PhD Written Candidacy Examination Part II: Statistical Modeling of fMRI Data

Rick Farouni

Tuesday 15$^{\text{th}}$ December, 2015

Examination Committee

Dr. Bob Cudeck Dr. Steve MacEachern Dr. Zhong-Lin Lu

Contents

Part I

Dr Lu: Statistical modeling of fMRI data

1 Question Part A: What is the Bayesian approach to fMRI data analysis?

1.1 Frequentist vs Bayesian Statistics

The difference between frequentist and Bayesian statistics can be summarized by how the two paradigms deal with the decision-theoretic problem of minimizing error. From a decision-theoretic perspective, statistical inference consists of the following steps:

1. Define a statistical model that involves three spaces: the **sample space** $\mathcal{Y}$, the **parameter space** Θ, and the **decision space** $\mathcal{D}$. Also, let $y \in \mathcal{Y}$ denote an **outcome** from a random experiment, and let $\theta \in \Theta$ denote a **parameter** that indexes a family of probability distributions $\mathcal{G} = \{G_\theta \mid \theta \in \Theta\}$. For example, we can choose the statistical model to be the family of univariate normal distributions with unknown location parameter $\mathcal{G} = \{N(\theta, 1) : \theta \in \Theta\}$. The y and θ are related such that observation $y \sim G_\theta$ provides evidence about $\theta \in \Theta$.

2. Define a decision $\delta \in \mathcal{D}$. When the decision is an estimator, we have $\mathcal{D} = \Theta = \mathbb{R}$ where $\mathcal{D}$ is now the set of estimators and δ is an **estimator**, a function $\delta : \mathcal{Y} \to \Theta$, whose evaluation $\delta(y)$ is called an **estimate**. The estimate gives a proposed value for the unknown θ. For example, in the case of univariate normal models, the decision $\delta(y) = y$ is called the MLE estimator.

3. Define a **loss function** $L(\theta, \delta(y))$ that measures the proximity of the estimate to the true value θ. For example, the squared loss function $L(\theta, \delta(y)) = (\theta - \delta(y))^2$ is one of the most commonly used measures, one that penalizes large deviations heavily.

The Problem The problem at hand is how to minimize a random quantity $L(\theta, \delta(y))$. The major difficulty lies in the fact that both θ and y are unknowns factors. As will be shown below, the two approaches diverge in how the two unknowns get treated.

Frequentist Solution The frequentist paradigm is a pessimistic approach that averages loss given a θ over all values of y - proportionally to the density $g(y \mid \theta)$. The frequentist average loss thus treats y as random and *theta* as fixed. It is a pessimistic approach because it discards information about y by averaging over all possible values of y even though we have already have observed y. The average loss is given by

$$\textbf{Frequentist Expected Loss} \quad R(\theta, \delta) = \mathrm{E}_y[L(\theta, \delta(y)) \mid \theta] = \int_{\mathcal{Y}} L(\theta, \delta(y))\ g(y \mid \theta) dy$$

Bayesian Solution In contrast, the Bayesian approach is an optimistic conditional perspective. It considers model parameters as random variables with prior distributions $\pi(\theta)$ that quantify our initial uncertainty about their values and with posterior distributions $\pi(\theta \mid y)$ that quantify the residual uncertainty we are left with after we observe the data. Posterior average loss does not integrate over $\mathcal{Y}$ since y is known. Instead, it conditions on the data y and integrates over the space Θ since it is θ that is unknown. The posterior expected loss thus treats θ as random and y as fixed and is given by

$$\textbf{Bayesian Expected Loss} \quad \rho(y) = \mathrm{E}_\theta[L(\theta, \delta(y)) \mid y] = \int_{\Theta} L(\theta, \delta(y)) \pi(\theta \mid y) d\theta$$

