Structural adaptive segmentation for statistical parametric mapping

Jörg Polzehl a,⁎, Henning U. Voss b, Karsten Tabelow a,1

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

A R T I C L E   I N F O
Article history:
Received 23 February 2010
Revised 12 April 2010
Accepted 16 April 2010
Available online 24 April 2010

Keywords:
Image enhancement
Functional MRI
Structure adaptive smoothing
Multiscale testing

A B S T R A C T
Functional Magnetic Resonance Imaging inherently involves noisy measurements and a severe multiple test problem. Smoothing is usually used to reduce the effective number of multiple comparisons and to locally integrate the signal and hence increase the signal-to-noise ratio. Here, we provide a new structural adaptive segmentation algorithm (AS) that naturally combines the signal detection with noise reduction in one procedure. Moreover, the new method is closely related to a recently proposed structural adaptive smoothing algorithm and preserves shape and spatial extent of activation areas without blurring their borders.

© 2010 Elsevier Inc. All rights reserved.

Introduction

The challenges of the measurement and analysis of functional Magnetic Resonance Imaging (fMRI) data (Friston et al., 2007; Lazar, 2008) are manifold. These include noisy measurements and an inherently severe multiple-comparison problem. For a recent discussion see Kriegeskorte and Bandettini (2007). These two challenges and their interplay will be in the focus of investigation in the current manuscript.

Since the functional signal is spatially distributed smoothing is frequently applied to increase the signal-to-noise ratio. Simultaneously it reduces the number of independent tests for signal detection. However, classical nonadaptive smoothing comes at the cost of decreased effective spatial resolution. A large number of algorithms has been developed in the past to overcome this limitation, based on different methodology. While it is impossible to name all of them we here refer only to a small and certainly incomplete selection which is closely related to the approach of this manuscript and use scale space ideas or adaptive region growing (Poline and Mazoyer, 1994ab; Worsley, 2001; Lu et al., 2003; Kriegeskorte et al., 2006; Harrison et al., 2008).

We recently proposed an algorithm based on the structural adaptation principle (Tabelow et al., 2006), which has been proven to be especially useful for high-resolution fMRI (Tabelow et al., 2009). There the statistical parametric map from a linear model is smoothed adaptively to preserve the shape and spatial extent of the activation areas. Since under the null hypotheses the standardized parameter values approximately form a random t-Field, signal detection can then be performed using thresholds obtained from Random Field Theory (RFT) (Adler, 2000; Worsley, 1994, 2003). Although this sequential procedure is statistically correct it is desirable to combine adaptive smoothing and signal detections in a way that solves the multiple-comparison problem and to directly tackle the noise and multiple test problem of fMRI analysis at once. Such integration is possible since the information used to generate weighting schemes in the structural adaptive procedure (Tabelow et al., 2006) can also be used for signal detection. The new proposed method will lead to similar signal detection results, but is conceptionally more coherent and uses much less approximating assumptions concerning the signal detection. It also provides a much more computationally efficient algorithm. See Section 5 for a comprehensive discussion.

The manuscript is organized as follows: In Section 2 we present the problem, some preliminaries from extreme value theory and multiscale tests and a proposal for multiscale tests on fMRI data. Section 3 introduces our new structural adaptive segmentation (AS) procedure and discusses the necessary parameter settings. In Section 4 we present results on two artificial and one experimental high resolution fMRI data set. We summarize results in the conclusions.

Theoretical background

Structural adaptive segmentation (AS) combines ideas from scale space analysis (Chaudhuri and Marron, 2000), multiscale testing (Dümbgen and Spokoiny, 2001; Poline and Mazoyer, 1994a) and structural adaptive smoothing (Polzehl and Spokoiny, 2006; Tabelow et al., 2006). The approach described below directly provides an
efficient solution to the problems of structure preserving denoising and signal detection in fMRI.

Within our approach we try to solve the following test problem, which occurs in signal detection in fMRI. Let $V$ be a set of voxels within a predefined region of interest (ROI) or the whole data cube. Denote by $\mathbb{I} = \{\gamma_i\}_{i=1}^N$ a statistical parametric map (SPM) of estimated parameters or contrasts $\gamma_i$ and by $\mathbb{D}(\gamma)$ their corresponding variance estimates. Our interest is in deciding if and in which region there is an activation related to the experiment. Usually, an activation in voxel $i \in V$ is detected if a suitable test rejects the hypothesis $H: \gamma_i = 0$, i.e., if $\gamma_i$ significantly deviates from zero. Here, we take a more general approach by assigning an activation to voxel $i \in V$ if the value $\gamma_i$ (or $|\gamma_i|$) exceeds a prescribed value $\delta \geq 0$, i.e., we consider voxels with contrast smaller than $\delta$ as uninteresting. Note that the commonly used hypothesis $H$ is included for $\delta = 0$.

