

# Statistical Modeling of Time-Dependent fMRI Activation Effects

Stefanie Kalus,<sup>1\*</sup> Ludwig Bothmann,<sup>1</sup> Christina Yassouridis,<sup>2</sup>  
Michael Czisch,<sup>3</sup> Philipp G. Sämann,<sup>3</sup> and Ludwig Fahrmeir<sup>1</sup>

<sup>1</sup>Department of Statistics, Ludwig-Maximilians-University, Ludwigstr. 33, 80539 Munich, Germany

<sup>2</sup>Institute of Applied Statistics and Computing, University of Natural Resources and Life Sciences, Peter-Jordan-Str. 82, 1190 Vienna, Austria

<sup>3</sup>Neuroimaging Research Group, Max Planck Institute of Psychiatry, Kraepelinstr. 2, 80804 Munich, Germany

---

**Abstract:** Functional magnetic resonance imaging (fMRI) activation detection within stimulus-based experimental paradigms is conventionally based on the assumption that activation effects remain constant over time. This assumption neglects the fact that the strength of activation may vary, for example, due to habituation processes or changing attention. Neither the functional form of time variation can be retrieved nor short-lasting effects can be detected by conventional methods. In this work, a new dynamic approach is proposed that allows to estimate time-varying effect profiles and hemodynamic response functions in event-related fMRI paradigms. To this end, we incorporate the time-varying coefficient methodology into the fMRI general regression framework. Inference is based on a voxelwise penalized least squares procedure. We assess the strength of activation and corresponding time variation on the basis of pointwise confidence intervals on a voxel level. Additionally, spatial clusters of effect curves are presented. Results of the analysis of an active oddball experiment show that activation effects deviating from a constant trend coexist with time-varying effects that exhibit different types of shapes, such as linear, (inversely) U-shaped or fluctuating forms. In a comparison to conventional approaches, like classical SPM, we observe that time-constant methods are rather insensitive to detect temporary effects, because these do not emerge when aggregated across the entire experiment. Hence, it is recommended to base activation detection analyses not merely on time-constant procedures but to include flexible time-varying effects that harbour valuable information on individual response patterns. *Hum Brain Mapp* 36:731–743, 2015. © 2014 Wiley Periodicals, Inc.

**Key words:** functional magnetic resonance imaging; time-varying activation and hemodynamic response function; varying coefficient model; penalized least squares estimation; event-related functional magnetic resonance imaging; auditory oddball

---

## INTRODUCTION

Most functional imaging experiments use the repeated presentation of event-related stimuli and averaging of resulting stimulus-related brain responses to increase the signal-to-noise-ratio over the experiment to detect stimulus-induced activation patterns. Aggregating the response across all stimulus occurrences relies on the assumption of a steady response over time—which may be superimposed by different sources of noise. It has, however, been shown by target detection paradigms that the

---

Contract grant sponsor: German Science Foundation; Contract grant number: FA 128/6-1.

\*Correspondence to: Dr. Stefanie Kalus, Department of Statistics, Ludwig-Maximilians-University, Ludwigstr. 33, 80539 Munich, Germany. E-mail: stefanie.kalus@stat.uni-muenchen.de

Received for publication 14 May 2014; Revised 17 September 2014; Accepted 6 October 2014.

DOI: 10.1002/hbm.22660

Published online 23 October 2014 in Wiley Online Library (wileyonlinelibrary.com).

positive electrophysiological deflection with latency 300 ms (ref. to as P300) underlies habituation, as, for example, obvious during an auditory oddball experiment in which regular nontarget tones alternate with fewer target tones [Ivey and Schmidt, 1993; Koelega et al., 1992; Lammers and Badia, 1989; Wesensten et al., 1990]. This standard paradigm has also been studied by functional magnetic resonance imaging (fMRI) blood oxygenation level dependent (BOLD) imaging to map the spatial pattern of this response and the exact areas in which habituation occurs [Kiehl and Liddle, 2003; Kiehl et al., 2005]. Attention is another neuropsychological factor that leads to time-on-task effects on the P300, for example, demonstrated by combined electroencephalography (EEG) and fMRI studies that correlated intertrial-variability of the P300 with the BOLD status of attention networks [Bénar et al., 2007; Mantini et al., 2009]. Indeed, there is general converging evidence that intrinsic fluctuations of cortical activity predict or “bias” the response to individual stimuli [Coste et al., 2011; Sadaghiani et al., 2009]. Mostly, in these studies, single-trial-BOLD responses in the peristimulus time-window are simply quantified through stratification, using an external criterium, for example, behavioral detection versus nondetection of a stimulus [Sadaghiani et al., 2009].

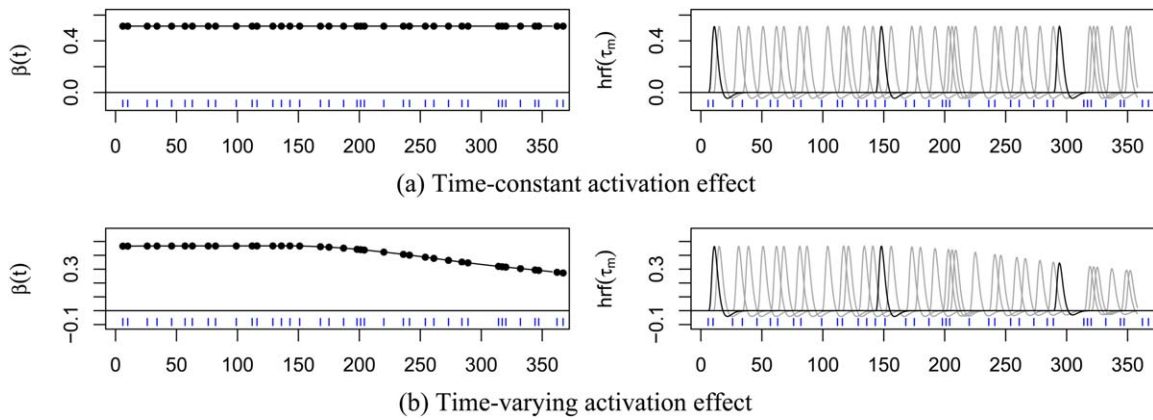
In typical activation studies that pursue the goal to derive generalizable maps of stimulus-induced activation, it is common practice to include the stimulus onset time point as stick function or neural activation of defined length and convolute this with the presumed hemodynamic response function (HRF) within the framework of a general linear model (GLM) [Friston et al., 1995]. As described in more detail below, modeling and estimation of the HRF still receives much attention, but it is usually assumed that its functional form as well as activation effects are time invariant and do not change during the course of an experiment. Although detection of activation changes caused by varying attention, habituation or other factors, is often of direct interest, reports on flexible models for capturing time-varying activation effects and HRFs are sparse or basic from a technical point of view. Most existing approaches specify deviations from time-constant activation effects through prechosen, simple parametric functions, such as an exponential decay or increase, see, for example, Büchel et al. [1998] or Rodriguez [2010] as well as some further references given there. Grindband et al. [2008] shares a similar problem, making an assumption about the course of time variation, that is, dependent on the globally (not space varying) measured reaction time. It is generally difficult, however, to choose adequate functional forms prior to some preliminary data analysis or some biomathematical model, in particular because activation patterns are spatially varying across brain regions. Moreover, inclusion of unknown hyperparameters leads to a nonlinear regression problem with possible problems of convergence due to nonconvexity and multiple modes.

Another choice would be to approximate time-varying effects through step functions. A typical case is reported in a recent study of human olfaction at the individual level [Morrot et al., 2013]. To measure olfactory habituation, fMRI time series are split into four consecutive periods and standard SPM analyses with the canonical HRF were applied within each period. It turned out that there are significantly different effects between these periods in certain regions of interest (ROIs). Similarly, we observed time-varying effects in our own experiments (Results section) when dividing fMRI time series into consecutive periods. Obviously, such step functions to account for time variation of activation effects and HRFs are only a very coarse approximation to underlying, smoothly varying temporal patterns. Therefore, more flexible approaches for modeling and recovering time-varying activation effects and HRFs are desirable. Such an attempt, based on smoothed step functions and Kalman filtering has been previously suggested in Gössl et al. [2000]. Other approaches for estimating trial-specific response estimates (e.g., based on Ridge Regression) have recently been proposed by Mumford et al. [2012] for optimizing classification tasks, however, the temporal proximity of consecutive stimuli is not exploited in their approaches.

