ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 33 (2006) 55–62

Analyzing fMRI experiments with structural adaptive smoothing procedures

Karsten Tabelow,[a,*] Jörg Polzehl,[a] Henning U. Voss,[b] and Vladimir Spokoiny[a,c]

[a]*Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany*
[b]*Citigroup Biomedical Imaging Center, Weill Medical College of Cornell University, 1300 York Avenue, Box 234, New York, NY 10021, USA*
[c]*Humboldt University Berlin, Germany*

Received 17 January 2006; revised 10 May 2006; accepted 26 June 2006
Available online 4 August 2006

Data from functional magnetic resonance imaging (fMRI) consist of time series of brain images that are characterized by a low signal-to-noise ratio. In order to reduce noise and to improve signal detection, the fMRI data are spatially smoothed. However, the common application of a Gaussian filter does this at the cost of loss of information on spatial extent and shape of the activation area. We suggest to use the propagation–separation procedures introduced by Polzehl and Spokoiny [Polzehl, J., Spokoiny, V. (2006). Propagation–separation approach for local likelihood estimation. Probab. Theory Relat. Fields 135, 335–362] instead. We show that this significantly improves the information on the spatial extent and shape of the activation region with similar results for the noise reduction. To complete the statistical analysis, signal detection is based on thresholds defined by random field theory. Effects of adaptive and non-adaptive smoothing are illustrated by artificial examples and an analysis of experimental data.

Keywords: Functional MRI; Spatially adaptive smoothing; Signal detection

Introduction

The aim of this paper is to offer an alternative approach for using spatial information in the statistical evaluation of fMRI experiments. fMRI data are characterized by a very low signal-to-noise ratio (SNR). Signal detection is to be performed in each voxel of a three-dimensional cube. This induces a severe multiple test problem. A voxelwise analysis, using individual critical values, produces a large portion of false-positive signals. At the same time small signals are concealed due to high variability of the parameter estimates. The use of a global critical value, that is, specification of an error probability to observe a false-positive in any voxel, usually leads to no signal detection at all. The situation is worsened with fMRI studies aiming for ever increased resolution that is subsequently lowering the SNR.

Several methods have been proposed to enhance signal detection using spatial information present in fMRI data. This includes cluster based methods using the spatial extent of detected activations at a given threshold, Friston et al. (1994), together with peak intensity, Poline et al. (1997), to test for activations. Others involve spatial smoothing, using a Gaussian filter, with the amount of smoothing chosen to match the spatial extent of the signal to be detected, Worsley et al. (1996a). Scale space methods, Poline and Mazoyer (1994) and Siegmund and Worsley (1995), attempt to achieve an optimal amount of smoothing at the expense of possible lower sensitivity. Thresholds for the tests are set by random field theory (Adler, 1981, 2000; Worsley, 1994). Especially spatial smoothing employing a non-adaptive Gaussian filter trades increased detectability with loss of information on spatial extent and shape of the activation areas.

Such loss is avoided by using a spatially adaptive method, like the propagation– separation (PS) approach (Polzehl and Spokoiny, 2006), that preserves sizes and shapes of the activated areas. In this paper we show how this method can be used in a fMRI analysis and how it improves the inferred information on the spatial extent and shape of the activated regions.

The use of such an idea, that is adaptive weights smoothing (AWS) (Polzehl and Spokoiny, 2000), for fMRI analysis has already been suggested in Polzehl and Spokoiny (2001). Several important questions were left open at that time. The approach was restricted to periodic activations. Furthermore it did not consider temporal and spatial correlation present in the data. And finally it failed to provide a formal solution to select appropriate thresholds for signal detection. In this paper we will overcome all these drawbacks and provide a complete procedure for structurally adaptive fMRI analysis using the PS approach.

The methodology does not use prior anatomical knowledge. Proposals in this direction, see, e.g., Andrade et al. (2001) or Kiebel and Friston (2002), can be combined with the smoothing procedures presented here.

* Corresponding author. Fax: +49 30 2044975.
E-mail address: tabelow@wias-berlin.de (K. Tabelow).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.06.029

The paper is organized as follows. The next section describes the voxelwise analysis of fMRI time series. In the Spatial smoothing and signal detection section, we outline our proposal to apply adaptive smoothing to enhance signal detection using spatial information. We then provide thresholds motivated by random field theory, see Worsley (1994) and Worsley et al. (1996b). The Results section provides results for experimental fMRI data as well as an artificial example to underline the advantages of the method in different situations.

Voxelwise analysis of time series

In fMRI the BOLD effect is used as a natural contrast employing the fact that voxels with increased neuronal activity are characterized by a higher oxygenation level (Ogawa et al., 1990, 1992). The expected BOLD response can be modeled by a convolution of the task indicator function with the hemodynamic response function. This function models the fact that blood oxygenation is subject to some delay and shows a more complicated structure than a simple indicator function. In fMRI experiments one finds a characteristic form for the measured BOLD response. Several suggestions have been made to model the hemodynamic response function $h(t)$. We give the $h(t)$ as the difference of two gamma functions, Glover (1999):

$$h(t) = \left(\frac{t}{d_1}\right)^{a_1} \exp\left(-\frac{t-d_1}{b_1}\right) - c\left(\frac{t}{d_2}\right)^{a_2} \exp\left(-\frac{t-d_2}{b_2}\right)$$

with $a_1=6$, $a_2=12$, $b_1=0.9$, $b_2=0.9$ and $d_i=a_ib_i$, $i=1, 2$, $c=0.35$ and t the time in seconds. Given the stimulus $s(t)$ as a task indicator function, we arrive at the expected BOLD response as convolution of $s(t)$ and $h(t)$:

$$x(t) = \int_0^\infty h(u)s(t-u)du.$$

The resulting function $x(t)$ is evaluated at the T scan acquisition times t. Approaches allowing for spatially varying latency of the hemodynamic response, Worsley and Taylor (2006), could be used alternatively.

