

Model Selection in Spatio-Temporal Electromagnetic Source Analysis

Lourens J. Waldorp*, *Member, IEEE*, Hilde M. Huizenga, Arye Nehorai, *Fellow, IEEE*,
Raoul P. P. Grasman, *Member, IEEE*, and Peter C. M. Molenaar

Abstract—Several methods [model selection procedures (MSPs)] to determine the number of sources in electroencephalogram (EEG) and magnetoencephalogram (MEG) data have previously been investigated in an instantaneous analysis. In this paper, these MSPs are extended to a spatio-temporal analysis if possible. It is seen that the residual variance (RV) tends to overestimate the number of sources. The Akaike information criterion (AIC) and the Wald test on amplitudes (WA) and the Wald test on locations (WL) have the highest probabilities of selecting the correct number of sources. The WA has the advantage that it offers the opportunity to test which source is active at which time sample.

Index Terms—Akaike criterion, Bayesian criterion, EEG, MEG, model order selection, source localization, testing source activity.

I. INTRODUCTION

TO describe the underlying processes of electroencephalogram (EEG) and magnetoencephalogram (MEG) data in terms of relatively few dipolar sources, an *a priori* estimate of the number of sources is required [1]–[3]. In [4], several methods [model selection procedures (MSPs)] were examined for instantaneous electromagnetic source analysis (EMSA) to obtain such an estimate, including the residual variance (RV). It was found that the RV had a high probability of selecting too many sources (over-fit). It was also found that alternatives like the adjusted Hotelling's test, the Bayesian information criterion (BIC), and the Wald tests (amplitudes and locations) were effective, that is they had a high probability to select too few sources (under-fit) when the true sources were close, and had a high probability of selecting the correct number of sources when the true sources were more distant.

In this paper, the effective MSPs from [4] will be extended to spatio-temporal EEG and MEG source analysis, if possible. The MSPs that will be extended are the RV (for comparison), the Akaike information criterion (AIC), the BIC, and the Wald tests, on amplitudes (WA) and locations (WL). Although the adjusted Hotelling's test performed well, its distribution cannot

be determined easily if the sample (nonparametric) estimate of the residual covariance matrix is unavailable. In spatio-temporal source analysis, the nonparametric residual covariance matrix is not available since the dimensions of this matrix are simply too large compared to the number of trials. The effective parameterized alternative offered in [9] and [10] do not lead to a similar distribution of the residual covariance matrix (Wishart distribution) as in the instantaneous case. Consequently, the adjusted Hotelling's test cannot be extended to a spatio-temporal analysis. (Asymptotically Hotelling's test would be equal to the chi-square test, which was seen to be ineffective in [4].)

As was seen in [4] the nonparametric estimate of the residual covariance matrix diminished the performance of several MSPs. It was, however, not clear to what extent the estimation errors in the covariance matrix were responsible for this effect. The parameterized covariance matrix offers the opportunity to investigate the effect of the estimation errors. To this end, the asymptotic distribution of the parameters of the residual covariance matrix (which is Gaussian) is determined to create samples of this matrix with estimation errors. This so called parametric bootstrap is similar to [11] in a Bayesian setting. The size of the estimation errors is expected to be reflected in the performance of the MSPs.

This paper is organized as follows. First, the spatio-temporal model, its assumptions, and the MSPs are specified in Section II. This is followed by a simulation in Section III showing the performance of the MSPs in spatio-temporal data and the effect of using an estimate of the covariance matrix on the performance of the MSPs. Finally, the results are discussed in section Section IV.

II. METHOD

Although the following methods are discussed in terms of MEG data, they can be applied to EEG and combined EEG and MEG data as well. The MEG data from m sensors and t samples are collected for each independent trial $j = 1, \dots, n$ in the $m \times t$ matrix $\mathbf{Y}_j = (\mathbf{y}_{1j}, \dots, \mathbf{y}_{tj})$, with $\mathbf{y}_{ij} = (y_{1ij}, \dots, y_{mij})'$. Denote the average over trials by $\bar{\mathbf{Y}} = (1/n) \sum_{j=1}^n \mathbf{Y}_j$. It is assumed that both the location and orientation of the d dipolar sources are fixed over time (samples), and only the amplitudes are allowed to vary over samples [1], [9]. The model for the averaged data is then

$$\bar{\mathbf{Y}} = \mathbf{G}\mathbf{A} + \mathbf{E}. \quad (1)$$

The $m \times d$ matrix \mathbf{G} contains the sensor gains of d sources with unit amplitudes. The matrix \mathbf{G} depends on the location and orientation parameters of the dipolar sources, denoted by $\boldsymbol{\tau}$ and

