

Model selection in electromagnetic source analysis with an application to VEF's

Lourens J. Waldorp, Hilde M. Huizenga, Raoul P.P.P. Grasman, Koen B.E. Böcker, Jan C. de Munck and Peter C.M. Molenaar

Abstract— In electromagnetic source analysis it is necessary to determine how many sources are required to describe the EEG or MEG adequately. Model selection procedures (MSP's, or goodness of fit procedures) give an estimate of the required number of sources. Existing and new MSP's are evaluated in different source and noise settings: two sources which are close or distant, and noise which is uncorrelated or correlated. The commonly used MSP residual variance is seen to be ineffective, that is it often selects too many sources. Alternatives like the adjusted Hotelling's test, Bayes information criterion, and the Wald test on source amplitudes are seen to be effective. The adjusted Hotelling's test is recommended if a conservative approach is taken, and MSP's such as Bayes information criterion or the Wald test on source amplitudes are recommended if a more liberal approach is desirable. The MSP's are applied to empirical data (visual evoked fields).

I. INTRODUCTION

ELECTROMAGNETIC source analysis yields an estimate of the sources of the electro-encephalogram (EEG) or magneto-encephalogram (MEG) [1]. Precise estimates of the sources can only be obtained if the number of sources is correct [2]. That is, if activity related to the stimulus, and no noise is modelled. Model selection procedures (MSP's) can be used to give an estimate of the required number of sources.

MSP's based on an eigenvalue decomposition have been studied extensively [3], [4], [5], [6], [7], [8], [9], [10]. A disadvantage of these MSP's is that they only give an estimate of the number of uncorrelated sources. Other MSP's, based on the residuals of a source analysis, have also been reported, for example the residual variance [11], the chi-square [1], [2], and lack of fit statistic [12]. These MSP's give an estimate of the number of uncorrelated or correlated sources.

In this paper we study the second class of MSP's more

*Lourens J. Waldorp is with the Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB, Amsterdam, The Netherlands, email: waldorp@psy.uva.nl. The Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged for funding this project. This research was conducted while L.J. Waldorp (527-25-013), R.P.P.P. Grasman (527-25-014), and K.B.E. Böcker (527-25-015) were supported by a grant of the NWO foundation for Behavioral and Educational Sciences of this organization awarded to H.M. Huizenga, P.C.M. Molenaar, and J.L. Kenemans. The research has also been made possible by a KNAW fellowship to H.M. Huizenga.

H.M. Huizenga, R.P.P.P. Grasman, and P.C.M. Molenaar are with the Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

K.B.E. Böcker is with the Department of Psychopharmacology, Utrecht University, the Netherlands

J.C. de Munck is with the MEG Center, Free University Medical Center, Amsterdam, the Netherlands

extensively, and propose several alternatives. More specifically, we present a set of theoretical criteria for an adequate MSP, and evaluate the MSP's according to these criteria. In addition their behavior is investigated in an MEG simulation study and in an empirical study on visual evoked fields (VEF's).

II. METHOD

In this section the most often used MSP, the residual variance (RV), is discussed. Based on this discussion, a set of theoretical criteria for MSP's is derived. Then alternative MSP's are discussed in terms of these criteria. First, however, some notation will be given.

A. Notation

Although only the instantaneous MEG model is considered, all methods are applicable to EEG, combined EEG/MEG, and a spatio-temporal model as well.

Let \mathbf{y}_i be the m -vector with MEG measured on m sensors for trial $i = 1, 2, \dots, n$, and let $\bar{\mathbf{y}} = \sum_{i=1}^n \mathbf{y}_i / n$ be the average. The average MEG $\bar{\mathbf{y}}$ is modeled by a function $\mathbf{f}(\boldsymbol{\theta})$, a vector of dimension m [1]. The dipole location and moment parameters are collected in the $p = 5d$ -vector $\boldsymbol{\theta}$, where d denotes the number of sources. Then

$$\bar{\mathbf{y}} = \mathbf{f}(\boldsymbol{\theta}) + \mathbf{e}, \quad (1)$$

where \mathbf{e} is the total error (residual). The total error consists of pure error (noise) and modeling (lack of fit) error. Pure error is for example spontaneous MEG, and modeling error is for example error due to an incorrect source model.

If the pure error is 'white', that is, is uncorrelated and has equal variances (homoscedastic), then source parameter estimates can be obtained by ordinary least squares (OLS). An OLS estimate $\hat{\boldsymbol{\theta}}$ is obtained by minimizing $\mathbf{e}'\mathbf{e}$ with respect to $\boldsymbol{\theta}$. At the minimum, $s^2(\hat{\boldsymbol{\theta}}) = \mathbf{e}'\mathbf{e}/(m - p)$ gives an estimate of the total error variance. An unbiased estimate of the pure error variance for the mean can be obtained from trial variation around the mean: $s^2 = \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})'(\mathbf{y}_i - \bar{\mathbf{y}}) / mn(n - 1)$ [13, p. 33].

If the pure error is 'colored', that is, is correlated and has unequal variances (heteroscedastic), then source parameters can be estimated by generalized least squares (GLS) [14]. In GLS both data and model are prewhitened by the pure error covariance matrix $\boldsymbol{\Sigma}$. Generally, $\boldsymbol{\Sigma}$ is unknown and is therefore estimated by $\mathbf{S} = \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})' / n(n - 1)$, that is from trial variation around the mean. The estimate \mathbf{S} is an unbiased estimate for the averaged data. Let $\mathbf{S} = s^2\mathbf{V}$ and $(\mathbf{V}^{-1/2})'\mathbf{V}^{-1/2} = \mathbf{V}^{-1}$

be the Cholesky decomposition [15, p. 138], then the prewhitened data and model are respectively $\mathbf{V}^{-1/2}\bar{\mathbf{y}}$ and $\mathbf{V}^{-1/2}\mathbf{f}(\boldsymbol{\theta})$. The prewhitened total error is then $\tilde{\mathbf{e}} = \mathbf{V}^{-1/2}\mathbf{e} = \mathbf{V}^{-1/2}\bar{\mathbf{y}} - \mathbf{V}^{-1/2}\mathbf{f}(\boldsymbol{\theta})$. A GLS estimate $\hat{\boldsymbol{\theta}}$ is obtained by minimizing $\tilde{\mathbf{e}}'\tilde{\mathbf{e}} = \mathbf{e}'\mathbf{V}^{-1}\mathbf{e}$ with respect to $\boldsymbol{\theta}$. At the minimum $\tilde{s}^2(\hat{\boldsymbol{\theta}}) = \tilde{\mathbf{e}}'\tilde{\mathbf{e}}/(m-p)$ gives an estimate of the total error variance. The estimate of the pure error variance remains the same:

$$\begin{aligned}\tilde{s}^2 &= \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})'\mathbf{V}^{-1}(\mathbf{y}_i - \bar{\mathbf{y}})/mn(n-1) \\ &= \text{tr}(\mathbf{V}^{-1}\mathbf{S}/m) = s^2,\end{aligned}$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix.

