Vector adaptive weights smoothing with applications to MRI

Jörg Polzehl* and Vladimir G. Spokoiny†

September 23, 1999

Preprint No. 519

Berlin 1999

1991 Mathematics Subject Classification. 62G07, 62P15, 92C55.

Keywords. adaptive smoothing; spatial adaptation; functional MRI; signal detection; dynamic MRI.

*Weierstrass Institute for Applied Analysis and Statistics, Mohrenstr. 39, D-10117 Berlin, Germany. E-mail: polzehl@wias-berlin.de

†Universität Würzburg, Sanderring 2, D-97070 Würzburg

Abstract

We consider the problem of adaptive spatial smoothing for a time series of images. This type of data typically occurs in functional and dynamic Magnet Resonance Imaging (MRI). We propose a new method based on spatial smoothing with adaptively chosen weights. We show how this procedure can be used for efficient image estimation and classification in functional and dynamic MRI experiments. The performance of the procedure is illustrated by applications to simulated and real data.

1 Introduction

Polzehl and Spokoiny (1998) 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 now generalize this idea to more complicated problems. We especially consider experiments in which an image of the same object is recorded several times. This is e.g. the case if time series of images are recorded or if images are observed with different multispectral characteristics. Below in this section we present a detailed description of two classes of such problems arising in the Magnet Resonance Imaging (MRI): functional and dynamic MRI. Other examples are delivered by multispectral satellite imaging or for multichannel MRI. Section 2 discusses one extension of the original AWS procedure which is referred to as the *vector AWS* and which allows for multi-image data. In Section 3 we show how this vector AWS procedure can be adapted for signal detection and signal identification in functional MRI (fMRI). We first test the performance of the vector AWS on simulated data and compare it with some other methods. Then we present an analysis of a real fMRI dataset. Section 4 gives an application vector AWS to classification in dynamic MRI.

1.1 Signal identification in functional MRI

Functional Magnet 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 stimulations depending on the problem studied. This stimulation causes neural activity in some regions of the brain. Identification of these 'activated' regions is the main interest of the experiment. For an excellent introduction into fMRI see Lange (1996) or Turner and Friston (1997).

The fMRI methodology is based on the following physical phenomenon. Neural activity is expected to increase the blood flow in blood vessels in the activated regions of the cortex. This causes a relative decrease in 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 Turner and Friston (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).

Typical experiments give series of 60 - 1000 images with a spatial resolution of 2–4 mm. Simple designs involve periodic activation with about 8-20 images in one period and several periods observed. 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 $\varepsilon_{i,t}$, that is,

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

with $\boldsymbol{E}\varepsilon_{i,t} = 0$ and $\boldsymbol{E}\varepsilon_{i,t}^2 = \sigma_i^2$. 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 variances can be assumed homogeneous over time, but seem to be inhomogeneous in space due to the underlying anatomic structure, blood flow or properties of the MR device.

1.2 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 and artifact detection and removal. In a second step a Statistical Parametric Map (SMP) is constructed. This simply means that at each voxel an appropriate general linear model, corresponding to the experimental design, is applied providing a value of an F-statistic, indicating the significance of the observed signal, see e.g. Holmes and Friston (1997) for details. In a third step the theory of Gaussian random fields is used to determine significant signals, see e.g. Poline et.al. (1997). 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 the activated regions. More subtle techniques use the fact that the spatial extend of the regions of interest is significantly larger than the spatial resolution. This is used by either spatially smoothing, using e.g. a Gaussian kernel, of the images or by testing for the spatial extend of an activation. See again Poline et.al. (1997) or Worsley et.al. (1992) for the first and Poline and Mazoyer (1993) for the second approach. Both methods while increasing the power of signal detection suffer from loss of information about the exact location of the activated region. Another method using Markov Random Fields (MRF) to model spatial connectivity is proposed e.g. in Descombes et.al. (1998).

Taking into account that both the high sensitivity to the activated signal and the precise location of the activated regions are important for further analysis, the application of the AWS procedure (which is specially designed for such situations) seems to be reasonable.

1.3 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 with 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.

Standard techniques 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 characteristics of the series like location of the minimum or extend of the minimum which then can be used for voxel classification. Here we meet the same problem as in functional MRI: multiple testing approaches require high thresholds which leads to a poor quality of image classification. Sebastiani et. al. (1996) suggested to use spatial filtering of the single images in a preprocessing step for an additional noise reduction. The AWS method described in the next section exploits the same idea allowing simultaneously for preserving the shape of homogeneous regions within the underlying image.

2 Vector AWS smoothing

Polzehl and Spokoiny (1998) offered a new locally adaptive smoothing procedure which is especially designed for estimation of a regression function allowing a reasonable approximation by piecewise constant models. We now generalize this approach to more complex situations and problems.

2.1 Basic idea

In what follows we consider the model which can be described as

$$Y_i = f(X_i) + \varepsilon_i\,, \qquad i = 1, \ldots, n, \qquad X_i \in I\!R^d, \qquad Y_i \in I\!R^T. \tag{1}$$

Here $X_1, \ldots, X_n$ are design points which are usually assumed to be equispaced in the unit cube $[0,1]^d$. At each point X_i we observe the $I\!R^T$-valued regression function $f(X_i)$ with some additive error $\varepsilon_i \in I\!R^T$. We suppose the errors ε_i to be independent zero mean random vectors with unknown distribution which may depend on location:

$$\boldsymbol{E}\varepsilon_i = 0 \in I\!R^T, \qquad \mathbf{Var}\,\varepsilon_i = \operatorname{diag}\{\sigma^2_{i,t}\,,\, t = 1, \ldots, T\},$$

with $\boldsymbol{E}$ and $\mathbf{Var}$ denoting expectation and variance, respectively.

In the applications we have in mind the data $Y_{i,t}$, $i = 1, \ldots, n$, for fixed t, correspond to the recorded image at time t and $Y_{i,t}$, $t = 1, \ldots, T$ is the series of observed values at X_i during the observation time.

Our basic assumption is that the regression function f is supposed piecewise constant, possessing the same structure in each component. This means that the unit cube $[0,1]^d$ can be split into disjoint regions $A_1, \ldots, A_M$ and

$$f(x) = \sum_{m=1}^{M} a_m \mathbf{1}(x \in A_m) \tag{2}$$

where $a_1, \ldots, a_M \in I\!R^T$ are some vectors and $\mathbf{1}$ stands for the indicator function. Obviously the image vectors $f(x)$ are constant within each region A_m. The regions A_m, the vectors a_m and even the total number of regions M are unknown. Clearly this assumption is valid for an arbitrary series of T images, since each region A_m may consist of one point. We however assume that M is essentially smaller than n that means that the regions A_m are sufficiently large. Such kind of modelling is reasonable if, e.g., the target of the statistical analysis is a vector (curve) classification. Typical examples are: 'activated / non-activated' in functional MRI applications or 'pathologic / normal' in dynamic MRI.

We first recall the basic idea of the adaptive weights smoothing (AWS) introduced in Polzehl and Spokoiny (1998). The problem of estimating the function f of the form (2) can be treated as follows: to recover the values $a_1, \ldots, a_M$ 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 a_m are to be estimated. This leads to obvious estimates

$$\widehat{a}_m = \frac{1}{N_{A_m}} \sum_{X_i \in A_m} Y_i$$

where N_{A_m} denotes the number of design points in A_m, $m = 1, \ldots, M$. Then we simply set $\widehat{f}(X_i)$ equal to the mean $\widehat{a}_m$ 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 f.

