Gössl, Auer, Fahrmeir:

Bayesian spatio-temporal inference in functional magnetic resonance imaging

Sonderforschungsbereich 386, Paper 192 (2000)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

Bayesian spatio-temporal inference in functional magnetic resonance imaging

C.Gössl[1,*], D.P.Auer[1], L.Fahrmeir[2]

[1]Max-Planck-Institute of Psychiatry
Kraepelinstr. 2–10, 80804 Munich, Germany
[2]Institute of Statistics, Ludwig-Maximilians-University Munich
Ludwigstr. 33, 80539 Munich, Germany
[*] *email:* goessl@mpipsykl.mpg.de

SUMMARY: Mapping of the human brain by means of functional magnetic resonance imaging (fMRI) is an emerging field in medical sciences. Current techniques to detect activated areas of the brain mostly proceed in two steps. First, conventional methods of correlation, regression and time series analysis are used to assess activation by a separate, pixelwise comparison of the MR signal time courses to the reference function of a presented stimulus. Spatial aspects caused by correlations between neighboring pixels are considered in a second step, if at all. Aim of this article is to present hierarchical Bayesian approaches that allow to simultaneously incorporate temporal and spatial dependencies between pixels directly in the model formulation. For reasons of computational feasibility, models have to be comparatively parsimonious, without oversimplifying. We introduce parametric and semiparametric spatial and spatio-temporal models that proved appropriate and illustrate their performance by application to fMRI data from a visual stimulation experiment.

KEY WORDS: human brain mapping, functional magnetic resonance imaging, MCMC, semiparametric models, spatio-temporal models.

1 Introduction

At the beginning of this decade an important advancement in medical imaging has been made. Through a new non-invasive technique called functional magnetic resonance imaging (fMRI) it has been made possible to examine sensory and higher cognitive functions in a living human brain without using an external contrast agent or tracer. Utilizing the different magnetic properties of oxygenated and desoxygenated blood, non-invasive mapping of brain functions has become feasible. The physiological changes induced by neuronal activation, known as the neurovascular coupling, lead to a local increase in blood oxygenation, the so called BOLD (blood oxygenation level dependent) effect, that may be directly visualized with specially sensitized MR sequences. While acquiring a whole time series of MR images, through this

BOLD effect changes in regional brain activity induce a systematic variation in the MR signal that is related to a presented external stimulus. Thus, areas or pixels where the signal time courses show a significantly stimulus related variation are assumed to be activated by that particular stimulus. Application of this new and intriguing methodology for human brain mapping studies, however, is far from being simple encompassing a number of critical issues regarding the physiology, physics and statistics involved. For the physiological and physical fundamentals of this technique we refer to the excellent tutorial by Lange (1996). In this article we focus on the statistical part of fMRI.

Aim of the statistical analysis of fMRI experiments is the assessment of significantly stimulus related activated areas of the brain as the basis of functional mapping. For this purpose, a time series of T MR volumes of the brain is acquired during the presentation of a certain stimulus paradigm. These volumes consist usually of about 7 to 30 slices each with 64×64 or 128×128 pixels or voxels with dimensions of ca. $3 \times 3 \times 5$ mm. For each of these pixels a complete time series exists. In the classical fMRI experiments the stimulus is presented in a so called boxcar paradigm, a sequence of ON and OFF periods (e.g. 30s OFF, 30s ON, 30s OFF, ...), in which every 1-4 seconds an image is acquired. For illustration Figure 1 shows such a boxcar stimulus together with 5 representative MR signal time courses of selected pixels. Current standard methods for data evaluation are correlation (Bandettini et al., 1993) and regression models (Friston et al., 1995), where the statistical dependence between the MR signal and the stimulus is investigated and tested to be significant. Regression models can be summarized as follows. For each pixel $i (i = 1, \ldots, I)$, the time series $\{y_{it}, t = 1, \ldots, T\}$ of MR signals is assumed to obey a linear parametric relationship

$$y_{it} = w_t' a_i + z_t b_i + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma_i^2), \quad i = 1, \ldots, I, \quad t = 1, \ldots, T. \tag{1}$$

Here w_t is a known design vector, which is supposed to model the trend or baseline drift. Design vectors that contain linear and quadratic trends or the first few terms of a Fourier expansion are common examples. The variable z_t denotes the transformed stimulus at time t, in other words z_t is a function of the presented ON-OFF-stimulus $x_t, t = 1, \ldots, T$. With regard to the transformation, we consider a temporal shift of the original stimulus by a time-delay d and a convolution with a parametric hemodynamic response function (HRF) h, so that:

$$z_t \;=\; \sum_{s=0}^{t-d} h(s, \theta) x_{t-d-s}. \tag{2}$$

Generally Poisson ($\mathrm{Po}(\lambda)$) or Gamma ($\mathrm{Ga}(\lambda, u)$) densities are chosen for this purpose. The parameters λ or λ, u and the time lag d are either calculated in a pilot least squares estimation algorithm prior to fitting model (1) to the fMRI data or are set in advance according to past experience. The transformation (2) formalizes the fact that (a) due to hemodynamic latencies the cerebral blood-flow (CBF), the source of the MR signal, increases approximately 6–8 s after the onset of the stimulus, and that (b) the flow responses do not occur suddenly, but

rather continuously and delayed. Other specifications are also possible (Bullmore et al., 1996). The parameter $b_i, i = 1, \ldots, I$ is interpreted as the effect of activation at pixel i. Testing now whether b_i is zero or not for all pixels yields a map, in fMRI literature often called 'statistical parameter map' (SPM), that shows the 'activated' areas of the brain for this particular experiment. For a visual stimulation experiment such an activation map is displayed in Figure 3a. The white areas indicate the pixels whose test statistic exceeds a value of 5.0 that corresponds to significance level $< 10^{-6}$.

Substantial gain in flexibility is achieved in Gössl, Auer and Fahrmeir (2000) by a semiparametric Bayesian approach. Using state space modelling and Kalman filtering both the baseline and a time-varying stimulus effect are modelled and estimated without being constrained to any particular parametric form, but allowing for a temporally smooth evolution. The observation model (1) is generalized to a state space model:

$$y_{it} = a_{it} + z_{it} b_{it} + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma_i^2), \tag{3}$$

where second order random walks

$$\begin{aligned} a_{it} &= 2a_{it-1} - a_{it-2} + \zeta_{it}, \quad \zeta_{it} \sim N(0, \sigma^2_{\zeta_i}). \\ b_{it} &= 2b_{it-1} - b_{it-2} + \eta_{it}, \quad \eta_{it} \sim N(0, \sigma^2_{\eta_i}). \end{aligned} \tag{4}$$

enforce smoothness of the sequence $a_i = (a_{i1}, \ldots, a_{iT})'$ of the baseline trend and of the stimulus effect $b_i = (b_{i1}, \ldots, b_{iT})'$. Further, the transformed reference function z_{it} is estimated pixelwise in advance, allowing the parameters θ and the time lag d in (2) to depend on i and thus the reference function to differ from pixel to pixel. The main feature of this approach is the description of a temporally varying stimulus related activation within an fMRI experiment in contrast to the temporally constant parametric model (1), which assumes a time-constant activation. This grants insight into the dynamics of the response pattern within an fMRI experiment.

For all mentioned models, analysis is performed pixelwise. Dependencies between pixels are considered by a smoothing of the data prior to the analysis and the application of results of Gaussian random field theory (Poline et al. 1997) to correct the significance level for multiple comparisons or by simple cluster methods to reduce activations that occurred by chance (Forman et al. 1995). For the latter, a pixel is assumed to be activated if itself exceeds a certain threshold and also at least n of its nearest neighbors. The parameter n serves in this context as a kind of smoothing parameter that controls the degree of noise reduction. Both methods have in common that they have to be applied in a two step analysis and cannot be incorporated simultaneously into a statistical model of the MR time series.

To overcome this limitation, we present Bayesian modelling approaches for fMRI space-time data that explicitly allow for a simultaneous temporal and spatial analysis.

The paper is organized as follows. In Section 2 we introduce Bayesian formulations of fMRI models, where the key issue relates to the formulation of prior distributions for the unknown

model parameters. The models and the priors differ in complexity, ranging from spatial extensions of simple parametric models to spatio-temporal models based on Markov random field priors. Section 3 gives a short outline of posterior estimation using Gibbs sampling. Section 4 illustrates these models and points out the differences and advantages of the different approaches by application to fMRI data from a visual stimulation experiment. The article closes with a discussion of possible extensions of these models.

2 Hierarchical Bayesian models for fMRI experiments

Hierarchical Bayesian models consist of an observation model for the data, given the parameters, and priors for the unknown parameters. Inference is then based on the posterior distribution of the parameters given the data.

