Combined spatial and non-spatial prior for inference on MRI time-series

Adrian R. Groves *, Michael A. Chappell, Mark W. Woolrich

FMRIB Centre, Department of Clinical Neurology, John Radcliffe Hospital, Oxford, OX3 9DU, UK

ARTICLE INFO

ABSTRACT

When modelling FMRI and other MRI time-series data, a Bayesian approach based on adaptive spatial smoothness priors is a compelling alternative to using a standard generalized linear model (GLM) on presmoothed data. Another benefit of the Bayesian approach is that biophysical prior information can be incorporated in a principled manner; however, this requirement for a fixed non-spatial prior on a parameter would normally preclude using spatial regularization on that same parameter. We have developed a Gaussian-process-based prior to apply adaptive spatial regularization while still ensuring that the fixed biophysical prior is correctly applied on each voxel. A parameterized covariance matrix provides separate control over the variance (the diagonal elements) and the between-voxel correlation (due to off-diagonal elements). Analysis proceeds using evidence optimization (EO), with variational Bayes (VB) updates used for some parameters. The method can also be applied to non-linear forward models by using a linear Taylor expansion centred on the latest parameter estimates. Applying the method to FMRI with a constrained haemodynamic response function (HRF) shape model shows improved fits in simulations, compared to using either the non-spatial or spatial-smoothness prior alone. We also analyse multi-inversion arterial spin labelling data using a non-linear perfusion model to estimate cerebral blood flow and bolus arrival time. By combining both types of prior information, this new prior performs consistently well across a wider range of situations than either prior alone, and provides better estimates when both types of prior information are relevant.

© 2008 Elsevier Inc. All rights reserved.

Introduction

Hierarchical Bayesian methods provide a flexible framework for the analysis of functional MRI and other MRI time-series data (Penny et al., 2003; Woolrich et al., 2006). In particular, the use of priors on signal parameters provides a principled approach to incorporating prior physical information into statistical inference (Friston, 2002). In this paper the "signal parameters" are analogous to the regression coefficients in (Penny et al., 2005), and we wish to infer their true value at each voxel in the brain.

We consider two types of prior information in this paper: fixed non-spatial priors and spatial priors. Fixed non-spatial priors provide information on the plausible range of values a signal parameter could have, often based on our prior biophysical understanding of the underlying process. Since these are informative priors based on real *a priori* knowledge, it is important that they be included in the model to ensure that only sensible interpretations of the data are considered. They are not to be confused with global shrinkage priors, which provide regularization by automatically inferring the scale of each signal parameter from the data.

Another type of prior, the spatial smoothness prior, explicitly encodes the belief that the parameter values in one voxel should not be dramatically different from those in its neighbours. It thereby provides spatial regularization which is similar to spatial presmoothing, but it is adaptive on each signal parameter separately (Penny et al., 2005). If there is not enough information in the data to justify additional spatial detail, then a simpler, smoother parameter image is produced.

We are interested in applications in which a fixed biophysical prior provides useful information during inference, but spatial regularization is also desirable. There are many approaches in the literature for combining different types of prior within a Bayesian framework. These include many examples in neuroimaging applications:

- Model averaging, in which several independent models are evaluated and the posteriors are averaged based on global model evidence (Trujillo-Barreto et al., 2004). In the event that one model has much higher evidence than the others, it will dominate the posterior solution; otherwise the result is a weighted average of the separate results.
- Mixture models, in which the prior's covariance matrix is the weighted sum of several different types of covariance matrix, each with adaptively determined parameters (Mattout et al., 2006). For example, a signal parameter might have two sources of variance: a spatially-smooth component and an additive non-spatial

* Corresponding author. Fax: +44 1865 222729.
E-mail addresses: adriang@fmrib.ox.ac.uk (A.R. Groves), chappell@fmrib.ox.ac.uk (M.A. Chappell), woolrich@fmrib.ox.ac.uk (M.W. Woolrich).

component. The variance of each component would be determined from the data.

- Hierarchical models, in which a hierarchy of signal parameters is used, with each level having a generative model to predict the level below it, and each level introducing its own type of additive noise. This approach has been successfully used for the analysis of MEG data (Trujillo-Barreto et al., 2008; Sato et al., 2004), for example by having spatial regression coefficients, which predict the noisy spatio-temporal electrical activity, which predicts the noisy MEG data. The overall variance at the lowest level is effectively a sum of different types of structured covariance matrix, with the weights determined by the magnitude of the "noise" added at each level.

Notably, none of these attempt to incorporate fixed biophysical prior information as part of the inference. Instead, they use adaptive spatial priors in addition to adaptive non-spatial priors as a way of modelling different types of variance in the observed signal.

Our objective is different; we would like to ensure that a fixed biophysical prior is respected, by conditioning the inference on the prior knowledge that the signal parameters must remain near a physically-plausible range of values. In technical terms, we want to ensure that the marginal prior distribution in each voxel always corresponds to the fixed biophysical prior.

One possible approach to this would be to use a wavelet decomposition that describes the signal parameters using sparse spatial basis functions Flandin and Penny (2007). In principle, different priors could be applied at each level of the decomposition. By applying the biophysical prior to the coarsest level instead of ARD, it may be possible to achieve approximate enforcement of the biophysical prior. However, to the best of our knowledge this has not yet been attempted.