We adopt the common view (Friston et al., 1995; Worsley and Friston, 1995; Worsley et al., 2002) of a linear model for the time series $Y_i=(Y_{it})$ in each voxel i after reconstruction of the raw data and motion correction,

$$Y_i = X\beta_i + \varepsilon_i, \tag{1}$$

where X denotes the design matrix. The first q columns of X contain values of the expected BOLD response for the different stimuli evaluated at scan acquisition times. The other p–q columns are chosen to be orthogonal to the expected BOLD responses and to account for a slowly varying drift and possible other external effects. The error vector ε_i has zero expectation and is assumed to be correlated in time. In order to access the variability of the estimates of β_i correctly we have to take the correlation structure of the error vector ε_i into account. Here we follow Worsley (2005b) assuming an AR(1) model to be sufficient for commonly used MRI scanners. The autocorrelation coefficients ρ_i are estimated from the residual vector $r_i=(r_{i1}, \ldots, r_{iT})$ of the fitted Model (1) as

$$\bar{\rho}_i = \sum_{t=2}^{T} r_{it}r_{i(t-1)} \Big/ \sum_{t=1}^{T} r_{it}^2 .$$

This estimate of the correlation coefficient is biased due to fitting the linear Model (1) (Worsley et al., 2002). We therefore apply the bias correction given by Worsley et al. (2002) leading to an estimate $\tilde{\rho}_i$.

We then use prewhitening to transform Model (1) into a linear model with approximately independent errors. The prewhitened linear model is obtained by multiplying the terms in Eq. (1) with some matrix A_i depending on $\tilde{\rho}_i$. The prewhitening procedure thus results in a new linear model

$$\tilde{Y}_{it} = \tilde{X}_{it}\beta_i + \tilde{\varepsilon}_{it} \tag{2}$$

with $\tilde{Y}_i=A_iY_i$, $\tilde{X}_i=A_iX$ and $\tilde{\varepsilon}_i=A_i\varepsilon_i$. In the new model the distribution of the errors $\tilde{\varepsilon}_{it}$ are approximately independent of time t, such that Var $\varepsilon_i=\sigma_i^2 I_T$.

Finally least squares estimates $\tilde{\beta}_i$ are obtained from Model (2) as

$$\tilde{\beta} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{Y}.$$

The error variance σ_i^2 is estimated from the residuals $\tilde{r}_i$ of the linear Model (2) as $\tilde{\sigma}_i^2=\Sigma_1^T\ \tilde{r}_{it}^2/(T-p)$ leading to estimated covariance matrices

$$\mathrm{Var}\,\tilde{\beta} = \tilde{\sigma}^2(\tilde{X}^T\tilde{X})^{-1}.$$

Let c be a vector of contrasts that defines the effect of interest. This leaves us with three-dimensional arrays $\tilde{\Gamma}$, $\tilde{S}$ containing the estimated effects $\tilde{\gamma}_i=c^T\tilde{\beta}_i$ and their estimated standard deviations $\tilde{s}_i=(c^T\ \mathrm{Var}\ \tilde{\beta}_i c)^{1/2}$. The voxelwise quotient $\tilde{\theta}_i=\tilde{\gamma}_i/\tilde{s}_i$ of both arrays forms a statistical parametric map (SPM) $\tilde{\Theta}$. This map is approximately a random t-field, see Worsley (1994). All arrays carry a correlation structure induced by the spatial correlation in the fMRI data.

Spatial smoothing and signal detection

A voxelwise signal detection may now be based on the SPM Θ, that is, define a voxel as activated if the corresponding value in the SPM Θ exceeds a critical value or threshold. Such an analysis is inefficient in the sense that it either produces a large number of false-positive signals or fails to detect many of the activations. The first situation is typical for applying a voxelwise threshold to the large number of voxels in the data cube while the second case is characteristic if thresholds are controlled by a global significance level.

In situations where activations have a spatial extent, spatial smoothing of the array $\tilde{\Gamma}$ has the potential to improve both overall sensitivity and specificity of signal detection. With Gaussian smoothing we can achieve a reduction in the signal-to-noise ratio at the cost of a possible bias at the border of the activated regions. This leads to an increased power of the tests and therefore more sensitivity and specificity of signal detection in all voxel except in a neighborhood of the border of the region. The size of this neighborhood depends on the amount of smoothing that is applied. In this neighborhood, specificity of signal detection may decrease. The optimal amount of smoothing depends on both signal strength and spatial extent, see, e.g., Worsley and Friston (1995). We illustrate in this paper that adaptive smoothing is able to avoid the loss of specificity at the border of activated regions. The PS approach naturally leads to an optimal local variance reduction and effectively selects a locally optimal bandwidth.

Spatial smoothing is usually applied to the original images in the fMRI time series prior to parameter estimation in the linear model. We note that for parameter estimation, except for effects from prewhitening, the order in which non-adaptive spatial smoothing and evaluation of the linear model are performed is arbitrary. If temporal correlations are spatially homogeneous temporal and spatial smoothing can be interchanged. If temporal correlations are estimated the result may change slightly when the order of smoothing and parameter estimation is reversed. Nevertheless smoothing the effects $\tilde{\gamma}_i$ obtained from the original fMRI data allows for a better assessment for the variance of estimated effects, see the Properties of $\hat{\Gamma}$ and $\hat{S}$ section.

For spatial adaptive smoothing the order of both steps is important. The quality of adaptation heavily depends on the signal-to-noise ratio present in the data. Parameter estimation in the linear model serves as a variance and dimension reduction step prior to spatial smoothing and therefore allows for a much better adaptation.

We propose to use a spatial adaptive smoothing procedure based on the PS approach from Polzehl and Spokoiny (2006) on the array of estimated parameters.