Adjustment for multiple testing can be done by requiring that if the hypothesis $H: \gamma_i \leq \delta$ (or $H: |\gamma_i| \leq \delta$ in case of a two sided alternative) is true for all voxel $i \in V$. For a suitable test to probability the rejection to the hypothesis in any voxel $i \in V$ should be less or equal a prescribed significance level $\alpha$. In other words we need to construct a test for the hypothesis

$$H : \max_{i \in V} \gamma_i \leq \delta \quad \text{(or $\max_{i \in V} |\gamma_i| \leq \delta$).} \quad (1)$$

We will then use the values of the corresponding test statistic to provide a segmentation of the set $V$ into two (three) classes: The first class corresponds to regions where the test does not reject the hypothesis. The other two classes contain regions where the hypothesis is rejected and the observed values are significantly larger than $\delta$ (or smaller than $-\delta$).

**Preliminaries: Extreme value theory and multiscale tests**

We first introduce a basic result from extreme value theory, see e.g., Resnick (1987), that dates back to Fisher and Tippett (1928) and Gnedenko (1943) and plays a central role in the construction of the multiscale tests proposed here.

Let $X_{n;i}, n \geq 1$ be an i.i.d. sequence of random variables with cumulative distribution function $F(x)$ and let $M_n = \max_{i=1,\ldots,n} X_i$. The distribution function of $M_n$ is the $n$-th power $F^n(x)$ of $F$. Denote by $P_L$ the probability of some event. We then have, see e.g., Gnedenko (1943); Resnick (1987), the following:

**Proposition.** (Gnedenko (1943); Resnick (1987)). Suppose there exist $a_n \to 0$ and $b_n \uparrow (n \geq 1)$ such that

$$P \left[ \frac{M_n - b_n}{a_n} \leq x \right] = F^n(a_n x + b_n) \to G(x)$$

weakly as $n \to \infty$ where $G$ is nondegenerate. Then the limiting distribution $G$ is of one of the following three classes:

(i) $\Phi_\alpha(x) = \begin{cases} 0 : & x < 0 \\ \exp(-\alpha x) : & x \geq 0 \end{cases}$ for some $\alpha > 0$

(ii) $\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) : & x < 0 \\ 1 : & x \geq 0 \end{cases}$ for some $\alpha > 0$

(iii) $\Lambda(x) = \exp(-e^{-x}) \quad x \in \mathbb{R}$

These functions characterize the three types of possible extreme value distributions depending on the tail behavior of the distribution function $F$.

Our proposal is motivated by a class of multiscale tests proposed in Dümbgen and Spokoiny (2001). The authors consider as a special case the problem of testing qualitative hypotheses on the regression function $f$ in the univariate regression model

$$Y_i = f(x_i) + \epsilon_i = f_i + \epsilon_i \quad \text{for} \quad i = 1, \ldots, n$$

with i.i.d. Gaussian errors $\epsilon_i \sim N(0, \sigma^2)$ and equidistant design $x_i = (i-1/2)/n$. In Section 2 of their paper the authors propose a multiscale test for the hypothesis $H_0: f_i \equiv 0 \forall i$, against the alternative $H_1: f_i \equiv \gamma$ for some $\gamma \neq 0$, that is rate optimal over arbitrary Hölder smoothness classes with respect to the supremum norm.

Let $K$ be a kernel with bounded total variation and compact support $[-1, 1]$. For the regression case define $K_n(x) = K((x-t)/h)$ and a standardized estimate

$$\hat{\Psi}_n(t, h) = \frac{1}{\sigma} \left( \sum_{i=1}^n K_n(x_i) \right)^2 / \left( \sum_{i=1}^n K_n(x_i)^2 \right)$$

The random variables $\hat{\Psi}((2j-1)h), h \to 0$, are independent Gaussian with variance 1 and, under the hypothesis $H_0$, expectation 0. Therefore, under $H_0$ and for suitable $a_n(h)$ and $b_n(h)$, the asymptotic distribution of the normalized maxima, $\hat{\Psi}_n((2j-1)h), b_n(h)$ of these $n(h) = (2h)^{-1}$ variables is in the domain of attraction of the Gumbel distribution $\Lambda$. The test statistic in Dümbgen and Spokoiny (2001) is given as

$$T(Y) = \sup_{h \in [0, 1/2]} \sup_{c \in (0, 1-1/h)} (\hat{\Psi}(Y(h), c))$$

with $c(h) = \sqrt{2 \ln(1/h)}$.

The additive correction term $C(2h)$ is obtained as the normalizing constant $b_n(h), n(h) = \frac{1}{\sqrt{n}}$, while $a_n(h) = b_n(h)^{-1} \to 0$ for $h \to 0$ secures that $\sup_{i \in [0, 1/h]} \hat{\Psi}_i(th), h \to 0 \equiv C(2h) + o_n(h)^2$. Therefore the supremum over $t$ for different $h$ are comparable in size and information on different scales $h$ is synchronized. Note that in (2) it suffices to consider pairs $(t, h)$ such that $t = j/n$ and $h = d/n$ for integers $d = [1, n/2]$ and $j = [d, n-d]$. Critical values for this test are well defined and obtained by simulation.

