Goessl, Auer, Fahrmeir:

Dynamic models in fMRI

Sonderforschungsbereich 386, Paper 136 (1998)

Online unter: http://epub.ub.uni-muenchen.de/

Dynamic models in fMRI

C.Gössl[1], D.P.Auer[1], L.Fahrmeir[2]

[1]Max-Planck-Institute of Psychiatry, Munich

[2]Institute of Statistics, University of Munich

Address for correspondence:

Christoff Gössl

Max-Planck-Institut für Psychiatrie

Kraepelinstr. 10

80804 München

Germany

Tel.: ++49 89 30 622 359

Fax: ++49 89 30 622 520

email: goessl@mpipsykl.mpg.de

Key words: fMRI, state space models, Kalman filtering, time varying activation

Abstract

Most statistical methods for assessing activated voxels in fMRI experiments are based on correlation or regression analysis. In this context the main assumptions are that the baseline can be described by a few known basis-functions or variables and that the effect of the stimulus, i.e. the activation, stays constant over time. As these assumptions are in many cases neither necessary nor correct, a new dynamic approach that does not depend on those suppositions will be presented. This allows for simultaneous nonparametric estimation of the baseline as well as the time-varying effect of stimulation.

This method of estimating the stimulus related areas of the brain furthermore provides the possibility of an analysis of the temporal and spatial development of the activation within an fMRI-experiment.

1 Introduction

In the analysis of functional MRI time series the main goal is the assessment of stimulus-related activated areas of the brain on the basis of a stimulus-induced MR-signal variation. This is equivalent to the detection of a dependency between the time-courses of the measured MR signal of the voxels in the brain, and a stimulus, that has been presented to the subject. For this purpose the most commonly used techniques are correlation methods (1) or linear regression analysis (2). But even if these methods are well known and established, due to a huge number of noise sources the analysis appears easier than it actually is. Whereas the physical noise produced by the scanner (e.g. Quantum noise, thermal noise) and the environmental sources can usually be described as a stationary white-noise process, the characterisation of other sources is far more difficult. Respiratory and cardiac cycles as well as other kinds of physiological noise are well known non-stationary confounds, which have to be considered. A common measure to do this is a detrending of the signal with a few polynomial or trigonometric basis-functions (3, 4). But this imposes two questions: up to which order of the basis-functions should one detrend and what is the interpretation of the estimated elements of the trend? Another difficulty lies in the principle of the above methods. They describe in a simple parametric way the relation between the signal and a reference stimulus. The latter is assumed to be similar to the time-course of the evoked activation and to be nearly always identical for all voxels in the brain. The estimated parameters are then, after standardisation and thresholding at a certain value, displayed in a so-called activation map. This should show the regions, where the stimulus has a significant influence on the MR signal thought to be activation-related via the BOLD mechanism. But there are no apparent reasons for the evoked activation, and consequently for the temporal delay or the shape of the signals increase, to be identical over the entire brain (3). Also, the assumption that one single parameter can describe the stimulus-related effect in a voxel is questionable. In many cases it can be observed, that the amplitude of the activation varies quite strongly and unpredictably between the stimulation periods. Here an adequate description with a prespecified reference function and a single temporally constant parameter is nearly impossible. But, ignoring this variability

leads to biassed estimates of the parameters and activation maps.

Even though these facts are well known, the insufficiencies of the existing methods have usually been neglected until now. The aim of our studies was to evaluate a new approach that can handle these problems in a very intuitive and elegant way and thus, provide an appropriate description of the measured time-series. By means of a dynamic model we present a nonparametric method that works without introducing any new covariables or basis-functions or constraining the parameters to any prespecified parametric form. Furthermore a solution for omitting the restrictions applied to the lag and shape parameters will be proposed, before some examples will illustrate this approach.

2 Theory

The Model

In the following we look at a familiar fMRI experiment. A certain number T of MR images $Y^t, t = 1, \ldots, T$ was obtained from a subject during the presentation of any kind of ON-OFF-Stimulus X (e.g. 30s of rest, 30s stimulation, 30s rest, ...). Each of these T images consists of I pixels or voxels, for which a whole time-series $Y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})', i = 1, \ldots, I$ of measured MR signals exists. Aim of the analysis is to assess those voxels, whose time-series show a significant connection to the presented stimulus.
Without any loss of generality, our considerations start from an ordinary linear regression model, which under certain conditions is transformable to a correlation model and can be formulated as follows:

$$Y_i = Wa_i + Zb_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_i^2 I), \quad i = 1, \ldots, I, \qquad [1]$$

or in componentwise notation

$$y_{it} = w_t' a_i + z_t b_i + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma_i^2), \quad i = 1, \ldots, I.$$

