

# Solving Incrementally the Fitting and Detection Problems in fMRI Time Series

Alexis Roche, Philippe Pinel, Stanislas Dehaene, and Jean-Baptiste Poline

CEA, Service Hospitalier Frédéric Joliot, Orsay, France  
Institut d’Imagerie Neurofonctionnelle (IFR 49), Paris, France  
`roche@shfj.cea.fr`

**Abstract.** We tackle the problem of real-time statistical analysis of functional magnetic resonance imaging (fMRI) data. In a recent paper, we proposed an incremental algorithm based on the extended Kalman filter (EKF) to fit fMRI time series in terms of a general linear model with autoregressive errors (GLM-AR model). We here improve the technique using a new Kalman filter variant specifically tailored to the GLM-AR fitting problem, the *Refined Kalman Filter* (RKF), that avoids both the estimation bias and initialization issues typical from the EKF, at the price of increased memory load. We then demonstrate the ability of the method to perform online analysis on a “functional calibration” event-related fMRI protocol.

## 1 Introduction

One of the current challenges in functional magnetic resonance imaging (fMRI) is to display reconstructed volumes and map brain activations in real time during an ongoing scan. This will make it possible to interact with fMRI experiments in a much more efficient way, either by monitoring acquisition parameters online depending on subject’s performance, or by designing paradigms that incorporate neurophysiological feedback. To date, the feasibility of real-time fMRI processing has been limited by the computational cost of both the three-dimensional reconstruction of MR scans and their statistical analysis.

This paper addresses the latter item, and is therefore focused on the feasibility of online fMRI statistical analysis. In this context, our goal is to fit, on each scan time, the currently available fMRI time course in terms of an appropriate model of the BOLD response, and further test for brain regions that are significantly correlated with the model. We will focus here on general linear models (GLM) [1] as they are by far the most common in the fMRI processing community.

Although many detection algorithms have been proposed so far, most of them are intended to work offline in the sense that they process a complete fMRI sequence once at a time, with computational cost and memory load proportional to the sequence length. Applying such methods online would imply that the incremental computation time increases on each new scan, which is clearly a serious drawback when considering real-time constraints. To overcome this problem, some techniques were proposed that compute the correlation between the signal

and the model, either in an incremental fashion [2], or by restricting computations to a sliding time window [3].

Such methods have the ability to process each new scan in a constant amount of time, but, being based on standard correlation, they work under the implicit assumption that the errors in the signal are temporally uncorrelated. Should this assumption be incorrect, the significance level of activation clusters may be substantially biased (over- or under-estimated). The importance of correcting inferences for temporal autocorrelations is widely recognized owing to the following facts: (i) errors may be found to be severely autocorrelated in some regions, especially when the model lacks flexibility; (ii) since autocorrelation is spatially dependent, it cannot be accounted for by a global threshold correction.

We recently advocated Kalman filtering techniques as good candidates for online fMRI analysis [4]. In its standard form, the Kalman filter is an incremental solver for ordinary least-square (OLS) regression problems, and is therefore well-suited for GLM fitting when assuming uncorrelated errors. In the more general case where the noise autocorrelation is unknown and is therefore to be estimated, the regression problem becomes nonlinear, a situation that may be handled using an extended Kalman filter (EKF) [4]. This technique's main drawback is that it requires parameter initialization to work; we observed from practical experience that good initialization is difficult to tune, and is very much machine-dependent.

To work around these issues, we design here a new Kalman filter variant to solve the GLM-AR fitting problem incrementally. Rather than using the linearization mechanism underlying the EKF, our basic idea is to rely on the standard Kalman filter to provide first parameter guesses on each iteration, and then refine the result using a simple optimization scheme. We will show that the algorithm outperforms the EKF in that it is insensitive to initialization, and provides asymptotically unbiased parameter estimates.