**Variance estimates in the linear model for fMRI**

Let $Y_{it}$ be the observed fMRI data, $i = V$ the voxel index in the image volume (or ROI), $t = 1, \ldots, T$ the index of observation times and $X = (X_1, \ldots, X_T)$ a design matrix describing both stimulus effects and possible trends in time. Let us assume that prewhitening in time has been done as part of data preprocessing. We use the linear model

$$Y_{it} = X_i \beta_t + \epsilon_{it}$$

with $\text{Var}_{it} = \sigma_t^2 \beta_t \in \mathbb{R}^T$ to describe the experiment. Let $c$ be a contrast vector and $\gamma_t = c^T \beta_t$. Let us assume that we have weighting schemes $W_t = (w_{ij})_{i,j}$ containing weights associated with a pair of voxel $i$ and $j$. For the moment we will consider $w_{ij} = K_{ij}(\delta h)$ where $K$ is a positive kernel function integrating to 1 and with finite second moment, $\delta(i,j)$ is the spatial distance between the centers of voxel $i$ and $j$, and $h$ is a bandwidth. We define spatially smoothed versions of $Y_{it}$ as

$$Y_{it}^{(h)} = \sum_{j \in V} w_{ij} Y_{jt} / \sum_{j \in V} w_{ij}$$

and denote by $\gamma_{it}^{(h)} = c^T \beta_{it}^{(h)}$ and $\hat{\delta}_{it}^{(h)} = (c^T \hat{\beta}_{it}^{(h)})^c$ the estimated contrast and its estimated standard variance obtained from the model

$$Y_{it}^{(h)} = X_i \gamma_{it}^{(h)} + \epsilon_{it}^{(h)}.$$

\footnote{The proof of this property, given for a more general form of (2) in Dümbgen & Spokoiny, 2001, relies on an extension of Levy’s modulus of continuity for Brownian motion, cf. (Shorack & Wellner, 1986).}
Note that estimating $\hat{\gamma}_i^{(h)}$ from model (4) coincides with smoothing estimates of $\gamma_i$ obtained from the model (3) using the weighting scheme $W_i$ due to the linearity of the equations. The quantities

$$\frac{\hat{\gamma}_i^{(h)}}{\sqrt{D \hat{\gamma}_i^{(h)}}}$$

are (approximately) $t$-distributed random variables with some $v$ degrees of freedom, with $v$ depending on $T$ and the correlation of $Y_i$ in time. Note, that $D \hat{\gamma}_i^{(h)} = c^2(X^T X)^{-1} c \hat{\delta}_h$, where $\hat{\delta}_h$ is the residual variance.

**Multiscale tests on fMRI data**

Statistical parametric maps (SPM) obtained in fMRI studies usually form a random field of $t$-distributed variables with $v$ degrees of freedom. According to the proposition in Section 2.1 the limiting distribution in this case is a Fréchet distribution with shape parameter $v$ and cumulative distribution function $\Phi(x) = \exp(-x^{-v})$ for $x \geq 0$. The sequences $a_n(v)$ and $b_n(v)$ here depend on $v$.

Following the idea from Dümbgen and Spokoiny (2001) we are now ready to propose multiscale tests for this situation. Let us consider the one sided hypothesis $H_{0i}: \gamma_i \leq \delta$ $\forall i$ (alternative $H_{1i}: \exists i \gamma_i > \delta$) where $\delta \geq 0$ would correspond to a maximum contrast value that is considered uninteresting. In the classical fMRI analysis $\delta$ equals 0.

Our proposal for a test statistic for this setting is

$$T_{\{i}^{(m)} = \max_{n \in N} \max_{v \leq \sqrt{n}} \frac{\left(\hat{\gamma}_i^{(h)} - \delta\right)}{a_n(h)(v) \sqrt{D \hat{\gamma}_i^{(h)}}} - \frac{b_n(h)(v)}{a_n(h)(v)}$$

(5)

where $\mathcal{N}$ denotes a set of bandwidths and $n(h)$ denotes a number of independent experiments that carries the same information as the random field $\hat{\Gamma} = \{\hat{\gamma}_i^{(h)}\}_{i \in \mathcal{V}}$. More formally, for homogeneous $\sigma_i^2$, $n(h)$ can be chosen as $n(h) = \frac{\hat{\delta}}{D \hat{\gamma}_i^{(h)}}$ $\forall n(h)$, i.e. depending on the variance reduction achieved at bandwidth $h$. Here $\hat{\delta} \mathcal{N}\text{Cor} \rightarrow \mathcal{N}$ corresponds to a set of independent resolution elements within a specified region of interest. It depends on both the size of this region as well as the spatial correlation observed in the original fMRI time series. Again normalization by $a_n(h)$ and $b_n(h)$ is used to make contributions for different bandwidths $h$ to the test statistic comparable.

Note that the critical values for this test are obtained under the least favorable situation $\gamma_i \equiv \delta$ $\forall i$ and therefore do not depend on $\delta$. If any $\gamma_i < \delta$ the resulting test will be conservative.