Issues with the Frequentist Approach for fMRI Data Analysis In fMRI data analysis, the frequentist perspective can be problematic in the context of hypothesis testing because what it offers is the likelihood of getting the data given a θ_0 (i.e. there is no activation). Thus, the null hypothesis states that response of a voxel to a stimulus is exactly zero. However, we know that no such voxel is possible in practice because a voxel will always have some activity. Moreover, with enough data we will be able to reject the null hypothesis for every voxel in the brain. The Bayesian approach instead gives us the probability of activation given the observed data and allows us to compute the probability that the activation was greater than a given threshold. Moreover, the frequentist expected loss attempts to find the best estimator that minimizes error given any value of θ. This assumes that we will encounter i.i.d. repetitions of exactly the same fMRI experiment infinitely many times! In practice, however, we care more about the fMRI experiment we already have conducted, not about other hypothetical experiments that are supposed to be conducted under unrealistic assumptions of experimental replicability. Another issue with the frequentist approach worth mentioning is the inflexibility of the inference procedure to deal with highly complex fMRI data. The reason is that since y

in $\delta(y)$ is treated as an unknown, finding an estimator δ that uniformly minimizes the average loss over all y is only possible for an artificially restricted set of possible estimators.

1.2 Model Setup

fMRI Data Let the vector $\mathbf{y}_v$ denote the fMRI time-series data for voxel v consisting of a magnetic resonance (MR) signal measured at T time points. The time-series data across all voxels $v = 1, \ldots, V$ can be collected in a single matrix as such

$$\begin{aligned}\mathbf{Y} &= \begin{bmatrix} \mathbf{y}_1 & \ldots & \mathbf{y}_v & \ldots & \mathbf{y}_V \end{bmatrix} \\ &= \begin{bmatrix} y_{11} & \ldots & y_{1v} & \ldots & y_{1V} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{t1} & \ldots & y_{tv} & \ldots & y_{tV} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{T1} & y_{T2} & y_{Tv} & \ldots & y_{TV} \end{bmatrix}\end{aligned}$$

The matrix shows the data as a set of V-dimensional multivariate time series of voxels. The data can also be viewed as as a time series of MR images where each row in the matrix is an image that contains a 3D scan of an entire brain volume collapsed into a vector.

The Bayesian Conditional Perspective Let $\boldsymbol{\Theta}$ refer to the collection of all the parameters in the model and let $p(\mathbf{Y}, \boldsymbol{\Theta})$ denote the joint probability distribution of the data and parameters. The Bayesian approach is a conditional perspective on statistical inference that begins with us factoring $p(\mathbf{Y}, \boldsymbol{\Theta})$ into a *data distribution* $p(\mathbf{Y} \mid \boldsymbol{\Theta})$ and a *prior distribution* $p(\boldsymbol{\Theta})$

$$p(\mathbf{Y}, \boldsymbol{\Theta}) = p(\mathbf{Y} \mid \boldsymbol{\Theta})\, p(\boldsymbol{\Theta})$$

and then through Bayes rule, gives us the joint posterior distribution of the parameters conditional on the observed data

$$p(\boldsymbol{\Theta} \mid \mathbf{Y}) = \frac{p(\mathbf{Y} \mid \boldsymbol{\Theta})\, p(\boldsymbol{\Theta})}{p(\mathbf{Y})}$$

Hierarchical Modelling The decomposition of the joint distribution can also be interpreted in terms of a two-stage hierarchical generative model

$$\boldsymbol{\Theta} \sim p(\boldsymbol{\Theta})$$

$$\mathbf{Y} \mid \boldsymbol{\Theta} \sim p(\mathbf{Y} \mid \boldsymbol{\Theta})$$

where for example, the activation effects for all the scans are sampled from a population (i.e. the prior) representing the subject and in the second stage, the data is sampled given the parameters.

We can stop here and use Empirical Bayes to obtain *point estimates* of the prior's parameters (hyperparameters), but Empirical Bayes is an approximation to a fully Bayesian method without the added benefits. A fully hierachical Bayesian model allows us to estimate the hyperparameters from the data and to fit a model with many parameters to a structured dataset without the risk of overfitting.

A fully Bayesian modelling approach assigns a hyperprior, a distribution that represents the uncertainty we have about the hyperparameter $\boldsymbol{\Xi}$. This can be expressed hierarchically as

$$\boldsymbol{\Xi} \sim p(\boldsymbol{\Xi})$$
$$\boldsymbol{\Theta} \mid \boldsymbol{\Xi} \sim p(\boldsymbol{\Theta} \mid \boldsymbol{\Xi})$$
$$\mathbf{Y} \mid \boldsymbol{\Theta} \sim p\left(\mathbf{Y} \mid \boldsymbol{\Theta}\right)$$

