Functional and dynamic Magnetic Resonance Imaging using vector adaptive weights smoothing

Jörg Polzehl and Vladimir G. Spokoiny

Weierstrass Institute for Applied Analysis and Stochastics, Germany

Summary. We consider the problem of statistical inference for functional and dynamic Magnetic Resonance Imaging (MRI). A new approach is proposed which extends the adaptive weights smoothing (AWS) procedure from Polzehl and Spokoiny (2000) originally designed for image denoising. We demonstrate how the AWS method can be applied for time series of images, which typically occur in functional and dynamic MRI. It is shown how signal detection in functional MRI and analysis of dynamic MRI can benefit from spatially adaptive smoothing. The performance of the procedure is illustrated using real and simulated data.

1. Introduction

Polzehl and Spokoiny (2000) introduced a new locally adaptive method for two and three dimensional image processing which we refer to as *adaptive weights smoothing (AWS)*. This method is especially fruitful in situations when the underlying image contains large homogeneous regions with sharp edges. We show how this idea can be applied to more complicated problems. We especially consider experiments in which several images of the same object are obtained. This is e.g. the case if time series of images are recorded or if images are observed with different multispectral characteristics. We present a description of two classes of such problems arising in Magnetic Resonance Imaging (MRI): functional and dynamic MRI. A detailed description of two fMRI and dMRI examples is given. Section 2 discusses an extension of the original AWS procedure which is referred to as *vector AWS* and which allows for multi-image data. In Section 3 we present the results obtained by the vector AWS procedures for the examples given in Section 1. We also test the performance of vector AWS for simulated data and illustrate its properties by comparison with other methods.

1.1. Signal identification in functional MRI

Functional Magnetic Resonance Imaging (fMRI) is a relatively new non-invasive technique used to study human brain function. The experiments conducted in this context have usually the following design. A time series of two or three dimensional MR images is recorded while a patient is exposed to some activating signal. This may be a series of visual or acoustic stimuli depending on the problem studied. This stimulation causes neural activity in some regions of the brain. Identification of these 'activated' regions is an important step in the evaluation of the experiment. For an excellent introduction into fMRI see e.g. Lange (1996) or Lange and Zeger (1997).

The fMRI methodology is based on the following physical phenomenon. Neural activity is expected to cause both a deoxygenation of blood and an increase of blood flow in vessels within the activated regions of the cortex. This results in a change of the concentration of paramagnetic deoxyhaemoglobin in this regions. This change can be observed as an

increase of the MR signal providing the necessary contrast in the experiment. The effect is called 'Blood Oxygenation Level Dependent Contrast' (BOLD). See again Lange and Zeger (1997) for details. Due to the indirect method of measurement we can expect the observed signal to be related to the activation by some transformation involving smooth changes in shape (convolution with the Hemodynamic response function) and a delay in time, see e.g. Rajapakse et al. (1998) or Genovese (2000).

Typical experiments give series of 60 - 1000 images with a spatial resolution of 2–4 mm. Simple situations involve periodic activation with about 8-20 images in one period and several periods observed. More involved problems often require experiments based on an irregular design of the stimulus. Images are recorded at equidistant times, with typical time differences between images of some seconds. The data, for one slice of the brain, therefore have the following structure: for every voxel i with coordinate X_i, we observe the gray value $Y_{i,t}$ which can be represented as a sum of the induced signal $f_t(X_i)$ and the random error $\epsilon_{i,t}$, that is,

$$Y_{i,t} = f_t(X_i) + \epsilon_{i,t} \qquad t = 1, \ldots, T \tag{1}$$

with $\mathbf{E}\epsilon_{i,t} = 0$ and $\mathbf{Var}\ \epsilon_{i,t} = \sigma_i^2$, $\mathbf{E}$ and $\mathbf{Var}$ standing for expectation and variance, respectively. Often random errors are assumed to be independent and approximately Gaussian. Short time correlation of errors are to be expected in case of short time intervals. Error variance may be assumed homogeneous over time, but seems to be inhomogeneous in space due to the underlying anatomic structure, blood flow or properties of the MR device.

Example 1 (Periodic functional MRI). We will use an example based on a data set kindly provided by F. Kruggel from the Max-Plank-Institute of Cognitive Neuroscience at Leipzig, Germany to illustrate and discuss our approach. The data consist of time series of 912 Magnetic Resonance images of four slices of the brain recorded every two seconds. Data are given as integer gray values ranging from 0 to 19611. The fMRI series have been corrected for artifacts and body movement. Additionally high resolution MRI of the same slices are given.

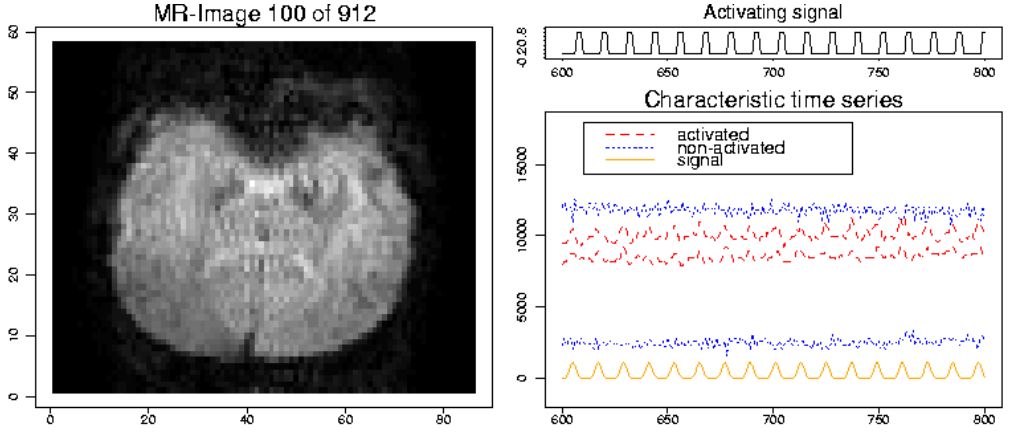

Fig. 1. Functional MRI data

The left plot in Figure 1 shows the 100th image of the time series from the third slice. Spatial resolution is about $2mm \times 4mm$. The patient is exposed to a periodic acoustic

signal, with no activation for the first 48 time intervals (upper right plot). The lower right of Figure 1 shows characteristic observed time series for voxel from both activated and non-activated regions. The data are generated using simple periodic experimental conditions.

1.2. Dynamic MRI

Dynamic MRI is used to study tissue perfusion within different organs of the body. A contrast agent (CA) is given to the patient and a temporal series of images is acquired using fast MR imaging techniques. The images are recorded over a suitable time interval starting before the injection of the CA and covering the expected impact of the CA. Each image is obtained using the same MR parameters and focusing on the same object (location). We therefore have at each voxel a time series of MR intensities reflecting the effect of the CA over time at the given location. See e.g. Sebastiani (1997) or Sebastiani et. al. (1996) for a more detailed presentation. Due to the short acquisition times of some hundreds of milliseconds spatial resolution is low and the observed noise level is high.

