Bayesian mapping of brain regions using compound Markov random field priors

By L. FAHRMEIR

Department of Statistics, Ludwig-Maximilians-University of Munich, Ludwigstr. 33, 80539 Munich, Germany
e-mail: fahrmeir@stat.uni-muenchen.de

C. GÖSSL

NMR Study Group, Max-Planck-Instiute of Psychiatry, Kraepelinstr. 2–10, 80804 Munich, Germany

and A. HENNERFEIND

Department of Statistics, Ludwig-Maximilians-University of Munich, Germany

Summary

Human brain mapping, i.e., the detection of functional regions and their connections, has experienced enormous progress through the use of functional magnetic resonance imaging (fMRI). The massive spatio–temporal data sets generated by this imaging technique impose challenging problems for statistical analysis. Many approaches focus on adequate modeling of the temporal component. Spatial aspects are often considered only in a separate postprocessing step, if at all, or modeling is based on Gaussian random fields. A weakness of Gaussian spatial smoothing is possible underestimation of activation peaks or blurring of sharp transitions between activated and non–activated regions. In this paper we suggest Bayesian spatio–temporal models, where spatial adaptivity is improved through inhomogeneous or compound Markov random field priors. Inference is based on an approximate MCMC technique. Performance of our approach is investigated through a simulation study, including a comparison to models based on Gaussian as well as more robust spatial priors in terms of pixelwise and global MSEs. Finally we demonstrate its use by an application to fMRI data from a visual stimulation experiment for assessing activation in visual cortical areas.

Some key words: Adaptive smoothing, Bayesian inference, human brain mapping, inhomogeneous Markov random fields, MCMC, spatio–temporal modeling.

1. Introduction

Functional magnetic resonance imaging (fMRI) has led to enormous progress in human brain mapping, i.e., the detection of regions activated by sensory, motor and cognitive functions. Compared to older techniques, such as positron emission tomography, fMRI has improved spatial and far better temporal resolution, and it is completely non–invasive. In fMRI experiments, a subject is exposed to controlled external stimuli. Local increase of neural activity is indicated by

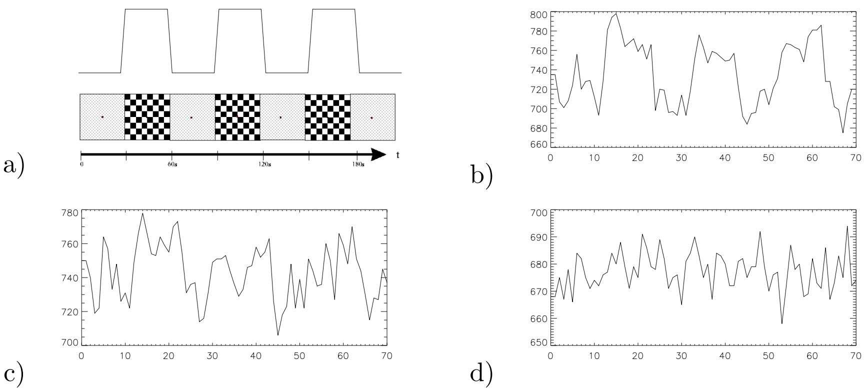

Fig. 1. Visual fMRI : a) an 8 Hz flickering rectangular checkerboard (ON) is presented to the subject alternating every 30 s with an uniformly dark background and a fixation point (OFF); an experiment consists of 4 OFFs and 3 ONs; additionally, representative MR signal time courses from strongly (b),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

a local increase of blood oxygenation in actived 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 often have a distance of several millimeters, and their images are obtained sequentially in time. Therefore, slices are usually 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 actived region (b), near to its boundary (c), and from a non–actived 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, a regression model is 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. Spatial correlation between pixels is often neglected or is taken into account by Gaussian random fields in a post–processing step (Friston et al., 1995) or as part of a Bayesian hierarchical model (Gössl, Auer and Fahrmeir, 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 therefore suggest semiparametric Bayesian spatio–temporal models which avoid underestimation of peaks or blurring of sharp transitions between activated and non–activated regions in activation surfaces. In our experience, time–variation of baseline trends and activation effects is relatively smooth, so that we use simple parametric forms for temporal components as described above. In contrast, activation effects are spatially inhomogeneous or unsmooth. Therefore, estimation of activation surfaces should be based on robust spatial priors. We suggest the use of inhomogeneous or compound Markov random field priors where the interaction weights, determining the degree of spatial variation between neighboring pixels, are allowed to vary stochastically. All model parameters, including hyper– or tuning parameters are estimated in a fully Bayesian setting, using MCMC techniques. For computational reasons, however, the exact MH updating step for stochastic interaction weights is replaced by an approximating Gibbs step. The quality of this ad hoc approximation is investigated in simulation studies and turns out to be satisfactory, but a theoretical justification is still missing.

The rest of the paper is organized as follows. The next section introduces semiparametric spatio–temporal models with compound MRF priors and outlines MCMC estimation. In Section 3, we investigate performance in simulation studies with artificial data generated from stylized activation surfaces, and compare it to results obtained with the Laplace or the Geman–Reynolds priors. An application to fMRI data from a visual stimulation experiment is presented in Section 4, and the concluding section discusses some possibilities for further development.