## 2 GLM-AR Model Fitting

Let us consider the time course vector  $\mathbf{y} = [y_1, \dots, y_n]^t$  associated with a given voxel in an fMRI sequence, where the acquisition times are numbered from 1 to  $n$ . In the remainder, it will be assumed that each incoming scan is spatially aligned with the first scan, which may necessitate a realignment procedure. In our usual processing pipeline, no spatial filtering is applied to the original scans (this enables us to use a slice-specific model to account for slice timing effects).

### 2.1 The GLM-AR Model

The general linear model states that the measured time course is a linear combination of known signals  $\mathbf{x}_1, \dots, \mathbf{x}_p$  called regressors, up to an additive noise:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where  $\mathbf{X} \equiv (\mathbf{x}_1^t, \dots, \mathbf{x}_p^t)$  is a  $n \times p$  matrix called the design matrix, which concatenates the different regressors columnwise,  $\boldsymbol{\varepsilon}$  is the outcome of the noise,

and  $\beta$  is the unknown  $p \times 1$  vector of regression coefficients, or “effect” vector. The design matrix contains paradigm-related regressors obtained, e.g., by convolving the different stimulation onsets with a canonical hemodynamic response function [1], as well as regressors that model the low-frequency drift, hence enabling us to “detrend” the signal (we use polynomials up to order three). Notice that the design matrix can be assembled incrementally since it involves either causal convolutions, or pre-specified detrending functions.

In this work, we assume that  $\varepsilon$  is a stationary Gaussian zero-mean AR(1) random process, i.e. it is characterized by:  $\varepsilon_i = a\varepsilon_{i-1} + n_i$ , where  $a$  is the autocorrelation parameter, and  $n_i$  is a “generator” white noise, with instantaneous Gaussian distribution  $N(0, \sigma^2)$ . Notice that the condition  $|a| < 1$  must hold for the AR noise to be stationary.

## 2.2 Offline Fitting

Solving the GLM-AR fitting problem means finding appropriate, somehow optimal, statistical estimators of the effect  $\beta$ , the noise autocorrelation  $a$  and scale parameter  $\sigma$ . A powerful estimation approach consists of maximizing the likelihood function or, equivalently, minimizing its negated logarithm given by [4]:

$$L(\beta, a, \sigma) = n \log \sqrt{2\pi}\sigma + \frac{1}{2} \log(1 - a^2) + \frac{1}{2\sigma^2} \left[ (1 - a^2)r_1^2 + \sum_{i=2}^n (r_i - ar_{i-1})^2 \right], \quad (1)$$

where  $r_i \equiv y_i - \mathbf{x}_i^t \beta$  denotes the residual at time  $i$ , and is a function of  $\beta$  only.

There is no closed-form solution to the minimization of equation (1), except when  $a$  is considered known beforehand, hence kept constant, in which case the problem boils down to traditional OLS regression. Based on this remark, maximum likelihood estimation may be implemented using an alternate optimization scheme, ensuring locally optimal parameter estimates [4]. However, because each iteration involves assembling and inverting a  $p \times p$  matrix, it may be hopelessly time consuming when dealing with large models. Alternative estimation strategies include pre-coloring [1], pre-whitening [5], bias-corrected OLS estimation [6], restricted maximum likelihood [7], and variational Bayesian techniques [8].

## 2.3 Online Fitting: The Refined Kalman Filter

In real-time context, we aim to solve the fitting problem each time a new measurement is available, i.e., at time  $i$ , process the partial sequence  $(y_1, y_2, \dots, y_i)$  as if it was the complete one. We discussed in section 1 the need for specific techniques to achieve such incremental analysis. We present here the refined Kalman filter (RKF) as an alternative to previous online fitting techniques [2,3,4].