2.1 Observation models

In principle, there are numerous options for spatio-temporal modelling of fMRI data. A conceptually straightforward option is to look at the data as a sequence $y_t = (y_{1t}, \ldots, y_{It})'$ of images and to introduce spatial correlation through the covariance matrix Σ of the corresponding error vectors $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{It})'$. For diagonal $\Sigma = diag(\sigma_1^2, \ldots, \sigma_I^2)$ we get back to the pixelwise linear model (1) or the state space model (3). For non-diagonal Σ however, this option becomes quickly computationally infeasible because of the huge dimension (I in the order of thousands). With the state space model (3) for example, one would have to run Kalman filters and smoothers with $(I \times I)$–matrices, causing insurmountable problems with memory and storage. Therefore we will not pursue this option, but introduce spatial correlation in the second stage of the hierarchy by assuming spatial or spatio-temporal Markov random field priors for the parameters.

In our experience a Gaussian assumption for the observations, conditional upon parameters, is not critical. Therefore the general form of the observation model for pixel $i, i = 1, \ldots, I$, is

$$y_{it}|a_{it}, b_{it}, \sigma_i^2 \sim N(a_{it} + z_{it}b_{it}, \sigma_i^2), \quad t = 1, \ldots, T, \tag{5}$$

with z_{it} as the transformed stimulus, a_{it} as the trend and b_{it} as the activation effect at time t. Given the parameters, observations y_{it} are conditionally independent.

Parametric models are obtained by modelling trend and activation effect as

$$a_{it} = w_t' a_i, \quad b_{it} = v_t' b_i,$$

with design vectors w_t and v_t as for the linear model (1), and time-constant parameter vectors $a_i = (\alpha_1, \ldots, \alpha_p)'$ and $b_i = (\beta_1, \ldots, \beta_q)'$ of fixed low dimension. For $v_t = 1$, the activation effect is assumed to be time-constant, while inclusion of time-varying components, like t, t^2 and sine/cosine terms into v allows for a time-varying activation effect.
Semiparametric models are obtained by assuming the whole sequence $a_i = (a_{i1}, \ldots, a_{iT})'$ and $b_i = (b_{i1}, \ldots, b_{iT})'$ of parameters as unknown and applying appropriate smoothness priors.

2.2 Prior distributions

The choice of prior distributions is less straightforward than formulation of the observation model. In the following we show how a wide range of models for fMRI experiments can be covered by simply applying different kinds of prior distributions. We proceed from simple parametric models to semiparametric spatio-temporal models, reflecting the trade off between computational simplicity and model complexity. We start with a reformulation of the pixelwise models (1) and (3, 4) fitting into our general Bayesian framework. Priors for a and b are mostly of the same type. Thus, to avoid unnecessary repetitions, we define them only for the stimulus effect b, where most interest lies on.

Pixelwise parametric modelling

For a Bayesian version of the parametric model (1) the most simple prior is to assume a highly dispersed diffuse Gaussian distribution for every parameter component. For scalar b_i, we get

$$b_i|\lambda \sim N(\mu, 1/\lambda) \quad \text{or} \quad p(b_i|\lambda) \propto \exp(-\frac{1}{2}\lambda b_i^2) \tag{6}$$

with μ as the prior mean and λ arbitrarily small to reduce the prior's influence. With $\lambda \to 0$ the prior is diffuse. Taking additionally the posterior mean as point estimate, the Bayes estimator and the least squares estimator become identical. Parameters are estimated pixelwise, i.e. for each pixel separately, and temporally constant.

Spatial parametric modelling

One of the main advantages of the Bayesian approach is that spatial correlations can easily be introduced into the modelling of the time series. For the parametric model (6) with scalar b_i this is achieved by assuming intrinsic autoregressive priors, also called pairwise difference priors (see e.g. Besag, York and Mollie, 1991). It can be seen as a kind of stochastic interpolation of

the adjacent neighbors and can be written down for $b = (b_1, \ldots, b_I)'$ as follows:

$$p(b|\lambda) \quad \propto \quad \exp(-\frac{1}{2}\lambda b'Qb). \tag{7}$$

The precision matrix Q has elements

$$Q_{ij} = \begin{cases} n_i & i = j \\ -1 & i \sim j \\ 0 & \text{else} \end{cases} \tag{8}$$

with n_i the numbers of neighbors of pixel i, or equivalently

$$p(b|\lambda) \quad \propto \quad \exp\{-\frac{1}{2}\lambda \sum_{i\sim j}(b_i - b_j)^2\}. \tag{9}$$

This prior can also be rewritten in terms of conditional distributions:

$$b_i|b_{j\neq i}, \lambda \quad \sim \quad N(\frac{1}{n_i}\sum_{j\sim i} b_j, \frac{1}{n_i\lambda}). \tag{10}$$

Compared to pixelwise parametric modelling, estimates are spatially smoothed. The number of neighbors used is essential for the amount of smoothing, the larger the neighborhood the spatially smoother becomes the parameter estimate. To avoid oversmoothing and blurring of edges in the activation surface, we use the four nearest neighbors.

Pixelwise semiparametric modelling

For the semiparametric state space model (3,4), Gaussian priors for the sequences $a_i = (a_{i1}, \ldots, a_{iT})'$ and $b_i = (b_{i1}, \ldots, b_{iT})'$ are defined by the random walk models in (4). Assuming diffuse initial priors for $a_{i1}, a_{i2}, b_{i1}, b_{i2}$, these priors can equivalently be rewritten in form of global Gaussian smoothness priors. For b_i we get

$$p(b_i|\lambda_i) \propto \exp(-\frac{1}{2}\lambda_i b_i' Q b_i), \tag{11}$$

with λ_i as the precision or inverse variance $\lambda_i = 1/\sigma^2_{\eta_i}$ of the prior. It controls the smoothness of the estimated curves. The precision matrix Q penalizes too rough estimates of the sequence $b_i = (b_{i1}, \ldots, b_{iT})'$. For a random walk of second order it has the following form:

$$Q = \begin{pmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{pmatrix} \tag{12}$$

This prior can also be seen as a stochastic quadratic interpolation in the temporal dimension. The effect of this prior is a model with time-varying coefficients, which vary slowly and smoothly in time. This is appropriate for describing slow temporal fluctuations of an inert underlying system as the convoluted blood flow is supposed to be.

Semiparametric spatio-temporal modelling

Up to now we introduced spatial and temporal smoothness priors for fMRI experiments separately. To obtain models that simultaneously consider these two aspects a combination of the above properties is necessary. This could be done by means of separable or non-separable time space interactions. We present two models that are a reasonable compromise between computational tractability and model complexity.
Separable time-space interactions can be thought of as a splitting of the particular parameter into several components. A simple model of this kind is obtained by splitting the activation effect b_{it} into

$$b_{it} = \alpha_i + \beta_{it}, \tag{13}$$

assuming a spatial smoothness prior (7) for the time constant part α_i and a pixelwise temporal random walk prior (11) for the time-varying effects β_{it}. For identifiability reasons, the sequence $\beta_i = (\beta_{i1}, \ldots, \beta_{iT})'$ has to be centered about zero. This model is useful, if the primary scientific goal is detection of activation areas. It separates spatially correlated static activation effects α_i and additional time-varying fluctuations β_{it}. In contrast, for the parametric spatial model (7) with time-constant activation effect, the fluctuations average over time and are implicitly already added to b_i.
However, as a result of the pixelwise modelling of the random effects the temporal variations in adjacent pixels are still more or less independent. If main interest also includes these fluctuations, *non-separable interactions* should be considered. In this case a splitting into temporal and spatial components is not possible. Both dependencies have to be incorporated into one prior. To do so, we modify a prior proposed by Clayton (1996) and applied by Knorr-Held (2000) for the interaction of random effects. These interactions are modelled by using the Kronecker product of two penalty matrices as a new penalty in the Markov random field prior. Transfered to our problem at hand, this means with Q^s as in (7) and Q^t as in (11):

$$p(b|\lambda) \propto \exp(-\frac{1}{2}\lambda b'Qb), \quad \text{with} \quad Q = Q^s \otimes Q^t \tag{14}$$

or equivalently:

$$p(b|\lambda) \propto \exp\{-\frac{1}{2}\lambda \sum_{i\sim j}\sum_t (\Delta^2 b_{it} - \Delta^2 b_{jt})^2\}, \tag{15}$$

with $\Delta^2 b_{it} = b_{it} - 2b_{it-1} + b_{it-2}$ denoting the second differences of b_{it}. Penalizing differences in the time courses of adjacent pixels, this prior should effect a temporally as well as a spatially smooth evolution of the parameters. But in our case, due to the dominating spatial structure of the fMRI data and a single global smoothness parameter λ, spatial smoothness is overestimated at the expense of the temporal. Therefore, we augment the above prior by a kind of main effect to control temporal smoothness. With it not only spatial differences in the time courses are penalized but also too rough evolutions itself. Additionally, to account for the high dimensions of the data ($\sim$ 2–300.000 observations) pixelwise precisions are introduced. This results in a spatio-temporal prior for fMRI experiments as follows:

$$p(b|\lambda) \propto \exp(-\frac{1}{2} b' Q_\lambda b), \quad \text{with} \quad Q_\lambda = (Q^s \otimes Q^t) + (\Lambda \otimes Q^t) = (Q^s + \Lambda) \otimes Q^t \tag{16}$$

with Q^t as defined in (11), Q^s with elements

$$Q^s_{ij} = \begin{cases} \sum_{k \in \partial i} (\lambda_i + \lambda_k) & i = j \\ -(\lambda_i + \lambda_j) & i \sim j \\ 0 & \text{else} \end{cases}, \tag{17}$$

and Λ as the diagonal matrix of pixelwise precisions $\lambda_i, i = 1, \ldots, I$. For a second order random walk this can be rewritten with the above definitions:

$$p(b|\lambda) \propto \exp[-\frac{1}{2} \sum_i \{\lambda_i \sum_t (\Delta^2 b_{it}^2 + \sum_{j \in \partial i} (\Delta^2 b_{it} - \Delta^2 b_{jt})^2)\}]. \tag{18}$$

