Brezger, Fahrmeir, Hennerfeind:

Adaptive Gaussian Markov Random Fields with Applications in Human Brain Mapping

Sonderforschungsbereich 386, Paper 456 (2005)

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

Projektpartner

Adaptive Gaussian Markov Random Fields with Applications in Human Brain Mapping

A. Brezger, L. Fahrmeir and A. Hennerfeind
Department of Statistics, Ludwig–Maximilians–University of Munich
Ludwigstr. 33, 80539 Munich, Germany
e–mail: andreas.brezger@stat.uni-muenchen.de
fahrmeir@stat.uni-muenchen.de
andreah@stat.uni-muenchen.de

Abstract

Functional magnetic resonance imaging (fMRI) has become the standard technology in human brain mapping. Analyses of the massive spatio–temporal fMRI data sets often focus on parametric or nonparametric modeling of the temporal component, while spatial smoothing is based on Gaussian kernels or random fields. A weakness of Gaussian spatial smoothing is underestimation of activation peaks or blurring of high–curvature transitions between activated and non–activated brain regions. In this paper, we introduce a class of inhomogeneous Markov random fields (MRF) with spatially adaptive interaction weights in a space–varying coefficient model for fMRI data. For given weights, the random field is conditionally Gaussian, but marginally it is non–Gaussian. Fully Bayesian inference, including estimation of weights and variance parameters, is carried out through efficient MCMC simulation. An application to fMRI data from a visual stimulation experiment demonstrates the performance of our approach in comparison to Gaussian and robustified non–Gaussian Markov random field models.

Key words: Adaptive weights, human brain mapping, inhomogeneous Markov random fields, MCMC, space–varying coefficient models, spatio–temporal modeling.

1 Introduction

Functional magnetic resonance imaging (fMRI) is the current standard technology in human brain mapping, i.e., the detection of regions activated by sensory, motor and cognitive functions. In fMRI experiments, a subject is exposed to controlled external stimuli. Local increase of neural activity is indicated by a local increase of blood oxygenation in activated areas, and this BOLD (blood oxygenation level dependent) effect can be visualized by fMRI. In classical experiments the stimulus is presented in a boxcar paradigm, i.e., a sequence of OFF and ON periods. The scanner records images of several slices of the brain. Each slice is about 5 mm thick and consists of 128×128 pixels. Slices usually have a distance of several millimeters, and their images are obtained sequentially in time. Therefore, slices are often analyzed separately. For each pixel of a slice, an fMRI experiment with a boxcar stimulus generates an MR signal time series, with an increase during the ON periods compared to the control or rest condition OFF. Our application in Section 4 analyzes data from a visual experiment. Figure 1 shows the boxcar stimulus and three MR time series of length 70, observed at three pixels, which are selected from the center of the activated region (b), near to its boundary (c), and from a non–activated region (d), respectively. Obviously, the activation effect of the stimulus on the MR signal is high in activated areas and is not present in non–activated areas.

To assess brain activity, separate regression models are applied at each pixel, with the MR signal as response and a transformed version of the stimulus as the regressor of primary interest. The value of the corresponding regression coefficient is considered as the "intensity" or "amplitude" of activation at the pixel. In the standard approach, spatial correlation between pixels is taken into account by stationary Gaussian random fields in a post–processing step (Friston et al., 1995), or it is incorporated as part of a spatio–temporal Bayesian hierarchical model based on Gaussian Markov random fields (Gössl et al., 2001). A potential weakness of Gaussian random field priors is underestimation of peaks and smoothing over edges, discontinuities or unsmooth parts of underlying functions.

In this paper, we combine separate pixelwise regressions through a Bayesian space–varying coefficient model. To enhance spatial adaptivity for the activation effects array, we introduce a new class of inhomogeneous or compound Markov random field priors where the interaction weights, determining the degree of spatial correlations between adjacent pixels, are allowed to vary stochastically. Conditional upon these weights, the prior is a Gaussian MRF, but marginally it is a non–Gaussian MRF with edge–preserving properties. All model parameters, including the adaptive interaction weights, are estimated in a fully Bayesian setting using MCMC