Manuscript received January 5, 2004; revised August 5, 2004. This work was supported in part by the Netherlands Organization for Scientific Research (NWO) foundation for Behavioral and Educational Sciences. The work of H. M. Huizenga was supported in part by a KNAW fellowship. *Asterisk indicates corresponding author.*

*L. J. Waldorp is with the Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands (e-mail: waldorp@psy.uva.nl).

H. M. Huizenga, R. P. P. Grasman, and P. C. M. Molenaar are with the Department of Psychology, University of Amsterdam, 1018 WB Amsterdam, The Netherlands.

A. Nehorai is with the Department of Electrical and Computing Engineering, University of Illinois at Chicago, Chicago, IL 60607 USA.

Digital Object Identifier 10.1109/TBME.2004.842982

ξ , respectively, [9]. The $d \times t$ matrix \mathbf{A} contains the time varying amplitudes of d sources, and the $m \times t$ matrix \mathbf{E} is the residual. The residual consists of pure error (e.g., spontaneous EEG or MEG) and modeling error (e.g., due to an incorrect number of sources). The number of free parameters in MEG for d sources and t samples is $p = 5d + td$.

If the pure error is white (i.e., uncorrelated and homoscedastic), then its covariance matrix is proportional to the identity matrix, and the source parameters can be obtained by ordinary least squares (OLS). An OLS estimate $\hat{\boldsymbol{\theta}} = (\text{vec}(\hat{\mathbf{A}})', \hat{\mathbf{r}}', \hat{\boldsymbol{\xi}})'$ minimizes the squared residuals $r^2(\boldsymbol{\theta}) = \text{tr}[(\bar{\mathbf{Y}} - \mathbf{GA})(\bar{\mathbf{Y}} - \mathbf{GA})']$, with $\text{vec}(\cdot)$ the operator that stacks the columns of a matrix [12], and $\text{tr}[\cdot]$ the trace of a matrix. At the minimum $s^2(\hat{\boldsymbol{\theta}}) = r^2(\hat{\boldsymbol{\theta}})/(mt - p)$ is an estimate of the residual variance.

If on the other hand the pure error is colored (i.e., correlated and heteroscedastic), then it is assumed that the covariance can be structured as $(1/n)(\boldsymbol{\Upsilon} \otimes \boldsymbol{\Sigma})$, where $\boldsymbol{\Upsilon}$ is the temporal correlation and $\boldsymbol{\Sigma}$ the spatial covariance matrix, and \otimes denotes the Kronecker product [9]. The temporal and spatial matrices can be further parameterized. The temporal correlation matrix $\boldsymbol{\Upsilon}(\beta)$ is assumed to be induced by an AR(1) process. Therefore, $\boldsymbol{\Upsilon}(\beta)$ has a Toeplitz structure with diagonals $v_k = \beta^k$ with $|\beta| < 1$ the coefficient of the AR(1) process, and $k = 0, \dots, t - 1$ the lag [13, p. 36]. For the spatial matrix $\boldsymbol{\Sigma}(\boldsymbol{\omega})$ the model $\sigma_{ij} = \kappa_i \kappa_j \exp(-d_{ij}/\alpha)$ is used, with $\kappa_j > 0$, $j = 1, \dots, m$, the standard deviation of the pure error at sensor j , d_{ij} the Euclidean distance between sensors i and j , $\alpha > 0$ the correlation strength [14], and $\boldsymbol{\omega} = (\kappa_1, \dots, \kappa_m, \alpha)'$. The parameters of both the spatial and temporal matrix are collected in the $m+2$ vector $\boldsymbol{\zeta} = (\beta, \boldsymbol{\omega})'$. There is some empirical evidence for the adequacy of both the Kronecker structure [10] and the further parametrization of the spatial [15] and temporal matrices [9].