B. RV and criteria for MSP's

The RV compares the squared total error to the squared data (data power) to give an estimate of how much of the data power remains unexplained by the model. The RV is defined as (e.g. [12], [16])

$$\text{RV} = 100 \left(\frac{\mathbf{e}'\mathbf{e}}{\bar{\mathbf{y}}'\bar{\mathbf{y}}} \right). \quad (2)$$

A model is said to fit the data if its value is close to 0. In practice a threshold value is chosen (e.g. [17]). The threshold of 5% is chosen here. When comparing several source models, the model with the smallest number of sources which has a value equal to or below the threshold is selected.

In the statistical literature the RV is in general not considered as a good MSP [18], [19]. Note that the RV is high if the modelling error is high, but also if the pure error is high. As a consequence, the MSP has a tendency to model pure error (i.e. over fit). This is due to the fact that the RV does not incorporate information concerning the pure error variance. One of the criteria for a good MSP is thus that: (C1) it should incorporate information concerning the pure error variance. Note however that this pure error variance is never known precisely. It should always be estimated. Therefore, a good MSP also incorporates the fact that: (C2) the pure error variance is not known exactly but is estimated.

The RV is highly dependent on data power: the RV will indicate more often a bad fit for low amplitude signals than for high amplitude signals. Therefore, the RV has a tendency to over fit for low amplitude signals. This can be accommodated by choosing a different threshold value, for example 10% instead of 5%, but these choices are quite arbitrary. Therefore, another criterion for a good MSP is that: (C3) the selection method should be based on an objective and sensible criterion. Note in addition that the RV decreases if the number of sources is increased, that is, there is no penalty for increasing the number of sources. As a consequence, the RV has again a tendency to over fit. Therefore, another criterion for a good MSP is that it should: (C4) account for the number of estimated parameters. Note moreover that the RV is based on white pure

error. It is not suited for colored pure error. Therefore, another criterion for a good MSP is that it should: (C5) account for colored pure error. Specifically, if the source parameter estimates are obtained by GLS, then both the data and model (and thus the total error) in the MSP should be prewhitened by the pure error covariance matrix. Note however, again that the pure error covariance matrix is not known exactly, but is estimated. Therefore, a good MSP should: (C6) also account for the fact that this matrix is estimated.

In summary, the criteria for good MSP's are:

- C1 Incorporate pure error variance
- C2 If C1 is satisfied and an estimate is used, then account for the errors in this estimate
- C3 A selection method based on an objective and sensible criterion
- C4 Incorporate the number of parameters
- C5 Incorporate pure error covariances
- C6 If C5 is satisfied and an estimate is used, then account for the errors in this estimate

C. Model selection procedures

An MSP consists of the test itself and the acceptance/rejection decision rule. For example, for the RV, the test is Eq. (2), and the decision rule is: accept the smallest model with an $\text{RV} \leq 5\%$. In the following, we discuss several MSP's and their decision rules in terms of the criteria C1-C6. An overview of all MSP's and the criteria is given in Table I.

Residual variance. As was shown above, the RV given in (2) does not meet C1-C6. If the pure error is colored, and the source estimates are obtained by GLS, then the data and model in (2) can be prewhitened [20]. In this manner C5 is satisfied but C6 is not.

Chi-square statistic. The chi-square statistic tests whether the total error variance exceeds the pure error variance. The chi-square statistic is defined for white pure error as [2], [1]

$$\chi^2 = \frac{\mathbf{e}'\mathbf{e}}{\sigma^2}, \quad (3)$$

where σ^2 is the true pure error variance. If it is assumed that the pure error is white and is multivariate normally distributed with known σ^2 , then the statistic is χ^2 distributed with $m-p$ degrees of freedom (df). A model fits if it is not significant, that is when the chi-square value is less than the value associated with $\chi^2(m-p, \alpha)$ with a chosen significance level $\alpha = 0.05$. When comparing source models, the smallest model with nonsignificant χ^2 is selected.

C1 is satisfied since σ^2 is used. However, in practice the true pure error variance is unknown and has to be estimated (by s^2). Consequently, the $\chi^2(m-p)$ distribution may not be correct, and so C2 is not satisfied. C3 is satisfied, since the criterion is based on hypothesis testing. C4 is also satisfied, for the degrees of freedom influence the significance level, which in turn affects the acceptance/rejection of a model. If the pure error is colored and the data and model are prewhitened, then C5 is met, but C6 is not.

TABLE I

A LIST OF THE CRITERIA FOR ALL THE MSP'S. A '-' INDICATES THAT THE MSP DOES NOT SATISFY THAT CRITERION, AND A '*' INDICATES THAT IT DOES.

	RV	χ^2	LOF	T^2	aT^2	LR	aLR	C_p	AIC	BIC	W
C1	-	*	*	*	*	*	-	*	*	*	-
C2	-	-	*	*	*	-	-	-	-	-	-
C3	-	*	*	*	*	*	*	*	*	*	*
C4	-	*	*	-	*	*	*	*	*	*	*
C5	*	*	*	*	*	*	*	*	*	*	*
C6	-	-	-	*	*	-	-	-	-	-	-

Lack of fit. The lack of fit (LOF) tests whether the total error variance exceeds the pure error variance. If the pure error is white, then the LOF is defined as [13], [14], [21]

$$\text{LOF} = \frac{\mathbf{e}'\mathbf{e}/(m-p)}{s^2}. \quad (4)$$

If the pure error is white and multivariate normally distributed, then the LOF has an F distribution with $m-p$ and $m(n-1)$ df. A model fits if the LOF is not significant at significance level α . When source models are compared, the smallest model with nonsignificant LOF is selected.

Criterion C1 is satisfied, since s^2 is incorporated. Both total error and pure error variance are treated as estimates. Consequently, an F distribution is used, and so C2 is also satisfied. C3 is satisfied since a hypothesis testing approach is taken. The number of parameters influences the acceptance/rejection region through the degrees of freedom, and C4 is therefore also satisfied. If the pure error is colored and the data and model are prewhitened C5 is met but C6 is not, since there is no adjustment to account for the errors in estimating the pure error covariances.