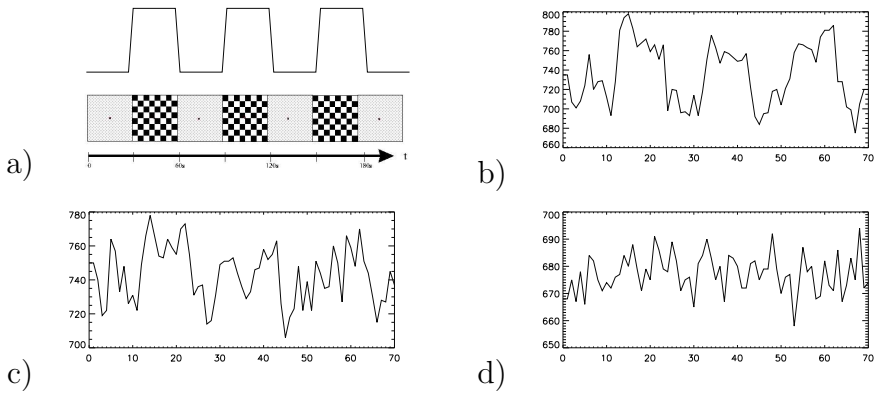

Figure 1: Visual fMRI : a) an 8 Hz flickering rectangular checkerboard (ON) is presented to the subject alternating every 30 s with an uniformly dark background and a fixation point (OFF); an experiment consists of 4 OFFs and 3 ONs; additionally, representative MR signal time courses from strongly (b), weak (c) and non–activated (d) pixels.

techniques. Efficient sampling from Gaussian MRFs as described in Rue (2001) and Rue and Held (2005) is combined with MH steps for updating interaction weights. Details of the sampling scheme are given in the Appendix. In Section 4, we apply our approach to fMRI data from a visual stimulation experiment and compare its performance with results obtained with commonly used (unweighted) Gaussian MRFs as well as with spatially more adaptive Laplace and Geman–Reynolds priors. We conclude that our spatially adaptive modelling approach performs well and should be a promising alternative in other applications too.

2 Space–varying coefficient models in fMRI

We start this section with a brief description of the standard approach for analyzing fMRI data. In its basic form, fMRI time series $\{y_{it}, t = 1, \ldots, T\}$ at each pixel or voxel i – visualized for three pixels in Figure 1b), c), d), – are fitted separately by a time series regression model. In its basic form, such a regression model for pixel i is defined through the measurement model

$$y_{it} = \mathbf{u}_t' \boldsymbol{\alpha}_i + z_{it}\beta_i + \varepsilon_{it}, \qquad \varepsilon_{it} \sim N(0, \sigma_i^2) \tag{1}$$

for $t = 1, \ldots, T$. The term $\mathbf{u}_t'\boldsymbol{\alpha}_i$ models a smooth baseline trend, where the vector $\mathbf{u}_t$ consists of a few simple functions, for example (piecewise) polynomials or some terms of a Fourier expansion, evaluated at $t = 1, \ldots, T$. The second term, sometimes called activation profile, is the product of the "activation effect" β_i at pixel i, and the covariate z_{it}, which is a smoothed and delayed version of the original ON–OFF–stimulus $x_t, t = 1, \ldots, T$, visualized in Figure 1a).

This transformation takes into account that the cerebral blood flow, the source of the MR signal, increases only approximately 6-8 s after the onset of the stimulus, and that flow responses do not occur suddenly, but more continuously and delayed. We will use transformations obtained by a delayed convolution with a so–called hemodynamic response function, i.e.,

$$z_{it} = \sum_{s=0}^{t-d_i} h(s; \theta_i) x_{t-d_i-s}. \tag{2}$$

Usually, Poisson ($\mathrm{Po}(\lambda_i)$) or Gamma ($\mathrm{GA}(\lambda_i, u_i)$) densities are chosen for h. The parameters $\theta_i = \lambda_i$ or $\boldsymbol{\theta}_i = (\lambda_i, u_i)'$ and the time lag d_i are estimated in a pilot step, see Gössl et al. (2001).