In the case of colored pure error the source parameters are estimated by (estimated) generalized least squares (GLS) [16]. In GLS, the data and model are prewhitened to render the pure error white. Assume that an estimate of the covariance parameters is available (see, e.g., [15]), and denote the ensuing temporal and spatial matrices, respectively, by $\mathbf{U} = \boldsymbol{\Upsilon}(\hat{\beta})$ and $s^2\mathbf{V} = \mathbf{S} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\omega}})$, with s^2 the average of the diagonal of \mathbf{S} [4]. The data and model can be prewhitened by the inverse of the Cholesky factor of the pure error covariance matrix. Using the Kronecker structure for the covariance matrix, the prewhitened data and model are $\mathbf{V}^{-1/2}\bar{\mathbf{Y}}(\mathbf{U}^{-1/2})'$ and $\mathbf{V}^{-1/2}\mathbf{GA}(\mathbf{U}^{-1/2})'$, with $\mathbf{U}^{-1/2}$ and $\mathbf{V}^{-1/2}$ the inverse of the Cholesky factors of the temporal and spatial matrices, respectively. A GLS estimate $\hat{\boldsymbol{\theta}}$ minimizes $\tilde{r}^2(\boldsymbol{\theta}) = \text{tr}[\mathbf{V}^{-1}(\bar{\mathbf{Y}} - \mathbf{GA})\mathbf{U}^{-1}(\bar{\mathbf{Y}} - \mathbf{GA})']$. At the minimum $\tilde{s}^2(\hat{\boldsymbol{\theta}}) = \tilde{r}^2(\hat{\boldsymbol{\theta}})/(mt - p)$ is an estimate of the residual variance.

A. Model Selection Procedures

In this section, the MSPs are briefly discussed. Each of the following MSPs is discussed more elaborately in [4]. For each MSP the version is given for white pure error. The version for colored pure error is obtained by prewhitening the data and model as outlined in the previous section.

The RV compares the squared residuals to the squared data for all sensors and samples simultaneously. The RV decreases as a function of the number of parameters, and will therefore over-fit easily. The RV is defined as

$$\text{RV} = 100 \frac{\text{tr}[(\bar{\mathbf{Y}} - \mathbf{GA})(\bar{\mathbf{Y}} - \mathbf{GA})']}{\text{tr}[\bar{\mathbf{Y}}\bar{\mathbf{Y}}]}. \quad (2)$$

A model is said to fit if the RV is below a certain threshold [17], for example 5%.

The AIC penalizes the log-likelihood function for additional parameters (i.c. source parameters) that are required to describe the data [18]. In doing so, the number of sources should be limited since additional sources will at one point hardly decrease the log-likelihood function but increase the penalty. The AIC is defined as

$$\text{AIC} = nmt \ln(\pi n^{-1} s^2) + n^{-1} s^{-2} \text{tr}[(\bar{\mathbf{Y}} - \mathbf{GA})(\bar{\mathbf{Y}} - \mathbf{GA})'] + 2p, \quad (3)$$

where the log-likelihood function has been multiplied by -2 . The model with the smallest AIC value is selected.

The BIC penalizes the log-likelihood function for additional parameters more severely than the AIC [19]. Consequently the BIC is more conservative than the AIC and should lead to less over-fitting. The BIC is defined as

$$\text{BIC} = nmt \ln(\pi n^{-1} s^2) + n^{-1} s^{-2} \text{tr}[(\bar{\mathbf{Y}} - \mathbf{GA})(\bar{\mathbf{Y}} - \mathbf{GA})'] + p \ln(mt). \quad (4)$$

The model with the smallest BIC value is selected.

The Wald test gives the opportunity to test a hypothesis on a specific subset of the parameters [20]. It can be tested whether the amplitudes of the sources are significantly larger than zero, in which case these sources should be included in the model. Alternatively, it can be tested whether the locations of sources differ from each other significantly, in which case these sources should also be included in the model.

Let \mathbf{r} be a q vector function of the source parameters (i.c. amplitudes or locations), \mathbf{r}_h the q vector of fixed hypothesized value of \mathbf{r} , \mathbf{R} the $q \times k$ Jacobian matrix of \mathbf{r} with respect to k source parameters, and \mathbf{C} the $k \times k$ covariance matrix of source parameters (amplitudes or locations, see [9]). Then the Wald test is defined as [21]

$$\mathbf{W} = \frac{1}{q} (\mathbf{r} - \mathbf{r}_h)' (\mathbf{RC}^{-1}\mathbf{R}')^{-1} (\mathbf{r} - \mathbf{r}_h). \quad (5)$$

If the hypotheses are correct then \mathbf{W} is approximately distributed as $F(q, mt - p)$ and is tested at significance level α .