In a similar way we may define a two sided test for $H_{0i}: \mid \gamma_i \mid \leq \delta$ $\forall i$ (alternative $H_{1i}: \exists i \mid \gamma_i \mid > \delta$) as

$$T_{\{i}^{(m)} = \max_{n \in N} \max_{v \leq \sqrt{n}} \frac{\left(\mid \hat{\gamma}_i^{(h)} \mid - \delta\right)}{a_n(h)(v) \sqrt{D \hat{\gamma}_i^{(h)}}} - \frac{b_n(h)(v)}{a_n(h)(v)}$$

The distribution of $T_{\{i}^{(m)}$ can, for given $v$ and $\mathcal{N}$ be approximated by simulation under the hypothesis $H_{0i}: \gamma_i = 0$ $\forall i$.

To get a good approximation of the distribution $F^\{\mathcal{N}\}$ of $M_n = \max_{v \leq \sqrt{n}} \hat{\gamma}_i^{(h)}$ by its limiting distribution $\Phi_v$, we select normalizing sequences $a_n(v)$ and $b_n(v)$ such that

$$\ln(F^\{\mathcal{N}\}(a_n(v)x_k + b_n(v))) = n \ln(F[a_n(v)x_k + b_n(v)])$$

$$= n \ln \left(1 - \frac{1}{2} + \frac{v}{\left|a_n(v)x_k + b_n(v)\right|} \left(\frac{v}{2} \frac{1}{2}\right)^2\right)$$

$$\approx \ln \Phi_v(x_k) = -X_k^v,$$

with $I_v(a, b)$ denoting the incomplete Beta function, is a good approximation for a set of quantiles $x_k$, $k = 1, \ldots, n$, of $\Phi_v$. This is achieved by minimizing a mean relative error

$$Q(a_n(v), b_n(v)) = \sum_{i=1}^n \left( \ln \left[1 - \frac{1}{2} + \frac{v}{\left|a_n(v)x_k + b_n(v)\right|} \left(\frac{v}{2} \frac{1}{2}\right)^2\right] \right) x_k^v + 1 \right)^2. \right.$$

The interval of quantile levels $(\alpha, 1 - \alpha)$ considered corresponds to the range of values needed to determine thresholds in the segmentation algorithm. Values $a_n(v)$ and $b_n(v)$ have been obtained numerically for a logarithmic grid of $n$, ranging from 250 to $2 \cdot 10^6$ and a dense grid of $v$, ranging from 40 to 1000, degrees of freedom. The resulting arrays $a_n(v)$ and $b_n(v)$ have then been approximated by a smooth function, see Fig. 1.

**Structural adaptive segmentation (AS)**

The Propagation-Separation approach introduced in Polzehl and Spokoiny 2006 provides an iterative smoothing procedure based on adaptive weighting schemes $W_i$. One essential property of this approach is that - in the case of no spatial structure - the resulting estimates behave like the nonadaptive kernel estimates used in the previous section. This case corresponds to the hypothesis of no activation in fMRI experiments. More precisely, the generated adaptive weighting schemes resemble their nonadaptive counterparts with high probability. This property will enable us to generalize the tests proposed in the last subsection to the use of such adaptive weighting schemes.

**The algorithm**

Let $K_{\alpha e}$ and $K_{\alpha}$ be nonincreasing kernel functions with compact support $[0,1]$. Let $\mathcal{N} = \{n_0, \ldots, n_h\}$ be a sequence of exponentially growing bandwidths, i.e. $h_b = c_n h_{b-1}$ for some $c_n > 1$. Nonadaptive smoothing with kernel $K_{\alpha e}$ and this sequence of bandwidths produces a sequence of SPM’s with a variance reduction of approximately $c_n^2$ from step to step. Note that for a Gaussian kernel $K_{\alpha e}$ non-adaptive weights coincide with the weights usually applied in conventional smoothing.

![Fig. 1. Approximation of the determined values for the sequences $a_n$ and $b_n$ by a smooth function and dependence on $v$ and $n$. Top: $1/(a_n(v))$. Bottom: $b_n(v)/a_n(v)$. Both plots are scaled by some arbitrary value for the sake of visualization.](image-url)
Adaptive weighting schemes $W_{ij}^{(h)}$ are generated as a product of two terms: The first term is a nonadaptive weight depending on the distance between voxel $i$ and $j$ weighted by a bandwidth $h$. The second is a stochastic term depending, e.g. on the Kullback-Leibler distance between voxel

\[ \frac{\hat{q}_{ij}^{(h)}}{c(X')X^{-1}c_{ij}^{T}} \]

Here $\sigma^2_{\hat{q}^{(h)}}$ is the variance of $Y_{ij}^{(h)}$. Note that this Kullback-Leibler distance is, for suitable large $v$, well approximated by

\[ \frac{\hat{q}_{ij}^{(h)}}{c(X')X^{-1}c_{ij}^{T}} \]

More formally the adaptive weights take the form

\[ W_{ij}^{(h)} = K_{loc}\left(\frac{\hat{q}(i,j)}{h}\right)K_{st}\left(\frac{\left(\hat{q}_{ij}^{(h)} - \hat{q}_{ij}^{(h-1)}\right)^2}{\lambda c^2(X')X^{-1}c_{ij}^{T}h_{loc}^{2}}\right)\]

for a suitably chosen value of $\lambda$ and $\sigma^2_{\hat{q}^{(h)}}$ the residual variance in (4).

