

Bayesian analysis of neuroimaging data in FSL[☆]

Mark W. Woolrich^{a,*}, Saad Jbabdi^a, Brian Patenaude^a, Michael Chappell^a, Salima Makni^a, Timothy Behrens^a, Christian Beckmann^{a,b}, Mark Jenkinson^a, Stephen M. Smith^a

^a University of Oxford Centre for Functional MRI of the Brain (FMRIB), UK

^b Imperial College, London, UK

ARTICLE INFO

Article history:

Received 2 September 2008

Accepted 15 October 2008

Available online 13 November 2008

ABSTRACT

Typically in neuroimaging we are looking to extract some pertinent information from imperfect, noisy images of the brain. This might be the inference of percent changes in blood flow in perfusion FMRI data, segmentation of subcortical structures from structural MRI, or inference of the probability of an anatomical connection between an area of cortex and a subthalamic nucleus using diffusion MRI. In this article we will describe how Bayesian techniques have made a significant impact in tackling problems such as these, particularly in regards to the analysis tools in the FMRIB Software Library (FSL). We shall see how Bayes provides a framework within which we can attempt to infer on models of neuroimaging data, while allowing us to incorporate our prior belief about the brain and the neuroimaging equipment in the form of biophysically informed or regularising priors. It allows us to extract probabilistic information from the data, and to probabilistically combine information from multiple modalities. Bayes can also be used to not only compare and select between models of different complexity, but also to infer on data using committees of models. Finally, we mention some analysis scenarios where Bayesian methods are impractical, and briefly discuss some practical approaches that we have taken in these cases.

Crown Copyright © 2008 Published by Elsevier Inc. All rights reserved.

Introduction

In a typical neuroimaging scenario we are looking to extract pertinent information about the brain from noisy data. Mapping measured data to brain characteristics is generally difficult to do directly. For example, since FMRI data is noisy we cannot simply use a rule that says “If the FMRI data looks exactly like X, then the brain is active in area Y”. However, it is comparatively easy to turn the problem around and specify “If the brain is active in area Y, then the FMRI data should look like X”, i.e., if we know what the brain is doing then we can predict what our neuroimaging data should look like. This is what we refer to as a *Generative Model*, which sits at the heart of all Bayesian neuroimaging analysis methods. Fig. 1 illustrates an example of a generative model for predicting FMRI data.

Generative models are a natural way for us to incorporate our understanding of the brain and of the neuroimaging modalities to make predictions about what neuroimaging data looks like. However, in practice we want to do the opposite. We want to be able to take acquired data (plus a generative model) and extract pertinent information about the brain (i.e., “infer” on the model and its

parameters). The classical approach to doing this is to *fit* the generative models to the data, for example by minimising the squared difference between the data and the generative model to estimate each parameter in the model¹. However, this approach has limitations. Firstly, extracting a single “best guess” (or *point estimate*) for a parameter completely ignores the presence of, or extent of, the uncertainty that we have in that parameter. Secondly, how do we systematically combine the information in the data with any prior knowledge that we have about the parameters in the model? Bayesian statistics offers a solution to these problems, and as we shall see also provides a framework in which we can do much more besides. For example, we can probabilistically combine information from multiple modalities, compare and select between models of different complexity, and infer on data using committees of models.

Bayes' rule

Bayesian statistics provide the only generic framework for the adjustment of belief (in the form of probability density functions (PDFs)) in the presence of new information (Cox, 1946). They give us a tool for inferring on *any* model we choose, and *guarantee* that uncertainty will be handled correctly.

[☆] The FMRIB Software Library (FSL) contains methods developed by the authors of this article.

* Corresponding author. Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK. Fax: +44 1865 222717.

E-mail address: woolrich@fmrib.ox.ac.uk (M.W. Woolrich).

¹ This approach assumes that the noise is Gaussian and is an example of a *maximum likelihood* approach.

Bayes' rule tells us how (for a model \mathcal{M}) we should use the data, \mathbf{Y} , to update our *prior* belief in the values of the parameters Θ , $p(\Theta|\mathcal{M})$ to a *posterior* distribution of the parameter values $p(\Theta|\mathbf{Y}, \mathcal{M})$:

$$p(\Theta|\mathbf{Y}, \mathcal{M}) = \frac{p(\mathbf{Y}|\Theta, \mathcal{M})p(\Theta|\mathcal{M})}{p(\mathbf{Y}|\mathcal{M})} \quad (1)$$

The term $p(\mathbf{Y}|\Theta, \mathcal{M})$ is the *likelihood* and typically corresponds to the generative model. Often in neuroimaging, data from different voxels are considered to be *conditionally independent*, i.e., conditioned on the parameters, the data is independent across voxels. This basically means that the likelihood can be factorised over voxels:

$$p(\mathbf{Y}|\Theta, \mathcal{M}) = \prod_i p(\mathbf{Y}_i|\Theta_i, \mathcal{M}) \quad (2)$$

The voxelwise likelihood, $p(\mathbf{Y}_i|\Theta_i, \mathcal{M})$, is then specified by a voxelwise generative model. Fig. 2 illustrates just such a case for the application of Bayesian inference on BOLD–fMRI data from Fig. 1.

Unfortunately, calculating the *posterior PDF* given in Eq. (1) is seldom straightforward. The denominator in Eq. (1) is:

$$p(\mathbf{Y}|\mathcal{M}) = \int_{\Theta} p(\mathbf{Y}|\Theta, \mathcal{M})p(\Theta|\mathcal{M})d\Theta \quad (3)$$

an integral which is often not tractable analytically. Furthermore, this *joint posterior PDF* on all parameters is often not the distribution that we are most interested in. We are often interested in the *posterior PDF* on a single parameter, or an interesting subset of parameters. Obtaining these *marginal* distributions again involves performing complicated, high-dimensional integrals:

$$p(\Theta_i|\mathbf{Y}, \mathcal{M}) = \int_{\Theta_{-i}} p(\Theta|\mathbf{Y}, \mathcal{M})d\Theta_{-i} \quad (4)$$

where Θ_i are the parameters of interest and Θ_{-i} are all other parameters. Again the integrals in Eq. (4) are seldom tractable analytically.

It is beyond the scope of this article to go into the technical details of how we can overcome the intractability of these integrals. However, in the Bayesian inference techniques section we consider some of the possibilities.

Priors

Bayesian statistics requires that we specify our prior probabilistic belief about the model parameters. This requirement has often been a source of criticism of the Bayesian approach. However, Bayesians support the view that we cannot infer from data without making assumptions; indeed, the act of choosing a generative model itself constitutes an assumption (that the model provides a good description of reality). It turns out that having a framework within which we can specify prior assumptions can be a big advantage. As we shall see, this can serve to augment the assumptions already made in the generative model with complementary knowledge of the system.

Biophysical priors

Biophysical priors are priors that encode what we understand as being biologically plausible. For example, we know that a value of 1.3 s at 3 T for the T1 of grey matter is plausible, whereas values of 0.3 s and 2.3 s are not. Within Bayes we can encode this information in the form of prior PDFs. This information will then be probabilistically combined with the information in the data within Bayes' rule.

A good example of the use of biophysical priors is in the analysis of Arterial Spin Labelling (ASL) data. ASL is a non-invasive MRI technique that measures regional cerebral blood flow (CBF) by magnetically tagging blood and using it as an endogenous contrast agent. A model often used for ASL is summarised in Fig. 3a. A traditional, non-Bayesian approach is to assume an exact, fixed value for the T1 of grey

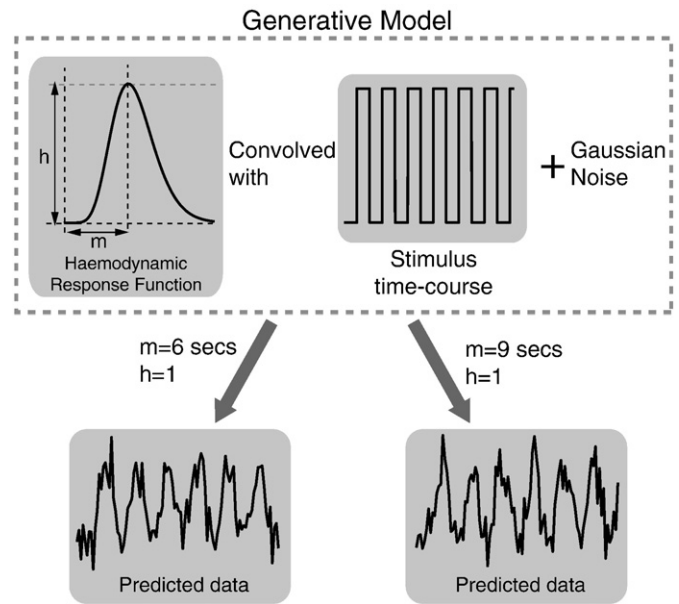


Fig. 1. Generative models are at the core of all Bayesian neuroimaging analysis techniques. This figure shows an example of a simple voxelwise generative model for predicting fMRI data. It consists of a parameterised haemodynamic response function (HRF) that is convolved with a time-course of the known experimental stimulus (in this case a boxcar stimulus) and then added to Gaussian noise. For the sake of simplicity we are assuming that the variance of the Gaussian noise is known. The only unknown parameters in the model are the time-to-peak, m , and the height, h , of the HRF. For different values for m and h , we can predicted what the fMRI data looks like in a voxel.

matter (e.g. 1.3 s at 3 T), and then fit the model to the data using non-linear least squares techniques to estimate the CBF, Δt and τ . However, this approach does not take advantage of what we know are biophysically plausible values for the bolus arrival time and length, and does not acknowledge that the T1 of grey matter may not be *exactly* 1.3 s. The Bayesian approach offers us the opportunity to encode biophysically realistic assumptions about these parameters. Fig. 3b shows the prior assumptions we can make about the parameters in the model (Chappell et al., 2008). We can see that when we have no prior information about a parameter such as the CBF, we can use a *non-informative* prior accordingly. At the other extreme we have quite strong knowledge about the plausible values for the T1 of grey matter, but we are not limiting ourselves to assuming an exact value. Fig. 3c demonstrates the result of Bayesian inference in the form of marginal posterior means for the different parameters in the model. This approach is implemented as part of the FABBER tool in the FMRIB Software Library (FSL).

Another example of the benefits of using biophysical prior information is in inferring on fMRI time-series models. fMRI analysis requires flexible haemodynamic response function (HRF) modelling, both across the brain and between subjects. As shown in Fig. 4 this flexibility can be achieved within the General Linear Model (GLM) framework by using basis functions. However, it is possible for a basis set to produce nonsensical HRFs. Priors can be placed on the basis function regression parameters such that we constrain the inference to only those combinations of parameters that give biophysically plausible HRF shapes (Woolrich et al., 2004b). As illustrated in Fig. 4d the inclusion of this prior information results in increased sensitivity.