2. Semiparametric spatio–temporal inference for fMRI data

Conventional standard analyses of fMRI data comprise pixelwise correlation, regression or time series analysis. Spatial correlation between neighboring pixels is considered only in a second step, if at all. In a regression setup, the MR signal $\{y_{it}, t = 1, \dots, T\}$ at pixel $i, i = 1, \dots, I$, is decomposed into the sum

$$y_{it} = \mathbf{a}_i' \mathbf{w}_t + b_i z_{it} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim N(0, \sigma_i^2) \tag{1}$$

of a parametric baseline trend $a_{it} = \mathbf{w_t}' \mathbf{a_i}$, an activation profile $z_{it} b_i$ and measurement errors. The vector $\mathbf{w_t}$ consists of a few simple functions of time, such as (piecewise) polynomials or the first terms of a Fourier expansion, evaluated at $t = 1, \dots, T$. The activation profile is the product of the activation effect b_i at pixel i, and the covariate z_{it} is a transformation of the original ON–OFF–stimulus $x_t, t = 1, \dots, T$.

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.

More flexible pixelwise fitting of baseline trends by semiparametric extensions of the basic model (1) has been suggested in Genovese (2000), using regression splines, and by Gössl, Auer and Fahrmeir (2000), who apply state space models. 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 this, we suggest a model with the following ingredients: Given a fixed pixel, the observation model is a slight extension of model (1), allowing the activation profile to vary over time during an fMRI experiment. In a second stage, regression parameters of neighboring pixels are correlated through inhomogeneous Markov random field priors. Inference using MCMC techniques is outlined in Section 2.2. An important issue is the trade–off between model complexity and computational feasibility. There are 64x59 voxels per slice, observed over 70 time points, resulting in a huge number of observations (264320).

2.1. Spatio–temporal models with inhomogeneous Markov random field priors

For each pixel i, the observation model is defined by slightly extending model (1) to

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

In addition to a time–varying baseline trend $a_{it} = \mathbf{w_t'}\boldsymbol{\alpha}_\mathbf{i}$, model (3) also admits the activation effect $b_{it} = \mathbf{v_t'}\boldsymbol{\beta}_\mathbf{i}$ at pixel i to vary over time, or, in other words, a time–varying activation profile $b_{it}z_{it}$. We assume here that the transformed stimulus z_{it} is available from a pilot estimation step, for example using the hemodynamic response model (2). As a special case, with $v_t \equiv 1$, we get back the conventional model (1). Some extensions of the parametric pixelwise model are indicated in the discussion.

Spatially adaptive smoothing is based on Markov random fields in a second stage of the hierarchy. We consider a family of priors for the space–varying coefficients $\boldsymbol{\beta}_i$, with the potential to avoid blurring edges or oversmoothing of high curvatures between activated and non–activated areas of the brain. Similar spatial priors could also be assigned to the baseline parameters $\alpha_\mathbf{i}$. Compared to the additional computational effort, however, there is no gain in the estimation of activation effects, which is of primary substantive interest. Therefore, diffuse independent priors $p(\alpha_{ik}) \propto \text{const}$, $i = 1, \ldots, I$, $k = 1, \ldots, \dim(\boldsymbol{\alpha}_i)$ are assigned to them.

In the following, we focus on a scalar component $\beta_i, i = 1, \ldots, I$, of the activation effect. Extensions to vectors are possible by assuming independent priors for its components or by multivariate versions of the following. The general 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)

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

where τ is a scale 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 w_{ij}'s are corresponding weights. It is assumed that the conditional distributions

$$p(\beta_i|\beta_{-i}, \tau, \mathbf{w}) \propto \tau \cdot exp\{-\sum_{j\in\partial_i} w_{ij}\Phi(\tau(\beta_i - \beta_j))\}$$

are well defined, where $\partial_i = \{j : j \sim i\}$.

If the weights 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, these are the well-known MRF priors described, e.g., in Besag et al. (1995). In the following we will also admit the w_{ij}'s to be random variables obeying a hyperprior. Then the prior $p(\boldsymbol{\beta}|\tau)$ is a mixture of pairwise interaction priors or a *compound* MRF prior. Such a prior 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.

For $\Phi(u) = \frac{1}{2}u^2$, this yields

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

Setting $w_{ij} = 1$ for regular grids, this reduces to the traditional *Gaussian priors* used in Gössl et al. (2001).

A compound or inhomogeneous Gauss MRF is obtained when the weights w_{ij} are allowed to vary stochastically in a further stage. Here, the w_{ij}'s are specified to be *i.i.d.* random variables following a Gamma hyperprior

$$w_{ij} \sim GA(\frac{\nu}{2}, \frac{\nu}{2}) \cdot \tag{4}$$

The resulting compound distribution $p(\beta_i|\beta_{-i}, \tau^2)$ is a *Student prior* with ν degrees of freedom. For $\nu = 1$ this results in a *Cauchy prior.* However, MCMC simulation will make explicit use of the latent Gamma prior (4).

For $\Phi(u) = |u|$, we obtain

$$p(\beta_i|\beta_{-i}, \tau, \mathbf{w}) \propto \tau \cdot exp\{-\tau \sum_{j \in \partial_i} w_{ij}|\beta_i - \beta_j|\} \cdot \tag{5}$$

For $w_{ij} = 1$, this is the traditional *Laplace prior*, considered to have improved edge-preserving properties compared to the Gaussian prior (see Besag, York and Mollie (1991)). It can be interpreted as a stochastic version of the median filter. Assuming again a Gamma hyperprior (4) for the w_{ij}'s, we define a *compound Laplace prior.*

In general, members of this family are specified by appropriate choices of Φ and of the hyperprior for the weights. For tomographic images, Geman and Mc Clure (1987) and Geman and Reynolds (1992) propose

$$\Phi(u) = -\lambda/(1 + |u|^p), \tag{6}$$

with $p = 2$ or $p = 1$, and λ as a tuning parameter. Further, Green (1990) suggests $\Phi(u) = \lambda c_1 \log\cosh(c_2 u)$, where c_i are chosen to match (6) closely. Künsch (1994) studies edge preserving posterior mode smoothers using Φ–functions proposed in robust statistics, for example truncated Gaussian or Huber priors. All these robust priors have been introduced in the related but simpler context of image analysis on regular lattices, thus assuming $w_{ij} = 1$ for all weights. A practical problem with these priors 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.