This model describes the signal Y_i at voxel i as a linear combination of the design matrices W and Z superimposed by a white-noise error term ϵ_i. In this context W denotes a known

vector or matrix, which is supposed to model the trend or the baseline drift. Matrices that contain linear and quadratic trends or the first few terms of a Fourier expansion are common examples. The variable Z denotes the transformed stimulus, in other words Z is a function of the presented ON-OFF-stimulus X. With regard to the transformation, we consider a temporal shift of the original stimulus by a time-delay d and a convolution with a parametric hemodynamic response function (HRF) h, so that:

$$z_t = \sum_{s=0}^{t-d} h(s,\theta) x_{t-d-s}. \qquad [2]$$

In general Poisson ($\mathrm{Po}(\lambda)$) or Gamma ($\mathrm{Ga}(\lambda, u)$) densities are chosen for this purpose. These transformations model that (a) due to hemodynamic latencies the cerebral blood-flow (CBF), the source of the MR signal, increases approximately 6-8s after the onset of the stimulus, and that (b) the flow responses do not occur suddenly, but rather continuously and delayed. The question, how the parameters d and θ can be estimated will be addressed later. At this point, we suppose that they are known. Finally, the unknown parameters a_i and b_i are a measure for the influence of the covariables W and Z on the MR signal.
After having estimated a_i and b_i, one can test whether b_i is zero or not, i.e. whether Z, respectively X, has an influence on Y. This yields the familiar activation maps, images of the standardised test statistics of all voxels thresholded at a certain value, that determines the p-value.

As mentioned above, we believe that some assumptions made in the above model are incorrect in many cases. Therefore, an alternative approach, based on the theory of dynamic linear models, which works without these suppositions, will be proposed in the following. As it is quite difficult to formulate an appropriate trend modelling matrix W for often highly nonlinear trends in the baselines of the MR-signals, in a first step the arbitrarily chosen matrix W will be omitted. We do not introduce any new artificial covariables, but we allow the parameter $a_i = (a_{i1}, \ldots, a_{iT})$ to be a function of time to describe the baseline directly without specifying any particular parametric form for it. This is achieved by modelling a_{it} nonparametrically. The only applied restriction is that a_{it} and thus the baseline should vary smoothly. In dynamic or state space models this is guaranteed by

imposing a so called random walk smoothness prior on a_{it}. For a single observation we can formulate this extension as follows:

$$y_{it} = a_{it} + z_t b_i + \epsilon_{it}, \qquad \epsilon_{it} \sim N(0, \sigma_i^2), \tag{3}$$

$$\text{with} \qquad a_{it} = 2a_{it-1} - a_{it-2} + \zeta_{it}, \quad \zeta_{it} \sim N(0, \sigma_{\zeta_i}^2).$$

The second term can be regarded as a kind of penalty term, which penalises all kinds of variation in the parameter a_i that deviate from a linear trend. For example, for a straight line this penalty is zero. So, a trade-off has to be found between data fit and smoothness of the curve. It should be mentioned, that in the estimation algorithm, which we will discuss later, this compromise is achieved entirely data-driven.

In a next step, the temporal variability in the amplitude of the activation should also be considered, i.e. the assumption that a single parameter can describe the whole fMRI time series is dropped. This effect will also be modelled nonparametrically. This means that we also allow the parameter b_i to vary over time, implying that the transformed stimulus Z may have a variable influence on the processes in a particular voxel. As for the parameter a_{it}, the only condition imposed is that b_{it} should vary smoothly. Thus, again a random walk prior on the parameter is assumed and the result is the new model:

$$y_{it} = a_{it} + z_t b_{it} + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma_i^2), \tag{4}$$

$$\text{with} \qquad \begin{aligned} a_{it} &= 2a_{it-1} - a_{it-2} + \zeta_{it}, \quad \zeta_{it} \sim N(0, \sigma_{\zeta_i}^2). \\ b_{it} &= 2b_{it-1} - b_{it-2} + \eta_{it}, \quad \eta_{it} \sim N(0, \sigma_{\eta_i}^2). \end{aligned} \tag{5}$$

At a first glance this extension might look a bit confusing, because it is not straightforward how to handle this parameter. But the underlying idea is quite simple and intuitive: When the amplitude of the response to a constant stimulus varies, then the parameter, which reflects the dependency between these two variables, should vary correspondingly. If the response to the stimulus becomes weaker or disappears altogether with time, then this particular parameter should be able to indicate this. An elegant way to do this (without prespecifying the temporal behaviour via a parametric form), is to have as many parameters as observations. These parameters show for every point in time the actual strength of the influence of the stimulus. As a consequence of this fact and because in this model the

estimated parameters $\hat{b}_{it}$ are distributed normally, doing this analysis for all voxels, one can calculate an activation map for every image or time point. This parameter map then shows the regions of the brain that are 'activated' (i.e. significantly stimulus-related) at this particular point in time. This is an important aspect of this model. To our knowledge it provides for the first time the possibility to investigate the temporal behaviour of a stimulus-induced signal variation within an fMRI experiment. Through the sequential display of all T activation maps the temporal and spatial evolution of the cortical activation during the experimental procedure can be studied.