Next we consider the inverse situation when the partition $A_1, \ldots, A_M$ is unknown but we are given a pilot estimate $\widehat{f}^{(0)}$ of the p-variate regression function f. 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 $\widehat{f}^{(0)}(X_i)$ and $\widehat{f}^{(0)}(X_j)$ whether they are in the same region. If the estimate $\widehat{f}^{(0)}(X_i)$ is significantly different from the estimate $\widehat{f}^{(0)}(X_j)$ these two points are almost definitely in different regions. Significance can be measured by performing a test for the hypothesis that $\widehat{f}^{(0)}(X_i) = \widehat{f}^{(0)}(X_j)$ based on some test statistic $T_{i,j}$. Let λ be an appropriate quantile of the distribution of T. 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 $\widehat{f}^{(1)}$ by

$$\widehat{f}^{(1)}(X_i) = \frac{\sum_{X_j \in \widehat{A}(X_i)} Y_j}{N_{\widehat{A}(X_i)}} = \frac{\sum_j w_{i,j}^{(1)} Y_j}{\sum_j w_{i,j}^{(1)}}$$

with

$$w_{i,j}^{(1)} = \mathbf{1}\left(T_{i,j} \le \lambda\right) \tag{3}$$

and N_S being the cardinality of the set S. Then we can repeat this calculation using $\widehat{f}^{(1)}$ in place of $\widehat{f}^{(0)}$ 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 $\widehat{f}^{(0)}(X_i)$ by averaging observations from a small neighborhood $U^{(0)}(X_i)$ of the point X_i (in many situations it can be the observation Y_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 $k \geq 1$,

$$\widehat{f}^{(k)}(X_i) = \frac{\sum\limits_{X_j \in U^{(k)}(X_i)} w_{i,j}^{(k)} Y_j}{\sum\limits_{X_j \in U^{(k)}(X_i)} w_{i,j}^{(k)}} \tag{4}$$

where the weights $w_{i,j}^{(k)}$ are computed by comparison of the preceding estimates $\widehat{f}^{(k-1)}(X_i)$ and $\widehat{f}^{(k-1)}(X_j)$. Secondly we use continuous weights $w_{i,j}^{(k)}$ instead of zero-one weights in (3).

2.2 Assessing significant differences of vectors

The essential element of the AWS procedure is a testing step: for two different points X_i and X_j we decide whether they belong to the same homogeneous region on the base of information obtained in previous iterations. Before going into details, it is worth noting that at each iteration k this procedure is carried over many times (for every pair X_i, X_j from $U^{(k)}(X_i)$) which makes the problem different from the classical testing problem with a single testing. The first type error for our testing procedure means that some point X_j from $U^{(k)}(X_i)$ with $f(X_j) = f(X_i)$ is classified as not belonging to $\widehat{A}^{(k)}(X_i)$. This might lead to random segmenting small subregions within a large homogeneous region. To avoid such kind of "undersmoothing", the joint error of the first kind for all tests together should be sufficiently small. This particularly means that each single test should be rather conservative. On the other side, an application of a too conservative testing procedure will include points X_j from other regions into the estimated region $\widehat{A}^{(k)}(X_i)$, which may produce some bias in the final estimate especially near the boundary between two neighbor regions.

The original AWS procedure from Polzehl and Spokoiny (1998) was designed for the case when we observe a single image, that is, $T = 1$ and $\widehat{f}^{(k-1)}(X_i)$ is an estimate of the value $f(X_i)$ obtained after $k-1$ iterations. With the use of the estimated variance

$\widehat{v}_i^{(k-1)}$ of $\widehat{f}^{(k-1)}(X_i)$, the test statistics $T_{i,j}^{(k)}$ can be taken in the form

$$T_{i,j}^{(k)} = \frac{|\widehat{f}^{(k-1)}(X_i) - \widehat{f}^{(k-1)}(X_j)|^2}{\widehat{v}_i^{(k-1)}}.$$

Simulated results and practical applications show an excellent performance of the procedure with the critical value λ about 9 which corresponds to the well know "rule of 3 sigma".

A natural generalization of this method corresponding to Hotelling's T^2, can be based on the L_2 distance of vectors, e.g.

$$T_{i,j}^{(k)} = \sum_{t=1}^{T} \frac{|\widehat{f}_t^{(k-1)}(X_i) - \widehat{f}_t^{(k-1)}(X_j)|^2}{\mathbf{Var}\, \widehat{f}_t^{(k-1)}(X_i)}.$$

Although this approach seems appropriate it has several drawbacks. It turns out being highly inefficient in high dimensional situations (large T), see e.g. Fan and Lin (1998). If additional information is available like smoothness of the curves or periodicity 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 an orthonormal set of functions satisfying

$$\frac{1}{T} \sum_{t=1}^{T} g_\ell(t)\, g_{\ell'}(t) = \delta_{\ell\ell'}. \tag{5}$$

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

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

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

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

A more efficient 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 $\varepsilon_{i,t}$ in (1), one obviously has

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

so that, if an estimate $\widehat{\sigma}_i$ of σ_i is available, this value can be estimated by $\widehat{\sigma}_i^2/T$. Now a reasonable test statistic, which is usually referred to as *Neyman smooth test*, Neyman (1937), can be defined as

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

Our approach combines this idea with the adaptive spatial smoothing approach of the AWS procedure. Namely, at every k-th step of the procedure we compare two curves (vectors) $f(X_i)$ and $f(X_j)$ on the basis of the previous step estimates $\widehat{f}^{(k-1)}(X_i)$ and $\widehat{f}^{(k-1)}(X_j)$. For this we again calculate the corresponding estimates of the coefficients $\beta_{i,\ell}$

$$\widehat{\beta}_{i,\ell}^{(k-1)} = \frac{1}{T} \sum_{t=1}^{n} \widehat{f}_t^{(k-1)}(X_i)\, g_\ell(t)$$

and apply the test statistics of the form

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

where $\widehat{v}_i^{(k-1)}$ is the estimate for $\mathbf{Var}\, \widehat{\beta}_{i,\ell}^{(k-1)}$ (which does not depend on ℓ).

This approach allows for the following simple interpretation: the original data (set of curves) are transformed into the set of the corresponding empirical coefficients $B_{i,\ell}$ and further the AWS procedure is carried over using these coefficients in place of the original data.

The choice of the set of basis functions g_ℓ is very important for the quality of the procedures. For some specific examples some prior information is available which helps to select this set in a reasonable way, see e.g. Section 3 below. For other situations, a data-driven methods can be recommended, see e.g. Ledwina (1994), Fan (1996), Spokoiny (1996), Ledwina and Kallenberg (1997) or Hart (1997). The idea is to consider simultaneously a collection of different basis sets $\{g_\ell\}$. For each of them, one can construct the corresponding test statistics and the resulting test rejects the hypothesis of similarity of two curves (vectors) if one of them does. Spokoiny (1996) showed that this adaptive test should be applied with slightly increased critical value (by a $\log\log n$-factor) then for each single test.

We will continue this discussion using several examples from dynamic and functional MRI in the following sections.

3 Applications to fMRI

The problem of signal detection or signal recovery in fMRI can be successfully attacked using the basic ideas of AWS. The goal of statistical analysis in fMRI is typically to identify the regions of activation and to obtain a description of the induced signal. Specific features of the problem which make it really complicated are:

- low image resolution,
- low intensity of the activated signal compared to the noise level,
- spatial noise heterogeneity,
- very indirect association between the, often periodic, activation and the induced (observed) signal, resulting in a change of the shape and a delay in time,
- presence of an underlying anatomic structure and of a slowly changing trend component which is typically spatially irregular.