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 (inverse) Gamma hyperpriors $\sigma_i^2 \sim IG(a, b)$ for observation variances, and $\tau^2 \sim GA(c, d)$ for Gauss and Cauchy priors and $\tau \sim GA(c, d)$ for the Laplace prior. As a standard option, we set $a = c = 1$ and $b = d = 0{\cdot}005$, yielding highly dispersed hyperpriors.

In principle, the same spatial priors could be chosen for the baseline parameters α_{ik}. 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 $k = 1, \ldots, \dim(\boldsymbol{\alpha}_i)$.

a) b) c)

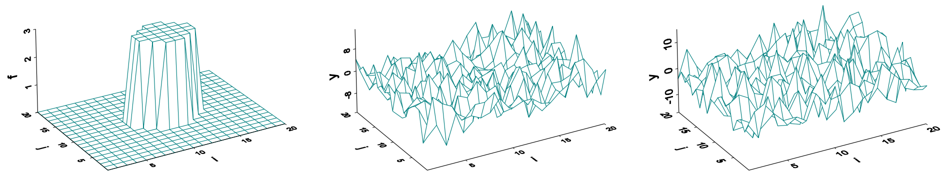

Fig. 2. a) activation area used for simulation study, b) observations for $t = 1$, c) observations for $t = 11$

2.2. Bayesian inference via MCMC

Gathering parameters in vectors $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_I)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_I)$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_I^2)$, $\mathbf{w} = (w_{ij}, i \sim j)$ and observations in $\mathbf{Y} = (y_{it}, i = 1, \ldots, I, t = 1, \ldots, T)$, fully Bayesian inference is based on the posterior

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

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. Focussing on the parametric observation model (3), the general strategy is as follows: The parameters $\alpha_{ik}, \beta_i, \sigma_i^2$ and τ are drawn from their corresponding full conditionals separately for each pixel. Vectors can be sampled univariately or as blocks. For Gaussian priors with known weights w_{ij} full conditionals are normals, Gamma or inverse Gamma distributions so that Gibbs sampling is possible. For non–Gaussian priors computationally more demanding Metropolis' steps are necessary for updating spatially correlated parameters. For compound Gaussian and Laplace priors with stochastic interaction weights w_{ij}, we replace the exact MH updating step for the weights by an approximating Gibbs step. We provide some details in the Appendix. These Gibbs steps are computationally much faster, and simulation studies support satisfactory quality of the approximation. A theoretical analysis is missing, however.

3. Simulation studies

A number of simulation experiments were carried out to compare different spatial smoothers and to explore their properties.

In a first step, we investigated pure spatial smoothing using Gauss, Laplace, Cauchy (=compound Gauss with $\nu = 1$) and compound Laplace priors. Data were generated according to Gaussian observation models $y_i = \mu_i + \varepsilon_i, i = 1, \ldots, I$ on regular two–dimensional lattices. For the true surfaces $\{\mu_i, i = 1, \ldots, I\}$ we chose two–dimensional step functions as in Künsch (1994), cylinders over a plane, and more smooth structures such as quadratic surfaces or planes. For unsmooth structures, and even for a quadratic surface, smoothing with the compound Laplace prior had the smallest mean square error, followed by the compound Gauss, the Laplace and the Gauss prior. For planes, the loss of efficiency when using compound priors instead of a Gaussian was quite small. Details are described in Hennerfeind (2000).

a) b) c)

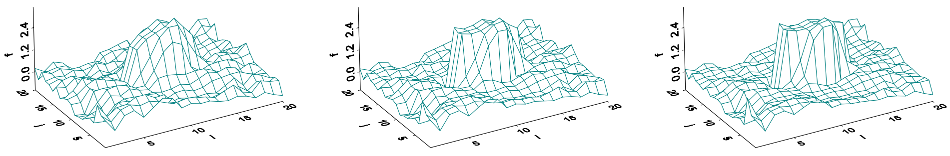

Fig. 3. Estimated surfaces a) Gauss, b) Cauchy, c) compound Laplace

a) b) c)

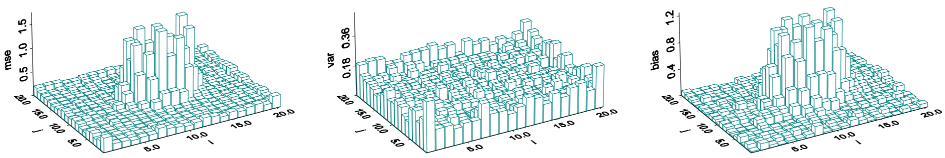

Fig. 4. Gauss a) MSE, b) variance, c) absolute bias

a) b) c)

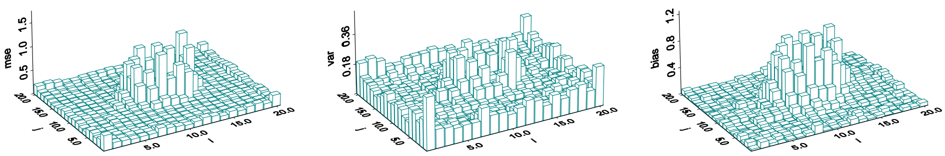

Fig. 5. Cauchy a) MSE, b) variance, c) absolute bias

a) b) c)