The introduction of the temporal main effect can also be accounted for as a downweighting of the spatial influence, that puts emphasis on the temporal smoothness of the estimated parameters.

In the last stage of the hierarchy, we assume priors for unknown hyperparameters, i.e. precisions λ_i and inverse variances $1/\sigma_i^2$ of the observation errors. A common choice are highly dispersed gamma distributions $\mathrm{GA}(\gamma_a, \gamma_b)$ where γ_a and γ_b have to be chosen appropriately.

Model specification is completed by assuming (conditional) independence between (blocks of) parameters. Gathering parameters in vectors $a = (a_1', \ldots, a_I')$, $b = (b_1', \ldots, b_I')$, $\lambda = (\lambda_1, \ldots, \lambda_I)'$, $\sigma^2 = (\sigma_1^2, \ldots, \sigma_I^2)$ and observations $Y = (y_{it}, i = 1, \ldots, I, t = 1, \ldots, T)$ the posterior distribution of the parameters given the observed data has the following form:

$$p(a, b, \sigma^2, \lambda|Y) \propto l(Y|a, b, \sigma^2)\ p(a|\lambda)\ p(b|\lambda)\ p(\lambda)\ p(\sigma^2). \tag{19}$$

The likelihood $l(Y|a, b, \sigma^2)$ is determined by the observation model, the other factors by the priors above.

3 Inference

Complete Bayesian inference is based solely on the posterior distribution (19). Even though this distribution is only known up to a normalizing constant, samples can be drawn by means of Markov chain Monte Carlo (MCMC) methods. This is achieved by iteratively drawing single parameters or groups of parameters from their full conditionals, i.e., the conditional distribution of these parameter given the rest and the data. For a thorough introduction to these methods see e.g. Tierney (1994), Gilks, Richardson and Spiegelhalter (1996) or Gamerman (1997). Due to the choice of normal and gamma priors for the fMRI models full conditionals can be written down in closed form and are itself normal or gamma distributions. Thus, Gibbs sampling can be used to draw from univariate or multivariate full conditionals. In the following we outline only the general strategy. Some details are provided in the Appendix. In all approaches the parameters a_i and b_i, respectively α_i and β_i in the separable model (13), are drawn for each pixel separately. Scalar parameters are sampled univariately, vectors as blocks. In both cases the full conditionals are normal distributions

$$\begin{aligned} p(a_i|.) &\sim N(\mu_a, \Sigma_a), \\ p(b_i|.) &\sim N(\mu_b, \Sigma_b), \end{aligned}$$

with means and covariances of adequate dimensions. Due to the use of conjugate priors, for the pixelwise parametric model (6) conditional means and variances are easy to calculate (see e.g. Robert, 1994). The spatially correlated parameters of model (7) and the α_i of model (13) can be sampled similarly, only replacing prior means and variances according to (10). Analogously one proceeds for the calculation of the mean vector and covariance matrix of the semiparametric model (11). Updating can be done by direct sampling, exploiting the band structure of the precision matrix. This method can also be applied to the parameters β_i in model (13). For the non-separable model (16) the mean vectors of the semiparametric model have to be augmented by the weighted spatial influence of the neighbors. Further, neighboring precisions have to be incorporated into the covariance matrix.
The variances of the observation errors are sampled separately from inverse Gamma distributions

$$p(\sigma_i^2|.) \sim IG(a_y, b_y),$$

where number of time points and sum of residual squares of the observation errors determine the parameters a_y and b_y. Global or pixelwise precisions are drawn from Gamma distributions

$$\begin{aligned} p(\lambda_{a(i)}^2|.) &\sim GA(a_a, b_a), \\ p(\lambda_{b(i)}^2|.) &\sim GA(a_b, b_b), \end{aligned}$$

where the rank of the precision matrix and the quadratic form in the prior determine the parameters a_a, a_b, b_a and b_b.

As noted in Section 1, the hemodynamic response function is determined in advance. In theory a fully Bayesian approach is conceivable where also the HRF parameters are modelled and estimated by MCMC. To keep computation time in a reasonable extent we decided to exclude these parameters from the MCMC algorithm and calculate them in a pilot estimate by least squares. However, with increasing computational power an incorporation can be considered.

4 Applications

The data set we use for illustration of the above approaches is a fMRI time series from a visual stimulation experiment. The set of $T2^*$ images was acquired on a 1.5 T system (Echospeed, GE Medical Systems, Milwaukee). Seven slices parallel to the intercommissural line with a voxel size of $2.9 \times 2.9 \times 5$ mm were positioned to cover the occipital lobes. A series of 73 images consisting of 128×128 pixels was acquired with the initial three images being discarded to avoid non-steady-state effects. With regard to the stimulation paradigm of the fMRI data, the subsequent 70 images were divided into four rest and three activation periods, with each period consisting of 10 images (30 s long). The reference box car is shown in Figure 1. During the visual stimulation periods a rectangular checkerboard that alternated at a frequency of 4 Hz was displayed with central fixation point. The fixation point was displayed with an uniformly dark background in the rest periods. To correct for subject's motion, an image registration (Jiang et al., 1995) was performed prior to the analysis.

All MCMC algorithms consisted of 6000 iterations with the first 1000 being discarded as burn-in and every 5th iteration included in the final sample. With this high-dimensional data, convergence diagnostics were reduced to a selection of randomly chosen parameter chains. Additionally, the sampling path of the Bayesian deviance was monitored. For all samples autocorrelations were less than 0.1 and almost independent of starting values. Parameters of the Gamma hyperpriors were both set to 1, respectively 1 and 10. First order neighborhoods were used in the spatial applications, i.e., only the four nearest neighbors entered the calculations. Figure 2 shows the sampling paths, autocorrelations and histograms of a representative parameter and the deviance of the non-separable model. The parameter histogram is superimposed with a normal distribution with same mean and variance. Exploiting the similarity of these distributions, for working memory reasons in each iteration only first and second moments of the approximately normal posteriors for parameters were updated. The posterior mean was used as point estimate.

Activation maps were calculated by testing whether the stimulus effect is zero or not. For a confidence level α this is equivalent to checking whether the α quantile covers the zero or

not. After standardizing the posterior to unit variance, this reduces to testing whether the transformed mean of the approximately normal posterior distribution with unit variance is sufficiently far apart from zero. For example, for such a distribution with mean 1.96, the 0.025 quantile does just not cover the zero. Thus, familiar activation maps can simply be calculated by thresholding the map of standardized posterior means at a certain value determined by the significance level α. The maps in our example were thresholded at different levels, to allow for a qualitative comparison of approaches.