The first three features motivate an application of methods based on spatial smoothing which allow to reduce the noise level while preserving the shape of the activated regions, so that the AWS procedure seems to be reasonable here. In addition, it should be designed sensitive to the activated (periodic) signal and insensitive to the slowly varying trend.

3.1 Elimination of the anatomic structure and slow time-dependent trends

In periodic fMRI often a special preprocessing step is used to remove slow time-dependent trends in each voxel. This roughly means, in each voxel separately, to subtract a non-parametric estimate of the time-dependent trend from each time series. As a result the anatomic structure is also eliminated. Kruggel et. al. (1998) propose to remove an moving average estimate of the baseline with a window of length $3p+1$ from the time series in each voxel (here p is the periodicity of the activating signal). The window length of $3p+1$ is chosen to avoid loss of the periodic structure. As a result one gets data

$$Z_{i,t} = Y_{i,t} - \frac{1}{3p+1} \sum_{k=-3p/2}^{3p/2} Y_{i,t+k} . \tag{6}$$

This approach seems to be very useful for practical applications although it introduces some time correlation in the data Z_i. Another possible approach avoids this step by selecting an appropriate Fourier or wavelet transform which automatically produces trend elimination.

We discuss shortly the last proposal for the case of applying the Fourier transform with the set of basic functions $g_{2\ell}(t) = \sqrt{2}\cos(\frac{2\pi\ell t}{p})$ and $g_{2\ell-1}(t) = \sqrt{2}\sin(\frac{2\pi\ell t}{p})$ for $l = 1,\dots,L/2$. Namely we utilize a well known fact that high order Fourier coefficients are almost insensitive to the slowly varying trend component of the signal.

To simplify our notation, we consider an imaginary situation when a function $f(t)$ is observed with noise and it can be represented as a sum of a periodic function $h(t)$ of periodicity p and an additional slowly varying component $s(t)$, $1,\dots,T$, where p is small compared to T. It is useful to introduce a continuous parameter $u = t/T$ with $0 \le u \le 1$. Then the mentioned property of the function $s(\cdot)$ can be reformulated as follows: the function $s_1(u) = s(uT)$ is smooth, e.g., in the sense that its second derivative is bounded: $|s_1''| \le M$. This implies that the corresponding Fourier coefficients $c_k = \int_0^1 s_1(u)\cos(2\pi ku)\mathrm{d}u$ decrease at rate k^{-2}, that is, $c_k \le Mk^{-2}$. This, in turn, yields

$$\frac{1}{T}\sum_{t=1}^{T} s(t)\cos(2\pi t/p) \approx \int_0^1 s_1(u)\cos(2\pi uT/p)\mathrm{d}u \le M(p/T)^2.$$

For sine basic functions, the coefficients $d_k = \int_0^1 s_1(u)\sin(2\pi ku)\mathrm{d}u$ decrease slowly, at rate k^{-1}, because of the boundary effect, unless the trend function s_1 satisfies the boundary condition $s_1(0) = s_1(1)$. The similar effect arises if the function s_1 is not smooth but only piecewise smooth with a finite number of jumps. In general, one may ensure that for all considered basic functions g_ℓ, the corresponding Fourier coefficients are at most of order p/T. Since the standard deviation of each empirical Fourier coefficient is of order $T^{-1/2}$, this leads to the following conclusion (see, e.g. Spokoiny, 1999): if the periodicity of the activated signal is small compared with the time of observation, that is, if $pT^{-1/2}$ is small , then the impact of the slowly varying trend component in the corresponding Fourier coefficients is negligible.

Of course, other basic functions, e.g. wavelets, can be used in place of the Fourier basis. The only requirement is that all basic functions are nearly orthogonal to a slowly varying time-dependent component.

After the transformation of the original data $Y_{i,t}$ into the set of corresponding empirical Fourier coefficients $B_{i,\ell}$ is done, under ideal conditions, we now have homogeneity in regions without an induced signal. For activated regions we expect to see some local homogeneity of the induced signal in terms of the corresponding Fourier coefficients $\beta_{i,\ell} = T^{-1}\sum_{t=1}^{T} f_t(X_i)g_\ell(t)$. This can be used to identify both regions without activation as well as the signal in activated regions by adaptive spatial smoothing.

3.2 fMRI procedure

We describe the steps of the analysis for a fMRI experiment with a periodic activation signal. We assume here that the induced (BOLD) signal is of same periodicity, lets say of p time intervals, but may be shifted and of different shape. We do not assume any prior information about delay and shape of the induced signal, although such kind of information could be naturally incorporated into our approach.

The whole analysis can be split into several preliminary steps, the AWS procedure and signal identification on the base of the AWS results.

3.2.1 Preliminary steps.

The presence of multiple data $Y_{i,t}$ at each voxel X_i with time homogeneous noise $\varepsilon_{i,t}$ allows for a voxelwise estimation of noise variance.

The noise variance at each voxel can easily be estimated from the corresponding time series $\{Y_{i,t}\}_{t\in 1,\dots,T}$ as

$$\widehat{\sigma}_i^2 = \frac{1}{6(T-2)} \sum_{t=2}^{T-1} [2Y_{i,t} - Y_{i,t-1} - Y_{i,t+1}]^2. \tag{7}$$

Using a robustified variance estimate or spatial smoothing of the variance estimates may be useful.

The next step transforms the data into the set of empirical Fourier coefficients.

Let $\{g_\ell\}$ be a set of orthonormal functions, $\ell = 1,\dots,L$ fulfilling (5). For periodic signals with a periodicity p a natural choice is $g_{2\ell-1}(t) = \sqrt{2}\sin(\frac{2\pi\ell t}{p})$ and $g_{2\ell}(t) = \sqrt{2}\cos(\frac{2\pi\ell t}{p})$ for $l = 1,\dots,L/2$.

For every voxel X_i, we next calculate the empirical coefficients $B_{i,\ell}$ as

$$B_{i,\ell} = \frac{1}{T}\sum_{t=1}^{T} Y_{i,t}\, g_\ell(t) \tag{8}$$

and use them instead of the original data.

3.2.2 Vector AWS for fMRI

We now apply the AWS procedure as discussed in Section 2. The procedure consists of successive iterations. At each step k and at every voxel X_i, for estimating the Fourier coefficients $\beta_{i,\ell}$, the empirical coefficients $B_{i,\ell}$ are averaged over the growing neighborhood $U^{(k)}(X_i)$ with weights $w_{i,j}^{(k)}$ computed on the base of the result of previous iterations. The procedure reads as follows.

Initialization: For each voxel X_i, we calculate initial estimates of $\beta_{i,\ell}$ and **Var** $\beta_{i,\ell}$ as

$$\begin{aligned}\widehat{\beta}^{(0)}_{i,\ell} &= \frac{1}{N^{(0)}(X_i)} \sum_{X_j \in U^{(0)}(X_i)} B_{j,\ell} \\ \widehat{v}^{(0)}_i &= \frac{1}{T|N^{(0)}(X_i)|^2} \sum_{X_j \in U^{(0)}(X_i)} \widehat{\sigma}^2_j\end{aligned}$$

and set $k = 1$.

