

Adaptive spatial smoothing of fMRI images

YU (RYAN) YUE, JI MENG LOH AND MARTIN A. LINDQUIST*

It is common practice to spatially smooth fMRI data prior to statistical analysis and a number of different smoothing techniques have been proposed (e.g., Gaussian kernel filters, wavelets, and prolate spheroidal wave functions). A common theme in all these methods is that the extent of smoothing is chosen independently of the data, and is assumed to be equal across the image. This can lead to problems, as the size and shape of activated regions may vary across the brain, leading to situations where certain regions are under-smoothed, while others are over-smoothed. This paper introduces a novel approach towards spatially smoothing fMRI data based on the use of nonstationary spatial Gaussian Markov random fields (Yue and Speckman, 2009). Our method not only allows the amount of smoothing to vary across the brain depending on the spatial extent of activation, but also enables researchers to study how the extent of activation changes over time. The benefit of the suggested approach is demonstrated by a series of simulation studies and through an application to experimental data.

KEYWORDS AND PHRASES: Spatially adaptive smoothing, Temporally adaptive smoothing, fMRI, Brain imaging, Smoothing.

1. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive imaging technique that can be used to study mental activity in the brain. It builds on repeatedly imaging a 3D brain volume and studying localized changes in oxygenation patterns. In the past decade fMRI has provided researchers with unprecedented access to the brain in action and provided countless new insights into the inner workings of the human brain. However, in order for fMRI to reach its full potential there is a need for principled statistical analysis of the resulting data; an issue which is complicated by complex spatio-temporal signal properties and the sheer amount of available data (Lindquist, 2008).

In an fMRI study each brain volume consists of a number of uniformly spaced volume elements, or voxels, whose intensity represents the spatial distribution of the nuclear spin density within that particular voxel. The actual signal measurements are acquired by the MR scanner in the

frequency-domain (k-space), which is typically sampled on a rectangular Cartesian grid, and then Fourier transformed into the spatial-domain (image-space). Prior to statistical analysis fMRI data typically undergo a series of preprocessing steps (e.g., registration and normalization) designed to validate the assumptions of the subsequent analysis. One such step is spatial smoothing, which is the focus of this paper. Smoothing typically involves convolving the functional images with a Gaussian kernel, often described by the full width of the kernel at half its maximum height (FWHM). Common values for kernel widths vary between 4 and 12 mm FWHM. Gaussian smoothing is implemented in major software packages such as SPM (Statistical Parametric Mapping, Wellcome Institute of Cognitive Neurology, University College London), AFNI (Analysis of Functional Imaging Data), and FSL (FMRIB software library, Oxford).

There are several reasons why it is popular to smooth fMRI data. First, it may improve inter-subject registration and overcome limitations in the spatial normalization by blurring any residual anatomical differences. Second, it ensures that the assumptions of random field theory (RFT, Worsley and Friston, 1995), commonly used to correct for multiple comparisons, are valid. A rough estimate of the amount of smoothing required to meet the assumptions of RFT is a FWHM of 3 times the voxel size (e.g., 9 mm for 3 mm voxels). Third, if the spatial extent of a region of interest is larger than the spatial resolution, smoothing may reduce random noise in individual voxels and increase the signal-to-noise ratio (SNR) within the region (Rosenfeld and Kak, 1982). Finally, spatial smoothing can be used to reduce the effects of ringing in the image due to the restriction of sampling to a finite k-space region (Lindquist and Wager, 2008).

While the use of a fixed Gaussian kernel is by far the most common approach towards smoothing fMRI data, a number of other studies have suggested alternative approaches. For example, Gaussians of varying width (Poline and Mazoyer, 1994; Worsley et al., 1996) and rotations (Shafie et al., 2003) have been proposed, as well as both wavelets (Van De Ville, Blu, and Unser, 2006) and prolate spheroidal wave functions (Lindquist and Wager, 2008; Lindquist et al., 2006). A common theme in all these methods is that the amount of smoothing is chosen *a priori* and independently of the data. Furthermore, the same amount of smoothing is applied throughout the whole image. This can potentially lead to problems, as the size and shape of activated regions are known to vary across the brain depending on the

*Corresponding author.

task, leading to a situation where certain regions are under-smoothed while others are over-smoothed. A number of fully Bayesian spatio-temporal models have been suggested (e.g., Penny, Trujillo-Barreto, and Friston, 2005) that use spatial priors in lieu of smoothing. In contrast to smoothing with a fixed kernel, these methods allow for the potential of space-varying averaging of voxels. Finally, Bowman et al. (2008) suggest not smoothing the data at all during preprocessing and instead smooth voxel-level estimates by modeling spatial correlations between voxels using a Bayesian hierarchical model.

In this paper we introduce an alternative method for spatial smoothing of fMRI data using nonstationary spatial Gaussian Markov random fields. The Gaussian Markov random field specifies the extent of the smoothing function fit to each voxel in the image by applying weights to neighboring voxels. If these weights are constant throughout the image, the result will be equivalent to Gaussian kernel smoothing as described above. In our method, however, we use the data to obtain voxel-specific weights. Thus the amount of smoothing is data-driven and allowed to vary spatially across the image. In addition, as the voxel-specific weights may change as a function of time our method allows the spatial extent of smoothing to not only vary across *space*, but also across *time*. Hence, the benefit of the suggested approach is two-fold. First, it allows the amount of smoothing to vary across the brain depending on the spatial extent of activation. Second, it allows researchers to study how the extent of activation varies as a function of time; something that to the best of our knowledge has not previously been possible in fMRI studies. Our approach is performed using a fully Bayesian setup and implemented with an efficient Markov Chain Monte Carlo algorithm.

We begin by discussing the theoretical aspects of the approach. We illustrate its utility and compare it to Gaussian smoothing in a series of simulation studies. Finally, we apply the method to experimental data collected during stimulation of the visual cortex.

2. SPATIALLY ADAPTIVE SMOOTHING

We will consider the following spatial model:

$$(1) \quad y_{jk} = f(u_j, v_k) + \varepsilon_{jk}, \quad j = 1, \dots, n_1, k = 1, \dots, n_2,$$

where y_{jk} are response values observed at locations $[u_j, v_k]$, f is an unknown bivariate function on a regular $n_1 \times n_2$ grid, and ε_{jk} are mean zero noise terms. In our context, y_{jk} represent the raw fMRI data and f the smoothed image. We will use this model to apply spatially adaptive smoothing to the raw fMRI image at each time point independently, using the Bayesian hierarchical spatial model devel-

oped in Yue and Speckman (2009). Details are provided in Sections 2.1 to 2.3, but briefly, this approach involves controlling the smoothness of f using a prior based on a discretized thin-plate spline. The prior contains parameters τ and δ_{jk} , and spatially adaptive smoothing is implemented by allowing δ_{jk} to vary spatially across voxels. Specifically, the parameter δ_{jk} controls the smoothness of f at voxel (u_j, v_k) , with larger δ_{jk} corresponding to more smoothing. The variation of δ_{jk} over the voxel locations is therefore related to the variation in the amount of smoothing applied.

