Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Structural adaptive segmentation for statistical parametric mapping

Jörg Polzehl[1], Henning U. Voss[2], Karsten Tabelow[1]

submitted: February 23th, 2010

[1] Weierstrass Institute
for Applied Analysis and Stochastics
Mohrenstr. 39
10117 Berlin
Germany
email: polzehl@wias-berlin.de
tabelow@wias-berlin.de

[2] Citigroup Biomedical Imaging Center
Weill Cornell Medical College
516 East 72nd Street
New York, NY 10021
USA

No. 1484

Berlin 2010

2000 *Mathematics Subject Classification.* 62P10, 92C55, 62G05, 62G10, 68U10, 62G08 .

Key words and phrases. Image Enhancement, Functional Magnetic Resonance Imaging, Structural Adaptive Smoothing, Multiscale Testing.

K. Tabelow was supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

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 the borders.

1 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). Both 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 surely incomplete selection which are in spirit closely related to the approach of this manuscript (Poline and Mazoyer, 1994a,b; Worsley, 2001; Lu et al., 2003; Kriegeskorte et al., 2006; Harrison et al., 2008) and use scale space ideas or adaptive region growing.

We recently proposed a new algorithm based on the structural adaptation principle (Tabe-

low 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 for each bandwidth in the structural adaptive procedure (Tabelow et al., 2006) can also be used for signal detection at the respective scale.

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.

2 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 $\Gamma = \{\widehat{\gamma}_i\}_{i\in V}$ a statistical parametric map (SPM) of estimated parameters or contrasts γ_i and by $\widehat{D}(\widehat{\gamma}_i)$ 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 $\widehat{\gamma}_i$ significantly deviates from zero. Here, we take a more general approach by assigning an activation to voxel $i \in V$ if the value γ_i (or $|\gamma_i|$) exceeds a prescribed value $\delta \geq 0$, i.e. we consider voxels with contrast smaller than δ 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 the probability to reject the hypothesis in *any* voxel $i \in V$ should be less or equal a prescribed significance level α. In other words we need to construct a test for the hypothesis

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

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 δ (or smaller than $-\delta$).

2.1 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 $X_{n,n\geq 1}$ be an i.i.d. sequence of random variables with common cumulative distribution function $F(x)$ and let $M_n = \max_{k=1,\ldots,n} X_k$. The distribution function of M_n is the n-th power $F^n(x)$ of F. Denote by $P[.]$ 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 $a_n > 0$ and b_n $(n \geq 1)$ 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:

$$
\begin{aligned}
(i) \quad \Phi_\alpha(x) &= \begin{cases} 0 & : \quad x < 0 \\ \exp\{-x^{-\alpha}\} & : \quad x \geq 0 \end{cases} \qquad \textit{for some} \quad \alpha > 0 \\
(ii) \quad \Psi_\alpha(x) &= \begin{cases} \exp\{-(-x)^{\alpha}\} & : \quad x < 0 \\ 1 & : \quad x \geq 0 \end{cases} \qquad \textit{for some} \quad \alpha > 0 \\
(iii) \quad \Lambda(x) &= \exp\{-e^{-x}\} \quad x \in R
\end{aligned}
$$

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 $H_1 : |f_i| \geq \rho$ for some $\rho > 0$, 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 $K_{t,h}(x) = K((x-t)/h)$ and a standardized estimate

$$\widehat{\Psi}_n(t,h) = \frac{\sum_{i=1}^n K_{t,h}(x_i)}{\sigma \left(\sum_{i=1}^n K_{t,h}(x_i)^2\right)^{1/2}} \frac{\sum_{i=1}^n K_{t,h}(x_i) Y_i}{\sum_{i=1}^n K_{t,h}(x_i)} = (\boldsymbol{D}\widehat{f}_h(t))^{-1/2} \widehat{f}_h(t)$$

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

$$T(Y) = \sup_{h \in (0,1/2)} \sup_{t \in [h, 1-h]} (|\widehat{\Psi}(t,h)| - C(2h)) \tag{2}$$

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

The additive correction term $C(2h)$ is obtained as the normalizing constant $b_n(h)$, $n(h) = \frac{1}{2h}$, while $a_n(h) = b_n(h)^{-1} \to 0$ for $h \to 0$ secures that $\sup_{t\in[h,1-h]} \widehat{\Psi}_n(th, h) = C(2h) + o_p(1)$.[1] Therefore the suprema 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 $d \in [1, n/2]$ and $j \in [d, n - d]$. Critical values for this test are well defined and obtained by simulation.