A hierarchical model allows us to incorporate additional information or constraints via second level predictors. The hierarchy can also be extended further, so instead of having two levels of scans-within-subject, we can model the activation effects at the session level to be informed by activation effects at the subject level, which in turn are constrained at the subject population level (i.e. a scans-sessions-subjects-population hierarchy)

An important benefit of the Bayesian approach is that it forces us to first consider the joint probability distribution of all observable and unobservable variables before we make any assumptions. A second look at the data matrix $\mathbf{Y}$ reveals that although our data is a matrix, it is actually a four dimensional array with three space and one time dimensions. The question arises on how to model the data. Are my samples a series of 3-dimensional tensors correlated in time? Or are they just simply a series of T observations on V voxels?
If the latter, then we can use a favorite trick the statistician's and assume the conditional independence of all measurements. Accordingly, we obtain the simplest sampling model possible

$$p(\mathbf{Y} \mid \boldsymbol{\Theta}) = \prod_{v=1}^{V} \prod_{t=1}^{T} \; p(y_{tv}|\theta)$$

a model that does not include any explanatory variables, that assumes neither spatial nor temporal dependency structure, and that ignores the grouping structure of the data, treating

all univariate data samples as coming from a single population governed by a one dimensional parameter. For example, the population distribution can be $p(y_{tv}|\theta) \equiv N(y_{tv}|\sigma^2)$. Of course, such a model is not much of use since it is an oversimplification. Next, we introduce an additive model of fMRI data consisting of two components: a BOLD response signal component $\mathbf{B}$ that combines with a noise component $\mathbf{E}$ to give rise to the data $\mathbf{Y} = \mathbf{B} + \mathbf{E}$. But first we begin with the design matrix.

Design Matrix Let $\mathbf{x}_q$ be a binary T dimensional **stimulus pattern vector** of zeros and ones indicating the occurrence of a presentation event (e.g. $\mathbf{x}_q = [0, 1, \ldots, 0, 1, 0]^\intercal$). An fMRI experiment is specified using a **design matrix X** with T rows and $(Q \times k)$ columns that are determined both by the number of distinct experimental conditions Q used in the experiment and by the time of stimuli presentation. If the relationship between stimulus pattern and measured MR signal is linear and instantaneous, then the columns of the design matrix $\mathbf{X}$ would just consist of the vectors $\mathbf{x}_q$ of $q = 1, \ldots, Q$, one for each experimental condition type. However, the relationship between the temporal stimulus pattern $x(t)$ and the MR signal $y(t)$ is mediated by several processes that delay the MR signal hemodynamically, thus requiring an additional k columns for each condition type to be added to the design matrix in order to model the **hemodynamic response function** (HRF). The design matrix can be expressed as a collection of Q sub-matrices, each of dimension $T \times k$ with columns consisting of shifted binary indicator vectors

$$\mathbf{X} = [\mathbf{X_1} \ldots \mathbf{X_q} \ldots \mathbf{X_Q}]$$

BOLD Response An experimental stimulus pattern $x(t)$ evokes, within several milliseconds, changes in **neural activity** $N(t)$, which lead to changes in local **cerebral blood flow** $c(t)$. Changes in blood flow affect the relative ratio of deoxyhemoglobin to oxyhemoglobin, which combine with a few other physiological variables (i.e. local blood volume changes) to form the **hemodynamic response** $h(t)$. Using fMRI, $h(t)$ can be detected as a percent change in blood oxygenation level dependent (BOLD) signal $b(t)$. Finally, $b(t)$ mixes with error components $\epsilon(t)$ such as physiological and scanner noise to produce the measured **MR signal** $y(t)$. Taking a less dynamic perspective, we can consider the sequence of events as arising from changes from a baseline. Accordingly, a stimulus presentation, gives rise to increased neuronal activation (i.e. neural response), leading to a blood flow response, which then produces a hemodynamic response.

