Likelihood-Based Hypothesis Tests for Brain Activation Detection From MRI Data Disturbed by Colored Noise: A Simulation Study

A. J. den Dekker*, D. H. J. Poot, R. Bos, and J. Sijbers

Abstract—**Functional magnetic resonance imaging (fMRI) data that are corrupted by temporally colored noise are generally preprocessed (i.e., prewhitened or precolored) prior to functional activation detection. In this paper, we propose likelihood-based hypothesis tests that account for colored noise *directly* within the framework of functional activation detection. Three likelihood-based tests are proposed: the generalized likelihood ratio (GLR) test, the Wald test, and the Rao test. The fMRI time series is modeled as a linear regression model, where one regressor describes the task-related hemodynamic response, one regressor accounts for a constant baseline and one regressor describes potential drift. The temporal correlation structure of the noise is modeled as an autoregressive (AR) model. The order of the AR model is determined from practical null data sets using Akaike's information criterion (with penalty factor 3) as order selection criterion. The tests proposed are based on exact expressions for the likelihood function of the data. Using Monte Carlo simulation experiments, the performance of the proposed tests is evaluated in terms of detection rate and false alarm rate properties and compared to the current general linear model (GLM) test, which estimates the coloring of the noise in a separate step. Results show that theoretical asymptotic distributions of the GLM, GLR, and Wald test statistics cannot be reliably used for computing thresholds for activation detection from finite length time series. Furthermore, it is shown that, for a fixed false alarm rate, the detection rate of the proposed GLR test statistic is slightly, but (statistically) significantly improved compared to that of the common GLM-based tests. Finally, simulations results reveal that all tests considered show seriously inferior performance if the order of the AR model is not chosen sufficiently high to give an adequate description of the correlation structure of the noise, whereas the effects of (slightly) overmodeling are observed to be less harmful.**

Index Terms—**Functional magnetic resonance imaging (fMRI), generalized likelihood ratio test, Rao test, statistical parametric maps, time series analysis, Wald test.**

I. INTRODUCTION

FUNCTIONAL magnetic resonance imaging (fMRI) is a noninvasive technique used to detect brain activity. By utilizing the fact that the magnetic resonance signal intensity is correlated with the cerebral blood flow, which in turn is correlated with neural activity [1], fMRI can localize brain regions that show significant neural activity upon stimulus presentation, where the stimulus is designed to activate the sensory, motor. or cognitive task under study. fMRI data sets typically consist of time series associated with the voxels of the brain. For each voxel, the significance of the response to the stimulus is assessed by statistically analyzing the associated fMRI time series. In this way, brain activation maps, or statistical parametric maps (SPMs), reflecting brain activity can be constructed.

Nowadays, fMRI time series are commonly modeled by a general linear model (GLM) disturbed by Gaussian distributed noise [2], [3]. Such a model is capable of including potential time trends by adopting extra linear terms. Furthermore, the GLM contains one or more activation related parameters of interest. Statistical parametric maps (SPMs) are obtained by testing the significance of the activation related GLM parameter(s) using standard statistical tools such as the two-sided t-test (in the one parameter case) or the F-test (in the case of more than one parameter). This method is also used in the "Estimate" and "Inference" steps of the well-known SPM software package [4].

The fMRI recordings are contaminated by noise from sources such as the MRI scanner, residual motion, and unrelated "spontaneous" brain activations [5]. It is reasonable to assume that this noise is colored (i.e., correlated) in the time direction. Unlike white noise, colored noise does not have a uniform (i.e., flat) power spectral density function. Since the underlying correlation structure is unknown, current methods deal with temporally correlated noise by prewhitening the data based on an *estimated* correlation matrix of the noise [3]. This correlation matrix is usually estimated by fitting an autoregressive (AR) time series model to the residuals obtained after fitting the general linear model to the fMRI time series in least squares sense [6]. This introduces a, usually small, bias in the correlation estimates [7]. Since an estimate of the correlation matrix instead of the unknown, true correlation matrix of the noise is used for prewhitening, the assumption that the test statistic has a student's t or F distribution (upon which inference on the significance of the response is based) is only approximately valid.

In this paper, an alternative approach is proposed. This approach is also based on a GLM with correlated noise modeled as an AR process, but unlike the common GLM approach, it does not require a prewhitening step. Instead, statistical inference is based on the exact likelihood function that describes the statistics of the data including the temporal correlation structure

*A. J. den Dekker is with the Delft University of Technology, Delft Center for Systems and Control, 2828 CD Delft, The Netherlands (e-mail: a.j.dendekker@tudelft.nl).

R. Bos is with the Delft University of Technology, Delft Center for Systems and Control, 2828 CD Delft, The Netherlands.

D. H. J. Poot and J. Sijbers are with the Visionlab, University of Antwerp, 2610 Antwerp, Belgium..

of the noise. No approximations are made. The order of the AR process is determined from practical null data sets, acquired in the absence of activity. Three likelihood-based statistical binary hypothesis tests are proposed: the generalized likelihood ratio test (GLRT), the Wald test, and the Rao test. In each case, the null hypothesis H_0 that no activation is present, is tested against the alternative hypothesis H_1 (activation is present). In the context of fMRI, the use of the GLRT has previously been proposed by Nan and Nowak [8]. However, they consider complex valued fMRI data contaminated with white noise while in the present work, we consider magnitude fMRI data and colored noise.

For the computation of the test statistics proposed, the maximum likelihood (ML) estimates of the unknown parameters under H_0 (Rao), H_1 (Wald), or both H_0 and H_1 (GLRT) are needed. They are obtained by maximizing the likelihood function with respect to all unknown parameters (including the parameters of the AR model) simultaneously. In this paper, the performance of the proposed tests are evaluated in terms of detection rate and false alarm rate.

It is known that the tests proposed have favorable asymptotic statistical properties [9]. The asymptotic statistical distributions of the test statistics under H_0 do not depend on any unknown parameters. Therefore, independent of the noise power, tests can be constructed that have a (specified) constant false alarm rate. Such a test is referred to as a constant false alarm rate (CFAR) test [9]. Whether these asymptotic properties also apply to a finite number of observations is investigated by means of simulation experiments. The performance of the proposed tests is also compared to that of the widely used t-test (which is based on the GLM approach).

The paper is organized as follows. Section II-A describes a general, statistical model of fMRI time series. Section II-B discusses the general linear model (GLM) approach, assuming correlated noise described by an autoregressive (AR) process. In Section II-C, the joint probability density function (PDF) of the data is derived. In Section II-D, some optimizations are introduced to efficiently compute the ML estimate. The Sections II-E–II-I describe the different test statistics. In Section III, experimental results are described. Section III-A describes how to determine the order of the AR process from null data sets. In the Sections III-B and III-C, the tests are applied to simulated and experimental data. Finally, conclusions are drawn in Section IV.