Propagation–separation approach

We shortly explain the main idea. Let us assume that for each voxel with coordinates $i=(i_x, i_y, i_z)$ the parameter γ_i can be well approximated by a constant within a local vicinity $\mathcal{U}(i)$ of voxel i. This serves as our structural assumption.

Our estimation problem can now be viewed as consisting of two parts. In order to efficiently estimate the parameter γ_i in a voxel i, we need to describe a local model, that is, to assign weights $W_i=\{w_{i1}, \ldots, w_{in}\}$. If we knew the neighborhood $\mathcal{U}(i)$ we would define local weights as $w_{ij} = I_{j \in \mathcal{U}(i)}$ and use the weighted least squares estimate

$$\hat{\gamma}_i = \sum_j \tilde{w}_{ij}\tilde{\gamma}_j / \sum_j \tilde{w}_{ij} \text{ with } \tilde{w}_{ij} = w_{ij}/\tilde{s}_j^2 \tag{3}$$

as an estimate of γ_i. Since γ_i and therefore $\mathcal{U}(i)$ are unknown, the assignments will have to depend on the information on Γ that we can extract from the estimates in $\tilde{\Gamma}$ and their estimated variances. If we have good estimates $\hat{\gamma}_j$ of γ_j, we can use this information to infer on the set $\mathcal{U}(i)$ by testing the hypothesis

$$H : \gamma_j = \gamma_i. \tag{4}$$

A weight w_{ij} can be specified based on the value of a test statistic T_{ij}, assigning zero weights if γ_j and γ_i are significantly different. This provides us with a set of weights $W_i=\{w_{i1}, \ldots, w_{in}\}$ that determines a local model in voxel i. These weights can then be used to obtain new estimates of the parameter function γ in each voxel i by Eq. (3).

We utilize the two steps in an iterative procedure. We start with a very local model in each voxel i given by weights

$$w_{ij}^{(0)} = K_{\mathrm{loc}}(\Delta(i,j,h^{(0)})),$$

where $\Delta(i, j, h)=((i_x-j_x)^2/h_x^2+(i_y-j_y)^2/h_y^2+(i_z-j_z)^2/h_z^2)^{1/2}$ is a weighted distance between voxels i and j. The initial vector of bandwidths $h^{(0)}=(h_x^{(0)}, h_y^{(0)}, h_z^{(0)})$ is chosen very small, with its components indirectly proportional to the size of a voxel in the three coordinate directions. K_{loc} is a Gaussian kernel with FWHM=1, truncated at $4/\sqrt{8\ln 2}$, that is, $K_{\mathrm{loc}}(x) = e^{-4\ln 2x^2} I_{x\le 4/\sqrt{8\ln 2}}$. Except from truncation this is the common choice in the fMRI and random fields literature and used here for comparability. Initial estimates are generated using Eq. (3).

We iterate two steps, refining the local models and estimation of Γ. In the kth iteration new weights are generated as

$$w_{ij}^{(k)} = K_{\mathrm{loc}}(\Delta(i,j,h^{(k)}))K_{\mathrm{st}}(\zeta_{ij}^{(k)}) \text{ with } \zeta_{ij}^{(k)} = T_{ij}^{(k)}/\lambda$$

and a monotone non-decreasing kernel K_{st}. We use $K_{\mathrm{st}}(x)=(1-x)_+$ as a default. The term

$$T_{ij}^{(k)} = N_i^{(k-1)}(\hat{\gamma}_i^{(k-1)} - \hat{\gamma}_j^{(k-1)})^2$$

with $N_i^{(k-1)}=\Sigma_j \tilde{w}_{ij}^{(k-1)}$ is used to specify the penalty $\zeta_{ij}^{(k)}$, see Polzehl and Spokoiny (2006). The parameter λ is the main parameter in our approach. It's choice will be explained in the Choice of parameters—propagation condition section.

Then we recompute the estimates employing the just defined weights as

$$\hat{\gamma}_i^{(k)} = \frac{1}{N_i^{(k)}} \sum_j \tilde{w}_{ij}^{(k)} \tilde{\gamma}_j$$

where $\tilde{w}_{ij}^{(k)}=w_{ij}^{(k)}/\tilde{s}_j$ and $N_i^{(k)}=\Sigma_j \tilde{w}_{ij}^{(k)}$. The bandwidth $h^{(k)}$ is increased by a constant factor with each iteration k.

The resulting procedure is essentially a multiscale procedure. In each iteration, we allow for a different amount of smoothing by increasing the bandwidth $h^{(k)}$. The resulting weighting scheme excludes observations in voxel j from being used in estimating at voxel i as soon as they are detected to have significantly different expectations in one of the iteration steps. The weighting scheme ensures that, with a high probability, such decisions are kept within the following iterations. We refer to Polzehl and Spokoiny (2006) for detailed properties of the weighting scheme.

Without spatial correlation $\hat{\gamma}_i(k)$ has variance

$$V_i^{(k)} = Q_i^{(k)}/(N_i^{(k)})^2 \le 1/N_i^{(k)} \text{ with } Q_i^{(k)} = \sum_j (\tilde{w}_{ij}^{(k)})^2 \tilde{s}_j^2,$$

that is, the term $1/N_i^{(k)}$ approximately reflects the variability of $\hat{\gamma}_i^{(k)}$.

Correction for spatial correlation

In our situation we have to adjust for the spatial correlation present in the data. The correlation in each direction is estimated as a global value using the residuals from Model (2). This may also be done locally, see, for example, Kiebel et al. (1999).