Modelling the HRF The evoked hemodynamic response elicited by a neural event can be modelled by a hemodynmaic response function (HRF). Assuming that the dependence between the BOLD response and the stimulus can be modelled as a **linear time invariant** (LTI) system, we can express the BOLD signal $b(t)$ as the convolution of the HRF $h(t)$ with the stimulus pattern $x(t)$. That is $b(t) = (h * x)(t)$. If we assume a unique HRF vector $\mathbf{h}_q$ of length k for each stimulus condition q, the convolution can then be expressed in matrix form as $\mathbf{B} = \mathbf{XH}$, where the HRFs for all Q condition are gathered into one vector

$$\mathbf{H} = \left[\mathbf{h}_1^\intercal \dots \mathbf{h}_q^\intercal \dots, \mathbf{h}_Q^\intercal\right]^\intercal$$

If we assume additive noise and a voxel dependent HRF that, we can model the time series MR signal for a voxel v as

$$\mathbf{y}_v = \mathbf{X}\mathbf{H}_v + \boldsymbol{\epsilon}_v$$

Under certain assumption, we can assume that $\mathbf{h}_q = \beta_q \bar{\mathbf{h}}_q$, where $\bar{\mathbf{h}}_q$ is the shape of the HRF and β_q is the amplitude of the neural response. We obtain

$$\mathbf{H} = \left[\bar{\mathbf{h}}_1^\intercal \beta_1 \dots \bar{\mathbf{h}}_q^\intercal \beta_q \dots, \bar{\mathbf{h}}_Q^\intercal \beta_Q\right]^\intercal$$

If we further assume that the HRF does not vary across conditions, we get

$$\mathbf{H} = [\mathbf{h}^\intercal \beta_1 \dots \mathbf{h}^\intercal \beta_q \dots, \mathbf{h}^\intercal \beta_Q]^\intercal = \boldsymbol{\beta} \otimes \mathbf{h}$$

so we can write the voxelwise model as $\mathbf{y}_v = \mathbf{X}(\boldsymbol{\beta}_v \otimes \mathbf{h}_v) + \boldsymbol{\epsilon}_v$, and the model for $\mathbf{Y}$ as

$$\begin{bmatrix}\mathbf{y}_1 & \dots & \mathbf{y}_V\end{bmatrix} = \mathbf{X}\begin{bmatrix}(\boldsymbol{\beta}_1 \otimes \mathbf{h}_1) & \dots & (\boldsymbol{\beta}_V \otimes \mathbf{h}_V)\end{bmatrix} + \begin{bmatrix}\boldsymbol{\epsilon}_1 & \dots & \boldsymbol{\epsilon}_V\end{bmatrix} \tag{1}$$

2 Question Part B: What are some of the more important applications utilizing the Bayesian Approach?

A cursory examination of Model 1 shows that the possible targets of estimation in an fMRI data analysis are the following:

- The $Q \times V$ activation coefficients
- The $p \times V$ basis of the HRF functions, where $p \leq k$
- The parameters that govern the distribution of the $T \times V$ error terms

In general, the number of parameters that needs to be estimated can be quite large, even more than the data points, if no restrictive assumptions are made regarding the following aspects of the model

- The shape of the HRF's
- The voxelwise temporal dependency structure of the error vector $\boldsymbol{\epsilon}_v$ for voxel v
- The spatial dependency structure of the V voxels at time t.
- The distributional form of the random elements in the model

In what follows, we discuss some noteworthy Bayesian models that have succeeded in improving statistical inference with regards to several of the above-mentioned issues

2.1 Temporal Models

Posterior Probability Mapping The first paper to apply a fully Bayesian analysis to the computation of brain activation maps was by Frank, Buxton, and Wong (1998). The paper put forward a GLM-based voxelwise temporal model of fMRI data and demonstrated for the first time the basic principle of posterior probability mapping (PPM) through the thresholding of the posterior distributions of activations at a set confidence level.

Biologically Plausible HRF Shapes Models that can adequately model and capture the HRF signal reduce the amount of temporal correlation present in the voxelwise time series. Several Bayesian approaches to modelling the HRF have been proposed. For example, Genovese (2000) developed parametric HRF models whose parameters govern certain characteristics of the HRF's shape (e.g. the time-to-peak). Their Bayesian formulation allowed them to give

the parameters restrictive priors that can effectively rule out shapes that are not biological plausible.

Robust HRF Estimator Since the shape of the hemodynamic response is highly variable across the brain, the assumption that the HRF's shape of is known and similar across brain regions can be problematic. Some methods allow the HRF to have any shape, but the flexibility comes at a cost. By estimating the magnitude of the HRF at each sampled time point (e.g. every 2s), we would need to estimate 15 parameters at each voxel if we assume the HRF to last 30 seconds. The large number of parameters renders any inference unstable. One Bayesian approach to robustly estimate the HRF was proposed by Marrelec, Benali, Ciuciu, Pélégrini-Issac, and Poline (2003) in which a regularizing prior on the parameters is used. Their proposed prior assumes two things: that the HRF starts and ends at 0, and that the HRF is smooth. Their Bayesian estimator was proven to give more accurate and robust results than the ML estimator for both activation detection and HRF estimation.