II. METHOD

A. Statistical Model of the fMRI Time Series

An fMRI time series $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ (the superscript T denotes matrix transposition) of equidistant observations can in general be modelled as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{v} \tag{1}$$

in which $\boldsymbol{X}$ is an $n \times m$ design matrix [2], [5]. It consists of m columns that model signals of interest and nuisance signals such as potential drift. Furthermore, $\boldsymbol{\theta}$ is an $m \times 1$ vector of unknown regression parameters and $\boldsymbol{v}$ is an $n \times 1$ vector that

represents stochastic noise contributions. The noise $\boldsymbol{v}$ is modelled as a stationary stochastic AR process of order r (i.e., an AR(r) process) [10]

$$v_t = e_t - \alpha_1 v_{t-1} - \alpha_2 v_{t-2} - \cdots - \alpha_r v_{t-r} \tag{2}$$

with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)^T$ the vector of AR parameters, t the time index and e_t independent, zero mean Gaussian distributed white noise with variance σ_e^2. Let $\sigma_e^2 \boldsymbol{V}$ be the $n \times n$ covariance matrix of the AR process (2), that is

$$\sigma_e^2 \boldsymbol{V} = \mathbb{E}[\boldsymbol{v}\boldsymbol{v}^T] \tag{3}$$

with $\boldsymbol{v} = (v_1, \ldots, v_n)^T$ and $\mathbb{E}[\cdot]$ the expectation operator. For observations of stationary stochastic processes, the covariance matrix has a Toeplitz structure. Therefore, the covariance matrix of the AR(r) process v_t may be written as

$$\sigma_e^2 \boldsymbol{V} = \sigma_v^2 \begin{pmatrix} \rho(0) & \rho(1) & \ldots & \rho(n-1) \\ \rho(1) & \rho(0) & \ldots & \rho(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(n-1) & \rho(n-2) & \ldots & \rho(0) \end{pmatrix} \tag{4}$$

where $\rho(k) = \mathbb{E}[v_t v_{t+k}]/\sigma_v^2$ and σ_v^2 is the variance of v_t. Notice that it follows from this definition that $\rho(0) = 1$. The elements of the matrix $\boldsymbol{V}$ can be expressed in terms of the AR parameters through the Yule Walker relations [11]

$$\rho(k) + \alpha_1 \rho(k-1) + \cdots + \alpha_r \rho(k-r) = 0,$$
$$k > 0, \quad \rho(-k) = \rho(k). \tag{5}$$

Furthermore, it can be shown that [10]

$$\sigma_v^2 = \frac{\sigma_e^2}{1 - \sum_{k=1}^{r} \alpha_k \rho(k)}. \tag{6}$$

Several authors have performed analyses that indicate that AR models give an accurate description of the actual temporal autocorrelation structure of the noise that contaminates fMRI data [6], [12]. The validity of the model will be assessed using experimental data in Section III-A.

In order to derive the different test statistics in the Sections II-F–II-I, first the generalized least squares (GLS) estimator, the joint probability density function of the data and the ML estimator are derived in the Sections II-B, II-D, and II-D, respectively.

B. Common GLM Approach

The widely used GLM approach, for example by SPM [4], consists of two steps. First, an estimate of the parameter vector $\boldsymbol{\theta}$ is obtained by least squares fitting of the model described by the right hand side of (1) to the data $\boldsymbol{y}$. This so-called ordinary least squares (OLS) estimator can be expressed in closed form by

$$\hat{\boldsymbol{\theta}}_{\text{OLS}} = (X^T X)^{-1} X^T \boldsymbol{y}. \tag{7}$$

Although not fully efficient, this estimator is unbiased [13]. Therefore, the residuals $\varepsilon_{\text{OLS}} = \boldsymbol{y} - X\hat{\boldsymbol{\theta}}_{\text{OLS}}$ have zero expectation values and a correlation structure that is approximately

equal to that of the noise $\boldsymbol{v}$. Assuming that the noise is generated by an AR(r) model, the parameters of this model and hence the matrix $\boldsymbol{V}$ can be estimated from the residuals [3]. In the simulation experiments described in Section III of this paper, the `sig2ar` function of the ARMASA Matlab toolbox [14], was used for this estimation. The estimated covariance matrix will be denoted as $\hat{\boldsymbol{V}}$.

Second, $\hat{\boldsymbol{V}}^{-1}$ is used as weighting matrix in a generalized least squares (GLS) estimator of $\boldsymbol{\theta}$, which results in

$$\hat{\boldsymbol{\theta}}_{\mathrm{GLS}} = (\boldsymbol{X}^T\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{V}}^{-1}\boldsymbol{y} = \hat{\boldsymbol{W}}\boldsymbol{X}^T\hat{\boldsymbol{V}}^{-1}\boldsymbol{y} \quad (8)$$

where the $m \times m$ matrix $\hat{\boldsymbol{W}} = (\boldsymbol{X}^T\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}$ is an estimator of the covariance matrix of $\hat{\boldsymbol{\theta}}_{\mathrm{GLS}}$. Notice that estimator (8) is equivalent to prewhitening the data and model with $\hat{\boldsymbol{V}}^{-1/2}$. That is, with $\tilde{\boldsymbol{X}} = \hat{\boldsymbol{V}}^{-1/2}\boldsymbol{X}$ and $\tilde{\boldsymbol{y}} = \hat{\boldsymbol{V}}^{-1/2}\boldsymbol{y}$ the GLS estimator can be written as

$$\hat{\boldsymbol{\theta}}_{\mathrm{GLS}} = (\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{y}}. \quad (9)$$

In principle, the procedure can be iterated by repeating both steps described above, that is, by re-estimating the covariance matrix $\boldsymbol{V}$ from the residuals

$$\varepsilon_{\mathrm{GLS}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\theta}}_{\mathrm{GLS}} \quad (10)$$

and substituting the result in (8). However, this procedure was not implemented in the simulation experiments described in Section III, since it was observed that iterating did not change the results significantly. Notice that if $\boldsymbol{V}$ is known, an unbiased estimator of σ_e^2 is given by

$$\widetilde{\sigma_e^2} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y})^T\boldsymbol{V}^{-1}$$
$$\times (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y})/(n - m) \quad (11)$$

where $\boldsymbol{W} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ and $(n-m)\widetilde{\sigma_e^2}/\sigma_e^2$ is χ^2 distributed with $n - m$ degrees-of–freedom. If we substitute the estimator $\hat{\boldsymbol{V}}$ for $\boldsymbol{V}$ in (11), we yield the estimator

$$\hat{\sigma_e^2} = \varepsilon_{\mathrm{GLS}}^T\hat{\boldsymbol{V}}^{-1}\varepsilon_{\mathrm{GLS}}/(n - m) \quad (12)$$

of which the statistics are not known exactly. However, as we will see later, the validity of the assumption that the test statistic (29) associated with the widely used F-test (described in Section II-F) has indeed an F-distribution is subject to the validity of the assumption that $\hat{\sigma_e^2}$ has the same distribution as estimator $\widetilde{\sigma_e^2}$. Obviously, this assumption is questionable.

Note that the GLM method described above can be implemented for any AR model order on a voxel by voxel basis. This differs from its implementation in the SPM software package [4], where only a single, iteratively estimated, global AR(1) model for all brain voxels is used.