Biophysical information for the use in priors can be derived from first principles, from the literature or directly from data in the form of training datasets. A good example of the latter is in FSL's subcortical segmentation tool, FIRST, which combines Bayesian concepts with active appearance models (see Fig. 5) (Patenaude et al., 2007). This segments subcortical structures, such as the thalamus, as parameterised surfaces (tessellated meshes). To constrain this segmentation,

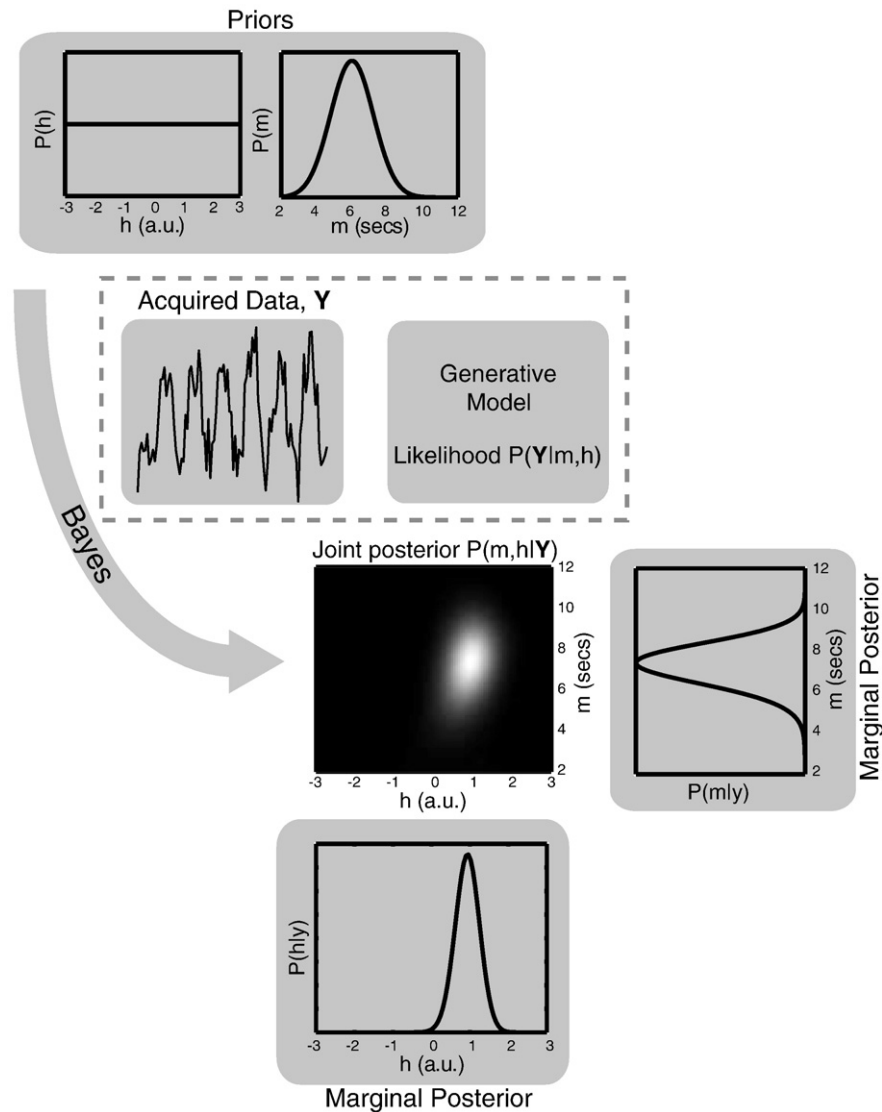


Fig. 2. Schematic of the process of Bayesian inference on the generative model in Fig. 1. The ingredients consist of the acquired FMRI data, the generative model, and the priors. The generative model provides us with the likelihood term in Bayes' rule. Bayes combines these ingredients to give us probabilistic inference on the model parameters (Eq. (1)) in the form of the joint posterior distribution across all parameters in the model. However, we are often interested in the posterior distribution on a single parameter, or a subset of parameters. This can be obtained by integrating (i.e., averaging) over parameters to obtain the “marginal” distribution (Eq. (4)).

biophysical prior information is derived from a training set consisting of manually segmented structural images. Shape and intensity are extracted from these training images by generating surface parameterisations of the individual structures and sampling the image intensities along the surface normals. Principal modes of variation in shape and intensity are computed from this to form basis functions for use in the generative model (these modes describe how shapes and intensities vary and co-vary across all of the subjects in the training dataset). However, as with the use of priors on basis sets for HRF modelling in FMRI (Fig. 4), we can also use information (extracted from our training set) to put priors on the sorts of combinations of these basis functions that are plausible. In particular, Bayes naturally provides a solution to the problem of how to weight the information between shape and intensity, such that for each combination of basis functions the intensity distribution conditioned on the shape may be generated. This is a particular problem as the shape and intensity are in completely different units. Without Bayes this weighting is heuristic, lacking a principled motivation. In addition, the Bayesian formulation takes into account the finite nature of the training set. Typically the dimension of the learnt data (shape coordinates and

intensities) is tens of thousands, while the number of training sets is in the hundreds (around 300 here) and consequently it is not possible to infer all aspects of the distribution of the data from the training set. Using Bayesian methods and applying a non-informative prior allows a natural way to apply techniques for estimating the distribution properties without resorting to arbitrary methods.

Regularisation priors

We already have seen how prior information can take the form of biophysically or empirically informed priors. We now consider priors that can be used to “regularise” (or to improve the stability in the estimation of) parameters. For example, a spatial regularisation prior captures the assumption that the value of a particular parameter at one voxel is likely to be similar to the value of the corresponding parameter at a neighbouring voxel.

A good example where such prior information can be crucial is in the FSL tool FAST, which segments structural MR images into different brain tissue types. As illustrated in Fig. 6, at the core of the approach is a generative model that predicts the structural MR data

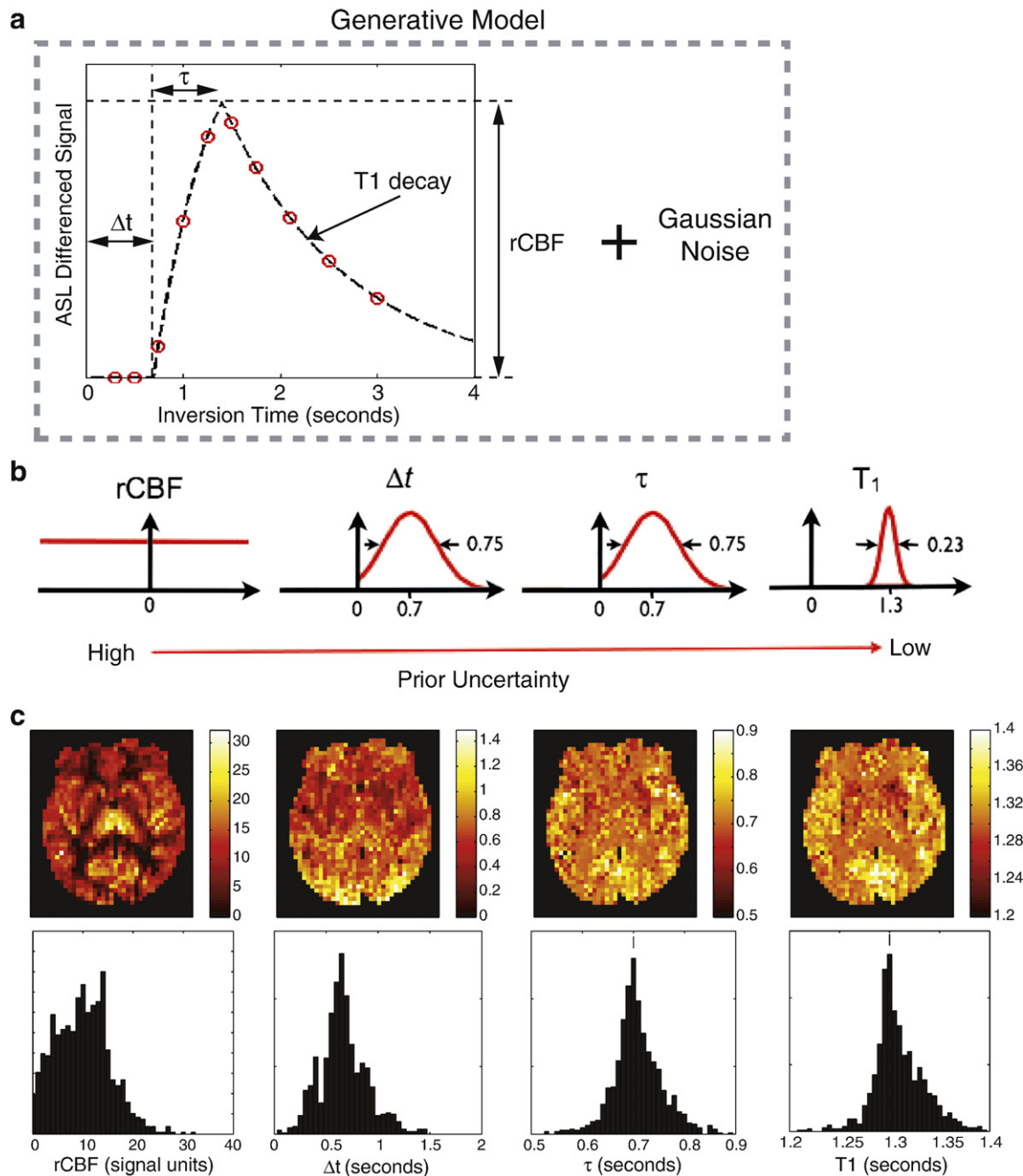


Fig. 3. The use of biophysical priors and Bayes to infer on Arterial Spin Labelling (ASL) data. The key parameter of interest in ASL is the regional cerebral blood flow (rCBF). (a) The model makes assumptions about the delivery of magnetically tagged blood into an imaged voxel to predict the ASL (tag minus control) differenced signal (indicated by red circles) at different times after the creation of the tag, referred to as inversion times. The model includes parameters describing the bolus arrival time, Δt , and the bolus length, τ , of magnetically tagged blood, and the T_1 of grey matter. The generative model consists of this model added to a Gaussian noise model. (b) A Bayesian approach allows us to place priors encoding the strength of our prior knowledge about different parameters in the model. (c) The result of Bayesian inference showing the voxelwise marginal posterior means for the different parameters in the model in the form of spatial maps [top] and histograms [bottom].

in the form of its histogram. Specifically, the histogram of image intensities is modelled as being made up of a mixture of distributions (known as a *mixture model*), where each tissue type is modelled as a Gaussian distribution. The means and variances of these Gaussian distribution “classes” are parameters in the model alongside the discrete classification parameters that label which class each voxel belongs to.