The online estimation problem may be formulated in terms of maximizing the likelihood function (1) as applied to the sequence available at time  $i$ . For

better computational tractability, we will however consider a slightly modified version of the likelihood criterion:

$$\begin{aligned} \tilde{L}_i(\boldsymbol{\beta}, a, \sigma) &= i \log \sqrt{2\pi}\sigma + \frac{1}{\sigma^2} C_i(\boldsymbol{\beta}, a) \\ \text{with } C_i(\boldsymbol{\beta}, a) &= (1 + a^2) \underbrace{\frac{1}{2} \sum_{k=1}^i r_k^2(\boldsymbol{\beta})}_{C_i^0(\boldsymbol{\beta})} - 2\gamma_i a \underbrace{\frac{1}{2} \sum_{k=2}^i r_k(\boldsymbol{\beta}) r_{k-1}(\boldsymbol{\beta})}_{C_i^1(\boldsymbol{\beta})}, \end{aligned} \tag{2}$$

where we define  $\gamma_i \equiv i/(i-1)$ . It may be shown that this modified likelihood is asymptotically equivalent to the genuine likelihood in the sense that the average difference  $(\tilde{L}_i - L_i)/i$  converges uniformly towards zero (on any bounded open set) as  $i$  approaches infinity. Therefore, the minimizers of (2) inherit the general maximum likelihood property of being asymptotically unbiased. Notice that for the parameter  $\boldsymbol{\beta}$ , the property holds not only asymptotically, but for any sample size. We introduce the correction factor  $\gamma_i$  to further reduce the estimation bias on  $a$  and  $\sigma$ . The RKF principles then arise from the following remarks:

- From equation (2), we observe that the estimation of  $\sigma$  may be completely decoupled from that of  $(\boldsymbol{\beta}, a)$ ; clearly, the optimal scale is determined from the minimum of  $C_i$  by:  $\sigma_i^2 = (2/i) \min_{\boldsymbol{\beta}, a} C_i(\boldsymbol{\beta}, a)$ .
- The criterion  $C_i(\boldsymbol{\beta}, a)$  is a weighted sum of two functions of  $\boldsymbol{\beta}$  only,  $C_i^0(\boldsymbol{\beta})$  and  $C_i^1(\boldsymbol{\beta})$ , the first of which is the classical OLS criterion, and may be calculated incrementally using a standard Kalman filter. A similar incremental calculation may be used for the second term  $C_i^1(\boldsymbol{\beta})$  as it is also quadratic.
- From the calculation of both  $C_i^0(\boldsymbol{\beta})$  and  $C_i^1(\boldsymbol{\beta})$ , an alternate minimization scheme similar to that described in section 2.2 can be used to iteratively estimate the autocorrelation  $a$ , and refine the OLS estimate of  $\boldsymbol{\beta}$ .

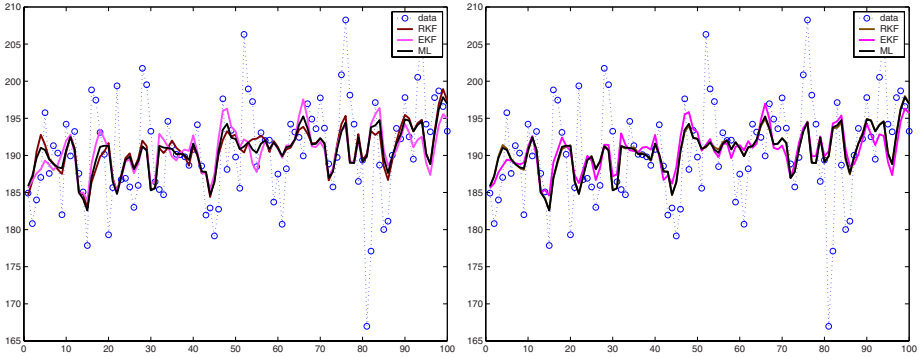
The RKF algorithm is detailed in table 1, and commented here below.