2.2 Variance estimates in the linear model for fMRI

Let Y_{it}, be the observed fMRI data, $i \in V$ the voxel index in the image volume (or ROI), $t = 1, \ldots, T$ the index of observation times and $X = (x_1, \ldots, x_p)$ 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

$$Y_{it} = X\beta_i + \epsilon_{it} \tag{3}$$

with $\operatorname{Var} \epsilon_{it} = \sigma_i^2$, $\beta_i \in R^p$ to describe the experiment. Let c be a contrast vector and $\gamma_i = c^T\beta_i$. Let us assume that we have weighting schemes $W_i = \{w_{ij}\}_{j\in V}$ containing weights associated with a pair of voxel i and j. For the moment we will consider $w_{ij}^{(h)} = K\left(\frac{\delta(i,j)}{h}\right)$ 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

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

and denote by $\widehat{\gamma}_i^{(h)} = c^T\widehat{\beta}_i^{(h)}$ and $\widehat{\boldsymbol{D}}\widehat{\gamma}_i^{(h)} = (c^T\widehat{\boldsymbol{D}}\widehat{\beta}_i^{(h)}c)$ the estimated contrast and its estimated standard variance obtained from the model

$$Y_{it}^{(h)} = X\beta_i^{(h)} + \epsilon_{it}^{(h)}. \tag{4}$$

Note that estimating $\widehat{\gamma}_i^{(h)}$ from model (4) coincides with smoothing estimates of $\widehat{\gamma}_i$ obtained from the model (3) using the weighting scheme W_i due to the linearity of the equations.

[1] The proof of this property, given for a more general form of (2) in Dümbgen and Spokoiny (2001), relies on an extension of Levy's modulus of continuity for Brownian motion, cf. Shorack and Wellner (1986).

The quantities

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

are (approximately) t-distributed random variables with some ν degrees of freedom, with ν depending on T and the correlation of Y_{it} in time. Note, that $\widehat{\boldsymbol{D}}\widehat{\gamma}_i^{(h)} = c^T(X^TX)^{-1}c\,\widehat{\sigma}_{i,h}^2$ where $\widehat{\sigma}_{i,h}^2$ is the residual variance.

2.3 Multiscale tests on fMRI data

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

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 $H_0 : \gamma_i \leq \delta \quad \forall i$ (alternative $H_1 : \exists i \quad \gamma_i > \delta$) where $\delta \geq 0$ would correspond to a maximum contrast value that is considered uninteresting. In the classical fMRI analysis δ equals 0.

Our proposal a test statistic for this setting is

$$T_1(\widehat{\Gamma}^{\mathcal{H}}) = \max_{h\in\mathcal{H}} \max_{i\in V} \frac{\left(\widehat{\gamma}_i^{(h)} - \delta\right)}{a_{n(h)}(\nu)\sqrt{\widehat{\boldsymbol{D}}\widehat{\gamma}_i^{(h)}}} - \frac{b_{n(h)}(\nu)}{a_{n(h)}(\nu)} \tag{5}$$

where $\mathcal{H}$ denotes a set of bandwidths and $n(h)$ denotes a number of independent experiments that carries the same information as the random field $\widehat{\Gamma} = \{\widehat{\gamma}_i^{(h)}\}_{i\in V}$. More formally, for homogeneous σ_i^2, $n(h)$ can be chosen as $n(h) = \frac{\widehat{\boldsymbol{D}}\widehat{\gamma}}{\widehat{\boldsymbol{D}}\widehat{\gamma}^{(h)}} n_{ROI}$, i.e. depending on the variance reduction achieved at bandwidth h. Here n_{ROI} corresponds to a number 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 $\gamma_i \equiv \delta \quad \forall i$ and therefore do not depend on δ. If any $\gamma_i < \delta$ the resulting test will be conservative.

In a similar way we may define a two sided test for $H_0 : |\gamma_i| \leq \delta \quad \forall i$ (alternative $H_1 : \exists i \quad |\gamma_i| > \delta$) as

$$T_2(\widehat{\Gamma}^{\mathcal{H}}) = \max_{h\in\mathcal{H}} \max_{i\in V} \frac{\left(|\widehat{\gamma}_i^{(h)}| - \delta\right)}{a_{n(h)}(\nu)\sqrt{\widehat{D}\widehat{\gamma}_i^{(h)}}} - \frac{b_{n(h)}(\nu)}{a_{n(h)}(\nu)}.$$

The distribution of $T_k(\widehat{\Gamma}^{\mathcal{H}})$ can, for given ν and $\mathcal{H}$ be approximated by simulation under the hypothesis $H_0 : \gamma_i = 0 \quad \forall i$.

To get a good approximation of the distribution F^n of $M_n = \max_{i\in V} \widehat{\gamma}_i^{(h)}$ by its limiting distribution Φ_ν we select normalizing sequences $a_n(\nu)$ and $b_n(\nu)$ such that

$$\begin{aligned} \ln(F^n[a_n(\nu)x_k + b_n(\nu)]) &= n\ln(F[a_n(\nu)x_k + b_n(\nu)]) \\ &= n\ln\left[1 - \frac{1}{2} I_{\frac{\nu}{\nu+[a_n(\nu)x_k+b_n(\nu)]^2}}\left(\frac{\nu}{2}, \frac{1}{2}\right)\right] \\ &\approx \ln\Phi_\nu(x_k) \\ &= -x_k^{-\nu}, \end{aligned} \tag{6}$$

with $I_x(a, b)$ denoting the incomplete Beta function, is a good approximation for a set of quantiles $x_{k,k=1,\ldots,K}$, of Φ_ν. This is achieved by minimizing a mean relative error

$$Q(a_n(\nu), b_n(\nu)) = \sum_{k=1}^{K} \left(n\ln\left[1 - \frac{1}{2} I_{\frac{\nu}{\nu+(a_n(\nu)x_k+b_n(\nu)))^2}}\left(\frac{\nu}{2}, \frac{1}{2}\right)\right](x_k)^\nu + 1\right)^2. \tag{7}$$

