}{2\sigma_{v}^{2}}\left(\tilde{\mathbf{y}}_{(s)}^{v}\right)'\tilde{\mathbf{X}}T^{-1}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}_{(s)}^{v}\right) \\
&\qquad \times \left|\left(1+T^{-1}\right)^{-1}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\right|^{1/2} \exp\left(-\frac{1}{2\sigma_{v}^{2}}\left(\tilde{\mathbf{y}}_{(s)}^{v}\right)'\tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\left(1+T^{-1}\right)\tilde{\mathbf{X}}'\tilde{\mathbf{y}}_{(s)}^{v}\right) \\
&\quad \propto \left(\sigma_{v}^{2}\right)^{-T} T^{-q_{(s)}^{v}} \left(T/(1+T)\right)^{q_{(s)}^{v}} \exp\left(-\frac{1}{2\sigma^{2}}\left(\left(\tilde{\mathbf{y}}_{(s)}^{v}\right)'\tilde{\mathbf{y}}_{(s)}^{v} - \left(\tilde{\mathbf{y}}_{(s)}^{v}\right)'\tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}_{(s)}^{v}\right)\right).
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{A} &:= \int_{\boldsymbol{\gamma}_{(1)}^{1}} \cdots \int_{\boldsymbol{\gamma}_{(n)}^{V}} \cdots d\boldsymbol{\gamma}_{(1)}^{1} \cdots d\boldsymbol{\gamma}_{(n)}^{V} \\
&= \prod_{s=1}^{n}\prod_{v=1}^{V}\left(\sigma_{v}^{2}\right)^{-(T)}\left(1+T\right)^{-q_{(s)}^{v}} \exp\left(-\frac{1}{2\sigma^{2}}\underbrace{\left(\left(\tilde{\mathbf{y}}_{(s)}^{v}\right)'\tilde{\mathbf{y}}_{(s)}^{v} - \left(\tilde{\mathbf{y}}_{(s)}^{v}\right)'\tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}_{(s)}^{v}\right)}_{M_{(s)}^{v}}\right) \\
&= \prod_{v=1}^{V}\left(\sigma_{v}^{2}\right)^{-(nT)}\left(1+T\right)^{-\sum_{s=1}^{n} q_{(s)}^{v}} \exp\left(\frac{-1}{2\sigma_{v}^{2}}\sum_{s=1}^{n} M_{(s)}^{v}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int \mathbb{A}\cdot\pi\left(\boldsymbol{\sigma}^{2}\right) d\boldsymbol{\sigma}^{2} &= \int \prod_{v=1}^{V}\left(\sigma_{v}^{2}\right)^{-(nT+a_{\sigma}+1)}\left(1+T\right)^{-\sum_{s=1}^{n} q_{(s)}^{v}} \exp\left(\frac{-\left(\frac{1}{2}\sum_{s=1}^{n} M_{(s)}^{v} + b_{\sigma}\right)}{\sigma_{v}^{2}}\right) d\boldsymbol{\sigma}^{2} \\
&\propto \prod_{v=1}^{V}\left(\frac{1}{2}\sum_{s=1}^{n} M_{(s)}^{v} + b_{\sigma}\right)^{(nT+a_{\sigma})} \left(1+T\right)^{-\sum_{s=1}^{n} q_{(s)}^{v}}.
\end{aligned}$$

Hence,

$$p\left(\tilde{\mathbf{y}} \mid \boldsymbol{\psi}\right) \propto \prod_{v=1}^{V} \left(\frac{1}{2} \sum_{s=1}^{n} M_{(s)}^{v} + b_{\sigma} \right)^{(nT+a_{\sigma})} (1+T)^{-\sum_{s=1}^{n} q_{(s)}^{v}}$$

As a result, the marginal posterior is

$$q\left(\boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{a}, \boldsymbol{\Omega} \mid \mathbf{y}\right) \propto p\left(\mathbf{y} \mid \boldsymbol{\psi}\right) \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi, \boldsymbol{a}\right) \pi\left(\mathbf{w}\right) \pi\left(\phi\right) \pi\left(\boldsymbol{a} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}\right) \pi\left(\boldsymbol{\mu}\right) \pi\left(\boldsymbol{\Omega}\right)$$