In Figure 3 an activation map (threshold 5.0) and surface plots of absolute and standardized fitted values for the parametric pixelwise approach (6) are displayed. The results of the classical parametric model (1) fitted by least squares are almost equivalent to the above, differing only marginally. Main activation is sited in the primary visual cortex, also called V1, in the occipital lobe. Additionally, small activated areas in the extrastriate cortex or secondary visual areas can be found in the post parietal areas on both sides of the brain. Further, anterior or in front of V1, a few pixels in the precuneus show a significant reaction to the stimulus. Primary visual areas are the first cortical relay station for general processing of all visual stimuli. Conversely, the secondary or associate regions coactivated in our experiment are supposed to handle specific information about e.g. colour and shape of objects. Episodic memory, a part of the declarative memory, is assumed to be located in the precuneus.
The results of the parametric spatial approach (7) are shown in Figure 4. It can be seen that in comparison to the pixelwise model estimated regions appear smoother and more connected. This is particularly so for the absolute fitted values. Single peaks are strongly reduced. This result was to be expected because one of the main effects of the spatial prior is smoothing of the parameter surface and reduction of single isolated peaks. Apart from that, spatial and pixelwise activation maps are quite similar, differing only in the level (threshold 6.5 for the spatial model). This is thought to be the consequence of the strong dependence between neighboring parameters and the resulting narrower credibility regions or posterior distributions.
Figure 5 shows the estimate for the dynamic model (3) with the parameter prior (11). Each of the respective three maps represents one time point in the stimulation cycle indicated by the asterisk on the boxcar on bottom of the figure. We confine to these three maps because activation varies only slowly in time and most interesting differences can be found between stimulation periods in this activation paradigm. In all maps a static occipital activation in V1 can be seen. Additionally, semiparametric modelling reveals a remarkable temporal variation in the secondary visual areas on both sides. Further, fluctuations in the precuneus can be observed. For illustration the effects of parametric and dynamic modelling are plotted in Figure 6 for a selected time series from the extrastriate visual cortex. The different quality of modelling is apparent, with a substantial gain in fitting accuracy for the dynamic model. The series of activation maps was thresholded at a level of 3.0. Indicated by the lower threshold, in comparison to the parametric approaches standardized values are noticeable reduced. The

dramatically increased number of parameters seems to lead to a splitting of information and thus to broadened credibility regions.
Comparable maps (Figure 7) are also obtained for the separable spatio-temporal model (thresholds 4.0). Comparing the spatial average effect to the temporally constant model (7), primary visual areas are similarly well detected whereas extrastriate activation differs remarkably. This is consistent with the dynamic model, where temporal fluctuations occur mostly in extrastriate areas whereas V1 activation is temporally relatively invariant. Figure 7c shows the spatial effect superimposed with the temporal random effect. Temporal fluctuations in the mentioned areas are described in accordance to the dynamic approach, whereas underlying activated areas show the typical spatial smoothness induced by the prior. The surface plots reflect the strong influence of the spatial effect. Even though more parameters than in the dynamic model are estimated, V1 activation is striking as well as extrastriate in the last stimulation cycle. As already mentioned, in this approach the temporal variations of neighboring pixels are still assumed to be independent.
Interactions between dynamic effects can be accounted for by the non-separable model (16). The results are shown in Figure 8. The spatial effect of the spatio-temporal prior can be seen especially in the surface plot. Through the principle of 'borrowing strength', posterior variances can be reduced and thus inference improved. The difference between the V1 values and the rest of the brain is appreciably increased, allowing for a more reliable discrimination of activated and non-activated regions. But, the dependence between adjacent time courses also implies reduced temporal variation found by this approach, especially in the secondary visual areas and in the precuneus where activation is suppressed strongly. Because non-activated pixels show no or few variation of the stimulus effect in time, the spatial prior also constrains the effect of activated pixels to perform only minor changes. This property could perhaps be eliminated by introducing robust versions of this type of prior that allow explicitly for edges between adjacent activated and not activated pixels.

When comparing the results it has to be stated that activations are estimated very consistently especially in the primary visual cortex. Also temporal fluctuations are covered in all dynamic approaches similarly. The only differences observed are the degrees of temporal smoothness estimated for the effects. In our experience these differences are not only a consequence of the choice of the hyperparameters but are also model inherent. Even though the hyperparameters have an effect on the results, data information mostly outweighs prior specifications. Consequently, this leads us directly to the problem of model comparison.
Table 1 contains posterior median, mean and standard deviation of the deviance as a global measure of goodness of fit for the models considered. It can be seen that all three parameters for the separable model are noticeable larger than for the other models, which are more or less indistinguishable in terms of goodness of fit. The increased mean and variation might be a consequence of the augmented parameter space for the baseline and the activation effect and

thus the higher complexity. Regarding the other approaches, the deviance alone is of limited relevance for the problem of choosing the adequate model. Model choice will therefore need to be based on the specific objectives of each study. If detection of activation areas averaged over time is the primary goal, simpler models with time constant activation parameters seem to be appropriate and taking care of spatial correlation is recommended to robustify segmentation of activated and non-activated areas. If, however, additional interest exists in dynamic effects of activation, the non-separable spatio-temporal model is a good choice.
Recently, Spiegelhalter, Best and Carlin (1998) proposed a deviance-based model selection criterion (DIC) also penalizing model complexity comparable to Akaike's information criterion for frequentistic approaches. But because experience with the performance of DIC is still limited, we do not use it here.
Currently, we are exploring the suitability of these models on other data sets and different experimental designs to allow for more reliable conclusions with respect to the problem of model comparison.

5 Conclusion and Outlook

We believe that Bayesian hierarchical modelling of the time-space structure underlying fMRI experiments offers new and intriguing possibilities for human brain mapping. We have shown that this approach is not only computationally feasible even with massive fMRI data sets, comprising several thousands of spatially correlated time series, but that it has great potential for detecting and analyzing spatio-temporal effects due to its flexibility. Some extensions and modifications for future work are outlined in the following.
Gaussian priors as applied here are appropriate for comparably smooth underlying functions of time or surfaces. However, they may blur edges or regions with high curvature between areas of high and low activation. To avoid this, spatially robust priors could be incorporated in spatio-temporal models, without destroying the ability of modelling smooth trends and time-varying effects. By penalizing larger differences between adjacent pixels less severely than normal distributions such priors do not smooth over large systematic gaps but only small irregularities induced by noise. In a purely spatial context Higdon (1994) suggested to extend normal priors by introducing gamma distributed weights for the precision. This results in heavier tailed student priors. Other robust distributions such as Laplace, Huber or truncated Gaussian (Künsch, 1994) priors could also be useful. However, when using robust priors, full conditionals are no longer explicitly given so that more time consuming Metropolis steps would have to be introduced instead.
Further, parametric modelling of time-varying effects with a parsimonious number of basis functions should be kept in view. Reducing the number of parameters and therefore computation

time, they represent an alternative to the computationally very expensive fully nonparametric Bayesian methods.
Apart from the priors for time and space, Bayesian modelling of the hemodynamic response function should be attempted, too, even though computation time would increase considerably. Another extension might be the incorporation of further substantial prior information. Knowledge about the functional organisation of the brain has dramatically evolved over the last decade by use of various mapping techniques such as positron emission tomography, EEG and fMRI. The Bayesian approach is ideally suited to account for such comprehensive prior knowledge. The incorporation of this information may range from applying simple vague probability gradients up to restricting activation to certain predefined sulci and gyri. However, plasticity of the brain, the fact of a displacement of functional areas, should always be kept in mind and carefully considered. An information that surely has to be included is that activation only occurs in gray matter. Thus, results of brain segmentation into distinct classes of tissue, certainly hold promise to improve the reliability of estimation in terms of reducing artefacts in white matter or cerebro spinal fluid. A positive side effect would also be the reduction of computation time by the exclusion of a substantial part of the data.

Acknowledgment

This work was supported by a grant from the German National Science Foundation (DFG), Sonderforschungsbereich 386. The authors want to express thanks to Harvard Rue from the Department of Mathematics at the Norwegian University of Science and Technology for providing valuable advice and support for the MCMC implementation.

Appendix

Full conditionals of parameters:

Full conditionals are proportional to the posterior distribution of the parameters (19) given the data

$$p(a, b, \sigma^2, \lambda|Y) \propto l(Y|a, b, \sigma^2)\ p(a|\lambda)\ p(b|\lambda)\ p(\lambda)\ p(\sigma^2).$$

with the likelihood determined by the observation model

$$y_{it}|a_{it}, b_{it}, \sigma_i^2 \sim N(a_{it} + z_{it}b_{it}, \sigma_i^2), \quad t = 1, \ldots, T, \quad i = 1, \ldots, I,$$

where for time-constant models $a_{it} = w_t' a_i$ and $b_{it} = v_t' b_i$ with adequate w_t and v_t. The MR signal for pixel $i, i = 1, \ldots, I$ is denoted by $y_i = (y_{it}, t = 1, \ldots, T)$ and $z_i = (z_{it}, t = 1, \ldots, T)$ is the transformed reference function. The prior distributions are specified in Section 2.2. Thus, full conditionals can be calculated as follows.