Here, we present a novel method that models time-varying effects nonparametrically, allowing for the extraction of trial-specific response estimates, and exemplify how these coefficients may serve as incoming information for secondary analyses, such as functional clustering of activation patterns. In more technical terms, within the regression framework of the GLM approach, the stimulus effect at a certain voxel is usually modeled as the convolution of the observed stimulus function  $u(t)$  where  $t$  denotes time starting with the beginning of the experiment, and a HRF  $\text{hrf}(\tau)$ , where  $\tau$  denotes time after stimulus onset, see the description of the fMRI regression model with time-constant effects in the Method section. In many studies,  $\text{hrf}(\tau)$  is assumed to have a fixed shape, such as SPMs canonical HRF which is the difference of two gamma functions [Worsley and Friston, 1995]. However, to increase flexibility, the HRF is often modeled as a linear combination of  $p$  basis functions  $B_k(\tau), k=1, \dots, p$ , that is,

$$\text{hrf}(\tau) = \sum_{k=1}^p \beta_k B_k(\tau), \quad (H)$$

where the unknown basis function coefficients  $\beta_k, k=1, \dots, p$ , have to be estimated from the fMRI time series  $y(t)$  observed at a specific voxel. For  $p=1$ , HRFs with a fixed shape  $B_1(\tau)$  and amplitude  $\beta_1$  are a special case of (H). A popular alternative is to choose a small to moderate number of basis functions, such as  $p=3$  gamma densities with prechosen shape and scale parameters [Friston et al., 1998] or the canonical basis function set, including the canonical HRF plus its temporal and dispersion derivatives [Friston



**Figure 1.**

Comparison of time-constant and time-varying activation effects. On the left side, the underlying effect curve  $\beta(t)$  is plotted over time where dots denote the  $\beta(\tau_m)$  values to presented stimuli  $= 1, \dots, M$  (blue stick functions). On the right side, the corresponding  $\text{hrf}(\tau_m)$  functions in response to the presented stimuli are shown.

et al., 2008, pp. 181]. In some cases, the basis functions contain additional unknown parameters, such as the inverse logit model, which consists of the superposition of three inverse logit functions [Descamps et al., 2012; Lindquist and Wager, 2007].

The most flexible HRF models are finite impulse response (FIR) models. They are step functions, allowing estimation of the height of the HRF at each time point within a window of time following stimulation. FIR models can be written in the form  $(\text{hrf})$  with 0–1 basis function  $B_k(t)$  at every time point  $\tau$  and  $\beta_k$  as the height of the HRF at  $\tau$ , see, for example, Ollinger et al. [2001]. However, the increased flexibility comes at the cost of a large number  $p$  of parameters, with the risk of overfitting the data, fewer degrees of freedom, and decreased power [Lindquist et al., 2009]. Therefore, additional constraints are introduced, for example, in the time-event separable model of Kay et al. [2008] or in form of regularization or smoothing techniques [Badillo et al., 2013; Makni et al., 2008; Zhang et al., 2008; Zhang et al., 2012].

Lindquist et al. [2009] investigate the performance of several HRF models  $(H)$  and the recent review of Monti [2011] on the GLM approach also provides information about the current state of modeling and estimating HRFs. HRFs of the form  $(H)$  are also used beyond the GLM framework, for example, in spatial-temporal hidden process models [Hutchinson et al., 2009; Shen et al., 2014].

So far, however, in all HRF models the basis function coefficients  $\beta_1, \dots, \beta_p$  are assumed to be time constant during the experiment, relying on the assumption of a steady brain response over time. This induces a time constant shape of the HRF during the experiment as visualized in Figure 1a. To incorporate flexible, but still smoothly varying time patterns, it is necessary to replace time-constant coefficients through time-varying coeffi-

cients  $\beta(t) = (\beta_1(t), \dots, \beta_p(t))$ , inducing time-varying shapes of the HRF as visualized in Figure 1b. To avoid too restrictive parametric functional forms, such as simplistic step functions, exponential functions, or low-dimensional polynomials, we follow a flexible semiparametric approach for specifying and estimating  $\beta(t)$ . We approximate time-varying coefficients through flexible penalized regression splines, introduced in Eilers and Marx [1996] and meanwhile often preferred to smoothing splines, see Fahrmeir et al. [2013] for a recent review. We show in our methodological derivation that this leads to a linear model with penalized least squares estimation. We outline the proposed method in the context of event-related experiments and apply it to SPMs canonical basis function set. However, it is important to note that our approach is quite general and extends any fMRI model  $(H)$  by replacing time-constant coefficients through time-varying coefficients.

In the Result section, we use exploratory tools, such as confidence bands and functional clustering to capture flexible activation patterns. The development of more formal tests or model choice tools—on the voxel and in particular on the whole brain level—are subject of current research and beyond the scope of this article.

The time-varying fMRI activation effects algorithm described in the Methods section has been implemented in a user-friendly software package. The software is freely available as add-on package RfmriVC to the R system of statistical computing [R Core Team, 2013].

The remainder of the article is organized as follows: In the Methods section, we first describe the basic fMRI GLM used conventionally for activation detection. This model relies on the assumption that the activation effects are time constant. Next, we show how this model can be extended to incorporate time-varying effect coefficients and give an outline to model inference. After providing

the methodological background, we present results obtained on an event-related fMRI dataset from an acoustic two-tone oddball design. In the Discussion section, we outline the main qualities and findings of our model and suggest starting points for further work.

## METHODS

### An fMRI Regression Model with Time-Constant Effects

fMRI data consist of signal time series  $y_{it}, t=1, \dots, T$ , recorded at each voxel  $i=1, \dots, N$  of a three-dimensional brain image. Detection of brain activity is usually based on voxelwise regression models of the form

$$y_{it} = f_{\text{base}}(i, t) + f_{\text{con}}(i, t) + f_{\text{stim}}(i, t) + \varepsilon_{it}. \quad (1)$$

In (1),  $f_{\text{base}}(i, t)$  is the baseline trend,  $f_{\text{con}}(i, t)$  is the effect of confounding covariates,  $f_{\text{stim}}(i, t)$  is the effect of the (transformed) stimulus, and  $\varepsilon_{it}$  is the random error term at voxel  $i$  and time  $t$ . In the following, we discuss the components of the fMRI regression model (1) in detail.

The baseline term  $f_{\text{base}}(i, t)$  corrects for slow periodic variations and drift either inherent to the scanning procedure or connected to nonparadigm correlated periodic variations. Thus,  $f_{\text{base}}(i, t)$  serves as a highpass filter. In this work, it is chosen to consist of a discrete cosine transform set [Friston et al., 2008, p. 123] as in SPM. In contrast to SPM, but conceptually equivalent, the highpass filter in our modeling approach enters the regression stage directly as a linear combination of basis functions  $f_{\text{base}}(i, t) = \mathbf{w}^{(b)}(t)' \boldsymbol{\alpha}_i^{(b)}$  with basis functions  $\mathbf{w}^{(b)}(t) = (w_k^{(b)}(t), k=1, \dots, p_b)'$  and weights  $\boldsymbol{\alpha}_i^{(b)} = (\alpha_{ik}^{(b)}, k=1, \dots, p_b)'$ .

The second term  $f_{\text{con}}(i, t)$  accounts for further confounding effects like, for example, movement related artifacts or brain tissue-specific properties capturing effects from cardiac and respiratory cycles, which are not captured by the highpass filter. It is assumed, that according information is available in form of several univariate, global variables with value  $\mathbf{w}^{(c)}(t) = (w_k^{(c)}(t), k=1, \dots, p_c)'$  at time  $t$ . The corresponding voxelspecific effect vector  $(\alpha_{ik}^{(c)}, k=1, \dots, p_c)'$  is denoted as  $\boldsymbol{\alpha}_i^{(c)}$ , so that  $f_{\text{con}}(i, t) = \mathbf{w}^{(c)}(t)' \boldsymbol{\alpha}_i^{(c)}$ .

The regression component  $f_{\text{stim}}(i, t)$  includes the transformed stimulus time series of a given type. The fMRI signal represents aggregated and time delayed neuronal activity, which in turn has a close correspondence to the stimulus presentation. Therefore, the stimulus time series has to be transformed to the level of the fMRI response to obtain models that more closely resemble the observed fMRI signal, thus, yielding a better fit. In this work, the focus is on modeling event related stimuli, although extensions to block designs are conceivable for the proposed methodology.

We follow an approach proposed by Josephs et al. [1997] using the concept of mathematical convolution:

$$f_{\text{stim}}(i, t) = \int_0^{\tau_{\text{max}}} \text{hrf}(i, \tau) u(t-\tau) d\tau$$

where the function  $u(t)$  describes the given time course of stimulation and  $\text{hrf}(i, \tau)$  is the unknown HRF at voxel  $i$ . To estimate  $\text{hrf}(i, \tau)$ , a flexible modeling strategy with basis functions  $B_k$  and corresponding voxelspecific weights  $\beta_{ik}$  is applied:

$$\text{hrf}(i, \tau) = \sum_{k=1}^p B_k(\tau) \beta_{ik}.$$

This approach leads to a flexible and data driven estimation of the voxelspecific functional form of the hemodynamic response. Different choices of basis sets exist [Henson et al., 2001]. In the application, we focus on the canonical basis function set [Friston et al., 2008, pp. 181], which is the default choice in SPM.