Hotelling's T^2 . Hotelling's T^2 tests whether the total error deviates from zero. If the pure error is white, and the estimate $\mathbf{f}(\hat{\boldsymbol{\theta}})$ is considered as a fixed hypothesis, then Hotelling's T^2 is defined as

$$T^2 = \frac{\mathbf{e}'\mathbf{e}}{s^2} \frac{n-m}{m(n-1)}. \quad (5)$$

If the pure error is white and multivariate normally distributed, and there is a fixed hypothesis, then T^2 has an F distribution with m and $n-m$ df. A model fits if the test is not significant at significance level α . If models are compared, then the smallest model with nonsignificant T^2 is selected.

The pure error variance s^2 is incorporated, and so, C1 is satisfied. Since the F distribution is used, C2 is also met. C3 is also satisfied since a hypothesis testing approach is taken. Criterion C4, however, is not satisfied, since the number of model parameters is not accounted for. If the pure error is colored then the prewhitened version of Hotelling's T^2 [22, p. 98], [23, p. 101] satisfies both C5 and C6.

Adjusted Hotelling's T^2 . The degrees of freedom m in the denominator in Hotelling's T^2 (5) reflects the idea that the estimated source model is a hypothesis and therefore

a constant [23, p. 105]. In our case, however, the source parameters are estimated which affects the degrees of freedom. To correct this, m in the denominator can be replaced by $m-p$ in Hotelling's T^2 , which gives the adjusted Hotelling's T^2 (aT^2). aT^2 is F distributed with $m-p$ and $n-m$ df [24]. Criteria C1-C6 are now all satisfied.

Likelihood ratio test. The likelihood ratio (LR) test evaluates if one model is more likely than another. The LR is defined as -2 times the difference between two log-likelihood functions corresponding to different models [25]. If the pure error is white and multivariate normal, and if the pure error variance is estimated by s^2 for both models, then the LR for a source model with d and $d+1$ sources is

$$\text{LR} = \frac{\mathbf{e}(\hat{\boldsymbol{\theta}}_d)' \mathbf{e}(\hat{\boldsymbol{\theta}}_d) - \mathbf{e}(\hat{\boldsymbol{\theta}}_{d+1})' \mathbf{e}(\hat{\boldsymbol{\theta}}_{d+1})}{s^2}. \quad (6)$$

The LR has a χ^2 distribution with $p_{d+1} - p_d$ df. If the LR is not significant, then the larger model is not more likely than the smaller model, and the smallest model is accepted.

Criterion C1 is satisfied since s^2 is incorporated. However, it is not considered as an estimate, so C2 is not met. C3 is satisfied for it is a hypothesis test. C4 is satisfied since the difference in the number of parameters is reflected in the degrees of freedom. If the pure error is colored and the data and model are prewhitened (with the same estimate \mathbf{V} for both models), then C5 is met but C6 is not.

Adjusted likelihood ratio test. The LR can be adjusted to a test which is more robust against violations of assuming a multivariate normal distribution. The aLR for white pure error is [13, p. 198]

$$\text{aLR} = \frac{m-p_{d+1}}{p_{d+1}-p_d} \frac{\mathbf{e}(\hat{\boldsymbol{\theta}}_d)' \mathbf{e}(\hat{\boldsymbol{\theta}}_d) - \mathbf{e}(\hat{\boldsymbol{\theta}}_{d+1})' \mathbf{e}(\hat{\boldsymbol{\theta}}_{d+1})}{\mathbf{e}(\hat{\boldsymbol{\theta}}_{d+1})' \mathbf{e}(\hat{\boldsymbol{\theta}}_{d+1})} \quad (7)$$

If the pure error is multivariate normally distributed, then aLR is F distributed with $p_{d+1} - p_d$ and $m - p_{d+1}$ df [26, p. 167]. The decision rule is the same as that of the LR.

Note that the estimate s^2 is not incorporated so that C1 and C2 are not satisfied. C3 and C4 are still satisfied. If the pure error is colored and the data and model are prewhitened, then C5 is met but not C6.

Mallow's C_p . Mallow's C_p compares the ratio of the total error variance to the pure error variance and a penalty term. If we use s^2 as an approximation to the total error variance obtained from a model with as many sources as

possible ('saturated model'), then C_p can be defined for white pure error as [27, p. 224]

$$C_p = \frac{\mathbf{e}'\mathbf{e}}{s^2} - m + 2p. \quad (8)$$

The penalty term is such that the expected value of C_p is p . Consequently, when comparing models, the model with the smallest value of $C_p - p$ is selected [27, p. 226].

Criterion C1 is satisfied since s^2 is used. However, s^2 is not considered as an estimate, and so C2 is not satisfied. C3 is met since the expected value of C_p is p . Criterion C4 is satisfied through the penalty term $2p$. If the pure error is colored and the data and model are prewhitened, then C_p satisfies C5 but not C6.

Akaike information criterion. The Akaike information criterion (AIC) compares the loglikelihood function of the total error to a penalty term for different models [28]. If the pure error is white and its variance is estimated by s^2 , then

$$\text{AIC} = \ln(\pi s^2) + \frac{\mathbf{e}'\mathbf{e}}{s^2} + 2p. \quad (9)$$

The model with the smallest AIC value is selected [29].

Criterion C1 is satisfied by s^2 . However, C2 is not satisfied. C3 is met, since the decision rule is a minimization over source models. The penalty term contains the number of parameters and so C4 is satisfied. If the pure error is colored and the data and model are prewhitened, then C5 is met but not C6.

Note that the AIC resembles the LR. The acceptance/rejection region of the hypothesis test is translated by $2(p_{d+1} - p_d)$ [30]. The AIC also resembles C_p , since the number of sensors m in C_p is fixed, and the first term in the AIC is also constant.

Bayes information criterion. The Bayes information criterion (BIC, or minimum description length) is similar to the AIC except for the penalty term. The BIC is defined in terms of the loglikelihood function [31], [32]. If the pure error is white, and if its variance is estimated by s^2 , then

$$\text{BIC} = \ln(\pi s^2) + \frac{\mathbf{e}'\mathbf{e}}{s^2} + \ln(m)p. \quad (10)$$

The source model with the smallest BIC value is selected.

Criterion C1 is satisfied since s^2 is incorporated. C2 is not met. Criterion C3 is satisfied in the same sense as the AIC. C4 is satisfied by the penalty term. If the pure error is colored and the data and model are prewhitened, then C5 is met but not C6.