We introduce the spatially adaptive smoothing method in three steps. First, we define an intrinsic Gaussian Markov random field (IGMRF) that is a discretization of the solution of non-adaptive thin-plate splines. Based on this definition, we set an IGMRF prior on f (Section 2.1). Next, we extend this prior to a spatially adaptive prior (Section 2.2). This is the key component that allows for spatially adaptive smoothing of fMRI images. Finally, we complete the specification of the model by defining the hyperpriors used in the model (Section 2.3).

2.1 Thin-plate spline prior

The spatially adaptive prior on the function space of f used in this work is based on intrinsic Gaussian Markov random fields (IGMRF), an important class of models in Bayesian hierarchical modeling (see e.g., Rue and Held, 2005). The specific IGMRF that we use is motivated by discretizing the thin-plate spline solution to smoothing functions.

Given data y_{jk} , a thin-plate spline estimator is the solution to the minimization problem

$$(2) \quad \hat{f} = \arg \min_f \left[\sum_{k=1}^{n_2} \sum_{j=1}^{n_1} (y_{jk} - f(u_j, v_k))^2 + \lambda J_2(f) \right],$$

where the penalty term $J_2(f)$ can be represented by the *biharmonic* differential operator (under certain boundary conditions),

$$(3) \quad \iint_{\mathbb{R}^2} \left[\left(\frac{\partial^4}{\partial u^4} + 2 \frac{\partial^4}{\partial u^2 \partial v^2} + \frac{\partial^4}{\partial v^4} \right) f(u, v) \right] dudv.$$

If we assume that h is a small distance between any two spatial locations, the second partial derivative of f at $[u_j, v_k]$ can be approximated by

$$\begin{aligned} \frac{\partial^2}{\partial u^2} f(u_j, v_k) &\approx h^{-2} \nabla_{(1,0)}^2 f(u_j, v_k) \quad \text{and} \\ \frac{\partial^2}{\partial v^2} f(u_j, v_k) &\approx h^{-2} \nabla_{(0,1)}^2 f(u_j, v_k), \end{aligned}$$

where $\nabla_{(1,0)}^2$ and $\nabla_{(0,1)}^2$ denote the second order backward difference operators

$$\begin{aligned}\nabla_{(1,0)}^2 f(u_j, v_k) &= f(u_{j+1}, v_k) - 2f(u_j, v_k) + f(u_{j-1}, v_k), \\ \nabla_{(0,1)}^2 f(u_j, v_k) &= f(u_j, v_{k+1}) - 2f(u_j, v_k) + f(u_j, v_{k-1}).\end{aligned}$$

Letting $z_{jk} = f(u_j, v_k)$, the differential operator in $J_2(f)$ can thus be discretized by

$$(4) \quad h^{-4} \left(\nabla_{(1,0)}^2 + \nabla_{(0,1)}^2 \right)^2 f(u_j, v_k) = h^{-4} [(z_{j+1,k} + z_{j-1,k} + z_{j,k+1} + z_{j,k-1}) - 4z_{j,k}]^2$$

at location $[u_j, v_k]$. The increment in (4) can be regarded as an extension of the difference operator defined for univariate random walks. As a result, one straightforward approximation of the thin-plate spline penalty $J_2(f)$ is

$$(5) \quad \frac{1}{h^4} \sum_{j=2}^{n_1-1} \sum_{k=2}^{n_2-1} [(z_{j+1,k} + z_{j-1,k} + z_{j,k+1} + z_{j,k-1}) - 4z_{j,k}]^2.$$

After including some boundary terms in (5) to fix rank deficiency (see [Yue and Speckman, 2009](#), for details), an improved approximation of the penalty (3) has a quadratic expression $\mathbf{z}'\mathbf{A}\mathbf{z}$, where \mathbf{z} is a vector of z_{jk} and \mathbf{A} is a semidefinite *structure matrix* whose entries are the coefficients of (5) plus those boundary terms. The detailed specification of \mathbf{A} can be found in [Yue and Speckman \(2009\)](#). If we let $\lambda_h = \lambda/h^4$, it is easy to see that the vector $\hat{\mathbf{z}}$ defined by

$$(6) \quad \hat{\mathbf{z}} = \arg \min_{\mathbf{z}} [(\mathbf{y} - \mathbf{z})'(\mathbf{y} - \mathbf{z}) + \lambda_h \mathbf{z}'\mathbf{A}\mathbf{z}]$$

is a discretized thin-plate spline, similar to \hat{f} in (2). The optimization criterion (6) suggests using a Gaussian likelihood, i.e., $\varepsilon_{jk} \stackrel{iid}{\sim} N(0, \tau^{-1})$, and thus, in a Bayesian formulation, we can set a prior on \mathbf{z} of the form

$$(7) \quad [\mathbf{z} | \delta] \propto \delta^{(n-1)/2} \exp\left(-\frac{\delta}{2} \mathbf{z}'\mathbf{A}\mathbf{z}\right),$$

where τ and δ are two precision (inverse variance) parameters. The posterior distribution of \mathbf{z} can be shown to be $MVN(\mathbf{S}_{\lambda_h} \mathbf{y}, \tau^{-1} \mathbf{S}_{\lambda_h})$, where the smoothing parameter is $\lambda_h = \delta/\tau$, and the smoother matrix is $\mathbf{S}_{\lambda_h} = (\mathbf{I}_n + \lambda_h \mathbf{A})^{-1}$. The posterior mean $\mathbf{S}_{\lambda_h} \mathbf{y}$ is then a Bayesian estimator of the discretized thin-plate spline.

Clearly, the random vector \mathbf{z} in (7) is an IGMRF because it follows an improper Gaussian distribution and satisfies the *Markov conditional independence* assumption. More specifically, the null space of \mathbf{A} is spanned by a constant vector, and the conditional distribution of each z_{jk} is Gaussian and only depends on its neighbors. Using graphical notation, the conditional expectation of an interior z_{jk} can be expressed as

$$(8) \quad E(z_{jk} | \mathbf{z}_{-jk}) = \frac{1}{20} \begin{pmatrix} \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \bullet & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\ \circ & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \circ & \circ & \bullet & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ & \circ \end{pmatrix},$$

where the locations denoted by a ‘•’ represent those values of \mathbf{z}_{-jk} that the conditional expectation of z_{jk} depends on, and the number in front of each grid denotes the weight given to the corresponding ‘•’ locations. Therefore the conditional mean of z_{jk} is a particular linear combination of the values of its neighbors, and the conditional variance is given by $\text{Var}(z_{jk} | \mathbf{z}_{-jk}) = (20\delta)^{-1}$ for all z_{jk} . Not only is this proposed IGMRF able to capture a rich class of spatial correlations, the matrix \mathbf{A} is also sparse, allowing the use of efficient algorithms for computation ([Rue and Held, 2005](#)).

A potential weakness of the IGMRF described above, however, is that the same amount of smoothing (determined by δ) is applied at every voxel. For efficient signal detection, we need less smoothing at activated voxels and relatively more smoothing on non-activated areas. Standard smoothing techniques (e.g., Gaussian kernel filter with fixed width) involves a trade-off between increased detectability and loss of information about the spatial extent and shape of the activation areas. See [Tabelow et al. \(2006\)](#), [Smith and Fahrmeir \(2007\)](#) and [Brezger et al. \(2007\)](#). Such loss of information can be avoided by using a spatially adaptive IGMRF extended from (7), as described in the next section.