The time series of neuronal activity  $u(t)$  is set equal to the signal time series, which is modeled as follows: Suppose we have event related stimuli of one type at times  $\tau_1, \tau_2, \dots, \tau_M$ . A stimulus at time  $\tau_m$  is modeled via a dirac delta function  $\delta(t-\tau_m)$ , that is, a stick-function, so that  $u(t) = \sum_{m=1}^M \delta(t-\tau_m)$ .

With this modeling strategy, the stimulus predictor for all presented stimuli is linearized with respect to unknown HRF effects:

$$f_{\text{stim}}(i, t) = \sum_{k=1}^p \underbrace{\sum_{m=1}^M B_k(t-\tau_m)}_{=z_k(t)} \beta_{ik} = \mathbf{z}(t)' \boldsymbol{\beta}_i, \quad (2)$$

where  $\mathbf{z}(t)' = (z_k(t), k=1, \dots, p)$  and  $\boldsymbol{\beta}_i = (\beta_{ik}, k=1, \dots, p)'$ .

Linearization of additive regression components leads to a reformulation of model (1) into voxelwise linear models

$$y_{it} = \mathbf{w}^{(b)}(t)' \boldsymbol{\alpha}_i^{(b)} + \mathbf{w}^{(c)}(t)' \boldsymbol{\alpha}_i^{(c)} + \mathbf{z}(t)' \boldsymbol{\beta}_i + \varepsilon_{it} \\ = \mathbf{w}(t)' \boldsymbol{\alpha}_i + \mathbf{z}(t)' \boldsymbol{\beta}_i + \varepsilon_{it} \quad (3)$$

where  $\mathbf{w}(t)' = (\mathbf{w}^{(b)}(t)', \mathbf{w}^{(c)}(t)')$  and  $\boldsymbol{\alpha}_i' = (\boldsymbol{\alpha}_i^{(b)'}', \boldsymbol{\alpha}_i^{(c)'}')$  are concatenated vectors. Collecting all observations and design vectors for voxel  $i$  in observation vectors and design matrices, we obtain a voxelwise linear model with time-constant stimulus effects of the form

$$y_i = \mathbf{W} \boldsymbol{\alpha}_i + \mathbf{Z} \boldsymbol{\beta}_i + \varepsilon_i.$$

Our model depends on the assumption  $\varepsilon_i \sim N_T(\mathbf{0}, \sigma_i^2 \mathbf{I})$  where  $\mathbf{I}$  is the identity matrix of size  $T$ . So far, our model does not account for serial correlations, because the added value is expected to be low. That is, our explorative postprocessing analysing strategy does not rely on the exact modeling of the error process traditionally needed for significance testing.

### An fMRI Regression Model with Time-Varying Effects

A basic assumption in modeling the stimulus predictor in (2) is that the effect  $\beta_i = (\beta_{i1}, \dots, \beta_{ip})$  is time constant and does not vary during the experiment. This implies that  $\text{hrf}(\tau_m)$  is the same for all  $M$  presented stimuli in the experiment (cf. Fig. 1a). This may be questionable and we relax it by allowing the stimulus effect vector to be a function of time, that is,  $\beta_i(t) = (\beta_{i1}(t), \dots, \beta_{ip}(t))$ . To avoid too restrictive parametric forms, we assume that each effect component can be expressed as a flexible linear combination of spline basis functions  $s_j(t), j=1, \dots, q$ , defined on a prechosen grid of time points, that is,

$$\beta_{ik}(t) = \sum_{j=1}^q s_j(t) \gamma_{ik,j} = \mathbf{s}(t)' \boldsymbol{\gamma}_{ik}. \quad (4)$$

We choose a cubic B-spline basis with a generous number  $q$  of basis functions (10–30) to guarantee flexibility of the unknown functions of time.

We now replace the time-constant effects  $\beta_{ik}$  in the stimulus predictor in (2) with the values of time-varying effects at stimulus times  $\tau_m$ , leading to an extension of the fMRI model (3) into

$$y_{it} = \mathbf{w}(t)' \boldsymbol{\alpha}_i + \sum_{k=1}^p \sum_{m=1}^M B_k(t - \tau_m) \beta_{ik}(\tau_m) + \varepsilon_{it} \quad (5)$$

with time-varying stimulus effects. With this, we allow the stimulus effect—and thus the hemodynamic response—to vary smoothly over time (cf. Fig. 1b for an exemplified decreasing effect). Inserting the spline representation (4) leads to

$$\begin{aligned} y_{it} &= \mathbf{w}(t)' \boldsymbol{\alpha}_i + \sum_{k=1}^p \underbrace{\sum_{m=1}^M B_k(t - \tau_m) \mathbf{s}(\tau_m)' \boldsymbol{\gamma}_{ik}}_{=\mathbf{x}_k(t)'} + \varepsilon_{it} \\ &= \mathbf{w}(t)' \boldsymbol{\alpha}_i + \sum_{k=1}^p \mathbf{x}_k(t)' \boldsymbol{\gamma}_{ik} + \varepsilon_{it} = \mathbf{w}(t)' \boldsymbol{\alpha}_i + \mathbf{x}(t)' \boldsymbol{\gamma}_i + \varepsilon_{it} \end{aligned}$$

with design vectors  $\mathbf{x}(t)' = (\mathbf{x}_1(t)', \dots, \mathbf{x}_p(t)')$  and according effect vectors  $\boldsymbol{\gamma}_i = (\boldsymbol{\gamma}_{i1}, \dots, \boldsymbol{\gamma}_{ip})$ .

Collecting all observations and design vectors for voxel  $i$  in observation vectors and design matrices, we obtain a voxelwise linear model with time-varying stimulus effects of the form

$$\begin{aligned} \mathbf{y}_i &= \mathbf{W} \boldsymbol{\alpha}_i + \sum_{k=1}^p \mathbf{X}_k \boldsymbol{\gamma}_{ik} + \boldsymbol{\varepsilon}_i \\ &= \mathbf{W} \boldsymbol{\alpha}_i + \mathbf{X} \boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i. \end{aligned}$$

However, estimation of  $\boldsymbol{\alpha}_i$  and  $\boldsymbol{\gamma}_i$  will not be based on the common least squares criterion for two related reasons: First, the dimension  $p$  of basis function coefficients,  $\boldsymbol{\gamma}_{ik}, k=1, \dots, p$ , is comparably high to enhance flexibility of spline representation for the time-varying effects  $\beta_{ik}(t)$ .

Unrestricted least squares estimation would suffer from instability. Second, the time-varying effects are assumed to be smooth, excluding rough functions with abrupt, local changes. This requires that neighboring coefficients  $\boldsymbol{\gamma}_{ik,j}$  and  $\boldsymbol{\gamma}_{ik,j-1}$  in the spline representation (4) do not differ too much from each other. Both goals, reduction of the effective dimension and smoothness, can be achieved through penalized least squares estimation outlined in the following.

### Penalized Least Squares Estimation

The least squares criterion is replaced by the penalized least squares criterion

$$LS_{\text{pen}}(\boldsymbol{\alpha}_i, \boldsymbol{\gamma}_i) = (\mathbf{y}_i - \mathbf{W} \boldsymbol{\alpha}_i - \mathbf{X} \boldsymbol{\gamma}_i)' (\mathbf{y}_i - \mathbf{W} \boldsymbol{\alpha}_i - \mathbf{X} \boldsymbol{\gamma}_i) + \sum_{k=1}^p \lambda_k \boldsymbol{\gamma}_{ik}' \mathbf{P}_k \boldsymbol{\gamma}_{ik}$$

The smoothing parameters  $\lambda_k$  control the trade-off between the least squares criterion as a measure of goodness-of-fit and the quadratic roughness penalty term with penalty matrix  $\mathbf{P}_k$ . We follow the Penalized-spline approach of Eilers and Marx [1996], where  $\mathbf{P}_k$  is chosen such that the penalty is the sum of squares of first or second order differences of successive basis function coefficients. For first order differences, we have

$$\boldsymbol{\gamma}_{ik}' \mathbf{P}_k \boldsymbol{\gamma}_{ik} = \sum (\boldsymbol{\gamma}_{ik,j} - \boldsymbol{\gamma}_{ik,j-1})^2,$$

with  $\mathbf{P}_k = \mathbf{D}_1' \mathbf{D}_1$  and  $\mathbf{D}_1$  the matrix of first order differences.