C. Joint Probability Density Function of the Data

In order to derive the ML estimator of $\boldsymbol{\theta}, \boldsymbol{\alpha}$ and σ_e^2, the joint probability density function (PDF) of the fMRI data is needed. This joint PDF $p(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2)$ can be factorized as

$$p(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2) = p\left(\boldsymbol{y}_{r+1:n}|\boldsymbol{y}_{1:r}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2\right) p\left(\boldsymbol{y}_{1:r}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2\right) \quad (13)$$

with $\boldsymbol{y}_{1:r} = (y_1, \ldots, y_r)^T$ and $\boldsymbol{y}_{r+1:n} = (y_{r+1}, \ldots, y_n)^T$. With (1) and (2) it can be shown that

$$e_t = y_t - \boldsymbol{x}_t\boldsymbol{\theta} + \alpha_1(y_{t-1} - \boldsymbol{x}_{t-1}\boldsymbol{\theta}) + \cdots + \alpha_r(y_{t-r} - \boldsymbol{x}_{t-r}\boldsymbol{\theta}) \quad (14)$$

where $\boldsymbol{x}_t$ denotes the tth row of the design matrix $\boldsymbol{X}$. Therefore, the conditional PDF of the observations $\boldsymbol{y}_{r+1:n}$, given that the first r observations $\boldsymbol{y}_{1:r}$ remain fixed at their observed values, may be written as [10, p. 347]

$$p(\boldsymbol{y}_{r+1:n}|\boldsymbol{y}_{1:r}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2) = \left(\frac{1}{2\pi\sigma_e^2}\right)^{(n-r)/2}$$
$$\times \exp\left(-\frac{1}{2\sigma_e^2}\sum_{t=r+1}^{n}\{y_t - \boldsymbol{x}_t\boldsymbol{\theta} + \alpha_1(y_{t-1} - \boldsymbol{x}_{t-1}\boldsymbol{\theta}) + \cdots \right.$$
$$\left. +\alpha_r(y_{t-r} - \boldsymbol{x}_{t-r}\boldsymbol{\theta})\}^2\right). \quad (15)$$

The joint PDF of the data $\boldsymbol{y}_{1:r}$ may be written as [10, p. 350]

$$p(\boldsymbol{y}_{1:r}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2) = \left(\frac{1}{2\pi\sigma_e^2}\right)^{r/2}|\boldsymbol{V}_r|^{-1/2}$$
$$\times \exp\left(-\frac{1}{2\sigma_e^2}(\boldsymbol{y}_r - \boldsymbol{X}_{1:r}\boldsymbol{\theta})^T\boldsymbol{V}_r^{-1}(\boldsymbol{y}_r - \boldsymbol{X}_{1:r}\boldsymbol{\theta})\right) \quad (16)$$

where $\boldsymbol{X}_{1:r}$ denotes the $r \times m$ matrix consisting of the first r rows of the design matrix $\boldsymbol{X}$. Furthermore, $\boldsymbol{V}_r$ denotes the $r \times r$ covariance matrix of $\boldsymbol{v}_{1:r} = (v_1, \ldots, v_r)^T$ divided by σ_e^2 and $|\boldsymbol{V}_r|$ denotes the determinant of $\boldsymbol{V}_r$. By multiplying the conditional PDF in (15) by (16), the exact joint PDF of the data $\boldsymbol{y}$ may be written as [10]

$$p(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2) = \left(\frac{1}{2\pi\sigma_e^2}\right)^{n/2}|\boldsymbol{V}_r|^{-1/2}$$
$$\times \exp\left(-\widetilde{Q}(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\alpha})/2\sigma_e^2\right) \quad (17)$$

where

$$\widetilde{Q}(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\alpha}) = (\boldsymbol{y}_r - \boldsymbol{X}_{1:r}\boldsymbol{\theta})^T\boldsymbol{V}_r^{-1}$$
$$\times (\boldsymbol{y}_r - \boldsymbol{X}_{1:r}\boldsymbol{\theta}) + Q(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\alpha}) \quad (18)$$

and

$$Q(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{t=r+1}^{n}\{y_t - \boldsymbol{x}_t\boldsymbol{\theta} + \alpha_1(y_{t-1} - \boldsymbol{x}_{t-1}\boldsymbol{\theta})$$
$$+ \ldots + \alpha_r(y_{t-r} - \boldsymbol{x}_{t-r}\boldsymbol{\theta})\}^2 \quad (19)$$

are defined for convenience.

D. Maximum Likelihood Estimator

When the data $\boldsymbol{y}$ are given, the PDF given in (17) is a function of the parameters $\boldsymbol{\alpha}, \boldsymbol{\theta}$ and σ_e^2 only and it is called the likelihood function. In order to compute the likelihood-based tests, the ML estimate of the unknown parameters has to be found, both under the null hypothesis H_0 and the alternative hypothesis H_1. For that purpose, the likelihood function has to be maximized with respect to the unknown parameters $\boldsymbol{\tau} = (\boldsymbol{\theta}^T \boldsymbol{\alpha}^T \sigma_e^2)^T$. Note that maximization of the likelihood function is equivalent to maximization of the (natural) logarithm of the likelihood function

because the logarithmic function is monotomic. It follows from (17) that the natural logarithm of the likelihood function, which is called the log-likelihood function, is given by

$$\ln\left(p\left(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha},\sigma_e^2\right)\right) = -\frac{n}{2}\ln\left(2\pi\sigma_e^2\right)$$
$$-\frac{1}{2}\ln(|\boldsymbol{V}_r|) - \frac{1}{2\sigma_e^2}\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}). \quad (20)$$

The noise variance σ_e^2 can be eliminated from the optimization problem since the value of σ_e^2 that maximizes the likelihood function $p(\boldsymbol{y};\boldsymbol{\alpha},\boldsymbol{\theta},\sigma_e^2)$ can easily be solved from

$$\frac{\partial \ln p(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha},\sigma_e^2)}{\partial \sigma_e^2} = -\frac{n}{2\sigma_e^2} - \frac{1}{2\sigma_e^4}\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}) = 0 \quad (21)$$

and is equal to

$$\hat{\sigma}_e^2 = \frac{1}{n}\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}). \quad (22)$$

Substituting (22) in (17) yields the so-called concentrated likelihood function

$$p(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}) = \left(\frac{n}{2\pi\tilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})}\right)^{n/2}|\boldsymbol{V}_r|^{-1/2}\exp\left(-\frac{n}{2}\right). \quad (23)$$

Notice that $\boldsymbol{V}_r$ depends on the parameters $\boldsymbol{\alpha}$. The ML estimates $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}})$ of the parameters $(\boldsymbol{\alpha}, \boldsymbol{\theta})$ can now be found by maximizing (23) with respect to $(\boldsymbol{\alpha}, \boldsymbol{\theta})$, both with and without the H_0 constraints. The maximization of the likelihood function is a nonlinear optimization problem that can be solved numerically. Substituting $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}})$ for $(\boldsymbol{\alpha}, \boldsymbol{\theta})$ in (22) yields the ML estimate of σ_e^2.

For computational reasons, the logarithm of the concentrated likelihood function

$$\ln(p(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})) = -\frac{n}{2}\ln\left(\frac{2\pi}{n}\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})\right) - \frac{1}{2}\ln(|\boldsymbol{V}_r|) - \frac{n}{2} \quad (24)$$

is maximized. Since numerical optimization is much more efficient when the gradient (and Hessian) is available, one can also compute the first (and second) derivative of the concentrated log-likelihood function with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$.

In the simulation experiments described in Section III of this paper, the ML estimator was implemented in MATLAB, using a built-in unconstrained optimization routine which uses a subspace trust region method and is based on the interior-reflective Newton method [15].