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

$$w^{(k)}_{i,j} = K\left(\frac{1}{\lambda}\sum_{\ell=1}^{L} \frac{\left(\widehat{\beta}^{(k-1)}_{i,\ell} - \widehat{\beta}^{(k-1)}_{j,\ell}\right)^2}{\widehat{v}^{(k-1)}_i}\right) \tag{9}$$

for all points X_j in $U^{(k)}(X_i)$ and compute new estimates of $\beta_{i,\ell}$ and **Var** $\beta_{i,\ell}$ as

$$\widehat{\beta}^{(k)}_{i,\ell} = \frac{\sum\limits_{X_j \in U^{(k)}(X_i)} w^{(k)}_{i,j} B_{j,\ell}}{\sum\limits_{X_j \in U^{(k)}(X_i)} w^{(k)}_{i,j}} \tag{10}$$

$$\widehat{v}^{(k)}_i = \frac{\sum\limits_{X_j \in U^{(k)}(X_i)} \left|w^{(k)}_{i,j}\right|^2 \widehat{\sigma}^2_j}{T\left(\sum\limits_{X_j \in U^{(k)}(X_i)} w^{(k)}_{i,j}\right)^2} \tag{11}$$

for all X_i.

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

$$CI^{(k)}_\ell = \left(\widehat{\beta}^{(k)}_{i,\ell} - \eta\sqrt{\widehat{v}^{(k)}_{i,\ell}}, \widehat{\beta}^{(k)}_{i,\ell} + \eta\sqrt{\widehat{v}^{(k)}_{i,\ell}}\right) = \left(CL^{(k)}_\ell, CU^{(k)}_\ell\right) \tag{12}$$

where η^2 is an appropriate quantile of the distribution of the maximum of L χ^2_1 random variables.

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

$$|\widehat{\beta}^{(k)}_{i,\ell} - \widehat{\beta}^{(m)}_{i,\ell}| \le \eta\sqrt{\widehat{v}^{(m)}_{i,\ell}} \tag{13}$$

keeping the previous estimates otherwise.

Stopping: Stop if $k = k^*$ or if $\widehat{\beta}^{(k)}_i = \widehat{\beta}^{(k-1)}_i$ 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 (1998) or Section 3.3 below for a proposal. The parameter λ controlling the probability to reject the hypothesis of two voxel to belong to the same region, can be chosen as a quantile of a χ^2_L distribution. These 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^2_1-distributed random variables. For a comprehensive discussion about the choice of the parameters λ and η see Polzehl and Spokoiny (1998).

3.2.3 Signal detection

It is natural to base the signal detection and identification on the results of the previously described AWS procedure, namely, on the estimates $\widehat{\beta}_{i,\ell} = \widehat{\beta}^{(k^*)}_{i,\ell}$ obtained at the last iteration of the AWS.

One possibility is to compare the estimates $\widehat{\beta}_{i,\ell}$ with the corresponding standard deviation $\widehat{v}^{1/2}_{i,\ell}$ which leads to the test statistic

$$T_i = \sum_{l=1}^{L} \frac{\widehat{\beta}^2_{i,\ell}}{\widehat{v}_{i,\ell}}. \tag{14}$$

The AWS procedure often provides a significant noise reduction so that even very small activations are detected. Sometimes it would be reasonable to show only regions where the activated signal is sufficiently large. For that purpose, we calculate the values

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

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

The parameter τ has the same meaning as the parameter λ in the AWS procedure but we recommend to take it a bit smaller than λ. When carrying over the AWS procedure, we perform the test with the critical value λ many times, which requires a very conservative choice. The final classification is to be done one time, and the choice τ as the usual 0.95 quantile of the χ^2_L distribution leads to reasonable signal identification.

3.2.4 Estimation of the activated signal

In some applications, it is of interest to recover the shape of the induced signal at every point where we detect an activation. This can be done using the spatial smoothing with adaptive weights $w_{i,j} = w^{(k^*)}_{i,j}$ applied in the last step of the AWS procedure. Namely, we first eliminate the trend component from the data using the proposal of Kruggel et al.

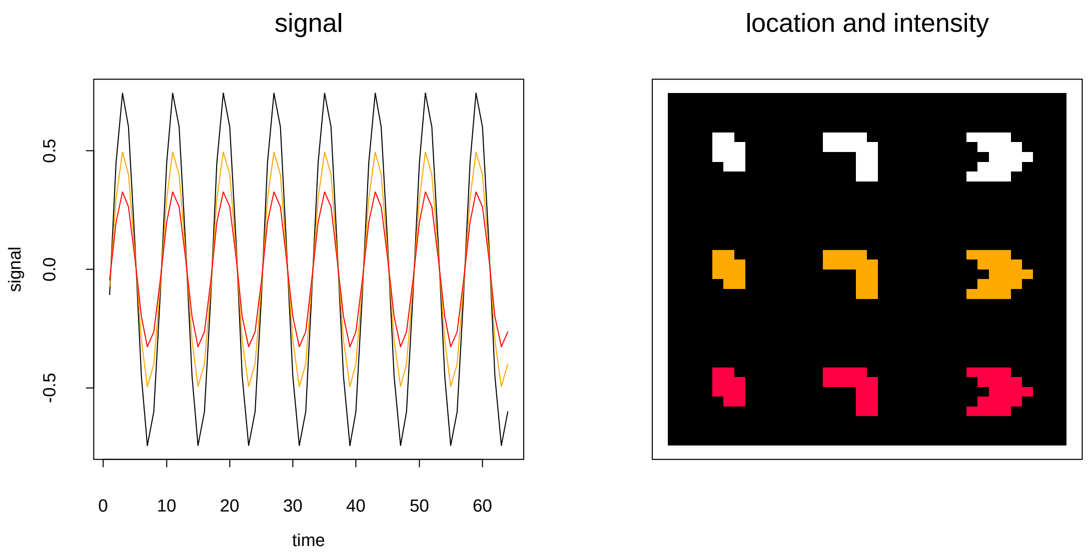


Figure 1: Simulation experiment, periodic signals (left plot) and their location (right image)

(1998), see (6). Then we calculate an estimate $\widehat{f}_t(X_i)$ of the periodic signal by averaging over all periods and over design points X_j with weights $w_{i,j}$:

$$\widehat{f}_t(X_i) = \frac{\sum\limits_j w_{i,j} \left(\frac{p}{T} \sum\limits_{m=1}^{T/p} Z_{i,t+(m-1)p} \right)}{\sum\limits_j w_{i,j}}. \tag{15}$$

3.3 Simulations

We conducted a small simulation study to illustrate the superiority of our approach in idealized i.i.d. situations. 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

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

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 1 illustrates the form of the signals (one period) as 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 $g_1(t) = \sqrt{2}\sin(\frac{2\pi t}{p})$ and $g_2(t) = \sqrt{2}\cos(\frac{2\pi t}{p})$, which

is appropriate in this situation. AWS is performed with $\lambda = 10.6$ corresponding to a 0.995-quantile of χ^2_2, $\eta = 3.5$ 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 were assumed to be known.

We conducted 200 simulation experiments to estimate the pointwise probability of signal detection using our AWS approach. A signal is detected if $T_i > \chi^2_{2,.95} \approx 6$ and $S_i > C_{AWS}$. The threshold C_{AWS} is chosen to give a mean voxelwise detection error of 0.01 for voxel with a distance of more than 2δ from activated regions, with δ being the distance between neighboring points.

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 $\sum_{l=1}^{L} B_{i,l}^2$ exceeds a threshold C_1, with C_1 again selected to provide $\boldsymbol{P}$(signal detected in X_i|no signal in X_i) ≤ 0.01. The second and third alternative involve a preliminary spatial smoothing of the $B_{i,l}$ 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_{AWS}.