For given smoothing parameters  $\lambda_k$ , the penalized least squares estimators  $\hat{\boldsymbol{\alpha}}_i, \hat{\boldsymbol{\gamma}}_i$  are obtained by minimizing  $LS_{\text{pen}}(\boldsymbol{\alpha}_i, \boldsymbol{\gamma}_i)$  with respect to  $\boldsymbol{\alpha}_i, \boldsymbol{\gamma}_i$ . The smoothing parameters,  $\lambda_k, k=1, \dots, p$ , can be estimated from the data through generalized cross-validation or restricted maximum likelihood. Furthermore, the (estimated) covariance matrices  $\widehat{\text{Cov}}(\hat{\boldsymbol{\alpha}}_i)$  and  $\widehat{\text{Cov}}(\hat{\boldsymbol{\gamma}}_i)$  are also available from the algorithm solving the penalized least squares minimization problem. More details are described in Ruppert et al. [2003], Wood [2006] and in Fahrmeir et al. [2013], and algorithms are implemented in the R-package `mgcv()`. P-splines assume only that the true underlying beta time course has no jumps and is differentiable, otherwise their shape can take arbitrary forms. To avoid oversmoothing in regions with very high curvature so-called adaptive splines [Krivobokova et al., 2008] may be applied as an extension of our proposed model.

Inserting the estimated basis coefficients  $\hat{\boldsymbol{\gamma}}_{ik}$  into the spline representation (4), estimated effect curves  $\hat{\beta}_{ik}(t)$  together with (pointwise) confidence bands with significance level  $\alpha$

$$\hat{\beta}_{ik}(t) \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\beta}_{ik}(t))}$$

are available. Note that,  $z_{1-\alpha/2}$  denotes the  $1-\alpha/2$  quantile of the standard Gaussian distribution. In addition, the

stimulus predictor as well as predictors for the fMRI signals  $y_{it}$  are available by plugging the stimulus effects  $\beta_{ik}$  into (5).

Simultaneous confidence bands are available by replacing the quantiles of the standard Gaussian distribution with the quantiles of the maximum absolute standardized deviation between the true effect curves and their estimates, see Ruppert et al. [2003, Ch. 6.5]. Since these quantiles are difficult to obtain analytically, even asymptotically, they are typically approximated through simulation, increasing computation times substantially.

Formal tests on the functional form of effect curves, for example, to decide whether an effect  $\beta_i(t)$  is really nonlinear or just linear, or constant as in model (2), or even zero, are still an area of active research, in particular for additive models which are closely related to time-varying coefficient models considered here. They are mostly based on (restricted) likelihood ratio tests derived from the mixed model representation of penalized splines. The null distribution of the test statistics is generally rather difficult to obtain, even asymptotically, see Greven and Crainiceanu [2013] and references given there. Therefore—as with simultaneous confidence bands—computation of (approximate)  $P$ -values is usually simulation-based, see Greven et al. [2008] and Scheipl et al. [2008]. Transferring these tests to varying coefficient models is conceptually easy, but we feel that more experience with artificial and real data is needed before relying on them in our framework. A particular issue is that for significance maps,  $P$ -values would have to be adjusted to take into account spatial correlation in a similar fashion as in SPM for time-constant activation effects.

Here, we are primarily interested in detecting specific patterns of stimulus effects, such as increasing or decreasing or non-monotonical effects, where no formal tests are available anyway. Therefore, we apply more exploratory tools, such as confidence bands or functional clustering (see Results section) to detect brain regions with such patterns.

## RESULTS

In this section, we present results obtained from an event-related fMRI dataset. The fMRI data were acquired on a clinical 3-Tesla scanner (General Electric MR750) from nine healthy male volunteers during an active two-tone-oddball paradigm [Kiehl et al., 2005]. In this paradigm, rare (high-pitched) odd tones (1500 Hz, duration 50 ms) appeared with 10% probability against the background of frequent (low-pitched) tones (1,000 Hz, duration 50 ms). The interstimulus interval was set to an average of 1,000 ms. The subjects were instructed to continuously pay attention to the tones and press the response button with their right index finger immediately after recognizing an odd (high-pitched) tone. Whole brain fMRI time series were acquired using an echoplanar imaging (EPI) sequence (time of repetition 2,000 ms, time of echo 40 ms, slice orientation according to anterior-commissure/posterior-commissure

landmarks, 28 slices, slice thickness 3.5 mm, 0.5 mm gap, in-plane resolution  $3.125 \times 3.125 \text{ mm}^2$ ) while the acoustic oddball paradigm was applied. A total of 307 image volumes per subject were recorded over 10.4 minutes with the first five images being disregarded due to unequilibrated T1 effect, leaving 302 images for the further analysis.

The dataset was postprocessed using the SPM software (<http://www.fil.ion.ucl.ac.uk/spm>, version SPM8). First, data were corrected for slice-timing effects caused by the bottom-up-interleaved acquisition scheme over 2 s. Second, motion correction was performed using rigid-body realignment with transformation parameters stored for each file. Third, images were spatially normalized using linear and nonlinear transformations to an EPI wholehead template in standard Montreal Neurological Institute (MNI) space with default settings of the SPM8 distribution. Intrinsic to the spatial normalization step is an interpolation step that was set to gain voxels sized  $3 \times 3 \times 3 \text{ mm}^3$ . Eventually, images underwent edge-preserving nonlinear spatial filtering (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/SUSAN>) with a kernel comparable to a 3D-Gaussian kernel sized  $6 \times 6 \times 6 \text{ mm}^3$  full width half maximum.

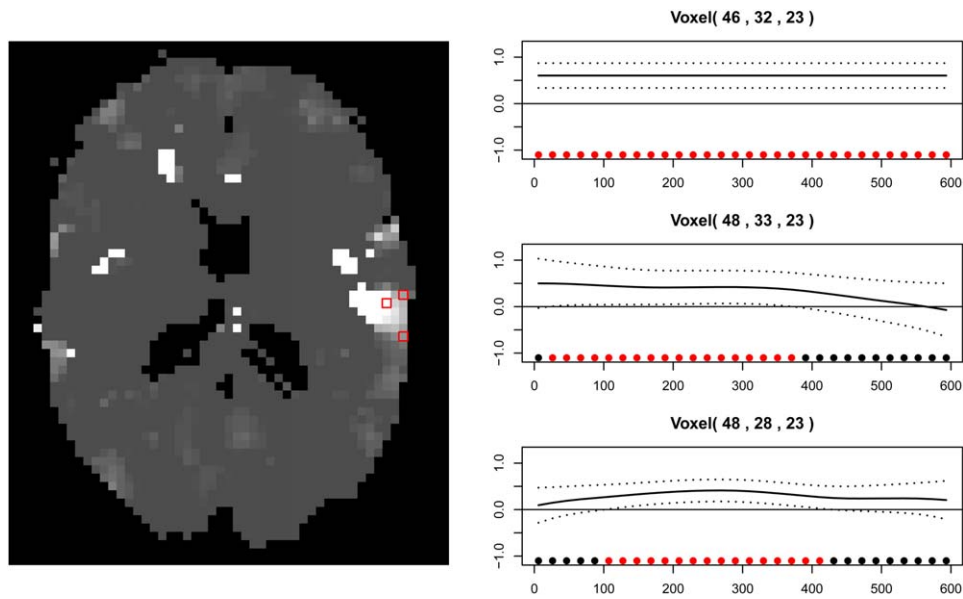
For the time-varying fMRI activation analysis, we set the highpass filter threshold for  $f_{\text{base}}(i, t)$  to a low value (20 s) to remove as many nonparadigm related fluctuations as possible. Additionally, a set of movement parameters from the rigid-body realignment were included as confounders to ensure that resulting activation effects are not motion-related.

In total, the voxelwise time-varying effects of  $M = 60$  presented odd stimuli (per subject) were estimated. For this, we chose a B-spline basis with  $q = 10$  basis functions to capture habituation and attention effects. Preruns with a larger number  $q$  of basis functions (not shown here) did not add value to the present analysis.

To detect voxels that are at least partly activated or deactivated over time, the following heuristic was used: Per voxel, an equally spaced grid  $T$  of 30 pointwise confidence intervals (cf. Penalized least squares estimation section) with significance level  $\alpha = 0.001$  was calculated. As a measure of activation strength over time, we then computed the relative frequency of points that do not cover zero

$$\kappa_i = \frac{|\{t \in T | 0 \notin CI(t)\}|}{|T|}, i = 1, \dots, N$$

where  $|\cdot|$  denotes the cardinality of the according set and  $\kappa$  is the corresponding map of voxelwise  $\kappa_i$ -values. In Figure 2, an exemplary  $\kappa$  slice is plotted together with the effect trajectories and confidence intervals for three selected voxels located temporally within the oddball network. The red dots denote confidence intervals that do not cover zero and black dots vice versa. The voxel trajectory on the top is located inside the activation focus. It has a strong constant effect, whereas  $\beta(t)$  curves of boundary voxel indicate time variation, that is, a decreasing effect for the middle time series (red voxel square on the right/top) and an inverse U-shaped effect for the time series on the bottom (red voxel square on the right/bottom).