E. Statistical Inference

Brain activation can now be detected on a voxel-by-voxel basis by testing the significance of the task-related parameter(s). To determine whether a voxel is active or not, one can distinguish one-sided and two-sided tests. One-sided tests should be applied when the sign of the activation parameter(s) is known *a priori*. Since this is usually not justified in fMRI experiments [16], we will restrict our analysis to two-sided tests. However, the methods presented can be easily extended to one-sided tests.

Since all considered tests can easily be stated for the multiparameter case, this section is not restricted to single parameter testing, but to a more general linear hypothesis test. Suppose that we wish to test if $\boldsymbol{\tau}$ satisfies the linear equations $\boldsymbol{C}\boldsymbol{\tau} = \boldsymbol{b}$, where $\boldsymbol{C}$ is a known full rank $j \times (m + r + 1)$ matrix and $\boldsymbol{b}$ is a known $j \times 1$ vector. Then a two-sided hypothesis test can be specified by

$$H_0 : \boldsymbol{C}\boldsymbol{\tau} = \boldsymbol{b} \quad (25)$$
$$H_1 : \boldsymbol{C}\boldsymbol{\tau} \neq \boldsymbol{b}. \quad (26)$$

The hypothesis test decides H_0 when $\boldsymbol{C}\boldsymbol{\tau}$ is not statistically significantly different from $\boldsymbol{b}$ and H_1 otherwise. For testing the presence of activation, usually $j = 1$, $\boldsymbol{b} = \boldsymbol{0}$, and $\boldsymbol{C}$ reduces to a row vector in which only the element corresponding to the activation parameter (e.g., θ_1) is nonzero. For some tests, the ML estimator of the parameters with and/or without the constraints imposed by (25) are needed. When we substitute the acquired data (numbers) $\boldsymbol{y}$ in the expression for the joint PDF of the data, given by (17), the resulting function is a function of the unknown parameters $\boldsymbol{\tau}$ only. The ML estimates under H_0 and H_1 are then given by, respectively

$$\hat{\boldsymbol{\tau}}_0 = \arg\max_{\boldsymbol{\tau}} p(\boldsymbol{y}|\boldsymbol{\tau}), \text{ subject to } \boldsymbol{C}\boldsymbol{\tau} = \boldsymbol{b} \quad (27)$$
$$\hat{\boldsymbol{\tau}}_1 = \arg\max_{\boldsymbol{\tau}} p(\boldsymbol{y}|\boldsymbol{\tau}). \quad (28)$$

In the next section, the GLM based F-test is reviewed. Subsequently, three likelihood-based tests are described: the GLRT, the Rao test, and the Wald test. All these tests are based on the joint PDF of the data, described in Section II-C. Furthermore, the Rao test and the Wald test are based on the Fisher information matrix, derived in the Appendix 1.

F. F-Test

For the GLM based F test, where only the linear regression parameters $\boldsymbol{\theta}$ can be tested, $\boldsymbol{\tau} = \boldsymbol{\theta}$ and $\boldsymbol{C}$ is a $j \times m$ matrix. The test statistic of the F-test is then given by

$$T_F = \frac{(\boldsymbol{C}\hat{\boldsymbol{\theta}}_{\text{GLS}} - b)^T(\boldsymbol{C}\hat{\boldsymbol{W}}\boldsymbol{C}^T)^{-1}(\boldsymbol{C}\hat{\boldsymbol{\theta}}_{\text{GLS}} - b)}{\hat{\sigma}_e^2} \quad (29)$$

where $\hat{\boldsymbol{\theta}}_{\text{GLS}}$ denotes the GLS estimator and $\hat{\sigma}_e^2$ is given by (12). Under H_0, the test statistic T_F has approximately an F distribution with j and $n - m$ degrees-of-freedom. If $\boldsymbol{V}$ would be known, T_F would be exactly F distributed with the specified degrees-of-freedom. The F-test, decides H_1 if $T_F > \gamma$, with γ some user specified threshold. This threshold is usually computed using the F distribution and balancing the false alarm rate (probability of deciding H_1 when H_0 is true) against the detection rate (probability of deciding H_1 when H_1 is true).

G. GLRT

The generalized likelihood ratio (GLR) is given by [9]

$$\lambda = \frac{p(\boldsymbol{y}|\hat{\boldsymbol{\tau}}_0)}{p(\boldsymbol{y}|\hat{\boldsymbol{\tau}}_1)}. \quad (30)$$

The GLRT principle now states that H_0 is to be rejected if and only if $\lambda \leq \lambda_0$, where λ_0 is some user specified threshold. It can

be shown that, asymptotically (i.e., for $n \to \infty$), the modified GLR test statistic

$$T_{\text{LR}} = -2 \ln \lambda \qquad (31)$$

possesses a χ_j^2 distribution with j degrees-of-freedom when H_0 is true.

Notice that, if $\boldsymbol{X}$ is not of full rank, the optimization implicit in (30) might be numerically difficult. In that case, the use of Bayes factors may be considered [17].

H. Wald Test

The Wald test statistic is given by [9]

$$T_{W2} = (\boldsymbol{C}\hat{\boldsymbol{\tau}}_1 - \boldsymbol{b})^T (\boldsymbol{C}\boldsymbol{F}^{-1}(\hat{\boldsymbol{\tau}}_1)\boldsymbol{C}^T)^{-1} (\boldsymbol{C}\hat{\boldsymbol{\tau}}_1 - \boldsymbol{b}) \qquad (32)$$

where $\boldsymbol{F}^{-1}(\hat{\boldsymbol{\tau}}_1)$ is the inverse of the Fisher information matrix, evaluated at $\hat{\boldsymbol{\tau}}_1$. The (two-sided) Wald test decides H_1 if $T_{W2} > \gamma$, where γ is some user specified threshold. Asymptotically, the test statistic T_{W2} has a χ_j^2 distribution, that is, a χ^2 distribution with j degrees-of-freedom under H_0.

I. Rao Test

The Rao test statistic is given by [9]

$$T_{R2} = \left.\frac{\partial \ln p(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\right|_{\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}_0}^T \boldsymbol{R}\boldsymbol{C}\boldsymbol{F}^{-1}(\hat{\boldsymbol{\tau}}_0)\boldsymbol{C}^T\boldsymbol{R}^T \left.\frac{\partial \ln p(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\right|_{\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}_0} \qquad (33)$$

where $\boldsymbol{F}^{-1}(\hat{\boldsymbol{\tau}}_0)$ is the inverse of the Fisher information matrix, evaluated at $\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}_0$ and the $(m + r + 1) \times j$ matrix $\boldsymbol{R}$ is the pseudoinverse of $\boldsymbol{C}$. The Rao test decides H_1 if $T_{R2} > \gamma$, where γ is some user specified threshold. Asymptotically, the test statistic T_{R2} has a χ_j^2 distribution under H_0.