2.2 Spatially adaptive IGMRF prior

Following [Yue and Speckman \(2009\)](#), the constant precision δ is replaced by locally varying precisions δ_{jk} to achieve an adaptive extension of (7). As a result, the full conditional distribution of an interior z_{jk} remains Gaussian with the same mean as in (8) but with adaptive variances $\text{Var}(z_{jk} | \mathbf{z}_{-jk}) = (20\delta_{jk})^{-1}$. Thus, a small value of δ_{jk} (large variance) corresponds to less smoothing of z_{jk} , appropriate when z_{jk} shows increased local variation. With such a modification, the resulting IGMRF becomes spatially adaptive and retains the nice Markov properties.

To complete the construction of the adaptive IGMRF, we first set $\delta_{jk} = \delta e^{\gamma_{jk}}$, so that δ is a scale parameter and $\gamma_{jk} \in \mathbb{R}$ serves as the adaptive precision for δ_{jk} . An additional prior needs to be defined for γ_{jk} . We set the prior on the vector $\boldsymbol{\gamma} = \text{Vec}([\gamma_{jk}])$ to be a first order IGMRF on a regular lattice ([Besag and Higdon, 1999](#); [Rue and Held, 2005](#)) subject to a constraint for identifiability,

$$(9) \quad [\boldsymbol{\gamma} | \eta] \propto \eta^{(n-2)/2} \exp\left(-\frac{\eta}{2} \boldsymbol{\gamma}' \mathbf{M} \boldsymbol{\gamma}\right) I_{(\mathbf{1}'\boldsymbol{\gamma}=0)},$$

where \mathbf{M} is a constant matrix with rank $n - 2$. The prior on \mathbf{z} now has the form

$$(10) \quad [\mathbf{z} | \delta, \boldsymbol{\gamma}] \propto \delta^{(n-1)/2} \exp\left(-\frac{\delta}{2} \mathbf{z}' \mathbf{A}_{\boldsymbol{\gamma}} \mathbf{z}\right),$$

where $\mathbf{A}_{\boldsymbol{\gamma}} = \mathbf{B}' \boldsymbol{\Lambda}_{\boldsymbol{\gamma}} \mathbf{B}$ is an adaptive structure matrix, \mathbf{B} is a full rank matrix, and $\boldsymbol{\Lambda}_{\boldsymbol{\gamma}} = \gamma_{jk} [e^{\gamma_{jk}}]$. Compared to the non-adaptive version given in (7), the prior in (10) features

the term $\exp(\gamma_{jk})$, which represents the adaptive variance of the voxel located at the location $[u_j, v_k]$. We refer to (10) and (9) together as a spatially adaptive IGMRF prior, which has appealing properties for Bayesian inference and computation as demonstrated in Yue and Speckman (2009).

2.3 Hyperpriors and computation

The hyperpriors on the precision components τ , δ , and η are required for a fully Bayesian inference. Defining a new parametrization with $\xi_1 = \delta/\tau$ and $\xi_2 = \eta/\delta$, the spatially adaptive IGMRF prior becomes

$$(11) \quad [z | \tau, \xi_1, \gamma] \propto (\tau \xi_1)^{(n-1)/2} \exp\left(-\frac{\tau \xi_1}{2} z' \mathbf{A}_\gamma z\right),$$

$$[\gamma | \tau, \xi_1, \xi_2] \propto (\tau \xi_1 \xi_2)^{(n-2)/2} \exp\left(-\frac{\tau \xi_1 \xi_2}{2} \gamma' \mathbf{M} \gamma\right) I_{(1', \gamma=0)}.$$

It is not hard to see that the adaptive smoothing parameters for z_{jk} have expression $\xi_1 \exp(\gamma_{jk})$, while ξ_2 can be considered as a smoothing parameter for γ . Therefore, the new parametrization is more interpretable. To be more specific, the value of ξ_1 controls the degree of global smoothing taken on the whole z field: a smaller ξ_1 would yield a less smooth field, indicating that there is a region (perhaps an activated region) where the voxel values vary significantly at the local level. The value of ξ_2 determines the degrees of smoothing put on the γ : a smaller ξ_2 implies more adaptive (i.e. more variable) precisions γ_{jk} and therefore more adaptive smoothing applied to the z field.

As for hyperpriors, we use an invariance prior on τ , a Pareto prior for ξ_1 , and an inverse gamma prior on ξ_2 , i.e.,

$$[\tau] \propto \frac{1}{\tau}, \quad [\xi_1 | c] = \frac{c}{(c + \xi_1)^2},$$

$$[\xi_2 | a, b] \propto \xi_2^{-(a+1)} \exp\left(-\frac{b}{\xi_2}\right),$$

where $\xi_1 > 0$, $c > 0$, $\xi_2 > 0$, $a > 0$, $b > 0$. Following Yue and Speckman (2009), the values of a , b and c are chosen to yield flexible priors and proper posterior distributions.

Besides adaptive spatial smoothing, efficient MCMC computation is another big advantage of using the IGMRF model described above. Since the priors are sparse, the Gibbs sampler is fairly efficient using the sampling techniques of Yue and Speckman (2009).

3. SIMULATION

We performed two different simulation studies to investigate the performance of our method. The first is intended to illustrate our method's ability to alter the amount of smoothing as a function of the spatial extent of activation across the image. The second is intended to illustrate the method's ability to alter the amount of smoothing as a function of how the spatial extent varies across time.

Simulation I: In the first simulation, we constructed a 40×40 phantom image containing nine regions of activation – circles with varying radius (see left panel of Figure 1). To simulate a dynamic image series, this base image was recreated 200 times. In each circle, activation was simulated