**Figure 2.**

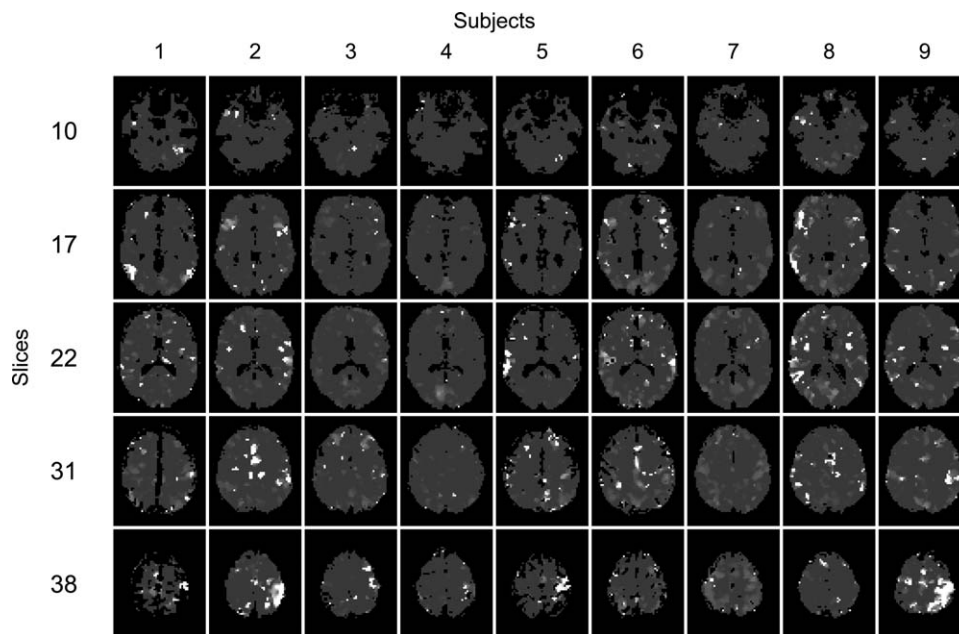
Exemplary  $\kappa$  slice plotted together with selected effect trajectories  $\beta(t)$  and confidence intervals of three voxels located temporally within the oddball network. The red dots denote pointwise confidence intervals that do not cover zero, and black dots vice versa. The corresponding gray-scale value in  $\kappa$  visualizes the relative frequency of red dots ranging from black ( $\kappa_i=0$ ) to white ( $\kappa_i=1$ ).

In Figure 3, selected  $\kappa$ -slices for all subjects of the sample are plotted to give an overview over individual activation patterns. Generally, we observe substantial variation between the individual activation maps. In six of nine cases, temporolateral, frontoinsular, and left motor cortex activation can be detected. In contrary, dorsal anterior cingulate activation is more heterogeneous over the group. Yet, there are subgroups that share common stimulus-response pattern, for example, in frontotemporal areas (subjects 2, 5, 6, 8, and 9). Closer inspection demonstrated that, for example, in subject 3, this pattern was evenly present, yet with more unstable or time-varying responses (dark gray).

In the following, we illustrate the gain in knowledge, which one can achieve by examining time-varying effect trajectories, on results from subject 2. For simplicity, we refer to methods relying on time-constant effect estimation as conventional. On the one hand, voxels with time-varying activation profiles that are not found by conventional activation detection methods (based on time-constant effect estimation) are of special interest. On the other hand, the form of time-varying effect profiles is worth to be examined. From Figure 4a, we observe the following: SPM activation foci (middle column) are rather small and contain voxels with time-varying effect profiles. Further, SPM-based activation foci are surrounded by light gray areas which suggests that in these areas, time-varying effects are too short-lived as to be detected by conventional methods. The activation foci found by iMRF (a

Bayesian activation detection model proposed by Kalus et al., 2014 utilizing a spatial regularization scheme based on an intrinsic Gaussian Markov Random field) are larger compared with SPM activation foci, likely by borrowing strength from spatial information. Hence, most iMRF activation foci mostly cover the corresponding time-varying effect foci, still, with less mismatch left compared with SPM (light gray areas). Note that there is an artifact in  $\kappa$ -maps that seems to be spurious at the first sight. There are clusters of voxels with reasonable sized, time-constant effect trajectories that are non-zero over the whole time range (e.g., cluster shown in slices 10 and 22) but still not detected by either of the conventional methods. One likely reason for this is that the statistical threshold criteria of the heuristic analysis method cannot be directly compared to the multiple test corrected and hence, more rigorous thresholds of the conventional procedures. In line with this, application of a more stringent threshold for the heuristic procedure would lead to the disappearance of these clusters (data not shown).

In Figure 5, the  $\beta(t)$ -trajectories of two ROIs sized  $4 \times 4$  voxels from a single subject (subject 2, as in Fig. 3) are depicted: one located in the attention network (see Fig. 4b, top) and one within the oddball network (see Fig. 4b, bottom). The zoom into the  $\beta(t)$ -trajectories reveals that SPM and iMRF reliably declare voxels as significant that display mostly constant effect profiles or time-varying effect profiles that are reasonable strong for most parts of the time window (dark orange voxels). The iMRF algorithm is



**Figure 3.**  
Selected  $\kappa$  slices for all examined subjects.

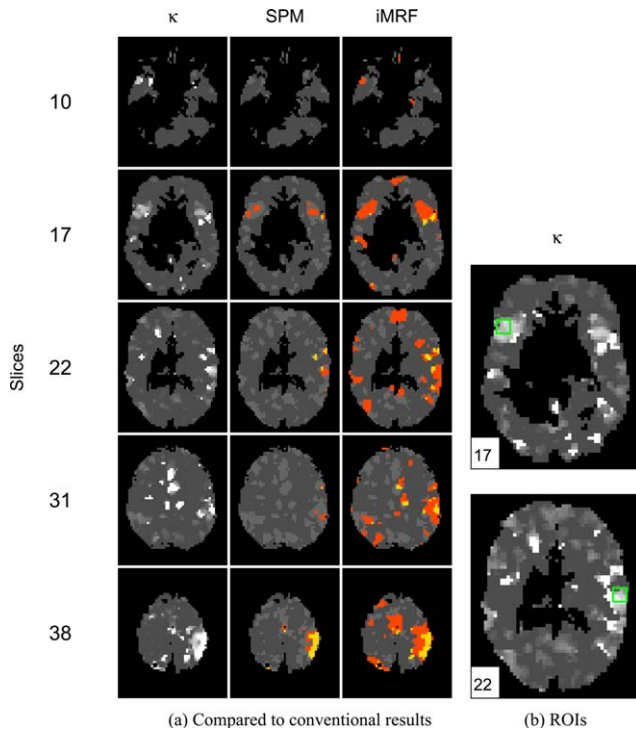
more sensitive in detecting effects that appear over a shorter time range (light orange voxels). Especially, SPM does not detect voxels with shorter but still strongly peaking activation as, for example, seen in voxel [44, 11] in Figure 5a. Both methods do not detect adjacent voxels that exhibit a substantial response at the beginning with steady decrease over time as, for example, seen in voxel [45, 10].

The display of a limited number of spatial clusters of functions on a map provides a convenient summary of results from a time-dependent fMRI analysis. The development of functional cluster algorithms, which cluster curves according to the similarity of their functional form in high-dimensional settings, is subject of current research. A basic approach, however, is the following [Abraham et al., 2003]: For each voxel-specific curve a basis function decomposition can be calculated. Then, given a prespecified number of clusters, the *K*-Means algorithm can be applied to the voxelwise basis function coefficients. These coefficients are directly available as  $\hat{\gamma}_i = \hat{\gamma}_{i,1}, i=1, \dots, N$  from Eq. (4) from our analysis. To decrease the computational burden, we preselected voxels with constant curves that have standard deviations that are numerically equal to zero. These preselected voxels were split into two clusters: one with  $\kappa_i$  values equal to one (“significant curves”) and another with  $\kappa_i$  values equal to zero (“nonsignificant curves”). Note, that  $\kappa_i$  has a binary outcome when curves are constant. Then, we clustered remaining  $\hat{\gamma}_i$ s requesting a solutions with five clusters, because this configuration resulted in interpretable and stable clusters—whereby stability was measured by the Rand-Index for cluster solu-

tions 2 to 12 [derived via Bootstrapping as described in Hubert and Arabie, 1985; Leisch, 2006]. It should be emphasized that, other than usually used in brain mapping reports, the term “cluster” as used refers to the result of the *K*-Means-clustering algorithm. That is, clusters are not necessarily spatially coherent but consist of areas disperse in anatomical space.