2.2 Spatial Models

Adaptive Spatial Smoothing The previous paragraph discussed methods that deal with the temporal aspect, not the spatial aspect of the hemodynamic response. Now, since neighbouring voxels tend to have similar level of activity (i.e. activity occurs in clusters), fMRI data tends to be characterized by spatial dependence. In frequentist analysis of fMRI data, the spatial dependency aspect of the hemodynamic response is usually dealt with by spatially smoothing the data using a Gaussian kernel during preprocessing. In contrast, the Bayesian framework is able to explicitly model spatial dependence, whether its due to the distributed neural activation or to the spatial extent of the hemodynamic response, by means of specifying an adaptive spatial prior on the activation coefficients. For example, Penny, Trujillo-Barreto, and Friston (2005) and Gössl, Auer, and Fahrmeir (2001) used a Gaussian Markov random field prior that functions as smoothing process of parameter estimates, thus enhancing the detection of spatial activation. Harrison, Penny, Ashburner, Trujillo-Barreto, and Friston (2007) extended their framework by using Gaussian process priors that essentially incorporate a spatially nonstationary smoothing process into the generative model. Spatial smoothing thus becomes part and parcel of posterior inference and estimation. In the approaches discussed, the strength of the smoothing effect are automatically controlled by the hyperparameters of the spatial prior. In a fully Bayesian model, the hyperparameters are estimated from the data and thus do not require any fine tuning. As a result, each activation coefficient gets smoothed according to the amount of uncertainty

remaining, given the data.

2.3 Spatiotemporal Models

Models with Increased Validity Although fMRI data exhibit complex spatio-temporal dependence, most commonly used models resort to oversimplifications for the sake of computational feasibility. Several attempts have been made to model both the spatial and temporal dependencies of the noise and the signal components of the data. Woolrich, Jenkinson, Brady, and Smith (2004) was the first to propose a fully Bayesian spatiotemporal framework that models the noise as non-separable space-time vector autoregressive process. The framework also includes a parametrized model of the HRF signal consisting of four bases functions with a regularizing prior on them, thus allowing the variability of the HRF shape across brain areas and for different experimental conditions to be captured. Flandin and Penny (2007) proposed another fully Bayesian nonsperarble spatiotemporal model that instead uses a Bayesian wavelet approach to spatially constrain the regression parameters. The proposed model produced adaptively regularised parameter estimates of the HRF signal allowing it to capture non-stationary variations in smoothness across brain regions. Moreover, the estimates were shown to be more accurate than estimates obtained using a Gaussian Markov random field prior or OLS estimates.

3 Question Part C: What are the advantages and limitations of the Bayesian approach?

3.1 Advantages

Bayesian Models Eliminate the Need for Multiple Comparisons In frequentist hypothesis testing, the multiple comparison problem is a major concern of fMRI statistical analysis given the large number of voxels presents in a typical dataset. The problem involves finding significant effects when searching through many voxels concurrently. As an analogy , assume that you are willing to decide that a coin is biased (i.e. there is an effect) if ten out ten times it turns up heads. The probability of that happening is low, but if now you flip 1000 fair coins ten times and consider all the outcomes simultaneous, then the probability that you will witness a series of ten heads is much higher. Using a Bayesian approach, multiple comparisons are not a problem because the parameters have a joint distribution which allows to compute the posterior probability of any combination of events we wish. Moreover, a hierarchical Bayesian model automatically addresses this concern by shrinking the estimates of the parameters toward each other as determined by the prior.

Biophysically Informed Prior Improve Inference Within the GLM approach, the shape of the HRF is modelled using a set of basis functions that are flexible enough to capture shape variation across voxels and subjects. However, this approach has the drawback of producing HRFs with nonsensical shapes. By imposing a biophysically informed prior on the parameters of the basis function, we can constrain the possible HRF shapes to those that are biologically plausible and improve the sensitivity of the estimates.