FAST places the generative model in the Bayesian framework and augments the histogram model with a spatial regularisation prior on the classification parameters (effectively, this is “smoothing” the final

segmentation). This greatly reduces the effect of noise on the segmentation. The spatial prior takes the form of a discrete Markov random field (MRF) prior:

$$p(x = \kappa | \phi_x) \propto f(\phi_x) \exp\left(-\frac{\phi_x}{4} \sum_i \sum_{j \in N_i} I[x_i \neq x_j]\right) \quad (5)$$

where x_i is the classification label at voxel i , N_i is the spatial neighbourhood of i , $I[x_i \neq x_j]$ is a function that is 1 if $x_i \neq x_j$ and is 0 otherwise, and $f(\phi_x)$ is some unknown function of ϕ_x . Note that ϕ_x is an important parameter that controls the strength of the spatial prior,

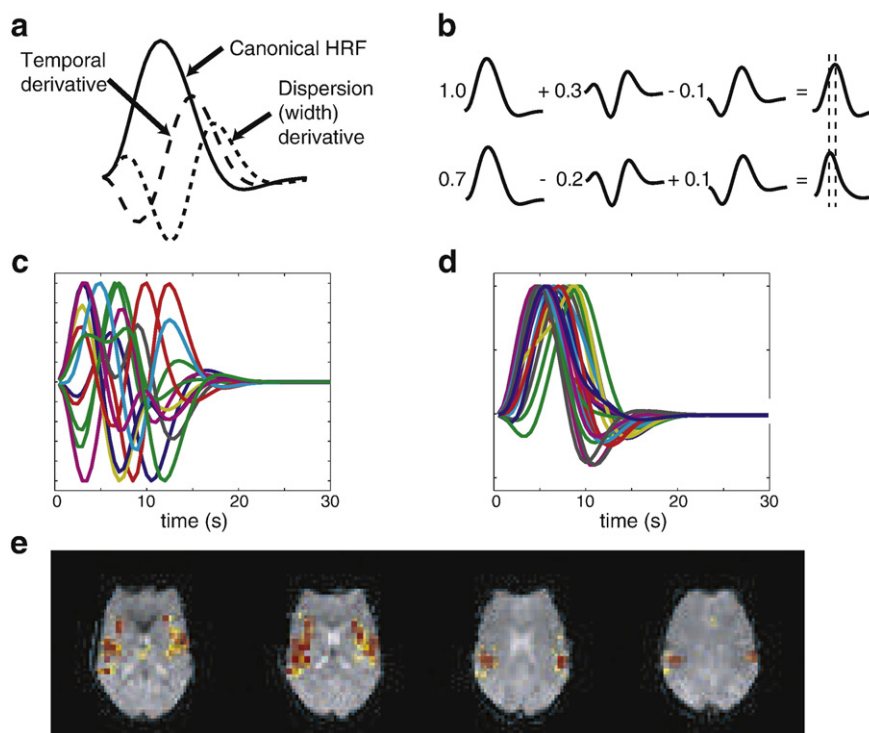


Fig. 4. The use of biophysical prior information on the shape of the haemodynamic response in FMRI data can provide increased sensitivity in detecting activity. (a) Example of a haemodynamic response function (HRF) basis set containing three basis functions. (b) Different linear combinations of the basis functions can be used to obtain different HRF shapes. (c) It is possible for many of the possible linear combinations of the basis set to produce nonsensical looking HRFs. (d) Bayesian biophysical priors can be used such that we constrain the inference to only those sets of parameters that give biophysically plausible HRF shapes. (e) Difference in thresholded activation between the HRF model with Bayesian priors and the HRF model without Bayesian priors for a single-event pain stimulus dataset: [red] voxels are active for both models, [yellow] voxels are active for just the constrained HRF model, and [blue] voxels are active for just the unconstrained HRF model.

i.e., how smooth the posterior classification image looks. Due to the fairly consistent granularity of the tissue classes and noise levels in the structural images, the classification is fairly robust to the actual value of ϕ_x , and so this can be tuned heuristically without too much difficulty.

As shown in Fig. 7, another application of spatially regularised mixture models is in segmenting FMRI statistical images into “active” and “non-active” voxels (Woolrich et al., 2005; Woolrich and Behrens, 2006). However, unlike FAST, this approach needs to be able to deal with a wide range of potential classification maps with different amounts of granularity. For this reason, the parameter that controls the amount of spatial smoothing (ϕ_x in Eq. (5)) needs to be determined adaptively from the data. This is achieved by approximating the discrete-valued MRF with a continuous MRF. All parameters in the model can then be adaptively determined from the data, and hence we can infer regions of brain activity objectively with regards to the amount of spatial smoothing that should be applied. Being able to determine the strength of the spatial regularisation adaptively from the data itself demonstrates an important feature of fully Bayesian approaches. It allows for the development of analysis methods which are less dependent on user-defined variables that might otherwise get tweaked to give an infinite number of possible solutions. The use of mixture modelling on statistical images also has the potential to offer greater inference flexibility. We can still look to control the false positive rate (FPR) by thresholding using the estimated PDF of the “non-active” voxels. However, now we could also look to approximately control the true positive rate (TPR) by thresholding using the probability of a voxel being “active”. This may be of real importance when using FMRI for pre-surgery planning (Bartsch et al., 2006). Currently, the wider-spread use of mixture modelling is hampered by the presence of structured noise artefacts (e.g. spontaneous networks of activity, stimulus correlated motion) that violate the distributional assumptions that need to be made. However, in the future FEAT will

incorporate an approach that combines the GLM with Independent Component modelling of this structured noise within a Bayesian framework, which should render the distributional assumptions valid (Makni et al., 2008b).

Priors can also be used to perform temporal regularisation. In Makni et al. (2008a) a temporal smoothness prior is used on the parameters describing the height of the haemodynamic response function (HRF) as it evolves over time. This is an attractive alternative to parameterised HRFs (e.g., Fig. 1) and basis functions (e.g., Fig. 4) as it makes no assumptions about the HRF shape other than that it is temporally smooth. Crucially, Bayes allows the amount of temporal smoothness to be adaptively determined from the data.

Bayesian versus frequentist inference

Until recently nearly all inference on functional neuroimaging data (e.g., PET, FMRI) was carried out in a frequentist framework; only in the last five years or so has Bayes started to be used either alongside or to replace frequentist inference. In classical frequentist statistics, probabilities refer to the frequency of outcomes. In contrast, Bayesians use probabilities to express degrees of belief. Care should be taken to not confuse P -values with Bayesian probabilities: Bayesian methods can be used to give the posterior probability that the null hypothesis is true, conditional on the data; classical P -values, in contrast, are the probability of the data², conditional on the null hypothesis. We can also use Bayes to infer about the absolute value of the parameters of interest. i.e., we may ask questions of our parameters such as, “What is the probability that θ lies in the interval $[\theta_0, \theta_1]$?”.

² Strictly speaking, classical P -values are the probability of a more extremal sample data than was observed.

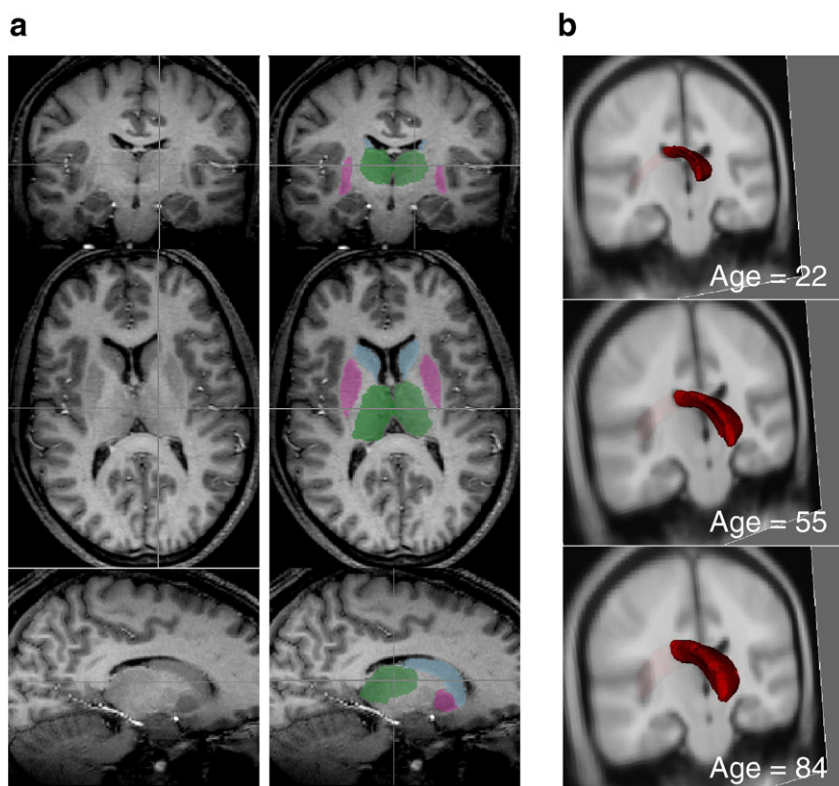


Fig. 5. FIRST is an FSL tool that can be used to segment subcortical structure. (a) Shows the segmentation of the left/right thalamus, caudate, and putamen from a T_1 -weighted image. A key component is the use of biophysical prior information about the shape and signal intensities for different sub-cortical structures. This is obtained from manually segmented training data. For example, (b) shows the mean shape of the left lateral ventricle given age, determined from a training set containing 87 subjects ranging from age 22 to 84.

A great advantage of Bayesian statistics is that it gives us a tool for inferring on *any* model we choose, and *guarantees* that uncertainty will be handled correctly. In contrast, only in certain special cases is it possible to derive analytical or asymptotic forms for the null distributions required by frequentist statistics. In their absence, frequentist solutions rely on null distributions derived from the data

(e.g., permutation tests), losing the statistical power gained from educated assumptions about, for example, the distribution of the noise.