*Standard Kalman iterations.* The Kalman filter is used to incrementally update the OLS criterion  $C_i^0$  defined in equation (2) so as to provide a starting guess of  $\boldsymbol{\beta}$  on each scan time. One motivation for this strategy is that the OLS estimator is at least unbiased despite it is not optimal for the GLM-AR model [1,6]. On each scan time  $i$ , the Kalman filter updates the minimizer  $\boldsymbol{\beta}_i^0$  of  $C_i^0$ , the minimum criterion value  $c_i^0 \equiv C_i^0(\boldsymbol{\beta}_i^0)$ , as well as its inverse Hessian  $\mathbf{S}_i^0$ . Since the Hessian  $\mathbf{H}_i^0$  is later needed in the refinement loop, we also update its value recursively in order to avoid inverting  $\mathbf{S}_i^0$ .

*Refinement loop.* After performing one Kalman iteration, we update the “correction” function  $C_i^1(\boldsymbol{\beta})$  involved in equation (2), which is quadratic, hence fully specified by its derivatives up to order two. Let  $c_i^1 \equiv C_i^1(\boldsymbol{\beta}_i^0)$ ,  $\mathbf{g}_i^1 \equiv \partial C_i^1 / \partial \boldsymbol{\beta}(\boldsymbol{\beta}_i^0)$  and  $\mathbf{H}_i^1 \equiv \partial^2 C_i^1 / \partial \boldsymbol{\beta}^2$  denote respectively the function value, gradient and Hessian computed at the current OLS estimate  $\boldsymbol{\beta}_i^0$ . Those quantities are easily related to their previous values using equation (3) in table 1.

At the stage where both  $C_i^0(\beta)$  and  $C_i^1(\beta)$  are calculated, it becomes possible to minimize  $C_i(\beta, a)$  as defined in equation (2), which is the actual estimation criterion we are interested in. To that end, we perform an alternate minimization of  $C_i(\beta, a)$ . When  $\beta_i$  is held fixed, the optimal autocorrelation is clearly given by  $a_i = \gamma_i C_i^1(\beta_i) / C_i^0(\beta_i)$ . On the other hand, when  $a_i$  is fixed, re-estimating  $\beta$  amounts to minimizing the sum of two quadratic functions, yielding a closed-form solution given by equation (4) in table 1. The formula involves  $\mathbf{S}_i \equiv (\partial^2 C_i / \partial \beta^2)^{-1}$ , the inverse Hessian of  $C_i(\beta, a)$  w.r.t.  $\beta$ , which is a function of  $a_i$  only as it is independent of  $\beta_i$ . This matrix plays a key role at the detection stage as it closely relates to the covariance of  $\beta_i$  (see section 2.4).

*Comparison with EKF.* The key feature of the RKF is that its incremental updates do not involve any approximation, unlike the EKF [4] which proceeds by successive linearizations. This property is achieved exploiting the specific form of the estimation criterion (2), and jointly updating the two quadratic functions  $C_i^0(\beta)$  and  $C_i^1(\beta)$  exactly. Hence, the data information is fully preserved by the RKF, whereas it is unavoidably degraded across iterations using an EKF.



**Fig. 1.** Comparative fitting example using the RKF and the EKF in an event-related paradigm (see section 3). From left to right, RKF (dark curve) and EKF (bright curve) results using respectively 1 and 2 local iterations are compared with the maximum likelihood result (black curve) computed using the offline algorithm described in section 2.2. In this case, the model contains 15 regressors.

## 2.4 Online Detection

On each scan time  $i$  the RKF provides a current estimate  $\beta_i$  of the effect in each voxel. However, to test whether the effect is significant, we also need to evaluate some kind of measure of uncertainty on this estimate. Based on the remark that our estimation criterion is an asymptotically valid likelihood function (see section 2.3), its inverse “Fisher information” is a natural approximation of the variance matrix of  $\beta_i$ :  $\text{Var}(\beta_i) \approx (\frac{\partial^2 \bar{L}_i}{\partial \beta^2})^{-1} = \sigma_i^2 \mathbf{S}_i$ , where  $\sigma_i$  is the current scale estimate, and  $\mathbf{S}_i$  is the inverse Hessian defined in section 2.3.