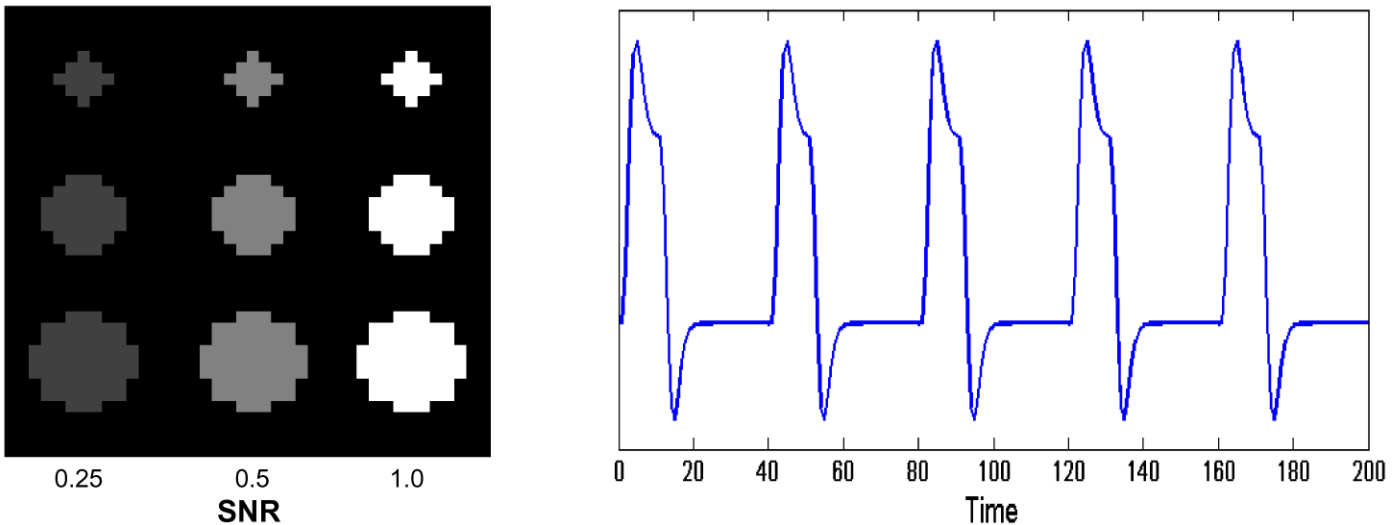


Figure 1. Illustration of Simulation I: A phantom image consisting of nine activation circles (left) and a plot of the activation profile (right). The amplitude of the response varies between 0.25, 0.5 and 1 depending on the column the circle lies in. Hence, circles in the left-most column have the smallest SNR, while circles in the right-most column have the highest.

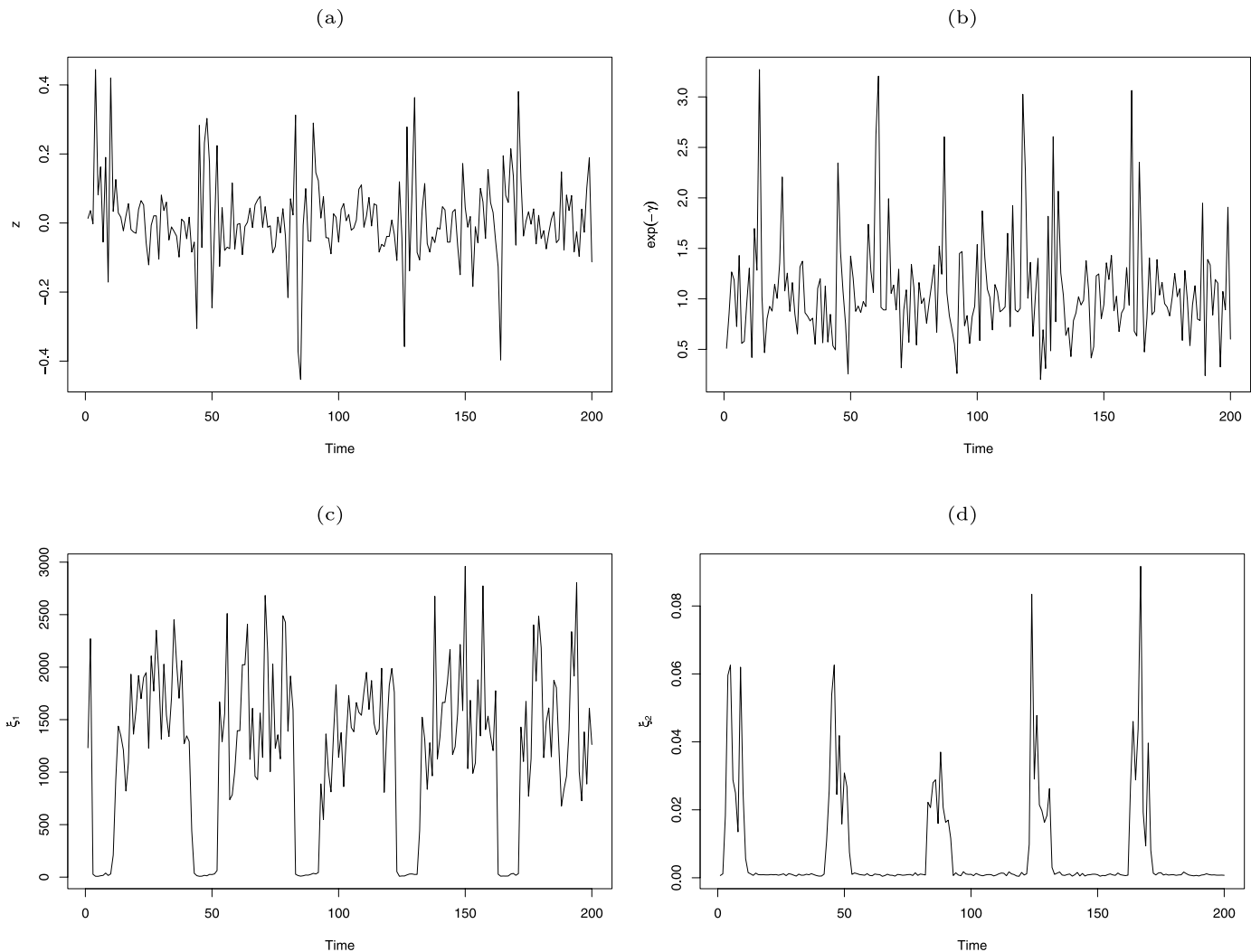


Figure 2. Time series plots for Simulation 1: (a) estimated response z from a particular voxel; (b) adaptive variance for that voxel; (c) estimated ξ_1 ; (d) estimated ξ_2 . It appears that values of ξ_1 track the periods of activation and non-activation closely, with smaller ξ_1 during activation and vice versa. Thus, less smoothing of z is performed during activation. Values of ξ_2 , which controls the smoothness of the γ field, also tracks the activation periods quite well, but in a reverse fashion. The time series plot for the adaptive variance shows that the voxel considered has larger variance during activation.

according to a boxcar paradigm convolved with a canonical hemodynamic response function (Boynton et al., 1996). The boxcar consisted of five repetitions of a 10 s stimulus followed by a 30 s rest period. The resulting activation profile is shown in the right panel of Figure 1. The amplitude of the response was allowed to vary between 0.25, 0.5 and 1 depending on which column of the image the circle lay in. Hence, in each circle, there are periods of activation and non-activation with varying signal strength, smallest for the regions on the left and largest for the regions on the right. Standard Gaussian noise was added to each image.

We applied our adaptive smoothing method to the simulated data set. Figure 2 shows time series plots of the smoothing parameters ξ_1 and ξ_2 , as well as plots of z_i and

γ_i for a particular voxel in one of the activation circles. Recall that ξ_1 controls the global smoothness of the z field, with larger ξ_1 corresponding to a smoother z field. We find that the values of ξ_1 track the periods of activation and non-activation rather closely, with smaller ξ_1 during activation and vice versa. Thus, less smoothing of z is performed during activation, since some voxels will have very different z values from the other voxels. We find that ξ_2 , which controls the smoothness of the γ field, also tracks the activation periods very well, but in a reverse fashion, larger (thus smoother γ) during periods of activation. The time series plot for adaptive variance $\exp(-\gamma_i)$ shows that the particular voxel considered has larger variance during activation.