The idea is that the (estimated) activation effect $\hat{\beta}_i$ should be large in strongly activated pixels (Fig. 1b)), medium for weakly activated pixels (Fig. 1c)), and close to zero ('non–significant') for non–activated pixels (Fig. 1d)). The estimates $\{\hat{\beta}_i, i = 1, \ldots, I\}$ for all pixels in a slice can be represented by and "activation map" or, after standardization, by a "t–map" $\{t_i = \hat{\beta}_i / se(\hat{\beta}_i), i = 1, \ldots, I\}$. In the standard SPM approach (see Friston et al., 1995), spatial correlation induced by neighboring pixels is taken into account (if at all) by applying Gaussian random field theory to the t–map, leading to an adjusted t–map. An adaptive smoothing procedure for pixelwise activation maps with good edge–preserving properties is suggested in recent work by Tabelow et al. (2005).

The basic model (1) has been refined by relaxing assumptions on the time series structure, for example by introducing autoregressive error terms, or using flexible regression splines (Genovese, 2000) and state space models (Gössl et al., 2000), for nonparametric estimation of baseline trends and even time–varying activation effects. Still, however, fMRI data are fitted separately at each pixel i based on (seemingly) unrelated regression or time series models.

In this paper we look at the measurement model (1) as a joint model for all pixels $i = 1, \ldots, I$. Then the coefficients $\{\boldsymbol{\alpha}_i, i = 1, \ldots, I\}$ and $\{\beta_i, i = 1, \ldots, I\}$ are spatially varying over the grid of pixels in a certain slice. To allow for activation effects to vary smoothly with time during the fMRI experiment, we slightly extend

model (1) to

$$y_{it} = \mathbf{u}_t'\boldsymbol{\alpha}_i + z_{it}\mathbf{v}_t'\boldsymbol{\beta}_i + \varepsilon_{it}, \qquad \varepsilon_{it} \sim N(0, \sigma_i^2), \tag{3}$$

for $t = 1, \ldots, T$ and all pixels $i = 1, \ldots, I$. In addition to a time–varying baseline trend $\mathbf{u}_t'\boldsymbol{\alpha}_i$, model (3) also admits the activation effect $\mathbf{v}_t'\boldsymbol{\beta}_i$ at pixel i to vary over time, or, in other words, a time–varying activation profile. Model (3) can be seen as a joint spatio–temporal regression model for fMRI data, with space–varying activation coefficients $\{\beta_i, i = 1, \ldots, I\}$. Spatial models for the coefficients will be based on Markov random fields. In Gössl et al. (2001) a hierarchy of spatial and spatio–temporal Bayesian models based on homogeneous Gaussian Markov random field smoothness priors has been presented and applied to fMRI data. A drawback of Gaussian priors is that they tend to over–smooth peaks and to blur edges or areas of high curvature between activated and non–activated areas. To avoid these problems, we suggest Markov random field priors with spatially adaptive interaction weights for adjacent pixels.

3 Spatially adaptive Markov random field priors

We will assume independent MRF priors for the components $\{\beta_{ik}, i = 1, \ldots, I\}$ of the space–varying coefficient vectors $\{\boldsymbol{\beta}_i, i = 1, \ldots, I\}$ in (3). For notational convenience, we focus on a scalar component $\{\beta_i, i = 1, \ldots, I\}$ as in the simpler model (1) with a time–constant activation effect. Multivariate MRFs for vectors $\boldsymbol{\beta}_i$ are conceivable, but will not be considered here.

The usual form of the prior for the vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_i, \ldots, \beta_I)'$ is a pairwise interaction Markov random field (MRF)

$$p(\boldsymbol{\beta}|\tau, \boldsymbol{w}) \propto \tau \cdot \exp\{-\sum_{i \sim j} w_{ij}\Phi(\tau(\beta_i - \beta_j))\}, \tag{4}$$

where τ is a scale or precision parameter, Φ is symmetric with $\Phi(u) = \Phi(-u)$, the summation is over all pairs of pixels $i \sim j$ that are neighbors, and the interaction weights w_{ij}'s are known or given. For regular grids as in image analysis, the weights are all set equal to 1, i.e. $w_{ij} = 1$. For $\Phi(u) = \frac{1}{2}u^2$, these are the most popular (intrinsic) Gaussian MRFs, see Rue and Held (2005) for a comprehensive presentation. The Laplace prior with $\Phi(u) = |u|$ is considered to have improved edge–preserving properties as well as the Geman–Reynolds prior

$$\Phi(u) = -\lambda/(1 + |u|^p), \tag{5}$$

with $p = 2$ or $p = 1$, and λ as a tuning parameter, introduced by Geman and Reynolds (1992). A practical problem with this prior is the appropriate choice of hyperparameters by data driven methods. The normalizing constant $c(\lambda, p)$ is analytically intractable, making inclusion into a fully Bayesian MCMC difficult. In the context of reconstruction in emission tomography, Higdon et al. (1997) propose a simulation method for precomputing normalizing constants on a grid of values. Conceptually, this approach might be adapted to our problem, but the computational burden increases dramatically.