As it is sometimes useful or even necessary to get only a single activation map out of an experiment, defining $b_{i0} = b_{i1}$ and $\sigma_\eta^2 = 0$ in the above model results in the familiar approach with a temporally constant parameter b_i, so that this case also can be considered without leaving this particular framework. Another possibility is to take the semiparametric model (3) and to apply an estimation algorithm for mixed models as proposed in Biller (5).

The last assumption to be omitted, is the one of an identical reference function for the whole brain. As well the lag and the shape of the activation should be allowed to differ between the voxels and therefore, pixelwise parameters d_i and θ_i for the stimulus transformations are introduced. Thus, the final model can be formulated as follows:

$$y_{it} = a_{it} + z_{it} b_{it} + \epsilon_{it}, \qquad [6]$$

with the error distributions and priors of the parameters a and b as defined in Eq[5].

Estimation

This section will give a brief description of the method of estimation in the presented model. Because most algorithms used for the above model have already been described very well and detailed in other publications, we refer for a more comprehensive introduction to Fahrmeir/Tutz (6), Harvey (7), De Jong (8) and Press et al. (9).

The estimation of the parameters is carried out in two major steps. The first, a more technical one, is calculating pixelwise pilot-estimates of the lag and shape parameters d_i and θ_i, while the second one yields the estimates of the model parameters and variances.

For the pilot-estimate a modified Gauss-Newton algorithm is used, which for each voxel minimises the quadratic distance between the MR signal Y_i and a reference function $z(d_i, \theta_i)$ as defined in Eq.[2]. As hemodynamic response function a Poisson density $\mathrm{Po}(\lambda)$ is assumed, so that $\theta_i = \lambda_i$ and

$$z_{it} = \sum_{s=0}^{t-d_i} \frac{\lambda_i^s}{s!} e^{\lambda_i} x_{t-d_i-s}.$$

We adapted the original algorithm (9) by first applying the ordinary algorithm for the parameter λ_i, and then heuristically minimising the resulting sum of squares (SSQ) after each iteration with regard to the discrete lag d_i. This was done by comparing the SSQs for the current value of d_i and its direct neighbours $d_i - 1, d_i + 1$. An upper bound for d_i is defined naturally by the beginning of the next stimulation period. Moreover for λ_i a limit was set to the value of 15s, based on plausibility considerations and to keep the computational effort acceptable. Once the parameters λ or d reached these bounds, they were set equal to them for the further analysis.

Two alternative ways of estimating the model parameters could be considered. First, Markov chain Monte Carlo (MCMC) methods. That means to sample from the posterior distribution of the parameters, which can be calculated as a combination of the distribution of the measured data for a given set of parameters and the prior distribution of the parameters according to Bayes' Theorem. Due to the huge amount of data this would be a very time-consuming procedure. The second possibility is a linear Kalman filter and smoother. Under the assumptions that the error terms are normal it also estimates the posterior mean together with the standard deviations and is computationally more efficient. Therefore, we decided on the filter. To apply the latter, the model only has to be transformed into state space form:

$$y_{it} = v_t \alpha_{it} + \epsilon_{it}, \qquad \epsilon_{it} \sim N(0, \sigma_i^2), \qquad [7]$$

$$\alpha_{it} = F \alpha_{it-1} + \zeta_{it}^*, \qquad \zeta_{it}^* \sim N(0, Q), \qquad [8]$$

$$\text{with } v_t = (1, 0, z_{it}, 0), \ \alpha_t = (a_{it}, a_{it-1}, b_{it}, b_{it-1})',$$

and

$$F = \begin{pmatrix} 2 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; Q_i = \begin{pmatrix} \sigma^2_{\zeta_i} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{\eta_i} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The filter estimates in an recursive procedure the means and the variances of the posterior distribution of the parameters α_{it}, the effects we are mainly interested in. To minimise the influence of arbitrarily chosen starting values, an extended version of the Kalman filter and smoother with diffuse initial value priors can be used, as proposed in De Jong (6). Further, to consider the unknown variances σ_i^2 and Q_i, the filter is embedded in an EM-Type algorithm (6) for estimating the unknown hyperparameters. This algorithm is of an iterative structure. In a first step the conditional expectation of the log-likelihood for a given set of observations and hyperparameters is built and then in a second one maximised with regard to these hyperparameters. This procedure is repeated until convergence. For the presented approach the expectation step is carried out via the Kalman filter and smoother and the maximisation can be solved analytically. For a detailed outline of the applied methods we refer the interested reader to Fahrmeir/Tutz (4) and Harvey (5).
As mentioned already the algorithm finds a trade-off between data fit and smoothness of the curves of the fitted effects. This compromise is mainly controlled by the relation of the variances of the observation model (Eq.[7]) and the transition model (Eq.[8]). Consequently, because the EM algorithm estimates the unknown variances on the basis of the data, it performs this trade-off completely data-driven simultaneously to the estimation of the time-varying effects. A supervision by the user is not necessary.