Spatial Regularizing Priors Produce Spatial Smoothing Regularization priors such as the Gaussian markov field priors encode the assumption that the values of neighbouring voxels are similiar (i.e. smooth). These priors allow the clustering of brain activity regions into functionally uniform parcels.

Hyperpriors Allow Remote Dependency Between Voxels The exact value of the prior's variance hyperparameter determines the extent of *information pooling* within the voxels. A hyperparameter set to infinity is equivalent to imposing an uninformative prior. An uninformative flat prior corresponds to *no pooling* of information and is equivalent to conducting a separate regression for each voxel's time course. On the other hand, a hyperparameter variance set

to zero corresponds to a *complete pooling* of information that ignores any differences between the voxels, effectively treating the entire brain volume as consists of multiple observations of one single voxel. Alternativly, if we assign a prior distribution on the hyperparameter (i.e. a hyperprior), we transform the model into a fully Bayesian hierarchical model that strikes a compromise between the two extremes of *complete pooling* and *no pooling*, allowing both information between distant voxels to be shared and the value of the hyperparameter to be estimated from the data.

Bayesian Hierarchical Models are less Dependent on Parameter Tweaking Fully Bayesian hierarchical models are less sensitive to the choice of user-defined parameters. In the case of spatial smoothing, for example, the hyperparameters that regulate the degree of smoothing can be adaptively estimated from the data.

3.2 Limitations

The limitation of the Bayesian approach to fMRI data are two fold.

High Expertise Barrier and Commitment to Objective Truth Bayesian modelling requires a great deal of thought and expertise to be applied correctly. The researcher needs to have a solid knowledge in statistics and an adequate understanding of both the dependency structure of the data and the generative process that represents the structure of the causal mechanism giving rise to the observations. The main pitfall lies in people's tendency to accept those results that confirm their preconceptions and discard results that they didn't expect to find. The choice of informative prior distributions can vary from person to person, thus introducing strong assumptions that risk the objectivity of the scientific enterprise. This is especially the case when their choice is not informed by biophysics or neuroscience. A thoughtful consideration of the process that maps stimulus precept to neural activity and subsequently to the BOLD response is essential for disciplined modelling. That said, the logic of the Bayesian method is very much in agreement with the inductive process that underpins the scientific method, the process by which we update the strength of our beliefs after we observe new data and conduct new experiments. It is important to note that although informative priors can indeed be subjective priors, they can also be informative priors based on objective data.

Bayesian Computation can be Intensive In frequentist statistics, there exists several few cases where the analytical form of the asymptotic null distribution have been derived.

In contrast, the Bayesian approach allows you to fit any model you can come up with no matter how complicated it is. The flexibility however comes at a cost. Bayesian inference often needs to evaluate quantities that are generally not analytically tractable, requiring numerical integration and intricate computation. The computations can be complex, sometimes requiring a substantial coding effort on behalf of the researcher. This is especially true in the domain statistical analysis of fMRI data, where there is still does not yet exist a generic inference engine that can perform probabilistic inference on an arbitrary model of fMRI data.

4 Question Part D: what are the future directions of the Bayesian approach to fMRI analysis?

To be able to predict the direction of research the Bayesian approach might take, we should note that the main targets of inference in an fMRI analysis are:

1. The estimation of the hemodynamic response function (HRF), whose shape can vary by condition, voxel, or subject; and whose spatial extent induces dependencies across neighbouring voxels

2. The detection of brain activations in response to experimental stimuli and the identification of patterns of dependencies across voxel timecourses (i.e. connectivity)

A principled statistical Bayesian framework for modelling fMRI signal data begins with incorporating what is already known about the spatio-temporal dependence properties of both the signal and the noise components of fMRI data and then proceeds to construct models that can efficiently capture important patterns in the data. Two research venues that address the two main targets mentioned above and where Bayesian modelling seems to show promise are summarized next.

Nonlinear Modelling of the BOLD Response The models that have been discussed so far have been all based on a general linear model in which the BOLD response signal component was assumed to be linear. Yue, Loh, and Lindquist (2010) proposed a fully Bayesian model that attempts to capture the non-linearities that are usual present in the BOLD response such as when brain vasculature tends to overreact to activation. Their model imposes an intrinsic Gaussian Markov random field (IGMRF) spatial prior on a bivariate function $f(u_i, u_j)$ that represents the image. The prior acts as a Gaussian smoothing kernel, but one that varies across space and time with the amount of spatial smoothing dependent on the strength of activation. However, their model lacks a time component and does not includes a design matrix.