In Figure 6, the 7-cluster solution (2 preselected plus 5 *K*-Means clusters) is plotted with corresponding center curves and cluster sizes. C1 denotes the preselected cluster with constant significant curves and C7 denotes the cluster with constant nonsignificant curves. In addition to C1, C2 also contains voxels with constant activation over the whole analysis window, yet with some minor fluctuations. C1 and C2, when taken together, represent activations typically elicited by an active acoustic oddball experiment. The areas (light and dark yellow) comprise temporolateral areas (secondary acoustic cortex) and the cingulate cortex and bordering medial premotor cortex. In these areas effects are (approximately) time constant. Therefore, these areas largely match the pattern known from conventional SPM analyses. C7, as mentioned, and C6 represent areas of no activation. Of most interest for our methodology are clusters C3 to C5 that exhibit temporally variant activation patterns. A strong deactivation pattern, although not presented as enduringly as the activation patterns of C1 and C2, is seen in cluster C5 (dark blue) with a slow turn to activation later in the experiment. The areas of C5 match the default mode network (DMN) (comprising a posterior midline node, two parietal lateral nodes (slices 26–28), a





**Figure 4.**

Left: Single subject comparison of  $\kappa$  map (left column) with conventional time-constant activation detection methods (middle and right column for SPM and iMRF, respectively). In the left column,  $\kappa$  is plotted with the gray-white intensity indicating the degree of activation over time. The middle and right columns depict time-varying effects within areas of activation as detected by SPM and iMRF, respectively. Conventionally active voxels are depicted in hot colors, with red indicating partial nonactivation and orange indicating full activation. Light gray represents voxels with partial activation in the  $\kappa$  map but no activation in the respective conventional approach. Right: ROIs for further result presentations.

midline frontopolar node and further more lateral prefrontal areas) that typically deactivates in terms of amplitude during the performance of goal-directed tasks. Cluster C4 is particular in that it shows high fluctuations over the experiment (about one every 2 min) that are slower than typical resting state network activity [Auer, 2008], yet not representing simple up- or down-drifts. Also, its location in the posterior midline area of the DMN render it functionally neighbored to the DMN (see also Discussion section). Last, cluster C3 (red) exhibits a temporal course mostly mirroring that of C5, with activation at the beginning and a turn to deactivation later. Anatomically, it is a complex assembly of medial parietal, lateral parietal, medial and lateral prefrontal and insular areas, possibly representing a higher order frontoparietal control network, parts of a general task positive network [Fox et al., 2006] or inconstant (e.g., habituating) elements of the acoustic oddball network [Kiehl et al., 2005].

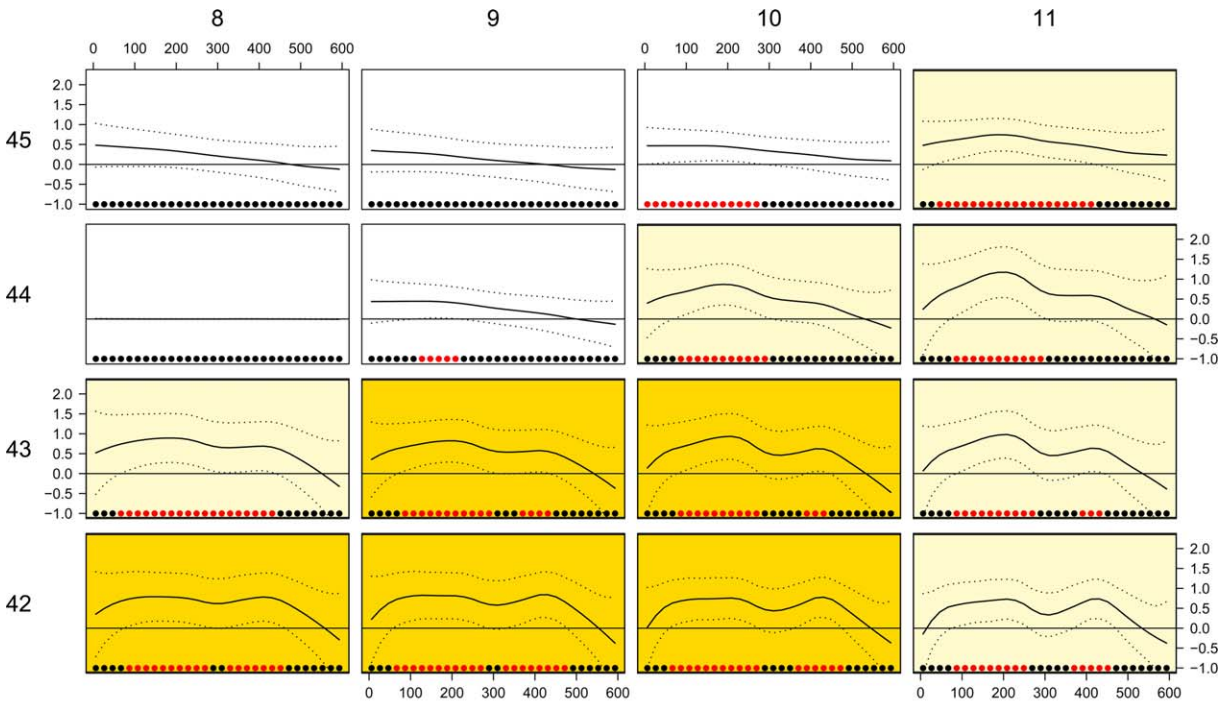
## DISCUSSION

Although we have seen that the proposed model is sufficiently sophisticated to reveal time-varying effects, the procedure can be renewed in several ways. For the time being, spatial dependencies are ignored. To borrow strength from the spatial neighborhood, the development of a Bayesian extension of the model that incorporates a suitable spatial prior seems conceivable. This spatial prior should enhance similarities of curves within one brain cluster, but it has to be ensured that it provides sufficient edge preserving properties as well. Hence, development of such a spatial model is not straightforward and an issue for future research.

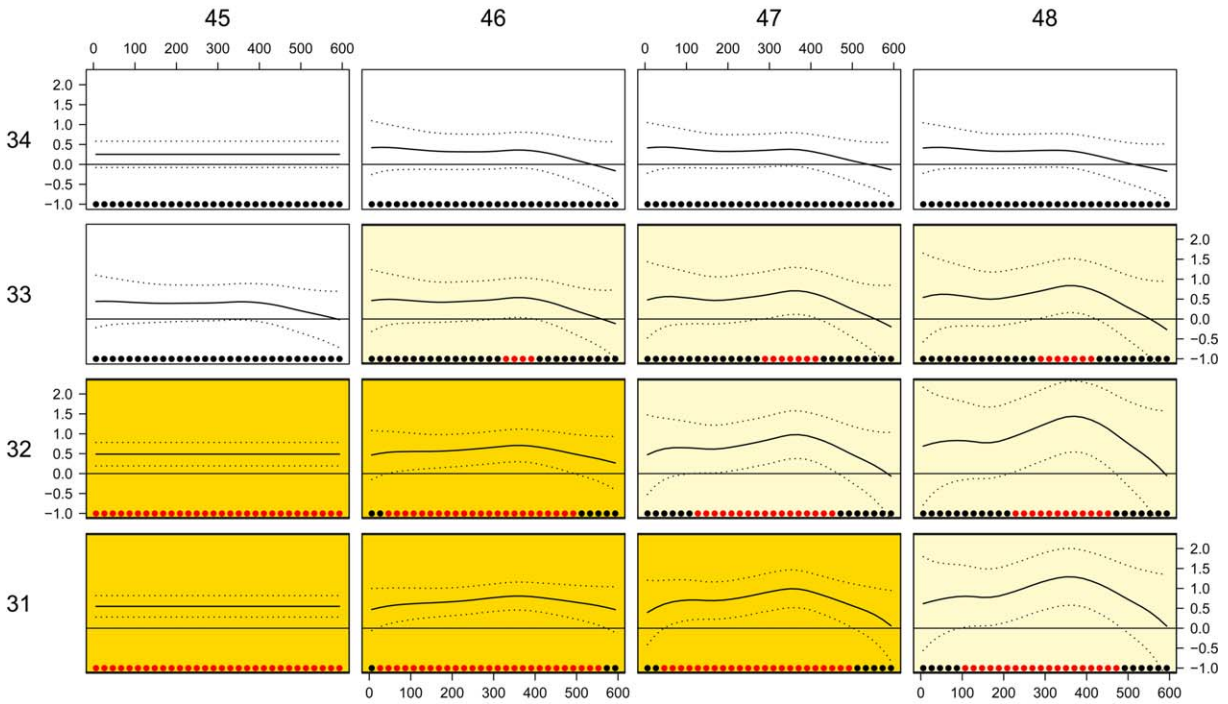
Beyond these options for an optimization of the estimation itself, the extraction of voxelwise time-varying coefficients as such is a valuable intermediate result level. This type of information on the course of the experiment is novel as current widely used GLM approaches still assume that the basis function coefficients are constant over the experiment—although offering different ways of modeling and estimating the HRF. As demonstrated, variation in the coefficients can be detected and may serve as input for different types of clustering algorithms. Here, we have exemplified that the coefficients, which contain the full information on the temporal development of the stimulus-associated HRF response, can be forwarded to a clustering algorithm. By this, groups of voxels can be identified that share a common dynamical pattern. Alternatively, similar to BOLD raw signals, spatial [Beckmann and Smith, 2004, 2005] or temporal [Smith et al., 2012] independent component analysis (ICA) in its extension to probabilistic group ICA may be utilized in this respect. These methods have proven to be helpful to extract anatomical areas that combine to functional units [Damoiseaux et al., 2006; Smith et al., 2012]. In contrast to standard models that focus on the mapping of a stimulus-specific response, the possibility to explore time-varying effects may, therefore, help to elucidate regulatory networks of processes that modulate stimulus response.