The idea that Bayes gives us a mechanism for inferring on any model means that we can tackle problems for which there is no parametric frequentist solution. A good example of this is in the FSL tool FLAME (FMRIB's Linear Analysis of Mixed Effects), which is used

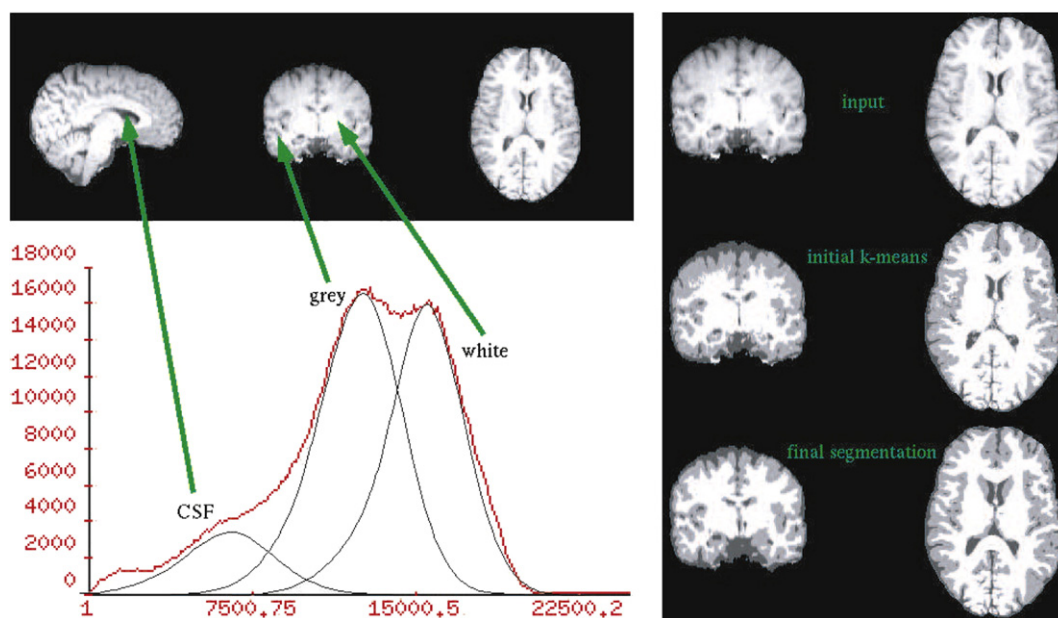


Fig. 6. Example of segmentation of structural MR images into tissue types using FAST. [Left] At the core of the approach is a generative model that predicts the data in the form of its histogram as being made up of a mixture of distributions where each tissue type is modelled as a Gaussian distribution. As part of the model there are also classification parameters that label which class each voxel belongs to. These parameters have a spatial regularisation prior on them to robustify the segmentation. [Right] Shows the MR data (top), the result of the K -means clustering used to initialise the Bayesian inference (middle), and the mode of the posterior classification parameter maps output from the Bayesian inference (bottom).

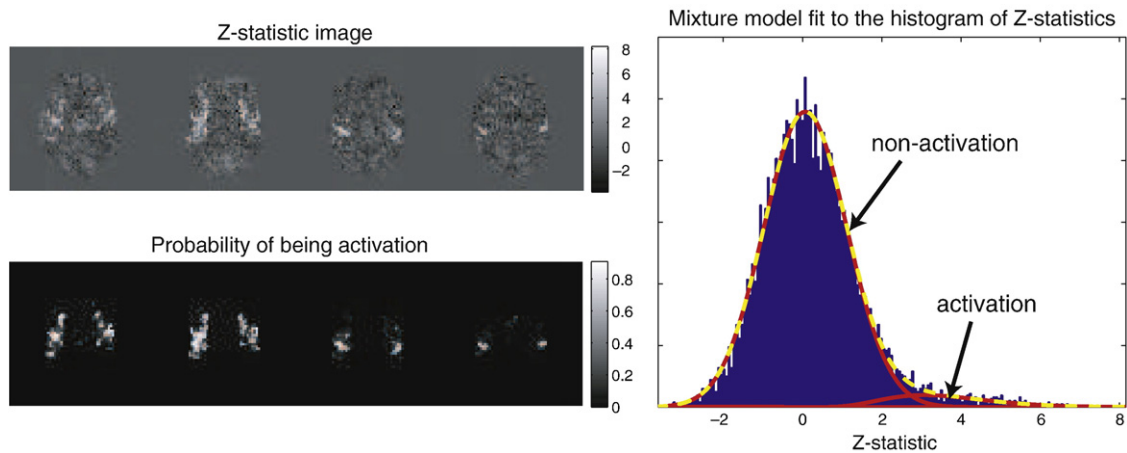


Fig. 7. Mixture modelling of a Z-statistic image. [Top-left] Four slices of a Z-statistic image obtained from fitting a GLM to the fMRI data for a single subject for a pain stimulus condition. [Right] Mixture model fit to the histogram of Z-statistics. Non-activating voxels are modelled as coming from a close-to-zero mean Gaussian distribution and activating voxels as coming from a Gamma distribution. [Bottom-left] Image showing the probability that a voxel is activating – this information can be used in thresholding to approximately control the true positive rate (TPR) as an alternative to null hypothesis testing.

to infer on multi-subject group studies in fMRI (Woolrich et al., 2004a). As illustrated in Fig. 8a, the majority of fMRI studies are multi-subject and/or multi-session where we make hypothesis tests at the group level. The human and computational costs involved in data analysis are relatively high, and so it is desirable to be able to make group-level inferences using the results of separate first-level (single session) analyses, where the first-levels are represented by *summary statistics* (Holmes and Friston, 1998). The problem is that it is not clear what or how summary statistics should be used to ensure that the inference at the group level is valid, particularly when variance components, such as the between-subject variance, are unknown.

The traditional approach in the frequentist literature is to use the regression parameter estimates from the first level as summary statistics (Holmes and Friston, 1998). However, using Bayes on the multi-level group model reveals that the summary statistics should include not only the regression parameter estimates from the first-level but also the variance of these regression parameter estimates. It also provides us with the means to use these summary statistics to infer at the group level in the face of unknown between-subject variance.

This Bayesian summary statistics approach used in FLAME has a number of distinct advantages over the frequentist-derived approach that only uses the regression parameter estimates from the first-level as summary statistics. Firstly, FLAME does not require balanced designs and so permits the analysis of fMRI data where the first-level design matrices have different structure from each other (e.g., contain behavioural scores as regressors). Secondly, as illustrated in Fig. 8b, FLAME can provide more accurate variance estimation and can also increase the ability to detect real activation.

Bayes and multiple comparisons

In null hypothesis testing of fMRI data there is typically one test carried out for each voxel in the brain. If the intention is to control the false positive rate across the whole brain then one should use a Bonferroni correction to adjust *P*-value thresholds by the number of tests carried out. The problem is that, due to the large number of voxels, these corrections are severe. What happens in the Bayesian setting with regards to multiple comparisons when we look to threshold marginal posterior probability maps of GLM regression parameters? In particular, there is often confusion about whether or not Bayes avoids the need for the multiple comparison corrections. The answer is the same as it is in the frequentist setting: it depends on what one wants to control when you threshold.

In the Bayesian setting, if one wants to label voxels as “active” by thresholding marginal posterior probability maps at a probability of

95% that the regression parameter is greater than zero, then in a completely non-activating brain, on average 5% of the voxels will be labelled as active. In this particular case, Bayes is no different to frequentist null hypothesis testing. That is, if one wants to control for wrongly labelling voxels across multiple tests then it is necessary to apply multiple comparison corrections on Bayesian posterior probability maps as well. This is perhaps not surprising given the numerical equivalence of *certain* posterior probabilities and *P*-values discussed in the last section.

Note that the need to use multiple comparison correction using Bayes is particular to when we are looking to threshold posterior probability maps of GLM regression parameters, but is not generally true (e.g. when the thresholding is explicitly contained within the model itself via classification parameters; Scott and Berger, 2006; Woolrich et al., 2005; Woolrich and Behrens, 2006).

Multi-modality fusion

In neuroimaging we can often obtain complementary information about some key underlying physiological phenomenon from more than one imaging modality. A key question is: how do we go about optimally combining this information? For example, if information extracted from a T1-weighted structural MR image tells us that a particular voxel is white matter, whereas a proton density MR structural tells us that the same voxel is grey matter, then what should we infer? The answer is to use Bayes.

As illustrated in Fig. 9, we start with a generative model that predicts how the T1-weighted MR structural and the proton density MR structural data should look like if we know what the tissue types are in all voxels. Bayesian inference will then infer on the parameters in the model in such a way as to probabilistically weight the information from the different modalities by implicitly assessing the quality and relevance of the different information available. In the case where we have two modalities, Bayes' rule gives us³:

$$\begin{aligned} p(\theta_1, \theta_2, \theta_{12} | \mathbf{Y}_1, \mathbf{Y}_2, \mathcal{M}) &\propto p(\mathbf{Y}_1 | \theta_1, \theta_{12}, \mathcal{M}) \\ p(\mathbf{Y}_2 | \theta_2, \theta_{12}, \mathcal{M}) p(\theta_1 | \mathcal{M}) p(\theta_2 | \mathcal{M}) p(\theta_{12} | \mathcal{M}) \end{aligned} \quad (6)$$

where \mathbf{Y}_1 and \mathbf{Y}_2 are the data from the first and second modalities respectively, θ_1 and θ_2 are the model parameters specific to the

³ Note that here we ignore the denominator of Bayes' rules, $p(\mathbf{Y} | \mathcal{M})$, as it does not depend on the model parameters. This is typically done, unless we are doing model selection or averaging.