Pixelwise parametric modelling

Prior distributions of scalar parameters a_i, b_i and σ_i^2:

$$\begin{aligned} a_i|\lambda &\sim N(\mu, 1/\lambda), \\ b_i|\lambda &\sim N(\mu, 1/\lambda), \\ \sigma_i^2 &\sim IG(\gamma_a, \gamma_b). \end{aligned}$$

Full conditionals:

$$\begin{aligned} a_i|Y, b_i, \lambda, \sigma_i^2 &\sim N(\frac{\sigma_i^2 \mu + 1/\lambda \sum_t (y_{it} - z_{it} b_i)}{\sigma_i^2 + T/\lambda}, \frac{\sigma_i^2/\lambda}{\sigma_i^2 + T/\lambda}), \\ b_i|Y, a_i, \lambda, \sigma_i^2 &\sim N(\frac{\sigma_i^2 \mu + 1/\lambda \sum_t z_{it}(y_{it} - a_i)}{\sigma_i^2 + z_{it}^2/\lambda}, \frac{\sigma_i^2/\lambda}{\sigma_i^2 + z_{it}^2/\lambda}), \\ \sigma_i^2|. &\sim IG(\gamma_a + T/2, \gamma_b + \frac{1}{2}\sum_t (y_{it} - a_i - z_{it} b_i)^2). \end{aligned}$$

Spatial parametric modelling

Prior distributions of parameters $a = (a_1, \ldots, a_I)'$ and $b = (b_1, \ldots, b_I)'$:

$$\begin{aligned} p(a) &\propto \exp(-\frac{1}{2}\lambda_a a' Q a), \\ p(b) &\propto \exp(-\frac{1}{2}\lambda_b b' Q b), \end{aligned}$$

with the precision matrix Q defined in (8), or for scalar parameters a_i and b_i:

$$\begin{aligned} a_i|a_{j \neq i}, \lambda_a &\sim N(\frac{1}{n_i}\sum_{j \sim i} a_j, \frac{1}{n_i \lambda_a}), \\ b_i|b_{j \neq i}, \lambda_b &\sim N(\frac{1}{n_i}\sum_{j \sim i} b_j, \frac{1}{n_i \lambda_b}). \end{aligned}$$

with n_i the numbers of neighbors of pixel i. Prior distributions of variances σ_i^2 and precisions λ_a and λ_b :

$$\begin{aligned} \sigma_i^2 &\sim IG(\gamma_a, \gamma_b), \\ \lambda_a &\sim GA(\gamma_a, \gamma_b), \\ \lambda_b &\sim GA(\gamma_a, \gamma_b). \end{aligned}$$

Full conditionals of $a_i|.$ and $b_i|.$ as for the pixelwise model with adequately replaced prior mean μ and precision λ. Further full conditionals:

$$\begin{aligned}
\sigma_i^2|. &\sim IG(\gamma_a + T/2, \gamma_b + \frac{1}{2}\sum_t (y_{it} - a_i - z_{it}b_i)^2), \\
\lambda_a|. &\sim GA(\gamma_a + rk(Q)/2, \gamma_b + \frac{1}{2}a'Qa), \\
\lambda_b|. &\sim GA(\gamma_a + rk(Q)/2, \gamma_b + \frac{1}{2}b'Qb).
\end{aligned}$$

Pixelwise semiparametric modelling

Prior distributions of parameters $a_i = (a_{i1}, \ldots, a_{iT})'$ and $b_i = (b_{i1}, \ldots, b_{iT})'$:

$$\begin{aligned}
p(a_i) &\propto \exp(-\frac{1}{2}\lambda_{a_i} a_i' Q a_i), \\
p(b_i) &\propto \exp(-\frac{1}{2}\lambda_{b_i} b_i' Q b_i),
\end{aligned}$$

with the precision matrix Q defined in (12). Prior distributions of variances σ_i^2 and precisions λ_{a_i} and λ_{b_i}:

$$\begin{aligned}
\sigma_i^2 &\sim IG(\gamma_a, \gamma_b), \\
\lambda_{a_i} &\sim GA(\gamma_a, \gamma_b), \\
\lambda_{b_i} &\sim GA(\gamma_a, \gamma_b).
\end{aligned}$$

Full conditionals for parameters a_i and b_i:

$$\begin{aligned}
a_i|Y, b_i, \lambda_{a_i}, \sigma_i^2 &\sim N((\frac{1}{\sigma_i}I + \lambda_{a_i}Q)^{-1}\{\frac{1}{\sigma_i}(y_i - \tilde{z}_i b_i)\}, (\frac{1}{\sigma_i}I + \lambda_{a_i}Q)^{-1}), \\
b_i|Y, a_i, \lambda_{b_i}, \sigma_i^2 &\sim N((\frac{1}{\sigma_i}\tilde{z}_i'\tilde{z}_i + \lambda_{b_i}Q)^{-1}\{\frac{1}{\sigma_i}\tilde{z}_i(y_i - a_i)\}, (\frac{1}{\sigma_i}\tilde{z}_i'\tilde{z}_i + \lambda_{b_i}Q)^{-1}),
\end{aligned}$$

with $\tilde{z}_i = diag(z_{it}, t = 1, \ldots, T)$.
Further full conditionals:

$$\begin{aligned}
\sigma_i^2|. &\sim IG(\gamma_a + T/2, \gamma_b + \frac{1}{2}\sum_t (y_{it} - a_{it} - z_{it}b_{it})^2), \\
\lambda_{a_i}|. &\sim GA(\gamma_a + rk(Q)/2, \gamma_b + \frac{1}{2}a_i'Qa_i), \\
\lambda_{b_i}|. &\sim GA(\gamma_a + rk(Q)/2, \gamma_b + \frac{1}{2}b_i'Qb_i).
\end{aligned}$$

Semiparametric spatio-temporal modelling

Full conditionals of the separable model (13) are straightforward modifications of the above sections.

For the non-separable model (16), prior distributions of parameters $a = (a_{it}, i = 1, \ldots, I, t = 1, \ldots, T)'$ and $b = (b_{it}, i = 1, \ldots, I, t = 1, \ldots, T)'$ run as follows:

$$\begin{aligned} p(a) &\propto \exp(-\frac{1}{2}\lambda_a a'Qa), \\ p(b) &\propto \exp(-\frac{1}{2}\lambda_b b'Qb), \end{aligned}$$

with the precision matrix Q defined in (16). For $a_i|a_{-i}$ this can be transformed into:

$$\begin{aligned} p(a_i|a_{-i}) &\propto \exp(-\frac{1}{2}\tilde{\lambda}_a(a_i - \mu_a)'Q^t(a_i - \mu_a)), \\ p(b_i|a_{-i}) &\propto \exp(-\frac{1}{2}\tilde{\lambda}_a(b_i - \mu_b)'Q^t(b_i - \mu_b)), \end{aligned}$$

with Q^t as in (11). Further, $\mu_a = (\mu_{a1}, \ldots, \mu_{aT})'$,

$$\begin{aligned} \mu_{ai} &= \{\lambda_i(n_i + 1) + \sum_{j\in\partial i}\lambda_j\}^{-1}\sum_{j\in\partial i}\{(\lambda_i + \lambda_j)a_{jt}\} \text{ and} \\ \tilde{\lambda}_a &= \{\lambda_i(n_i + 1) + \sum_{j\in\partial i}\lambda_j\}, \end{aligned}$$

with n_i as the number of neighbors of pixel i. The parameters μ_b and $\tilde{\lambda}_b$ are calculated equivalently. Prior distributions of variances σ_i^2 and precisions λ_{a_i} and λ_{b_i}:

$$\begin{aligned} \sigma_i^2 &\sim IG(\gamma_a, \gamma_b), \\ \lambda_{a_i} &\sim GA(\gamma_a, \gamma_b), \\ \lambda_{b_i} &\sim GA(\gamma_a, \gamma_b). \end{aligned}$$

Full conditionals for parameters a_i and b_i:

$$\begin{aligned} a_i|Y, a_{-i}, b, \lambda_a, \sigma_i^2 &\sim N((\frac{1}{\sigma_i}I + \tilde{\lambda}_a Q^t)^{-1}\{\frac{1}{\sigma_i}(y_i - \tilde{z}_i b_i) + \tilde{\lambda}_a Q^t\mu_a\}, (\frac{1}{\sigma_i}I + \tilde{\lambda}_a Q^t)^{-1}), \\ b_i|Y, b_{-i}, a, \lambda_b, \sigma_i^2 &\sim N((\frac{1}{\sigma_i}\tilde{z}_i'\tilde{z}_i + \tilde{\lambda}_b Q^t)^{-1}\{\frac{1}{\sigma_i}\tilde{z}_i(y_i - a_i) + \tilde{\lambda}_b Q^t\mu_b\}, (\frac{1}{\sigma_i}\tilde{z}_i'\tilde{z}_i + \tilde{\lambda}_b Q^t)^{-1}), \end{aligned}$$

with $\tilde{z}_i = diag(z_{it}, t = 1, \ldots, T)$.
Further full conditionals:

$$\begin{aligned} \sigma_i^2|. &\sim IG(\gamma_a + T/2, \gamma_b + \frac{1}{2}\sum_t(y_{it} - a_{it} - z_{it}b_{it})^2), \\ \lambda_{a_i}|. &\sim GA(\gamma_a + (n_i + 1)(T - 2)/2, \gamma_b + \frac{1}{2}[\sum_t\{\Delta^2 a_{it}^2 + \sum_{j\in\partial i}(\Delta^2 a_{it} - \Delta^2 a_{jt})^2\}]), \\ \lambda_{b_i}|. &\sim GA(\gamma_a + (n_i + 1)(T - 2)/2, \gamma_b + \frac{1}{2}[\sum_t\{\Delta^2 b_{it}^2 + \sum_{j\in\partial i}(\Delta^2 b_{it} - \Delta^2 b_{jt})^2\}]), \end{aligned}$$

with $\Delta^2 x_{it} = x_{it} - 2x_{it-1} + x_{it-2}$.