Our approach is based on Gaussian process priors (GPPs), which provide a flexible approach to spatial regularization. In particular, we use the same GPP form previously used in neuroimaging for analysing stimulus–response functions in neural recordings (Sahani and Linden, 2003) and for spatiotemporal analysis using functional distance metrics (Bowman, 2007). In these papers it is used to provide both spatial regularization and a global shrinkage prior. However, by fixing the global shrinkage parameter we can instead use the GPP to enforce the informative biophysical prior while also providing adaptive spatial regularization. The covariance matrix is parameterized to provide separate control over the variance (the diagonal elements) and the between-voxel correlation (due to off-diagonal elements). The mean and variance are set to match the fixed non-spatial prior, while the correlation between voxels drops exponentially with distance and is controlled by an adaptive correlation–distance parameter. This provides adaptive spatial regularization while ensuring the non-spatial prior still applies to each voxel.

This combined prior can be used as a drop-in replacement for the spatial smoothness prior in the hierarchical FMRI model developed by Penny et al. (2005). That paper used variational Bayes (VB) for inference, which is an iterative approach that provides fully probabilistic results. We use a hybrid approach for inference on the combined prior: VB estimates are used for inference on the noise parameters, while an empirical Bayes technique called evidence optimization (EO) is used to estimate the Gaussian process hyperparameters. EO also provides probabilistic estimates of the signal parameters, which are equivalent to the regression coefficients in a generalized linear model (GLM).

We also extend this technique to work with non-linear modelling problems. Non-linear time-series models are beneficial in neuroimaging applications such as multi-contrast FMRI (Woolrich et al., 2006). This is particularly useful because parameters in non-linear models usually have real biophysical meanings, and as a result often have informative priors associated with them. To achieve this we adapt our

previously published approaches for using VB with non-linear time-series models (Chappell et al., 2009).

We evaluate our technique using two applications in brain MRI. Both use non-linear forward models and have informative non-spatial priors on some parameters. In each application, we compare using the combined prior against two alternatives: using the non-spatial information only, or using the spatial smoothness prior that ignores the informative non-spatial prior.

The first of these applications is FMRI analysis using constrained linear basis sets. The non-spatial prior allows the haemodynamic response function (HRF) to undergo reasonable variations in its shape in order to fit the data, while avoiding unrealistic HRF shapes (Woolrich et al., 2004). The second application is to multi-inversion-time ASL, where we estimate blood flow and bolus arrival time from resting-state data by fitting an intrinsically non-linear model of cerebral perfusion.

Theory

Voxelwise models of MRI time-series

Four-dimensional MRI data is usually analysed as a collection of V separate time series (one for each brain voxel). A generative model $g(\cdot)$ predicts the $T \times 1$ vector of time-series data $\boldsymbol{y}_v$ in voxel v given a $K \times 1$ vector of parameter values $\boldsymbol{w}_v$ in that voxel:

$$\boldsymbol{y}_v = \boldsymbol{g}(\boldsymbol{w}_v) + \boldsymbol{e}_v \tag{1}$$

where $\boldsymbol{e}_v$ is the additive noise in that voxel. The forward model $\boldsymbol{g}$ could be linear, as used in most previous work (Penny et al., 2003; Roberts and Penny, 2002), but could also be a general non-linear function (Friston, 2002; Woolrich et al., 2006).

In addition to the forward model parameters $\boldsymbol{w}_v$, each voxel has noise parameters. We use the standard AR(1) autoregressive noise model, which is commonly used for FMRI analysis (Woolrich et al., 2001; Penny et al., 2003):

$$e_v^{(t)} = a_v e_v^{(t-1)} + \epsilon_v^{(t)}, \quad \epsilon_v^{(t)} \sim N\left(0, \phi_v^{-1}\right). \tag{2}$$

with noise precision (inverse variance) ϕ_v and AR parameter a_v. This can be condensed into a matrix form:

$$\mathbf{R}_v \boldsymbol{e}_v \sim N\left(0, \phi_v^{-1}\mathbf{I}\right) \tag{3}$$

where $\mathbf{I}$ is the $(T-1) \times (T-1)$ identity matrix and the $\mathbf{R}_v$ is a $T-1 \times T$ bidiagonal matrix we call the "noise-whitening matrix":

$$\boldsymbol{R}_v = \begin{bmatrix} 1 & -a_v & & & \\ & 1 & -a_v & & \\ & & \ddots & \ddots & \\ & & & 1 & -a_v \end{bmatrix} \tag{4}$$

This is equivalent to the "embedded" form of AR used by Penny et al. (2003), but this matrix formulation is more convenient for our purposes. For a white noise model, simply use $\mathbf{R}_v = \mathbf{I}$, the identity matrix.

The forward model and noise model define the likelihood for voxel v:

$$P(\boldsymbol{y}_v | \boldsymbol{w}_v, \phi_v, a_v) = N\left(\mathbf{R}_v \boldsymbol{y}_v; \mathbf{R}_v \boldsymbol{g}(\boldsymbol{w}_v), \phi_v^{-1}\mathbf{I}\right). \tag{5}$$

Hierarchical models for fMRI time-series

Hierarchical Bayesian models provide a way to perform inferences on each voxel while maintaining some higher-level priors; these can help to pool information and provide regularization. In the hierarchical