J. Discussion

Knowledge of the PDF of the test statistic under H_0 allows one to compose tests with a desired *false alarm rate*. The false alarm rate is the probability that the test will decide H_1 when H_0 is true. The *detection rate* is the probability that the test will decide H_1 when H_1 is true. Throughout this paper, we will denote the false alarm rate and the detection rate by P_f and P_d, respectively. Furthermore, a test has the so-called constant false-alarm rate (CFAR) property if the threshold required to maintain a constant P_f can be found independently of the signal-to-noise ratio (SNR) [9], which is usually unknown beforehand. Since the asymptotic PDFs of the likelihood-based test statistics discussed in this section are known and parameter and SNR invariant, the tests will all have the CFAR property at least asymptotically. Whether or not the tests have the CFAR property for a finite number of observations can be found out by means of simulations. For more details on likelihood-based tests, see [9].

III. EXPERIMENTS AND RESULTS

Experimental fMRI data sets were obtained from a healthy human volunteer, male, age 32 years. An informed consent was signed by the participant. All human experiments were performed on a 1.5 T scanner with high-performance 40 mT/m gradients (Siemens Sonata, Erlangen, Germany). Gradient-recalled multishot EPI sequences (TE 50 ms, TR 3000 ms) were used with 30 slices covering the whole brain. The voxel dimensions were 3 mm $\times$ 3 mm $\times$ 3 mm. Head movement was restricted by foam-padded cushions and the subject wore earplugs and noise reducing headphones throughout the entire experiment.

A. Estimation of the Order of the AR and Trend Models

Experimental fMRI resting (null) data were used to determine relevant orders of the AR noise process and trend model.

The trend model we used was a polynomial of order m (to be selected)

$$\text{trend} = \sum_{k=0}^{m} b_k t^k. \qquad (34)$$

The noise was modeled by an AR process as in (2). AR models of orders 0–8 and trend models of orders 0–4 were evaluated for a random selection of 10 000 brain voxel traces with 90 time points from an fMRI null data set. The polynomial order and AR order of each voxel was selected using Akaike's information criterion (AIC) [18], where a penalty factor of 3 instead of 2 was chosen [19]

$$\text{AIC}_3 = 2 \ln p(\boldsymbol{y}|\hat{\boldsymbol{\tau}}_1) + 3(m + r). \qquad (35)$$

A histogram of the selected orders is plotted in Fig. 1(a). For most traces, the selected order of the polynomial was 0 or 1 (linear trend). Also, for most traces the AR order selected was between 0 and 4. Due to the statistics involved in the order selection, it is unlikely that for all traces the selected order of the model equals the model of the underlying process. In order to get an impression of the orders selected for a given model, a simulation of AR(4) noise with a linear trend was set up. The simulation also had 90 time points per voxel trace. The parameters of the AR noise generating process and the trend used are given by

$$\boldsymbol{\text{AR}} = \begin{bmatrix} 1 & -0.177 & -0.164 & -0.115 & -0.130 \end{bmatrix}$$
$$\text{trend} = 0.2t \qquad (36)$$

where t is the time index of the simulated volume. In Fig. 1(b), the results of the order selection, again using the AIC order selection procedure with a penalty factor 3, of the simulated data are plotted. The parameters of the simulation were chosen to give approximately the same selection results, as can be seen by comparing Fig. 1(a) with (b).

The most interesting parts of the histograms are those parts where the order of the trend model exceeds 1 or where the order of the AR model exceeds 4. In these parts, the orders are selected approximately equally often from the measured and simulated data. Therefore, we think that a model with a linear trend (polynomial order 1) and an AR(4) noise model gives a sufficiently accurate description of the data. A linear trend and AR models up to order 4 were therefore used in the simulation experiments of the next section. Note that we do not claim that the process
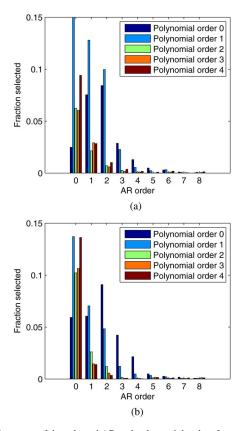
(a)

(b)

Fig. 1. Histogram of the selected AR and polynomial orders from measured as well as simulated data, using the AIC criterion with penalty factor 3. See (36) for the parameters of the simulation model. The 10 000 traces inside the brain were used for the measurements and 10 000 generated independent traces were used for the simulation. Note that for AR orders 5–8 the selection frequency for the measured and simulated data is approximately equal. (a) Measured fMRI data. (b) Simulated fMRI data.
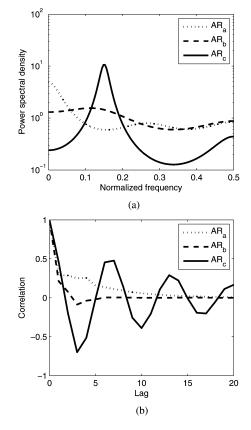

(a)

(b)

Fig. 2. (a) Power spectral density of the noise processes used in the simulations as function of the normalized frequency. (b) Correlation functions of the noise processes used in the simulations.

underlying the data actually consists of a linear trend and AR(4) noise process, but only that there is not enough evidence to assume that higher order parameters are significantly present in the data. For AR orders lower than 4 or trend orders lower than 1 this could not be concluded since the histograms were always substantially different.

This analysis has been performed for several other data sets (results not shown). In these other data sets, the linear trend was generally present, but for some data sets AR(3) models turned out to give a sufficiently accurate description of the data. AR orders higher than 4 or trend orders higher than 1 were not needed to describe the data sets considered.

B. Simulation Experiments

Simulation experiments were set up to detect brain activation. A simple block design activation scheme was used in which traces of 100 time points were generated with period equal to 20 (10 stimulus on, 10 stimulus off). This block stimulus was convolved with a standard HRF function [3] (fmridesign with default parameters and $TR = 2$) to get the activation pattern. Also, for each voxel a small linear trend increasing 0.1 per time point and a baseline of 100 were introduced. A linear trend model $(m = 1)$ was used in the model, as well as the activation pattern. Note that when the trend (including baseline) is modeled correctly (as it is here), changing trend parameters does

not influence the value of the likelihood function in its maximum, and thus the likelihood-based test values are independent of the actual trend parameters. Several different noise models, based upon results of the previous section and selected to investigate different properties of the estimators, were used to generate fMRI data. These noise models were

$$\mathbf{AR}_a = [1 \quad -0.177 \quad -0.164 \quad -0.115 \quad -0.130]$$
$$(37)$$
$$\mathbf{AR}_b = [1 \quad -0.208 \quad -0.056 \quad 0.115] \quad (38)$$
$$\mathbf{AR}_c = [1 \quad -0.400 \quad 0.118 \quad 0.568] \quad (39)$$