References

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., and Hyde, J. S. (1993). Processing Strategies for Time-Course Data Sets in Functional MRI of the Human Brain. *Magnetic Resonance in Medicine*, **30**, 161–173.

Besag, J., York, J., and Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43, 1**, 1–59.

Bullmore, E., Brammer, M., Williams, S. C. R., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., and Sham, P. (1996). Statistical Methods of Estimation and Inference for Functional MR Image Analysis. *Magnetic Resonance in Medicine*, **35**, 261–277.

Clayton, D. (1996). Generalized linear mixed models. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (eds), pp. 275–301. Chapman & Hall, London.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., and Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine*, **33**, 636–647.

Friston, K. J., Holmes, A. P., Poline, J.-B., Grasby, P., Williams, S. C. R., Frackowiak, R. S. J., and Turner, R. (1995). Analysis of fMRI Time-Series Revisited. *Neuroimage*, **2**, 45–53.

Gamerman, D. (1997). Efficient Sampling from the posterior distribution in generalized linear models. *Statistics and Computing*, **7**, 57–68.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Gössl, C., Auer, D. P., and Fahrmeir, L. (2000). Dynamic models in fMRI. *Magnetic Resonance in Medicine*, **43**, 72–81.

Higdon, D. (1994). *Spatial applications of Markov chain Monte Carlo for Bayesian Inference*. Ph.D. thesis, Department of Statistics, University of Washington.

Jiang, A. P., Kennedy, D. N., Baker, J. R., Weisskoff, R., Tootell, R. B. H., Woods, R. P., Benson, R. R., Kwong, K. K., Brady, T. J., Rosen, B. R., and Belliveau, J. W. (1995). Motion detection and correction in functional MR imaging. *Human Brain Mapping*, **3**, 224–235.

Knorr-Held, L. (2000). Bayesian Modelling of Inseparable Space-Time Variation in Disease Risk. *Statistics in Medicine*, **to appear**.

Künsch, H. R. (1994). Robust priors for smoothing and image restoration. *Annals of the Institute of Statistical Mathematics*, **46, 1**, 1–19.

Lange, N. (1996). Statistical Approaches To Human Brain Mapping By Functional Magnetic Resonance Imaging. *Statistics in Medicine*, **15**, 389–428.

Poline, J. B., Worsley, K. J., Evans, A. C., and Friston, K. J. (1997). Combining Spatial Extent and Peak Intensity to Test for Activations in Functional Imaging. *Neuroimage*, **5**, 83–96.

Robert, C. P. (1994). *The Bayesian Choice*. Springer Verlag, New York.

Spiegelhalter, D. J., Best, N. G., and Carlin, B. P. (1998). Bayesian deviance, the effective number of parameters and the comparison of arbitrarily complex models. Technical report, MRC Biostatistics Unit, Institute of Public Health, Cambridge.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, **22, 4**, 1701–1762.

Model	Median	Mean	STD
pixelwise parametric model	211536	211527	622
spatial parametric model	211539	211543	638
pixelwise semiparametric model	211458	211444	630
separable spatio-temporal model	221730	221709	1722
non-separable spatio-temporal model	211511	211515	643

Table 1: Posterior summaries of the deviance.

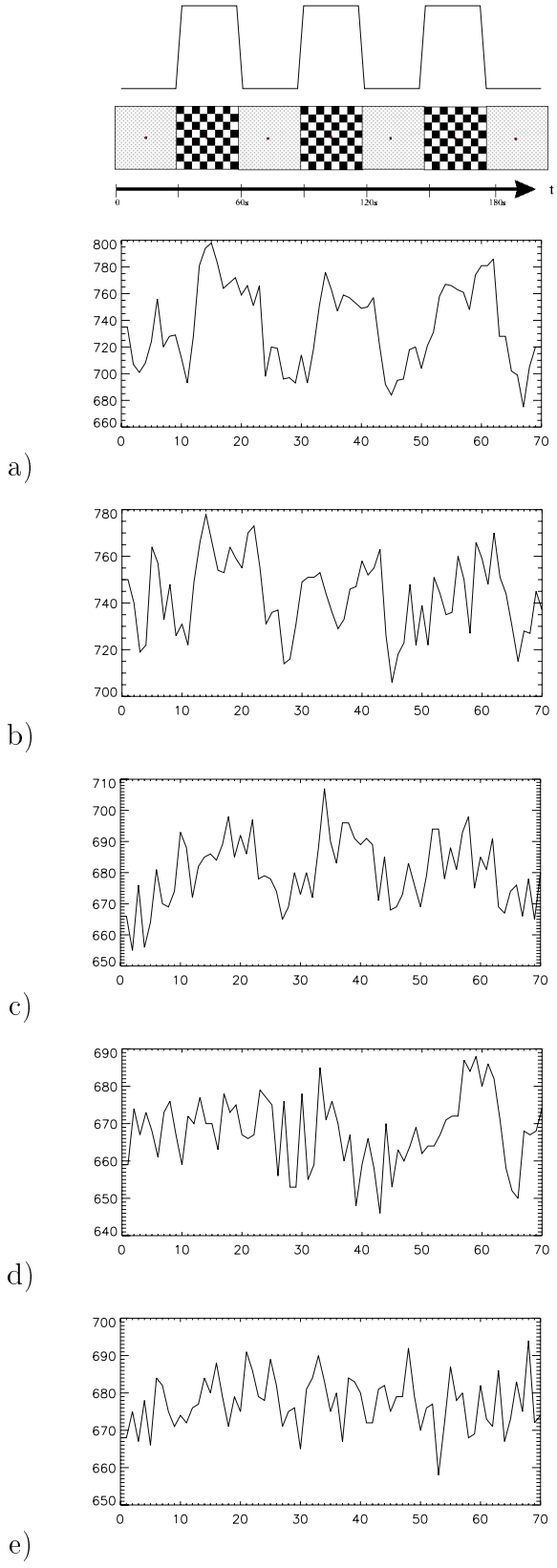
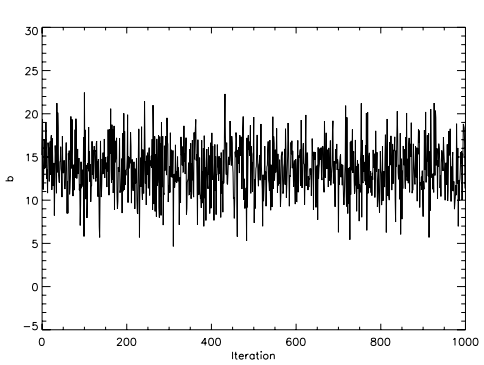
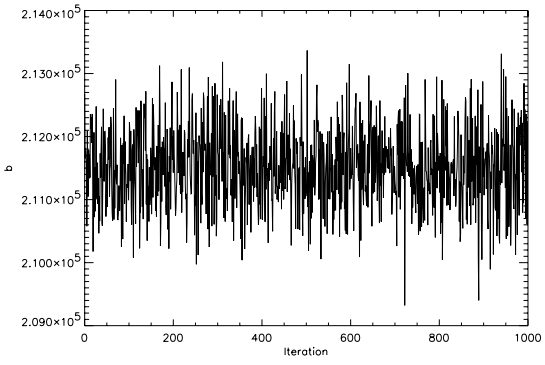


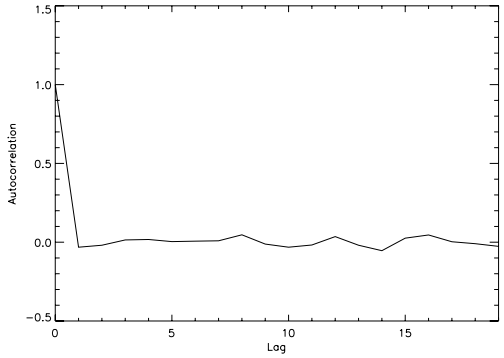
Figure 1: Stimulus paradigm and representative time series ranging from strongly activated (a,b) over weak activation (c,d) to no activation (e).