The interval of quantile levels $(.9, .995)$ considered corresponds to the range of values needed to determine thresholds in the segmentation algorithm. Values $a_n(\nu)$ and $b_n(\nu)$ have been obtained numerically for a logarithmic grid of n, ranging from 250 to $2 \cdot 10^6$ and a dense grid of ν, ranging from 40 to 1000, degrees of freedom. The resulting arrays $1/a_n(\nu)$ and $b_n(\nu)/a_n(\nu)$ have then been approximated by a smooth function, see Figure 1.

3 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

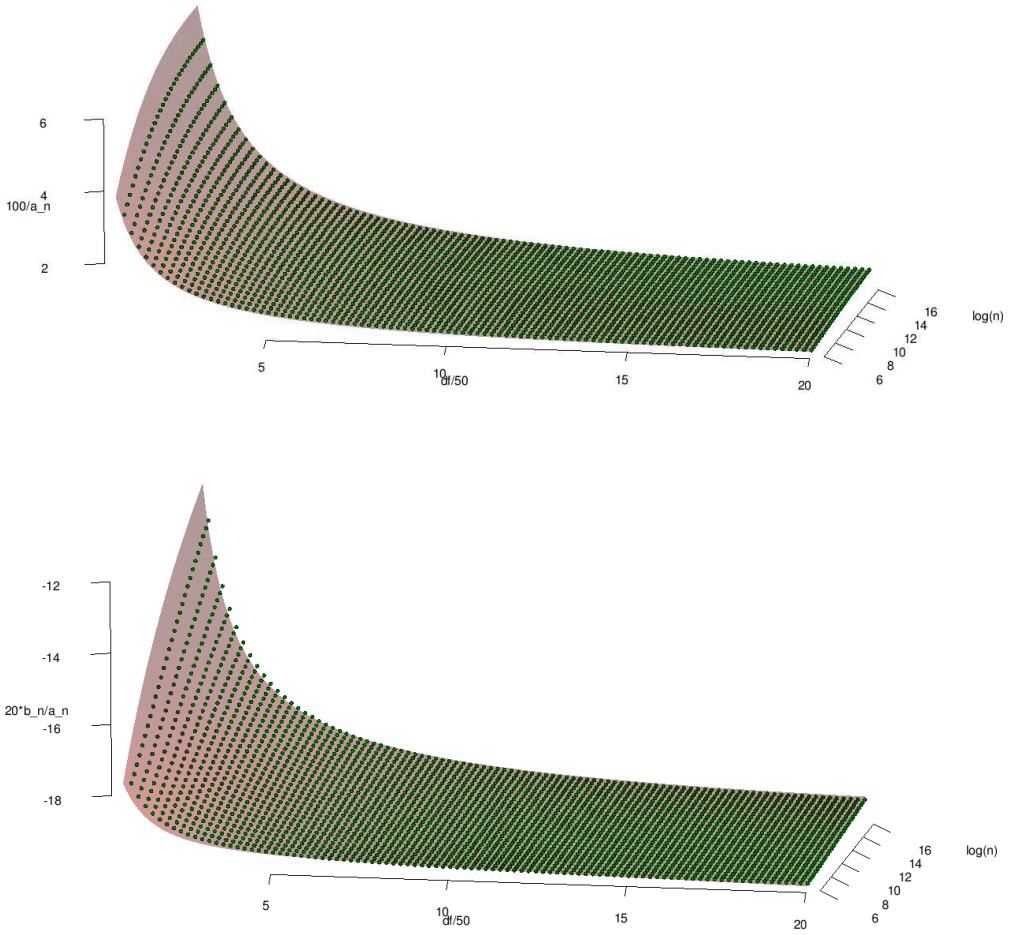

Figure 1: Approximation of the determined values for the sequences a_n and b_b by a smooth function and dependence on ν and n. Top: $1/a_n(\nu)$. Bottom: $b_n(\nu)/a_n(\nu)$. Both plots are scaled by some arbitrary value for the sake of visualization.

high probability. This property will enable us to generalize the tests proposed in the last subsection to the use of such adaptive weighting schemes.