EXAMPLE 2 (DYNAMIC MRI). Sebastiani et. al. (1996) use a data set consisting of 30 images of a slice of the brain of a rat recorded in intervals of 0.6 seconds. A part of the brain is known to be damaged. The effect of the contrast agent can be observed starting with the 7th image.

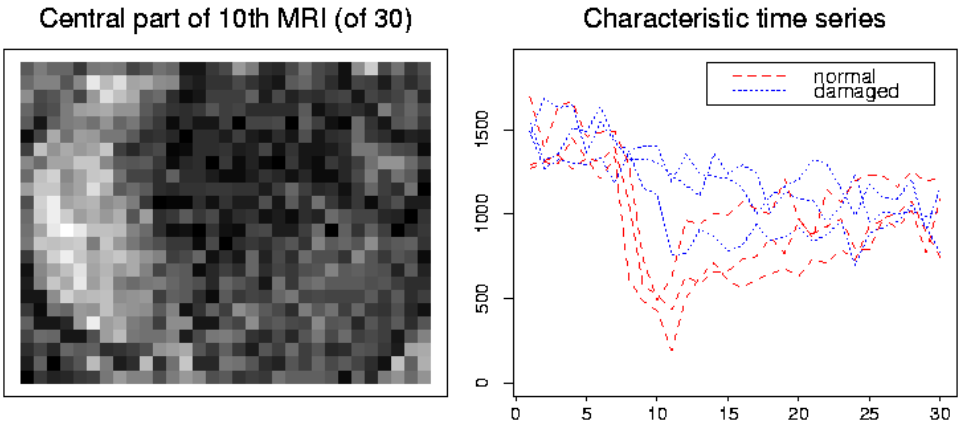

Fig. 2. Central part of first image from a dMRI series of the brain of a rat (upper plot) and characteristic time series from both normal and pathologic areas (lower plot)

Figure 2 displays the central part of the 10th image (left) together with characteristic time series (right) from both pathologic (dotted) areas and normal (dashed) tissue. In regions with full functionality the observed time series are expected to show a sharp decrease in MR intensity from image 7 to 10 and to nearly return to the initial values at the end of the observation period, i.e. possessing a distinguished minimum. In pathologic areas the minimum, or peak, is either inferior or not observed. This means that both peak delay, i.e. the location of the minimum, and peak intensity, i.e. the difference between the size of the minimum and the end value of the time series, can be used to discriminate between pathologic and normal tissue.

1.3. Traditional approaches

Traditional approaches to analyze functional neuroimaging data involve three steps. The first consists of several preprocessing tasks, including e.g. correction for body movement, artifact detection and removal. In a second step a Statistical Parametric Map (SPM) is constructed. This simply means that at each voxel a regression of the observed time series with an expected signal is performed. See e.g. Lange and Zeger (1997) or Genovese (2000) on how to model the expected signal. Specifying time delay and shape of the hemodynamic response this leads to a linear model, corresponding to the experimental design, and therefore provides a value of an F-statistic, indicating the significance of the observed signal. In a third step the theory of maxima of random fields, see e.g. Adler (1981), is used to determine significant signals. The simplest procedure of this sort is to test the hypothesis of no signal at each voxel independently. This corresponds to a multiple comparison problem and requires high thresholds which leads to a low sensitivity in detecting an activated region.

Better methods for detecting the presence of a signal in a specified region use the fact that the spatial extent of the regions of interest is often significantly larger than the spatial resolution. Therefore spatial smoothing, using e.g. a Gaussian kernel, of the images can be used to improve the sensitivity of the analysis. Adaptive approaches, e.g. the scale-space approach of Siegmund and Worsley (1995), use the maximum of a smoothed random field and different scales to detect the presence of an activation. Related techniques are based on the expected Euler characteristic of excursion sets of random fields, see e.g. Worsley (1994) and Worsley (1995). These methods focus on the decision whether there is a signal of a predefined shape in a certain region or not, and they are very well designed to answer just this question. Another method using Markov Random Fields (MRF) to model spatial connectivity is proposed e.g. in Descombes and Kruggel (1999).

Standard techniques in dynamic MRI again focus on voxel by voxel analysis of the time series. This includes computation of characteristics of the time series based on parametric regression, see e.g. Rosen et.al. (1990), or the analysis of temporal correlation to an expected time series, see e.g. Rogowska and Wolf (1992). Sebastiani et. al. (1996) propose to use nonparametric smoothing in time to estimate some parameters of the time series like location of the minimum or extend of the minimum which then can be used for voxel characterization. Here we meet the same problem as in functional MRI: multiple testing approaches require high thresholds which leads to a poor quality of decisions. Sebastiani et. al. (1996) suggested to use spatial filtering of the single images in a preprocessing step for an additional noise reduction.

2. Adaptive spatio-temporal smoothing

In contrast to the traditional approaches the method proposed in this paper will not concentrate on detecting a signal of a given form, but on generating a spatial structure corresponding to similarities and significant differences between time series in neighboring locations. This will provide information on both identification of the shape of signals and of the shape and spatial extend of regions the signals can be allocated in.

The procedure which will be discussed further originates from Polzehl and Spokoiny (2000). There a new locally adaptive smoothing procedure was offered which is especially designed for estimation of a regression function that allows for a reasonable piecewise constant approximation. Our approach to functional and dynamic MRI uses a multivariate generalization of this method.

At the first step, as usual in functional and dynamic MRI analysis, some data preprocessing (transformation) is carried out, see e.g. Holmes and Friston (1997), Chapter 3.4, Lange and Zeger (1997) or Genovese (2000). This transformation will include removal of anatomic structure and slowly varying trends in time, and an aggregation in time to reduce the dimensionality of the problem.

After transformation the regions of interest are supposed to have some spatial extension, showing a similar behavior of the transformed data, e.g. similar shape, time delay and size of the observed signal, within the region. The adaptive weights approach introduced in Polzehl and Spokoiny (2000) can now be used to employ these similarities in order to identify the regions of activation.

2.1. Data aggregation

Adaptive weights smoothing as many other smoothing methods heavily relies to comparing estimated signals in neighboring locations. It is well known, that tests based on differences of vectors in high dimensional spaces, i.e. of two time series in neighboring voxels, are highly inefficient, see e.g. Fan and Lin (1998). If additional information is available, like smoothness of the curves, periodicity or shape of the expected signal, this information can be used to increase the power of the test. One way to achieve this is to base the test on aggregated data like wavelet or Fourier coefficients for each curve. Which aggregation method or which coefficients of an orthogonal series expansion to use mainly depends on the properties of the curves. Heuristically the method has to be chosen to reduce the dimensionality of the problem while preserving the main information about the characteristics of the curves.