Table 1: Mean probability of signal detection in activated regions

Method	$T \times C$	left			central			right		
		top	central	bottom	top	central	bottom	top	central	bottom
AWS	.120	.986	.897	.659	.984	.905	.670	.979	.888	.758
Raw data	9.21	.915	.492	.168	.914	.492	.162	.915	.485	.177
smoothed (0.5δ)	3.80	.994	.815	.354	.996	.816	.362	.994	.810	.398
smoothed (1δ)	.767	1.00	.990	.766	1.00	.984	.774	.999	.969	.832

Table 2: Mean probability of signal detection in a neighborhood $U_4(S)/S$ of the activated regions (no signal).

Method	$T \times C$	left			central			right		
		top	central	bottom	top	central	bottom	top	central	bottom
AWS	.120	.008	.022	.058	.008	.029	.089	.018	.043	.079
Raw data	9.21	.012	.011	.009	.011	.011	.009	.008	.009	.009
smoothed (0.5δ)	3.80	.023	.017	.011	.023	.015	.012	.024	.018	.011
smoothed (1δ)	.767	.498	.220	.069	.571	.297	.112	.513	.284	.118

Tables 1 to 3 provide mean values of $\boldsymbol{P}$(signal detected in X_i) for the different approaches and regions. The results clearly show a better overall behaviour of AWS. Signal detection without spatial smoothing suffers from high threshold values, resulting in prob-

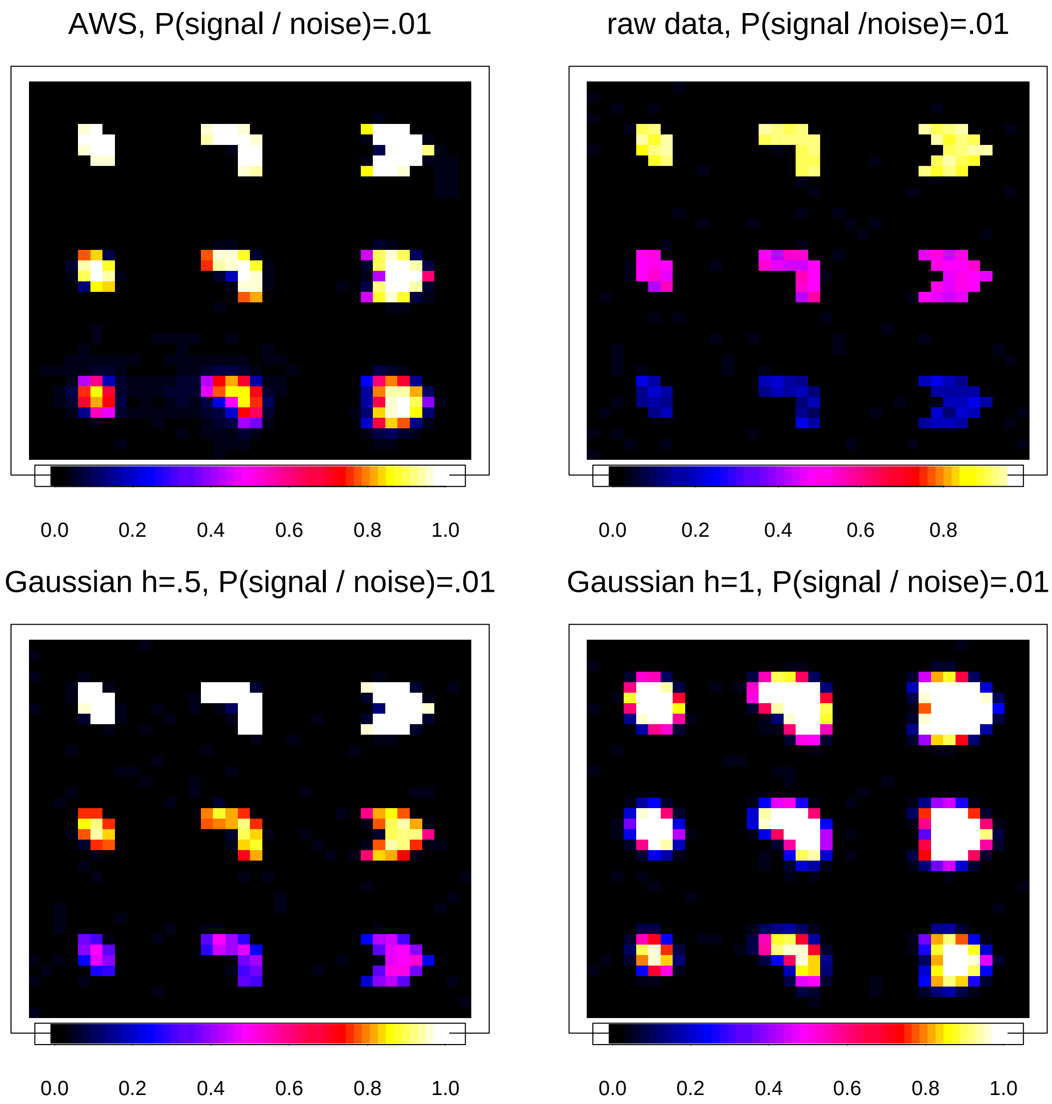


Figure 2: Pointwise signal detection probabilities obtained from 200 simulations.

Table 3: Mean probability of false detection for subimages of 12x12 voxel including one activated region

Method	$T \times C$	left			central			right			over
		top	central	bottom	top	central	bottom	top	central	bottom	all
AWS	.120	.007	.017	.045	.006	.022	.063	.014	.030	.055	.028
Raw data	9.21	.016	.044	.068	.017	.064	.095	.019	.079	.123	.058
smoothed (0.5δ)	3.80	.012	.024	.055	.011	.030	.075	.011	.037	.093	.038
smoothed (1δ)	.767	.070	.037	.033	.104	.059	.048	.110	.068	.053	.065

lems to detect weak signals. Non-adaptive spatial smoothing improves signal detection within the activated regions but reduces regional specificity, see also Poline et.al. (1997).

Figure 2 illustrates the pointwise probabilities of signal detection obtained from 200 simulation runs. Figure 3 provides the detection results for a typical time series of images,

