A Likelihood Ratio Test for Functional MRI Data Analysis to Account for Colored Noise

[1] University of Antwerp, Vision Lab, Groenenborgerlaan 171, U316, B-2020 Antwerpen, Belgium

[2] Delft University of Technology, Delft Center for Systems and Control, Mekelweg 2, 2628 CD Delft, The Netherlands

Abstract. Functional magnetic resonance (fMRI) data are often corrupted with colored noise. To account for this type of noise, many prewhitening and pre-coloring strategies have been proposed to process the fMRI time series prior to statistical inference. In this paper, a generalized likelihood ratio test for brain activation detection is proposed in which the temporal correlation structure of the noise is modelled as an autoregressive (AR) model. The order of the AR model is determined from experimental null data sets. Simulation tests reveal that, for a fixed false alarm rate, the proposed test is slightly (2-3%) better than current tests incorporating colored noise in terms of detection rate.

1 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 neural activity [1], fMRI can localize brain regions that show significant neural activity upon stimulus presentation. 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, the most common approach is to model the time series of fMRI data by a general linear model (GLM) disturbed by Gaussian distributed noise [2,3]. Potential time trends can be included in the linear model by adopting extra linear terms. The model contains one or more activation related parameters of interest as well as nuisance parameters. Statistical parametric maps (SPMs) are obtained by testing the significance of the activation related parameter(s) of the linear model 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 parameters).

Current methods deal with temporally correlated noise by prewhitening the data based on the estimated correlation matrix of the noise [3]. This correlation matrix is 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 [4]. Since an estimate of the correlation matrix instead of

J. Blanc-Talon et al. (Eds.): ACIVS 2005, LNCS 3708, pp. 538–546, 2005.

the (unknown) true correlation matrix of the noise is used for prewhitening the data, the assumption that the test statistic has a Student's t distribution (upon which inference on the significance of the response is based) is only approximately valid. Obviously, this fact may harm the performance of the test.

In this paper, an alternative approach is proposed. This approach is also based on a general linear model with correlated noise modelled 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 (LF) that describes the statistics of the data including the temporal correlation structure of the noise. No approximations are made. The order of the AR process, which is fixed in the proposed test, is determined from practical null data sets (acquired in the absence of activity). The performance of the proposed tests is evaluated in terms of detection rate and false alarm rate properties.

The paper is organized as follows. In Section 2.1 and Section 3, statistical inference incorporating colored noise model is reviewed. Section 4 describes a novel approach for the construction of a statistical test that also accounts for colored noise. Simulation and experimental results are presented in Section 5.

2 Statistical Inference Incorporating Colored Noise

2.1 The Statistical Model of the fMRI Time Series

An fMRI time series $\boldsymbol{y} = (y_1, ..., y_n)^T$ (the superscript T denotes matrix transposition) of equidistant observations can in general be modelled as [2,5]

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

in which $\boldsymbol{X}$ is an $n \times m$ design matrix. 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 parameters and $\boldsymbol{v}$ is an $n \times 1$ vector that represents stochastic noise contributions. The noise is modelled as a stationary stochastic AR process of order p (i.e., an AR(p) process):

$$v_t + \alpha_1 v_{t-1} + \alpha_2 v_{t-2} + \cdots + \alpha_p v_{t-p} = e \tag{2}$$

with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^T$ the vector of AR parameters and e 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, i.e., $\sigma_e^2\boldsymbol{V} = \mathbb{E}[\boldsymbol{v}\boldsymbol{v}^T]$ with $\mathbb{E}$ the expectation operator. For observations of stationary stochastic processes, the covariance matrix of the AR(p) process v_t may be written as

$$\sigma_e^2\boldsymbol{V} = \sigma_v^2 \begin{pmatrix} \rho(0) & \rho(1) & \ldots & \rho(n-1) \\ \rho(1) & \rho(0) & \ldots & \rho(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(n-1) & \rho(n-2) & \ldots & \rho(0) \end{pmatrix} \tag{3}$$

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 the AR parameters through the Yule Walker relations [6]:

$$\rho(k) + \alpha_1\rho(k-1) + \cdots + \alpha_p\rho(k-p) = 0, \quad k > 0, \quad \rho(-k) = \rho(k). \tag{4}$$

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 [4,7]. The validity of the model will be assessed using experimental data in section 5.1.

In this paper, the noise is assumed to be Gaussian distributed. Although magnitude MR data are known to be Rician distributed, the Rice distribution is nearly Gaussian at high SNR [8]. Hence, the test derived in this paper will only be valid for high SNR fMRI magnitude data (i.e., SNR>10).