Let $g_\ell = g_\ell(t)\,,\ \ell = 1, \ldots, L\,,$ be a set of functions satisfying

$$\frac{1}{T}\sum_{t=1}^{T} g_\ell(t)\, g_{\ell'}(t) = \delta_{\ell\ell'}. \tag{2}$$

A specific example is produced by a set of Fourier or wavelet basis functions. The corresponding coefficients for every curve $f_t(X_i)$ with $t = 1, \ldots, T$ are defined by

$$\beta_{i,\ell} = \frac{1}{T}\sum_{t=1}^{T} f_t(X_i)\, g_\ell(t).$$

Using the observations $Y_{i,t}\,,$ following the model (1), these values can be estimated by the empirical coefficients

$$B_{i,\ell} = \frac{1}{T}\sum_{t=1}^{T} Y_{i,t}\, g_\ell(t).$$

A reasonable test for the hypothesis that two curves (vectors) $f(X_i)$ and $f(X_j)$ coincide, can be based on these empirical coefficients. Assuming independent and time homogeneous noise $\epsilon_{i,t}$ in (1), one obviously has

$$\mathbf{Var}\, B_{i,\ell} = \frac{1}{T^2}\sum_{t=1}^{T} \sigma_i^2 |g_\ell(t)|^2 = \frac{1}{T}\sigma_i^2$$

so that, if an estimate $\widehat{\sigma}_i$ of σ_i is available, this value can be estimated by $\widehat{\sigma}_i^2/T$. In case of known correlation in the time series an appropriate adjustment in (3) should be used.

Now a statistic, which is usually referred to as *Neyman smooth test*, Neyman (1937), can be defined as

$$T_{i,j} = T \sum_{\ell=1}^{L} \frac{|B_{i,\ell} - B_{j,\ell}|^2}{\widehat{\sigma}_i^2} \tag{3}$$

allowing to compare the time series in voxel X_i and X_j.

The choice of the set of basis functions g_ℓ is very important for the quality of the resulting estimates. For some specific examples prior information is available which helps to select this set in a reasonable way, see e.g. the examples below.

Alternatively parametric models for the hemodynamic response, see e.g. Lange and Zeger (1997) or Genovese (2000), can be used for the aggregation step.

2.1.1. Functional MRI (Example 1)

Having in mind the periodic structure of the fMRI experiment we aggregate the data using a two term Fourier approximation with basis functions

$$g_1(t) = \sqrt{2}\sin(2\pi t/p) \quad \text{and} \quad g_2(t) = \sqrt{2}\cos(2\pi t/p)$$

leading to empirical Fourier coefficients $B_{i,l}, l = 1, 2$ for each voxel i. This reduces the time series information to amplitude and time delay of a sine function with appropriate periodicity, i.e. similarity of induced signals is measured in terms of these two characteristics. This also eliminates information corresponding to anatomic structure and slowly varying trends.

2.1.2. Dynamic MRI Example

For the dynamic MRI example in a preliminary transformation step we remove the effect of the anatomic structure subtracting a baseline estimate obtained from the first 6 images. These images do not reflect any effect from the contrast agent. This provides the transformed time series

$$Z_{i,t} = Y_{i,t} - 1/6 \sum_{s=1}^{6} Y_{i,s}, \qquad t = 7, \ldots, 30.$$

An appropriate set of orthonormal basis functions $g_l(t)$ can be chosen recalling the expected form of the time series and keeping in mind the different behavior within pathologic areas. We use the following elementary set of basic functions,

$$g_1(t) = \tfrac{1}{\sqrt{3}} I_{t\in(6,9]}, \qquad g_2(t) = \tfrac{1}{\sqrt{3}} I_{t\in(9,12]}, \qquad g_3(t) = \tfrac{1}{\sqrt{6}} I_{t\in(12,18]},$$

$$g_4(t) = \tfrac{1}{\sqrt{6}} I_{t\in(18,24]}, \quad g_5(t) = \tfrac{1}{\sqrt{6}} I_{t\in(24,30]},$$

with I_A denoting the indicator function on set A. This simply means aggregation is done averaging observations from certain time intervals, giving empirical coefficients

$$B_{i,\ell} = \sum_{t=7}^{30} Z_{i,t} g_\ell(t).$$

The first two intervals are chosen smaller to reflect the expected higher variation of the curves at times 7–12. The minimum of the curves is expected within the second interval for normal tissue, while for pathologic areas the peak will be in one of the last intervals. Peak intensity can be easily estimated from the aggregated data, see Section 3.2 for more details.

2.2. Variance estimation

The special design of functional and dynamic MRI allows voxelwise variance estimating at every voxel X_i from the corresponding time series $\{Y_{i,t}\}_{t\in 1,\ldots,T}$. Below we describe one possible approach for our fMRI and dMRI examples.

2.2.1. Functional MRI Example

In case of the periodic fMRI experiment it is natural to base the estimate of the error variance in voxel X_i on residuals with respect to a nonparametric estimate of the periodic induced signal in X_i:

$$\widehat{\sigma}_i^2 = C * IQR^2(Y_{i,t} - S_{i,t}), \qquad t = 1, \ldots, T,$$

with $S_{i,t}$ being a periodic signal with period length p and

$$S_{i,t} = \frac{p}{T} \sum_{m=1}^{T/p} Y_{i,t+(m-1)p}, \qquad t = 1, \ldots, p. \tag{4}$$

The use of the inter-quartile range (IQR), with constant $C = T/(T-p)/1.35^2$, gives a robust variance estimate.

In general the method of variance estimation has to take account of possible temporal correlation. We estimated a mean first order correlation of .06 for this example. The bias of the variance caused by this correlation is relatively small. We therefore did not correct for this bias, although such a correction could be incorporated into the estimate.

2.2.2. Dynamic MRI Example

A variance estimate can be based on residuals of the $Z_{i,\cdot}$ with respect to a local linear regression estimate $\widehat{Z}_{i,\cdot}$ (bandwidth $h = 5$), see e.g. Fan and Gijbels (1996), of the time series in each voxel. It is expected that the contrast agent results in a sharp decrease of the observed signal in the beginning and then in an increase in a smooth way, see Figure 2. This means that residuals up to $t = 13$ may contain a non negligible bias and should not be used to estimate the variance. We therefore use only the residuals $Z_{i,t} - \widehat{Z}_{i,t}$ for $t = 14, \ldots, 30$ for variance estimation which leads to the estimate

$$\widehat{\sigma}_i^2 = C * \mathbf{Var}\,(Z_{i,14:30} - \widehat{Z}_{i,14:30}),$$

with a constant C adjusting for the loss in degrees of freedom by local linear regression. (In our example $C = 1.28$.)

Voxelwise variance estimates show no significant variance inhomogeneity. We therefore assume a homogeneous variance in this example. The resulting variance estimate is

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \widehat{\sigma}_i^2.$$

Tests for time correlation based on the residuals $Z_{i,t} - \widehat{Z}_{i,t}$ show no significant time correlation in this example.