Wald tests. A Wald test offers the possibility to test a hypothesis about a set of parameters. One test is a test on the source amplitudes. The basic idea is that a source should be incorporated only if its amplitude is significantly larger than zero. Another test is a test on the source location parameters. The basic idea here is that a source should be incorporated only if its location is significantly different from other source locations. First the general Wald test is presented and then the two specific formulations.

Let $\mathbf{r}(\hat{\boldsymbol{\theta}})$ be a vector of q (non)linear functions of $\hat{\boldsymbol{\theta}}$ which represent the hypotheses about the source parameters, and the hypothesized values of these functions are collected in the q -vector \mathbf{r}_h . Furthermore, let $s^2(\hat{\boldsymbol{\theta}})\mathbf{C} = s^2(\hat{\boldsymbol{\theta}})(\mathbf{F}'\mathbf{F})^{-1}$ be the covariance matrix of $\hat{\boldsymbol{\theta}}$, where the $m \times p$ matrix \mathbf{F} contains the first order partial derivatives of the model $\mathbf{f}(\hat{\boldsymbol{\theta}})$ with respect to the parameters [13, p. 26], [33]. The covariance matrix of $\mathbf{r}(\hat{\boldsymbol{\theta}}) - \mathbf{r}_h$ is then $s^2(\hat{\boldsymbol{\theta}})\mathbf{R}(\hat{\boldsymbol{\theta}})\mathbf{C}\mathbf{R}(\hat{\boldsymbol{\theta}})'$, with $\mathbf{R}(\hat{\boldsymbol{\theta}})$ a $q \times p$ matrix of first order partial derivatives of $\mathbf{r}(\hat{\boldsymbol{\theta}})$ with respect to the p source parameters. The Wald test for white pure error is then defined as [34]

$$W = \frac{(\mathbf{r}(\hat{\boldsymbol{\theta}}) - \mathbf{r}_h)'[\mathbf{R}(\hat{\boldsymbol{\theta}})\mathbf{C}\mathbf{R}(\hat{\boldsymbol{\theta}})']^{-1}(\mathbf{r}(\hat{\boldsymbol{\theta}}) - \mathbf{r}_h)}{qs^2(\hat{\boldsymbol{\theta}})}. \quad (11)$$

W is approximately F distributed with q and $m - p$ df.

The Wald amplitude (WA) test, tests if source amplitudes deviate from zero. The $q = d$ vector of functions then is $\mathbf{r}(\hat{\boldsymbol{\theta}}) = (\hat{\theta}_1, \dots, \hat{\theta}_d)'$, where $\hat{\theta}_i$ is the amplitude parameter of source i . The derivative matrix $\mathbf{R}(\hat{\boldsymbol{\theta}})$ is defined accordingly and the hypothesis vector is $\mathbf{r}_h = \mathbf{0}$. A source is included if both multivariate (testing all amplitudes simultaneously) and subsequent univariate tests (for all parameters individually) are significant. The largest model for which this is true is selected.

The Wald location (WL) test, tests if source location parameters differ. The following $q = 3d(d-1)/2$ vector of functions is used $\mathbf{r}(\hat{\boldsymbol{\theta}}) = (\hat{\theta}_{11} - \hat{\theta}_{12}, \hat{\theta}_{21} - \hat{\theta}_{22}, \dots, \hat{\theta}_{3(d-1)} - \hat{\theta}_{3d})'$, where $\hat{\theta}_{ji}$ is a location parameter j of source i . The matrix $\mathbf{R}(\hat{\boldsymbol{\theta}})$ is defined accordingly. The hypothesis vector is $\mathbf{r}_h = \mathbf{0}$. A source is included in the model if both multivariate (all pairs of source locations simultaneously) and subsequent tests for all source location-pairs individually, are significant. The largest model for which this is true is selected.

Criterion C1, and thus C2, are not met since s^2 is not incorporated. The Wald tests satisfy C3 for they are hypothesis tests. C4 is also satisfied since the degrees of freedom contains the number of parameters. If the pure error is colored and the model and data are prewhitened, then the parameter covariance matrix is $\tilde{s}^2(\hat{\boldsymbol{\theta}})(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}$ [33]. C5 is then satisfied but C6 is not.

III. SIMULATIONS

In order to evaluate the MSP's, instantaneous MEG simulations were carried out in which the distance between two sources was varied and in which the pure error was either white or colored. The measure of effectiveness of the MSP's was the percentage of correct decisions. Incorrect decisions were categorized as under- or over fitting.

Forward computations. The signal was generated by two dipolar sources inside a sphere with a radius of 10 cm. Both sources were located in the midcoronal plane (crossing both ears and the vertex) with a varying angle γ between the location vectors. The angle was varied by 10° steps between 10° and 50° , yielding distances between 1.39 and 6.76 cm. The moment vectors were completely tangential and perpendicular to the coronal plane at all 5 angles.