3.1 The algorithm

Let $\mathcal{H} = \{h_0, \ldots, h_{k^*}\}$ be a sequence of exponentially growing bandwidths, and K_{loc} and K_{st} be nonincreasing kernel functions with compact support $[0, 1]$. Adaptive weighting schemes $W_i^{(h)}$ 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 of the non-central t-distributions with noncentrality parameters $\frac{\widehat{\gamma}_i^{(h)}}{c^T(X^TX)^{-1}c\sigma^2_{i,(h)}}$ and $\frac{\widehat{\gamma}_j^{(h)}}{c^T(X^TX)^{-1}c\sigma^2_{j,(h)}}$. Here $\sigma^2_{.,(h)}$ is the variance of $Y^{(h)}_{.t}$. Note that this Kullback-Leibler distance is, for suitable large ν, well approximated by $\frac{\left(\widehat{\gamma}_i^{(h)}-\widehat{\gamma}_j^{(h)}\right)^2}{c^T(X^TX)^{-1}c\sigma^2_{i,(h)}}$.

More formally the adaptive weights take the form

$$w_{ij}^{(h)} = K_{\text{loc}}\left(\frac{\delta(i,j)}{h}\right) K_{\text{st}}\left(\frac{\left(\widehat{\gamma}_i^{(h)} - \widehat{\gamma}_j^{(h)}\right)^2}{\lambda c^T(X^TX)^{-1}c\widehat{\sigma}^2_{i,(h)}}\right)$$

for a suitably chosen value of λ and $\widehat{\sigma}^2_{i,(h)}$ the residual variance in (4).

Let e_{it} denote the residuals from estimating β in model (3). We will use $\zeta_i \in \{-1, 0, 1\}$ 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 $w_{ij}^{(h_0)} = K_{\text{loc}}\left(\frac{\delta(i,j)}{h_0}\right)$. For all $i \in V$ compute smoothed estimates of γ_i as

  $$\widehat{\gamma}_i^{(h_0)} = \sum_{j\in V} w_{ij}^{(h_0)}\widehat{\gamma}_j / \sum_{j\in V} w_{ij}^{(h_0)}$$

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

  $$\widehat{\sigma}^2_{i,(h_0)} = \frac{1}{T-p}\sum_{t=1}^{T}\left(e_{it}^{(h_0)}\right)^2 \tag{8}$$

  $$e_{it}^{(h_0)} = \sum_{j\in V} w_{ij}^{(h_0)} e_{jt} / \sum_{j\in V} w_{ij}^{(h_0)}. \tag{9}$$

  Set $\zeta_i = 0$ for all voxel and $k = 1$.

- Generate weighting schemes $W_i^{(h_k)} = \{w_{ij}^{(h_k)}\}_{j\in V}\ \forall i$, as

  $$w_{ij}^{(h_k)} = \begin{cases} K_{\text{loc}}\left(\frac{\delta(i,j)}{h_k}\right) K_{\text{st}}\left(\frac{(\widehat{\gamma}_i^{(h_{k-1})}-\widehat{\gamma}_j^{(h_{k-1})})^2}{\lambda c^T(X^TX)^{-1}c\widehat{\sigma}^2_{i,(h_{k-1})}}\right) & : \ \zeta_i\zeta_j = 0 \\ K_{\text{loc}}\left(\frac{\delta(i,j)}{h_k}\right) & : \ \zeta_i\zeta_j = 1 \\ 0 & : \ \zeta_i\zeta_j = -1 \end{cases} \tag{10}$$

- Compute new (smoothed) parameter estimates as

  $$\widehat{\gamma}_i^{(h_k)} = \sum_j w_{ij}^{(h_k)}\widehat{\gamma}_j / \sum_j w_{ij}^{(h_k)} \tag{11}$$

and corresponding variance estimates as $\widehat{D}\widehat{\gamma}_i^{(h_k)} = c^T(X^TX)^{-1}c\,\widehat{\sigma}^2_{i,(h_k)}$ with

$$\widehat{\sigma}^2_{i,(h_k)} = \frac{1}{T-p}\sum_{t=1}^{T}(e_{it}^{(h_k)})^2 \tag{12}$$

$$e_{it}^{(h_k)} = \sum_{j\in V} w_{ij}^{(h_k)} e_{jt} / \sum_{j\in V} w_{ij}^{(h_k)} \tag{13}$$

- If $\zeta_i = 0$ and

$$\frac{(|\widehat{\gamma}_i^{(h_k)}| - \delta)}{a_{n(i,h_k)}(\nu)(\widehat{\boldsymbol{D}}\widehat{\gamma}_i^{(h_k)})^{1/2}} - \frac{b_{n(i,h_k)}(\nu)}{a_{n(i,h_k)}(\nu)} > \tau \tag{14}$$

  with $n(i,h_k) = \frac{\boldsymbol{D}(e_{it}^{(h_k)})}{\boldsymbol{D}(e_{it})}\, n_{ROI}$ set

$$\zeta_i = \operatorname{sign}\widehat{\gamma}_i^{(h_k)} \tag{15}$$

- Iterate: Set $k := k+1$. If $k > k^*$ 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 voxel.

3.2 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 λ in (10), and the sequence of bandwidths $\mathcal{H}$. The second group consists of δ and the critical value τ in (14). A third group includes the choice of kernels K_{loc} and K_{st}.