2.3. Vector adaptive weights smoothing

Adaptive spatial smoothing will be performed using the multivariate regression model

$$B_i = \beta(X_i) + \varepsilon_i\,, \quad i = 1, \ldots, n, \quad X_i \in I\!R^d, \quad B_i \in I\!R^L$$

which is the result of transformation (preprocessing) of the original data Y_i. Here $X_1, \ldots, X_n$ are design points which are assumed to be equispaced in the unit cube $[0,1]^d$. At each point X_i we observe the $I\!R^L$-valued transformed regression function $\beta(X_i)$ with some additive error $\varepsilon_i \in I\!R^L$. We suppose the errors ε_i to be independent zero mean random vectors with unknown distribution which may depend on location and

$$\mathbf{E}\varepsilon_i = 0 \in I\!R^L, \qquad \mathbf{Var}\ \varepsilon_i = \mathrm{diag}\{s_i\,,\ \ell = 1, \ldots, L\}. \tag{5}$$

In our examples the data $B_{i,\ell}\,,\ i = 1, \ldots, n$, for fixed ℓ, correspond to the image of the ℓ-th empirical coefficients and $s_i = \sigma_i^2/T$. The variance s_i of the errors $\varepsilon_{i,\ell}$ is unknown but its estimate $\widehat{s}_i$ can be computed, as described in Section 2.2, as $\widehat{s}_i = \widehat{\sigma}_i^2/T$.

This is exactly the situation one meets in image analysis, i.e. adaptive smoothing methods for image processing can be applied here. An overview on such methods is fare beyond the scope of this paper, we refer to Polzehl and Spokoiny (2000) for some discussion.

Our approach is based on the structural assumption that the regression function β is piecewise constant. This means that the unit cube $[0,1]^d$ can be split into disjoint regions $A_1, \ldots, A_M$ and

$$\beta(x) = \sum_{m=1}^{M} b_m I_{\{x \in A_m\}} \tag{6}$$

where $b_1, \ldots, b_M \in I\!R^L$ are some vectors and I stands for the indicator function. The regions A_m, the vectors b_m and even the total number of regions M are unknown. This assumption is valid for an arbitrary series of L images, since each region A_m may consist of one point. However, if the structural assumption is approximately valid with an M essentially smaller than n, i.e. with sufficiently large regions A_m, it can be used to get improved estimates of $\beta(x)$.

Such kind of modeling is reasonable if, e.g., the target of the statistical analysis is a vector (curve) characterization, i.e. classification into 'activated / non-activated' regions in the fMRI example or characterization as 'pathologic / normal' in the dMRI example.

The problem of estimating the function β of the form (6) can be treated as follows: to recover the values $b_1, \ldots, b_M$ and to decide for each point X_i in which region A_m it is. To explain the idea of the method, we imagine for a moment that the regions $A_1, \ldots, A_M$ are known and only the vectors b_m are to be estimated. This leads to obvious estimates

$$\widehat{b}_m = \frac{1}{N_{A_m}} \sum_{X_i \in A_m} B_i$$

where N_{A_m} denotes the number of design points in $A_m\,,\ m = 1, \ldots, M$. We can simply set $\widehat{\beta}(X_i)$ equal to the mean $\widehat{b}_m$ of Y_j's over the region A_m containing X_i. Therefore, given a partition $A_1, \ldots, A_M$, we can easily estimate the underlying function β.

Next we consider the inverse situation when the partition $A_1, \ldots, A_M$ is unknown but we are given a pilot estimate $\widehat{\beta}^{(0)}$ of the L-variate regression function β. It is natural to use this estimate to recover for every point X_i the corresponding region A_m. Namely, for each pair of points X_i and X_j, we may decide on the basis of the estimates $\widehat{\beta}^{(0)}(X_i)$ and $\widehat{\beta}^{(0)}(X_j)$ whether they are in the same region. If the estimate $\widehat{\beta}^{(0)}(X_i)$ is significantly different from the estimate $\widehat{\beta}^{(0)}(X_j)$ these two points are almost definitely in different regions. Significance can be measured by performing a test for the hypothesis $\beta(X_i) = \beta(X_j)$ based on a test statistic $T_{i,j}$, see (3). Let λ be an appropriate quantile of the distribution of $T_{i,j}$. For each design point X_i, the set $\widehat{A}(X_i)$ with

$$\widehat{A}(X_i) = \{X_j : T_{i,j} \le \lambda\}$$

estimates the region A_m containing X_i. Using these estimated regions, we may define the new estimate $\widehat{\beta}^{(1)}$ by

$$\widehat{\beta}^{(1)}(X_i) = \frac{\sum\limits_{X_j \in \widehat{A}(X_i)} B_j}{N_{\widehat{A}(X_i)}} = \frac{\sum\limits_j w_{i,j}^{(1)} B_j}{\sum\limits_j w_{i,j}^{(1)}}$$

with

$$w_{i,j}^{(1)} = I\left(T_{i,j} \le \lambda\right) \tag{7}$$

and N_S being the cardinality of the set S. Then we can repeat this calculation using $\widehat{\beta}^{(1)}$ in place of $\widehat{\beta}^{(0)}$ and so on.

Our adaptive procedure mostly realizes this idea with two modifications. First of all, at each iteration k, we restrict the estimated region $\widehat{A}(X_i)$ to some local neighborhood $U^{(k)}(X_i)$ of the point X_i such that the size of $U^{(k)}(X_i)$ grows with k. This means that we calculate the initial pilot estimate $\widehat{\beta}^{(0)}(X_i)$ by averaging observations from a small neighborhood $U^{(0)}(X_i)$ of the point X_i (in many situations it can be the observation B_i itself). Then we recalculate this estimate by averaging over a larger neighborhood $U^{(1)}(X_i)$ but now using only data points where there are no essential differences between values of the initial estimates. We continue in this way, increasing each time the considered neighborhood $U^{(k)}(X_i)$, that is, for each $k \ge 1$,

$$\widehat{\beta}^{(k)}(X_i) = \frac{\sum\limits_{X_j \in U^{(k)}(X_i)} w_{i,j}^{(k)} B_j}{\sum\limits_{X_j \in U^{(k)}(X_i)} w_{i,j}^{(k)}} \tag{8}$$

where the weights $w_{i,j}^{(k)}$ are computed by comparison of the estimates $\widehat{\beta}^{(k-1)}(X_i)$ and $\widehat{\beta}^{(k-1)}(X_j)$. Secondly we use continuous weights $w_{i,j}^{(k)}$ instead of zero-one weights in (7).

For a much more detailed discussion of the basic idea of AWS we refer to Polzehl and Spokoiny (2000). We now present the formal description of the method.