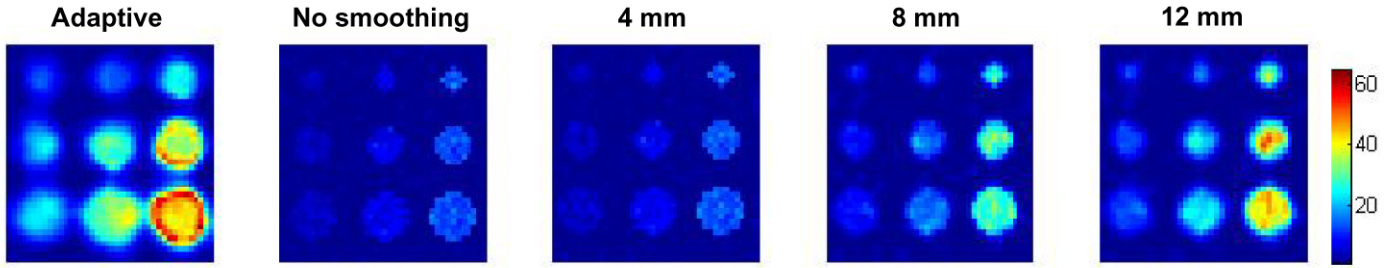


Figure 3. Results of Simulation I: t -maps obtained using spatially adaptive smoothing and fixed Gaussian kernels with various widths. The adaptive smoothing achieves the best balance between smoothing the image and retaining details about boundaries.

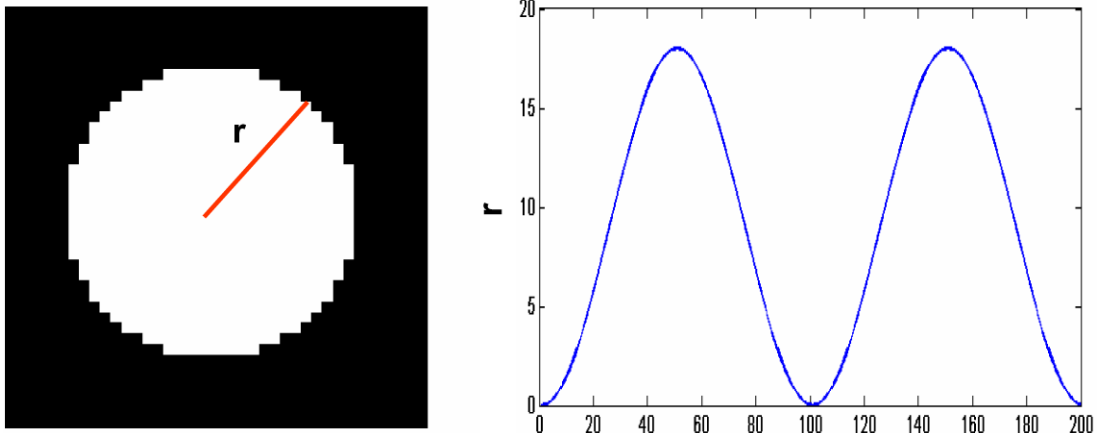


Figure 4. Illustration of Simulation II: A phantom image consisting of a circular region of activation with radius r (left), where the radius varies across time according to a sinusoidal function (right).

Next, we performed statistical analysis on the data processed using our method and using fixed Gaussian kernels with a FWHM of 0, 4, 8 and 12 mm. The analysis on the 5 resulting data sets was performed using the general linear model (GLM) (Worsley and Friston, 1995). The first column of the design matrix consisted of a baseline function, while the second column corresponded to the true response profile shown in Figure 1. The model was fit voxel-wise and a t -test was performed to determine the significance of the component related to the response. The first panel in Figure 3 shows the t -map obtained using our method. Also shown are similar images obtained using a Gaussian kernel with different widths. We find that our adaptive smoothing method achieves a good balance between smoothing the image and retaining the details, and seems to be better at detecting regions with lower signal (left side of the image).

Simulation II: In the second simulation, we again constructed a series of 200 phantom images of size 40×40 . Here the activation region is a single circle whose radius varies according to a sinusoidal function. Thus the activated region grows and shrinks as a function of time. Figure 4 shows a

plot of the radius of the activated circle versus time. The activation profile within the circle is assumed to take the same shape as in Simulation I (see right panel of Figure 1) and standard Gaussian noise is added to each image.

Figures 5 and 6 show the results obtained using our adaptive smoothing method. Figure 5 contains several time series plots of estimates of z at the center voxel for the various smoothing techniques. This figure clearly shows the effectiveness of our adaptive smoothing method for recovering the signal. Figure 6 shows plots of z and $\exp(-\gamma)$ for the center voxel as well as plots of ξ_1 and ξ_2 . The plot of ξ_1 shows the overall smoothing applied to the z field. The amount of smoothing varies sinusoidally with time, as expected based on the setup of the simulated data. Consistent with the first simulation study, there is less smoothing of z when the spatial extent of the activated region increases. Similarly, we find that ξ_2 varies roughly sinusoidally, with more smoothing of the γ field (indicating less adaptivity) during periods when a larger region is activated. The plot for $\exp(-\gamma)$ shows the adaptive variance at the center voxel. We find that the adaptive variance is high during periods with the greatest change in the number of activated voxels. On the other

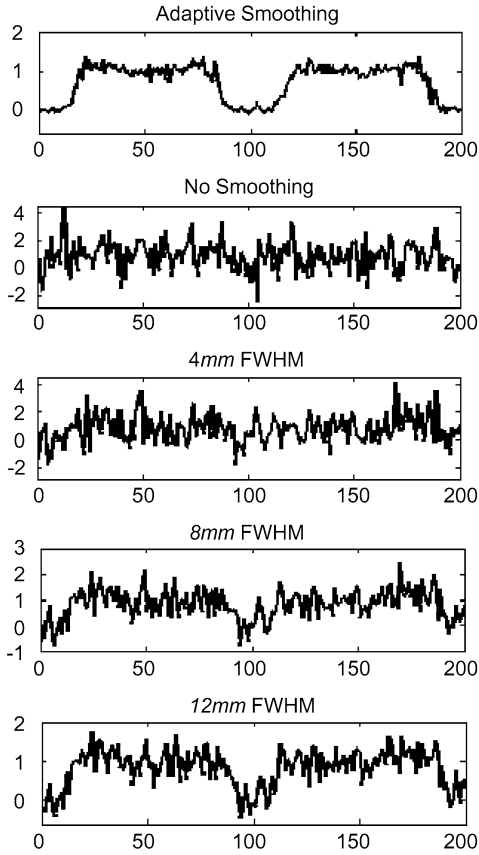


Figure 5. Results of Simulation II: Time series plots of the estimated response z for a central voxel obtained by spatially adaptive smoothing and fixed Gaussian kernels with various widths. The adaptive smoothing method does the best job in removing noise while retaining signal.

hand, the variance is small at more stable periods when most of the voxels are activated or deactivated. This can be seen by matching the plot for $\exp(-\gamma)$ with that for z .