Let us assume that the spatial correlation present in $\tilde{\Gamma}$ results from spatial smoothing using the location kernel K_{loc} and employing bandwidths $g=(g_x, g_y, g_z)$ in the three coordinate directions. This means that, for a interior voxel i, we assume the elements of $\tilde{\Gamma}$ to be generated from a spatially uncorrelated field $\check{\Gamma}$ with Var $\check{\gamma}_l=\check{s}_l^2$ as

$$\tilde{\gamma}_i = \sum_j K_{\mathrm{loc}}(\Delta(i,j,g))\check{\gamma}_j/N_g \text{ with } N_g = \sum_j K_{\mathrm{loc}}(\Delta(i,j,g)).$$

This results in

$$\tilde{\gamma}_i^{(k)} = \frac{1}{N_i^{(k)} N_g} \sum_j \tilde{w}_{ij}^{(k)} \sum_j K_{\mathrm{loc}}(\Delta(j,l,g))\check{\gamma}_l.$$

Then the variance of $\hat{\gamma}_i^{(k)}$ is given by

$$\widehat{V}_i^{(k)} = \frac{\sum_l \check{s}_l^2 \left[\sum_j \tilde{w}_{ij}^{(k)} K_{\mathrm{loc}}(\Delta(j,l,g))\right]^2}{(N_i^{(k)})^2 N_g^2}$$

$$= \frac{\sum_l \check{s}_l^2 \left[\sum_j \tilde{w}_{ij}^{(k)} K_{\mathrm{loc}}(\Delta(j,l,g))\right]^2}{\sum_j (\tilde{w}_{ij}^{(k)})^2 \tilde{s}_j^2 N_g^2} V_i^{(k)}.$$

Let $Q_g = \Sigma j\, K_{\mathrm{loc}}(\Delta(i,j,g))^2$. Note that, except at the boundaries of the data cube, the sums Q_g and N_g do not depend on the voxel i. Then for spatially homogeneous variances $\tilde{s}_i^2 = \check{s}^2$, that is, $\tilde{s}_i^2 = Q_g / N_g^2 \check{s}^2$,

$$\widehat{V}_i \approx \frac{\sum_l \left[\sum_j \tilde{w}_{ij}^{(k)} K_{\mathrm{loc}}(\Delta(j,l,g))\right]^2}{\sum_j (\tilde{w}_{ij}^{(k)})^2 Q_g} V_i^{(k)}.$$

If the statistical penalty is negligible, that is, $w_{ij}^{(k)} \approx K_{\mathrm{loc}}(\Delta(i,j,h^{(k)}))$, then

$$\widehat{V}_i^{(k)} \approx \frac{\sum_l \left[\sum_j K_{\mathrm{loc}}(\Delta(i,j,h^{(k)})) K_{\mathrm{loc}}(\Delta(j,l,g))\right]^2}{\sum_j K_{\mathrm{loc}}^2(\Delta(i,j,h^{(k)})) Q_g} V_i^{(k)}$$
$$= C\left(g, h^{(k)}\right) V_i^{(k)}.$$

The factor $C(g, h^{(k-1)})^{-1}$ will be used as an adjustment to $N_i^{(k-1)}$ in the definition of $T_{ij}^{(k)}$ to account for spatial correlation.

PS algorithm for heteroscedastic and spatially correlated data

We now formally describe the resulting algorithm.

- Initialization: Set the initial bandwidth $h^{(0)}$ and compute, for every i, the statistics

$$N_i^{(0)} = \sum_j K_{\mathrm{loc}}(\Delta(i,j,h^{(0)}))/\tilde{s}_j^2,$$

and

$$U_i^{(0)} = \sum_j K_{\mathrm{loc}}(\Delta(i,j,h^{(0)}))\tilde{\gamma}_j/\tilde{s}_j^2,$$

and the estimates

$$\widehat{\gamma}_i^{(0)} = U_i^{(0)}/N_i^{(0)}.$$

Set $k=1$ and $h^{(1)} = c_h h^{(0)}$ for some $c_h > 1$.

- Adaptation: For every pair i, j, compute the penalty

$$\zeta_{ij}^{(k)} = (\lambda C(g, h^{(k-1)}))^{-1} T_{ij}^{(k)}$$
$$= (\lambda C(g, h^{(k-1)}))^{-1} N_i^{(k-1)} (\hat{\theta}_i^{(k-1)} - \hat{\theta}_j^{(k-1)})^2.$$

Compute weights $w_{ij}^{(k)}$ as

$$w_{ij}^{(k)} = K_{\mathrm{loc}}(\Delta(i,j,h^{(k)})) K_{\mathrm{st}}(\zeta_{ij}^{(k)}) \text{ and } \tilde{w}_{ij}^{(k)} = w_{ij}^{(k)}/\tilde{s}_j^2.$$

- Local estimation: Now compute new local MLE estimates $\tilde{\gamma}_i^{(k)}$ of γ_i as

$$\widehat{\gamma}_i^{(k)} = U_i^{(k)}/N_i^{(k)} \text{ with } N_i^{(k)} = \sum_j \tilde{w}_{ij}^{(k)}, U_i^{(k)}$$
$$= \sum_j \tilde{w}_{ij}^{(k)} \tilde{\gamma}_j.$$

- Stopping: Stop if $k=k^*$, otherwise set $h^{(k)} = c_h h^{(k-1)}$, increase k by 1 and continue with the adaptation step.