Using a series of neighborhood $U^{(k)}(X_i)$, $k = 1, k^*$ with $U^{(k-1)}(X_i) \subset U^{(k)}(X_i)$ and a threshold value λ we get the following *Vector Adaptive Weights Smoothing* procedure:

Initialization: For each voxel X_i, calculate initial estimates $\widehat{\beta}_{i,\ell}$ of $\beta_{i,\ell}$ and its variance as

$$\widehat{\beta}_{i,\ell}^{(0)} = \frac{1}{N^{(0)}(X_i)} \sum_{X_j \in U^{(0)}(X_i)} B_{j,\ell} \qquad \text{and} \qquad \widehat{v}_i^{(0)} = \frac{1}{|N^{(0)}(X_i)|^2} \sum_{X_j \in U^{(0)}(X_i)} \widehat{s}_j$$

and set $k = 1$.

Adaptation: Compute weights $w_{i,j}^{(k)}$ as

$$w_{i,j}^{(k)} = K\left(\frac{1}{\lambda} \sum_{\ell=1}^{L} \frac{\left(\widehat{\beta}_{i,\ell}^{(k-1)} - \widehat{\beta}_{j,\ell}^{(k-1)}\right)^2}{\widehat{v}_i^{(k-1)}} \right)$$

for all points X_j in $U^{(k)}(X_i)$, with K denoting a monotone decreasing function, e.g. $K(x) = \exp(-x)$. Compute new estimates of $\beta_{i,\ell}$ and **Var** $\beta_{i,\ell}$ as

$$\widehat{\beta}_{i,\ell}^{(k)} = \frac{\sum\limits_{X_j \in U^{(k)}(X_i)} w_{i,j}^{(k)} B_{j,\ell}}{\sum\limits_{X_j \in U^{(k)}(X_i)} w_{i,j}^{(k)}} \qquad \text{and} \qquad \widehat{v}_i^{(k)} = \frac{\sum\limits_{X_j \in U^{(k)}(X_i)} \left| w_{i,j}^{(k)} \right|^2 \widehat{s}_j}{\left(\sum\limits_{X_j \in U^{(k)}(X_i)} w_{i,j}^{(k)} \right)^2}$$

for all X_i.

Control: Under the condition that $\widehat{\beta}_{i,\ell}^{(k)}$ is unbiased we can compute a L-dimensional confidence region that contains β_i with probability $1 - \alpha$ as

$$CI_\ell^{(k)} = \left(\widehat{\beta}_{i,\ell}^{(k)} - \eta\sqrt{\widehat{v}_{i,\ell}^{(k)}}, \widehat{\beta}_{i,\ell}^{(k)} + \eta\sqrt{\widehat{v}_{i,\ell}^{(k)}} \right) = \left(CL_\ell^{(k)}, CU_\ell^{(k)} \right)$$

where η^2 is an appropriate quantile of the distribution of the maximum of L χ_1^2 random variables.

The new estimate $\widehat{\beta}_i^{(k)}$ is accepted if, for each m with $m < k$ and every $\ell \le L$, the ℓ-th estimate $\widehat{\beta}_{i,\ell}^{(k)}$ belongs to the interval $(CL_\ell^{(m)}, CU_\ell^{(m)})$, that is,

$$|\widehat{\beta}_{i,\ell}^{(k)} - \widehat{\beta}_{i,\ell}^{(m)}| \le \eta\sqrt{\widehat{v}_{i,\ell}^{(m)}}$$

keeping the previous estimates otherwise.

Stopping: Stop if $k = k^*$ or if $\widehat{\beta}_i^{(k)} = \widehat{\beta}_i^{(k-1)}$ for all i, otherwise increase k by 1 and continue with the adaptation step.

The parameters of the procedure are chosen using the same considerations as for the original AWS. The set of neighborhoods $U^{(k)}(X_i)$ should contain an exponentially (in k) growing number of voxel, see Polzehl and Spokoiny (2000) or Section 3.3 below for a proposal. The parameter λ controlling type I error of a test for the hypothesis of two voxel to belong to the same region, can be chosen as a quantile of a χ_L^2 distribution. The tests have to be performed at a very high significance level, our experience suggesting to use a 0.995-quantile or an even larger value. A suitable value for η^2 is the 0.999-quantile of the distribution of a maximum of L χ_1^2-distributed random variables. For a comprehensive discussion about the choice of the parameters λ and η see Polzehl and Spokoiny (2000).

3. Application to the examples and simulations results

We first demonstrate how the AWS procedure described in Section 2 can be applied to the examples from functional and dynamic MRI:

3.1. Signal detection in fMRI

Let B_i be the transformed data for the fMRI example from Section 1.1. It is natural to base the signal detection and identification on the results of the previously described AWS procedure, namely, on the estimates $\widehat{\beta}_{i,\ell} = \widehat{\beta}_{i,\ell}^{(k^*)}$ obtained at the last iteration of the AWS.

To exclude insignificant signals we can compare the estimates $\widehat{\beta}_{i,\ell}$ with the corresponding standard deviation $\widehat{v}_{i,\ell}^{1/2}$ which leads to the statistic

$$Q_i = \sum_{l=1}^{L} \frac{\widehat{\beta}_{i,\ell}^2}{\widehat{v}_{i,\ell}}. \tag{9}$$

The AWS procedure often provides a significant noise reduction so that even very small differences from a zero mean are detected. These may be due to systematic effects caused by measuring and preprocessing, making it sometimes reasonable to show only regions where the activated signal is sufficiently large. For that purpose, we calculate the values

$$R_i = \sum_{l=1}^{L} \widehat{\beta}_{i,\ell}^2. \tag{10}$$

A signal will be detected in voxel i if Q_i exceeds a given threshold τ and if R_i is larger than some constant C.

The parameter τ has the same meaning as the parameter λ in the AWS procedure. We recommend to chose the value of τ as a proper quantile of the χ^2_L distribution.

To recover the shape of the induced signal in each voxel where we detect an activation we perform a spatial smoothing of the mean periodic signal $S_{i,t}$ from (4) using the adaptive weights $w_{i,j} = w_{i,j}^{(k^*)}$ coming from the last step of the AWS procedure.

$$\widehat{f}_t(X_i) = \sum_j w_{i,j} S_{i,t} \Big/ \sum_j w_{i,j}, \quad t = 1, \ldots, p \tag{11}$$

Signal detection is based on the statistics Q_i and R_i from (9) and (10). Figure 3 illustrates the output of our detection procedure. The central plot shows the intensity $\sqrt{R_i}$ of all signals detected using the threshold $\tau = \chi^2_{2,0.9995} \approx 15.2$ for Q_i and meeting $R_i > C = 1250$. This choice excludes signals with amplitude smaller than 50. Our tools allow to select a signal interactively from the intensity map. The left plot shows all detected signals, possessing a correlation larger than 0.6 with a signal selected in the central plot, mapped into the high resolution anatomical image. The right plot contains the graphs of $\widehat{f}_t(X_i) - \bar{Y}_i$ for all voxel marked in the left plot, $\bar{Y}_i = T^{-1} \sum_{t=1}^{T} Y_{i,t}$ being the mean of the observations $Y_{i,t}$ at X_i.