2.2 Statistical Inference

In the next two sections, two-sided as well as one-sided hypothesis testing will be considered. If the test is two-sided, the null hypothesis H_0 that the task-related i^{th} component θ_i of $\boldsymbol{\theta}$ equals zero is tested against the alternative hypothesis H_1 that $\theta_i \neq 0$. If it is known that $\theta_i > 0$ (under H_1), one may use a one-sided test in which H_0 that $\theta_i = 0$ is tested against H_1 where $\theta_i > 0$:

	H_0	H_1
one-sided test	$\theta_i = 0$	$\theta_i > 0$ or $\theta_i < 0$
two-sided test	$\theta_i = 0$	$\theta_i \neq 0$

3 The Common GLM Approach

The widely used GLM approach 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 Eq. (1) to the data $\boldsymbol{y}$. A closed form expression of this so-called ordinary least squares (OLS) estimator is given by:

$$\widehat{\boldsymbol{\theta}}_{\text{OLS}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}. \tag{5}$$

Although not fully efficient, this estimator is unbiased [9]. Therefore, the residuals $\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}_{\text{OLS}}$ 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(p) model, the parameters of this model and hence the matrix $\boldsymbol{V}$ can be estimated from the residuals [3]. The estimated covariance matrix will be denoted as $\widehat{\boldsymbol{V}}$. Second, $\widehat{\boldsymbol{V}}^{-1}$ is used as weighting matrix in a generalized least squares (GLS) estimator of $\boldsymbol{\theta}$, which results in:

$$\widehat{\boldsymbol{\theta}}_{\text{GLS}} = \widehat{\boldsymbol{W}}\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{y} \tag{6}$$

where the $m \times m$ matrix $\widehat{\boldsymbol{W}} = (\boldsymbol{X}^T \widehat{\boldsymbol{V}}^{-1} \boldsymbol{X})^{-1}$ is an estimator of the covariance matrix of $\widehat{\boldsymbol{\theta}}_{\mathrm{GLS}}$ described by Eq. (6). Notice that estimator (6) is equivalent to applying the matrix $\widehat{\boldsymbol{V}}^{-1}$ to the model given by Eq. (1) before applying an ordinary least squares estimator. This is known as prewhitening of the data.

Finally, an estimator of σ_e^2 is given by

$$\widehat{\sigma_e^2} = \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}_{\mathrm{GLS}}\right)^T \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}_{\mathrm{GLS}}\right) / (n-m) \tag{7}$$

of which the statistics are not known exactly.

3.1 Statistical Inference

Brain activation can now be detected by testing the significance of the task-related parameter, say, θ_i of the linear model using standard statistical tools such as the t-test or the F-test. The Student's-t test statistic is given by

$$T_t = \left[\widehat{\boldsymbol{\theta}}_{\mathrm{GLS}}\right]_i / \sqrt{\widehat{W}_{ii}\widehat{\sigma_e^2}} \quad , \tag{8}$$

where $\left[\widehat{\boldsymbol{\theta}}_{\mathrm{GLS}}\right]_i$ denotes the i^{th} element of $\widehat{\boldsymbol{\theta}}_{\mathrm{GLS}}$, $\widehat{\sigma_e^2}$ is given by Eq. (7), and $\widehat{W}_{ii}$ denotes the i^{th} diagonal element of the $m \times m$ matrix $\widehat{\boldsymbol{W}}$. The one-sided t-test decides H_1 if $T_t > \gamma$, whereas the two-sided t-test decides H_1 if $T_t < -\gamma$ or $T_t > \gamma$, with γ a user specified, positive threshold. In practice, this threshold is chosen in function of a false positive rate that the user allows in case the null hypothesis H_0 is true. Approximately, the test statistic T_t has a t distribution with $n-m$ degrees of freedom (exact if $\boldsymbol{V}$ would be known) under H_0. Alternatively, one may use the test statistic

$$T_F = \left(\left[\widehat{\boldsymbol{\theta}}_{\mathrm{GLS}}\right]_i\right)^2 / \left(\widehat{W}_{ii}\widehat{\sigma_e^2}\right) \quad , \tag{9}$$

which has an approximate F distribution with 1 and $n-m$ degrees of freedom (exact if $\boldsymbol{V}$ is known) under H_0. The F-test, which is a two-sided test, decides H_1 if $T_F > \gamma$, with γ some user specified threshold.