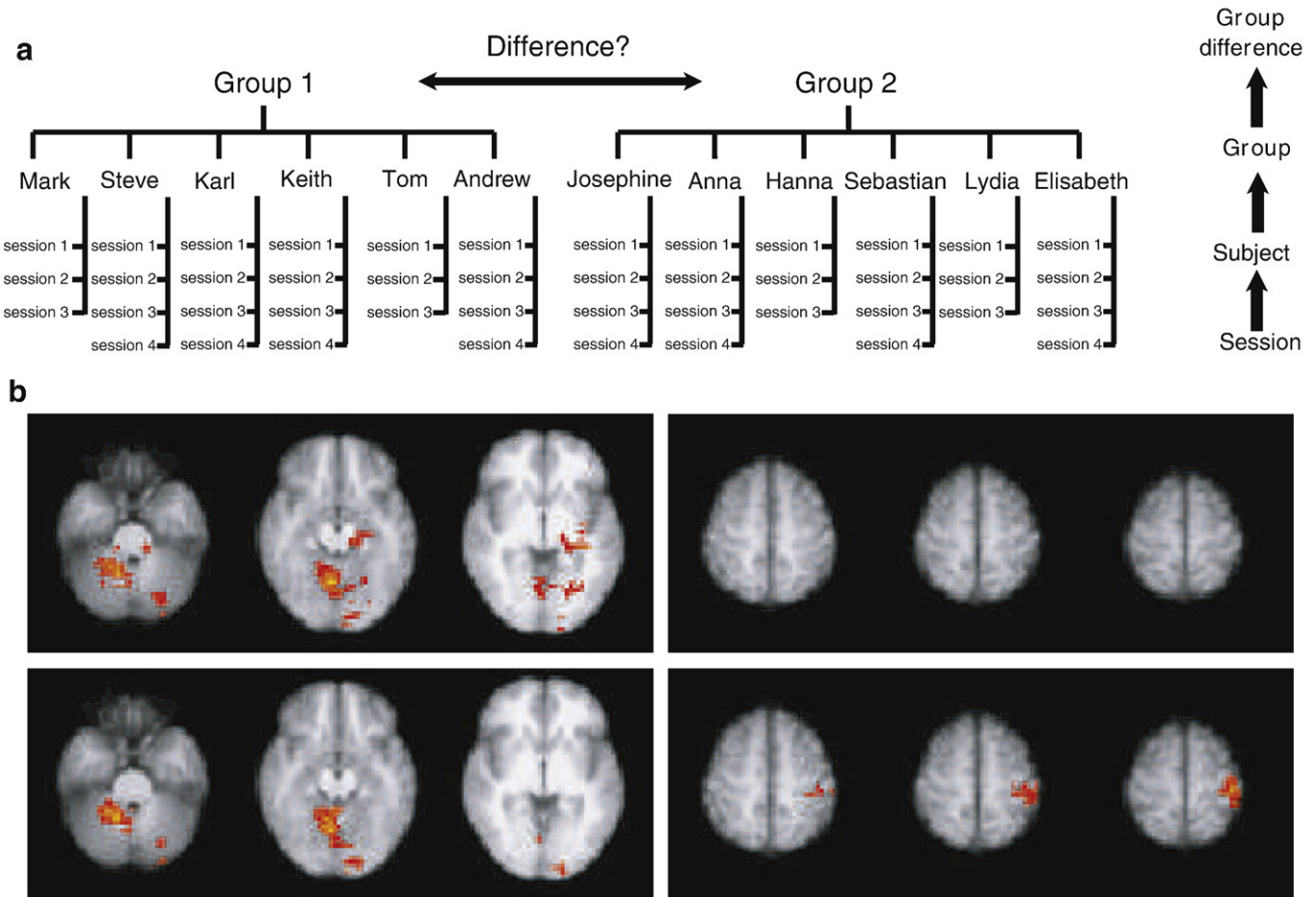


Fig. 8. (a) Hierarchical GLM for the analysis of group FMRI data. With a summary statistics approach the GLMs are estimated one level at a time and summary statistics are passed up to the next level of the hierarchy. (b) Thresholded group activation for two different group analysis approaches and for two different motor tasks. [top] A frequentist-derived approach that only uses first-level regression parameter estimates as summary statistics, and [bottom] the Bayesian-derived FLAME approach that augments the first-level regression parameter estimate summary statistics with their variances. [left] Index finger tapping vs rest. There is a general decrease in z-statistics in the FLAME approach, demonstrating the increased accuracy of the cross-subject random-effects variance estimation, due to the use of the first-level variances. [right] Sequential finger tapping vs index finger tapping. There is a general increase in z-statistics in the FLAME approach, demonstrating increased efficiency in parameter estimation by down-weighting subjects that have high first-level variance.

generative model for the first and second modalities respectively, and Θ_{12} are the generative model parameters common to both modalities. It is then often desirable to marginalise over the parameters specific to each modality to infer on the marginal posterior for the common parameters:

$$p(\Theta_{12} | \mathbf{Y}_1, \mathbf{Y}_2, \mathcal{M}) = \int \theta_1 \int \theta_2 p(\theta_1, \theta_2, \Theta_{12} | \mathbf{Y}_1, \mathbf{Y}_2, \mathcal{M}) d\theta_1 d\theta_2 \quad (7)$$

The structural segmentation tool in FSL, FAST, is a good example of a Bayesian tool that can deal with multi-modality data in this way, as illustrated in Fig. 9. The generative models for each modality are the same as they are in the single modality case, i.e., we use a mixture model (see Fig. 6), and priors are used to spatially regularise the classification. In this way FAST can allow for two or more input images, giving much improved results, for example, in the deep grey structures where T1-only segmentation often has problems due to the intermediate (between white and cortical grey) intensities of some subcortical grey matter.

Another example of multi-modality Bayes is the inference of haemodynamic changes from simultaneous BOLD and ASL FMRI data (Woolrich et al., 2006). As shown in Fig. 10, the traditional approach to dealing with such data is to separately analyse the BOLD–FMRI and ASL–FMRI, to infer percent BOLD and CBF changes respectively. However, this fails to acknowledge the presence of percent BOLD and

CBF changes (and also static magnetisation, M , changes) in *both* modes of data. In contrast, the multi-modality Bayesian approach results in increased sensitivity and reduced cross-contamination in the inference of the different haemodynamic changes.

Hierarchical models – combining local and global information

In the Bayes' rule section, we saw how spatial priors can be used to spatially regularise model parameters, in other words, how to model dependency between neighbouring imaging voxels. Bayesian techniques can also be used to allow data from *remote* voxels to depend on each other, via the use of *hierarchical priors*. Inference made on local parameters specific to one voxel, θ_i , may then be influenced by inference made in other voxels, if the parameters from these voxels share a *hierarchical prior*. For example, we can use Bayes' rule to get:

$$p(\theta^l, \theta^g | \mathbf{Y}, \mathcal{M}) \propto \prod_i p(\mathbf{Y}_i | \theta_i^l, \mathcal{M}) p(\theta^l | \theta^g, \mathcal{M}) p(\theta^g | \mathcal{M}) \quad (8)$$

where $p(\mathbf{Y}_i | \theta_i^l, \mathcal{M})$ is the voxelwise likelihood (generative model), and θ^g are the global parameters controlling the behaviour of the hierarchical prior, $p(\theta^l | \theta^g, \mathcal{M})$. The joint posterior can be marginalised to infer on either the local or global parameters.

An example of the use of a hierarchical model is in probabilistic global tractography using diffusion MRI data. Diffusion MRI is a

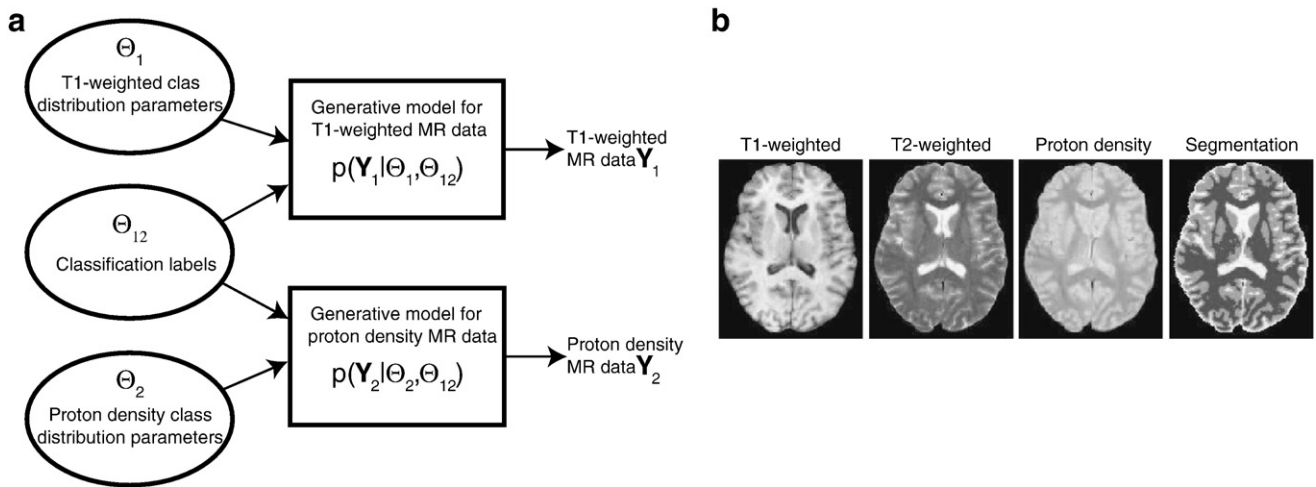


Fig. 9. (a) Generative model used to infer tissue type from two different modalities of structural MRI data in the FSL segmentation tool FAST. (b) Result of FAST segmentation run on three different modalities of structural MRI data. This gives improved results when compared with single modality segmentation, due to the probabilistic Bayesian weighting of the complementary information available in the different modalities.

modality that is sensitive to the diffusion of water molecules in the brain. The diffusion of water is influenced by the underlying microstructure, and in particular by the orientation of axonal fibres in brain white matter. Local (voxelwise) models of diffusion can be inferred on using Bayes to obtain PDFs on local fibre direction. Probabilistic tractography can then be used to track through these local fibre directions, to infer on anatomical white matter connections between two distant points (Behrens et al., 2003). A deficiency in this

approach is that any knowledge of the global structure is not used to improve the inference of the local information. As tracking proceeds, local errors are compounded, so any errors in local modelling have a dramatic effect on the global result (i.e. the inferred connection). This deficiency can be overcome by the use of a hierarchical model where a global connection between two brain regions act as hierarchical prior on voxels along the path of the connection (see Fig. 11) (Jbabdi et al., 2007). In this example, the local parameters, Θ^l , include the voxelwise