3 F-MRI data analysis

The Data

The data sets to illustrate the above method were fMRI time series of photic and acoustic stimulation experiments. They were acquired on a 1.5 T system (Echospeed, GE Medical Systems, Milwaukee) using a single shot Gradient Echo EPI sequence with echo and

repetition times of 60 ms and 3000 ms, respectively. Seven slices parallel to the intercommissural line with a nominal voxel size of 2.2 × 2.2 × 5 mm were positioned to cover the occipital lobes in the visual and the temporal lobes in the acoustic stimulation. 73 images were acquired with the initial three images being discarded to avoid non-steady-state effects. With regard to the stimulation paradigm, the subsequent 70 images were divided into four rest and three activation periods, with each period consisting of 10 images (30s long). During the visual stimulation periods a 32 × 24 element rectangular checkerboard that alternated at a frequency of 4 Hz was displayed, whereas a uniformly dark screen with a small fixation point in the centre was displayed in the rest periods. The stimulation pattern was projected into the magnet room through the observation window using an LCD projector. For acoustic stimulation, a tape recorded novel read by a professional speaker was presented via headphones that also reduced scanner noise in the rest periods. Data were transferred to a separate workstation (Sparc 20, Sun Microsystems, CA) for processing and calculation. For the data of the acoustic experiment a 3D motion correction was performed using version 3.08 of the public domain image registration program AIR (10, 11).

The algorithm

The starting values for the estimation of the pixelwise lag and shape parameters were set to $d = 6s$ and $\lambda = 6s$, the convergence criterion was a relative change of λ smaller than 0.001 in subsequent iterations.

The variances of the observation and transition model at the beginning of the EM-Type algorithm were set to 100 and 1. The convergence criterion for the variance of the observation model was chosen as above.

A diffuse Kalman filter and smoother was used and the time series were mean corrected prior to the analysis. For the activation maps, the estimated means of the posterior distribution of the effects were taken as point estimators. They were normalised to unit variance and then tested to be non-zero. This was simply carried out by thresholding the parametric map at a certain value, without defining a particular p-value. We renounced the latter, due to the non-existence of a sufficient solution of this highly complex problem.

The arbitrary threshold was set to the value of 3.5, which seemed to be an appropriate value to suppress most of the noise artefacts on the one hand and on the other not to be too restrictive with respect to the 'activated' voxels. At this point we did not apply any kind of spatial analysis, for not biassing the results. These aspects will be discussed later.

Results

Figure (1) shows the time series of a voxel from the visual cortex and its estimated effects. In the upper row the most simple regression model is displayed with an intercept term in the second column and a fixed effect for the boxcar reference stimulus in the third. The resulting fitted curve is superimposed on the data. The residuals are plotted in the last column. The second row displays an approach with the nonparametrically modelled baseline (Eq.[3]) but still a fixed effect for the pixelwise estimate of the reference function. In the last row the effects of the fully dynamic model (Eq.[6]) are plotted. Here, for illustration, the estimated dynamic effect of the stimulus was multiplied with the reference function.

As it is indicated through the residual plots, in both models with a single fixed parameter for the influence of the stimulus, an adequate fit to the data is not given. The only model that provides an appropriate description of the MR-signal is the completely nonparametric approach (Eq.[6]).

It can be seen that the dynamic effect varies quite strongly over time. This means that for a single voxel the significance of the stimulus influence can change within the experiment. As mentioned above the nonparametric approach generates as many parameters as there are observations for one effect. Doing the same analysis for all voxels of the brain, standardising and thresholding the estimated stimulus effects results in T parametric maps which display for every time-point all voxels that at this time exceed the threshold. One would expect that these maps reflect the temporal variation. Three activation maps taken from each stimulation period at 9s, 15s, and 21s after stimulus onset clearly show (Fig. 2a) a major variability of the extent of activation between periods especially in extrastriatal areas. The anatomical location of detected activation is fairly comparable to standard correlation maps (Fig. 2b). Compared to the well defined simple photic stimulus, presentation of a

spoken text also implies cognitive components that may furthermore fluctuate over time depending on the story and the subject's interest. The same nine representative maps (Fig. 3a) taken from fixed time points within the stimulation periods illustrate a remarkable variation of activation in frontal cortex, right thalamus and secondary acoustic cortex including posterior language areas, while the extent of activation in the primary acoustic cortex stays rather constant. The single correlation map (Fig. 3b) shows similar anatomical activation, but time resolution proves the transient nature of thalamic activation which does not occur in the third stimulation period.
Furthermore, it could be stated that the motion correction had no influence on the spatial variability in these and 10 other tested data sets.