Let $e_i$ denote the residuals from estimating $\beta$ in model (3). We will use $\hat{z}_i \in \{-1, 0, 1\}$ to denote if voxel $i$ has been classified to have significantly negative, nonsignificant or significantly positive contrast. The resulting algorithm, for a two sided alternative, has the following form

- **Initialization:** Start with an nonadaptive weighting scheme $W_{ij}^{(h_0)} = K_{loc}\left(\frac{\hat{q}(i,j)}{h_0}\right)$. For all $i \in V$ compute smoothed estimates of $Y_i$ as

\[ \hat{Y}_i^{(h)} = \sum_{j \in V} w_{ij}^{(h)} \hat{q}_j / \sum_{j \in V} w_{ij}^{(h)} \]

and generate estimates of $\sigma^2_{\hat{q}(h)}$ from smoothed residuals

\[ \hat{\sigma}^2_{\hat{q}(h)} = \frac{1}{T-P} \sum_{t=1}^{T} e_{it}^{(h)} \]

Set $\hat{z}_i = 0$ for all voxel and $k = 1$.

- **Generate weighting schemes $W_{ij}^{(h)} = \{W_{ij}^{(h)}\}_{j \in V} \forall i$, as**

\[ W_{ij}^{(h)} = \begin{cases} K_{loc}\left(\frac{\hat{q}(i,j)}{h_{k}}\right)K_{st}\left(\frac{\left(\hat{q}_{ij}^{(h)} - \hat{q}_{ij}^{(h-1)}\right)^2}{\lambda c^2(X')X^{-1}c_{ij}^{T}h_{loc}^{2}}\right): \hat{z}_i \hat{z}_j = 0 \\ K_{loc}\left(\frac{\hat{q}(i,j)}{h_{k}}\right): \hat{z}_i \hat{z}_j = 1 \\ 0: \hat{z}_i \hat{z}_j = -1 \end{cases} \]

- **Compute new (smoothed) parameter estimates as**

\[ \hat{Y}_i^{(h)} = \sum_{j \in V} W_{ij}^{(h)} \hat{q}_j / \sum_{j \in V} W_{ij}^{(h)} \]

and corresponding variance estimates as $\hat{D}\hat{Y}_i^{(h)} = c^T(X')X^{-1}c\sigma^2_{\hat{q}(h)}$ with

\[ \hat{\sigma}^2_{\hat{q}(h)} = \frac{1}{T-P} \sum_{t=1}^{T} e_{it}^{(h)} \]

\[ e_{it}^{(h)} = \sum_{j \in V} W_{ij}^{(h)} e_{jt} / \sum_{j \in V} W_{ij}^{(h)} \]

- **Iterate:** Set $k := k + 1$. If $k \geq h^*$ stop, else continue with second step.

The use of the intermediate segmentation results (15) in the definition of weights (10) leads to a nonadaptive estimate within detected segments. Adaptive weights are used if the hypothesis has not been rejected for one of the voxels.

We summarize the main steps of the proposed structural adaptive segmentation algorithm:

**Overview over algorithm (for each voxel $i$)**

1. Pre-process fMRI data $Y_{it}$ (motion-correction, pre-whitening, no smoothing, ...).
2. Generate SPM $\hat{Y}_i = c^T\hat{\beta}_i$ and its variance $\hat{D}\hat{Y}_i$ from the linear model (3).
3. Choose parameter $h_{k^*}$ (and $\hat{\delta}$). This determines $\mathcal{H} = \{h_{0}, ..., h_{k^*}\}$.
4. Iterate over $k = 0, ..., k^*$:
   - Determine weighting schemes $W_{ij}^{(h_k)}$ using Eq. (10) based on $\hat{Y}_i^{(h_k)}$ and $\hat{z}_i$ from previous iteration step.
   - Use $W_{ij}^{(h_k)}$ to compute smoothed SPM $\hat{Y}_i^{(h_k)}$ and variance $\hat{D}\hat{Y}_i^{(h_k)}$ using Eqs. (11)–(13).
   - If $\hat{z}_i = 0$ use $\hat{Y}_i^{(h_0)}$ (and its variance $\hat{D}\hat{Y}_i^{(h_0)}$) to perform test Eq. (14) to set $\hat{z}_i$.
5. After iteration: if $\hat{z}_i = 0$ the hypotheses (1) of no activation could not be rejected. If $\hat{z}_i = \pm 1$ voxel $i$ belongs to an activated region.