In the WA test, the td vector $\mathbf{r} = \text{vec}(\hat{\mathbf{A}})$, $\mathbf{R} = \mathbf{I}_{td}$, $\mathbf{r}_h = \mathbf{0}$, and $\mathbf{C} = \mathbf{C}(\hat{\mathbf{A}})$ is the covariance matrix of the amplitudes with dimensions $td \times td$. Subsequent semi-univariate tests are performed at α/d (for each source separately but for all samples simultaneously), and univariate tests at α/td (for each source and sample separately). A source is included if the following is true: the multivariate test is significant, the semi-univariate tests are significant for all sources, and the univariate tests are significant at least at one sample for all sources.

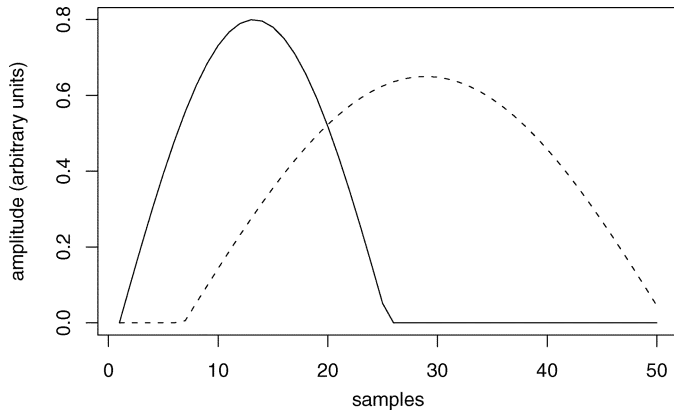


Fig. 1. The amplitude functions used in the simulation for the two sources.

In the WL test, the $3d(d-1)/2$ vector $\mathbf{r} = \text{vec}(\hat{\boldsymbol{\tau}}_1 - \hat{\boldsymbol{\tau}}_2, \hat{\boldsymbol{\tau}}_1 - \hat{\boldsymbol{\tau}}_3, \dots, \hat{\boldsymbol{\tau}}_{d-1} - \hat{\boldsymbol{\tau}}_d)$, where $\hat{\boldsymbol{\tau}}_i$ denotes the location vector of the i th source, the $3d(d-1)/2 \times 3d$ Jacobian matrix $\mathbf{R} = \partial \mathbf{r} / \partial \boldsymbol{\tau}$, with $\boldsymbol{\tau}$ the locations of all sources, $\mathbf{r}_h = \mathbf{0}$, and $\mathbf{C} = \mathbf{C}(\hat{\boldsymbol{\tau}})$ is the covariance matrix of the source locations with dimensions $3d \times 3d$. Subsequent semi-univariate tests are performed at $\alpha/[d(d-1)/2]$ (for each source-pair separately). A source is included if the following is true: the multivariate test is significant, and the semi-univariate tests are significant for all sources.

III. SIMULATIONS

Similar to the simulations in [4], MEG data are generated by two sources with increasing distance, and the pure error is either white or colored. If the pure error is colored a bootstrap covariance matrix is used to examine the effect of estimation errors of the covariance matrix on the performance of the MSPs. The MSPs are evaluated according to the probabilities of three types of decisions: correct, under-fit, and over-fit.

A. Design

The signal on 61 sensors (see [22]) is generated by two dipoles inside a unit sphere. Both sources are in the midcoronal plane (crossing both ears and the vertex) with a varying angle $\gamma = 5^\circ - 20^\circ$, in degrees, with steps of 5° , between the location vectors. As an indication of the distance between the sources, consider a sphere with radius 10 cm. Then 5° corresponds to 0.17 cm and 20° to 3.5 cm. Both sources were at eccentricity 0.80 in the unit sphere. The orientation of both dipoles is toward the nose and completely tangential. This configuration of the sources corresponds to reasonably well separable sources [23], and ensures that that the angle between the source locations determines the separability of the sources. The amplitude functions for both dipoles for 50 samples are depicted in Fig. 1. The absolute temporal correlation between the sources is 0.31.