4. DATA ANALYSIS

Nine students at the University of Michigan were recruited and paid \$50 for participation in the study. All human participant procedures were conducted in accordance with Institutional Review Board guidelines. The experimental data consisted of a visual paradigm conducted on the 9 subjects, specifically, of a blocked alternation of 11 s of full-field contrast-reversing checkerboards (16 Hz) with 30 s of open-eye fixation baseline. Blocks of unilateral contrast-reversing checkerboards were presented on an in-scanner LCD screen (IFIS, Psychology Software Tools). Spiral-out gradient echo images were collected on a GE 3T fMRI scanner. Seven oblique slices were collected through visual and motor cortex, $3.12 \times 3.12 \times 5$ mm voxels, TR = 0.5 s, TE = 25 ms, flip angle = 90, FOV = 20 cm, 410 images. Data from all images were corrected for slice-acquisition timing

differences using 4-point sinc interpolation and corrected for head movement using 6-parameter affine registration prior to analysis.

For each subject Gaussian filters with FWHM of 0, 4, 8 and 12 mm were applied to the slice of the data which contained the largest signal over the visual cortex (slice #3). Thereafter, we applied our method to the same data set. Next a standard GLM analysis was performed on each of the smoothed data sets, as well as the non-smoothed data, to create t-maps (see Figure 7). We then thresholded the data at $\alpha = 0.01$ using Bonferroni correction to account for multiple comparisons. The thresholded images are also shown in Figure 7. Similar to the results of our simulation study, we find that adaptive smoothing yields a good balance between smoothing the image and retaining detail in the activated regions.

Figure 8 shows time series plots of the raw and smoothed data for a particular voxel, as well as a plot of the adaptive variance $\exp(-\gamma)$ at that voxel. We find that the smoothed version shows more clearly the signal in the data. Also, we find that adaptive variance is high (low) corresponding to peaks (troughs) in z , suggesting less smoothing of the data during activated periods. In addition, we found that the smoothing parameters ξ_1 and ξ_2 (not shown) for each time point are pretty similar suggesting that at each time point the same amount of global smoothing was applied to most of the voxels.

5. DISCUSSION

This paper introduces a novel approach towards spatially smoothing fMRI data based on the use of non-stationary spatial Gaussian Markov random fields. A novel feature of our approach is that it allows the spatial extent of smoothing to vary not only across *space*, but also across *time*. The benefit of the suggested approach is therefore two-fold. First, it allows the amount of smoothing to vary across the brain depending on the spatial extent of activation. This will help circumvent problems with over/under-smoothing active brain regions of varying size that may occur if smoothing is performed using a Gaussian kernel of fixed width. Also, adaptive smoothing is more in line with the matched filter theorem (Rosenfeld and Kak, 1982) which states a filter that is matched to the signal will give optimum signal to noise. Second, our method allows researchers to study how the extent of activation varies as a function of time. We are not aware of any other fMRI study that does this. Functional MRI is based on studying localized changes in oxygenation patterns. However, it is well known (Malonek and Grinvald, 1996) that brain vasculature tends to overreact to calls for oxygenated blood in response to neuronal activity, giving rise to oxygenation patterns that will eventually exceed the area of neural activity. Using our approach we can study how regions of activation vary spatially as a function of time. This has important implications as it may potentially allow

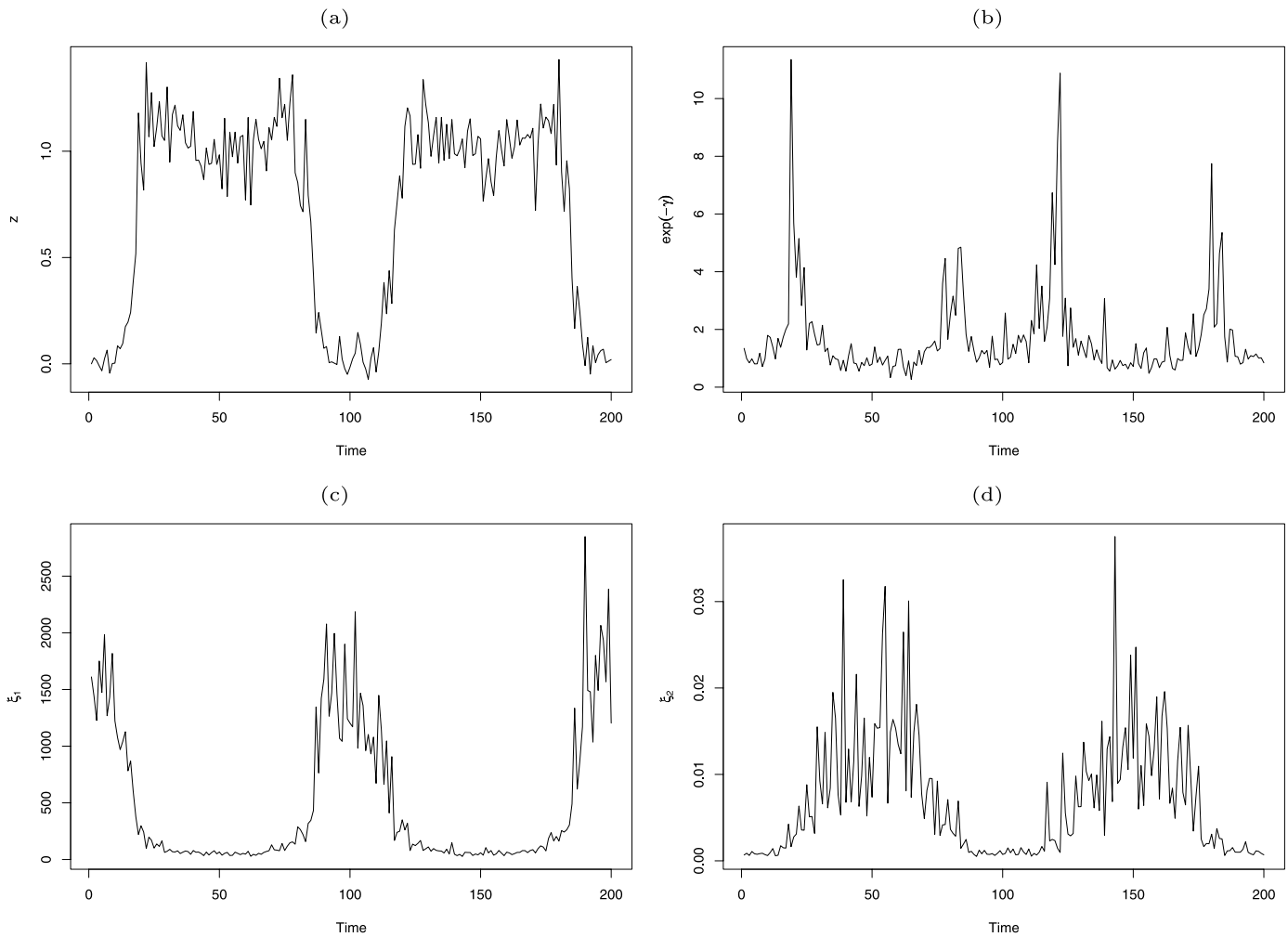


Figure 6. Time series plots for Simulation II: (a) The estimated response z at a center voxel; (b) adaptive variance for that voxel; (c) estimated ξ_1 ; (d) estimated ξ_2 . The plot of ξ_1 shows the overall smoothing applied to the z field varies sinusoidally with time as expected. Similarly, ξ_2 also varies roughly sinusoidally, with more smoothing of the γ field (indicating less adaptivity) during periods when a larger region is activated. Finally, the adaptive variance is high during periods with the greatest change in the number of activated voxels and low at more stable periods when most of the voxels are activated or deactivated.

researchers to discriminate between areas of true activation and those simply adjacent to activation.