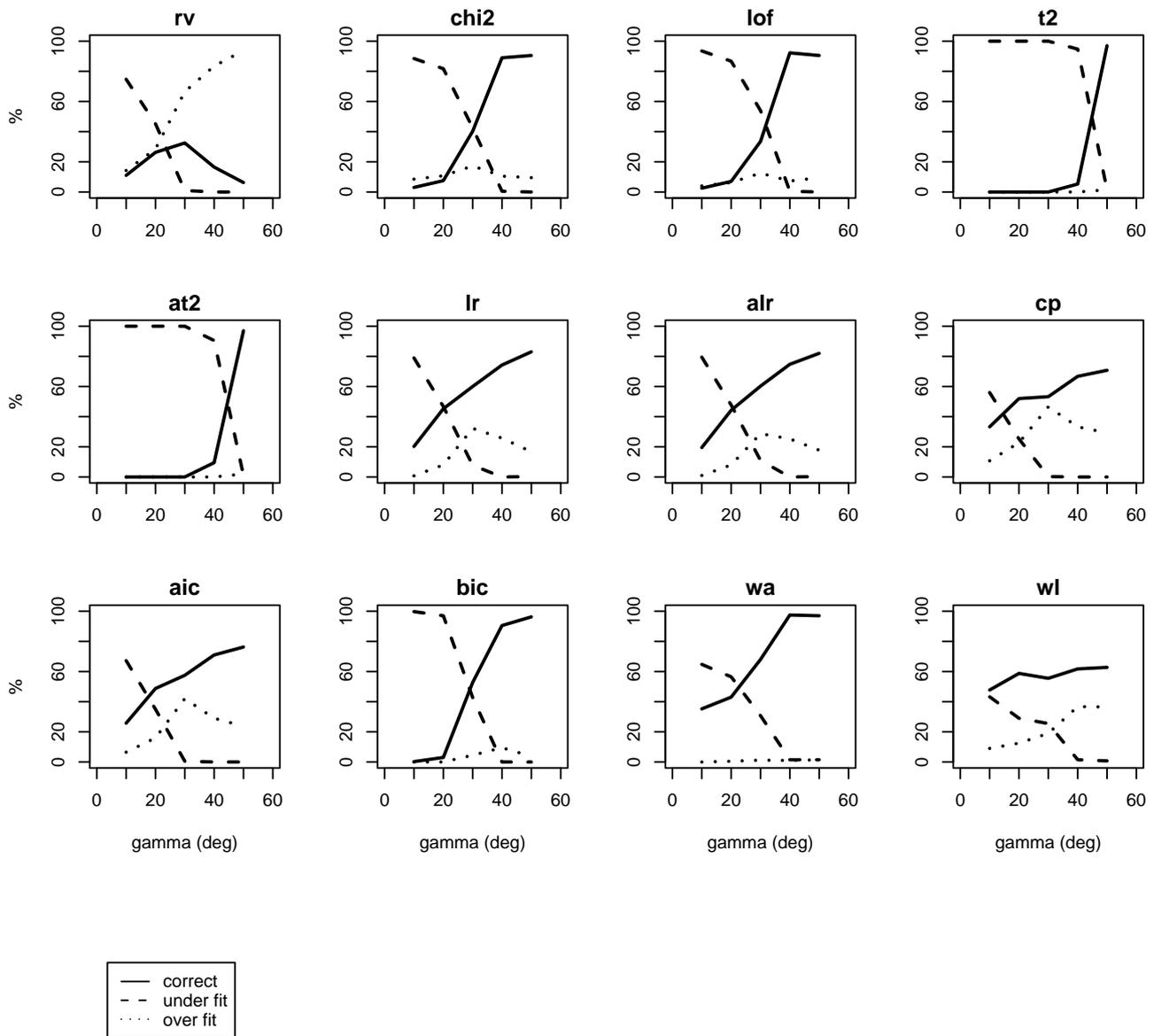


Fig. 1. The percentages of correct and incorrect decisions of all MSP's for white pure error are shown at the angles $\gamma = 10^\circ - 50^\circ$. There are three different kind of percentages: the unbroken line represents the correct decisions (two source model), the broken line represents under fit (one source model), and the dotted line represents over fit (three or more sources).

Normally distributed pure error, white or colored, was added to each sensor in 500 trials; the trials were then averaged. 400 of these replications were generated. The pure error standard deviation was set at 10% of the maximal signal of all sensors of the averaged data, when the sources were at an angle of 25° (3.46 cm). Two pure error conditions were generated: white and colored. The colored pure error covariance matrix was obtained from empirical data. The data were recorded with a whole head CTF device, with 151 axial gradiometers, at a sampling frequency of 250 Hz. The subject received no stimulation and had his eyes open. The data consisted of 100 trials of 100 samples. One sensor was excluded from analysis. The 150×150 spatial pure error covariance matrix was computed with the algorithm described in [35]. The resulting pure error absolute correlations ranged from 0.00 to 0.80 with an average

of 0.16.

Inverse computations. The correct head model was used for the inverse computations. When the pure error was white, the source parameters were estimated by OLS. When the pure error was colored the source parameters were estimated by GLS. The pure error covariance matrix was estimated from the variation of 500 trials around their mean (Sec. II-A). All 400 replicates were modelled by 1, 2, and 3 sources.

A. Results

The decisions of the MSP's in the white and colored pure error cases are shown in respectively Fig. 1 and Fig. 2. If the sources are estimated by GLS then the prewhitened MSP (denoted by the prefix 'g') is used. Three different percentages are depicted: correct decisions (2 source