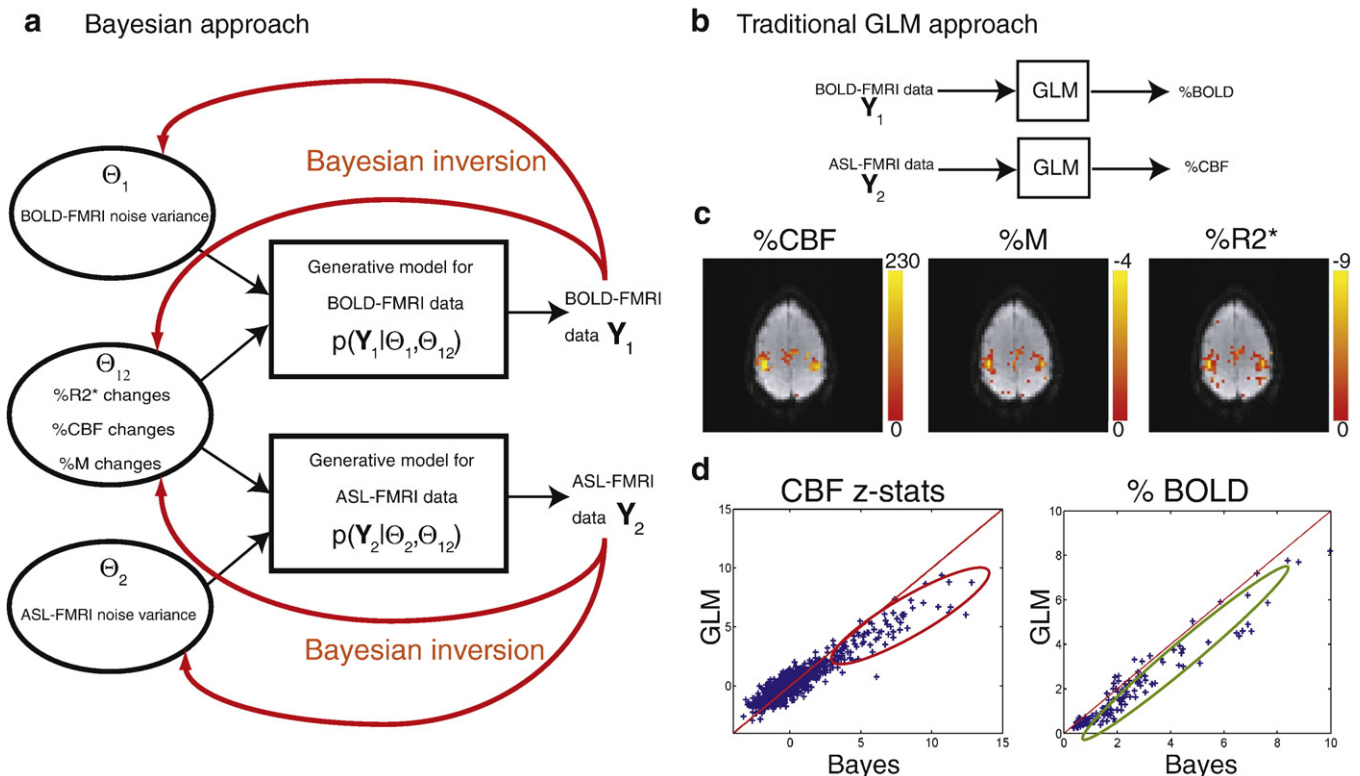


Fig. 10. (a) Bayesian approach to inferring on multi-modality BOLD and ASL FMRI data. The biophysical generative model predicts what the two data modalities look like given we know the haemodynamic changes. Bayesian “inversion” (of the generative, or forward, model) is then used to probabilistically infer all of the haemodynamic changes. This is in contrast to a traditional approach (b) where the BOLD and ASL-FMRI are analysed using separate GLMs to infer percent BOLD and CBF changes respectively. The Bayesian approach uses information from both modalities to infer all the haemodynamic changes, as shown in (c) for a subject doing a finger tapping task. The main benefits of the Bayesian approach is increased sensitivity in inferring CBF (red ellipse), and the correction of an under-estimation of the percent BOLD by accounting for the percent M changes (M is the static magnetisation in a voxel and is a function of blood flow and volume changes) (green ellipse).

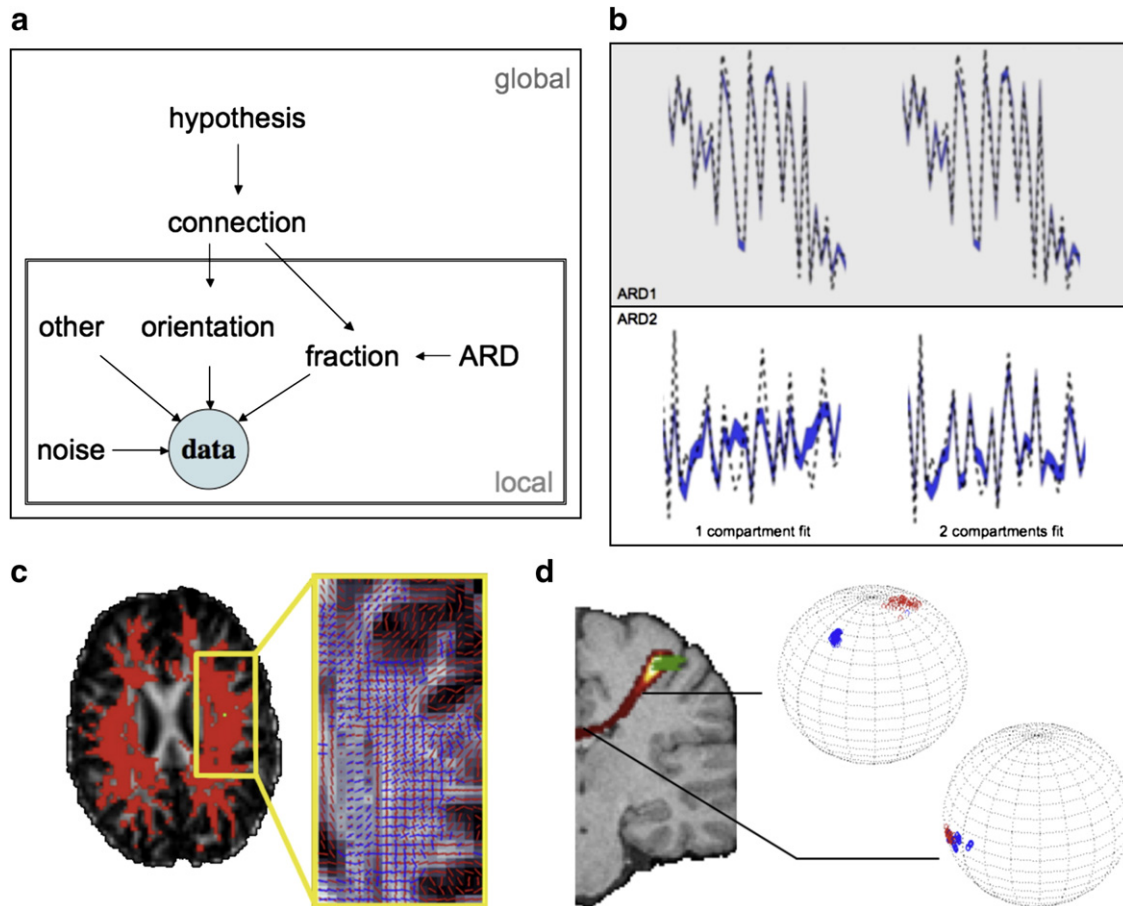


Fig. 11. (a) Schematic of a generative model for both a local and global Bayesian model for diffusion MRI data. The automatic relevance determination (ARD) prior relates to the relative contribution of each compartment (fibre bundle) to the signal. Global hierarchical priors influence the local parameters that include the orientations and volume fractions of fibre bundles at each voxel. The connection is driven by a hypothesis on the existence or not of a connection, which allows us to test for the existence of a connection using model comparison. (b) Real data (dotted line) and fitted signal (with one standard deviation calculated using the posterior PDFs) for two voxels where the ARD prior supported the existence of one (top) vs two compartments (bottom). Note that while one compartment fit explains the signal quite well on the top row, two compartments are required to explain the data on the bottom row. The ARD prior allows us to choose between the two models in an objective way. (c) Distribution of voxels (red) for which the ARD prior supported the existence of two compartments. The box shows the local fibre orientations for the two populations (compartments). (d) Posterior distribution of the local fibre orientations using local (red) and global (blue) modelling. Note that for a voxel supporting one fibre population, the two posteriors are similar, while in a voxel supporting a mixture of fibre populations, the global model switches the orientation distribution to the population supporting the hypothesised connection.

fibre bundle volume fractions and orientations; the global parameters, Θ^g , include parameters describing the path that the connection takes; and the hierarchical prior, $p(\theta^l | \Theta^g, \mathcal{M})$, includes a description of the constraint on the voxelwise fibre bundle orientations given the overall path of the connection.

The effect of this kind of hierarchical prior is that errors are not compounded along the pathway. Instead information from the entire pathway is considered simultaneously so that the pathway is unlikely to be diverted by a single erroneous measurement. Importantly, the subset of voxels on which the spatial prior acts (i.e., the voxels along the path of the connection) is also inferred from the data. In other words, inference is made both on the local scale (voxels) and global scale (connections). This method was termed *constrained tractography*, as it allows us to infer on connections for which the end points are imposed as priors. This constraint influences the posterior distribution on voxel-wise parameters along the course of a connection, increasing the robustness of the local estimates of, e.g., tissue orientations.

Model selection

Bayesian data analysis is the process of fitting a probabilistic model to a set of data, and encompasses the following three main steps: (i) setting up a full probabilistic model, including a data generative model

(likelihood function) and a set of priors on model parameters; (ii) conditioning on observed data (posterior distribution); and (iii) evaluating the performance of the model. The last step is one of the strengths of Bayesian techniques. Given that any model used within a Bayesian framework requires making a certain number of assumptions, it is somewhat reassuring that the same framework contains its own procedure for evaluating the assumptions that have been made against actual data.

In order to evaluate a model using Bayes, the quantity that needs to be evaluated is the (up to this point ignored) denominator of the posterior PDF given in Eq. (3), i.e., the probability of the data given the model⁴. This quantity, often termed the *marginal likelihood*, or *model evidence*, accounts for both *accuracy* (data fit) and *complexity* of the model given the observed data. Model selection consists of calculating this quantity for a given number of models, and selecting the model with highest marginal likelihood. Commonly two models are compared by computing the Bayes Factor, which is the ratio of the evidences for the two competing models. However, it is worth noting

⁴ Although a pure Bayesian would actually calculate the posterior distribution $p(\mathcal{M} | \mathbf{Y}) \propto p(\mathbf{Y} | \mathcal{M}) p(\mathcal{M})$ provided an appropriate choice for the priors $p(\mathcal{M})$. Typically, it is implicitly assumed that all models are a priori equally likely, i.e., $p(\mathcal{M})$ is equal for all models. $p(\mathcal{M} | \mathbf{Y})$ is then proportional to the evidence, $p(\mathbf{Y} | \mathcal{M})$.