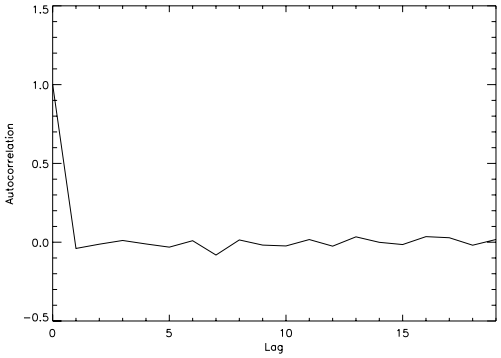
a) sampling path (selected parameter)



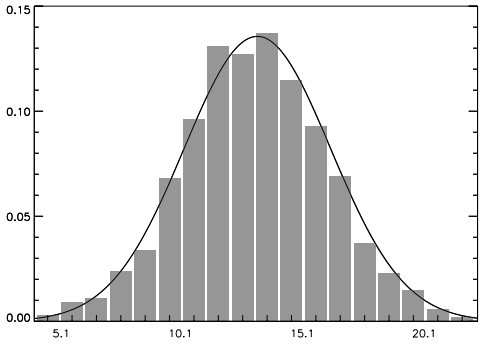
b) sampling path (deviance)



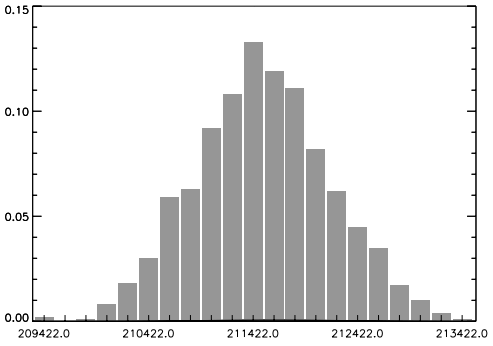
c) autocorrelation (selected parameter)



d) autocorrelation (deviance)

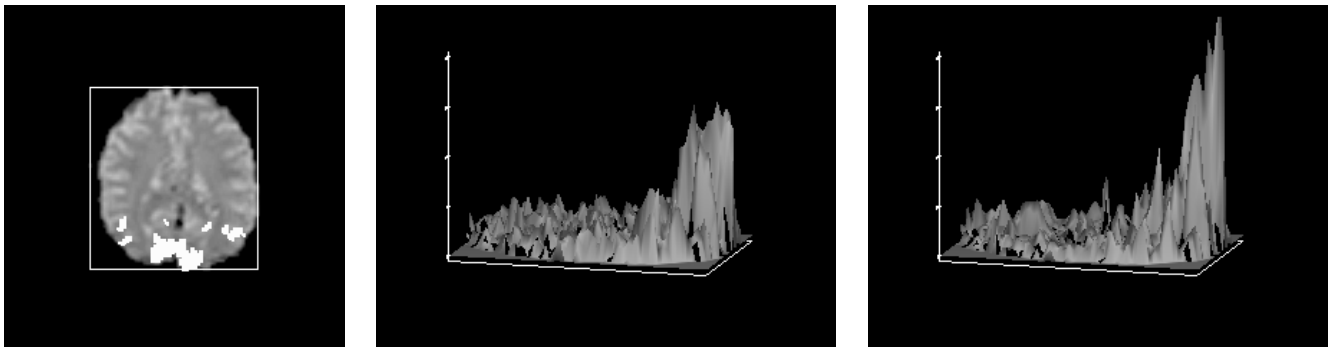


e) histogram (selected parameter)



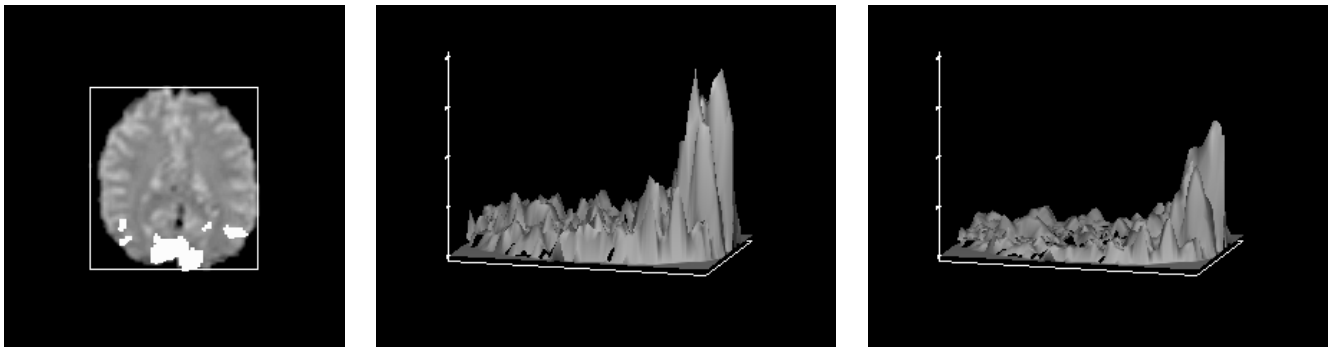
f) histogram (deviance)

Figure 2: MCMC diagnostics.



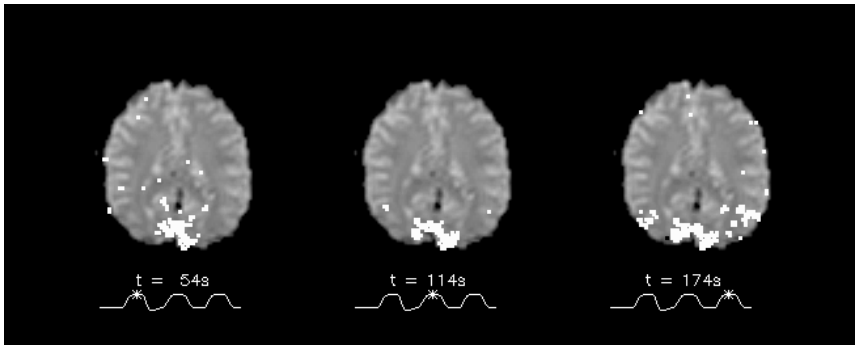
a) activation map b) standardized values c) absolute values

Figure 3: Results for the pixelwise parametric model.

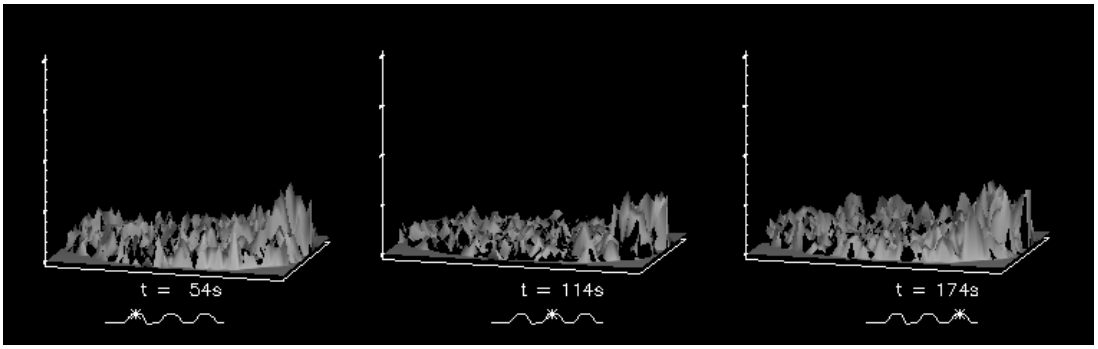


a) activation map b) standardized values c) absolute values

Figure 4: Results for the spatial parametric model.