**Table 1.** Refined Kalman Filter (RKF) synopsis.

Initialize with:  $\beta_0^0 = \mathbf{0}$ ,  $\mathbf{H}_0^0 = \mathbf{0}_p$ ,  $\mathbf{S}_0^0 = \lambda \mathbf{I}_p$ , with  $\lambda$  large enough (e.g.  $\lambda = 10^{10}$ ).  
For  $i \in \{1, 2, \dots, n\}$ ,

1. Update the OLS estimate (standard Kalman iteration). Compute the auxiliary variables:  $\rho_i = y_i - \mathbf{x}_i^t \beta_{i-1}^0$ ,  $\mathbf{k}_i = \mathbf{S}_{i-1}^0 \mathbf{x}_i$ ,  $v_i = \mathbf{x}_i^t \mathbf{k}_i$ , in order to perform the following recursion:

$$\begin{aligned}\beta_i^0 &= \beta_{i-1}^0 + \Delta \beta_i^0 \quad \text{with} \quad \Delta \beta_i^0 = \frac{\rho_i}{v_i} \mathbf{k}_i \\ \mathbf{S}_i^0 &= \mathbf{S}_{i-1}^0 - \frac{1}{v_i} \mathbf{k}_i \mathbf{k}_i^t \\ \mathbf{H}_i^0 &= \mathbf{H}_{i-1}^0 + \mathbf{x}_i \mathbf{x}_i^t \\ c_i^0 &= c_{i-1}^0 + \frac{\rho_i^2}{2v_i}\end{aligned}$$

2. Compute the value, gradient and Hessian of  $C_i^1(\beta)$  at the new OLS estimate  $\beta_i^0$ . Using the residuals  $r_i = y_i - \mathbf{x}_i^t \beta_i^0$  and  $r_{i-1} = y_{i-1} - \mathbf{x}_{i-1}^t \beta_{i-1}^0$ , do:

$$\begin{aligned}c_i^1 &= c_{i-1}^1 + (\mathbf{g}_{i-1}^1)^t \Delta \beta_i^0 + \frac{1}{2} (\Delta \beta_i^0)^t \mathbf{H}_{i-1}^1 \Delta \beta_i^0 + \frac{1}{2} r_i r_{i-1} \\ \mathbf{g}_i^1 &= \mathbf{g}_{i-1}^1 + \mathbf{H}_{i-1}^1 \Delta \beta_i^0 - \frac{1}{2} (r_{i-1} \mathbf{x}_i + r_i \mathbf{x}_{i-1}) \\ \mathbf{H}_i^1 &= \mathbf{H}_{i-1}^1 + \frac{1}{2} (\mathbf{x}_i \mathbf{x}_{i-1}^t + \mathbf{x}_{i-1} \mathbf{x}_i^t)\end{aligned} \quad (3)$$

3. Refinement loop. Initialize:  $\beta_i = \beta_i^0$  and  $\mathbf{S}_i = \mathbf{S}_i^0$ , then repeat the following two-pass routine a fixed number of times:

- Estimate the autocorrelation, using the values of  $C_i^0(\beta)$  and  $C_i^1(\beta)$  at the current estimate  $\beta_i$ , and the deviation from the OLS estimate  $\Delta \beta_i = \beta_i - \beta_i^0$ ,

$$\begin{aligned}\tilde{c}_i^0 &= c_i^0 + \frac{1}{2} \Delta \beta_i^t \mathbf{H}_i^0 \Delta \beta_i \\ \tilde{c}_i^1 &= c_i^1 + (\mathbf{g}_i^1)^t \Delta \beta_i + \frac{1}{2} \Delta \beta_i^t \mathbf{H}_i^1 \Delta \beta_i \\ a_i &= \gamma_i \frac{\tilde{c}_i^1}{\tilde{c}_i^0}\end{aligned}$$