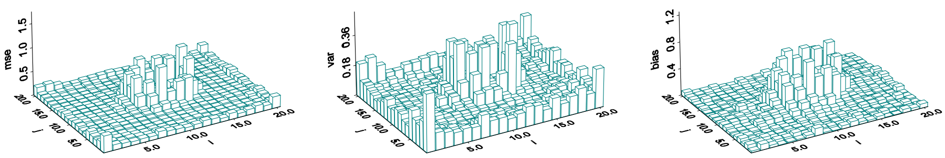

Fig. 6. compound Laplace a) MSE, b) variance, c) absolute bias

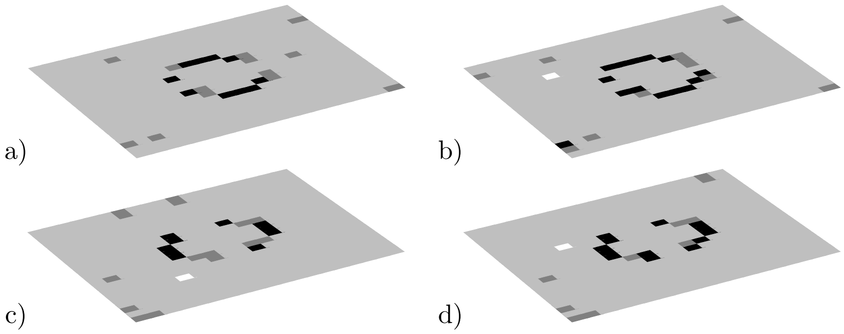

Fig. 7. Estimated weights w_{ij} a),b) horizontal, c),d) vertical, a),c) Cauchy, b),d) compound Laplace

In the following simulation experiment, spatio–temporal data were generated according to a stylized structure resembling fMRI data. The surface $f(\cdot)$ in Figure 2a stylizes an activation area by a cylinder. Its height is the "activation effect". Outside of the cylinder, activation is zero. For each pixel i, data were generated by

$$y_{it} = f(i) \cdot z_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma_i^2), \quad t = 1, \ldots, 70,$$

where z_{it} is 0-1-stimulus, with $z_{it} = 0$ for $t = 1, \ldots, 10$, $z_{it} = 1$ for $t = 11, \ldots, 20$, etc. To achieve a realistically low signal-to-noise ratio, we set $\sigma_i^2 = 25 + 2 \cdot N(0, 1)$. For the simulation study, spatio–temporal data $y_{it}^{(s)}$ where generated in this way for each simulation run $s = 1, \ldots, 50$. For a typical data set, Figures 2b and 2c display the observations for $t = 1$ ($z_{it} = 0$) and $t = 11$ ($z_{it} = 1$). It is more or less impossible to recover the surface by eye. Aim of the simulation study was to compare the performance of spatial smoothers with Gauss, compound Gauss (=Cauchy), compound Laplace and Geman–Reynolds priors (6). Figures 3a–c show the average of the posterior mean estimates of the surface using a Gauss, Cauchy and compound Laplace prior with the standard option $a = c = 1$, $b = d = 0{\cdot}005$ for highly dispersed hyperpriors for observation variances and smoothing parameters. Compared to the Gauss prior, the compound Gauss and Laplace priors have improved edge preserving properties. This is also confirmed by Figures 4 a–c, 5 a–c and 6 a–c, which show the MSE, variance and absolute bias at each pixel, and by corresponding values averaged over all pixels (see Table 1). It can be seen that improved MSE performance is primarily caused by reducing bias and less by reducing variance. Moreover, for the compound priors the variance is higher near the edges of the (stylized) activation surface. This is a consequence of correspondingly decreased weights w_{ij} for neighbors near the border of the activation surface. A smaller weight w_{ij} makes larger jumps of the surface between pixel i and j more likely, but - as can be seen from full conditionals for variances - implies an increase in variance. Figure 7 visualizes smaller values of weights by dark grey or black colored pixels, for both compound priors and for weights between horizontal and vertical neighbors, respectively. It is seen, that the jump in the surface is quite well reflected by a decrease of corresponding weights. For the Geman–Reynolds prior, a straightforward fully data driven choice of the hyperparameters λ, τ and p is not possible. Therefore, surface estimation was carried out for various combinations of λ and τ, separately for $p = 1$ and $p = 2$ (see Table 1 for some selected results). It is shown that performance in terms of global MSE, variance and absolute bias is very sensitive with respect to these hyperparameters. Figures 8a and 8b show average surface estimates for a "good" and a "bad" choice of λ and τ. Obviously, a "good" choice is crucial and can be achieved

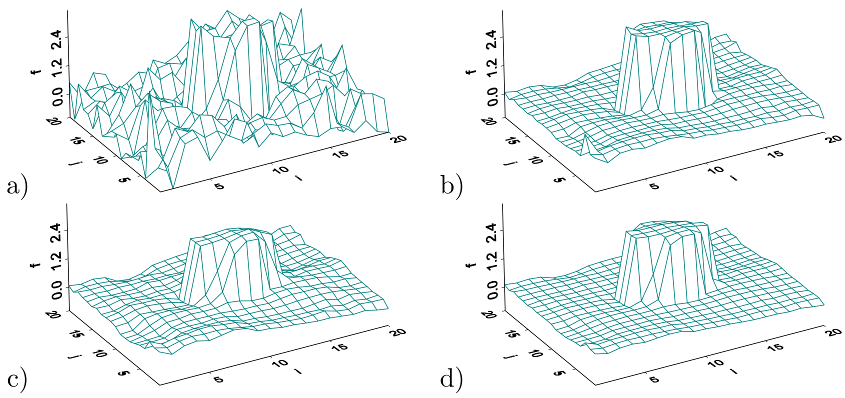

Fig. 8. Estimated surfaces a) Geman-Reynolds with $p = 2$, $\lambda = 1$, $\tau = 1$, b) Geman-Reynolds with $p = 2$, $\lambda = 3$, $\tau = 2$, c) Cauchy with c=100 d=10, d) compound Laplace with c=100 d=20