4 Likelihood Based Tests

In this section, two new tests (a one-sided as well as a two-sided likelihood ratio test) for brain activation detection is presented with incorporation of colored noise. Thereby, the significance of the task-related parameter θ_i of the linear model is tested.

4.1 The Joint Probability Density Function of the Data

In order to use likelihood based tests, the joint probability density function (PDF) of the fMRI data $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\alpha})$ is required. From Bayes' theorem, we have:

$$p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2) = p(\boldsymbol{y}_p|\boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2)\, p(\boldsymbol{y}_{n-p}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{y}_p) \tag{10}$$

with $\boldsymbol{y}_p = (y_1, \ldots, y_p)^T$ and $\boldsymbol{y}_{n-p} = (y_{p+1}, \ldots, y_n)^T$. The second part of the right hand side is the conditional PDF of the observations $\boldsymbol{y}_{n-p}$ given that the initial observations $\boldsymbol{y}_p$ remain fixed at their observed values. Under the assumed AR model (2), where e is Gaussian distributed, it may be written as [10]

$$p(\boldsymbol{y}_{n-p}|\boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2, \boldsymbol{y}_p) = \left(\frac{1}{2\pi\sigma_e^2}\right)^{(n-p)/2} \times$$
$$\exp\left(-\frac{1}{2\sigma_e^2}\sum_{t=p+1}^{n}\{y_t - \boldsymbol{x}_t\boldsymbol{\theta} + \alpha_1(y_{t-1} - \boldsymbol{x}_{t-1}\boldsymbol{\theta}) + \ldots + \alpha_p(y_{t-p} - \boldsymbol{x}_{t-p}\boldsymbol{\theta})\}^2\right) \tag{11}$$

where $\boldsymbol{x}_t$ denotes the t-th row of the design matrix $\boldsymbol{X}$. The joint PDF of the data $\boldsymbol{y}_p$ may be written as [10]

$$p(\boldsymbol{y}_p|\boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2) = \left(\frac{1}{2\pi\sigma_e^2}\right)^{p/2} \times$$
$$|\boldsymbol{V}_p|^{-1/2} \exp\left(-\frac{1}{2\sigma_e^2}(\boldsymbol{y}_p - \boldsymbol{X}_{1:p}\boldsymbol{\theta})^T \boldsymbol{V}_p^{-1}(\boldsymbol{y}_p - \boldsymbol{X}_{1:p}\boldsymbol{\theta})\right) \tag{12}$$

where $\boldsymbol{X}_{1:p}$ denotes the $p \times m$ matrix consisting of the first p rows of the design matrix $\boldsymbol{X}$. $\boldsymbol{V}_p$ denotes the $p \times p$ covariance matrix of $\boldsymbol{v}_p = (v_1, \ldots, v_p)^T$ and $|\boldsymbol{V}_p|$ denotes the determinant of $\boldsymbol{V}_p$.

4.2 Statistical Inference

If we substitute the acquired data $\boldsymbol{y}$ in the expression for the joint PDF of the data (10), the resulting function is a function of the unknown parameters $(\boldsymbol{\alpha}, \boldsymbol{\theta}, \sigma_e^2)$ only. By regarding these parameters as variables, the LF $p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2; \boldsymbol{y})$ is obtained. Then, the generalized likelihood ratio (GLR) is given by [11]:

$$\lambda = \frac{\sup_{\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_m,\boldsymbol{\alpha},\sigma_e^2} p\left(\theta_1, \ldots, \theta_{i-1}, 0, \theta_{i+1}, \ldots, \theta_m, \boldsymbol{\alpha}, \sigma_e^2; \boldsymbol{y}\right)}{\sup_{\boldsymbol{\theta},\boldsymbol{\alpha},\sigma_e^2} p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2; \boldsymbol{y})} \quad . \tag{13}$$

The denominator of λ is the LF evaluated at the maximum likelihood (ML) estimator under H_0, whereas the numerator of λ is the LF evaluated at the ML estimator under H_1. From the GLR statistic, a one-sided as well as a two-sided likelihood ratio test can be constructed.

Two-sided likelihood ratio test. The generalized likelihood ratio test (GLRT) principle states that H_0 is to be rejected if and only if $\lambda \geq \lambda_0$, where λ_0 is some user specified threshold. It can be shown that, asymptotically (i.e., for $N \to \infty$), the modified GLR statistic

$$T_{LR} = 2\log\lambda \tag{14}$$

possesses a χ_1^2 distribution, that is, a chi-square distribution with 1 degree of freedom, when H_0 is true [11].