We have presented the time-varying effect curves in (5) through flexible spline functions defined in (4). A conceptually different strategy might be functional principal component analysis (FPCA), expanding the effect curves in (5) into an orthogonal series of eigenfunctions (the Karhunen–Loeve expansion) and extracting the main directions of time-variability of effect curves. FPCA has been recently developed as a framework for high-dimensional imaging data, see Greven et al. [2010], Zipunnikov et al. [2011], and Zipunnikov et al. [in press], and appears to be a promising alternative to explore time-varying activation effects.

Our approach can be extended to include time-varying effects for additional conditions as separate regressors: Time-constant effects in conventional regression models for this situation are replaced through time-varying effect curves based on penalized splines in the same way as in



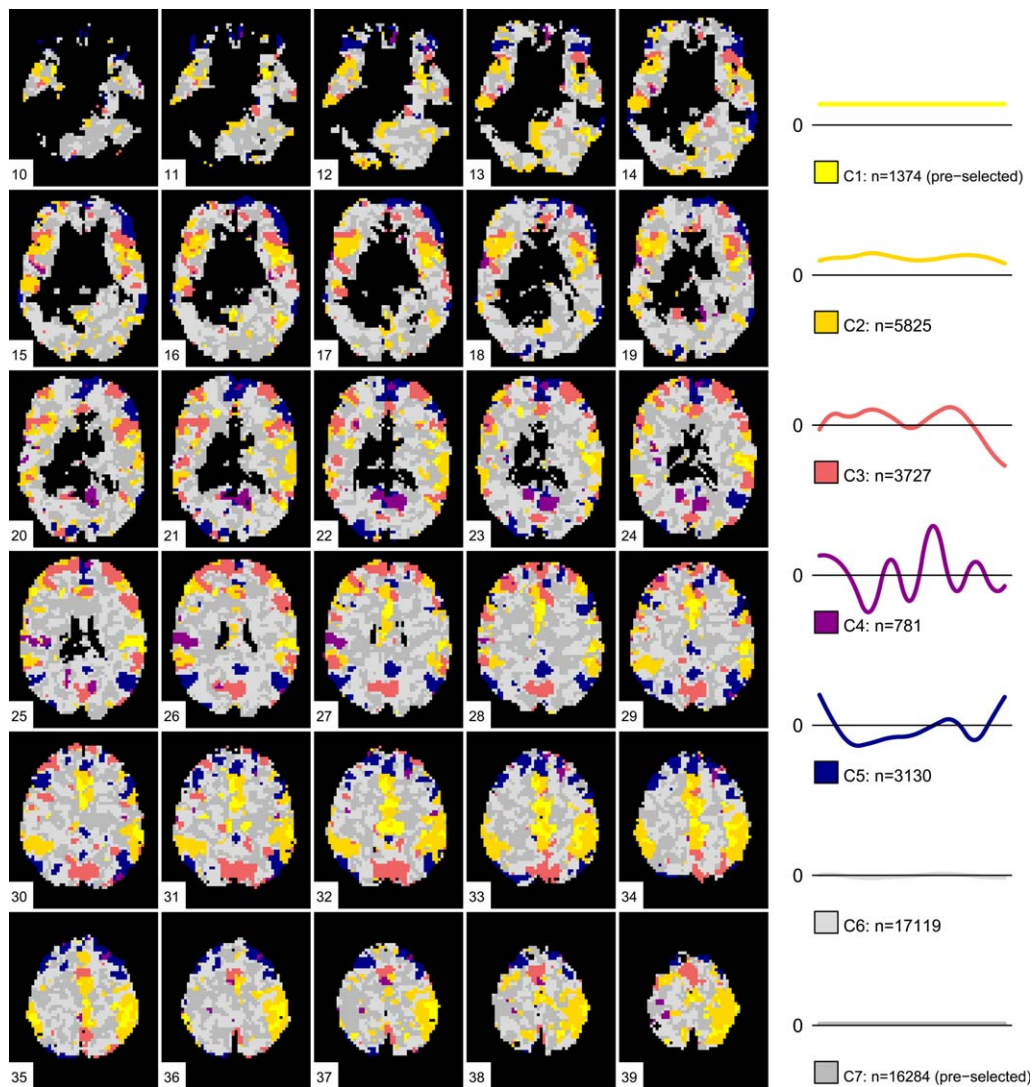
(a) Attention network



(b) Oddball network

**Figure 5.**

Selected effect trajectories for two different ROIs (cf. Fig. 4b) for subject 2. White backgrounds denote voxels that are not found to be activated by either of the two constant detection models, light orange backgrounds denote voxels that are found solely by iMRF, dark orange backgrounds denote voxels found by both iMRF and SPM. Numbers in the figure margins indicate the voxel IDs in x and y direction in one slice.



**Figure 6.**  
Results of functional clustering algorithm (slices 10–39).

our model for a single condition. However, it has to be checked whether multicollinearity issues are introduced by the experimental design. For example, in the presented auditory oddball design, odd and even tones present an almost regular series of stimuli. If modeled simultaneously, the odd and even regressors are highly collinear with the intercept. Hence, strategies for coping with this have to be applied.

A straightforward way to perform group-level analysis can follow SPMs approach of second level analysis to yield mean beta time series-estimates for a given group of subjects. However, it must be ensured that aggregation over a group is sensible at the first place. Time-variation profiles may possess a large intersubject variability so that mean profiles are not interpretable.

Besides this, it is of interest to identify factors that cause the variation in effect strengths. For the time variable that may represent a surrogate for other unobserved variables, putative causes for the variation can be examined by incorporating other types of varying effects, for example, vigilance levels as determined from parallel EEG-measurements. Then, effect strength variations can directly be associated with a hypothesized factor at the voxel level. Such incorporation of EEG-based varying coefficients is subject of current work [Bothmann, 2012]. One ideal application of such technique is pharmacological MRI in which modulatory effects of pharmacological substances on signal processing are expected [Jenkins, 2012]. Another application area is the clinical field in which habituation patterns to certain stimuli may be pathologically altered [Gordeev, 2008].

In our analyses, we have seen that the time-varying effect trajectories still contain a low amount of default state fluctuations that are not directly caused by the experimental paradigm and could not be regressed out by the conventional form of the highpass filter (although set to a very low value). Therefore, it is quite likely that some of the identified dynamical response patterns (see Fig. 6, K-Means clustering result) represent true modulations of the task response over time, whereas others may represent remaining effects of very low frequency components of resting state networks that would be equally detected in task-free data. As the DMN is deactivated during all kinds of cognitive stimuli and as it is also intrinsically anticorrelated to a task positive network during rest [Fox et al., 2005], both reactive patterns of the DMN and detection of spontaneous DMN patterns are well conceivable. Separating intrinsic activity from stimulus response is an important but complex problem, as the brain's responses to stimuli are not independent from ongoing, intrinsic activity. Statistical modeling of this issue, for example, based on P-splines or wavelets, is a challenging task for future research.

In conclusion, we present a new statistical model that allows for the identification of time-varying effects in event-related fMRI. Based on an acoustic oddball experiment, we observed that both stable and unstable, that is, temporally dynamic, response patterns can be separated. Direct comparisons with SPM demonstrate that the assumption of time-constant effects only applies for some of the typically retrieved activation maps. Consideration of time-varying effects may increase the sensitivity to detect individual response patterns, which is useful for applications in several domains of clinical neuroscience including pharmacological fMRI.

## ACKNOWLEDGMENTS

We are grateful to Sara A. Kiem for supporting data acquisition of the acoustic oddball paradigm. Remarks by the editorial board and two anonymous referees were extremely helpful in revising an earlier draft.