Normally distributed pure error, white or colored, is added to the signal at each sensor. The pure error standard deviation for the mean (over trials) is set at 10% of the maximal signal (without error) from the 61 sensors and 50 samples when the angle between the dipoles is 12.5° . A total of 300 simulations with each 300 trials is generated. Instead of estimating the parameters of the covariance model in each simulation, a parametric bootstrap sample is generated, which is computationally

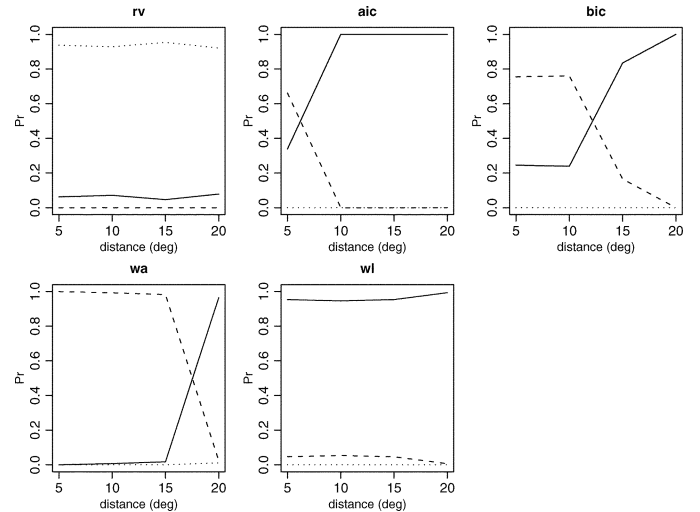


Fig. 2. The probabilities of correct and incorrect decisions in the white pure error case as a function of the distance between the two true sources (in degrees). Solid line: correct (two sources), dashed line: under fit (one source), dotted line: over fit (three sources).

much more convenient for simulation purposes. Colored pure error is obtained from such a parametric bootstrap sample of the covariance parameters for the Kronecker covariance matrix. The true covariance parameters are $\beta = 0.8$, $\alpha = 0.7$, and the κ_i are all equal and are determined according to the 10% criterion for the signal to noise ratio [16]. By using a sample $(\beta_*, \boldsymbol{\omega}'_*)'$ from the asymptotic distribution of the covariance parameters, estimation error is created in the temporal correlation matrix $\boldsymbol{\Upsilon}(\beta_*)$ and the spatial covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\omega}_*)$. The estimator of the covariance parameters is normally distributed with the true covariance parameters as its mean and covariance matrix $\boldsymbol{\Omega}$ (see the Appendix). Then $\psi^2 \boldsymbol{\Omega}$, with ψ equal to the predefined standard deviation, determines the deviation of the parametric bootstrap sample $(\beta_*, \boldsymbol{\omega}'_*)'$ from the true covariance parameters. The difference between the true and parametric bootstrap temporal and spatial matrices reflects the estimation error encountered in standard estimation of the residual covariance matrix. The size of created estimation error is expressed by the average of the absolute difference between the Frobenius norm of the true and the parametric bootstrap covariance matrices. The parametric bootstrap covariance matrix is used in the GLS function to estimate the source parameters and in the MSPs for colored pure error. Two standard deviation levels for the asymptotic distribution are used: $\psi = 1$ and 10, and the corresponding size of the created estimation errors is 0.0017 and 0.0171.

B. Results

To evaluate the MSPs three types of decision are considered: correct (two sources), under-fit (one source), and over-fit (three sources). It should be borne in mind that if the decision is incorrect then under-fit is preferred to over-fit when the sources are close. A desirable pattern is then that when the sources are close the decision of one source has high probability and when they are more distant, the decision of two sources has high probability. In Figs. 2–4, the probabilities of the decisions of the MSPs are shown for white and colored pure error with $\psi = 1$ and 10.

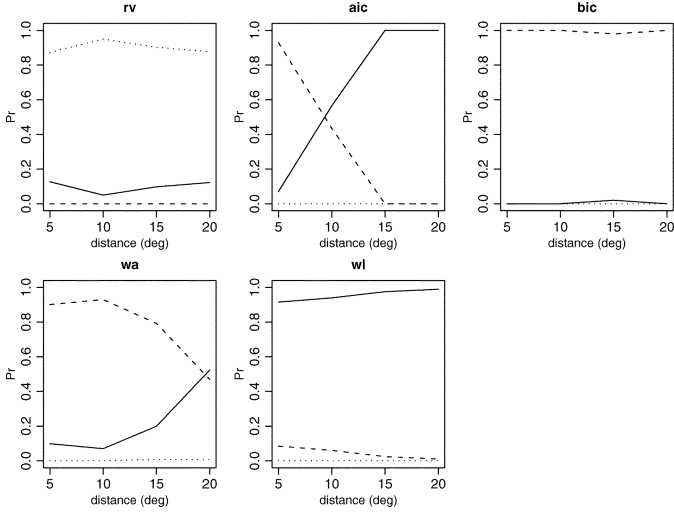


Fig. 3. The probabilities of correct and incorrect decisions in the colored pure error case with $\psi = 1$ as a function of the distance between the two true sources (in degrees). Solid line: correct (two sources), dashed line: under fit (one source), dotted line: over fit (three sources).