One-sided likelihood ratio test. The signed likelihood ratio test statistic is given by [12]

$$T_{LR1} = \text{Sign}\left(\widehat{\theta}_i\right)\sqrt{2\log\lambda}. \tag{15}$$

The test decides H_1 if $T_{LR1} > \gamma$, with γ some user specified threshold. Asymptotically, the test statistic T_{LR1} has a standard normal distribution under H_0.

4.3 Computational Considerations

To obtain the likelihood ratio λ, the ML estimates of the unknown parameters under the null hypothesis H_0 and the alternative hypothesis H_1 have to be found. For that purpose, the LF has to be maximized with respect to the unknown parameters $(\boldsymbol{\alpha}, \boldsymbol{\theta}, \sigma_e^2)$. The noise variance σ_e^2 can be eliminated from this optimization problem since it can be shown that the value of σ_e^2 that maximizes the LF $p(\boldsymbol{\alpha}, \boldsymbol{\theta}, \sigma_e^2; \boldsymbol{y})$ with respect to σ_e^2 is given by

$$\sigma_e^2 = \frac{1}{n}\left[\sum_{i=1}^{p}\sum_{j=1}^{p}[\boldsymbol{V}_p^{-1}]_{ij}(y_i - \boldsymbol{x}_i\boldsymbol{\theta})(y_j - \boldsymbol{x}_j\boldsymbol{\theta}) + \sum_{t=p+1}^{n}\{y_t - \boldsymbol{x}_t\boldsymbol{\theta} + \alpha_1(y_{t-1} - \boldsymbol{x}_{t-1}\boldsymbol{\theta}) + \ldots + \alpha_p(y_{t-p} - \boldsymbol{x}_{t-p}\boldsymbol{\theta})\}^2\right], \tag{16}$$

$[\boldsymbol{V}_p^{-1}]_{ij}$ being the (i,j)th element of $\boldsymbol{V}_p^{-1}$. Substituting (16) in (10) yields the so-called concentrated LF. The ML estimates $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\theta}})$ of the parameters $(\boldsymbol{\alpha}, \boldsymbol{\theta})$ can now be found by maximizing the concentrated LF with respect to $(\boldsymbol{\alpha}, \boldsymbol{\theta})$, which is a nonlinear optimization problem that can be solved numerically.

5 Experiments

Experimental fMRI data sets were obtained from small animal as well as from human subjects. The experiments for the small animals (3 rats) were done on a 7T MRI system (SMIS, Guildford, UK) with an 80 mm aperture and self-shielded gradients. Images were taken with size 256×128, maximum gradient strengths $G_r = 0.017$ T/m, $G_p = 0.027$ T/m, $G_{sl} = 0.07$ T/m, and ramp time 100 μs. All human experiments were performed on a 1,5 T scanner with high-performance 40 mT/m gradients (Siemens Sonata, Erlangen, Germany). Subjects were three healthy volunteers (mean age 33 years). Gradient-recalled multi-shot EPI sequences (TE 50 ms, TR 3000 ms) were used with 30 slices covering the whole brain. The voxels dimensions were $3 \times 3 \times 3$ mm.

5.1 Order of the AR Model of fMRI Noise Structures

From the experimental fMRI null data, the order of the AR model was determined. Previous work by Woolrich et al. examined the necessary AR order from

six null data sets. They concluded that AR(6) was sufficient for their data [4]. In our work, various null data sets were acquired from humans as well as from small animals. The null data were modelled with a second order polynomial model: $b_0 + b_1 t + b_1 t^2$ along with an AR(p) model of which the order was estimated using Akaike's information criterion (AIC) [13], where a penalty factor of 3 instead of 2 was chosen [14]. Evaluation of AR order maps, constructed from these data revealed that an AR(3) model is conservative with enough freedom to accommodate even more complex AR processes than expected.

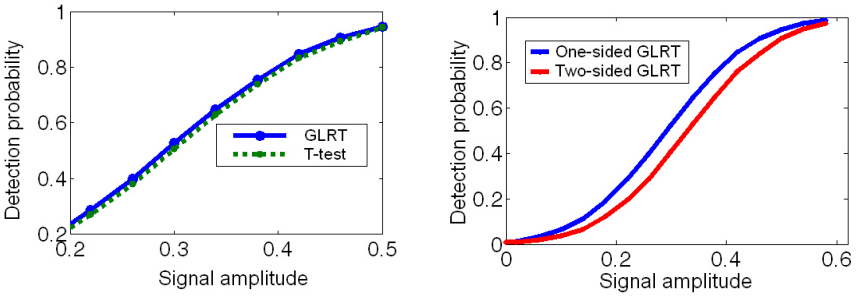
5.2 Simulation Experiments

For a fixed false alarm rate of 1%, the likelihood ratio tests proposed were compared to the GLM tests with respect to detection 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.