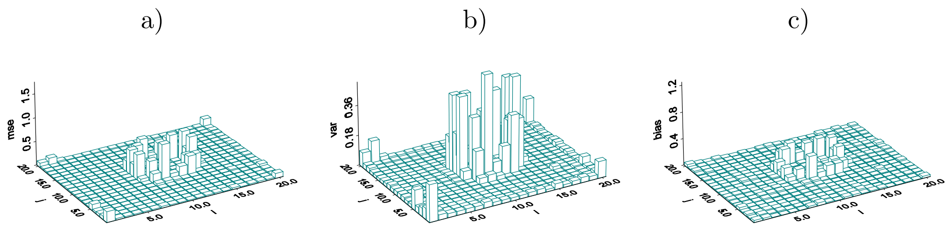

Fig. 9. Geman-Reynolds with $\lambda = 3$, $\tau = 2$ a) MSE, b) variance, c) absolute bias

when the "true" scene is known. For a fair comparison, we also experimented with a number of different more informative hyperpriors for the variance τ^2 or scale τ of the compound priors. Figures 8c and 8d show estimation results for $c = 100$, $d = 10$ and $c = 100$, $d = 20$, giving additional improvements compared to Figures 3b and 3c. Figures 9a–c and Figures 10a–c display MSE, variance and absolute bias at each pixel for the "best" Geman–Reynolds prior and for the compound Laplace prior.

The conclusion is as follows: If we know the "true" scene or we have a good idea of it through designed experiments, Geman–Reynolds priors with carefully selected parameters perform quite well. Our results show that the compound Laplace prior with a carefully chosen hyperprior for smoothing can lead to comparable performance in recovering the surface. In real applications to individual human brain mapping however, there is no "true" activation surface which is known, at least approximately. In this situation, compound Gauss or Laplace priors (with highly dispersed hyperpriors) are preferable, because they make a straightforward fully Bayesian choice of smoothness parameters and automatic surface smoothing possible. In applications, as in the following, one may first apply spatial smoothing with compound Gauss or Laplace priors and, in a second step, try to obtain improvements with a Geman–Reynolds or a more informative compound Laplace prior.

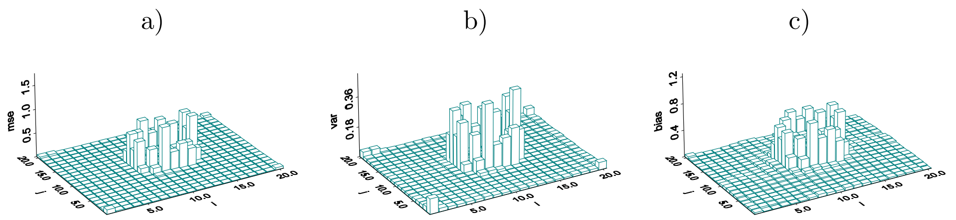

Fig. 10. compound Laplace with c=100, d=20 a) MSE, b) variance, c) absolute bias

prior	c	d	MSE	variance	|bias|
Gauß	1	0.005	0.199 (0.253)	0.112 (0.033)	0.148 (0.254)
Cauchy	1	0.005	0.145 (0.168)	0.102 (0.049)	0.109 (0.179)
compound Laplace	1	0.005	0.109 (0.130)	0.086 (0.065)	0.080 (0.129)
Gauß	100	10	0.334 (0.845)	0.019 (0.007)	0.300 (0.475)
Cauchy	100	10	0.104 (0.233)	0.036 (0.039)	0.133 (0.224)
compound Laplace	100	20	0.048 (0.129)	0.025 (0.056)	0.075 (0.131)
prior	λ	τ	MSE	variance	|bias|
G-R, p=1	1	1	0.538 (0.135)	0.526 (0.133)	0.088 (0.068)
G-R, p=1	3	1	0.132 (0.117)	0.123 (0.095)	0.057 (0.077)
G-R, p=1	5	1	0.048 (0.111)	0.034 (0.062)	0.057 (0.106)
G-R, p=1	6	1	0.046 (0.127)	0.027 (0.064)	0.067 (0.119)
G-R, p=1	4	2	0.040 (0.098)	0.034 (0.076)	0.033 (0.067)
G-R, p=2	1	1	0.408 (0.132)	0.398 (0.128)	0.079 (0.062)
G-R, p=2	2	1	0.154 (0.121)	0.145 (0.101)	0.059 (0.073)
G-R, p=2	4	1	0.078 (0.158)	0.050 (0.069)	0.085 (0.148)
G-R, p=2	5	1	0.146 (0.351)	0.055 (0.093)	0.157 (0.258)
G-R, p=2	3	2	0.041 (0.096)	0.038 (0.083)	0.023 (0.051)
G-R, p=2	4	3.33	0.233 (0.665)	0.109 (0.300)	0.195 (0.294)
G-R, p=2	8	0.33	0.175 (0.283)	0.078 (0.025)	0.158 (0.270)

Table 1. *global MSE, variance and absolute bias for different priors and different choices for the hyperparameters*