An estimate of the variance of the final estimate $\hat{\gamma}_i^{(k^*)}$ is given by

$$\hat{s}_i^2 = V_i^{(k^*)} = C(g, h^{(k^*)}) Q_i^{(k^*)} / (N_i^{(k^*)})^2$$

$$\text{with } Q_i^{(k^*)} = \sum_j (\tilde{w}_{ij}^{(k^*)})^2 \tilde{s}_j^2.$$

Choice of parameters—propagation condition

The proposed procedure involves several parameters. The most important one is the scale parameter λ in the statistical penalty ζ_{ij}. The special case $\lambda = \infty$ simply leads to a kernel estimate with bandwidth $h^{(k^*)}$. We propose to chose λ as the smallest value satisfying a propagation condition. This condition requires that, if the local assumption is valid globally, that is, $\gamma_i \equiv \gamma$ does not depend on i, then at each step of the algorithm the adaptive estimate approximately behaves like its non-adaptive counterpart that employs the same bandwidth. Particularly the final estimate for $k^* = \infty$ has approximately the same quality as the global estimate. More formally we request that in this case for each k

$$E|\tilde{\gamma}^{(k)} - \tilde{\gamma}^{(k)}| < \alpha E|\tilde{\gamma}^{(k)} - \gamma| \tag{5}$$

for a specified constant $\alpha > 0$. Here

$$\tilde{\gamma}_i^{(k)} = \frac{\sum_j K_{\mathrm{loc}}(\Delta(i,j,h^{(k)}))/\tilde{s}_j^2 \tilde{\gamma}_j}{\sum_j K_{\mathrm{loc}}(\Delta(i,j,h^{(k)}))/\tilde{s}_j^2}$$

denotes the non-adaptive kernel estimate employing the bandwidth $h^{(k)}$ from step k. The value λ provided by this condition does not depend on the unknown model parameter γ and can therefore be approximately found by simulations. We set a default value for λ using $\alpha = 0.1$.

The second parameter of interest is the number of iterations k^*, or equivalently the maximal bandwidth $h^{(k^*)}$, which controls both numerical complexity of the algorithm and smoothness within homogeneous regions. The initial bandwidth is chosen as $h^{(0)} = \left(1, \frac{v_y}{v_x}, \frac{v_z}{v_x}\right)/\sqrt{8\ln 2}$ in units of $v_x$mm. The bandwidth is increased after each iteration by a default factor $c_h = 1.25^{1/3}$.

Choosing parameters in this procedure by the propagation condition (Eq. (5)) ensures the algorithm, under the hypothesis H: $\gamma = 0$, to behave like a corresponding non-adaptive smoothing algorithm, that is a Gaussian filter employing the largest bandwidth $h^{(k^*)}$. This is important for the definition of thresholds using random field theory. At the same time the separation property of PS allows to clearly separate activated areas, characterized by values of γ different from zero, from regions not affected by the experiment. The maximum bandwidth of the kernel used in the iterative procedure can be chosen as FWHM of 2–3 times voxel size as usual.

Properties of $\hat{\Gamma}$ and $\hat{S}$

Adaptive spatial smoothing results in arrays $\hat{\Gamma}=(\hat{\gamma}_i)$ and $\hat{S}=(\hat{s}_i)$. If no activation is present in any voxel, i.e., the hypothesis H: $\gamma_i=0$ holds for all i, choosing the smoothing parameter λ by the propagation condition (Eq. (5)) ensures the properties of $\hat{\Gamma}$ and $\hat{S}=(\hat{s}_i)$ to approximately coincide with properties of the corresponding arrays obtained by non-adaptive smoothing employing the kernel K_{loc} with bandwidth $h^{(k^*)}$.

We now shortly explain the effect of spatial smoothing on properties of $\hat{S}=(\hat{s}_i)$. Let us assume that $S^2=(s_i^2)$ is an array of i.i.d. χ^2 variables with d_f degrees of freedom. Let weights w_{ij} be fixed. Then the distribution of

$$C_s\bar{s}_i = C_s\frac{\sum_j w_{ij}^2 s_j^2}{\left(\sum_j w_{ij}\right)^2}$$

with $C_s = \sum_j w_{ij}^2(\sum_j w_{ij})^2 \sum_j w_{ij}^4$ can be well approximated by a χ^2-distribution with $(\sum_j w_{ij}^2)^2 df / \sum_j w_{ij}^4$ degrees of freedom. A similar behavior is, for sufficiently large d_f, observed for

$$\tilde{C}_s\hat{s}_i = \tilde{C}_s\frac{\sum_j \tilde{w}_{ij}^2 s_j^2}{\left(\sum_j \tilde{w}_{ij}\right)^2}$$

with $\tilde{C}_s$ depending on both d_f and the weighting scheme. We see that the degrees of freedom for $\hat{S}$ increase with its smoothness induced by spatial smoothing. Note that this is in strict contrast to the situation where spatial smoothing is applied to the original observed images prior to estimation of parameters in the linear Model (1), in which case the degrees of freedom are not affected by the spatial smoothing.

Spatial smoothing provides, in each voxel, a sufficient number of degrees of freedom, which allows to approximate the distribution of the array $\hat{\Theta}=(\hat{\gamma}_i/\hat{s}_i)$ by a (spatially inhomogeneous) Gaussian field.

Defining p values

To define appropriate thresholds for rejecting the hypotheses H: $\hat{\Theta}=0$, we follow the argumentation in Worsley et al. (1996a). For a spatially homogeneous Gaussian random field with smoothness $g=(g_x, g_y, g_z)$ in the coordinate directions, measured in FWHM, an appropriate threshold for signal detection can be defined approximating the p value of an extreme event by its expected Euler characteristic (EC). Worsley et al. (1996a) show that an approximate p value of the maximum Z_{max} of a 3D Gaussian random field over a specified search region is given by

$$P(Z_{max} > z) \approx \sum_{d=0}^{3} R_d(V,g)\rho_d(z) \tag{6}$$

where R_d is the d-dimensional resel count depending on the volume V of the search region and the smoothness measured by the FWHM bandwidths g in mm. $\rho_d(z)$ is the d-dimensional EC density for the case of a Gaussian random field, see Worsley et al. (1996a). As an alternative, the DLM method of Worsley (2005a) may be used.