Usually the weights in (4) are specified deterministically, e.g. by setting them equal to one for regular grids or by measuring the distance between neighboring sites in irregular lattices, see Besag et al. (1995). As an alternative, we suggest a conditionally Gaussian MRF prior

$$\beta_i|\beta_{-i}, \tau, \boldsymbol{w} \sim N\left(\sum_{j\in\partial_i} \frac{w_{ij}\beta_j}{w_{i+}}, \frac{1}{\tau^2 w_{i+}}\right), \tag{6}$$

where the interaction weights w_{ij} are allowed to vary stochastically in a further stage of the hierarchical model. The precision (or inverse variance) τ^2 acts as a smoothing parameter. Based on experience in the one–dimensional case of locally adaptive function estimation (Lang et al., 2002), we specify the weights to be i.i.d. random variables following a Gamma hyperprior

$$w_{ij} \sim GA\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \tag{7}$$

with small degrees of freedom, in particular $\nu = 1$. The resulting marginal or compound MRF prior $p(\boldsymbol{\beta}|\tau^2)$ is non–Gaussian and does not admit a simple closed analytical form. It gives additional flexibility when pixel i is near the border of an activated area, where some neighbors $j \in \partial_i$ have similar activation effects and others may be only weakly or not activated. At first glance, it seems more natural to assume a spatial prior for the field of weights w_{ij}. Own experience in the closely related situation of one–dimensional locally adaptive function estimation shows, however, that i.i.d. priors combine computational advantages with desirable edge–preserving properties. Spatial priors are only marginally better for highly–oscillating functions, while computational efforts increase drastically.

The Bayesian specification is completed by priors for the variances σ_i^2 and the precision parameter τ. We follow the common choice and assume weakly inverse Gamma hyperpriors $\sigma_i^2 \sim IG(a, b)$ for observation variances, and $\tau^2 \sim GA(c, d)$ for Gauss and adaptive Gauss priors, and $\tau \sim GA(c, d)$ for the Laplace prior. As a standard option, we set $a = b = c = d = 0.001$, yielding highly dispersed hyperpriors.

In principle, the same spatial priors could be chosen for the baseline parameters $\boldsymbol{\alpha}_i = \{\alpha_{ik}, k = 1, \ldots, \dim(\boldsymbol{\alpha}_i)\}$. However, because the focus is placed on the activation effect, we assign only separate, independent diffuse priors $p(\alpha_{ik}) \propto const$ or highly dispersed normal priors for each pixel $i = 1, \ldots, I$ and each $k = 1, \ldots, \dim(\boldsymbol{\alpha}_i)$.

Gathering parameters in vectors $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_I)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_I)$, where $\boldsymbol{\beta}_i = \{\beta_{ik}, k = 1, \ldots, \dim(\boldsymbol{\beta}_i)\}$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_I^2)$, $\boldsymbol{\tau}^2 = (\tau_1^2, \ldots, \tau^2_{\dim(\boldsymbol{\beta}_i)})$, $\boldsymbol{w} = (w_{ij}, i \sim j)$ and observations in $\mathbf{Y} = (y_{it}, i = 1, \ldots, I, t = 1, \ldots, T)$, fully Bayesian inference is based on the posterior

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}^2, \boldsymbol{w}|\mathbf{Y}) \propto L(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) \cdot p(\boldsymbol{\alpha}) \cdot p(\boldsymbol{\beta}|\boldsymbol{\tau}^2, \boldsymbol{w}) \cdot p(\boldsymbol{\sigma}^2) \cdot p(\boldsymbol{\tau}^2) \cdot p(\boldsymbol{w}).$$

The likelihood $L(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2)$ is determined by the observation model, the other factors by the priors, together with conditional independence assumptions.