4 Discussion

The work presented here demonstrates that dynamic linear models are an appropriate means to model fMRI time series. They are capable of accounting accurately for signal-drifts and other non-stationary baseline effects using a simple and intuitive nonparametric approach, without requiring any prior knowledge of the underlying processes. Because the baseline is modelled with only one parameter, also the interpretation of the estimated effect is simplified. It is not necessary to explain the meaning of several trend modelling covariables or basis-functions. Moreover, such a splitting does not seem to be advantageous. The source of background signal changes in fMRI time series is not completely understood, apart from breathing, cardiac cycles and CSF pulsations, also residual motion has to be considered. Further, the complexity of the modelled trend is not restricted by the number of included basis-functions, but is adjusted dynamically. Thus, it is independent of the operators supervision and therefore provides an interesting tool for a more comprehensive study of the nature of baseline changes.
The dynamically modelled stimulus effect reflects in a very flexible way the variability of stimulus-induced signal variations. Especially this variability is quite informative and important for the temporal investigation of the stimulus-related processes. This model provides the possibility to really examine the temporal evolution of activated brain areas

and thus to analyse the spatio-temporal development of a given activation, which is impossible with a single parameter. Whereas the commonly used correlation and regression techniques calculate a single activation map, i.e. they compress the time series to a single parameter, the nonparametric approach yields a parametric map for every image, keeping the temporal dimension of the experiment. This does not only offer the possibility to investigate the temporal differences in the responses of several voxels as by the approach of Friston et al. (3), but also to analyse the amplitudinal variation.

As shown for the language task, temporal resolution of the fMRI maps enables the detection of transient variations in the activation pattern possibly reflecting differential cognition during prolonged listening to a complex story. Thus, a potentially much finer detection level for specific answers to short stimuli can be foreseen. Further studies will be necessary to define the physiological limitation of temporal resolution, i.e. necessary minimum persistence of activation in order to be detected with this approach. However, this time-resolved approach together with carefully designed temporally varying stimuli affords a better exploitation of the inherent information from fMRI time series. Another advantage that can be expected for robustness of results in simple sensory or motor tasks, where habituation effects or other subject dependant sources of variation like attention, drowsiness or motor weakness may affect BOLD responsiveness. Systematic comparison of this model with correlation techniques for analysis of event-related versus prolonged steady state stimulation designs will finally provide insight into optimal adaptation of temporal stimulation schedules for specific neurobiological applications.

Up to now an important aspect of the analysis of fMRI experiments has not been taken into account. Critical in the analysis of fMRI experiments is the treatment of spatial dependencies between the voxels. The whole analysis was carried out pixelwise. Even though the spatial connectivity is a highly nontrivial problem, its consideration should not be neglected. To regard the voxels in the analysis as independent individuals is certainly not correct. But it is questionable whether the present approaches, cluster-analysis (12) or applying the theory of Gaussian random fields (13), can meet the requirements of this task. Whereas cluster-analysis more or less keeps existing contours in the parametric map, it does not account for the quantitative differences of the adjacent voxels and is strongly

dependent on the choice of the threshold. Further, it cannot provide any objective confidence level. Applying some distributional assumptions and the theory of Gaussian random fields yield such p-values. But this imposes another problem. The necessary blurring prior to the analysis smoothes out the contours and edges of the map. Also, it is not evident whether activation maps fulfil the assumption of a stationary Gaussian random field.

Possible ways to overcome these problems might be approaches that incorporate spatial aspects already on the stage of the time series models. This will get close to the idea of a really spatio-temporal analysis. In other areas of research such models exist already. In the field of epidemiological investigations fully Bayesian approaches for a spatio-temporal analysis are actually made (14). Here the spatial connectivity is explicitly considered in Markovian random fields defining neighbourhood systems. The temporal and spatial interactions enter in the model simultaneously and are equally weighted. However, the challenging task of a transfer to fMRI data is not trivial. The huge amount of data and some unpleasant characteristics of the data, like the small signal to noise ratio, partial volume effects in the signal intensities or spatial inhomogeneities, are aspects that complicate this endeavour.

Until its realisation more extensive validation studies will have to be performed to prove that the described positive properties of the presented model have not only occurred by chance. A comparison with other imaging techniques that have a higher temporal resolution would be very informative. First approaches in the direction of an inclusion of EEG methods are done but have to be looked at in greater detail.

Acknowledgement

This work was supported by a grant from the German National Science Foundation (DFG), Sonderforschungsbereich 386.

References

1. P. A. Bandettini, A. Jesmanowicz, E. C. Wong, J. S. Hyde. Processing strategies for time-course data sets in functional MRI of the brain. *Magn. Reson. Med.* **30**, 161–173 (1993).

2. K. J. Friston, A. P. Holmes, J.-B. Poline, P.J. Grasby, S. C. R. Williams, R. S. J. Frackowiak. Analysis of fMRI Time-Series Revisited. *Neuroimage* **2**, 45–53 (1995).

3. K. J. Friston, C. D. Frith, R. Turner, R. S. J. Frackowiak. Characterizing evoked hemodynamics with fMRI. *Neuroimage* **2**, 173–181 (1995).

4. K. J. Friston, C. D. Frith, R. S. J. Frackowiak, R. Turner. Characterizing dynamic brain responses with fMRI: A multivariate approach. *Neuroimage* **2**, 166–172 (1995).