3.2.1 Smoothing parameters

The sequence of bandwidths $\mathcal{H}$ can be chosen as $h_k = c_h^k h_0$ with initial bandwidth h_0 such that $K_{\text{loc}}(\delta(i,j)/h_0) > 0 \Leftrightarrow i \equiv j$ and $c_h = 1.25^{1/3}$. 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 λ that controls the sensitivity of adaptation. Note, that this parameter does not depend on the signal and noise variance within the data. The parameter λ can 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* $\check{\gamma}_i^{(k)} = \sum_j K_{\mathrm{loc}}\left(\frac{\delta(i,j)}{h^{(k)}}\right)\gamma_j / \sum_j K_{\mathrm{loc}}\left(\frac{\delta(i,j)}{h_k}\right)$ *be the non-adaptive estimate using kernel* K_{loc} *and bandwidth* h_k*. We require that for a SPM* Γ *with* $\mathbf{E}\gamma_i \equiv 0$ *and* $\tau = \infty$ *the adaptive estimates* $\widehat{\gamma}_i^{(k)}$ *fulfill the propagation condition*

$$\boldsymbol{E}|\widehat{\gamma}^{(k)} - \check{\gamma}^{(h_k)}| < \kappa \boldsymbol{E}|\check{\gamma}^{(h_k)} - \gamma| \tag{16}$$

for some prespecified $\kappa > 0$.

We use a value corresponding to $\kappa = 0.05$.

3.2.2 Parameters of the multiscale test

The normalizing sequences $a_{n(i,h)}(\nu)$ and $b_{n(i,h)}(\nu)$ are specified using the approximation described in the end of Subsection 2.3. Under the hypothesis (for $\delta = 0$) $n(i,h) \approx n(h)$ while in case of deviations from H_0 we usually observe $n(i,h) \geq n(h)$ 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 ν, λ, k^* and n_{ROI}. The critical value τ mainly depends the significance level and degrees of freedom ν. In the simulated empirical distributions we observe a slight dependence on the chosen value of the smoothing parameter λ, the number of iteration steps k^*, or equivalently the maximum bandwidth, and the field of view n_{ROI}. Values for τ 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.

3.2.3 Kernels

We use

$$K_{\mathrm{loc}}(x) = \begin{cases} 1 & : \quad x < 1/2 \\ 2(1-x) & : \quad 1/2 \leq x < 1 \\ 0 & : \quad x \geq 1 \end{cases} \quad \text{and} \quad K_{\mathrm{st}}(x) = \begin{cases} 1 & : \quad x < 1/4 \\ 4/3(1-x) & : \quad 1/4 \leq x < 1 \\ 0 & : \quad x \geq 1 \end{cases}$$

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

4 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). All datasets were processed using the package *fmri* for R (R foundation for Statistical Computing) (Polzehl and Tabelow, 2007; R Development Core Team, 2006). Typical computation times for a complete analysis from linear modeling to the final activation maps range in the order of one minute depending on the dataset resolution.

4.1 Artificial data I

For the creation of an artificial dataset we use a numerical phantom simulating a situation where the fMRI response depends on two stimuli, left/right (L/R), and analyze their contrast. This mimics e.g. the situation in imaging ocular dominance columns in the visual cortex. The phantom consists of 10 slices with only the fifth slice containing activation. We designed stripes with different widths (1-4) at an in-plane matrix of 64×64, see Figure 2.

White and black stripes correspond to voxel where $R > L$ and $L > R$, respectively. Figure 3 illustrates the expected BOLD signals for the two regions. Voxel size in z-direction is held constant and set to twice the within slice voxel size of the medium resolution dataset. At each voxel a time series was created with 174 samples, stimulus onset times at the 7th, 31st, 55th, 79th, 103rd, 127th, 151st sample for the L stimulus and at the 19th, 43rd, 67th, 91st, 115th, 139th, 163rd sample for the R stimulus with a stimulus

duration of 6 samples and 1.5 s between two samples. We added autocorrelated Gaussian noise ($AR(1) = 0.3$) with a standard deviation of 10 for the independent noise variables. We aggregated the artificial data at lower resolutions of 32×32 and 16×16 matrix to mimic partial volume effects and SNR dependence. Here, clearly partial voluming removes the small stripes completely from the data. Even for the wider stripes, there are some voxel with no signal contrast in the data. Hence, we do not expect them to be detected by any method at the lower resolution.

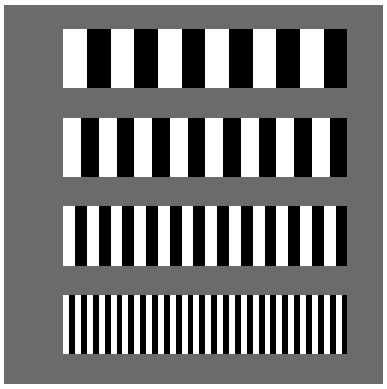

Figure 2: Phantom design.