Inference is performed by MCMC simulation through repeated drawings from univariate or multivariate full conditionals. The general strategy is as follows:

1. Draw parameters α_{ik} from the Gaussian full conditionals.
2. Draw the blocks $\boldsymbol{\beta}_k = \{\beta_{1k}, \ldots, \beta_{Ik}\}$ from the (multivariate) Gaussian full conditionals, i.e. given current iterates for the weights $\boldsymbol{w}$.
3. Draw the weights w_{ij} via the MH–step described in the Appendix.
4. Draw the variance parameters σ_i^2 and the hyperparameters $\boldsymbol{\tau}^2$ from their corresponding (inverse) Gamma full conditionals.

4 Application

We illustrate our approach based on adaptive Gaussian MRF priors (shortly adaptive Gauss) by application to an fMRI data set from a visual stimulation experiment as described in the introduction. Visual paradigms are known to elicit great activation amplitudes in the visual cortical areas, which are sharply separated from other functional areas.

In a first step we apply the parametric observation model (1) with time–constant activation effect β_i, where the transformed stimulus z_{it} was determined through a pilot estimate, see Gössl (2001, p.33) for details. The baseline trend was modeled parametrically by $\mathbf{u}_t'\boldsymbol{\alpha}_i = \alpha_{i0} + \alpha_{i1} \cdot t + \alpha_{i2} \cdot \sin(\pi/16 \cdot t) + \alpha_{i3} \cdot \cos(\pi/25 \cdot t) + \alpha_{i4} \cdot \cos(\pi/40 \cdot t)$. In a second step we apply observation model (3) with a time–varying activation

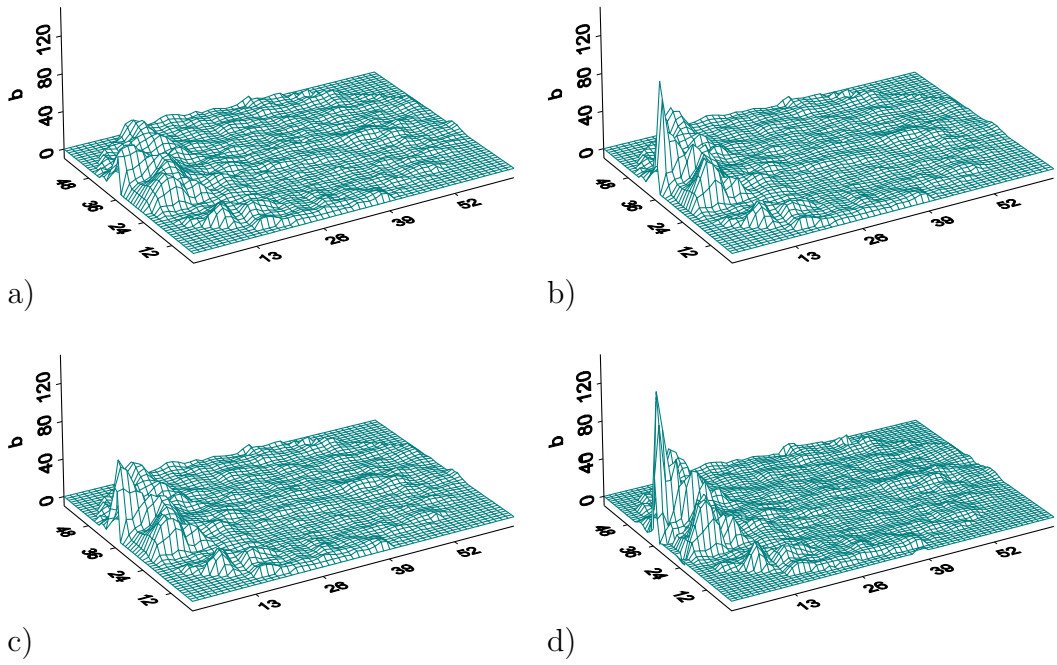

Figure 2: estimated surfaces for the time–constant model: a) Gauss, b) adaptive Gauss, c) Laplace, d) Geman–Reynolds

profile $z_{it}\mathbf{v}_t'\boldsymbol{\beta}_i$. We replace β_i by $\mathbf{v}_t'\boldsymbol{\beta}_i = \beta_{i0}+\beta_{i1}\cdot t+\beta_{i2}\cdot\cos(\pi/25\cdot t)+\beta_{i3}\cdot\cos(\pi/40\cdot t)$. All frequencies were selected through stepwise selection procedures.