An effect often observed in fMRI experiments with many observed periods of activation is a change in the shape and size of the induced (BOLD) signal over time. This can be explained e.g. by learning or by getting accustomed to the stimulus. Our approach easily allows to incorporate this by selecting an appropriate aggregation of the time series. In

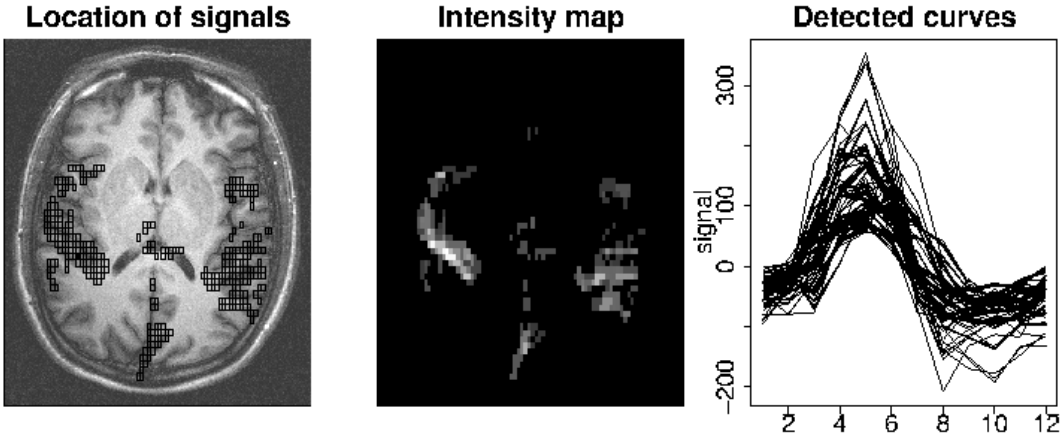

Fig. 3. Signal detection with AWS. Central plot: intensity map of detected signals. Left plot: Position of selected signals in the corresponding high resolution image. Right plot: mean periodic signals in selected voxel.

order to test for time inhomogeneity in our example we divide the time series into three segments of length 288 and compute empirical Fourier coefficients for each part as

$$B_{i,\ell} = 1/288 \sum_{t=49+(l-1)*288}^{48+\ell*288} Y_{i,t} g_1(t) \quad \text{and} \quad B_{i,\ell+3} = 1/288 \sum_{t=49+(l-1)*288}^{48+\ell*288} Y_{i,t} g_2(t),$$

for $\ell = 1, 2, 3$.

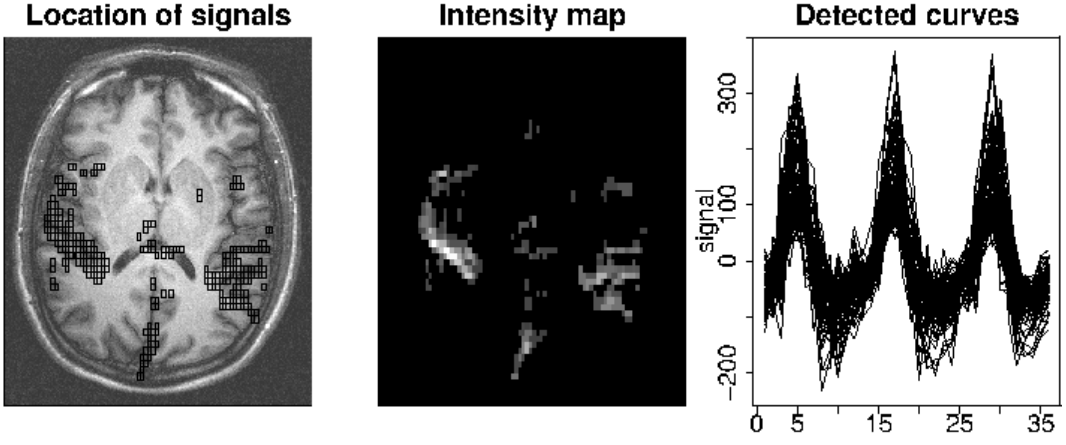

Fig. 4. Signal detection with AWS for possibly inhomogeneous situations. Central plot: intensity map of detected signals. Left plot: Position of selected signals in the corresponding high resolution image. Right plot: mean periodic signals (from three segments) in selected voxel.

We now carry out the adaptive weights smoothing based on six Fourier coefficients. Signal detection based on Q_i and R_i with threshold $\tau = \chi^2_{6,.9995} \approx 24$ and $C = 3750$ gives the results shown in Figure 4.

The curves displayed in the right plot of Figure 4 are now generated stringing the mean periodic curves $\widehat{f}_t(X_i) - \bar{Y}_i$ from the three segments together. The findings are essentially the same as in the previous setting indicating no time heterogeneity.

3.2. Analysis of dynamic MRI

We start with data aggregation and variance estimation as suggested in Sections 2.1.2 and 2.2. Next the vector AWS procedure from Section 2.3 is carried out. The final estimates of the target coefficients $\beta_{i,\ell}$ are denoted by $\widehat{\beta}_{i,\ell}$.

For the example, a natural characteristic of every curve $f_t(X_i)$ at the location X_i are the peak delay D_i and the peak intensity Δ_i defined as

$$D_i = \arg\min_t f_t(X_i), \qquad \Delta_i = \min_t f_t(X_i) - f_T(X_i).$$

They can be approximated by the introduced coefficients $\beta_{i,\ell}$:

$$D_i \approx \arg\min_\ell \frac{\beta_{i,\ell}}{c_\ell} \quad \text{and} \quad \Delta_i \approx \min_\ell \frac{\beta_{i,\ell}}{c_\ell} - \frac{\beta_{i,5}}{c_5},$$

with $c_\ell = \sqrt{3}$ for $\ell = 1, 2$, and $c_\ell = \sqrt{6}$ for $\ell = 3, 4$ and 5, respectively.

This naturally leads to the following statistics for curve characterization:

$$\widehat{D}_i = \arg\min_\ell \frac{\widehat{\beta}_{i,\ell}}{c_\ell} \quad \text{and} \quad \widehat{\Delta}_i = \min_l \frac{\widehat{\beta}_{i,\ell}}{c_\ell} - \frac{\widehat{\beta}_{i,5}}{c_5}.$$

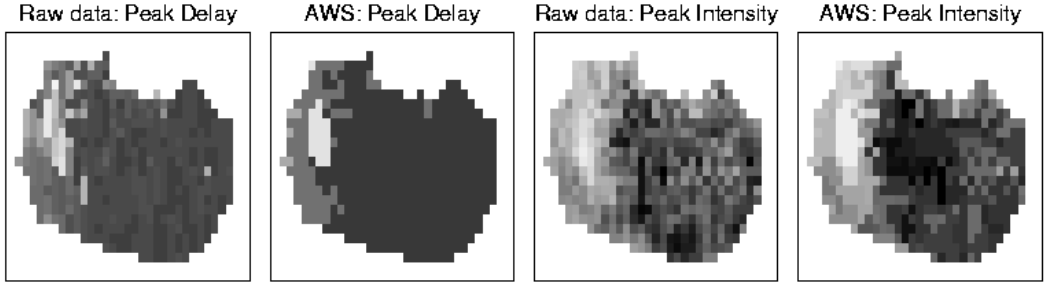

Fig. 5. Peak delay (upper row) and peak intensity (lower row) maps obtained from the original time series (left column) and the AWS estimates (right column).