It can be seen that the RV tends to over-fit for both white and colored pure error: it has a probability of around 0.90 to select three sources at any angle in each pure error condition.

The AIC and BIC show the desirable pattern when the pure error is white: a high probability of selecting one source when the true sources are close and a high probability of selecting two sources when the true sources are more distant. The AIC shows the same pattern for colored pure error (both conditions) as for white pure error, except that it has a lower probability of a correct decision in colored pure error at $5^\circ - 10^\circ$. The estimation errors influenced the AIC only slightly at 10° : the probability of a correct decision is lower when $\psi = 10$ than when $\psi = 1$. The BIC on the other hand, has a high probability to under-fit at all angles when the pure error is colored for both $\psi = 1$ and 10. The BIC breaks down in colored pure error because the penalty is based on the assumption of independent samples and sensors. The penalty of $p \ln(mt)$ is too severe if the samples and sensors are not completely independent (there are estimation errors in the prewhitening matrix).

The WA shows a similar pattern as the AIC for white and colored pure error, except that the probability of correct decisions starts to increase at 15° (colored pure error) or 20° (white pure error) for the WA instead of 10° for the AIC. The WA seems to be influenced the most by the size of estimation error (ψ). The WL shows the best overall performance: correct decision at all angles and in all pure error conditions.

An advantage of using the WA in spatio-temporal analysis is the possibility of checking the univariate significance levels (p-values) to determine at which samples the sources are active. In Fig. 5, is an example of estimated amplitudes (by GLS) of two sources (left panel) and the corresponding univariate p-values from the WA and the (Bonferroni-corrected) significance level at $0.05/100 = 0.0005$ (right panel), of a single simulation with the sources at $\gamma = 20^\circ$ and with colored pure error ($\psi = 1$). The WA selected two sources in this particular simulation. If the amplitude is high compared to its standard deviation then the

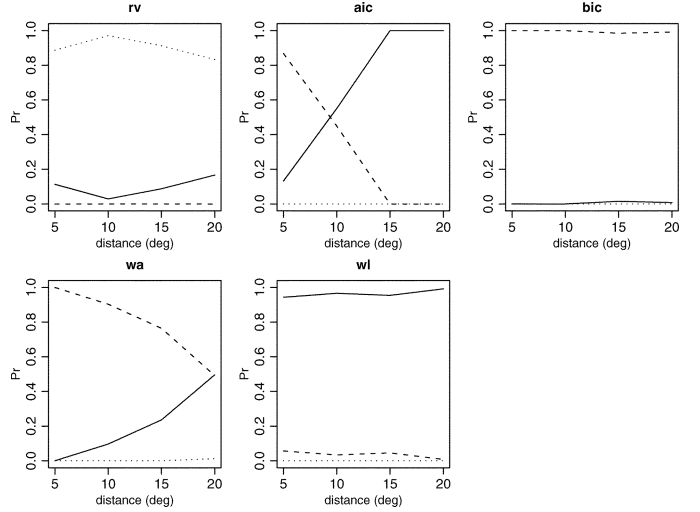


Fig. 4. The probabilities of correct and incorrect decisions in the colored pure error case with $\psi = 10$ as a function of the distance between the two true sources (in degrees). Solid line: correct (two sources), dashed line: under fit (one source), dotted line: over fit (three sources).