## REFERENCES

- Abraham C, Cornillon PA, Matzner-Løber E, Molinari N (2003): Unsupervised curve clustering using B-splines. *Scand Stat Theory Appl* 30:581–595.
- Auer DP (2008): Spontaneous low-frequency blood oxygenation level-dependent fluctuations and functional connectivity analysis of the 'resting' brain. *Magn Reson Imaging* 26:1055–1064.
- Badillo S, Vincent T, Ciuciu P (2013): Group-level impacts of within- and between-subject hemodynamic variability in fMRI. *NeuroImage* 82:433–448.
- Beckmann C, Smith S (2004): Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging* 23:137–152.
- Beckmann CF, Smith SM (2005): Tensorial extensions of independent component analysis for multisubject FMRI analysis. *NeuroImage* 25:294–311.
- Bénar C, Schön D, Grimault S, Nazarian B, Burle B, Roth M, Badier J, Marquis P, Liegeois-Chauvel C, Anton, J (2007): Single-trial analysis of oddball event-related potentials in simultaneous EEG-fMRI. *Hum Brain Mapp* 28:602–613.
- Bothmann L (2012): Statistische Modellierung von EEG-abhängigen Stimuluseffekten in der fMRT-Analyse. Diploma thesis. Munich: Ludwig-Maximilians-University.
- Büchel C, Holmes A, Rees G, Friston K (1998): Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *NeuroImage* 8:140–148.
- Coste C, Sadaghiani S, Friston K, Kleinschmidt A (2011): Ongoing brain activity fluctuations directly account for intertrial and indirectly for intersubject variability in stroop task performance. *Cereb Cortex* 21:2612–2619.
- Damoiseaux J, Rombouts S, Barkhof F, Scheltens P, Stam C, Smith S, Beckmann C (2006): Consistent resting-state networks across healthy subjects. *Proc Natl Acad Sci USA* 103:13848–13853.
- Descamps B, Vandemaële P, Reyngoudt H, Deblaere K, Leybaert L, Paemeleire K, Achten E (2012): Quantifying hemodynamic refractory BOLD effects in normal subjects at the single-subject level using an inverse logit fitting procedure. *Magn Reson Imaging* 35:723–730.
- Eilers PHC, Marx BD (1996): Flexible smoothing with B-splines and penalties. *Stat Sci* 11:89–121.
- Fahrmeir L, Kneib T, Lang S, Marx BD (2013): *Regression*. Springer, Berlin Heidelberg.
- Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005): The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci USA* 102:9673–9678.
- Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME (2006): Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc Natl Acad Sci USA* 103:10046–10051.
- Friston K, Holmes A, Worsley K, Poline J, Frith C, Frackowiak R (1995): Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp* 2:189–210.
- Friston KJ, Josephs O, Rees G, Turner R (1998): Nonlinear event-related responses in fMRI. *Magn Reson Med* 39:41–52.
- Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD (2008): *Statistical parametric mapping—the analysis of functional brain images*. Academic Press, London.
- Gordeev SA (2008): Clinical-psychophysiological studies of patients with panic attacks with and without agoraphobic disorders. *Neurosci Behav Physiol* 38:633–637.
- Gössl C, Auer DP, Fahrmeir L (2000): Dynamic models in fMRI. *Magn Reson Med* 43:72–81.
- Greven S, Crainiceanu C (2013): On likelihood ratio testing for penalized splines. *AStA Adv Stat Anal* 97:387–402.
- Greven S, Crainiceanu C, Küchenhoff H, Peters A (2008): Restricted likelihood ratio testing for zero variance components in linear mixed models. *J Comput Graph Stat* 17:870–891.
- Greven S, Crainiceanu C, Caffo B (2010): Longitudinal functional principal component analysis. *Electron J Stat* 4:1022–1054.
- Grindband J, Wager TD, Lindquist M, Ferrera VP, Hirsch J (2008): Detection of time-varying signals in event-related fMRI designs. *Neuroimage* 43:509–520.
- Henson R, Rugg MD, Friston KJ (2001): The choice of basis functions in event-related fMRI. *NeuroImage* 13:149.
- Hubert L, Arabie P (1985): Comparing partitions. *J Classifi* 2:193–218.
- Hutchinson R, Niculescu RS, Keller TA, Rustandi I, Mitchell TM (2009): Modeling fMRI data generated by overlapping

- cognitive processes with unknown onsets using hidden process models. *NeuroImage* 46:87–104.
- Ivey R, Schmidt H (1993): P300 response: Habituation. *J Am Acad Audiol* 4:182–188.
- Jenkins B (2012): Pharmacologic magnetic resonance imaging (phMRI): Imaging drug action in the brain. *NeuroImage* 62:1072–1085.
- Josephs O, Turner R, Friston KJ (1997): Event-related fMRI. *Hum Brain Mapp* 5:243–248.
- Kalus S, Sämann PG, Fahrmeir L (2014): Classification of brain activation via spatial Bayesian variable selection in fMRI regression. *Adv Data Anal Classif* 8:63–83.
- Kay KN, David SV, Prenger RJ, Hansen KA, Gallant JL (2008): Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. *Hum Brain Mapp* 29:142–156.
- Kiehl K, Liddle P (2003): Reproducibility of the hemodynamic response to auditory oddball stimuli: A six-week test-retest study. *Hum Brain Mapp* 18:42–52.
- Kiehl KA, Stevens MC, Laurens KR, Pearlson G, Calhoun VD, Liddle PF (2005): An adaptive reflexive processing model of neurocognitive function: Supporting evidence from a large scale (n=100) fMRI study of an auditory oddball task. *NeuroImage* 25:899–915.
- Koelega HS, Verbaten M, van Leeuwen TH, Kenemans JL, Kemner C, Sjouw W (1992): Time effects on event-related brain potentials and vigilance performance. *Biol Psychol* 34:59–86.
- Krivobokova T, Crainiceanu C, Kauermann G, (2008): Fast adaptive penalized splines. *J Comput Graph Stat* 17:1–20.
- Lammers W, Badia P (1989): Habituation of p300 to target stimuli. *Physiol Behav* 45:595–601.
- Leisch F (2006): A toolbox for k-centroids cluster analysis. *Comput Stat Data Anal* 51:526–544.
- Lindquist M, Wager T (2007): Validity and power in hemodynamic response modeling: A comparison study and a new approach. *Hum Brain Mapp* 28:764–784.
- Lindquist M, Loh JM, Atlas LY, Wager, TD (2009): Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage* 45:S187–S198.
- Makni S, Idier J, Vincent T, Thirion B, Dehaene-Lambertz G, Ciuciu P (2008): A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *NeuroImage* 41:941–969.
- Mantini D, Corbetta M, Perrucci MG, Romani GL, Del Gratta C (2009): Large-scale brain networks account for sustained and transient activity during target detection. *NeuroImage* 44:265–274.
- Monti M (2011): Statistical analysis of fMRI time-series: A critical review of the GLM approach. *Front Hum Neurosci* 5(28):1–13.
- Morrot G, Bonny JM, Lehallier B, Zanca M (2013): fMRI of human olfaction at the individual level: Interindividual variability. *Magn Reson Imaging* 37:92–100.
- Mumford J, Turner, BO, Ashby FG, Poldrack, RA (2012): Deconvolving BOLD activation in event-related designs for multi-voxel pattern classification analyses. *NeuroImage* 59:2636–2643.
- Ollinger JM, Shulman GL, Corbetta M (2001): Separating processes within a trial in event-related functional fMRI. *NeuroImage* 13:210–217.
- R Core Team (2013): R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
- Rodriguez PF (2010): Using conditional maximization to determine hyperparameters in model-based fMRI. *NeuroImage* 50:472–478.
- Ruppert D, Wand M, Carroll R (2003): Semiparametric Regression. Cambridge: Cambridge University Press.
- Sadaghiani S, Hesselmann G, Kleinschmidt A (2009): Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection. *J Neurosci* 29:13410–13417.
- Scheipl F, Greven S, Küchenhoff H (2008): Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Comput Stat Data Anal* 52:3283–3299.
- Shen Y, Mayhew SD, Kourtzi Z, Tino P (2014): Spatial-temporal modeling of fMRI data through spatially regularized mixture of hidden process models. *NeuroImage* 84:657–671.
- Smith SM, Miller KL, Moeller S, Xu J, Auerbach EJ, Woolrich MW, Beckmann CF, Jenkinson M, Andersson J, Glasser MF, Van Essen DC, Feinberg DA, Yacoub ES, Ugurbil K (2012): Temporally-independent functional modes of spontaneous brain activity. *Proc Natl Acad Sci USA* 109:3131–3136.
- Wesensten N, Badia P, Harsh J (1990): Time of day, repeated testing, and interblock interval effects on p300 amplitude. *Physiol Behav* 47:653–658.
- Wood SN (2006): Generalized Additive Models: An Introduction with R. Chapman & Hall, Boca Raton, Florida.
- Worsley KJ, Friston KJ (1995): Analysis of time-series revisited—Again. *NeuroImage* 2:173–181.
- Zhang C, Lu Y, Johnstone T, Oakes T, Davidson RJ (2008): Efficient modeling and inference for event-related fMRI data. *Comput Stat Data Anal* 52:4859–4871.
- Zhang T, Li F, Beckes L, Brown C, Coan JA (2012): Nonparametric inference of the hemodynamic response using multi-subject fMRI data. *NeuroImage* 63:1754–1765.
- Zipunnikov V, Caffo B, Yousem D, Davatzikos C, Schwartz B, Crainiceanu C (2011): Functional principal component analysis for high-dimensional brain imaging. *Neuroimage* 58:772–784.
- Zipunnikov V, Greven S, Shou H, Caffo B, Reich D, Crainiceanu C: Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *Annals of Applied Statistics* (in press).