In Figure 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

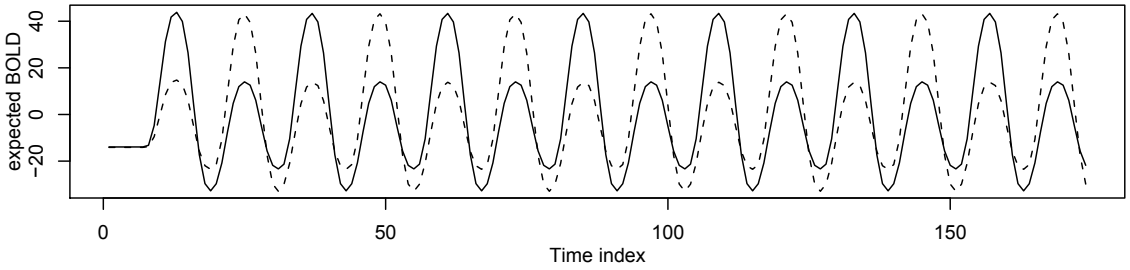

Figure 3: Expected BOLD signals ($L > R$ - solid, $R > L$ - dashed).

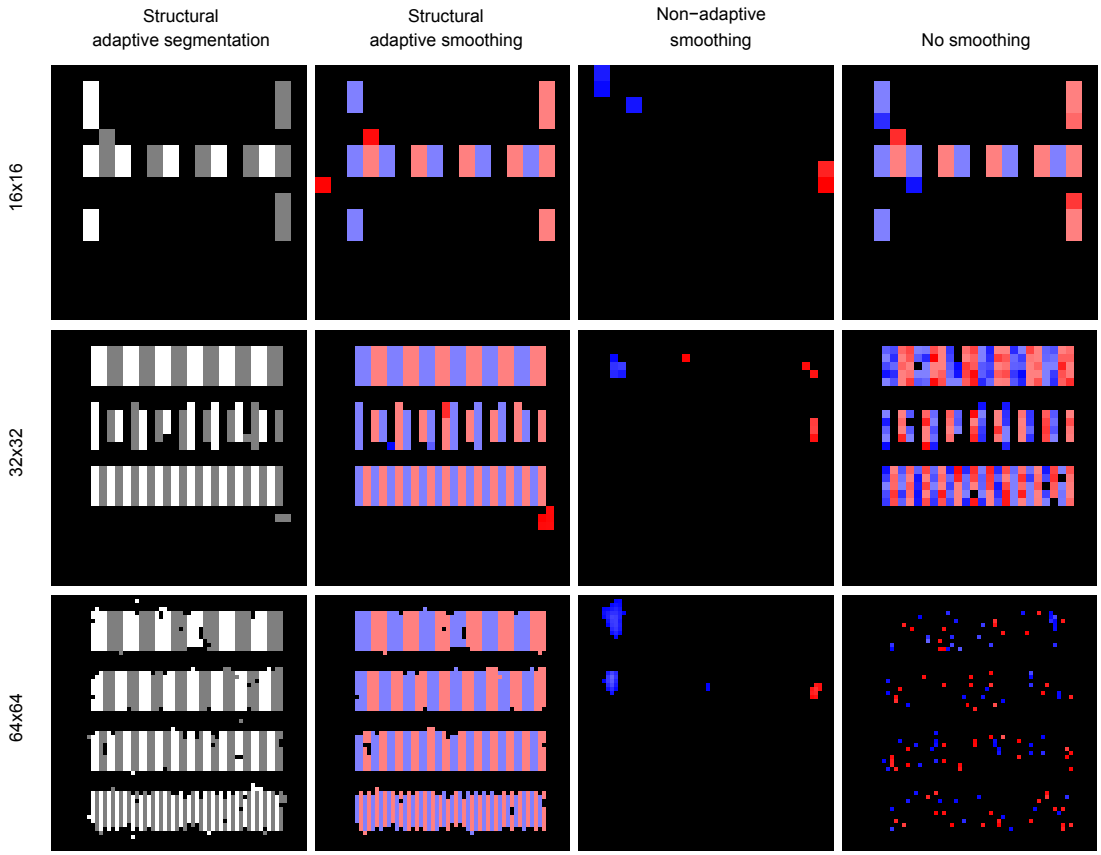

Figure 4: Signal detection for the artificial dataset I at different resolutions using different smoothing methods. The first column contains detections of the $R > L$ and $L > R$ contrast using the structural adaptive segmentation method described in this paper. The result is a segmentation of the brain into two segments where the hypotheses H_0 of no activation has been rejected and one set of voxel with no evidence to reject the hypothesis H_0. We use white, grey and black for the segments. The second column shows detected signals for both contrasts (blue for $R > L$ and red for $L > R$, the color has been arbitrarily chosen to provide a good visual contrast between both areas) using the structural adaptive smoothing method from Tabelow et al. (2006) with multiple test correction via RFT ($\alpha = 0.05$). Note that the color intensity corresponds to the p-values. The third column is the result for non-adaptive smoothing, while the right column provides voxelwise results (no smoothing). See online version for color.

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.