**Adjustment of parameters**

The procedure depends on several parameters that serve different purposes. The first class of parameters refers to properties of the adaptive smoothing procedure and consists of the scale parameter $\lambda$ in (10), and the sequence of bandwidths $\mathcal{H}$. The second group consists of $\delta$ and the critical value $\tau$ in (14). A third group includes the choice of kernels $K_{loc}$ and $K_{st}$.

**Smoothing parameters**

The sequence of bandwidths $\mathcal{H}$ can be chosen as $h_k := h_{0}^{\xi_k} h_0$ with initial bandwidth $h_0$ such that $K_{loc}(\hat{q}(i,j)/h_0) > 0 \iff i,j \in \mathcal{E}$ and $\xi_0 = 1.25$. This leads to a sequence of exponentially growing bandwidths and corresponds to the settings used in Polzehl and Spokoiny 2006; Tabelow et al. (2006, 2009). The maximal bandwidth $h_{k^*}$ may be chosen as a maximal expected diameter of activated regions.

The most important smoothing parameter is $\lambda$ that controls the sensitivity of adaptation. Note, that this parameter does not depend on the signal and noise variance within the data. The parameter $\lambda$ can
Table 1 Parameters of the proposed structural adaptive segmentation algorithm.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Interpretation</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa )</td>
<td>determines possible loss of efficiency by adaption under the hypothesis</td>
<td>fixed as ( \kappa = 0.05 )</td>
</tr>
<tr>
<td>( c_0 )</td>
<td>growth factor in series of bandwidths</td>
<td>fixed as ( c_0 = 1.25^{1/3} ) given by propagation condition (16)</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>scale parameter in Eq. (10), depends on ( \kappa )</td>
<td>fixed as quantiles of the distribution of the test statistic for given degrees of freedom ( v, \lambda^*, n_{\text{ROI}} ), see Subsection 3.2.2, user choice ( \hat{\kappa} \geq 0 ), see Section 2.3</td>
</tr>
<tr>
<td>( \tau )</td>
<td>critical values of the test statistic</td>
<td></td>
</tr>
<tr>
<td>( \delta )</td>
<td>maximal amplitude of the signal considered as uninteresting</td>
<td></td>
</tr>
<tr>
<td>( k^* ) (or ( h_0 ))</td>
<td>number of iterations / maximal bandwidth</td>
<td>maximal expected diameter of activated regions; user choice</td>
</tr>
</tbody>
</table>

Thus be chosen by simulation, as the smallest value that complies with the following propagation condition, see also Tabelow et al. (2006):

**Propagation condition.** Let \( \hat{\gamma}_i^{(k)} = \sum_j K_{\text{loc}} \left( \frac{u(i,j)}{h_k} \right) \gamma_j / \sum_j K_{\text{loc}} \left( \frac{u(i,j)}{h_k} \right) \) be the nonadaptive estimate using kernel \( K_{\text{loc}} \) and bandwidth \( h_k \). We require that for a SPM \( \Gamma \) with \( \gamma_i \equiv 0 \) and \( \tau = \phi \) the adaptive estimates \( \hat{\gamma}_i^{(k)} \) fulfill the propagation condition

\[
E[\hat{\gamma}_i^{(k)} - \gamma_i] < (1 + \kappa) E[\hat{\gamma}_i^{(k)} - \gamma] \tag{16}
\]

for some prespecified \( \kappa > 0 \).

We use a value corresponding to \( \kappa = 0.05 \).

**Parameters of the multiscale test**

The normalizing sequences \( a_{\text{null}}(t) \) and \( b_{\text{null}}(t) \) are specified using the approximation described in the end of Subsection 2.3. Under the hypothesis (for \( \delta = 0 \)) \( n(t,h) \approx n(h) \) while in case of deviations from \( H_0 \) we usually observe \( n(t,h) \geq n(h) \) due to the adaptive weighting scheme.

Empirical distributions for the test statistic \( T_2 \) have been obtained by simulation (simulation size 10000) under the hypothesis \( \delta = 0 \) for a grid of reasonable values for \( v, \lambda, k^* \) and \( n_{\text{ROI}} \). The critical value \( \tau \) mainly depends on the significance level and degrees of freedom \( v \). In the simulated empirical distributions we observe a slight dependence on the chosen value of the smoothing parameter \( \lambda \), the number of iteration steps \( k^* \), or equivalently the maximum bandwidth, and the field of view \( n_{\text{ROI}} \). Values for \( \tau \) used in our implementation are smooth interpolations between the appropriate quantiles of the empirical distributions. The same critical values are used in case of \( \delta > 0 \) although the resulting test will be conservative. The parameters of the proposed algorithm are summarized in Table 1.

Kernels

We use

\[
\begin{align*}
K_{\text{loc}}(x) &= \left\{ \begin{array}{ll}
1 & : x < 1/2 \\
2(1-x) & : 1/2 \leq x < 1 \\
0 & : x \geq 1
\end{array} \right. \quad K_{\text{st}}(x) = \left\{ \begin{array}{ll}
1 & : x < 1/4 \\
4/3(1-x) & : 1/4 \leq x < 1 \\
0 & : x \geq 1
\end{array} \right.
\end{align*}
\]

as kernel functions. They have only minor influence on the results and are chosen for speed of computation.