While the proposed method has many advantages over smoothing with a fixed kernel, there are certain disadvantages as well. Smoothing with a fixed Gaussian kernel has gained widespread use because of its speed and ease of implementation. Our model is significantly more complex and will therefore lead to increased computation time; an order of magnitude higher than when smoothing with Gaussian or wavelet filters. Also, it is important to note that in this work the model setup assumes that the input data is two-dimensional. In reality fMRI data are four-dimensional with three spatial dimensions and one temporal. Therefore, it may ultimately be more appropriate to smooth the three spatial dimensions directly, or al-

ternatively the full 4D data set. However, as is the case with other similar types of models (e.g., Brezger et al., 2007), computational constraints currently limit our approach to 2D. In spite of this shortcoming we still maintain that smoothing in 2D serves a useful purpose as fMRI data are often analyzed either slice-wise or using cortical surface-based techniques (Dale, Fischl, and Sereno, 1999; Fischl, Sereno, and Dale, 1999). We are currently working on alternative approaches that entail less computation, but constructing a practical 3D or 4D Gaussian Markov random field is non-trivial. Finally, fMRI data are often analyzed for multiple subjects and inference performed on the group (population) level. As we smooth each individual image separately, there are no guarantees of improved inference on the group level using our approach. For these reasons it may be

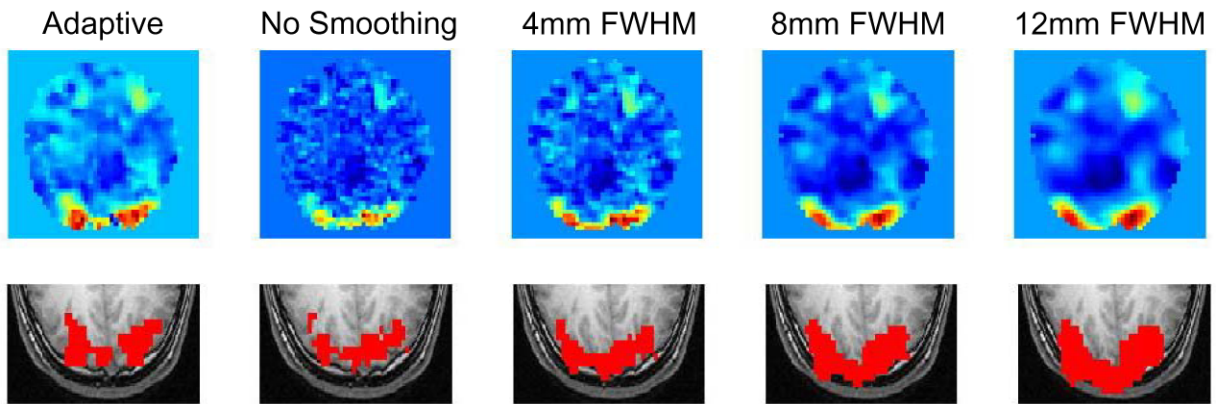


Figure 7. Results of the data analysis: t -maps (top) and thresholded images (bottom) obtained using spatially adaptive smoothing and fixed Gaussian kernels with various widths. Similar to the results of our simulation studies, we find that adaptive smoothing yields a good balance between smoothing the image and retaining detail in the activated regions.

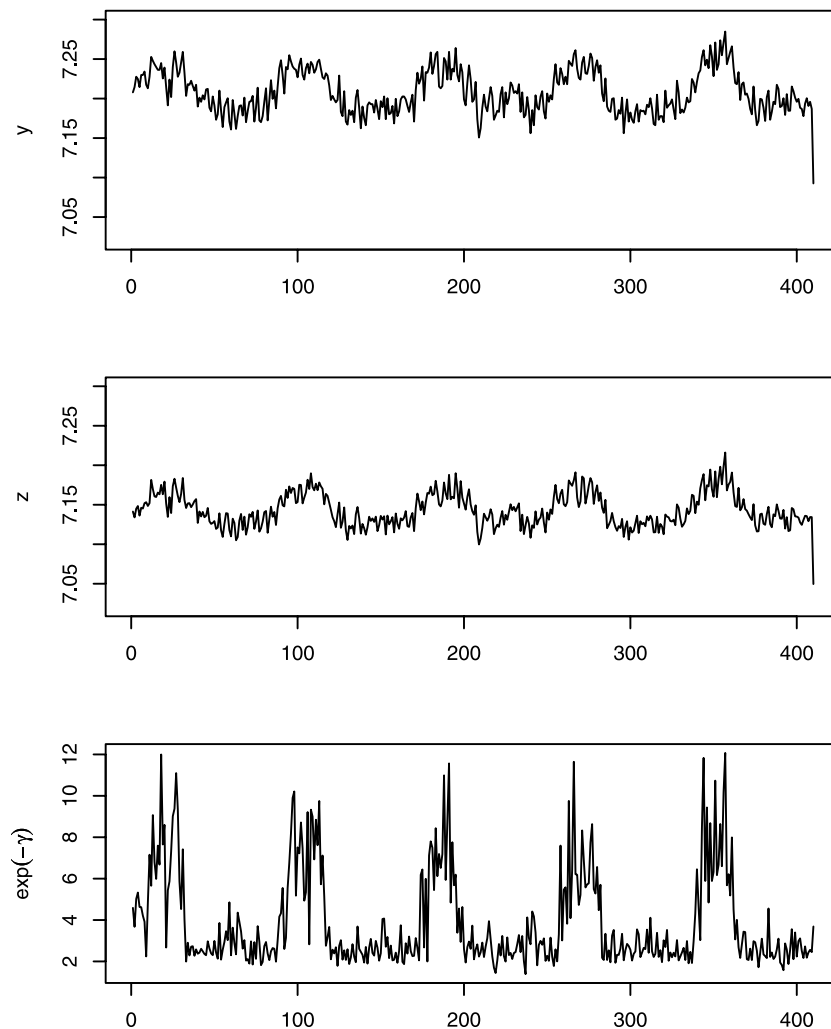


Figure 8. Time series plots of the original response y at a particular voxel (top), the estimated response z (middle), and the estimated adaptive variance (bottom). The smoothed version shows more clearly the signal in the data. Also, the adaptive variance is high (low) corresponding to peaks (troughs) in z , suggesting less smoothing of the data during activated periods.

optimal to smooth all of the images simultaneously, but for the reasons outlined above this is currently not considered feasible.