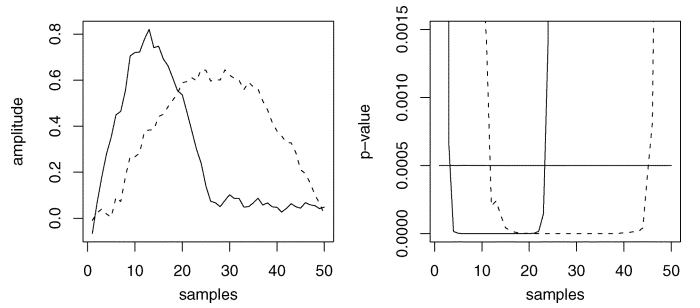


Fig. 5. In the left panel are the amplitude time functions of two estimated sources from data with colored pure error ($\psi = 1$) and estimated by GLS. In the right panel are the corresponding p-values for the sources computed with the WA for each source and each sample, and the Bonferroni-corrected significance level at $0.05/100 = 0.0005$. The line types in the left and right panel correspond to the same dipole.

p-value drops below the significance line, indicating that its amplitude is significantly larger than zero. It can be concluded that the first source is active between samples 4–24 and the second source is active between samples 12–45, which corresponds reasonably well to the true amplitude functions (cf. Fig. 1).

IV. DISCUSSION

The objective was to extend the results for MSPs obtained in [4] to spatio-temporal data in EMSA. Additionally, the influence of estimating the prewhitening matrix on the MSPs was investigated by using a parametric bootstrap procedure akin to a Bayesian method. The WL showed the best overall performance. The AIC and WA showed good performance in the sense that the probability of selecting one source was high when the true sources were close, and the probability of selecting two sources was high when the true sources were more distant. The WA has the advantage that the univariate tests can be used to determine at which samples the sources had amplitudes significantly larger than zero.

MSPs based on the eigenvalue decomposition (or, equivalently, the singular value decomposition, or the principal component analysis) [1], [5]–[7] were not investigated here. This is because it has been shown that this type of MSP is suboptimal when the sources are temporally correlated [8].

It could be argued that the over-fitting of the RV was caused by using a threshold value which was too low. That is, a higher threshold value than the 5% used in the present simulations, might have given good results. Indeed, several sources of additional information might be used to determine the threshold. For example, the baseline could be used in an estimate of the threshold, or it could be reasoned that the threshold should be higher since the pure error on the sensors is correlated. However, there is no clear procedure on how to set the threshold in the RV, and more importantly, how this threshold should depend on the data. Moreover, the RV is dependent on data power, i.e., the RV will more often indicate a bad fit for low amplitude signals than for high amplitude signals [4]. In the present simulation design, the increasing distance between the two true sources, produces a decrease in overall amplitude of the sources. In that case, the RV might have a higher probability of correct decisions when the true sources are close, which is then due to the increased overall amplitude, but it would still over-fit when the true sources are more distant.

The roles of the AIC and BIC are reversed compared to the results in [4] with instantaneous data. The likely cause is the difference in the penalty function of the AIC and BIC in the instantaneous and spatio-temporal model on the one hand, and the effect of estimating the pure error covariance matrix on the other. The penalty of the BIC in the instantaneous model is $5d \ln(m)$ and in the spatio-temporal model $(5d + td) \ln(mt)$. The penalties with two sources are approximately 50.17 and 882.52 for the instantaneous and spatio-temporal model, respectively. Combined with the fact that the residuals are not completely independent in colored pure error (since an estimate is used for prewhitening), the large penalty in spatio-temporal analysis leads to under-fitting. This follows from the fact that the BIC performs well when the pure error is white and performs poorly when the pure error is colored. The penalty of the AIC, on the other hand, depends only on the number of parameters and, therefore, leads to less under-fitting.

In instantaneous analysis, the WL showed a similar flat pattern across distances between the true sources as in spatio-temporal analysis, except that in the spatio-temporal analysis the probability of correct decisions was higher. The better performance of the WL in spatio-temporal analysis than in instantaneous analysis can be explained by the higher precision of the location parameters due to the additional information from the samples (see [9]). Accordingly, the null hypothesis that the locations were the same was more easily rejected when two sources were tested. The higher precision of the location parameters will remain intact even if the correlation between the source amplitude functions were higher. This is because of the definition of the spatio-temporal model, in which only the amplitude parameters are estimated at each time sample.

The created estimation errors in the prewhitening matrix diminished the performance of the AIC and WA, but only slightly, and was less extensive than anticipated. It is likely though that

larger values of ψ than used in the present simulation decreases the performance of the tests further.

These particularities raise the question of the generalizability of the results of the present simulations. For example, different source orientations, different source amplitude functions, or using less trials or sensors might yield other results of the MSPs. Generally, it is expected that the ordering of the MSPs will remain the same given that several factors might deteriorate the bias or variance of the source parameters. This is due to the fact that poorer estimation of the source parameters effects the MSPs generally in the same way.