5. C. Biller. Posterior mode estimation in generalized linear mixed models. *SFB-Discussion-Paper Nr. 70*, http://www.stat.uni-muenchen.de/sfb386/publikation.html (1997).

6. L. Fahrmeir, G. Tutz. "Multivariate Statistical Modelling Based on Generalized Linear Models". Springer Verlag New York, New York, 1994.

7. A. C. Harvey. "Forecasting, Structural Time Series Models and the Kalman Filter". Cambridge University Press, Cambridge, 1989.

8. P. De Jong. The Diffuse Kalman-Filter. *Annals of Statistics* **19**, No. 2 1073–1083 (1991)

9. W. H. Press, S. A. Teukolski, W. T. Vetterling, B. P. Flannery. "Numerical recipes in C, second edition". Cambridge University Press, Cambridge, 1992.

10. R. P. Woods, S. R Cherry, J. C. Mazziota. Rapid automated algorithm for aligning and reslicing PET images. *J. Comp. Assist. Tomogr.* **16**, 4, 620 (1992).

11. A. P. Jiang, D. N. Kennedy, J. R. Baker, R. Weisskoff, R. B. H. Tootell, R. P. Woods, R. R. Benson, K. K. Kwong, T. J. Brady, B. R. Rosen and J. W. Belliveau. Motion

detection and correction in functional MR imaging. *Hum. Brain Mapping* **3**, 224–235 (1995)

12. S. D. Forman, J. D. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintun, D. C. Noll. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* **33**, 636–647 (1995).

13. J. B. Poline, K. J. Worsley, A. C. Evans, K. J. Friston. Combining Spatial Extent and Peak Intensity to Test for Activations in Functional Imaging. *Neuroimage* **5**, 83–96 (1997).

14. A. Mollié. Bayesian mapping of disease. In "Markov Chain Monte Carlo in Practice"(Gilks, Richardson, Spiegelhalter, Eds.), pp. 441–464 Chapman & Hall, London, 1996.

Legends for figures

Figure 1: Fitted Models, effects and residual-plots for a selected example. The upper row shows a simple linear regression, the centre row a nonparametrically modelled baseline and the lower row the completely nonparametric approach.

Figure 2a: Time-dependent activation maps for visual stimulation, estimated with the nonparametric model.

Figure 2b: Cross-correlation map for visual stimulation

Figure 3a: Time-dependent activation maps for acoustic stimulation, estimated with the nonparametric model.

Figure 3b: Cross-correlation map for acoustic stimulation

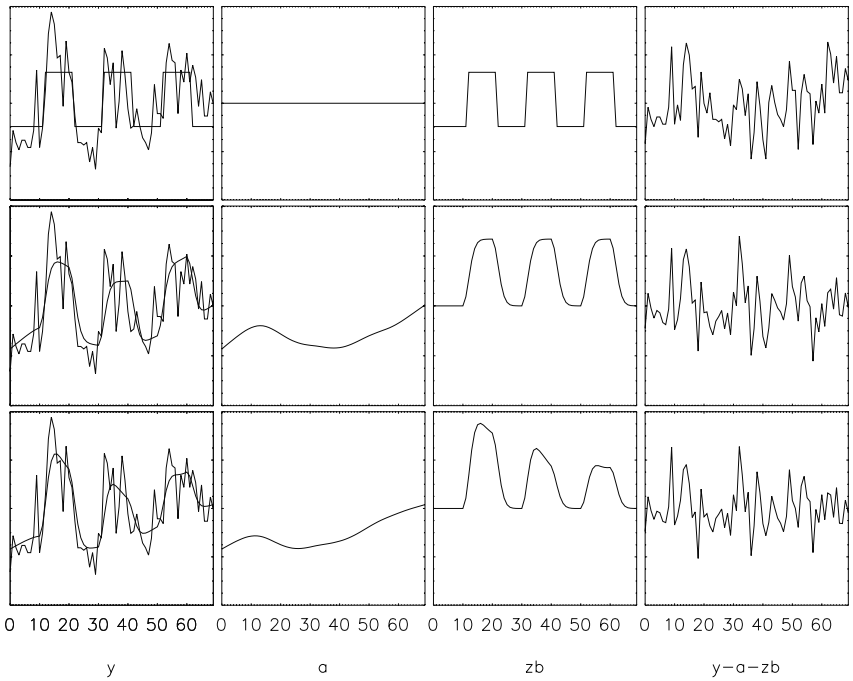

Figure 1: Fitted Models, effects and residual-plots for a selected example. The upper row shows a simple linear regression, the centre row a nonparametrically modelled baseline and the lower row the completely nonparametric approach.