In this method the extent of activation is used to help determine the degree of smoothing. Since, the smoothed images are subsequently used to obtain statistical inferences about localized activations this may create concern that significant activations may stem in part from the differences in the degree of smoothing applied to the image. However, because the method has no temporal aspect and contains no information about the design matrix used in the subsequent statistical analysis, voxel-level p-values (either uncorrected or corrected at the voxel-level) will be valid, just as they would be when performing standard smoothing.

Although it is advantageous to smooth data for a variety of reasons, there are also obvious costs in spatial resolution. With larger sample sizes, higher field strengths, and other advances in imaging technology, many groups may wish to take advantage of the high potential spatial resolution of fMRI data and minimize the amount of smoothing. The process of spatially smoothing an image is equivalent to applying a low-pass filter to the sampled k-space data prior to reconstruction. This implies that much of the acquired data is discarded as a byproduct of smoothing and temporal resolution is sacrificed without gaining any benefits. Additionally, acquiring an image with high spatial resolution and thereafter smoothing the image does not lead to the same results as directly acquiring a low resolution image. The signal-to-noise ratio during acquisition increases as the square of the voxel volume, so acquiring small voxels means that some signal is lost that can never be recovered. Hence, it is optimal in terms of sensitivity to acquire images at the desired resolution and not employ smoothing. Some recent acquisition schemes have been designed to acquire images at the final functional resolution desired (Lindquist et al., 2008a,b; Zhang et al., 2008). This allows for much more rapid image acquisition, as time is not spent acquiring information that will be discarded in the subsequent analysis. An interesting side note is that since adaptive smoothing effectively alters the extent of the applied low-pass filter there is less inefficiency in data collection as certain voxels are not smoothed and thus use all available data.

Acknowledgements

Yu Yue's research is supported by PSC-CUNY research award #60147-39 40. Ji Meng Loh's research is partially supported by NSF award AST-507687. Martin Lindquist's research is partially supported by NSF grant DMS-0806088. The authors thank Tor Wager for the data.

Received 6 July 2009

REFERENCES

- BESAG, J. AND HIGDON, D. (1999). Bayesian analysis of agricultural field experiments (with discussion). *Journal of the Royal Statistical Society, Series B*, **61**(4), 691–746. [MR1722238](#)
- BOWMAN, F., CAFFO, B., BASSETT, S., AND KILTS, C. (2008). Bayesian hierarchical framework for spatial modeling of fmri data. *NeuroImage*, **39**, 146–156.
- BOYNTON, G., ENGEL, S., GLOVER, G., AND HEEGER, D. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *J. Neurosci*, **16**, 4207–4221.
- BREZGER, A., FAHRMEIR, L., AND HENNERFEIND, A. (2007). Adaptive Gaussian Markov random fields with applications in human brain mapping. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **56**, 327–345. [MR2370993](#)
- DALE, A., FISCHL, B., AND SERENO, M. (1999). Cortical surface-based analysis i: Segmentation and surface reconstruction. *NeuroImage*, **9**, 179–194.
- FISCHL, B., SERENO, M., AND DALE, A. (1999). Cortical surface-based analysis ii: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, **9**, 195–207.
- LINDQUIST, M. (2008). The statistical analysis of fmri data. *Statistical Science*, **23**, 439–464. [MR2530545](#)
- LINDQUIST, M. AND WAGER, T. (2008). Spatial smoothing in fmri using prolate spheroidal wave functions. *Human Brain mapping*, **29**, 1276–1287.
- LINDQUIST, M., ZHANG, C., GLOVER, G., SHEPP, L., AND YANG, Q. (2006). A generalization of the two dimensional prolate spheroidal wave function method for non-rectilinear mri data acquisition methods. *IEEE Transactions in Image Processing*, **15**, 2792–2804. [MR2483118](#)
- LINDQUIST, M., ZHANG, C., GLOVER, G., AND SHEPP, L. (2008a). Acquisition and statistical analysis of rapid 3d fmri data. *Statistica Sinica*, **18**, 1395–1419. [MR2468274](#)
- LINDQUIST, M., ZHANG, C., GLOVER, G., AND SHEPP, L. (2008b). Rapid three-dimensional functional magnetic resonance imaging of the negative bold response. *Journal of Magnetic Resonance*, **191**, 100–111.
- MALONEK, D. AND GRINVALD, A. (1996). The imaging spectroscopy reveals the interaction between electrical activity and cortical microcirculation: implication for optical, pet and mr functional brain imaging. *Science*, **272**, 551–554.
- PENNY, W., TRUJILLO-BARRETO, N., AND FRISTON, K. (2005). Bayesian fmri time series analysis with spatial priors. *NeuroImage*, **24**(2), 350–362.
- POLINE, J. AND MAZOYER, B. (1994). Analysis of individual brain activation maps using hierarchical description and multiscale detection. *IEEE Transactions in Medical Imaging*, **4**, 702–710.
- ROSENFELD, A. AND KAK, A. (1982). *Digital Picture Processing*. New York: Academic Press, 2 edition.
- RUE, H. AND HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London. [MR2130347](#)
- SHAFIE, K., SIGAL, B., SIEGMUND, D., AND WORSLEY, K. (2003). Rotation space random fields with an application to fmri data. *Annals of Statistics*, **31**, 1732–1771. [MR2036389](#)
- SMITH, M. S. AND FAHRMEIR, L. (2007). Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *Journal of the American Statistical Association*, **102**, 417–431. [MR2370843](#)
- TABELOW, K., POLZEHL, J., VOSS, H. U., AND SPOKOINY, V. (2006). Analyzing fmri experiments with structural adaptive smoothing procedures. *NeuroImage*, **33**(1), 55–62.
- VAN DE VILLE, D., BLU, T., AND UNSER, M. (2006). Surfing the brain: An overview of wavelet-based techniques for fmri data analysis. *IEEE Engineering in Medicine and Biology Magazine*, **25**, 65–78.
- WORSLEY, K. J. AND FRISTON, K. J. (1995). Analysis of fmri time-series revisited-again. *NeuroImage*, **2**, 173–181.

WORSLEY, K. J., MARRETT, S., NEELIN, P., VANDAL, A. C., FRISTON, K. J., AND EVANS, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, **4**, 58–73.

YUE, Y. AND SPECKMAN, P. L. (2009). Nonstationary spatial Gaussian Markov random fields. *Journal of Computational and Graphical Statistics*. (to appear).

ZHANG, C., LINDQUIST, M., CHO, Z., GLOVER, G., AND SHEPP, L. (2008). Fast functional magnetic resonance imaging – a new approach towards neuroimaging. *Statistics and Its Interface*, **1**, 13–22. [MR2425341](#)

Yu (Ryan) Yue
Department of Statistics and CIS
Baruch College
New York, NY, 10010
E-mail address: yu.yue@baruch.cuny.edu

Ji Meng Loh
Department of Statistics
Columbia University
New York, NY, 10027
E-mail address: meng@stat.columbia.edu

Martin A. Lindquist
Department of Statistics
Columbia University
New York, NY, 10027
E-mail address: martin@stat.columbia.edu