Figure 5 illustrates the results based either on the $\widehat{D}_i$'s or on the $\widehat{\Delta}_i$'s. The left two images show a peak delay map calculated from the original data and from the AWS estimates $\widehat{\beta}_{i,\ell}$, respectively. The two right images give the corresponding peak intensity maps. Voxel outside the region of interest are masked (white).

Note that adaptive spatial smoothing using AWS clearly improves the interpretability of the peak delay map, allowing for a discrimination between pathologic and normal tissue. The peak intensity maps also show the effect of spatial smoothing by AWS. See also Sebastiani et. al. (1996) for results using other approaches.

3.3. Simulation

We conducted a small simulation study to illustrate the performance of our approach in idealized situations. Presenting the results we focus on the following features of the procedure: probability to estimate a signal in regions with no activation and probability of a wrong classification near the boundary of the activated regions. The latter can be treated as the quality of estimating the shape of the region. We also compared our procedure with three others to illustrate the strengths of our method and the differences to traditional approaches.

The simulation setup is as follows. We generated a time series of $T = 64$ images, with each image containing 50×50 voxel. We arranged periodic signals in 9 regions of varying shape and size. The signal is of the form

$$f_t(X_i) = c_i \left(0.45 \sin\Big(\frac{2\pi t}{p}\Big) - 0.6 \cos\Big(\frac{2\pi t}{p}\Big) \right), \qquad t = 1, \ldots, 64$$

with period $p = 8$ and c_i being 1, 2/3 and 4/9 for the different regions and equal to zero for voxel outside these regions. Figure 6 illustrates the form of the signals (one period) as

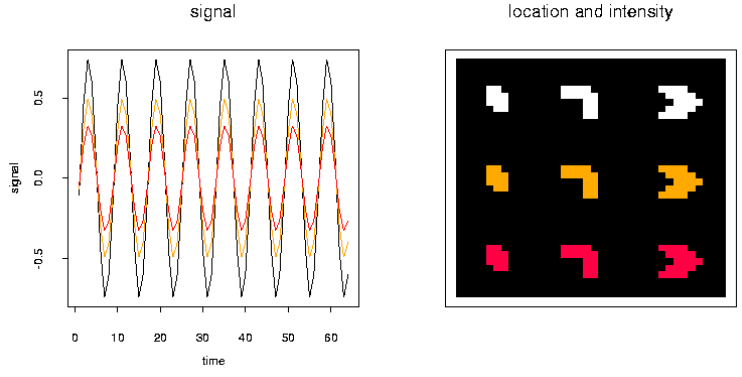

Fig. 6. Simulation experiment, periodic signals (left plot) and their location (right image)

well as their location, with the magnitude of signals decreasing from top to bottom and the size of regions increasing from left to right in the displayed image. We then added standard Gaussian white noise in each voxel.

We apply our vector AWS algorithm with the following specifications. We use two Fourier coefficients, i.e. basis functions $g_1(t) = \sqrt{2}\sin(2\pi t/p)$ and $g_2(t) = \sqrt{2}\cos(2\pi t/p)$, which is appropriate in this situation. AWS is performed with $\lambda = 10.6$ corresponding to a 0.995-quantile of χ^2_2, $\eta = 4$ and the sequence of neighborhoods U_k specified as circles with radii $\{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.4, 5, 6, 7, 8\}$ ($k^* = 13$). Error variances in each voxel are estimated from the data as described in Section 2.2.

We conducted 200 simulation experiments to estimate the pointwise probability of signal detection using our AWS approach. A signal is detected solely on the base of the statistics Q_i from (9) with threshold $\tau = \chi^2_{2,.9995} \approx 15.2$. This provides a mean voxelwise detection error of 0.0068 for voxel with a distance of more than 2δ from activated regions, with δ being the distance between neighboring voxel centers, i.e.

$$\mathbf{P}\,(\text{signal detected in } X_i | d(X_i) > 2\delta) \leq 0.0068, \tag{12}$$

Table 1. Mean probability of signal detection in activated regions

Method	left			central			right		
	top	central	bottom	top	central	bottom	top	central	bottom
AWS	.982	.887	.577	.983	.890	.617	.973	.875	.743
Raw data	.850	.404	.127	.846	.402	.117	.855	.381	.121
smoothed (0.5δ)	.992	.752	.288	.992	.751	.280	.988	.743	.312
smoothed (1δ)	1.000	.983	.734 1.000	.977	.727	.999	.958	.802	

Table 2. Mean type I error of signal detection in a neighborhood (distance of $\leq 2\delta$) of the activated regions.

Method	left			central			right		
	top	central	bottom	top	central	bottom	top	central	bottom
AWS	.003	.021	.051	.004	.027	.091	.014	.037	.074
Raw data	.008	.011	.007	.008	.007	.006	.004	.006	.007
smoothed (0.5δ)	.016	.014	.007	.015	.010	.007	.015	.011	.007
smoothed (1δ)	.453	.185	.058	.524	.256	.086	.485	.249	.097

where $d(X_i)$ denotes the distance from X_i to the closest activated region.

For a comparison we give the results for three alternative approaches. The first alternative is based on the raw data, i.e. a signal is detected in voxel i if $\sum_{\ell=1}^{L} B_{i,\ell}^2/s_i$ exceeds a threshold C_1, with C_1 again selected to provide (12). The second and third alternative involve a preliminary spatial smoothing of the $B_{i,\ell}/\sqrt{s_i}$ using a bivariate Gaussian kernel with bandwidths $h = 0.5\delta$ and $h = 1\delta$, respectively. Signal detection is performed as before with thresholds C_2 and C_3 chosen in analogy to C_1.

All methods we considered have the same type I error outside the regions of activation and we now look at the performance of each method near the boundary of the activated regions. We present separately the results describing the probability of detecting an existing activation (which can be naturally interpreted as the power of the test on the presence of signal), see Table 1, and the probability of an incorrect detection of a signal in a point without activation (type I error), see Table 2.

Table 3 provides mean probabilities of an incorrect decision, i.e. either detecting a signal at a point with no activation or non-detecting a signal from the activated region. The results are given for 9 subregions, containing 12×12 voxel each, of the whole image. The results clearly show a better overall behaviour of AWS. Signal detection without spatial smoothing suffers from high threshold values, resulting in problems to detect weak signals. Non-adaptive spatial smoothing improves signal detection within the activated regions but reduces regional specificity.