**Results**

We applied the new AS procedure to two artificial datasets and one experimental fMRI dataset. Experimental data was motion-corrected using afni (Cox, 1996). In all cases we pre-whitened the data using an AR(1) model for the temporal autocorrelation. All datasets were processed using the package fmri for R (R foundation for Statistical Computing) (Polzehl and Tabelow, 2007; R Development Core Team, 2006). As discussed in the preceding section most parameters of the algorithm do not depend on the actual data and are determined and fixed in advance by simulation (\( \lambda, \tau \)) or convenience (\( \kappa, c_0 \)). The user-defined parameters \( h_k \)- and \( \delta \) are given with the corresponding results.
In Fig. 4 we show the results of signal detection in these artificial datasets with different methods. The first column corresponds to the detection of the $R$-$L$, and $L$-$R$ contrast segments using the structural adaptive segmentation method described in this paper at significance level $\alpha = 0.05$. The second column provides the signals detected for both contrasts (blue for $R$-$L$ and red for $L$-$R$) using the structural adaptive smoothing method from Tabelow et al. (2006) with multiple test correction via RFT ($\alpha = 0.05$). The third column is the result for non-adaptive smoothing, while the right column shows results of a voxelwise analysis. Both adaptive smoothing methods are able to detect the simulated stripe structures at high resolution and low signal-to-noise, cf. Tabelow et al. (2009). This shows, that the adaptive segmentation method proposed in this paper has a similar sensitivity. Non-adaptive smoothing inherently is not able to detect such alternating structures and leads to significant blurring and a loss in information on the shape of the activation region. The low SNR prevents the detection without smoothing, or with only very small bandwidths that may be used to avoid blurring and partial voluming. Note that in all cases the signal detection is corrected for the large number of multiple tests. We used $\delta = 0$ and a maximum bandwidth $h_k^* = 6,3,2$ for the high, medium, and low resolution data, respectively.

**Artificial data II**

To compare the results with our previous work in Tabelow et al. (2009), we re-analyze the numerical phantom first created there. It
provides a rich spatial structure of activation areas and allows for a direct comparison of the signal detection using the algorithm proposed here and our former work Tabelow et al. (2006). The phantom consists of 30 slices with three slices containing activation alternated with two slices without. The in-plane resolution corresponds to a matrix size of $128 \times 128$, see Fig. 5(a). Voxel size in z-direction is held constant and set to twice the within slice voxel size. Gray values correspond to the signal strength in the corresponding position.

Structures are replicated with clockwise increasing SNR, i.e., the signal size is $1.25^k$ for $k = 0,...,7$. At each voxel a time series was created with 107 samples, stimulus onset times at the 18th, 48th, and 78th sample with a stimulus duration of 15 samples and 2 s between two samples. Errors are generated from white noise with a standard deviation of 10 by first applying an AR(1) model with parameter 0.3 and followed by a convolution with a Gaussian kernel with FWHM-bandwidths $(1, 1, 0.5)$ times voxel size.

Fig. 5 illustrates the phantom (a) and the relative frequency of signal detection in slices containing activation in the different locations within the phantom for both the structural adaptive smoothing algorithm Tabelow et al. (2009, 2006) and the proposed structural adaptive segmentation procedure. In this setting structural adaptive segmentation was able to detect 13% more activated voxel and to reduce the number of false positives by 42% compared to the adaptive segmentation procedure. In this setting structural adaptive smoothing fMRI analysis (Tabelow et al., 2006).

A somatosensory motor task fMRI scan was performed by one healthy adult male subject within a research protocol approved by the institutional review board of Weill Cornell Medical College. For functional MRI, a GE-EPI sequence with TE/TR = 40/2000 ms was used and 20 axial slices of 4 mm thickness were acquired. We used a field-of-view of 24 cm with a matrix size $128 \times 128$, yielding voxel dimensions of 1.88 mm, respectively. A task was performed in three blocks of 60 s duration; each block consisted of 30 s task and 30 s rest. The first 4 scans before these block were discarded, yielding total 105 scans. The task consisted of bimanual tapping of the thumb against all fingers of the same hand, one by one and in quick succession.

The result of our different signal detection methods for this experiment are shown in Fig. 7. These are three consecutive slices in the primary motor area. The first column reproduces the result of signal detection using classical structural adaptive smoothing and RFT (Tabelow et al., 2009) (multiple test corrected $\alpha=0.05$). Since in this paper our focus is not on comparing our structural adaptive smoothing method with non-adaptive methods, we generate our results using a Plateau-kernel for the localization kernel (see Tabelow et al. (2006) and Section 3.2.3 for details). We consider the differences to be caused by a slightly inaccurate assessment of the realized significance level for the multiple test problem within the original algorithm: The use of Random Field Theory for thresholding requires a correct description of the smoothness of the resulting random field. For experimental data this will always be a rough approximation.