Appropriate thresholds and p values are derived from the properties of the generated random field under the null hypothesis, i.e., the absence of any activation. In this case the properties of our smoothing procedure are determined by our propagation condition, see the Choice of parameters—propagation condition section, especially Eq. (5). This condition ensures that, under the hypothesis, the resulting random field has very similar properties as the random field generated by a Gaussian filter employing the bandwidth $h^{(k^*)}$ from the last iteration. The difference is that, due to the presence of the statistical penalty ζ_{ij}, the weights w_{ij} are random and slightly smaller compared to the kernel weights $K_{loc}(\Delta(i, j, h^{(k^*)}))$ applied in case of a Gaussian filter. This effect is very small and controlled by the propagation condition (Eq. (5)).

Under stationarity, approximative p values associated with a signal in voxel i can be defined as

$$p_i = \sum_{d=0}^{3} R_d(V, g^{(k^*)})\rho_d(\hat{\theta}_i). \tag{7}$$

where $g^{(k^*)} \approx \sqrt{g^2 + h^{(k^*)^2}}$ is vector of the effective FWHM bandwidths that measures the smoothness of the random field $\hat{\Theta}$. These bandwidths depend on the bandwidth $h^{(k^*)}$ from the last iteration of the PS procedure and on the vector of bandwidths g characterizing the smoothness of the random field $\tilde{\Theta}$. We use the p values (Eq. (7)) for thresholding.

The case of a non-stationary FWHM g can be addressed, see Worsley et al. (1999) or Taylor and Adler (2003).

Results

In this section, we demonstrate various aspects of the procedure introduced in the preceding sections in a series of examples.

Application to artificial data

We first start with an examination of artificial data to illustrate the different behavior of our adaptive PS method and the corresponding

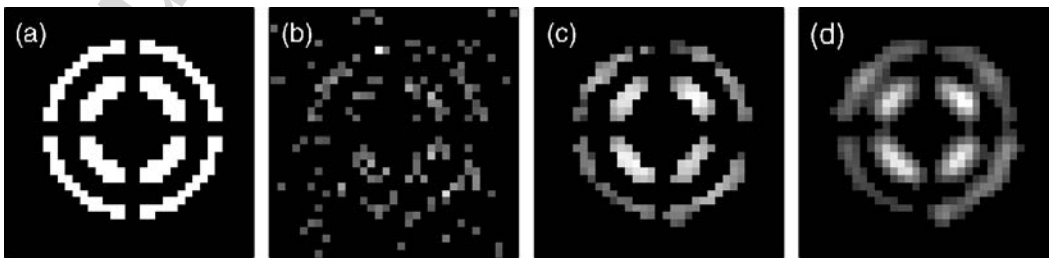

Fig. 1. Results for an artificial fMRI experiment, central slice (from left to right): (a) Signal size for the expected BOLD response assigned to the voxel. (b) Voxelwise analysis without smoothing, employing voxelwise thresholds, (c) PS approach and (d) Gaussian filter.

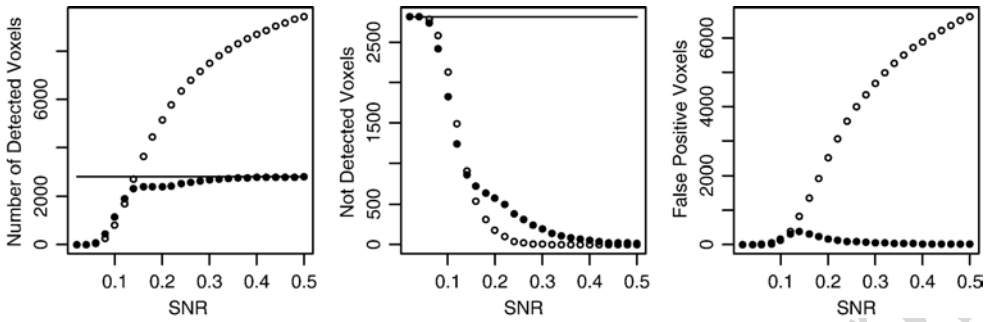

Fig. 2. Analysis of artificial data for the configuration from Fig. 1. SNR in the activation area is varied from 0.02 to 0.5 (x-axis). The left figure illustrates the dependence of the number of detected voxels on SNR. Circles and bullets correspond to non adaptive and adaptive smoothing, respectively. The (maximum) bandwidth $h^{(k^*)}$ is chosen according to the size of the activation area. The line indicates the true number of active voxels. The center figure shows the number of not detected voxels. In the right figure the corresponding number of false-positives is provided.

Gaussian filter in a known situation. We use a typical dimension of $64 \times 64 \times 26$ voxels for the data cube. The activation area consists of voxels with a distance between 5 and 7.5 or 10.5 to 12 from the center of the cube. No activation is assigned to voxels in the two central vertical slices, which induces gaps of width 2 between activation areas (see left image in Fig. 1 to view the activation area in a central slice). At each of these voxels a time series of 107 scans is generated using a convolution of the hemodynamic response function with a task indicator function with onset times at scans 18, 48 and 78 and a duration time of 15 scans. Heavy autocorrelated noise following an AR(1) model with an autocorrelation parameter of 0.3 is added at every voxel. Results for both methods, using a FWHM bandwidth of $h^{(k^*)} = 3.05 v_x$, that is $k^* = 15$, and a voxelwise analysis are shown in Fig. 1. A voxelwise analysis, employing voxelwise thresholds, produces a large portion of false-positives. On the other hand, if thresholds are controlled by a global significance level the analysis generally fails to detect many if not all activations. For the PS method we observe almost no false-positive activations (see Fig. 2 for details) and some none detected voxels. In contrast to this smoothing by a Gaussian filter results in loss of spatial information. Gaps between areas are only partly preserved in case of non-adaptive smoothing.

The adaptive smoothing procedure leads to a much better signal detection. In order to discuss this in more detail we show how the detection depends on signal strength. The same configuration of activated areas is used. We generate an artificial SPM $\tilde{\Theta}$ of estimated parameters for the BOLD response as signal plus standard normal noise. Furthermore we assign χ^2_T ($T = 100$) random variates to resemble variances of the estimated parameters.