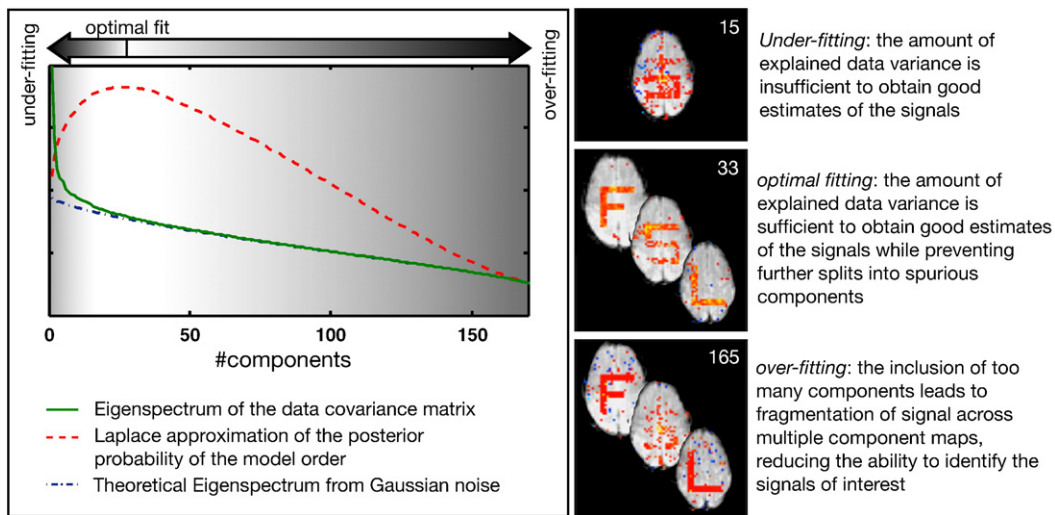


Fig. 12. Selecting the model order in ICA-based analysis of fMRI data: the observed Eigenspectrum (green) – obtained from a Singular Value Decomposition of the data covariance matrix – is compared to the expected Eigenspectrum from white Gaussian noise (blue). Using a Laplace approximation to the posterior probability of the model order, one defines the model order as the dimensionality with highest probability. From this point onwards the observed Eigenspectrum is not sufficiently different from an Eigenspectrum obtained from random noise. Reducing the number of components to below the optimal model order can lead to under-fitting, i.e., poor estimation of some of the underlying signals. Estimating too many components, on the other hand, can lead to over-fitting, where some signals split and get represented across multiple component maps (Beckmann and Smith, 2004).

that computing the model evidences, or the Bayes Factors, is not generally a trivial task.

Model selection has been used in many occasions in the context of neuroimaging, for example in the selection of the number of components in probabilistic Independent Component Analysis (ICA) of fMRI data used in the FSL tool MELODIC. ICA decomposes the data into (statistically) independent spatial maps (and associated time courses), which can represent neuronal activity without using a temporal model of the underlying neural activity as traditional GLM approaches do, and can also identify artefacts in the data (e.g., stimulus-correlated motion and spontaneous networks of neural activity).

As shown in Fig. 12, a crucial variable in the approach is how many components there are in the data. If too many or too few are assumed, then components get erroneously split or merged respectively and the results are hard to interpret. The solution is to use Bayesian model selection to adaptively determine the number of components from the data itself. Recently, an approach that combines ICA modelling with a traditional GLM approach has been proposed (Makni et al., 2008b). This has the potential to perform model averaging (rather than model selection) on the number of components using a technique called Automatic Relevance Determination (ARD). Model averaging and ARD will be described in more detail in the next section.

Model selection has also been used in a recent adaptation of multi-subject fMRI analysis, to allow for probabilistic outlier inference (Woolrich, 2008). Group studies can include outlier subjects whose effect sizes are completely at odds with the general population for reasons that are not of experimental interest. If ignored, these outliers can dramatically affect the inference results. As shown in Fig. 13, the outlier modelling assumes that the population distribution is a mixture of two Gaussians, one for the normal population and one for the outlier population (Woolrich, 2008). However, we do not want to automatically assume that there is an outlier population, as this may cause us to erroneously infer outliers when there is insufficient evidence that they exist. The solution is to fit both a traditional single Gaussian (no outlier) model and a two Gaussian (outlier) model, and then select the model with the highest Bayesian evidence. This approach has been incorporated into the FEAT tool FLAME.

Other applications of model selection includes the comparison of models of the structure of functional brain networks (e.g., Dynamic

Causal Modelling; Penny et al., 2004), or anatomical brain networks (Jbabdi et al., 2007). This last application uses the framework of global tractography. Two models are proposed, one which enforces an anatomical connection between two candidate brain regions, and one which enforces the absence of a connection. After model fitting, the evidence for each model is compared. This technique allows, for the first time, statistical inferences to be drawn about anatomical connections using diffusion imaging.

Model averaging

When performing Bayesian model selection, inference on the parameters is made for each model separately, and the evidences may subsequently be compared to select the appropriate model. An alternative method is Bayesian model averaging, where inference is derived from a committee of models. Posterior PDFs on model parameters are then a weighted average of the posteriors given each model, where the weights are the respective evidences of each model:

$$p(\theta|\mathbf{Y}, \{\mathcal{M}_1, \dots, \mathcal{M}_n\}) = \sum_{k=1}^n p(\theta|\mathbf{Y}, \mathcal{M}_k) p(\mathcal{M}_k|\mathbf{Y}). \quad (9)$$

While MCMC approaches that jump between models of different dimensionality, such as reversible jump MCMC (Green, 1995), can be used, the computational complexity of inferring using model averaging is often prohibitive. However, alternative approximate approaches do exist. In particular, we can use shrinkage priors, also known as Automatic Relevance Determination (ARD) (MacKay, 1995). ARD allows for averaging between models with and without particular model parameters. In its most straightforward form ARD consists of placing a Gaussian prior, with zero mean and unknown variance, on any model parameter in question, e.g., $P(\theta|\sigma) \sim N(0, \sigma^2)$. The key is that the variance, σ^2 , is probabilistically inferred alongside the parameter, θ , such that if $\sigma^2 \rightarrow 0$ then $\theta \rightarrow 0$ and the parameter is knocked out of the model, otherwise $\sigma^2 > 0$ and θ can be inferred as non-zero.

An example of the use of an ARD prior is in the modelling of crossing fibres in the local diffusion model of the FSL Diffusion Toolbox (FDT). Within any particular voxel there may be one, two or more fibre bundles passing through. Diffusion MRI data are modelled as a mixture of data generative processes, corresponding to different fibre bundles within a voxel (Behrens et al., 2007). An ARD prior on the

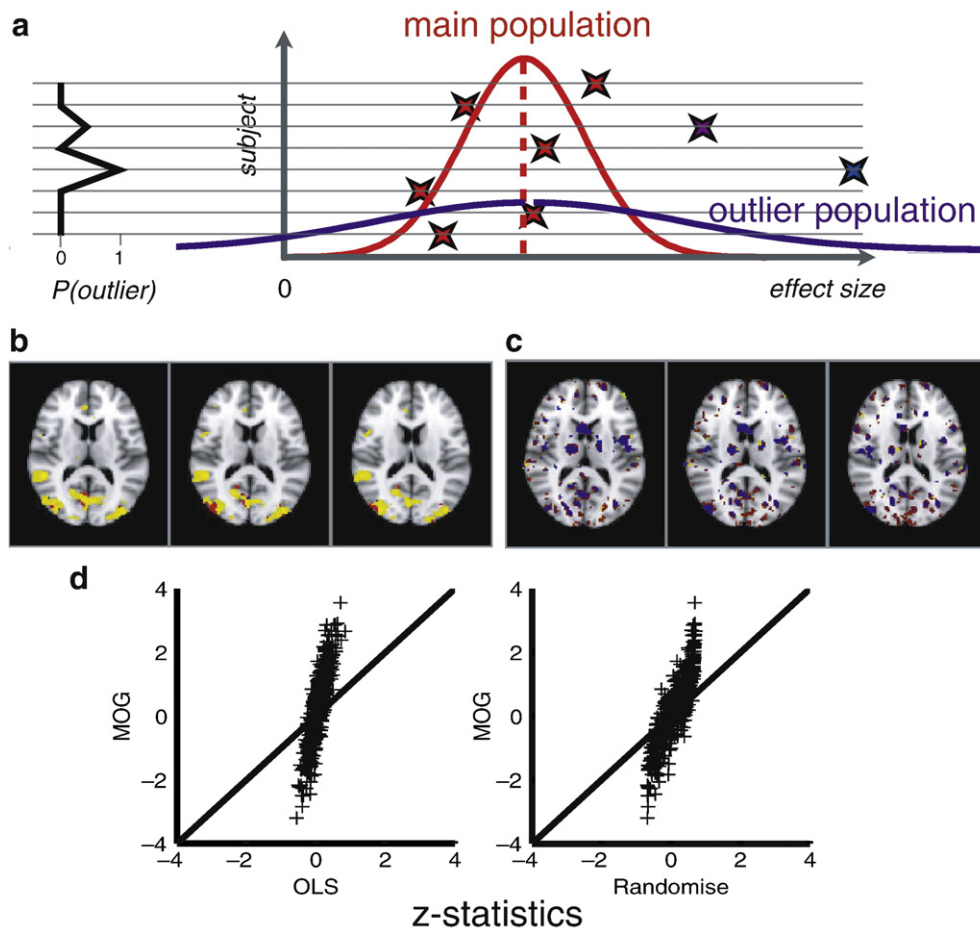


Fig. 13. Probabilistic Bayesian outlier inference in multi-subject fMRI analysis. (a) The generative model models the population distribution as a Mixture of Gaussians (MOG). In this example the fourth subject (from the bottom) is exhibiting strong outlier behaviour, and the sixth subject is an outlier with probability of approximately 0.5. (b) and (c) relate to inferring the group mean effect of the activation in the feedback period of an fMRI decision-making experiment. (b) Shows voxels passing threshold for the outlier approach only (red), for a traditional non-outlier approach (blue) and for both (yellow). (c) Shows the posterior maps (using the Bayesian outlier approach) that there are approximately one (red voxels) or two (blue voxels) or more (yellow voxels) outlier subjects. (d) Comparison of z-statistics between the outlier approach (MOG), a non-outlier approach (OLS, i.e. Ordinary Least Squares), and a permutation testing approach (Randomise) for data exhibiting outlier behaviour. The z-statistic differences indicate that while permutation testing can protect the false positive rate validity, the approach is less sensitive when compared with the MOG outlier approach (Woolrich, 2008).

relative contribution of each fibre bundle allows robust estimation of the number of crossing fibres within a voxel (Fig. 11). This is crucial information for successful diffusion-based tractography.

Bayesian inference techniques

Throughout this article we have considered some of the key concepts in Bayes and how they can be used to our advantage in performing neuroimage analysis. However, we have largely ignored one of the key difficulties in using Bayes: that is that the solution to Bayes' equation is seldom straightforward. In particular, the integrals in Eqs. (3) and (4) are not in general analytically tractable (i.e., we cannot solve them mathematically). As illustrated in Fig. 14, we briefly consider here some of the options available. We separate these into approaches which solve the integrals numerically and those that use parametric approximations to the posterior.

Numerical integration

One solution is to calculate the numerator in Eq. (1) at every point in a grid defined in the model parameter space. We refer to this as a *full evaluation* approach. This can be a very straightforward and accurate technique in models with low numbers of parameters. However, it is

difficult to define efficient grids for models with high numbers of parameters and can easily become prohibitively time-consuming.

Another solution is to draw samples in parameter space from the joint posterior distribution, implicitly performing the integrals numerically. For example, we may repetitively choose random sets of parameter values and choose to accept or reject these *samples* according to a criterion based on the value of the numerator in Eq. (1). Examples of schemes such as this are *rejection sampling* and *importance sampling* (Gamerman, 1997). However, these kinds of sampling schemes tend to be very slow, particularly in high dimensional parameter spaces, as samples are proposed at random, and thus each has a very small chance of being accepted.