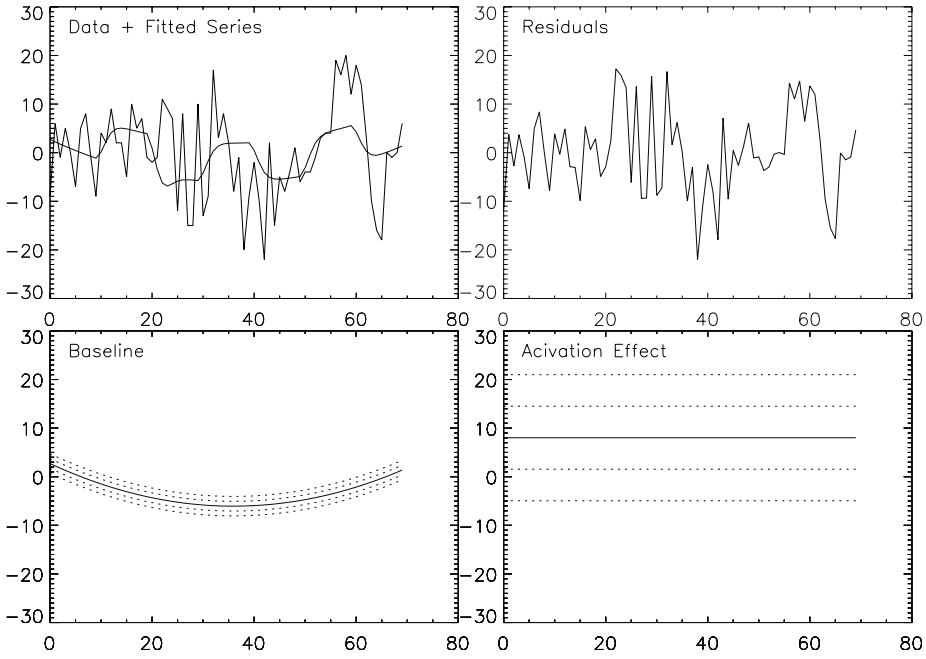


a) activation map

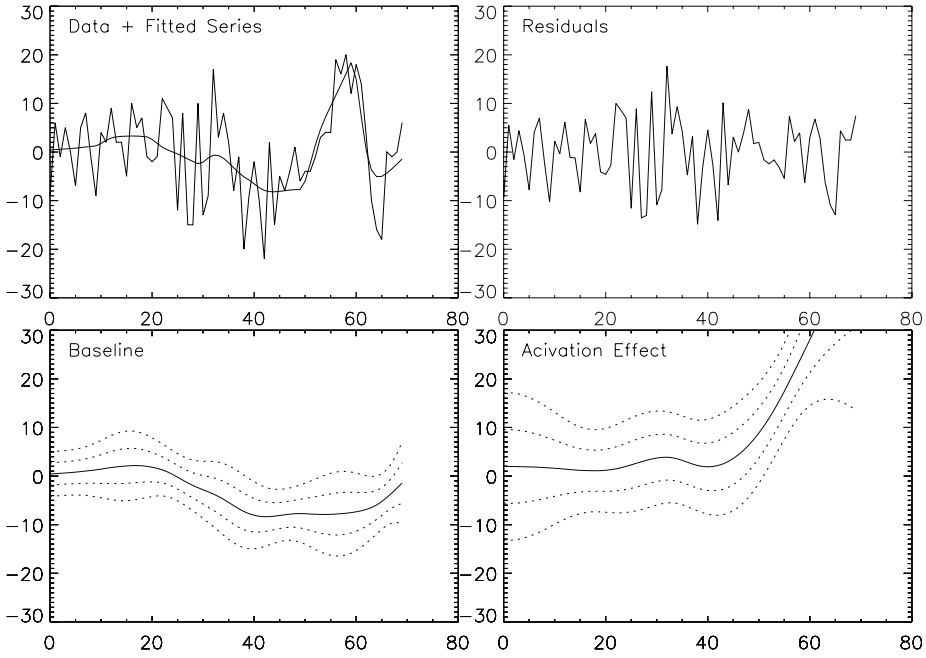


b) standardized values

Figure 5: Results for the pixelwise semiparametric model.

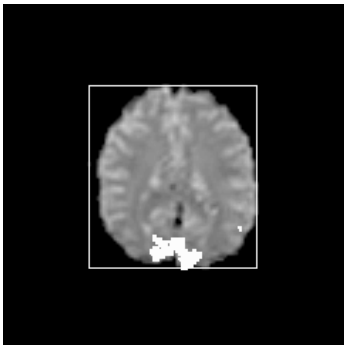


a) parametric model

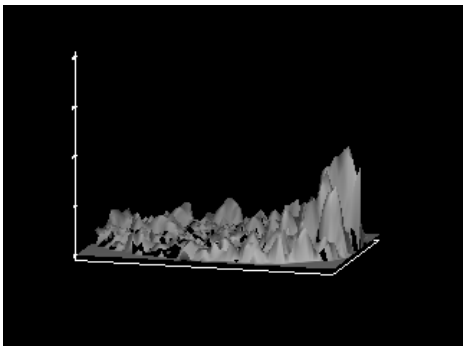


b) dynamic model

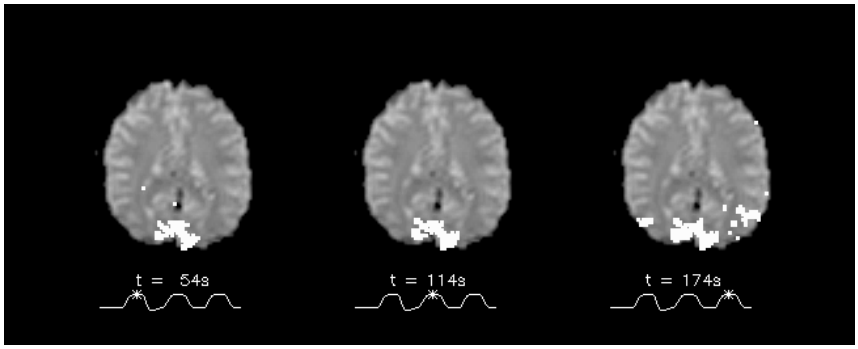
Figure 6: Fitted models.



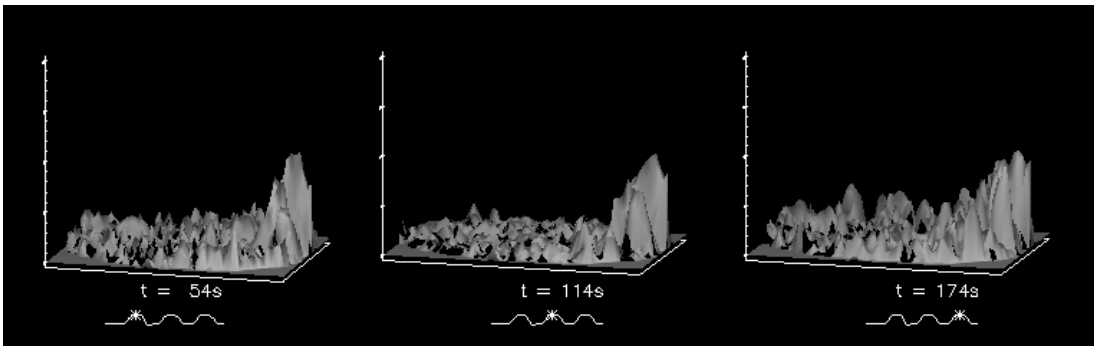
a) activation map (spatial effect α)



b) standardized values (spatial effect α)

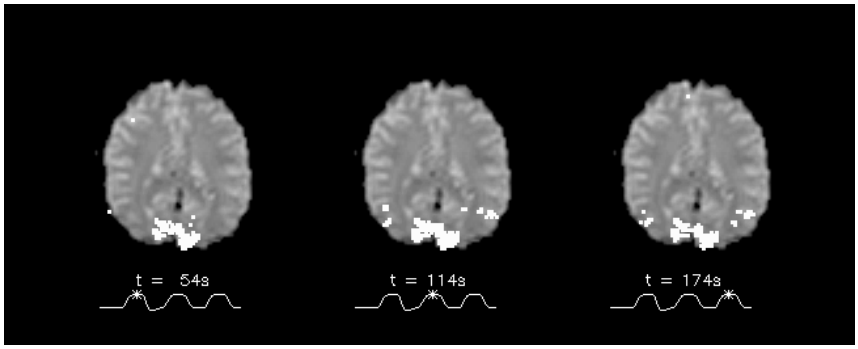


c) activation map (activation effect $\alpha + \beta$)

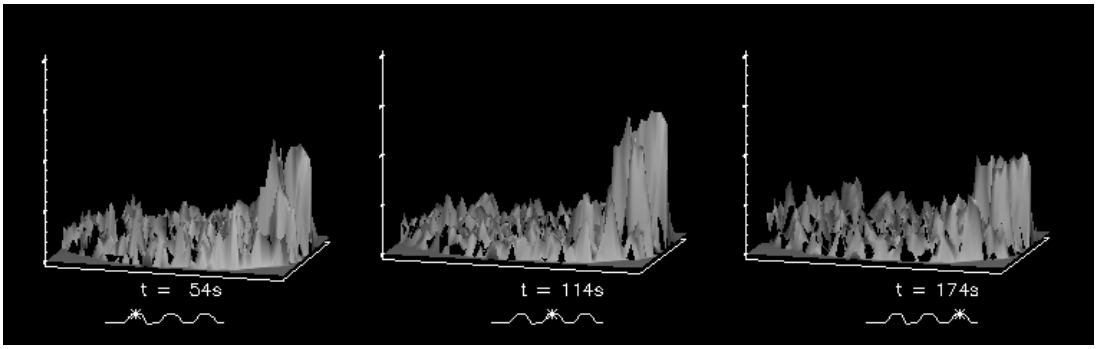


d) standardized values (activation effect $\alpha + \beta$)

Figure 7: Results for the separable spatio-temporal model.



a) activation map



b) standardized values

Figure 8: Results for the non-separable spatio-temporal model.