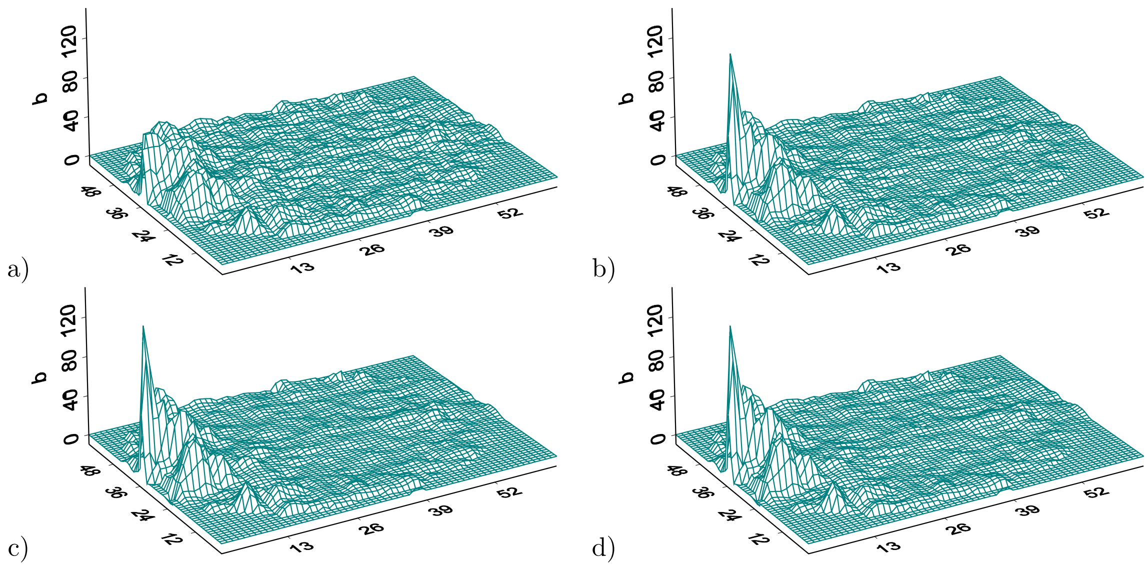

Fig. 11. estimated surfaces for the time-constant model: a) Gauss, b) Cauchy, c) compound Laplace, d) Geman-Reynolds

4. Application

We illustrate our approach 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 modelled parametrically by $a_{it} = \alpha_{i0} + \alpha_{i1} \cdot t + \alpha_{i2} \cdot \sin(\pi/16 \cdot t) + \alpha_{i3} \cdot \cos(\pi/25 \cdot t) + \alpha_{i4} \cdot \cos(\pi/40 \cdot t)$. In a second step we apply observation model (3) with a time–varying activation profile $b_{it}z_{it}$. We replace β_i by $b_{it} = \beta_{i0} + \beta_{i1} \cdot t + \beta_{i2} \cdot \cos(\pi/25 \cdot t) + \beta_{i3} \cdot \cos(\pi/40 \cdot t)$. 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, Cauchy and compound Laplace prior as well as the Geman-Reynolds prior with $p = 2$. Parameters of the inverse Gamma prior were set to $a = 1$ and $b = 0{\cdot}005$ and parameters of the spatial priors were set to $c = 1$ and $d = 0{\cdot}005$, $\lambda = 3$ and $\tau = 0{\cdot}2$, respectively. Every MCMC algorithm consisted of 30000 iterations, with the first 10000 being discarded as burn-in and every 20th iteration included in the final sample. Convergence was checked visually by sampling paths. For the neighborhoods in the spatial priors, the four nearest neighbors were chosen. Figures 11a-11d 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 results for the Cauchy, compound Laplace and Geman-Reynolds priors clearly illustrate improved smoothing and edge-preserving properties. Although these three surfaces are very similar, we prefer the Cauchy and compound

Laplace models, because they allow us to use non-informative priors without precomputation of normalizing constants. As the result for the Geman-Reynolds prior is sensitive to the choice of the parameters λ and τ, one has to be very careful when setting them and compare the results for different choices.

This disadvantage becomes even more crucial when applying the time-varying model (3), because one would have to specify two hyperparameters for each of the components of the parametervector β_i. Due to the effort it would take to compare a sufficiently large number of estimated activation surfaces, that result from different choices for the eight hyperparameters, we do not consider the Geman-Reynolds prior with fixes hyperparameters to be appropriate for this model. Besides the already mentioned necessity to precompute normalizing constants a fully Bayesian estimation with the Geman–Reynolds prior would have the disadvantage that Metropolis steps (instead of Gibbs steps) with eight appropriate proposals would be needed for updating the hyperparameters. Therefore we only use the Gauss, Cauchy and compound Laplace prior. With $\boldsymbol{\tau} = (\tau_0, \ldots, \tau_3)$ we get for the conditional distribution

$$p(\beta_{ik}|\beta_{-ik}, \boldsymbol{\tau}, \mathbf{w}) \propto \tau_k \cdot exp\{-\sum_{j \in \partial_i} w_{ij}^{(k)} \Phi(\tau_k(\beta_{ik} - \beta_{jk}))\}, \quad k = 0, \ldots, 3\cdot$$

With $\tau_k \sim GA(c, d), c = 1, d = 0{\cdot}005$ and $w_{ij}^{(k)} \sim GA(1/2, 1/2)$ we are able to estimate all parameters with the approximate MCMC technique outlined in Section 2.2 and in the appendix. Figures 12a-12f show the estimated activation surfaces $\{\hat{b}_{it}, i = 1, \ldots, I\}$ 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 clearly seen that the Cauchy and compound Laplace prior lead to reduced oversmoothing of edges and peaks.