Functional Parcellation using Nonparametric Bayesian Mixture Models To study the activity of a group of voxels, we can do one of three things: we can (1) use functional or anatomical regions of interest (ROIs) based on previous experiments (2) use a brain atlas that provides pre-defined labels based on anatomical landmarks (3) use functional parcellations models to clusters the voxels into groups with similar activation profiles. A successful approach to detecting activations and their spatial distribution uses mixture models that assume that the

brain map of voxel activations is made up of a mixture of clusters, where each cluster represents a distinct functional system. One problem with finite mixture models is that noise artefacts characterized by temporal trend invalidate the models assumptions. Moreover, the specification of the number of clusters beforehand creates a need for multiple model comparisons to determine the most likely number of clusters. Lashkari et al. (2012) proposed a nonparameteric Bayesian model that uses a hierarchical Dirichlet prior on the probability parameter of the Bernoulli activation variable such that voxels that belong in the same functional parcel have the same probabilities of activation. The hierarchical specification allows the model to learn patterns of functional specificity shared across a group of subjects. Zhang, Guindani, Versace, and Vannucci (2014) proposed wavelet-based spatiotemporal nonparametric Bayesian model with an MRF spatial smoothing prior on the activation parameter and a Dirichlet Process prior on the long-memory process parameter of the wavelet transformed error component. The Dirichlet Process prior induces a clustering of distant voxel timecourses that share similar profiles. It is interesting to note that whereas Lashkari applies a temporal filter to decorrelate the noise before imposing the DP prior on the activation parameters, Zhang et al. (2014) model imposes the DP prior on the error term parameters and fits the model to a single slice at a time. Their model can be extended to multiple subject analysis by using a hierachical Dirichlet process prior similiar to Zhang et al. (2014) approach and to the spatial priors that take into account the 3-dimensional spatial structure of the data. Alternatively, one can attempt a surface based approach (Fischl, Sereno, & Dale, 1999) that uses a geodesic distance metric to define neighbourhoods of voxels and that only includes voxels which are classified as grey matter in the analysis.

References

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system. *Neuroimage*, *9*(2), 195–207.

Flandin, G., & Penny, W. D. (2007). Bayesian fmri data analysis with sparse spatial basis function priors. *NeuroImage*, *34*(3), 1108–1125.

Frank, L. R., Buxton, R. B., & Wong, E. C. (1998). Probabilistic analysis of functional magnetic resonance imaging data. *Magnetic resonance in medicine*, *39*(1), 132–148.

Genovese, C. R. (2000). A bayesian time-course model for functional magnetic resonance imaging data. *Journal of the American Statistical Association*, *95*(451), 691–703.

Gössl, C., Auer, D. P., & Fahrmeir, L. (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, *57*(2), 554–562.

Harrison, L., Penny, W., Ashburner, J., Trujillo-Barreto, N., & Friston, K. (2007). Diffusion-based spatial priors for imaging. *NeuroImage*, *38*(4), 677–695.

Lashkari, D., Sridharan, R., Vul, E., Hsieh, P.-J., Kanwisher, N., & Golland, P. (2012). Search for patterns of functional specificity in the brain: a nonparametric hierarchical bayesian model for group fmri data. *Neuroimage*, *59*(2), 1348–1368.

Marrelec, G., Benali, H., Ciuciu, P., Pélégrini-Issac, M., & Poline, J.-B. (2003). Robust bayesian estimation of the hemodynamic response function in event-related bold fmri using basic physiological information. *Human Brain Mapping*, *19*(1), 1–17.

Penny, W. D., Trujillo-Barreto, N. J., & Friston, K. J. (2005). Bayesian fmri time series analysis with spatial priors. *NeuroImage*, *24*(2), 350–362.

Woolrich, M. W., Jenkinson, M., Brady, J. M., & Smith, S. M. (2004). Fully bayesian spatio-temporal modeling of fmri data. *Medical Imaging, IEEE Transactions on*, *23*(2), 213–231.

Yue, Y., Loh, J. M., & Lindquist, M. A. (2010). Adaptive spatial smoothing of fmri images. *Statistics and its Interface*, *3*, 3–13.

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.