Table 3. Mean probability of incorrect signal detection (for subregions of 12x12 voxel).

Method	left			central			right			over
	top	central	bottom	top	central	bottom	top	central	bottom	all
AWS	.003	.015	.047	.004	.021	.066	.010	.028	.054	.028
Raw data	.017	.048	.067	.022	.069	.097	.024	.092	.128	.063
smoothed (0.5δ)	.008	.025	.056	.008	.034	.081	.008	.043	.102	.041
smoothed (1δ)	.062	.030	.031	.093	.051	.047	.103	.061	.051	.059

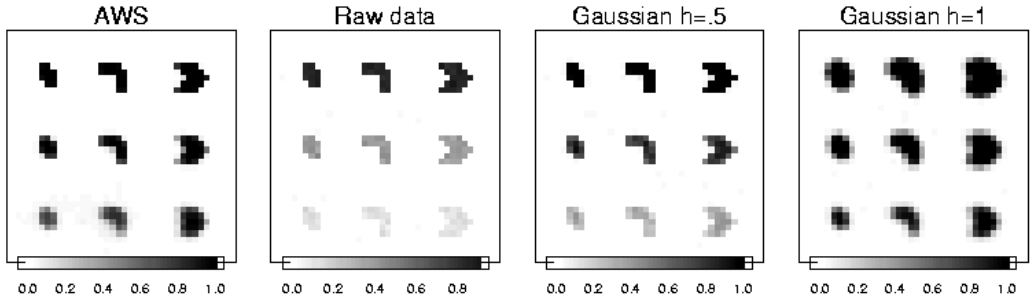

Fig. 7. Pointwise signal detection probabilities obtained from 200 simulations

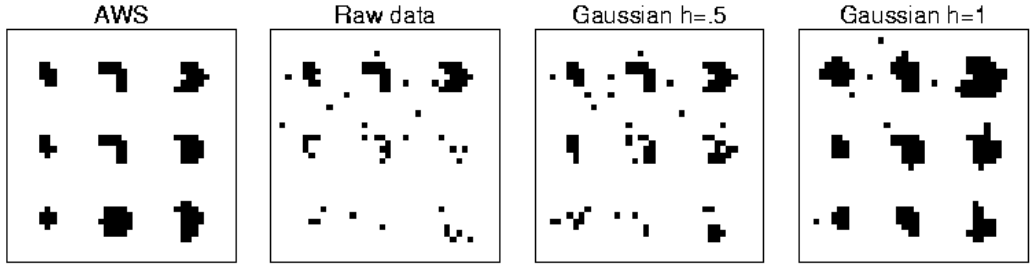

Fig. 8. Detected signals for a typical data set from the simulation study using AWS (upper left), raw data (upper right) and spatial smoothing (lower row) using comparable thresholds.

Figure 7 illustrates the pointwise probabilities of signal detection obtained from 200 simulation runs. Figure 8 provides the detection results for a typical time series of images, i.e. with a medium number of ill classified voxel for all approaches, from our simulations.

4. Conclusions

The present paper offers a data driven approach to some statistical problems in functional and dynamic MRI like signal detection, identification of activated regions and characterization of curves. Large homogeneous regions with similar curves are in favor of the procedure. We show how, for functional and dynamic MRI, the original problem can be transformed to meet such an assumption. The simulated results and the examples demonstrate the capabilities of the proposed procedure allowing both for detecting small signals and for precise estimation of its location. All these issues are in agreement with theoretical properties of the AWS procedure introduced in Polzehl and Spokoiny (2000). Theoretical properties of the method especially for applications to multiple data sets will be subject of further investigations. Application is not restricted to MRI. We expect the method to be useful whenever series of spatially registered images occur, one example being multispectral satellite imaging.

Acknowledgements:

We would like to thank Fritjof Kruggel, Fred Godtliebsen and Giovanni Sebastiani for their

introduction into MRI and for useful discussions. We are also grateful to Klaus Hahn, Gerhard Winkler and two anonymous referees for helpful remarks and discussion.

References

Adler, R. J. (1981). *The Geometry of Random Fields*, Wiley: New York.

Descombes, X. and Kruggel, F. (1999). A Markov Pixon Information Approach for Low-Level Image Description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**, 482–494.

Fan, J. and Gijbels (1996). *Local Polynomial Modelling and its Applications*, Chapman & Hall: London.

Fan, J. and Lin, S.-K. (1998). Test of Significance When Data Are Curves. *J. Amer. Statist. Ass.*, **93**, 1007–1021.

Genovese, C.R. (2000). A Bayesian Time-Course Model for Functional Magnetic Resonance Imaging Data (with discussion). *J. Amer. Statist. Ass.* **95** 691–719.

Holmes,A. and Friston, K.J. (1997). Statistical models and experimental design. In SPM-course, Institute of Neurology, Wellcome Dept of Cognitive Neurology, University College London, `http://www.fil.ion.ucl.ac.uk/spm/course/notes.html`

Lange, N. (1996). Tutorial in biostatistics: Statistical approaches to human brain mapping by functional magnetic resonance imaging. *Statistics in Medicine*, **15**, 389–428.

Lange, N. and Zeger, S. L. (1997). Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with discussion). *Applied Statistics*, **46**, 1-29.

Neyman, J. (1937). "Smooth test" for goodness of fit. *Scand. Aktuarietidskr.* **20** 149–199.

Polzehl, J. and Spokoiny, V. G. (2000). Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society*, Ser. B, **62**, 2, 335–354.

Rajapakse, J., Kruggel, F., Maisog, J. and von Cramon, D. (1998). Modeling Hemodynamic Response for Analysis of Functional MRI Time-Series,*Human Brain Mapping*,**6**, 283–300.

Rogowska, J. and Wolf, G.L. (1992). Temporal correlation images from sequential MR scans. *J. Comp. Assist. Tom.* , **16**, 784–788.

Rosen, B.R., Belliveau, J.W., Vevea, J.M. and Brady, T.J. (1990). Perfusion imaging with NMR contrast agents. *Magn. Reson. Med.*, **6**, 249–265.

Sebastiani, G. (1997). Mathematical and statistical methods for medical magnetic resonance imaging. PhD Thesis, The Norwegian University of Science and Technology, Trondheim, Norway.

Sebastiani, G., Godtliebsen, F., Jones, A. R., Haraldseth, O., Müller, T. B. and Rinck, P. A. (1996). Analysis of Dynamic Magnetic Resonance Images. *IEEE Transactions on Medical Imaging*, **15**, 268–277.

Siegmund, D. O. and Worsley, K. J. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Annals of Statistics*, **23**, no. 2, 608–639.

Spokoiny, V. (1996). Adaptive Hypothesis Testing using Wavelets. *Ann. Statist.*, **24**, no. 6, 2477–2498.

Worsley, K.J. (1994). Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Advances in Applied Probability*,**27**, 943–959.

Worsley, K.J. (1995). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Advances in Applied Probability*, **27**, 943–959.