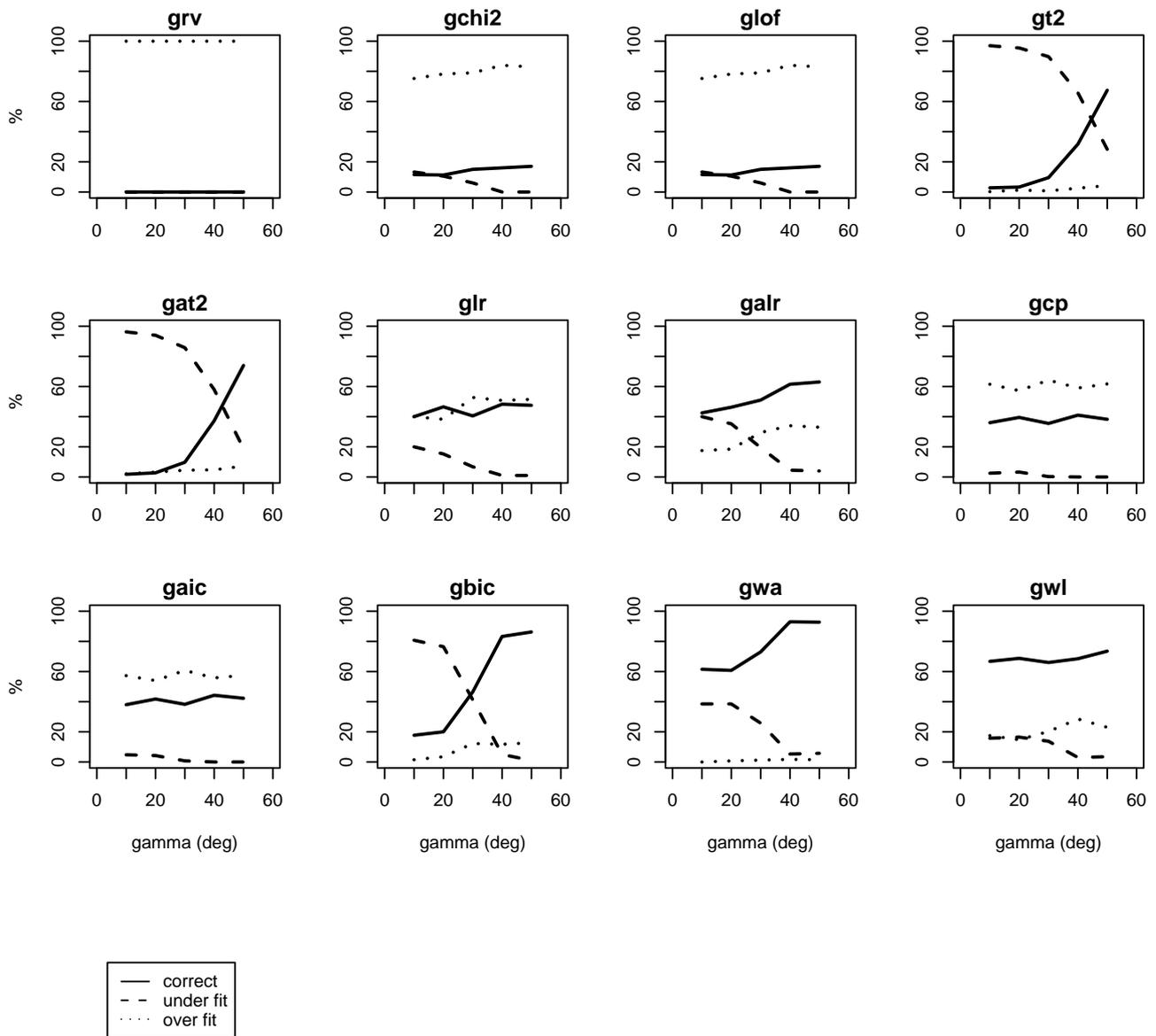


Fig. 2. The percentages of correct and incorrect decisions of all MSP's for colored pure error (denoted by the prefix 'g') are shown at the angles $\gamma = 10^\circ - 50^\circ$. There are three different kind of percentages: the unbroken line represents the correct decisions (two source model), the broken line represents under fit (one source model), and the dotted line represents over fit (three or more sources).

model), under fitting (1 source model), and over fitting (3 or more source model). To evaluate an MSP, three aspects should be considered: (1) The percentage of correct decisions should be as high as possible at all angles, (2) if sources are close (i.e. $\gamma = 10^\circ - 20^\circ$), under fitting is preferred to over fitting, and (3) if an MSP selects the correct number of sources, the source parameter estimates should be accurate.

The RV is inadequate and tends to over fit severely in both error conditions. This can be explained by the fact that the data power decreases as the distance between the sources increases. Therefore, the RV tends to over fit for high angles.

The χ^2 and LOF show equal adequate performance when pure error is white: under fitting for close sources and correct decisions for distant sources. The LOF was expected to

be better than the chi-square statistic, since the χ^2 assumes that the pure error variance is known exactly. However, 500 trials offer a good estimate of the pure error variance. Therefore, the χ^2 and LOF perform equally well. The severe tendency to over fit of both the chi-square statistic and LOF for the colored case might be attributed to the fact that the errors in the pure error covariance estimates are not accounted for.

The performance of T^2 and aT^2 in the white pure error case is conservative: they under fit mostly, and the correct decision is only given when sources are distant. When the pure error is colored they are slightly less conservative.

The performance of the LR and aLR is adequate in the white pure error case: under fitting when sources are close and correct decisions when sources are distant. However, when the pure error is colored, the number of correct de-

cisions remains fairly constant around 40-60%. This might be explained by the fact that the error in the estimate of the pure error covariance matrix is not accounted for.

The C_p and AIC are very similar in performance. The tendency to over fit is already present when pure error is white, and becomes dominant when the pure error is colored. This is a common finding. The AIC has been found to over fit in a variety of settings [29], [36]. The cause of this is the tolerant ‘penalty’ of $2p$. Note that the performance of the AIC is also similar to that of the LR (see Sec. II-C).

The performance of the BIC can be considered as good. It under fits when sources are close and hardly over fits. In addition it performs well in the colored case. Note that if less sensors are used, then the penalty of the BIC can reduce to that of the AIC. Therefore, to avoid over fitting the BIC should be used only when the penalty term is larger than that of the AIC [37].

The WA is also good: a tendency to under fit when the sources are close and correct decisions when sources are distant. The tendency to under fit for close sources is less for colored pure error. The correct decisions of the WL remain around 60% when pure error is white, and remain around 70% when the pure error is colored. The improvement of performance in both the WA and WL in the colored pure error condition could be caused by too ‘optimistic’ parameter standard error estimates. In GLS the standard errors are too small if the ratio of the number of trials (500) to the number of estimated covariances (11325) is poor [14]. Therefore, the WA and WL become more often significant and thus select the two source model more often if the pure error is colored.

If an MSP selected the correct number of sources, then the source parameter estimates should be accurate (unbiased). This is the case for all MSP’s. That is: if an MSP selects the correct number of sources, then the source parameters are unbiased.

In summary, the BIC and WA are the best candidates for model selection. First, because they under fit when sources were close. Second, they have the highest percentages of correct decisions at larger distances. Third, the percentage of over fitting remains very low at all angles and in both pure error conditions. This means that it is unlikely that pure error is modelled. The aT^2 also shows this pattern, although the percentage of correct decisions rises only at larger distances.

IV. APPLICATION OF MSP’S TO VEF’S

To illustrate the use of MSP’s in experimental research, the procedures are applied to visual evoked field (VEF) data. From previous research it was expected that two sources generated the VEF [38].

A. Method

MEG data were recorded with a CTF-gradiometer system with 151 sensors. Head position was monitored with the same 151 sensors and repeatedly activated coils at the fiducial points (nasion, left, and right ear). The subject was

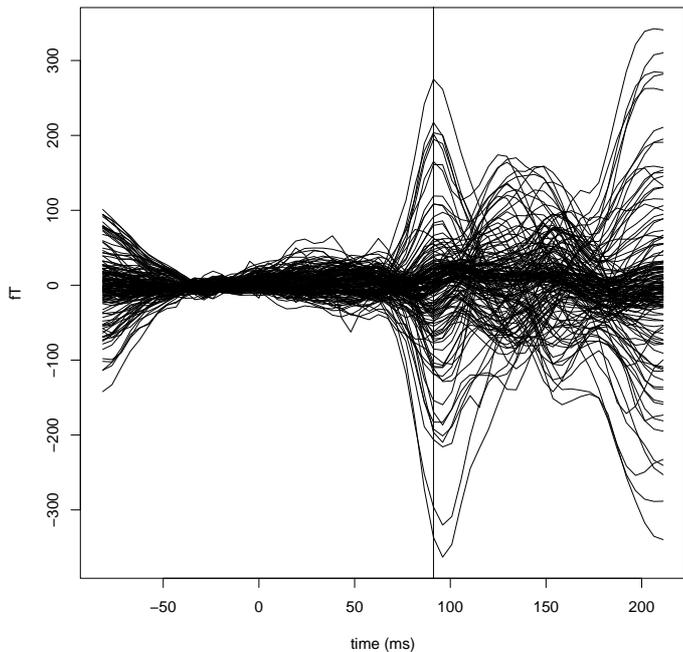


Fig. 3. The averaged magnetic brain responses of the 151 sensors.