We then infer the parameters $\boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{a}$ and $\boldsymbol{\Omega}$ via a MCMC algorithm that allows us to obtain posterior samples from this marginal posterior using Metropolis-Hastings steps and/or Gibbs steps as follows:

- Sampling $\boldsymbol{\psi}$: this involves npV updates. For $j = 1, \ldots, p$, $v = 1, \ldots, V$ and $s = 1, \ldots, n$, we have $Pr\left(\psi_{(s),j}^{v} = 1 \;\middle|\; \boldsymbol{\psi}_{(-s),(-j)}^{(-v)}, \mathbf{w}_{(j)}, \boldsymbol{a}, \phi, \mathbf{y}\right) \propto \theta\left(\psi_{(s),j}^{v} = 1\right)$, with

$$\theta\left(\psi_{(s),j}^{v} = 1\right) = \pi\left(\psi_{(s),j}^{v} = 1 \;\middle|\; S_{(s),j}^{v} = \mathbf{K}_{v}'(\phi)\mathbf{w}_{(j)}, a_{(s)}^{d}\right) \left(\frac{1}{2} \sum_{s=1}^{n} M_{(s)}^{v} + b_{\sigma} \right)^{(nT+a_{\sigma})} \times (1+T)^{-\sum_{s=1}^{n} q_{(s)}^{v}},$$

  where $\boldsymbol{\psi}_{(-s),(-j)}^{(-v)}$ are all binary indicator variables $\psi_{(s'),l}^{u}$, $l \neq j$, $u \neq v$ and $s' \neq s$. Therefore,

$$\left(\psi_{(s),j}^{v} \;\middle|\; \boldsymbol{\psi}_{(-s),(-j)}^{(-v)}, \mathbf{w}_{(j)}, \boldsymbol{a}, \phi, \mathbf{y}\right) \sim Bernoulli\left(\frac{\theta\left(\psi_{(s),j}^{v} = 1\right)}{\theta\left(\psi_{(s),j}^{v} = 1\right) + \theta\left(\psi_{(s),j}^{v} = 0\right)} \right).$$

- Sampling $\mathbf{w}$: this involves pG steps. For $j = 1, \ldots, p$ and for each $g = 1, \ldots, G$ we have that the conditional density of w_j^g is given by

$$\begin{aligned} q\left(w_j^g \mid \mathbf{w}_{(-j)}^{(-g)}, \boldsymbol{\psi}, \phi, \boldsymbol{a}, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{\psi}_{(j)} \mid \mathbf{w}_{(j)}, \phi, \boldsymbol{a}\right) \pi\left(w_j^g\right) \\ &\propto \left(\prod_{s=1}^{n} \prod_{v=1}^{V} \pi\left(\psi_{(s),j}^v \mid \mathbf{w}_{(j)}, \phi, \boldsymbol{a}\right)\right) \pi\left(w_j^g\right), \end{aligned}$$

where $\mathbf{w}_{(-j)}^{(-g)}$ are all region level spatial random effects w_l^h, $l \neq j$, $h \neq g$. A random walk proposal $\left(w_j^g\right)^* \sim N\left(\left(w_j^g\right)^{(m)}, \sigma_{w_j^g}^2\right)$ is used, where $\left(w_j^g\right)^{(m)}$ is the current state and $\sigma_{w_j^g}^2$ is a tuning parameter. Hence the Metropolis-Hastings ratio is

$$\begin{aligned} &\frac{q\left(\left(w_j^g\right)^* \,\middle|\, \mathbf{w}_{(-j)}^{(-g)}, \boldsymbol{\psi}, \phi, \boldsymbol{a}, \mathbf{y}\right)}{q\left(\left(w_j^g\right)^{(m)} \,\middle|\, \mathbf{w}_{(-j)}^{(-g)}, \boldsymbol{\psi}, \phi, \boldsymbol{a}, \mathbf{y}\right)} \\ &= \frac{\prod_{s=1}^{n} \prod_{v=1}^{V} \pi\left(\psi_{(s),j}^v \,\middle|\, S_j^v\left(\left(w_j^g\right)^*; \mathbf{K}_v(\phi), \mathbf{w}_{(-j)}^{(-g)}\right), a_{(s)}^d\right)}{\prod_{s=1}^{n} \prod_{v=1}^{V} \pi\left(\psi_{(s),j}^v \,\middle|\, S_j^v\left(\left(w_j^g\right)^{(m)}; \mathbf{K}_v(\phi), \mathbf{w}_{(-j)}^{(-g)}\right), a_{(s)}^d\right)} \times \frac{\pi\left(\left(w_j^g\right)^*\right)}{\pi\left(\left(w_j^g\right)^{(m)}\right)}. \end{aligned}$$