Markov chain Monte Carlo (MCMC) (Gilks et al., 1996) is a sampling technique which addresses the problem of having a large number of parameters by proposing samples preferentially in areas of high probability. Samples drawn from the posterior are no longer independent of one another, but the high probability of accepting samples allows for many samples to be drawn and, in many cases, for the posterior PDF to be built in a relatively short period of time (compared with other numerical integration approaches). While MCMC is a powerful technique that can be used on a wide variety of models, it is time consuming when compared with posterior approximation approaches.

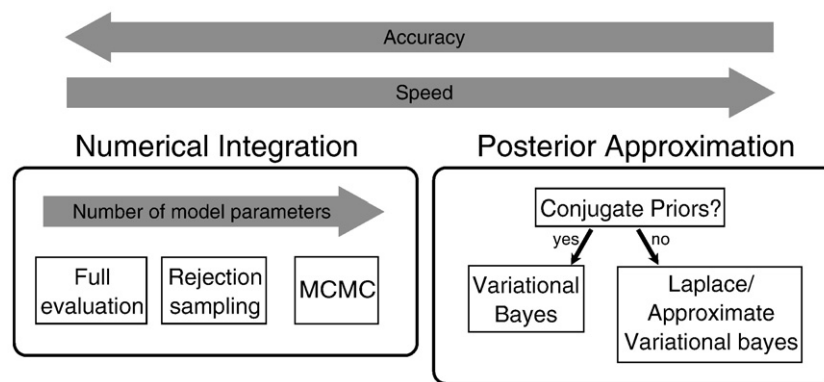


Fig. 14. The use of Bayes requires integrations to be performed that are seldom tractable, in which case there are broadly speaking two separate approaches used: (1) solve the integrals numerically, or (2) make approximations to the posterior distribution.

Posterior approximations

The intention here is to make approximations such that we make the integrals analytically tractable. These approaches tend to be less accurate but computationally faster than numerical integration, and typically require more work to be done mathematically rather than computationally. A key issue is whether or not the model has *conjugate priors*. Conjugacy exists for a model if the posteriors have the same distributional form as the prior, e.g., if the prior is Gaussian then so must the posterior be. If this is the case, then a technique known as *Variational Bayes* can be employed.

Variational Bayes approximates the true posterior distribution by estimating it using a posterior factorised over subsets of the model parameters. Minimising the difference between the true and approximate distributions results in tractable integrals when there are conjugate priors (Attias, 2000; Penny et al., 2003). However, models are often not conjugate, for example when we have nonlinear generative models with additive Gaussian noise (e.g., Figs. 1 and 10), and so other approximations are needed. One option is what we refer to as *Approximate Variational Bayes*, which works by approximating the posterior for non-conjugate parameters as being multivariate Gaussian. This can be achieved by applying first- or second-order Taylor expansions of the problem terms in Bayes' rule. The problem can then be solved using Variational Bayes in the normal way (Penny et al., 2004; Friston et al., 2007; Woolrich and Behrens, 2006; Chappell et al., 2008).

Non-Bayesian approaches

We have described the advantages in the Bayesian approach of estimating the full posterior distribution of model parameters of interest, as opposed to just obtaining a *point estimate*, or “best guess”, of each model parameter, such as one achieves with *maximum likelihood* approaches. However, there are many scenarios in the analysis of neuroimaging data where the Bayesian approach is impractical. Typically, this occurs when the model and the interaction between model parameters become too complex for posterior approximation approaches to be sufficiently accurate; and when the likelihood equation is so complex that numerical integration approaches become too slow. In these cases the only viable approach is to revert to point estimation approaches. A clear example of this is the nonlinear registration tool FNIRT (FMRIB's nonlinear image registration tool; Andersson et al., 2007). In FNIRT there are many thousands of model parameters describing the warp field mapping one subject to another. The idea of characterising the posterior distribution of these many thousands of parameters, which are highly dependent on each other, is not at all possible with current computational resources.

Furthermore, we may even have an analysis scenario that is sufficiently complex (or be considering an analysis strategy that is fundamentally ad hoc), that we do not even consider it useful to

explicitly formulate a likelihood equation. For example, BET (brain extraction tool; Smith, 2002) deforms a tessellated mesh model in order to estimate the exterior brain surface. The various procedures that have been developed to achieve this generally work accurately and robustly, but are a combination of different practical heuristics that interact together to optimise what could be seen as a very high-dimensional problem; at no point is a single overall likelihood equation formulated.

Other examples of tools in FSL where Bayes has been considered impractical up to now are TBSS (tract-based spatial statistics, for carrying out voxelwise multi-subject analysis of diffusion data; Smith et al., 2006), FLIRT (FMRIB's linear image registration tool; Jenkinson and Smith, 2001), PRELUDE/FUGUE (phase unwarping and MRI unwarping; Jenkinson, 2003), and the “unmixing” algorithm at the heart of the MELODIC ICA-based approach for model-free fMRI analysis (Beckmann and Smith, 2004, 2005).

Conclusion

In this article we have described the impact that Bayes has had on neuroimaging in the research experiences of the authors. In particular we have considered Bayesian concepts such as biophysical and regularisation priors, hierarchical modelling, and model selection and averaging. Bayes provides us with a framework which (computational practicalities aside) allows us to infer on any generative model that we can propose. As computational resources increase then so will the quality of information that we will be able to extract using Bayes, by the use of more complex, flexible and adaptive models.

Role of the funding source

Mark Woolrich was funded by the UK EPSRC, Saad Jbabdi by the Hadwen Trust, Michael Chappell by GlaxoSmithKline, Salima Makni by the UK EPSRC, Timothy Behrens by the UK MRC, and Mark Jenkinson by the UK BBSRC.

Conflict of interest statement

The authors declare that they have no conflict of interest relative to this research.

References

- Andersson, J., Jenkinson, M., and Smith, S. (2007). Non-linear registration, aka spatial normalisation. Technical Report TR07J2, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, Oxford University, Oxford, UK. Available at <http://www.fmrib.ox.ac.uk/analysis/techrep> for downloading.
- Attias, H., 2000. A Variational Bayesian framework for graphical models. Adv. Neur. Info. Proc. Sys. 12.
- Bartsch, A., Homola, G., Biller, A., Solymosi, L., Bendszus, M., 2006. Diagnostic functional MRI: illustrated clinical applications and decision-making. J. Magn. Reson. Imaging 23, 921–932.

- Beckmann, C., Smith, S., 2005. Tensorial extensions of independent component analysis for multisubject fMRI analysis. *Neuroimage* 25, 294–311.
- Beckmann, C., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. on Medical Imaging* 23 (2), 137–152.
- Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., Matthews, P., Brady, J., Smith, S., 2003. Characterisation and propagation of uncertainty in diffusion weighted MR imaging. *Magn. Reson. Med.* 50, 1077–1088.
- Behrens, T.E.J., Berg, H.J., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W., 2007. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* 34 (1), 144–155.
- Chappell, M., Grove, A., Whitcher, B., and Woolrich, M. (2008). Variational Bayesian inference for a non-linear forward model. *IEEE Sig Proc. In Submission*.
- Cox, R., 1946. Probability, frequency and reasonable expectation. *Am. J. Phys.* 14, 1–13.
- Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. *Neuroimage* 34, 220–234.
- Gamerman, D., 1997. Markov Chain Monte Carlo. Chapman and Hall, London.
- Gilks, W., Richardson, S., Spiegelhalter, D., 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall, London.
- Green, P., 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711–732.
- Holmes, A., Friston, K., 1998. Generalisability, random effects and population inference. Fourth Int. Conf. on Functional Mapping of the Human Brain, *NeuroImage*, volume 7, p. S754.
- Jbabdi, S., Woolrich, M., Andersson, J., Behrens, T., 2007. A Bayesian framework for global tractography. *Neuroimage* 37 (1), 116–129.
- Jenkinson, M., 2003. A fast, automated, n -dimensional phase unwrapping algorithm. *Magn. Reson. Med.* 49 (1), 193–197.
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5 (2), 143–156.
- Mackay, D., 1995. Developments in probabilistic modelling with neural networks — ensemble learning. *Proceedings of the third Annual Symposium on Neural Networks*. Springer, Nijmegen, Netherlands, pp. 191–198.
- Makni, S., Beckmann, C., Smith, S., Woolrich, M., 2008a. Bayesian deconvolution of fMRI data using bilinear dynamical system. *NeuroImage* Accepted.
- Makni, S., Beckmann, C., Smith, S., and Woolrich, M. (2008b). Combining ICA and GLM for fMRI data analysis. In Fourteenth Int. Conf. on Functional Mapping of the Human Brain.
- Patenaude, B., Smith, S., Kennedy, D., and Jenkinson, M. (2007). FIRST — FMRIB's integrated registration and segmentation tool. In Thirteenth Int. Conf. on Functional Mapping of the Human Brain.
- Penny, W., Kiebel, S., Friston, K., 2003. Variational Bayesian inference for fMRI time series. *NeuroImage* 19 (3), 1477–1491.
- Penny, W., Stephan, K., Mechelli, A., Friston, K., 2004. Comparing dynamic causal models. *Neuroimage* 22, 1157–1172.
- Scott, J., Berger, J., 2006. An exploration of aspects of Bayesian multiple testing. *Statistical Science*.
- Smith, S., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17 (3), 143–155.
- Smith, S., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T., Mackay, C., Watkins, K., Ciccarelli, O., Cader, M., Matthews, P., Behrens, T., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 31, 1487–1505.
- Woolrich, M., 2008. Robust group analysis using outlier inference. *NeuroImage* 41, 286–301.
- Woolrich, M., Behrens, T., 2006. Variational bayes inference of spatial mixture models for segmentation. *IEEE Trans. on Medical Imaging* 25 (10), 1380–1391.
- Woolrich, M., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S., 2004a. Multi-level linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage* 21 (4), 1732–1747.
- Woolrich, M., Behrens, T., Smith, S., 2004b. Constrained linear basis sets for HRF modelling using Variational Bayes. *NeuroImage* 21 (4), 1748–1761.
- Woolrich, M., Behrens, T., Beckmann, C., Smith, S., 2005. Mixture models with adaptive spatial regularisation for segmentation with an application to fMRI data. *IEEE Trans. on Medical Imaging* 24 (1), 1–11.
- Woolrich, M., Chiarelli, P., Gallichan, D., Perthen, J., Liu, T., 2006. Bayesian inference of haemodynamic changes in functional ASL data. *Magn. Reson. Med.* 56, 891–906.