presented with 3600 visual stimuli of 50 ms duration with stimulus onset asynchronies between 225 and 285 ms. Here the data are analyzed on 400 black-and-white checkerboard patterns with a 4.8 cycles per degree spatial frequency and subtending 6.7 degrees of visual angle. No task was performed by the subject. An interval of 300 ms was recorded, of which 82 ms were before stimulus presentation. The data were 70 Hz low pass-band filtered, downsampled to 4.8 ms intervals (208.3 Hz), and baseline corrected from -50 to 0 ms. In Fig. 3 the magnetic brain responses are shown after averaging over 383 trials. The time point of interest for analysis was 91.2 ms post stimulus onset (0 ms). The pure error was correlated with absolute correlations between 0.00 and 0.94 with an average of 0.24. Source parameters were estimated by GLS for one, two and three sources in a spherical head model of radius 9.5 cm. The center of the sphere was determined from an MRI-scan of the subject. Subsequently, the MSP’s were applied to the fitted models.

B. Results

In Table II are the source parameters for one, two and three sources. The two source model has two sources near theinion and on either side of the sagittal plane. It can be seen that if one source is estimated, then it is near one of the sources of the two source model (2-1) but more eccentric and with a lower amplitude. In the three source model, one source estimate (3-2) is similar to one of the two source model (2-2), one estimate is more anterior (3-1), and one is quite deep (3-3). All estimates of the three source model have lower amplitudes than the two source model.

The results of all MSP’s are given in Table III. It can be seen that the best MSP’s, the WA and BIC, agree: the number of sources should be 2. This supports the hypothe-

TABLE II

SOURCE PARAMETERS AS ESTIMATED BY GLS FROM THE VEF-DATA AT 91.2 MS FOR 1, 2, AND 3 SOURCES. SOURCE LOCATION PARAMETERS (x , y , z) ARE IN CM WITH RESPECT TO THE ORIGIN OF THE SPHERE, ORIENTATION PARAMETERS (x_0 , y_0 , z_0) ARE NORMALIZED, AND THE AMPLITUDE a IS IN nAM. THE DIRECTION OF THE AXES FOR BOTH THE LOCATION AND ORIENTATION PARAMETERS ARE: THE POSITIVE x IS DIRECTED ANTERIORLY, THE POSITIVE y IS DIRECTED TO THE RIGHT, AND THE POSITIVE z IS DIRECTED UPWARDS.

model	location (cm)			orientation			
	x	y	z	x_0	y_0	z_0	a
1	-8.18	-2.84	-1.04	-0.88	0.77	-0.62	0.21
2-1	-6.75	-1.74	-1.90	-0.05	0.82	-0.57	1.93
2-2	-6.47	1.14	-2.29	-0.07	-0.95	-0.29	2.46
3-1	-4.95	-5.14	-2.31	-0.50	0.09	0.86	0.46
3-2	-7.57	2.41	-2.42	-0.24	-0.95	-0.19	0.34
3-3	0.13	-3.77	5.46	-0.73	-0.57	-0.38	0.12

TABLE III

THE ESTIMATED NUMBER OF SOURCES FOR THE VEF DATA AT 90.8 MS. A MAXIMUM OF 3 SOURCES WAS ESTIMATED.

	RV	χ^2	LOF	T^2	aT^2	LR	aLR	C_p	AIC	BIC	WA	WL
GLS	3	3	3	3	3	3	3	2	2	2	2	2

sis that two sources are required to describe the VEF data adequately. The other tests either indicate that at least three sources are required (RV, χ^2 , LOF, T^2 , aT^2 , LR, and aLR) or confirm the hypothesis that two sources are necessary (C_p , AIC, and WL).

V. DISCUSSION

Several MSP's were compared in their effectiveness to determine the number of sources. From the theoretical analysis, it follows that the aT^2 should be the best candidate and RV the worst. The RV was indeed suboptimal in the simulations: it tended to over fit severely in both the white and colored pure error case. However, the aT^2 appeared to be very conservative in the simulation. The simulations indicated that the BIC and WA are less conservative and more trustworthy, even though not all theoretical criteria are satisfied. Therefore, it is recommended that the aT^2 is used if a conservative approach is taken, and that the BIC or WA are used if a more liberal approach is desirable.

The sources at angles $10^\circ - 20^\circ$ were difficult to separate. This was also found in [2] and [39] in similar circumstances. In [2] the authors suggest that modelling two close sources as one can be advantageous, in the sense that standard errors of the estimates are smaller. In this respect, it is preferable in empirical situations to under fit if it is reasonable to assume that sources are close. For the BIC and WA, 40° sources appeared to be separable. However, the graphs of correct decisions in Fig. 1 and 2 may shift to the right. For example, if the pure error variance were higher, if EEG were used, if fewer sensors were used, or if less trials were used.

The correct head model was used in the simulations. In practice, errors from the head model are to be expected. These errors will be added to the modelling error. This

will decrease the performance of the MSP's. However, the ranking of the MSP's according to their effectiveness is expected to remain the same.

Finally, estimating the pure error covariances resulted, in general, in a poorer performance of the MSP's. This effect is due to the errors of estimating the pure error covariances and not to prewhitening itself. This can be seen by inserting the true pure error covariance matrix (a constant) instead of an estimate. This renders the pure error uncorrelated, but has no effect on the relative performance of the MSP's. One way to reduce these estimation errors is by modelling the pure error covariances [40], [41].

ACKNOWLEDGEMENTS

The authors would like to thank the staff of the MEG center in Amsterdam for their support in recording the data.