5. Conclusion

Detection of activation areas in the living human brain using fMRI data offers challenging problems in biostatistical modelling. Simulation results and real applications show that regression models with compound Markov random field priors for activation effects have spatially adaptive smoothing properties as needed in this situation, and, at the same, keep a reasonable balance between model complexity and computational feasibility for analyzing massive fMRI data sets. From a theoretical point of view, the approximate Gibbs updating step for interaction weights has to be investigated analytically. As an alternative, the exact MH step might be implemented. Further developments to be considered in future research are compound MRFs with spatial priors for the weights, as proposed by Aykroyd (1998) in the simpler context of conventional image analysis, and the incorporation of substantive prior knowledge from human brain templates.

Acknowledgement: We thank Dorothee Auer (Max-Planck-Instiute of Psychiatry) for many discussions, motivating and guiding our work, and the German National Science Foundation for financial support through grants from the Sonderforschungsbereich 386 "Statistical Analysis of Discrete Structures".

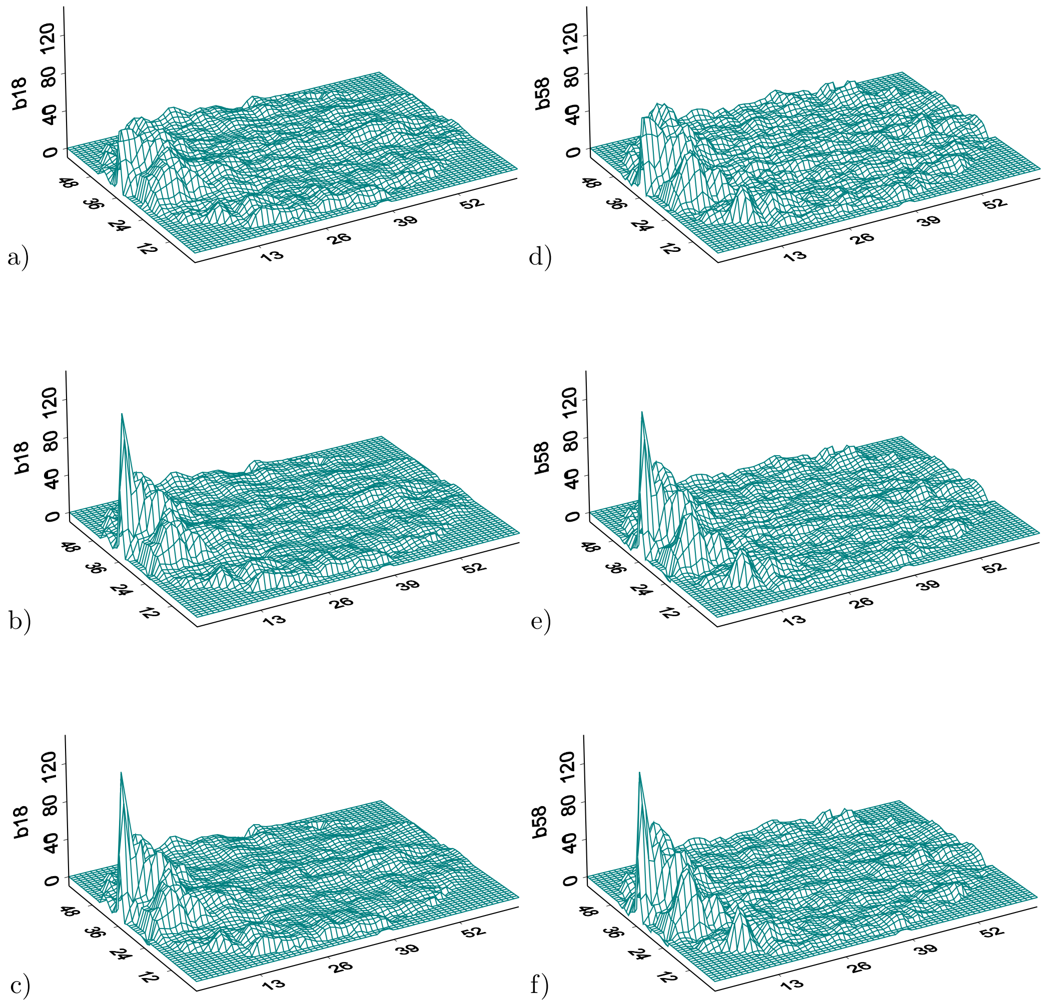

Fig. 12. Estimated activation effects at $t = 18$ (a-c) and $t = 58$ (d-f) for Gauss (a,d), Cauchy (b,e), and compound Laplace (c,f)

Appendix

In the following we provide some details on the estimation of the parameters when using the compound Laplace prior. To simplify the notation we only write down the posterior distributions for α_{i0} and β_{i0} (the posterior distributions of the other parameters are calculated analogously) and rewrite the observation model (3) as follows:

$$y_{it} = \alpha_{i0} + a_{it} + \beta_{i0} z_{it} + b_{it} z_{it} + \varepsilon_{it}.$$