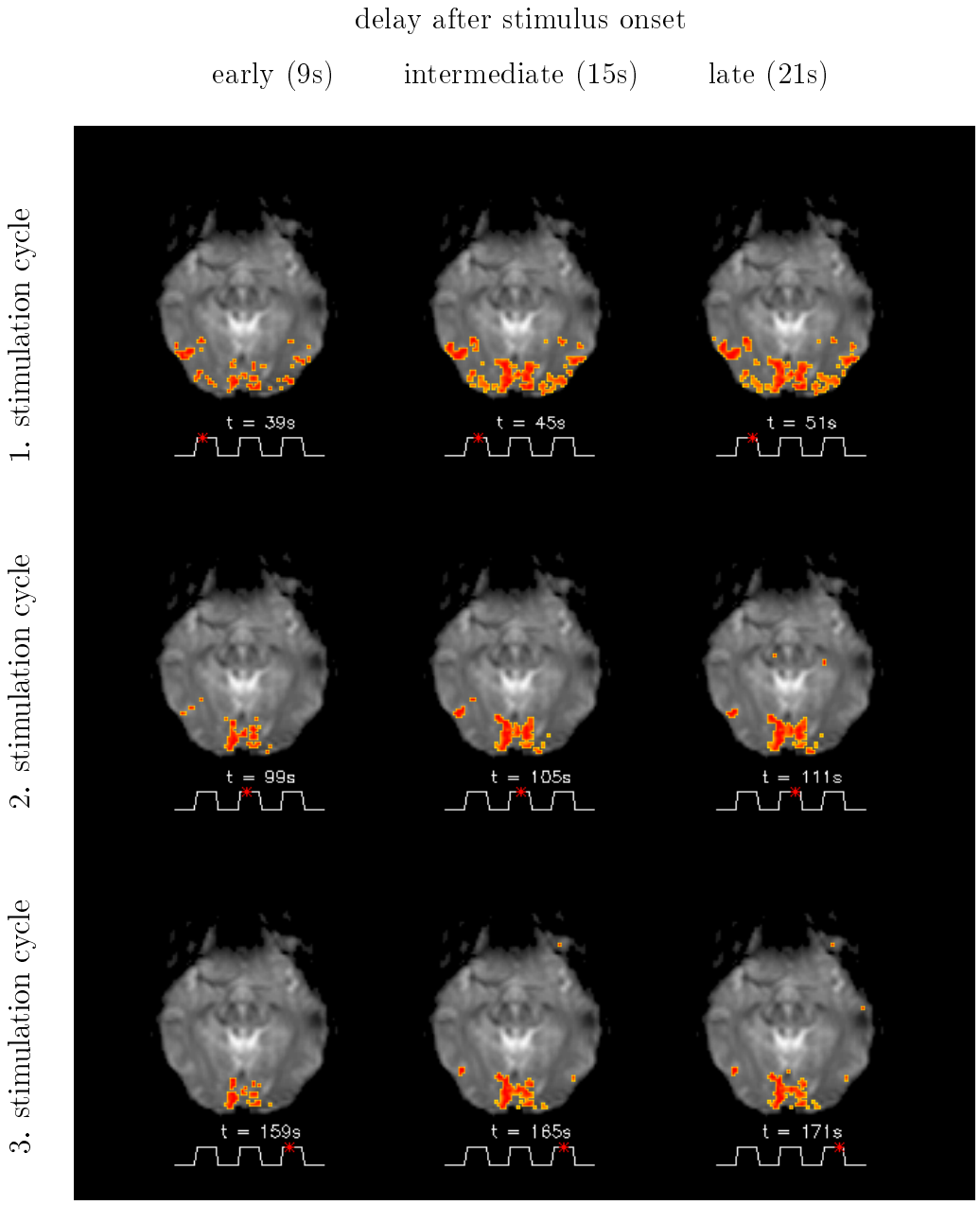

Figure 2a: Time-dependent activation maps for visual stimulation, estimated with the nonparametric model.