- Sampling ϕ: Let $\xi = \log(\phi)$. $\phi = e^{\xi}$. Then the full conditional of ξ is

$$\begin{aligned} q\left(\xi \mid \boldsymbol{\psi}, \mathbf{w}, \boldsymbol{a}, \mathbf{y}\right) &= q\left(\phi(\xi) \mid \boldsymbol{\psi}, \mathbf{w}, \boldsymbol{a}, \mathbf{y}\right) e^{\xi} \\ &= Ga\left(e^{\xi} \mid a_\phi, b_\phi\right) \cdot e^{\xi} \\ &\quad \times \prod_{s=1}^{n} \prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_{(s),j}^v \,\middle|\, S_j^v\left(\mathbf{K}_v(e^{\xi}); \mathbf{w}_{(j)}\right), a_{(s)}^d\right). \end{aligned}$$

We can use normal random walk proposal on ξ: $\xi^* \sim N\left(\xi^{(m)}, \sigma_\xi^2\right)$, where $\xi^{(m)}$ is current state value and σ_ξ^2 is a tuning parameter. Then the Metropolis-Hastings

ratio is

$$\frac{q\left(\xi^{*} \mid \boldsymbol{\psi}, \mathbf{w}, \boldsymbol{a}, \mathbf{y}\right)}{q\left(\xi^{(m)} \mid \boldsymbol{\psi}, \mathbf{w}, \boldsymbol{a}, \mathbf{y}\right)} = \frac{\prod_{s=1}^{n} \prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_{(s),j}^{v} \mid S_{j}^{v}\left(\mathbf{K}_{v}(e^{\xi^{*}}); \mathbf{w}_{(j)}\right), a_{(s)}^{d}\right)}{\prod_{s=1}^{n} \prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_{(s),j}^{v} \mid S_{j}^{v}\left(\mathbf{K}_{v}(e^{\xi^{(m)}}); \mathbf{w}_{(j)}\right), a_{(s)}^{d}\right)} \times \frac{Ga\left(e^{\xi^{*}} \mid a_{\phi}, b_{\phi}\right) \cdot e^{\xi^{*}}}{Ga\left(e^{\xi^{(m)}} \mid a_{\phi}, b_{\phi}\right) \cdot e^{\xi^{(m)}}}.$$

For the general case that $\phi = \left(\phi^{1}, \ldots, \phi^{G}\right)$, each ϕ^{g} can be updated via the scheme similar to the one described above.

- Sampling $\boldsymbol{a}$: this involves n steps. For $s = 1, \ldots, n$, the conditional density of $\boldsymbol{a}_{(s)} = \left(a_{(s)}^{1}, \ldots a_{(s)}^{D}\right)$ is given by

$$q\left(\boldsymbol{a}_{(s)} \mid \boldsymbol{a}_{-(s)}, \boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{y}\right) \propto \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi, \boldsymbol{a}\right) \pi\left(\boldsymbol{a} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}\right)$$
$$= \left(\prod_{s=1}^{n} \prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_{(s),j}^{v} \mid \mathbf{w}_{(j)}, \phi, \boldsymbol{a}_{(s)}\right)\right) N\left(\boldsymbol{a}_{(s)} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}\right),$$

where $\boldsymbol{a}_{-(s)}$ are all the connectivity parameters $\boldsymbol{a}_{(s')}$, $s' \neq s$. Block move update for $\boldsymbol{a}_{(s)}$ and a random walk proposal $\boldsymbol{a}_{(s)}^{*} \sim N_{D}\left(\boldsymbol{a}_{(s)}^{(m)}, \boldsymbol{\Sigma}_{\boldsymbol{a}_{(s)}}\right)$ is used, where $\boldsymbol{a}_{(s)}^{(m)}$ is the current state and $\boldsymbol{\Sigma}_{\boldsymbol{a}_{(s)}}$ is the tuning covariance matrix.