For the time–constant model we estimated the activation surface $\{\hat{\beta}_i, i = 1,\ldots,I\}$ using the Gauss, adaptive Gauss, Laplace and a Geman–Reynolds prior. Inverse Gamma priors with $a = b = 0.001$ were chosen for the error variances σ_i^2 in all four cases, as well as for the variance parameter τ^2 in the case of Gauss and adaptive Gauss priors. Based on simulation results for a grid of values of the tuning parameters p, τ and λ of the Geman–Reynolds prior as well as on visual inspection for the data at hand we set them to $p = 2$, $\tau = 0.2$ and $\lambda = 3$.

Figures 2a)–2d) show posterior mean estimates $\{\hat{\beta}_i, i = 1,\ldots,I\}$ for these four models. Obviously, the Gauss prior oversmoothes the sharp peaks and ridges as well as steep slopes in the area of the central visual cortex (on the left side of the activation surface), while it undersmoothes in non–activated areas, resulting in a comparably rough estimated surface. The result for the Laplace prior shows better local adaptivity, but still oversmoothes the central activation area. Both the adaptive Gauss and the Geman–Reynolds prior exhibit the desired features: non–activated

areas are smooth and close to zero, but activated areas, in particular the distinct physiologically known peak in the central visual cortex remains. The adaptive Gauss prior has the advantage that hyperparameters can be estimated easily from the data. This is much more difficult for the Geman–Reynolds prior with its complicated normalizing constant.

This disadvantage becomes even more crucial when applying the time–varying model (3), because one would have to specify hyperparameters for each of the components of the parameter vector $\boldsymbol{\beta}_i$. Therefore we carry out comparative analyses for Gauss and adaptive Gauss priors only. With the same choice of inverse Gamma priors as in the time–constant case, we are able to estimate all space–varying parameters with the technique outlined in Section 3 and in the appendix. Figures 3a)–3f) show the estimated activation surfaces $\{z_{it}\mathbf{v}_t'\hat{\boldsymbol{\beta}}_i, i = 1, \ldots, I\}$ for two different points of time which refer to the first and third activation period, respectively. Comparing the smoothing qualities of the different priors, it can again be seen that the adaptive Gauss prior has distinctly better local adaptivity properties.

5 Conclusion

Detection of activation areas in the living human brain using fMRI data offers challenging problems in spatio–temporal modeling. In particular, there is a need for locally adaptive surface smoothers which still can cope with the massive data sets and high–dimensional parameters arrays in the setting of space–varying regression models from a computational viewpoint. Gaussian Markov random fields with adaptive interaction weights as developed in this paper seem to be a promising framework. Although our methodological development was motivated by fMRI data, it should also be of use in other spatial problems where local adaptivity and edge–preservation are required.

Acknowledgement: We thank Dorothee Auer and Christoff Gössl (Max–Planck–Institute of Psychiatry) for providing the data and for many discussions, motivating and guiding this work, and the German National Science Foundation for financial support through grants from the Sonderforschungsbereich 386 "Statistical Analysis of Discrete Structures".