REFERENCES

- [1] M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knutila, and O.V. Lounasmaa. Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. *Review of Modern Physics*, 65:413–497, 1993.
- [2] S. Supek and C.J. Aine. Simulation studies of multiple dipole neuromagnetic source localization: model order and limits of source resolution. *IEEE Transactions on Biomedical Engineering*, 40:529–540, 1993.
- [3] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Signal Processing*, 33:387–392, 1985.
- [4] L.C. Zhao, P.R. Krishnaiah, and Z.D. Bai. On detection of the number of signals when the noise covariance matrix is arbitrary. *Journal of Multivariate Analysis*, 20:26–49, 1986.
- [5] J.C. de Munck. The estimation of time varying dipoles on the basis of evoked potentials. *Electroencephalography and Clinical Neurophysiology*, 77:156–160, 1990.
- [6] K.M. Wong, Q.-T. Zhang, J.P. Reilly, and P.C. Yip. On information theoretic criteria for determining the number of signals in high resolution array processing. *IEEE Transactions on Signal Processing*, 18:1959–1971, 1990.

- [7] Q.T. Zhang and K.M. Wong. Information theoretic criteria for the determination of the number of signals in spatially correlated noise. *IEEE Transactions on Signal Processing*, 41:1652–1663, 1993.
- [8] P. Stoica and M. Cedervall. Detection tests for array processing in unknown correlated noise fields. *IEEE Transactions on Signal Processing*, 45:2351–2362, 1997.
- [9] T.R. Knösche, E.M. Berends, H.R.A. Jagers, and M.J. Peters. Determining the number of independent sources of the eeg: A simulation study on information criteria. *Brain Topography*, 11:111–124, 1998.
- [10] J.C. Mosher and R.M. Leahy. Recursive music: A framework for eeg and meg source localization. *IEEE Transactions on Biomedical Engineering*, 45:1342–1354, 1998.
- [11] M. Scherg. Fundamentals of dipole source potential analysis. In F. Grandori, M. Hoke, and G.L. Romani, editors, *Auditory evoked magnetic fields and electric potentials*, pages 40–69. Basel: Karger, 1990.
- [12] H.M. Huizenga and P.C.M. Molenaar. Estimating and testing the sources of evoked potentials in the brain. *Multivariate Behavioral Research*, 29:237–267, 1994.
- [13] G.A.F. Seber and C.J. Wild. *Nonlinear regression*. Toronto: John Wiley and Sons, 1989.
- [14] H.M. Huizenga and P.C.M. Molenaar. Equivalent source estimation of scalp potential fields contaminated by heteroscedastic and correlated noise. *Brain topography*, 8:13–33, 1995.
- [15] J. R. Schott. *Matrix analysis for statistics*. New York: John Wiley & Sons, 1997.
- [16] B.N. Cuffin. A comparison of moving dipole inverse solutions using EEG's and MEG's. *IEEE Transactions on Biomedical Engineering*, 32:905–910, 1985.
- [17] M. Scherg. Functional imaging and localization of electromagnetic brain activity. *Brain Topography*, 5:103–111, 1992.
- [18] T.O. Kvålseth. Cautionary note about R^2 . *The American Statistician*, 39:279–285, 1985.
- [19] R. Anderson-Sprecher. Model comparison and R^2 . *The American Statistician*, 48:113–117, 1994.
- [20] J.B. Willet and J.D. Singer. Another cautionary note about R^2 : Its use in weighted least-squares regression analysis. *The American Statistician*, 42:236–238, 1988.
- [21] J.W. Neill. Testing for lack of fit in nonlinear regression. *The Annals of Statistics*, 16:733–740, 1988.
- [22] M. Bilodeau and D. Brenner. *Theory of multivariate statistics*. New York: Springer-Verlag, 1999.
- [23] T.W. Anderson. *An introduction to multivariate statistical analysis*. New York: John Wiley and Sons, 1958.
- [24] L.J. Waldorp. Statistical model selection procedures. Technical report, Department of psychology, University of Amsterdam, May 2002.
- [25] R.F. Engle. Wald, likelihood ratio, and lagrange multiplier tests in econometrics. In Z. Gilliches and M.D. Intriligator, editors, *Handbook of Econometrics, Volume*, pages 775–826. Amsterdam: Elsevier science publishers BV, 1984.
- [26] T.S. Ferguson. *A course in large sample theory*. Bury st Edmunds: Chapman and Hall, 1996.
- [27] T. Ryan. *Modern regression methods*. New York: John Wiley and Sons, 1997.
- [28] H. Akaike. Information and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csáki, editors, *Proceedings of the second international symposium on information theory. Supplement to problems of control and information theory*, pages 267–281. Akademiai Kiado: Budapest, 1973.
- [29] H. Bozdogan. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52:345–370, 1987.
- [30] M. Aerts, G. Claeskens, and J.D. Hart. Testing the fit of a parametric function. *Journal of the American Statistical Association*, 94:869–879, 1999.
- [31] G. Schwartz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.
- [32] G.C. Chow. A comparison of information and posterior probability criteria for model selection. *Journal of Econometrics*, 16:21–33, 1981.
- [33] H.M. Huizenga and P.C.M. Molenaar. Ordinary least squares dipole localization is influenced by the reference. *Electroencephalography and Clinical Neurophysiology*, 99:562–567, 1996.
- [34] H.M. Huizenga, D.J. Heslenfeld, and P.C.M. Molenaar. Optimal measurement conditions for spatiotemporal EEG/MEG source analysis. *Psychometrika (in Press)*.
- [35] J.C. de Munck, H.M. Huizenga, L.J. Waldorp, and R.M. Heethaar. Estimating stationary dipoles from meg/eeg data contaminated with spatially and temporally correlated background noise. *IEEE Transactions on Signal Processing*, 50:1565–1572, 2002.
- [36] J. Shao. An asymptotic theory for linear model selection. *Statistica Sinica*, 7:221–264, 1997.
- [37] P. Zhang. On the distributional properties of model selection criteria. *Journal of the American Statistical Association*, 87:732–737, 1992.
- [38] J.L. Kenemans, J.M.P. Baas, G.R. Mangun, M. Lijffijt, and M.N. Verbaten. On the processing of spatial frequencies as revealed by evoked-potential source modeling. *Clinical Neurophysiology*, 111:1113–1123, 2000.
- [39] B. Lütkenhöner. Dipole separability in a neuromagnetic source analysis. *IEEE Transactions on Biomedical Engineering*, 45:572–581, 1998.
- [40] L.J. Waldorp, H.M. Huizenga, C.V. Dolan, and P.C.M. Molenaar. Estimated generalized least squares electromagnetic source analysis based on a parametric noise covariance model. *IEEE Transactions on Biomedical Engineering*, 48:737–741, 2001.
- [41] H.M. Huizenga, J.C. de Munck, L.J. Waldorp, and R.P.P.P. Grasman. Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model. *IEEE Transactions on Biomedical Engineering*, 49:533–539, 2002.