The Metropolis-Hastings ratio is

$$\begin{aligned}
&\frac{q\left(\boldsymbol{a}_{(s)}^{*} \mid \boldsymbol{a}_{-(s)}, \boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{y}\right)}{q\left(\boldsymbol{a}_{(s)}^{(m)} \mid \boldsymbol{a}_{-(s)}, \boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{y}\right)} \\
&\quad = \frac{\prod_{s=1}^{n} \prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_{(s),j}^{v} \mid S_{j}^{v}\left(\mathbf{K}_{v}(\phi), \mathbf{w}_{(j)}\right), \boldsymbol{a}_{(s)}^{*}\right)}{\prod_{s=1}^{n} \prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_{(s),j}^{v} \mid S_{j}^{v}\left(\mathbf{K}_{v}(e^{\xi^{*}}); \mathbf{w}_{(j)}\right), \boldsymbol{a}_{(s)}^{(m)}\right)} \\
&\qquad \times \frac{N_{D}\left(\boldsymbol{a}_{(s)}^{*}, \boldsymbol{\Sigma}_{\boldsymbol{a}_{(s)}}\right)}{N_{D}\left(\boldsymbol{a}_{(s)}^{(m)}, \boldsymbol{\Sigma}_{\boldsymbol{a}_{(s)}}\right)}
\end{aligned}$$

- Sampling $\boldsymbol{\mu}$: The full conditional of $\boldsymbol{\mu}$ is

$$\begin{aligned}
q\left(\boldsymbol{\mu} \mid \boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{a}, \boldsymbol{\Omega}, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{a} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}\right) \pi\left(\boldsymbol{\mu}\right) = N_{D}\left(\boldsymbol{a}_{(s)} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}\right) N_{D}\left(\boldsymbol{\mu} \mid \mathbf{0}, \mathbf{I}\right) \\
&\sim N_{D}\left(\left(\mathbf{I}+n\boldsymbol{\Omega}\right)^{-1} \boldsymbol{\Omega} n \overline{\boldsymbol{a}}, \left(\mathbf{I}+n\boldsymbol{\Omega}\right)^{-1}\right),
\end{aligned}$$

where $\overline{\boldsymbol{a}} = \frac{1}{n}\sum_{s=1}^{n} \boldsymbol{a}_{(s)}$. Then $\boldsymbol{\mu}$ is directly sampled from the multivariate normal distribution above.

- Sampling $\boldsymbol{\Omega}$: The full conditional of $\boldsymbol{\Omega}$ is

$$\begin{aligned}
q\left(\boldsymbol{\Omega} \mid \boldsymbol{\psi}, \mathbf{w}, \phi, \boldsymbol{a}, \boldsymbol{\mu}, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{a} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}\right) \pi\left(\boldsymbol{\Omega}\right) \\
&= \left(\prod_{s=1}^{n} N_{D}\left(\boldsymbol{a}_{(s)} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}\right)\right) Wishart\left(\boldsymbol{\Omega} \mid r_{0}, \boldsymbol{\Omega}_{0}\right) \\
&\propto |\boldsymbol{\Omega}|^{-\frac{n+r_{0}-D-1}{2}} \exp\left(-\frac{1}{2} tr\left(\sum_{s=1}^{n} \boldsymbol{a}_{(s)}\left(\boldsymbol{a}_{(s)}\right)' + \boldsymbol{\Omega}_{0}^{-1}\right) \boldsymbol{\Omega}\right) \\
&\sim Wishart_{D}\left(n+r_{0}, \left(\sum_{s=1}^{n}\left(\boldsymbol{a}_{(s)}-\boldsymbol{\mu}\right)\left(\boldsymbol{a}_{(s)}-\boldsymbol{\mu}\right)' + \boldsymbol{\Omega}_{0}^{-1}\right)^{-1}\right).
\end{aligned}$$

$\boldsymbol{\Omega}$ is directly sampled from the Wishart distribution above.