The next three columns show the detected signal segments for different values of $\delta$ and significance level $\alpha$. Apparently, the total size of activation areas reduces with $\alpha$, while at $\alpha=0.05$ the result is comparable to the PS result. The exclusion of low signals $\delta=2$ is shown in column

![Fig. 5](image-url) (a) Phantom used in simulations. Gray values indicate the size of the signal at the corresponding location. Size and structure of activated areas vary radially while the size of the signal increases clockwise by a factor of 1.25 from spoke to spoke. The phantom was first used in Tabelow et al. (2009) and re-analyzed here to directly compare the former results with the new algorithm. (b) Probability of signal detection using the structural adaptive smoothing algorithm as described in Tabelow et al. 2009, 2006. (c) Probability of signal detection using the structural adaptive segmentation algorithm.

### Table 2

Signal detection results depending on signal size for structural adaptive smoothing [Tabelow et al., 2006] (PS) and the proposed segmentation method (AS) as shown in Fig. 5. For signal intensity 0 false positives are reported while other columns contain the number of detected voxels for clockwise increasing SNR.

<table>
<thead>
<tr>
<th>Signal intensity</th>
<th>0</th>
<th>1</th>
<th>1.25</th>
<th>1.25</th>
<th>1.25</th>
<th>1.25</th>
<th>1.25</th>
<th>1.25</th>
<th>1.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of voxel</td>
<td>449760</td>
<td>4860</td>
<td>5580</td>
<td>4860</td>
<td>5580</td>
<td>4860</td>
<td>5580</td>
<td>4860</td>
<td>5580</td>
</tr>
<tr>
<td>Detected voxel (PS)</td>
<td>1710</td>
<td>17</td>
<td>17</td>
<td>509</td>
<td>1530</td>
<td>2411</td>
<td>3889</td>
<td>4233</td>
<td>5499</td>
</tr>
<tr>
<td>Detected voxel (AS)</td>
<td>989</td>
<td>27</td>
<td>206</td>
<td>721</td>
<td>1973</td>
<td>2819</td>
<td>4672</td>
<td>4650</td>
<td>5555</td>
</tr>
</tbody>
</table>
Conclusions

Structural adaptive segmentation (AS) is a new algorithm for the analysis of functional MRI data which is based on the principle of structural adaptation. While achieving similar performance as or previously developed algorithm Tabelow et al. (2006) it has several major conceptual advantages:

- It does not rely on an assumption of a local constant contrast $\gamma$.
- Signal detection and structural adaptive smoothing are integrated within the consecutive steps of the iterative procedure.
- Signal detection is based on multiscale testing rather than Random Field Theory (RFT).
- The procedure provides a decision at a prespecified significance level rather than p-values. Color coding can be used to provide information on the estimated size of the signal $\hat{\gamma}(h_k^*)$, which can be easily interpreted, see Fig. 7.
- The utilized tests collect information on all scales $h_0, \ldots, h_k^*$ visited in the algorithm rather than providing a decision based only on the final smoothing result by RFT.
- In contrast to the PS algorithm the new approach avoids correction terms for spatial correlation within the fMRI data.

Fig. 6. Intermediate results for a single slice of the phantom for steps 1, 5, 10 and final segmentation result for the maximal bandwidth $h_k^* = 3.06$ (from left to right). The rightmost image provides the segmentation result for a maximal bandwidth $h_k^* = 5.85$.

Fig. 7. Signal detection in a sensorymotor experiment. The first column contains the signal detection using our structural adaptive smoothing algorithm (multiple test corrected $\alpha = 0.05$) for comparison. Colors correspond to p-values. The second, third, and fourth column contain the results of our proposed structural adaptive segmentation algorithm using different values for the minimal signal $\delta$, and the significance level $\alpha$. Colors code the size of the estimated signal. We used $h_k^* = 6$ for this analysis. For comparison we again provide the result obtained with non-adaptive Gaussian smoothing from Tabelow et al. (2009). The image orientation is R-L. See online version for color.
The last two characteristics lead to an improved sensitivity with respect to signal detection, see Table 2.

The algorithm provides noise reduction without blurring borders of activation areas. In addition, it gives an intrinsic solution to the multiple test problem. In contrast to multiple test adjustments by random field theory, as used e.g. in our original structural adaptive smoothing method, the present approach does not depend on the use of a Gaussian location kernel. Using kernels with compact support leads to a much more computationally efficient algorithm.

We think the approach can be useful in any situation where low signal-to-noise ratio requires smoothing. The new algorithm can be especially helpful in situations with sharp discontinuities and effects observed only at high spatial resolution. This can be the case in presurgical planning. Although we do not provide a corresponding dataset, the algorithm might prove to be useful for stripe like structures like ocular dominance columns, see Fig. 4 for a simulation.

Acknowledgments

This work is supported by the DFG Research Center Matheon. We thank two anonymous reviewers for their helpful comments.

References