In Fig. 2 signal detection using adaptive (PS) and Gaussian smoothing is compared. We vary the SNR in the activation area from 0.02 to 0.5. This is shown on the x-axis of the diagrams. The number of detected voxels is increasing with the SNR with both methods. However, the analysis using adaptive smoothing is capable to detect smaller signals. Furthermore while the number of false-positive detected voxels is practically zero, for Gaussian filtering the number of false-positives increases with SNR. The adaptive smoothing method naturally adapts to the different sizes and shapes of activation areas rather than over-smoothing them.

Application to experimental data

With the results of the preceding section in mind, we now consider experimental fMRI data. The images were acquired on a

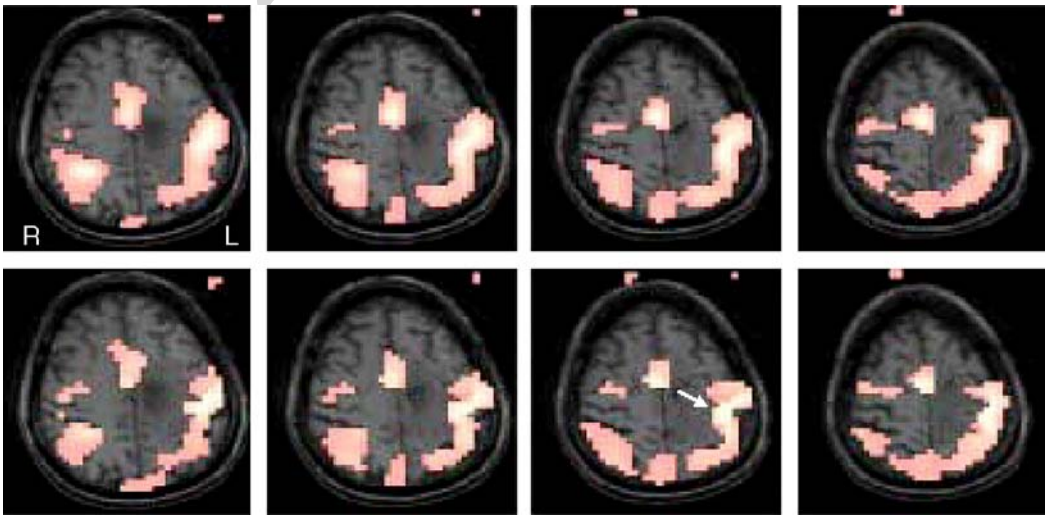

Fig. 3. Application to a right handed finger tapping fMRI experiment. Upper row: non-adaptive smoothing with bandwidth (FWHM) $h = 10$ mm. Lower row: adaptive smoothing with maximal bandwidth (FWHM) $h = 10$ mm.

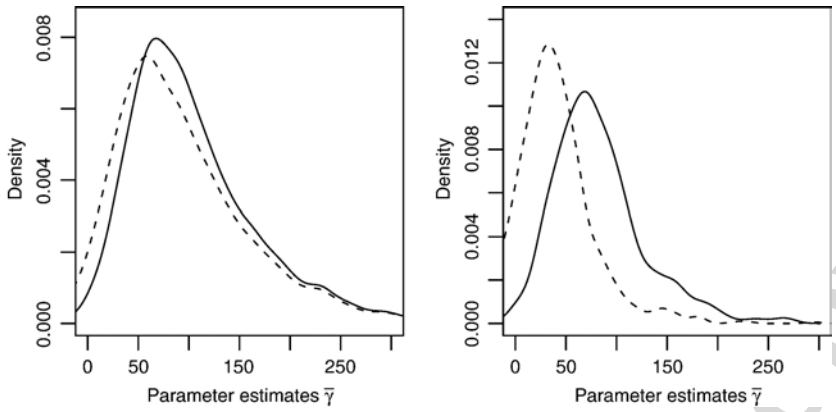

Fig. 4. The density of the estimated parameters $\tilde{\gamma}$ for all detected voxels (left) and for voxels detected only by the specified method (right). The dashed line corresponds to non adaptive smoothing while the solid line is the result from adaptive smoothing.

GE 3T scanner using 2D gradient echo EPI sequence with TE/TR = 40/2000 ms. 26 axial slices of 4-mm thickness and a matrix size of 64 × 64 were acquired. Task were performed in three blocks during 3.7 min of scanning time. The task for the data presented here was right hand finger tapping, where each finger of the right hand successively touched the thumb. The blocklength was 30 s on and 30 s of, repeated three times after an initial of block of 42 s. The first 4 scans were discarded, yielding 107 scans in total. The statistical analysis was applied after motion correction using AFNI, Cox (1996). The results of this analysis are compared for selected slices in Fig. 3, the upper row showing Gaussian and the lower row adaptive smoothing (PS). It is evident that the two results differ in the detection of activation areas. PS provides a much better rendering of the shape of activation areas. For example, the characteristic hand motor area (arrow) is recognized better than in the Gaussian filter result.

Additionally the density of the estimated signals $\tilde{\gamma}$ from Eq. (1) in detected voxels differs for voxels detected using adaptive and non adaptive smoothing. Fig. 4 illustrates this providing densities of $\tilde{\gamma}$ for detected voxels (left plot) and voxels detected by only one method (right plot). Note that the mean of $\tilde{\gamma}$ over voxels detected exclusively using adaptive smoothing is much larger than the corresponding quantity for the Gaussian filter.

Summary

We presented a general approach to integrate adaptive smoothing into the analysis of fMRI experiments. Whereas this approach yields a similar amount of noise reduction and sensitivity as Gaussian filtering, it has the advantage that information on shape and geometry of the activation areas is preserved. Our method can be seen as an extension of the PS approach of Polzehl and Spokoiny (2006) to spatially correlated data, in which the local correlation structure is taken into account to define corresponding local thresholds using random field theory. Additionally, it was shown that smoothing the map of estimated parameters rather than estimating the parameters from smoothed images increases the number of degrees of freedom, such that the array $\hat{\Theta}$ is better approximated by a Gaussian field.