The power spectral density (PSD) of these noise processes is plotted in Fig. 2(a) and the correlation functions in Fig. 2(b). $\mathbf{AR}_a$ is a low-frequency colored noise process. $\mathbf{AR}_b$ is almost white, but has slight excess power near one of the main frequencies present in the stimulus used. $\mathbf{AR}_c$ is stronger colored, also with the maximum power near one of the main frequencies present in the stimulus used. The simulations of the null-data (i.e., data containing no activation) with model $\mathbf{AR}_a$ used 20 000 independent traces, the simulations with model $\mathbf{AR}_b$ used 100 000 independent traces, and the simulations with model $\mathbf{AR}_c$ used 100 000 and 40 000 independent traces for the lengths 100 and 2500, respectively. To investigate the effect of changing SNR in the simulation experiments, the amplitude of the activation pattern was changed from 0 till 1.2, while the noise standard deviation was fixed to 1, which
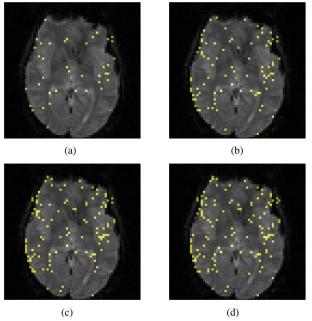
Fig. 3. In these figures, the yellow dots are the voxels that were detected as active in real fMRI null data, by each of the four different test statistics. The threshold was computed to have a theoretical asymptotic P_f of 1% for all test statistics. An AR(4) model was used for the noise model, and the fMRI time series had a length of 80 points. Note the difference in false alarm rate of the different test statistics. For this asymptotic P_f, even the Rao test has an observed P_f that is 2.0 (± 0.7) times higher than desired. (a) Test: Rao. (b) Test: LR. (c) Test: GLM. (d) Test: Wald.

are realistic values of SNR in fMRI [20]. At each activation level, the time courses of 1000 voxels were generated. In this simulation experiment, the null hypothesis is given by (25), with $\boldsymbol{b} = \boldsymbol{0}$ and $\boldsymbol{C} = [1\,0\,\dots\,0]$ (the first column of $\boldsymbol{X}$ contains the activation related regressor).

1) Null Distribution: The observed distribution of the test statistics under H_0, the null distribution, was compared with the theoretically known asymptotic null distribution. This is important since the asymptotic distribution might be used to compute thresholds for a given false alarm rate P_f. The comparison of the distributions was made by observing the actual P_f of null data as a function of the theoretical asymptotic P_f. To help visualizing this, Fig. 3 shows a slice with the (falsely) detected active voxels in a real fMRI null experiment.

2) Activation Sensitivity of the Test Statistics: For a fixed P_f of 0.1%, the detection rates (P_d) of the different tests were compared. Since it was observed that the observed null distribution of the test statistics was not equal to the asymptotic distribution, a correction was needed in order to make a fair comparison of the different test statistics and models. Therefore, in all experiments where different P_d values were compared, the observed null distribution of the simulated null data was used to compute threshold values to obtain a specified P_f.

C. Results of the Simulation Experiments

1) Null Distribution: The thresholds for detecting activation can be computed by using the theoretical asymptotic distribution of the test statistics under H_0. However, this does not necessarily lead to an accurate P_f for time series with a limited trace length. In Fig. 4, the observed P_f of the different tests,
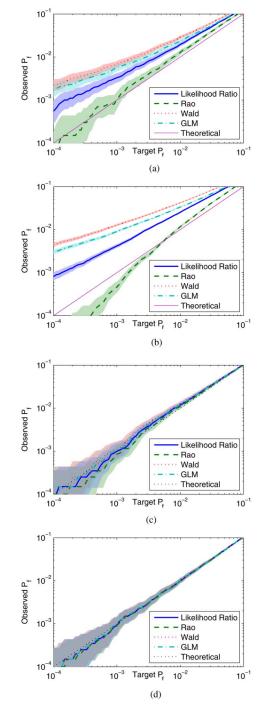

Fig. 4. Logarithmic plot of P_f as a function of the P_f computed from the asymptotic distribution for the different test statistics. The diagonal corresponds with the asymptotic distribution. The shaded areas indicate the 95% confidence regions of the observed P_f as computed from the binomial counting statistics. The uncertainty indicated by these regions is caused by the finite number of voxel time series used in the simulations. For each simulation the noise contaminated time trace was modelled by a linear trend and an AR(4) model. (a) Noise process $\mathbf{AR}_a$, trace length: 100. (b) Noise process $\mathbf{AR}_b$, trace length: 80. (c) Noise process $\mathbf{AR}_a$, trace length: 500. (d) Noise process $\mathbf{AR}_a$, trace length: 2500.

with a linear trend and AR(4) noise model, are plotted as a function of the theoretical asymptotic P_f. Model $\mathbf{AR}_a$ was used to generate the noise for Fig. 4(a)–(d) and model $\mathbf{AR}_b$ was used to generate the noise for Fig. 4(b). The first thing to note from Fig. 4(a) and (b) is that the distribution of the test statistics has

not reached the asymptotic distribution for 80 or 100 samples per time series.

It can be noted from these figures as well that the observed P_f is larger for the Wald test than for the LR test, which in turn is larger than the P_f of the Rao test statistic. The P_f of the GLM test is somewhere in between. For linear models the ordering of the Wald, LR, and Rao tests statistics can be proven to be as observed here, (see [21, p. 231]). Since the models used in this paper are nonlinear, this ordering might be different. However, we did not observe this in any data analyzed. Usually, as is the case in the presented figures, the Rao test statistic approximates the asymptotic distribution most accurately, especially in the most relevant region of false alarm rates between 0.01 and 0.001. However, even in this interval, the Rao test statistic has an actual P_f that might differ from the asymptotic value by a factor larger than 2. So for data series with 80–100 time points the asymptotic distribution cannot be reliably used to determine the thresholds of the test statistics.

When the length of the data series is increased, the asymptotic distribution is approached much more accurately, as is demonstrated in Fig. 4(c) and (d) were the trace length was 500 and 2500, respectively.

The main contribution to the difference between the observed and asymptotic distributions is the finite length of the time series. However, changing the regression model or noise process influences the distribution of the test statistic slightly. Therefore, the observed distributions shown in Fig. 4 cannot be reliably used for all different regression models and noise sources.

When the order of the noise model is below the order needed to give an accurate description of the data, the null distribution of the test statistics deviates from the asymptotic distribution as well, as might be expected since the model of the data is incomplete. In fact, this deviation can easily be much larger than the deviation caused by short data series. A demonstration of this is shown in Fig. 5, where the observed P_f is much larger or smaller than the P_f set using the asymptotic distribution for the AR(0), AR(1), and AR(2) noise models.

Apart from simulation experiments, the P_f of the tests under concern were computed for experimental fMRI null data sets. The asymptotic theoretical distribution was used to obtain the thresholds for the tests with a theoretical P_f of 1%. Fig. 3 shows the voxels that are detected as active with this threshold. For this threshold, the observed P_f of the Rao test statistic [Fig. 3(a)] is $2.0 (\pm 0.7)$ times the asymptotic theoretical P_f. As is clear from Fig. 3(b)–(d), the LR, GLM, and Wald tests were observed to have even higher false alarm rates of approximately 4.4%, 5.7%, and 6.4%, respectively. This clearly demonstrates the need for correction of the P_f to obtain reliable activation detection.