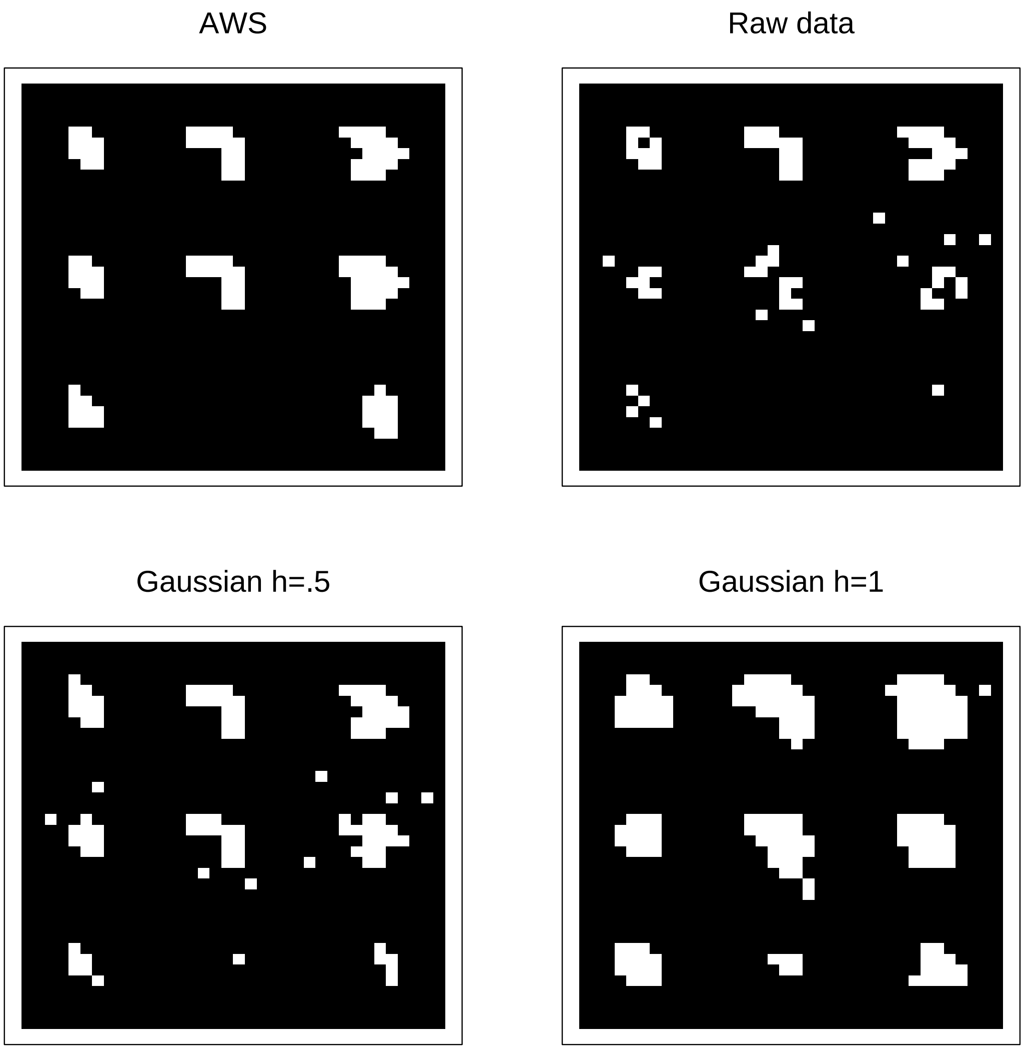


Figure 3: Detected signals for a typical dataset from the simulation study using AWS (upper left), raw data (upper right) and spatial smoothing (lower row) using comparable thresholds.

i.e. with medium number of detection errors for all approaches, from our simulations.

3.4 A real life example

The example we present here is based on a data set kindly provided by F. Kruggel from the Max-Plank-Institute of Cognitive Neuroscience at Leipzig, Germany. 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 255. The fMRI series are already corrected for artifacts and body movement. Additionally high resolution MRI of the same slices are given. The upper plot in Figure 4 shows the 100th image of the time series from the third slice. Spatial resolution is about $2mm \times 4mm$. The patient is exposed

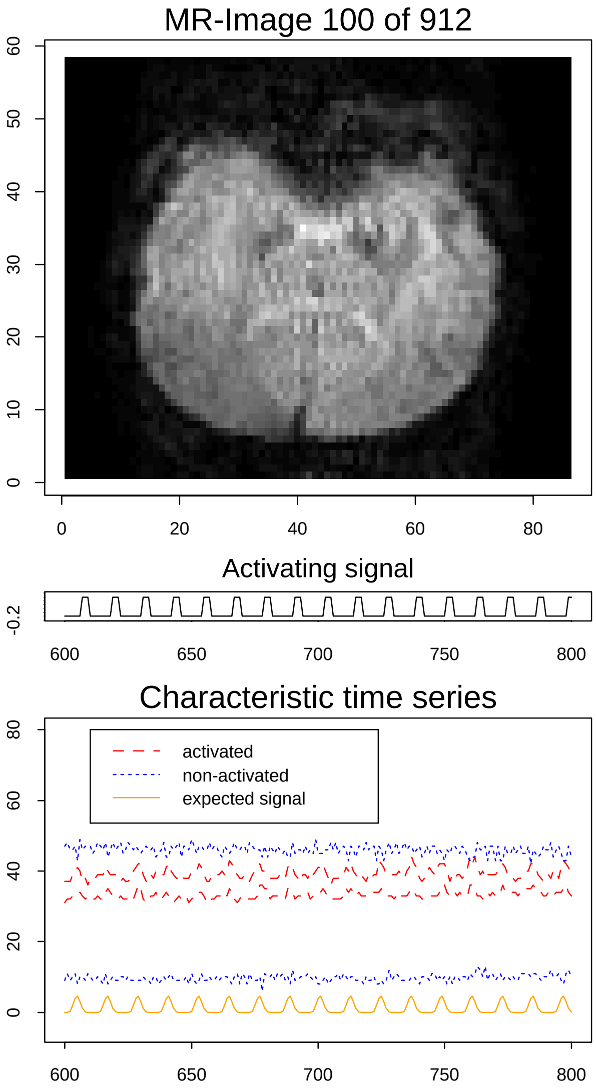


Figure 4: Functional MRI data

to a periodic signal, with no activation for the first 48 time intervals, see second plot from top. The lower part of Figure 4 shows characteristic observed time series for voxel from both activated and non-activated regions.

The data are generated using simple periodic experimental conditions suggesting a periodic induced (BOLD) signal of same periodicity $p = 12$. We first aggregate the data using a two term Fourier approximation with basis functions $g_1(t) = \sqrt{2}\sin(\frac{2\pi t}{p})$ and $g_2(t) = \sqrt{2}\cos(\frac{2\pi t}{p})$ leading to Fourier coefficients $B_{i,l}, l = 1, 2$ for each voxel i. Mean periodic signals $\widehat{f}_t(X_i)$ are computed according to equation (15). We then start our AWS procedure as described in Section 3.2.

Signal detection is based on the statistics T_i and S_i introduced in Section 3.2. Figure 5 illustrates the output of our detection procedure. The central plot shows the intensity of

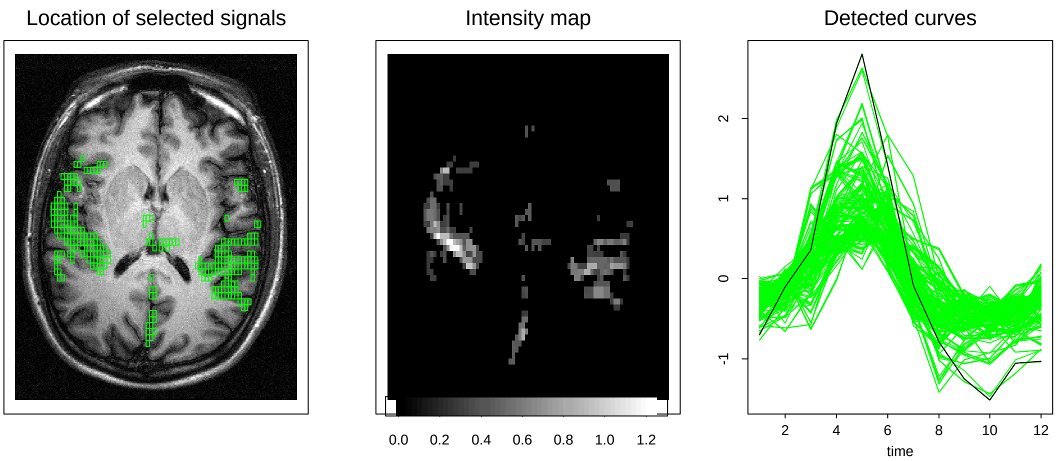


Figure 5: Signal detection with AWS. Central plot: intensity map of detected signals. Left plot: Position of selected signals in the corresponding high resolution image. Right plot: mean periodic signals in selected voxel.