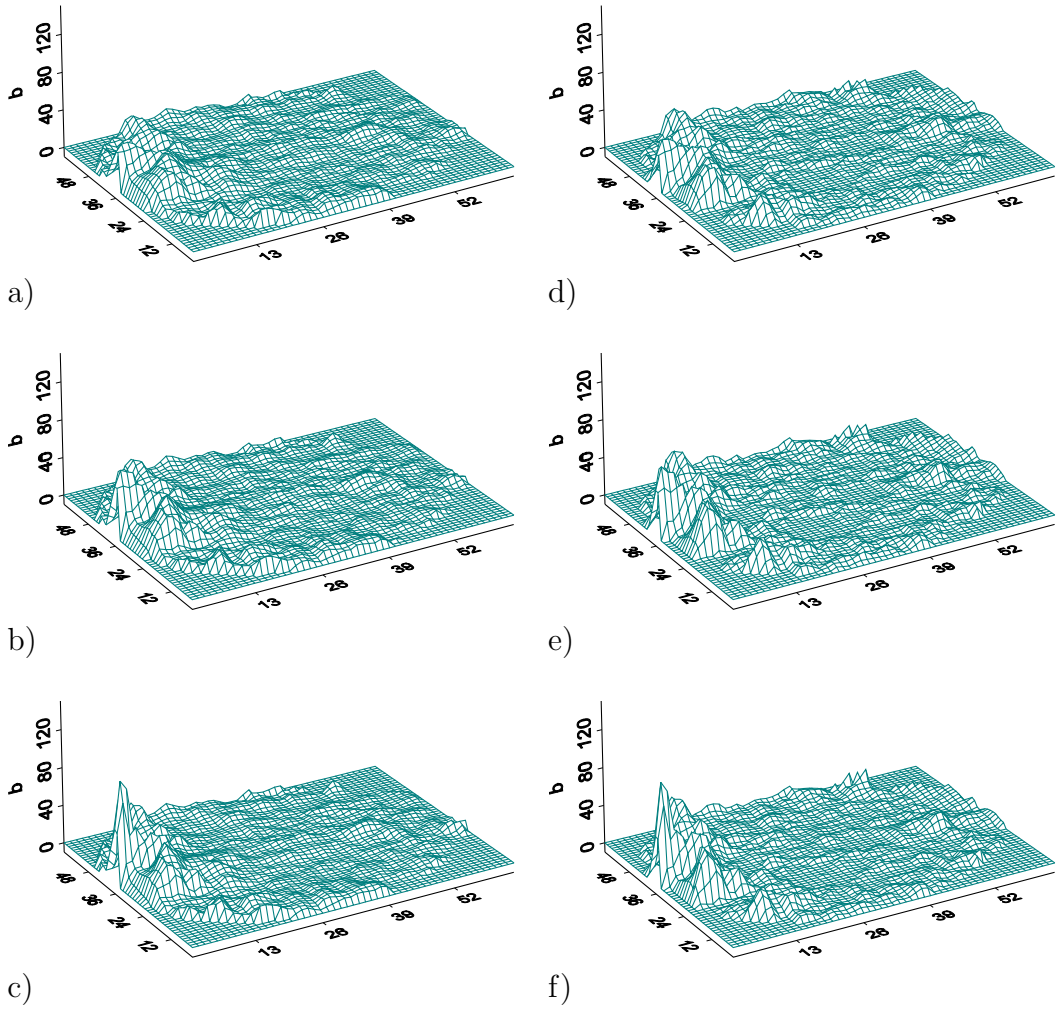

Figure 3: Estimated activation effects at $t = 18$ (a-c) and $t = 58$ (d-f) for Gauss (a,d), Laplace (b,e), and adaptive Gauss (c,f)

Appendix: Details on updating weights

Given the current iterates for the weights in the MCMC algorithm of Section 3, steps 1.), 2.) and 4.) are in complete analogy to estimation in spatial models with latent Gaussian MFRs, see for example Lang and Brezger (2004) and Rue and Held (2005). Thereby we make efficient use of sparse matrix operations for block updating in step 2.). In the following we focus on details of the MH–step 3.).

The full conditional for the weight $w_{ij}^{(k)}$ is

$$\begin{aligned} p(w_{ij}^{(k)}|\cdot) &\propto p(\boldsymbol{\beta}_k|\cdot)p(w_{ij}^{(k)}) \\ &\propto \left(\prod_{i=2}^{M^2}\lambda_i\right)^{1/2} (w_{ij}^{(k)})^{\frac{\nu}{2}-1} \exp\left\{-w_{ij}^{(k)}\left(\frac{\nu}{2}+\frac{(\beta_{ik}-\beta_{jk})^2}{2\tau_k^2}\right)\right\}. \end{aligned} \tag{8}$$

Here λ_i, $i = 2, \ldots, M^2$, are the non–zero eigenvalues of the penalty matrix $\boldsymbol{K}_k(\boldsymbol{w}^{(k)})$ corresponding to the joint prior $p(\boldsymbol{\beta}_k|\boldsymbol{w}^{(k)}, \tau_k^2)$ derived from the conditional prior (6). We explicitly denote the penalty matrix by $\boldsymbol{K}_k(\boldsymbol{w}^{(k)})$ to emphasize its dependency on the weights $\boldsymbol{w}$. Note that (8) is a Gamma $GA\left(a'_{w_{ij}^{(k)}}, b'_{w_{ij}^{(k)}}\right)$ density with parameters