- Refine  $\beta_i$  and the inverse Hessian  $\mathbf{S}_i$ ,

$$\begin{aligned}\mathbf{S}_i &= \frac{1}{1 + a_i^2} (\mathbf{I}_p + \frac{2\gamma_i a_i}{1 + a_i^2} \mathbf{S}_i^0 \mathbf{H}_i^1) \mathbf{S}_i^0 \\ \beta_i &= \beta_i^0 + 2\gamma_i a_i \mathbf{S}_i \mathbf{g}_i^1\end{aligned} \quad (4)$$

4. Estimate the scale:

$$\sigma_i^2 = 2(1 - a_i^2) \frac{\tilde{c}_i^0}{i}$$

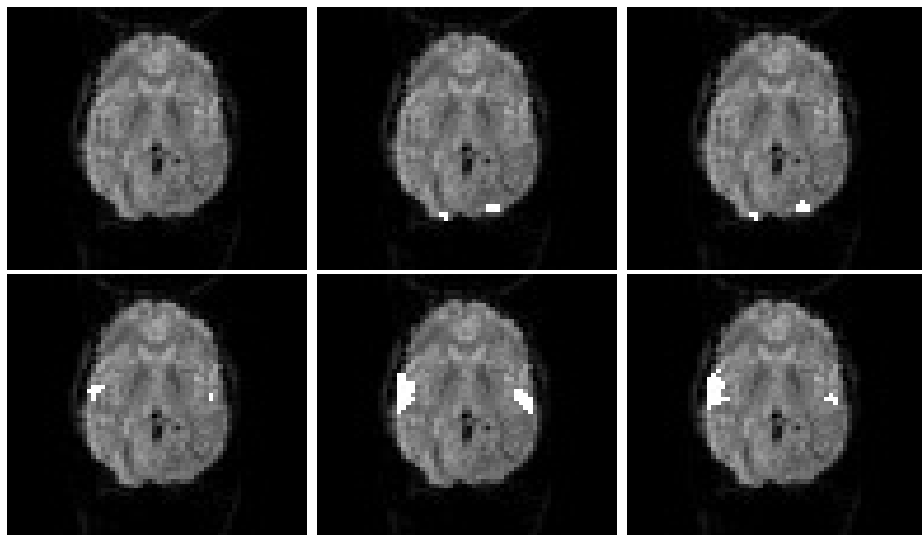
Given a contrast vector  $\mathbf{c}$ , we are interested in identifying the voxels that show a contrasted effect  $\mathbf{c}^t \beta$ , for instance, significantly positive. As a first-order approximation, we may assume that the effect's estimate is normally distributed around the true, unknown effect  $\beta^*$ , i.e.  $\beta_i \sim \mathcal{N}(\beta^*, \sigma_i^2 \mathbf{S}_i)$ . Hence, under the null hypothesis that  $\mathbf{c}^t \beta^* = 0$ , the statistic:

$$z_i = \text{Var}(\mathbf{c}^t \beta_i)^{-\frac{1}{2}} \mathbf{c}^t \beta_i = \sigma_i^{-1} (\mathbf{c}^t \mathbf{S}_i \mathbf{c})^{-\frac{1}{2}} \mathbf{c}^t \beta_i,$$

defines a  $z$ -score. Testing for positive activations may thus be achieved at any time  $i$  by thresholding the image of  $z$ -scores. Notice that this approach may also be interpreted in a Bayesian perspective [4]. As is standard in practice, we apply some spatial Gaussian smoothing to the  $z$ -score image before thresholding to improve the localization power of detection [1]. We usually set the threshold so as to match an uncorrected  $p$ -value of  $10^{-3}$ , although this should ideally be corrected for multiple comparisons.