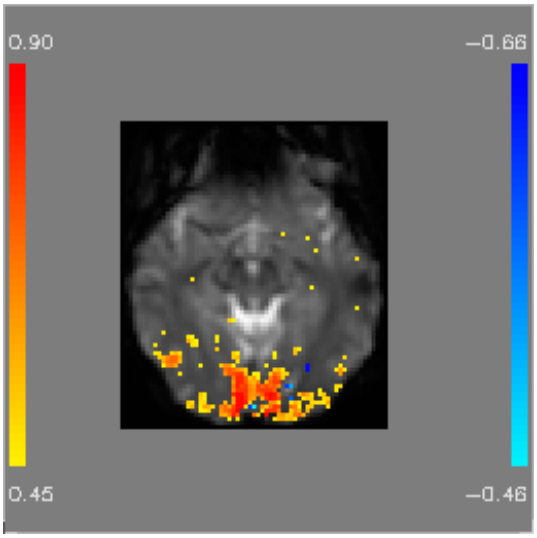

Figure 2b: Cross-correlation map for visual stimulation

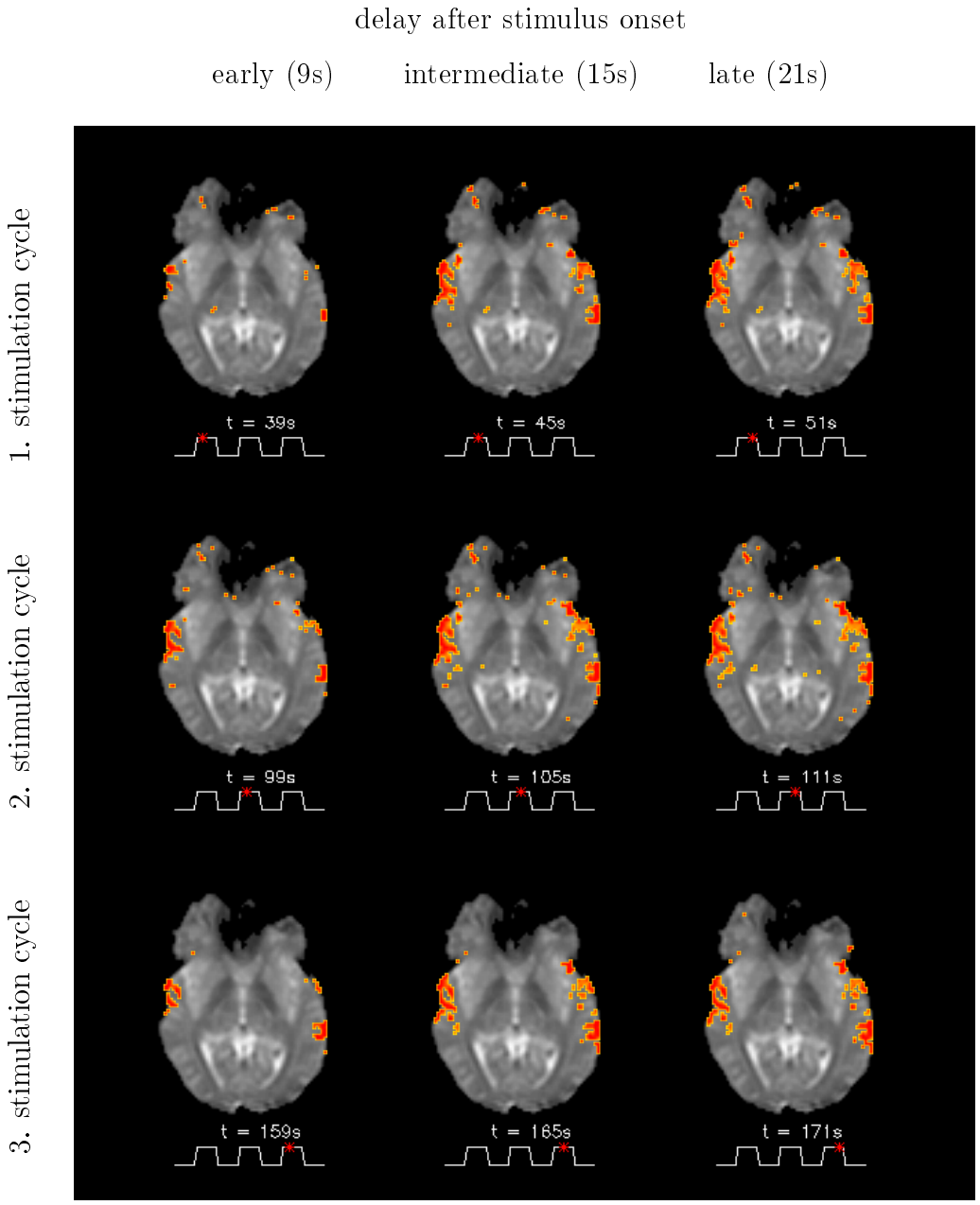

Figure 3a: Time-dependent activation maps for acoustic stimulation, estimated with the nonparametric model.

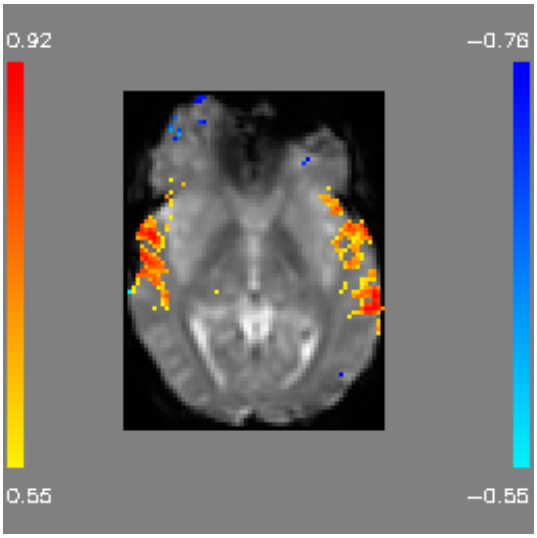

Figure 3b: Cross-correlation map for acoustic stimulation