all signals detected using the threshold $\tau = \chi^2_{2,0.95} \approx 6$ and meeting $S_i > C = 0.08$. This choice excludes signals with amplitude smaller than 0.4 . 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 $S_i(t)$ for all voxel marked in the left plot. 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 stimulation. Our approach easily allows to incorporate this by selecting an appropriate aggregation of the time series. In order to test for time inhomogeneity in our example we divide the time series into tree segments of length 288 and compute Fourier coefficients for each part as

$$B_{il} = 1/288 \sum_{t=49+(l-1)*288}^{48+l*288} Y_{i,t} g_1(t) \quad \text{and} \quad B_{i(l+3)} = 1/288 \sum_{t=49+(l-1)*288}^{48+l*288} Y_{i,t} g_2(t). \qquad (16)$$

We now carry out the adaptive weights smoothing based on six Fourier coefficients. Signal detection based on T_i and S_i with threshold $\tau = \chi^2_{6,.95} \approx 12.6$ and $C = 0.24$ gives the results shown in Figure 6. The curves $\widehat{f}_t(X_i)$ displayed in the right plot of Figure 6 are now generated stringing the mean periodic curves from the three segments together. The findings are essentially the same as in the previous setting indicating no time inhomogeneity here.

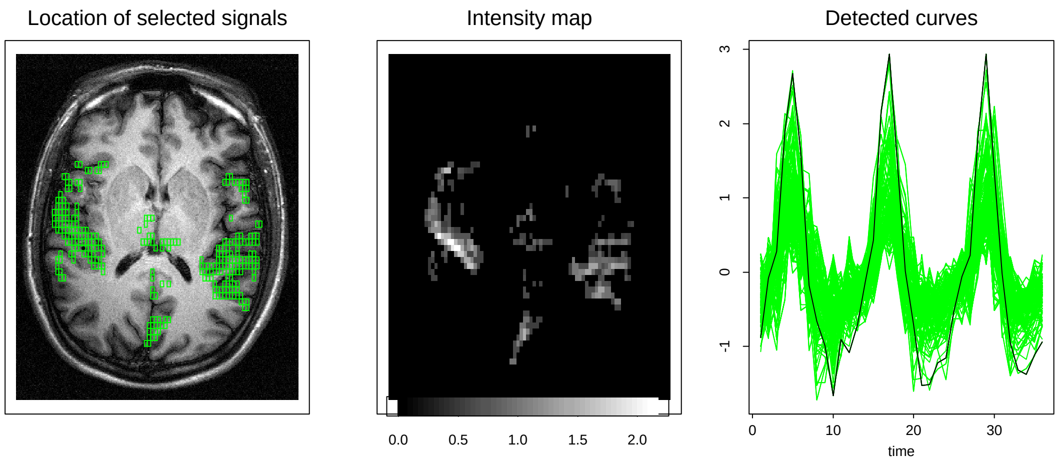


Figure 6: Signal detection with AWS for possibly inhomogeneous situations. Central plot: intensity map of detected signals. Left plot: Position of selected signals in the corresponding high resolution image. Right plot: mean periodic signals (from three segments) in selected voxel.

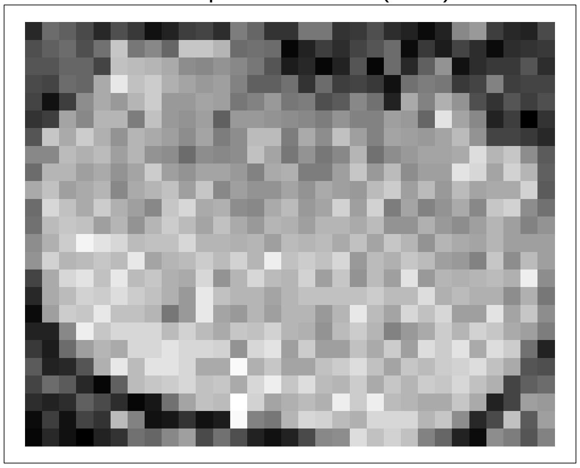
4 An application to dynamic MRI

The same approach can be used to analyze data from dynamic MRI (dMRI) experiments, see also Section 1.3.

We illustrate the use of AWS in this context using the same example as Sebastiani et. al. (1996). The data consist 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.

Figure 7 displays the central part of the first image (top) together with characteristic time series 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.

Let $Y_{i,t} = f_t(X_i) + \epsilon_{i,t}$ be the observed time series of MR intensities for voxel i. In a preliminary transformation step we remove the effect of the anatomic structure subtracting a baseline estimate obtained from the first 6 images. Recall that these images do not reflect any effect from the contrast agent. This provides transformed time series $Z_{i,t} = Y_{i,t} - 1/6 \sum_{s=1}^{6} Y_{i,s}$.



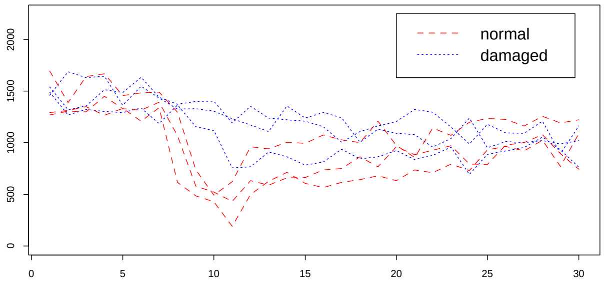


Figure 7: Central part of first image from a dMRI series of the brain of a rat (upper plot) and characteristic time series from both normal and pathologic areas (lower plot)

We now show how vector AWS can be used in this context. We again start with a data aggregation step. 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,

$$\begin{aligned} &g_1(t) \equiv \tfrac{1}{\sqrt{3}} I_{t\in(6,9]}, \quad g_2(t) = \tfrac{1}{\sqrt{3}} I_{t\in(9,12]}, \\ &g_3(t) = \tfrac{1}{\sqrt{6}} I_{t\in(12,18]}, \quad g_4(t) = \tfrac{1}{\sqrt{6}} I_{t\in(18,24]}, \quad g_5(t) = \tfrac{1}{\sqrt{6}} I_{t\in(24,30]}, \end{aligned} \tag{17}$$

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

$$B_{i,j} = \sum_{t=7}^{30} Z_{i,t} g_l(t). \tag{18}$$

The first two intervals are chosen to be smaller to reflect the expected higher variation of the curves at times $6-12$. 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 the last intervals. Peak intensity can be easily estimated from the aggregated data.