### 3 Results

The method was tested *offline* on several fMRI datasets acquired on our site from both GE Signa 1.5T and Bruker 3T whole-body scanners, always providing final results consistent with SPM'99. For illustration, we present here a “functional calibration” protocol designed to localize the main brain functions in about five minutes. The experimental paradigm contains 11 different conditions (labelled as ‘visual’, ‘motor’, ‘calculation’ and ‘language’), from which a total of 100 events are presented pseudo-randomly to the subject. The data was acquired on the Bruker 3T scanner using a 3s repetition time, for a total of 100 scans with  $64 \times 64 \times 26$  voxels of size  $3.75 \times 3.75 \times 4.5$  mm<sup>3</sup>.



**Fig. 2.** Incremental detection of visual and auditory regions in a functional calibration paradigm. From left to right, activation maps after respectively 2'00", 3'30" and 5'00".

The RKF algorithm was applied *offline* to the fMRI sequence. One regressor was associated with each condition by convolving its onsets with a canonical hemodynamic response function [1]. Three additional polynomial regressors were

used to model the low frequency drifts present in the signal. The number of iterations in the refinement loop was set to three. Using a C implementation, the computation time to process each time frame was about two tenth of second on a standard PC (1.80GHz processor).

The activation maps in figure 2 show in an axial slice the regions that were detected respectively after 40 (2'00"), 70 (3'30") and 100 (5'00") scans, for a contrast between visual and auditory sentences. As expected, positive effects are found laterally in the occipital lobe where is the visual cortex (top row), while negative effects are found in the temporal lobes (bottom row). After 2'00", no significant visual region is detected in this slice, whereas auditory regions are already appearing. Larger clusters are found after 3'30" without major changes until the end of the sequence. We notice a subtle loss of sensitivity in the right temporal lobe, which might be explained either by a late motion of the subject, or by a neuronal adaptation effect. Although rather qualitative, these results demonstrate the potential use of real-time fMRI, suggesting that functional regions may be detected significantly before the end of an experiment.

## 4 Conclusion

We have improved our previous incremental, EKF-based detection method for fMRI time series by designing an original Kalman variant called the refined Kalman filter (RKF). The new method achieves excellent statistical performances without requiring any initialization parameter unlike the EKF and classical variants such as the second-order EKF or the unscented Kalman filter [9]. The price to pay is essentially increased memory load, as the RKF tends to run even slightly faster than the EKF.

## References

1. Friston, K.J.: 2. In: *Human Brain Function*. Academic Press (1997) 25–42
2. Cox, R., Jesmanowicz, A., Hyde, J.: Real-Time Functional Magnetic Resonance Imaging. *Magnetic Resonance in Medicine* **33** (1995) 230–236
3. Gembris, D., Taylor, J., Schor, S., Frings, W., Suter, D., Posse, S.: Functional Magnetic Resonance Imaging in Real Time (FIRE). *Magnetic Resonance in Medicine* **43** (2000) 259–268
4. Roche, A., Lahaye, P.J., Poline, J.B.: Incremental activation detection in fmri series using kalman filtering. In: *Proc. 2st Proc. IEEE ISBI, Arlington, VA* (2004) 376–379
5. Woolrich, M., Ripley, B., Brady, M., Smith, S.: Temporal autocorrelation in univariate linear modelling of fMRI data. *Neuroimage* **14** (2001) 1370–1386
6. Worsley, K., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., Evans, A.: A general statistical analysis for fMRI data. *Neuroimage* **15** (2002) 1–15
7. Friston, K., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J.: Classical and Bayesian Inference in Neuroimaging: Theory. *NeuroImage* **16** (2002) 465–483
8. Penny, W., Kiebel, S., Friston, K.: Variational Bayesian Inference for fMRI time series. *NeuroImage* (2003) In press.
9. Julier, S., Uhlmann, J.: A new extension of the kalman filter to nonlinear systems. In: *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*. (1997)