Simulation experiments were set up to detect brain activation. Thereby, a simple on-off activation scheme was used in which traces of 100 time-points were generated with period equal to 20 (10 on, 10 off). Also, small linear and quadratic trends were introduced that were modelled along with the baseline and activation pattern. The amplitude of the activation pattern was gradually increased from 0 till 0.6; the noise standard deviation was fixed to 1. For each simulation experiment, 10^4 Monte Carlo simulations were run.

6 Results and Discussion

Typical results for the simulation experiments described in Subsection 5.2 are shown in Fig. 1. Fig. 1(a) shows the detection rate as a function of the amplitude of the activation pattern. Although results weakly depend on this amplitude, it may be concluded from the numerical outcomes that, for a fixed false alarm rate



(a) GLM vs. GLRT (onesided) (b) Onesided vs. two-sided GLRT

Fig. 1. Detection rates with a fixed false alarm rate of 1%

of 1%, the detection rate of the proposed one-sided GLRT is uniformly 2-3% better compared to the detection rate of the GLM test incorporating colored noise. Similar results were observed when comparing the two-sided tests.

Finally, Fig. 1(b) shows the results when comparing the one-sided test against the two-sided test in case the amplitude of the activation pattern was known to be positive. As expected, the one-sided test performs in that case over 10% better than the two-sided test.

7 Conclusions

In this paper, likelihood ratio tests for the detection of functional brain activity, one-sided as well as two-sided, have been presented. In contrast to the general linear model (GLM) tests, the proposed likelihood ratio tests allow direct incorporation of colored noise and do not require a prewhitening step. Simulation results showed that likelihood based detection results in systematic slightly improved detection probabilities compared to the currently popular GLM based tests.

Acknowledgements

This work has been financially supported by the FWO (Fund for Scientific Research - Belgium). The authors like to thank A. Smolders and N. Van Camp for providing them with experimental fMRI null data.

References

1. Cohen, M.S.: Real-time functional magnetic resonance imaging. Methods **25** (2001) 201–220
2. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S.J.: Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. **2** (1995) 189–210
3. Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., Evans, A.C.: A general statistical analysis for fMRI data. NeuroImage **15** (2002) 1–15
4. Woolrich, M.W., Ripley, B.D., Brady, J.M., Smith, S.M.: Temporal autocorrelation in univariate linear modelling of fMRI data. NeuroImage **14** (2001) 1370–1386
5. Worsley, K.J., Friston, K.J.: Analysis of fMRI time-series revisited – again. NeuroImage **2** (1995) 173–181
6. Kay, S.M., Marple, S.L.: Spectrum analysis – a modern perspective. In: Proceedings of the IEEE. Volume 69. (1981) 1380–1419
7. Carew, J.D., Wahba, G., Xie, X., Nordheim, E.V., Meyerand, M.E.: Optimal spline smoothing of fMRI time series by generalized cross-validation. NeuroImage **18** (2003) 950–961
8. den Dekker, A.J., Sijbers, J.: Estimation of signal and noise from MR data. In: Advanced Image Processing in Magnetic Resonance Imaging. Volume 26 of Signal Processing and Communications. Marcel Dekker (2005) ISBN: 0824725425.

9. van den Bos, A.: 8: Parameter Estimation. In: Handbook of Measurement Science. Volume 1. Edited by P. H. Sydenham, Wiley, Chichester, England (1982) 331–377
10. Priestley, M.B.: Spectral analysis and time series. Academic Press, London (1981)
11. Kay, S.M.: Fundamentals of statistical signal processing: estimation theory. Prentice-Hall, Inc. (1993)
12. Rowe, D.B., Logan, B.R.: A complex way to compute fMRI activation. Neuroimage **23** (2004) 1078–1092
13. Ardekani, B.A., Kershaw, J., Kashikura, K., Kanno, I.: Activation detection in functional MRI using subspace modeling and maximum likelihood estimation. IEEE Trans Med Imaging **18** (1999) 246–254
14. Broersen, P.M.T.: Finite sample criteria for autoregressive order selection. IEEE Trans Sig Proc **48** (2000) 3550–3558