2) Activation Sensitivity of the Test Statistics: In the second simulation experiment, the activation sensitivity of the test statistics is investigated. The results are plotted in Fig. 6. An upper limit to the detection rate is included in these plots. This upper limit is the theoretical detection rate for the case in which the noise generating AR process and the noise variance σ_e^2 are known. In this case, all evaluated test statistics are equivalent and equal to

$$T_t = \frac{(C\hat{\theta} - b)^T (CWC^T)^{-1} (C\hat{\theta} - b)}{\sigma_e^2} = \frac{\hat{\theta}_1^{\,2}}{\sigma_{\theta_1}^2} \qquad (40)$$
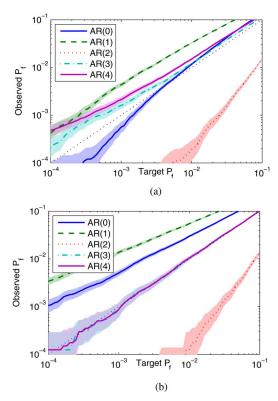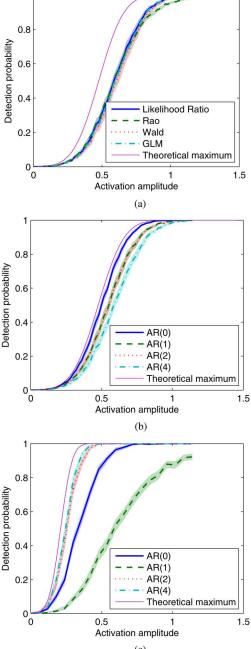

Fig. 5. The observed distributions of the LR test statistic for different AR orders in the model. The axes in this figure are the same as in Fig. 4. Note that even for long time series (b) the asymptotic distribution is not reached for the AR(1) and AR(2) models, since the noise is generated by an AR(3) process. Also note that for short time series none of the tests reaches the asymptotic χ_1^2 distribution. (a) Noise process $\mathbf{AR}_c$, 100 volumes. (b) Noise process $\mathbf{AR}_c$, 2500 volumes.

still with $\mathbf{b} = \mathbf{0}$, $j = 1$, and $\mathbf{C} = [1\ 0\ \dots\ 0]$. Note that $\sigma_{\theta_1}^2 = \sigma_e^2 W_{1,1}$ is the variance of $\hat{\theta}_1$. When the noise process is known, the tests are optimally sensitive, $\hat{\theta}_1$ will be normally distributed with mean value a (denoting the activation level used in the simulation), and the test statistic (40) has a noncentral chi-squared distribution with 1 degree-of-freedom and noncentrality parameter $\lambda = ((a)/(\sigma_{\theta_1}))^2$. The threshold value t_t, which can be computed from $\Pr(T_t > t_t; H_0) = \alpha$, can be used to compute the detection probability at each activation level, $p_d = \Pr(T_t > t_t; H_1)$.

In practice, the coloring of the noise and the noise variance are not known. Therefore, this theoretical limit is unreachable. In Fig. 6(a), it is visible that the Rao test statistic generally is the least sensitive to activation. The other three test statistics, LR, Wald, and GLM have approximately equal detection rates, although, by evaluating many simulations, it turns out that the LR test often has a slightly higher detection rate. However, it is far more important to use the correct noise model, as can be seen in Fig. 6(b) and (c). These figures contain the results of the LR test statistic. The other statistics are almost overlapping and are therefore not plotted. When no or little color is present in the noise, as is the case with noise process $\mathbf{AR}_b$, the optimal detection rate can be reached by an AR(0) model [Fig. 6(b)]. This is expected, since this is the model that can describe the data accurately with the lowest number of parameters. However, when the coloring of the data is stronger, as it is in noise model $\mathbf{AR}_c$, which is used for Fig. 6(c), the reduced precision due to the extra parameters of the

Fig. 6. Activation sensitivity of the different test statistics, with corrected false positive level. The theoretical maximum is reached when the true noise model and the noise variance are known. The number of timepoints is 80 for the (a) and (b) and 100 for (c). (a) The Rao test is less sensitive in detecting activation and that the LR test is slightly more sensitive than the others. (b) and (c) The order of the noise model should be chosen carefully so as to avoid inferior activation sensitivity. (a) Noise process $\mathbf{AR}_b$, modelled with AR(4) noise model. (b) Noise process $\mathbf{AR}_b$, LR test statistic. (c) Noise process $\mathbf{AR}_c$, LR test statistic.

AR models is more than compensated for by an increase in accuracy of the model and thus, for a given activation amplitude, the detection rates of the AR(2) and AR(4) models are higher. So for (nearly) white noise processes, using a high order AR model (AR(4)) results in a "modest" performance loss, but, using a low order AR model (AR(0)) when there is strongly colored noise results in a "large" performance loss. This suggests that the order of the AR model should not be chosen too low.

IV. CONCLUSION

In this paper, likelihood-based tests for the detection of functional brain activity were presented. In contrast to the GLM tests, the proposed likelihood ratio tests allow direct incorporation of colored noise and do not require an explicit prewhitening step. Simulation results showed that the detection rate of the proposed likelihood ratio test is slightly, but significantly improved compared to the detection rate of the currently popular GLM based tests. Furthermore, it was demonstrated that thresholds based on theoretical, asymptotically valid null distributions of test statistics cannot be reliably used when the data series does not have more than a few hundred time points per voxel. In that case, thresholds obtained from observed null distributions should be used instead. Finally, it was shown that undermodeling of the (correlation structure of the) noise leads to inferior test results.

APPENDIX

For the Rao and Wald tests the Fisher information matrix is needed. Therefore, the Fisher score vector and Fisher information matrix are computed in this Appendix.

The Fisher score vector of the data set $\boldsymbol{y}$ with respect to the parameters $\boldsymbol{\tau} = (\boldsymbol{\theta}^T \boldsymbol{\alpha}^T \sigma_e^2)^T$ is defined as the $(m + r + 1) \times 1$ vector

$$s(\boldsymbol{\tau}) = \frac{\partial \ln p}{\partial \boldsymbol{\tau}} \qquad (41)$$

with p the joint PDF of the observations described by (17). It can be shown that the expectation of the Fisher score (evaluated at the true values of the parameters) is equal to zero [22], that is

$$\mathbb{E}[s(\boldsymbol{\tau})] = \boldsymbol{0} \qquad (42)$$

with $\boldsymbol{0}$ the $(m + r + 1) \times 1$ null vector. The $(m + r + 1) \times (m + r + 1)$ covariance matrix of the Fisher score is therefore given by [22]

$$\boldsymbol{F} = \mathbb{E}[s(\boldsymbol{\tau}) s^T(\boldsymbol{\tau})] = \mathbb{E}\left[\frac{\partial \ln p}{\partial \boldsymbol{\tau}} \left(\frac{\partial \ln p}{\partial \boldsymbol{\tau}}\right)^T\right]. \qquad (43)$$

This covariance matrix is called the Fisher information matrix [22]. It can be shown that under certain regularity conditions $\boldsymbol{F}$ may alternatively be written as

$$\boldsymbol{F} = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\tau}^2}\right]. \qquad (44)$$