As any filtering method, the application of spatially adaptive smoothing increases the signal-to-noise ratio. In addition, it also increases the effective spatial resolution of activated areas. This may be of particular usefulness in experiments where different activation areas are intermingled in a complex way, like the ocular dominance columns in the visual cortex, or in experiments of brain plasticity where the relocation of functions is being determined (Dancause et al., 2005; Pascual-Leone et al., 2005).

Software

The procedures described in the Voxelwise analysis of time series section and the Spatial smoothing and signal detection section are implemented as a package for the R statistical environment (R Development Core Team, 2005). The package allows to read and write the AFNI (Cox, 1996) and ANALYZE (Biomedical Imaging Resource, 2001) file formats. The R package has been used to produce the results in the Results section.

Acknowledgments

This work is supported by the DFG Research Center Matheon. H.U.V. acknowledges financial support from the Cervical Spine Research Society.

We thank two anonymous referees for their very helpful comments.

References

Adler, R., 1981. The Geometry of Random Fields. Wiley, New York.
Adler, R., 2000. On excursion sets, tube formulae, and maxima of random fields (special invited paper) Ann. Appl. Probab. 10, 1–74.
Andrade, A., Kherif, F., Mangin, J., Worsley, K., Paradis, A., Simon, O., Dehaene, S., Le Bihan, D., Poline, J.-B., 2001. Detection of fMRI activation using cortical surface mapping. Hum. Brain Mapp. 12, 79–93.
Biomedical Imaging Resource, 2001. Analyze Program. Mayo Foundation.
Cox, R., 1996. Software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162.
Dancause, N., Barbay, S., Frost, S., Plautz, E., Chen, D., Zoubina, E., Stowe, A., Nudo, R., 2005. Extensive cortical rewiring after brain injury. J. Neurosci. 25 (44), 10167–10179.
Friston, K., Worsley, K., Frackowiak, R., Mazziotta, J., Evans, A., 1994. Assessing the significance of focal activations using their spatial extent. Hum. Brain Mapp. 1, 214–220.
Friston, K., Holmes, A., Worsley, K., Poline, J.-B., Frith, C., Frackowiak, R., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429.

Kiebel, S., Friston, K., 2002. Anatomically informed basis functions in multisubject studies. Hum. Brain Mapp. 16, 36–46.

Kiebel, S.J., Poline, J.B., Friston, K.J., Holmes, A., Worsley, K.J., 1999. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. NeuroImage 10, 756–766.

Ogawa, S., Lee, T., Kay, A., Tank, D., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. U. S. A. 87, 9868–9872.

Ogawa, S., Tank, D., Menon, R., Ellermann, J., Kim, S., Merkle, H., Ugurbil, K., 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc. Natl. Acad. Sci. U. S. A. 89, 5951–5955.

Pascual-Leone, A., Amedi, A., Fregni, F., Merabet, L., 2005. The plastic human brain cortex. Annu. Rev. Neurosci. 28, 377–401.

Poline, J., Mazoyer, B., 1994. Enhanced detection in brain activation maps using a multifiltering approach. J. Cereb. Blood Flow Metab. 14, 639–642.

Poline, J., Worsley, K., Evans, A., Friston, K., 1997. Combining spatial extent and peak intensity to test for activations in functional imaging. NeuroImage 5, 83–96.

Polzehl, J., Spokoiny, V., 2000. Adaptive weights smoothing with applications to image restoration. J. R. Stat. Soc. Ser., B 62, 335–354.

Polzehl, J., Spokoiny, V., 2001. Functional and dynamic magnetic resonance imaging using vector adaptive weights smoothing. J. R. Statist. Soc., Ser. C 50, 485–501.

Polzehl, J., Spokoiny, V., 2006. Propagation–separation approach for local likelihood estimation. Probab. Theory Relat. Fields 135, 335–362.

R Development Core Team, 2005. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 3-900051-07-0.

Siegmund, D., Worsley, K., 1995. Testing for a signal with unknown location and scale in a stationary Gaussian random field. Ann. Stat. 23, 608–639.

Taylor, J., Adler, R., 2003. Euler characteristics for Gaussian fields on manifolds. Ann. Probab. 31, 533–563.

Worsley, K., 1994. Local maxima and the expected Euler characteristic of excursion sets of χ^2, f and t fields. Adv. Appl. Probab. 26, 13–42.

Worsley, K., 2005a. An improved theoretical p-value for SPMs based on discrete local maxima. NeuroImage 28, 1056–1062.

Worsley, K., 2005b. Spatial smoothing of autocorrelations to control the degrees of freedom in fMRI analysis. NeuroImage 26, 635–641.

Worsley, K., Friston, K., 1995. Analysis of fMRI time series revisited—Again. NeuroImage 2, 173–181.

Worsley, K., Taylor, J., 2006. Detecting fMRI activation allowing for unknown latency of the hemodynamic response. NeuroImage 29, 649–654.

Worsley, K., Marrett, S., Neelin, P., Evans, A., 1996a. Searching scale space for activation in PET images. Hum. Brain Mapp. 4, 74–90.

Worsley, K., Marrett, S., Neelin, P., Friston, K., Evans, A., 1996b. A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. 4, 58–73.

Worsley, K., Andermann, M., Koulis, T., MacDonald, D., Evans, A., 1999. Detecting changes in non-isotropic images. Hum. Brain Mapp. 8, 98–101.

Worsley, K., Liao, C., Aston, J.A.D., Petre, V., Duncan, G., Morales, F., Evans, A., 2002. A general statistical analysis for fMRI data. NeuroImage 15, 1–15.