4.2 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 128×128, see Figure 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

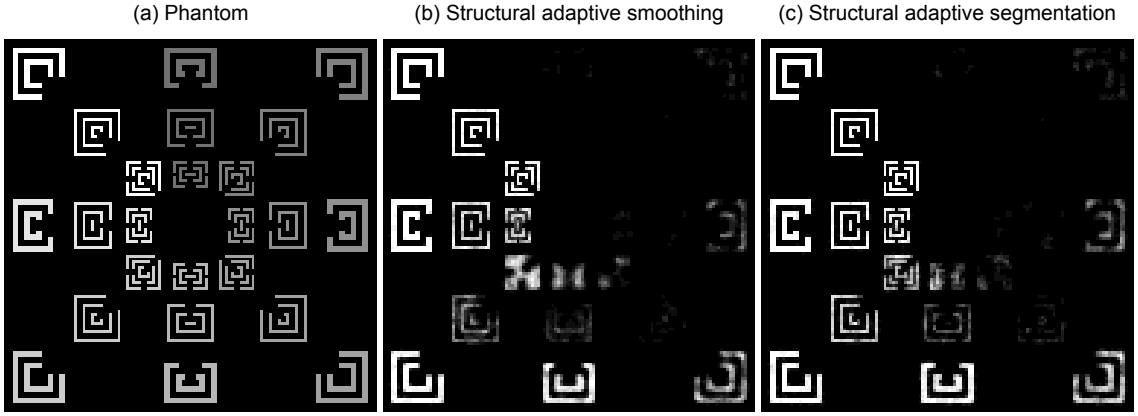

Figure 5: (a) Phantom used in simulations. Gray values indicate the size of the signal at the corresponding location. Size and structure of activated areas vary radially while the size of the signal increases clockwise by a factor of 1.25 from spoke to spoke. The phantom was first used in Tabelow et 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.

$k = 0, \ldots, 7$. 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

Table 1: 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 Figure 5. For signal intensity 0 false positives are reported while other columns contain the number of detected voxels for clockwise increasing SNR.

Signal intensity	0	1	1.25	1.25^2	1.25^2	1.25^3	1.25^4	1.25^5	1.25^6
Number of voxel	449760	4860	5580	4860	5580	4860	5580	4860	5580
detected voxel (PS)	1710	17	176	509	1530	2411	3889	4233	5499
detected voxel (AS)	989	27	206	721	1973	2819	4672	4650	5555

with a Gaussian kernel with FWHM-bandwidths $(1, 1, 0.5)$ times voxel size.

Figure 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 smoothing with RFT thresholding. For detailed results depending on signal strength see Table 1. Note that for both algorithms false positives almost exclusively occur adjacent to activated regions.

4.3 Experimental Data

We re-analyse an experimental fMRI dataset we used in one of our previous publications (Tabelow et al., 2009) to demonstrate the properties of the new structural adaptive segmentation algorithm proposed here and to compare its results with our 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 128×128, 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