The Fisher matrix may be written in the form

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\alpha}} & \boldsymbol{F}_{\boldsymbol{\theta}\sigma_e^2} \\ \boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\theta}} & \boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} & \boldsymbol{F}_{\boldsymbol{\alpha}\sigma_e^2} \\ \boldsymbol{F}_{\sigma_e^2\boldsymbol{\theta}} & \boldsymbol{F}_{\sigma_e^2\boldsymbol{\alpha}} & \boldsymbol{F}_{\sigma_e^2\sigma_e^2} \end{pmatrix} \qquad (45)$$

where

$$\boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\theta}^2}\right] \qquad \boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\alpha}} = \boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\theta}}^T = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\theta}\partial \boldsymbol{\alpha}}\right]$$

$$\boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\alpha}^2}\right] \qquad \boldsymbol{F}_{\boldsymbol{\theta}\sigma_e^2} = \boldsymbol{F}_{\sigma_e^2\boldsymbol{\theta}}^T = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\theta}\partial \sigma_e^2}\right]$$

$$\boldsymbol{F}_{\sigma_e^2\sigma_e^2} = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial (\sigma_e^2)^2}\right] \qquad \boldsymbol{F}_{\boldsymbol{\alpha}\sigma_e^2} = \boldsymbol{F}_{\sigma_e^2\boldsymbol{\alpha}}^T = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\alpha}\partial \sigma_e^2}\right]$$

with $\boldsymbol{F_{\theta\theta}} \in \mathbb{R}^{m \times m}$, $\boldsymbol{F_{\alpha\alpha}} \in \mathbb{R}^{r \times r}$, $\boldsymbol{F_{\sigma_e^2 \sigma_e^2}} \in \mathbb{R}^{1 \times 1}$, $\boldsymbol{F_{\theta\alpha}} \in \mathbb{R}^{m \times r}$, $\boldsymbol{F_{\alpha\theta}} \in \mathbb{R}^{r \times m}$, $\boldsymbol{F_{\theta\sigma_e^2}} \in \mathbb{R}^{m \times 1}$. It can be shown that all elements of $\boldsymbol{F_{\theta\alpha}}$, $\boldsymbol{F_{\alpha\theta}}$, $\boldsymbol{F_{\alpha\sigma_e^2}}$, $\boldsymbol{F_{\sigma_e^2\alpha}}$, $\boldsymbol{F_{\theta\sigma_e^2}}$ and $\boldsymbol{F_{\sigma_e^2\theta}}$ are equal to zero. This means that (45) simplifies to

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F_{\theta\theta}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F_{\alpha\alpha}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{F_{\sigma_e^2\sigma_e^2}} \end{pmatrix}. \tag{46}$$

It can be shown that

$$\boldsymbol{F_{\theta\theta}} = \frac{1}{\sigma_e^2} \boldsymbol{X}_{1:r}^T \boldsymbol{V}_r^{-1} \boldsymbol{X}_{1:r} + \frac{1}{\sigma_e^2} \sum_{t=r+1}^{n} \boldsymbol{X}_{t:t-r}^T \boldsymbol{A} \boldsymbol{X}_{t:t-r} \tag{47}$$

where

$$\boldsymbol{X}_{t:t-r} = \begin{pmatrix} \boldsymbol{x}_t \\ \boldsymbol{x}_{t-1} \\ \vdots \\ \boldsymbol{x}_{t-r} \end{pmatrix} \tag{48}$$

and

$$\boldsymbol{A} = \begin{pmatrix} 1 & \boldsymbol{\alpha}^T \\ \boldsymbol{\alpha} & \boldsymbol{\alpha}\boldsymbol{\alpha}^T \end{pmatrix}. \tag{49}$$

For activation detection only $\boldsymbol{\theta}$ is used in the test statistics. Therefore, and because $\boldsymbol{F}$ is block diagonal, the only term of $\boldsymbol{F}$ that is needed is $\boldsymbol{F_{\theta\theta}}$.

ACKNOWLEDGMENT

REFERENCES

[1] M. S. Cohen, "Real-time functional magnetic resonance imaging," *Methods*, vol. 25, pp. 201–220, 2001.

[2] K. J. Friston, A. P. Holmes, K. J. Worsley, J. B. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Map.*, vol. 2, pp. 189–210, 1995.

[3] K. J. Worsley, C. H. Liao, J. Aston, V. Petre, G. H. Duncan, F. Morales, and A. C. Evans, "A general statistical analysis for fMRI data," *NeuroImage*, vol. 15, pp. 1–15, 2002.

[4] SPM5. Wellcome Trust Centre for Neuroimaging [Online]. Available: http://www.fil.ion.ucl.ac.uk/spm/software/spm5/

[5] K. J. Worsley and K. J. Friston, "Analysis of fMRI time-series revisited—Again," *NeuroImage*, vol. 2, pp. 173–181, 1995.

[6] M. W. Woolrich, B. D. Ripley, J. M. Brady, and S. M. Smith, "Temporal autocorrelation in univariate linear modelling of fMRI data," *NeuroImage*, vol. 14, no. 6, pp. 1370–1386, 2001.

[7] J. L. Marchini and S. M. Smith, "On bias in the estimation of autocorrelations for fMRI voxel time-series analysis," *IEEE Trans. Med. Imag.*, vol. 18, no. 1, pp. 83–90, Jan. 2003.

[8] F. Y. Nan and R. D. Nowak, "Generalized likelihood ratio detection for fMRI using complex data," *IEEE Trans. Med. Imag.*, vol. 18, no. 4, pp. 320–329, Apr. 1999.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II Detection Theory*. Upper Saddle River, NJ: Prentice Hall, 1998.

[10] M. B. Priestley, *Spectral Analysis and Time Series*. London, U.K.: Academic, 1981.

[11] S. M. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.

[12] J. D. Carew, G. Wahba, X. Xie, E. V. Nordheim, and M. E. Meyerand, "Optimal spline smoothing of fMRI time series by generalized cross-validation," *NeuroImage*, vol. 18, no. 4, pp. 950–961, 2003.

[13] A. van den Bos, "Parameter estimation," in *Handbook of Measurement Science*, P. H. Sydenham, Ed. Chichester, U.K.: Wiley, 1982, vol. 1, ch. 8, pp. 331–377.

[14] P. M. T. Broersen, "Automatic spectral analysis with time series models," *IEEE Trans. Instrum. Meas.* vol. 51, no. 2, pp. 211–216, Apr. 2002.

[15] T. F. Coleman and Y. Li, "An interior, trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optimizat.*, vol. 6, pp. 418–445, 1996.

[16] E. Vandervliet, G. Nagels, A. Heinecke, W. Van Hecke, A. Leemans, J. Sijbers, and P. M. Parizel, "On the cause and mechanisms of the negative bold response in fmri," in *Proc. 23rd Annu. Sci. Meeting Eur. Soc. Magn. Reson. Med. Biol.*, Warsaw, Poland, 2006, p. 308.

[17] S. N. Goodman, "Toward evidence-based medical statistics. 2: The bayes factor," *Ann. Internal Med.*, vol. 130, no. 12, pp. 1005–1013, 1999.

[18] B. A. Ardekani, J. Kershaw, K. Kashikura, and I. Kanno, "Activation detection in functional MRI using subspace modeling and maximum likelihood estimation," *IEEE Trans. Med. Imag.*, vol. 18, no. 2, pp. 246–254, Feb. 1999.

[19] P. M. T. Broersen, "Finite sample criteria for autoregressive order selection," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3550–3558, Dec. 2000.

[20] Z. Zhanga, A. H. Andersena, M. J. Avisonb, G. A. Gerhardta, and D. M. Gasha, "Functional MRI of apomorphine activation of the basal ganglia in awake rhesus monkeys," *Brain Res.*, vol. 852, no. 2, pp. 290–296, Jan. 2000.

[21] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*. New York: John Wiley and Sons, 1989.

[22] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 2nd ed. New York: Hafner, 1967, vol. 2.