$$a'_{w_{ij}^{(k)}} = \frac{\nu}{2} \quad \text{and} \quad b'_{w_{ij}^{(k)}} = \frac{\nu}{2} + \frac{(\beta_{ik}-\beta_{jk})^2}{2\tau_k^2}, \tag{9}$$

multiplied by $\left(\prod_{i=2}^{M^2}\lambda_i\right)^{1/2}$. In order to sample from this distribution we employ a MH–step and use a Gamma distribution with the parameters in (9) as proposal density. Therefore the acceptance probability reduces to

$$\alpha = \min\left\{1, \left(\frac{\prod_{i=2}^{M^2}\lambda_i^*}{\prod_{i=2}^{M^2}\lambda_i}\right)^{1/2}\right\},$$

where the λ_i^* denote the non–zero eigenvalues of the penalty matrix $\boldsymbol{K}_k(w_{ij}^{*(k)})$ resulting from a proposed weight $w_{ij}^{*(k)}$. Acceptance rates are usually quite high. In our implementation we use the fact that

$$\frac{\prod_{i=2}^{M^2}\lambda_i^*}{\prod_{i=2}^{M^2}\lambda_i} = \frac{|\boldsymbol{K}_{11}^*|}{|\boldsymbol{K}_{11}|}, \tag{10}$$

where $|\boldsymbol{K}_{11}|$ and $|\boldsymbol{K}^*_{11}|$ denote the determinant of the sub–matrices of $\boldsymbol{K}_k(w_{ij}^{(k)})$ and $\boldsymbol{K}_k(w_{ij}^{*(k)})$, respectively, where the last row and the last column is deleted. The advantage arising from (10) is that, instead of an expensive computation of eigenvalues of order $O(n^3)$, the ratio can be obtained by the computationally much more efficient Cholesky decomposition of band matrices, which is of order $O(n)$. Additionally, we exploit the fact that it is sufficient to start the Cholesky decomposition in the row corresponding to the position where a proposed new weight is located. Block updating of several weights in one step speeds up computation considerably, since the ratio (10) has to be evaluated only once per block. Since the proposal densities $p(w_{ij}^{(k)})$ are independent no further difficulties are imposed by sampling from a block of weights. In our application to human brain mapping in Section 4 joint updating of 6 weights still yields an acceptance rate $> 50\%$. A proof of (10) can be found in Brezger (2005).

References

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995) Bayesian Computation and Stochastic Systems. *Statistical Science*, **10**, 3–66.

Brezger, A. (2005) Bayesian P–Splines in Structured Additive Regression Models. *Dr. Hut Verlag*, Dissertation, Univ. of Munich.

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.

Geman, D. and Reynolds, G. (1992) Constrained Restoration and the Recovery of Discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14-3**, 367–383.

Genovese, C. R. (2000) A Bayesian Time–Course Model for Functional Magnetic Resonance Imaging Data (with discussion). *Journal of the American Statistical Association*, **95**, 691–719.

Gössl, C. (2001) Bayesian Models in functional Magnetic Resonance Imaging: Approaches for Human Brain Mapping. *Shaker Verlag*, Dissertation, Univ. of Munich.

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

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

Higdon, D. M., Johnson, V. E., Bowsher, J. E., Turkington, T. G., Gilland, D. R., Jaszczack, R. J. (1997) Fully Bayesian estimation of Gibbs hyperparmaters for emission tomography data. *IEEE Transactions on Medical Imaging*, **16**, 516–526.

Lang, S. and Brezger, A. (2004) Bayesian P–splines, *Journal of Computational and Graphical Statistics*, **13**, 183–212.

Lang, S., Fronk, E.–M. and Fahrmeir, L. (2002) Function estimation with locally adaptive dynamic models, *Computational Statistics*, **17**, 479–500.

Rue, H. (2001) Fast sampling of Gaussian Markov random fields with applications. *Journal of the Royal Statistical Society B*, **63**, 325–338.

Rue, H. and Held, L. (2005) Gaussian Markov Random Fields; Theory and Applications, *CRC Press/Chapman and Hall*.

Tabelow, K., Polzehl, J. and Spokoiny, V. (2005) Analysing fMRI experiments with structural adaptive smoothing procedures. WIAS preprint No. 1079.