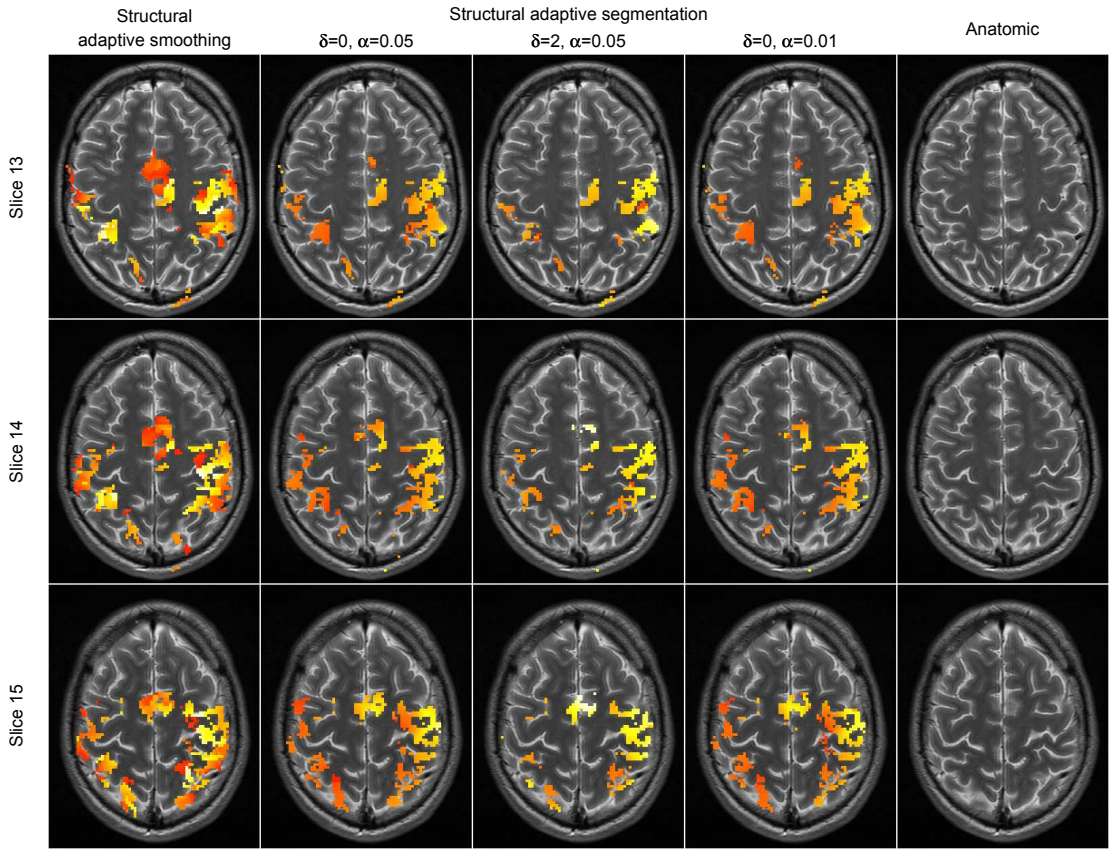

Figure 6: Signal detection in a sensorymotor experiment. The first column contains the signal detection using our structural adaptive smoothing algorithm (multiple test corrected $\alpha = 0.05$.) for comparison. Colors correspond to p-values. The second, third, and fourth column contain the results of our proposed structural adaptive segmentation algorithm using different values for the minimal signal δ, and the significance level α. Colors code the size of the estimated signal. The image orientation is R-L. See online version for color.

30 s rest. The first 4 scans before these block were discarded, yielding in 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 Figure 6. 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) and a more sensitive adaptation leading to slightly different areas.

The next three columns show the detected signal segments for different values of δ and significance level α. Apparently, the total size of activation areas reduces with α, while at $\alpha = 0.05$ the result is comparable to the PS result. The exclusion of low signals $\delta = 2$ is shown in column three. Within detected segments color codes the size of the estimated signal $\widehat{\gamma}_i^{(h_{k^*})}$. This provides, in contrast to p-values, *easily interpretable additional* information. The rightmost column shows T1 anatomic data.

5 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 conceptional advantages:

- It does not rely on an assumption of a local constant contrast γ.
- 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 $\widehat{\gamma}_i^{(h_{k^*})}$, which can be easily interpreted, see Figure 6.
- 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.

The last two characteristics lead to an improved sensitivity with respect to signal detection, see Table 1. The new algorithm can be especially helpful in situations with sharp

discontinuities and effects observed only at high spatial resolution.

Acknowledgments

This work is supported by the DFG Research Center MATHEON.

References

Adler, R. (2000). On excursion sets, tube formulae, and maxima of random fields, (special invited paper). *Ann. Appl. Probab.*, 10:1–74.

Chaudhuri, P. and Marron, J. S. (2000). Scale space view of curve estimation. *Ann. Statist.*, 28:408–428.

Cox, R. (1996). Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomed. Res.*, 29:162.

Dümbgen, L. and Spokoiny, V. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.*, 29:124–152.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math. Proc. Cambridge Philosophical Society*, 24:180–190.

Friston, K., Ashburner, J., and Kiebel, S., editors (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoiresur la distribution limite du terme maximum d'une serie aleatoire. *Ann. Math.*, 44:423–453.

Harrison, L., Penny, W., Daunizeau, J., and Friston, K. (2008). Diffusion-based spatial priors for functional magnetic resonance images. *NeuroImage*, 41:408–423.

Kriegeskorte, N. and Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fmri. *NeuroImage*, 38:649–662.

Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U.S.A.*, 103(10):3863–3868.

Lazar, N. A. (2008). *The Statistical Analysis of Functional MRI Data*. Statistics for Biology and Health. Springer.

Lu, Y., Jiang, T., and Zang, Y. (2003). Region growing method for the analysis of functional MRI data. *NeuroImage*, 20:455–465.

Poline, J. and Mazoyer, B. (1994a). Enhanced detection in brain activation maps using a multifiltering approach. *J. Cerebral Blood Flow and Metabolism*, 14:639–642.

Poline, J. B. and Mazoyer, B. M. (1994b). Analysis of individual brain activation maps using hierarchical description and multiscale detection. *IEEE Trans. Med. Imaging*.

Polzehl, J. and Spokoiny, V. (2006). Propagation-separation approach for local likelihood estimation. *Probab. Theory and Relat. Fields*, 135:335–362.

Polzehl, J. and Tabelow, K. (2007). fmri: A package for analyzing fmri data. *RNews*, 7(2):13–17.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag.

Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

Tabelow, K., Piëch, V., Polzehl, J., and Voss, H. U. (2009). High-resolution fmri: Overcoming the signal-to-noise problem. *J. Neuroscience Meth.*, 178:357–365.

Tabelow, K., Polzehl, J., Voss, H., and Spokoiny, V. (2006). Analyzing fmri experiments with structural adaptive smoothing procedures. *NeuroImage*, 33:55–62.

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. (2001). Testing for signals with unknown location and scale in a chi^2 random field, with an application to fmri. *Adv. Appl. Probab.*, 33:773–793.

Worsley, K. (2003). Detecting activation in fMRI data. *Statist. Meth. Med. Res.*, 12:401–418.