Appendix B

Description of the human CV-fMRI data and the approach of Karaman et al. (2014)

The human CV-fMRI data analyzed in Section 4.4 was recorded during a block designed fMRI experiment that consisted of a unilateral finger-tapping task performed with a visual cue indicating whether to tap or rest. The block designed experiment had an initial 20 s of rest followed by 16 epochs of 15 s of tapping (stimulus on) and 15 s of rest (stimulus off). Karaman et al. (2014) analyzed these data with 3 different models that we refer to in Section 4.4 as KBR14-CV, KBR14-MO and DeTeCT-ING. A brief description of the models follows:

KBR14-CV: this is a complex-valued constant phase activation model that linearly describes the temporally varying magnitude. Parameters are estimated via MLE and

activation images are thresholded at 5% family wise error (FWE) rate.

KBR14-MO: this model is also a constant phase activation model that linearly describes the temporally varying magnitude, but it is applied to magnitude-only fMRI data derived from the CV-fMRI data. Parameters are estimated via MLE and activation images are thresholded at 5% FWE rate.

DeTeCT-ING: this model is a more physically realistic model than the previous 2 models as it takes into account the effect that the task execution has in the magnetization equation. More specifically, the model incorporates the tissue and imaging parameters T_1 and T_2^* into the magnetic resonance equation. This is a non-linear model so parameter estimates are obtained numerically via least squares estimation. The first scans of the CV-fMRI data (with no stimulus present) are used to estimate T_1 and T_2^* and these estimates are then used to detect active voxels with the remaining scans. Activation images are also thresholded at 5% FWE rate.

Figure B.1 shows the results of the analyses in Karaman et al. (2014). From this figure it can be seen that the KBR14-CV and the DeTeCT-ING models have a higher power of detection than the KBR14-MO model, with KBR14-CV having the highest power. Activation for this experimental task is expected to occur in the left motor and suplementary area which is adequately detected by both, the KBR14-CV and the DeTeCT-ING models but mostly missed by the KBR14-MO model. With a 5% FWE rate, we also see that the KBR14-CV model leads to several false positives outside the brain, while the DeTeCT-ING model produces no false positives outside the brain. The KBR14-MO also has no false positives outside the brain, but, this model

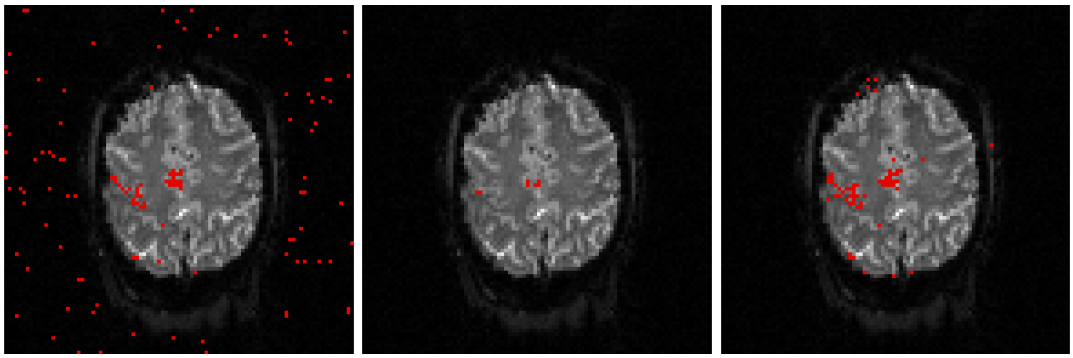


Figure B.1: Activation maps thresholded at 5% FWE rate from the Karaman et al. (2014) models: complex-valued model KBR14-CV (left plot); magnitude-only model KBR14-MO (center plot); and DeTeCT-ING model (right plot).

has low power identifying active sites in the left and supplementary motor area. The model DeTeCT-ING detects a few false positives in the upper left side of the brain, very close to the no signal area. Note that these models assume that the observations are independent in space and time.