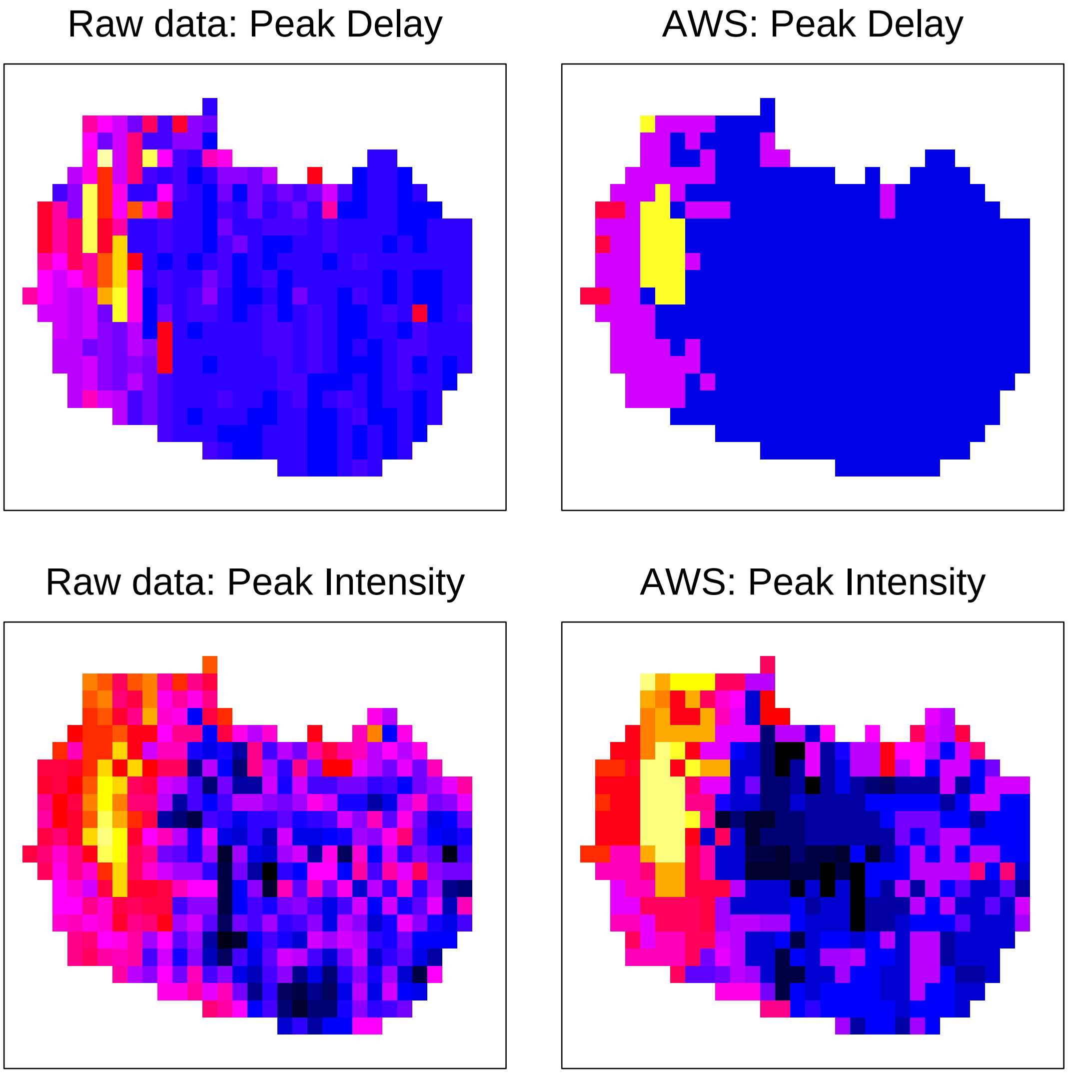


Figure 8: Peak delay (upper row) and peak intensity (lower row) maps obtained from the original time series (left column) and the AWS estimates (right column).

Variance estimates are again obtained using second order differences of the $Y_{i,t}$. The variance estimates show no significant spatial inhomogeneity. We therefore use a mean (homogeneous) variance estimate.

We perform a spatial smoothing using Vector AWS as described in Section 3. We then, for each voxel, calculate estimates of the peak delay D and peak intensity Δ as

$$\widehat{D}_i = \arg\min_l \frac{\widehat{\beta}_{i,l}}{c_l} \quad \text{and} \quad \widehat{\Delta}_i = \min_l \frac{\widehat{\beta}_{i,l}}{c_l} - \frac{\widehat{\beta}_{i,5}}{c_5}, \tag{19}$$

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

Figure 8 illustrates the results. The upper left image shows a peak delay map calculated from the original data. In the upper right we display the corresponding map $\widehat{D}$ obtained from the AWS estimates $\widehat{\beta}_{i,l}$. The bottom row gives 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 classification results using other approaches.

5 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 classification 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 (1998). Theoretical properties of the method especially for applications to multiple datasets 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 and Gerhard Winkler for helpful remarks and discussion.

References

[1] Banerjee, S. and Rosenfeld, A. (1993). MAP estimation of piecewise constant digital signals. *CVGIP: Image understanding*, **57**, 63–80.

[2] Descombes, X., Kruggel, F. and von Cramon, D. (1997). spatio-temporal fMRI Analysis using Markov Random Fields. Technical Report 4/97, MPI of Cognitive Neuroscience Leipzig, Germany.

[3] Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Ass.* **91** 674–688.

[4] Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications.* Chapman & Hall, London.

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

[6] Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C. and Evans, A.C. (1994). Assessing the Significance of Focal Activations Using Their Spatial Extent. *Human Brain Mapping* 1:210-220.

[7] Hart, J. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests* New York, Berlin, Heidelberg: Springer.

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

[9] Kallenberg, W.C.M. and Ledwina, T. (1997). Data driven smooth tests when the hypotheses is composite. *Journal of the American Statistical Association* **92**, 1094-1104.

[10] Kruggel, F., Descombes, X. and von Cramon, D. (1998). Preprocessing of fMR Datasets. Technical Report, MPI of Cognitive Neuroscience Leipzig, Germany.

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

[12] Ledwina, T. (1994). Data-driven version of Neyman's smooth test of fit. *J. Amer. Statist. Ass.* **89** 1000–1005.

[13] Lepski, O. and Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, **25**, no. 6, 2512–2546.

[14] Liang, Z., MacFall, J. R. and Harrington, D. P. (1994). Parameter estimation and tissue segmentation from multispectral MR Images. *IEEE Transactions on Medical Imaging*, **13**, 441–449.

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

[16] Poline, J.-B., Holmes,A., Worsley,K.J. and Friston, K.J. (1997). Statistical inference and the theory of Gaussian fields. In SMPcourse, Institute of Neurology, Wellcome Dept of Cognitive Neurology, University College London, `http://www.fil.ion.ucl.ac.uk/spm/course/notes.html`

[17] Poline, J.-B. and Mazoyer, B.M. (1993). Analysis of Individual Positron Emission Tomography Activation Maps by Detection of High-Signal-to-Noise Ratio pixel Clusters. *Journal of Cerebral Blood Flow and Metabolism*, **13**, 425-437.

[18] Polzehl, J. and Spokoiny, V. G. (1998). Adaptive weights smoothing with applications to image restoration. Preprint 405, WIAS Berlin.

[19] 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.

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

[21] 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.

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

[23] Sebastiani, G. and Barone, P. (1991). Mathematical principles of basic magnetic resonance imaging in medicine. *Signal Processing*, **25** , 227–250.

[24] Sebastiani, G. and Barone, P. (1995). Truncation artifact reduction in magnetic resonance imaging by Markov random field methods. *IEEE Transactions on Medical Imaging*, **24**, 434–441.

[25] 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.

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

[27] Spokoiny, V. (1999). Data-driven testing the fit of linear models. Preprint 472, WIAS Berlin.

[28] Turner, R. and Friston, K.J. (1997). Functional MRI. In SMPcourse, Institute of Neurology, Wellcome Dept of Cognitive Neurology, University College London, `http://www.fil.ion.ucl.ac.uk/spm/course/notes.html`

[29] Worsley, K.J., Evans, A.C., Marrett, S. and Neelin, P. (1992). A Three-Dimensional Statistical Analysis for CBF Activation Studies in Human Brain. *Journal of Cerebral Blood Flow and Metabolism* **12**, 900-918.