In our application $a_{it} = \alpha_{i1} \cdot t + \alpha_{i2} \cdot \sin(\pi/16 \cdot t) + \alpha_{i3} \cdot \cos(\pi/25 \cdot t) + \alpha_{i4} \cdot \cos(\pi/40 \cdot t)$ and $b_{it} = \beta_{i1} \cdot t + \beta_{i2} \cdot \cos(\pi/25 \cdot t) + \beta_{i3} \cdot \cos(\pi/40 \cdot t)$.
With a diffuse prior $p(\alpha_{i0}) \propto$ const the full conditional distributions of the parameters $\alpha_{ik}, i = 1, \ldots, I, k = 0, \ldots, 4$ are normal. Therefore these parameters can be estimated via Gibbs sampling. For α_{i0} we get:

$$\begin{aligned}
&p(\alpha_{i0}|\cdot) \propto \prod_t p(y_{it}|\cdot) \cdot p(\alpha_{i0}) \\
&\Rightarrow \alpha_{i0}|\cdot \sim N\left(\mu_{\alpha_{i0}}, \sigma^2_{\alpha_{i0}}\right) \\
&\mu_{\alpha_{i0}} = \frac{\sum\limits_t (y_{it} - (a_{it} + \beta_{i0} z_{it} + b_{it} z_{it}))}{\sum\limits_t 1} \\
&\sigma^2_{\alpha_{i0}} = \frac{\sigma_i^2}{\sum\limits_t 1}
\end{aligned}$$

The parameters $\beta_{ik}, i = 1, \ldots, I, k = 1, \ldots, 3$ are estimated via a single–move Metropolis–Hastings algorithm using random walk proposals. For the posterior distribution of β_{i0} we get:

$$\begin{aligned}
p(\beta_{i0}|\cdot) &\propto \prod_t p(y_{it}|\cdot) \cdot p\left(\beta_{i0}|\tau_0, \beta_{-i0}, w^{(0)}\right) \\
&\propto exp\left(-\frac{1}{2\sigma_i^2}\left(\sum_t (y_{it} - (\alpha_{i0} + a_{it} + \beta_{i0} z_{it} + b_{it} z_{it}))^2\right)\right) \cdot \\
&\quad exp\left(-\frac{1}{\tau_0}\sum_{j\in\delta_i} w_{ij}^{(0)} |\beta_{i0} - \beta_{j0}|\right) \\
&\propto exp\left(-\frac{1}{2\sigma_i^2}\left(\sum_t \left(-2y_{it}\beta_{i0} z_{it} + \beta_{i0}^2 z_{it}^2 + 2\beta_{i0} z_{it}\left(\alpha_{i0} + a_{it} + b_{it} z_{it}\right)\right)\right)\right) \\
&\quad \cdot exp\left(-\frac{1}{\tau_0}\sum_{j\in\delta_i} w_{ij}^{(0)} |\beta_{i0} - \beta_{j0}|\right)
\end{aligned}$$

The hyperparameters σ_i^2 and τ_k are updated via Gibbs sampling. The posterior distributions are as follows:

$$\sigma_i^2|\cdot \sim IG\left(a_i, b_i\right)$$

$$\begin{aligned} a_i &= a + \frac{T}{2} \\ b_i &= b + \frac{1}{2}\sum_t \left(y_{it} - \left(\alpha_{i0} + a_{it} + \beta_{i0}z_{it} + b_{it}z_{it}\right)\right)^2 \end{aligned}$$

$$\tau_k|\cdot \sim IG\left(c_k, d_k\right)$$

$$\begin{aligned} c_k &= c + (I-1) \\ d_k &= d + \sum_{i\sim j} w_{ij}^{(k)}|\beta_{ik} - \beta_{jk}| \end{aligned}$$

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

$$p(w_{ij}^{(k)}|\beta,\tau) \propto p(\beta_{ik}|\beta_{-ik},\tau_k,w^{(k)})p(w^{(k)}) \propto p(\beta_{ik}|\beta_{lk}, l \in \delta_i, \tau_k, w^{(k)})p(w^{(k)})\cdot$$

The conditional $p(\beta_{ik}|\beta_{lk}, l \in \delta_i, \tau_k, w^{(k)})$ is a multivariate Laplace distribution, and the normalizing constant is a (complicated) function of the weights. As an approximation, we condition only on the neighbor j of i. Then

$$p(w_{ij}^{(k)}|\beta,\tau) \propto p(\beta_{ik}|\beta_{jk},\tau_k,w_{ij}^{(k)})p(w_{ij}^{(k)})$$

and therefore

$$\begin{aligned} w_{ij}^{(k)}|\cdot &\overset{app\cdot}{\sim} GA\left(\nu_{1k}, \nu_{2k}\right) \\ \nu_{1k} &= \frac{1}{2} + 1 \\ \nu_{2k} &= \frac{1}{2} + \frac{1}{\tau_k}|\beta_{ik} - \beta_{jk}| \end{aligned}$$

We can view this Gibbs step as an approximate MH step, with $p(\beta_{ik}|\beta_{jk},\tau_k,w_{ij}^{(k)})p(w_{ij}^{(k)})$ as a proposal density, where we accept each proposal. An exact analytical investigation of this approximation would be desirable.

References

AYKROYD, R.G. (1998). Bayesian Estimation for Homogeneous and inhomogeneous Gaussian Random Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20-5**, 533–539.

BESAG J., GREEN P., HIGDON D. and MENGERSEN K. (1995). Bayesian Computation and Stochastic Systems. *Statistical Science*, **10**, 3–66.

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.

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 S. and MCCLURE D. E. (1987). Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute*, **LII-4**, 5-21.

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.

GREEN P. J. (1990). Bayesian Reconstructions From Emission Tomography Data Using a Modified EM Algorithm. *IEEE Transactions on Medical Imaging*, **9**, 84–93.

HENNERFEIND A. (2000). Bayesianische kantenerhaltende Glättung, Diplomarbeit, Dept. of Statistics, Univ. of Munich.

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.

KÜNSCH H. (1994). Robust Priors for Smoothing and Image Restoration. *Annals of the Institute of Statistical Mathematics*, **46 1**, 1–19.