APPENDIX

The objective in this appendix is to find the distribution of the parameters of the parameterized covariance matrix. This distribution can be used to create parametric bootstrap samples of the covariance parameters (see the text for the details).

The covariance parameters $\zeta = (\beta, \omega')$ in the matrix $\Upsilon(\beta) \otimes \Sigma(\omega)$ can be estimated by maximum-likelihood (ML) [15]. The asymptotic distribution of an ML estimate is normal [24]. The mean and covariance matrix [Cramér-Rao bound (CRB)] of the asymptotic distribution of the ML estimate $\hat{\zeta}$ are required to create the parametric bootstrap samples. Since the model for the covariance matrix $\Upsilon(\beta) \otimes \Sigma(\omega)$ is known, the mean is known and the CRB is proportional to the inverse of the expectation of the matrix of second-order partial derivatives (expected Hessian) of the ML function [12]. In the following, the dependence of the matrices on the parameters is suppressed for convenience of notation. The ML function with which to estimate the parameters is [25]

$$Q(\zeta) = \frac{1}{2} \text{tr}[(\mathbf{C} - \Upsilon \otimes \Sigma)(\Upsilon^{-1} \otimes \Sigma^{-1})]^2$$

where $\mathbf{C} = (1/(n-1)) \sum_{j=1}^n (\mathbf{y}_j - \bar{\mathbf{y}})(\mathbf{y}_j - \bar{\mathbf{y}})'$, and $\mathbf{y}_j = \text{vec}(\mathbf{Y}_j)$ [25]. Let the operator ∂_{φ} denote the first-order partial derivative with respect to φ . The expected Hessian is [25], [26, Prop. 6]

$$E\{\mathbf{H}(\zeta)\} = (\partial_{\zeta} \text{vec}(\Upsilon \otimes \Sigma))' \times [(\Upsilon \otimes \Sigma) \otimes (\Upsilon \otimes \Sigma)]^{-1} (\partial_{\zeta} \text{vec}(\Upsilon \otimes \Sigma)). \quad (\text{A1})$$

Let \mathbf{K}_{tm} be a $tm \times tm$ commutation matrix which allows matrices to commute in a Kronecker product, i.e., $\mathbf{K}_{tm}(\Sigma \otimes \Upsilon)\mathbf{K}_{tm} = \Upsilon \otimes \Sigma$ [12, p. 46]. Then, the first term in (A1) is [12, p. 47]

$$(\partial_{\zeta} \text{vec}(\Upsilon \otimes \Sigma)) = (\mathbf{I}_t \otimes \mathbf{K}_{mt} \otimes \mathbf{I}_m) \times [(\partial_{\zeta} \text{vec}(\Upsilon)) \otimes \text{vec}(\Sigma) + \text{vec}(\Upsilon) \otimes (\partial_{\zeta} \text{vec}(\Sigma))].$$

Let $\Upsilon_K^{-1} = \Upsilon^{-1} \otimes \Upsilon^{-1}$ and $\Sigma_K^{-1} = \Sigma^{-1} \otimes \Sigma^{-1}$. This gives together with the previous result and the above mentioned property of the commutation matrix (A1)

$$(\mathbf{I}_t \otimes \mathbf{K}_{tm} \otimes \mathbf{I}_m)(\Upsilon \otimes (\Sigma \otimes \Upsilon) \otimes \Sigma)^{-1} (\mathbf{I}_t \otimes \mathbf{K}_{mt} \otimes \mathbf{I}_m) = \Upsilon_K^{-1} \otimes \Sigma_K^{-1}.$$

Then, (A1) can be rewritten as

$$E\{\mathbf{H}(\zeta)\} = [(\partial_{\zeta} \text{vec}(\Upsilon)) \otimes \text{vec}(\Sigma) + \text{vec}(\Upsilon) \otimes (\partial_{\zeta} \text{vec}(\Sigma))] \times (\Upsilon_K^{-1} \otimes \Sigma_K^{-1}) \times [(\partial_{\zeta} \text{vec}(\Upsilon)) \otimes \text{vec}(\Sigma) + \text{vec}(\Upsilon) \otimes (\partial_{\zeta} \text{vec}(\Sigma))].$$

Let $\bar{\mathbf{H}}_\varphi$ be a partition of the expected Hessian $E\{\mathbf{H}(\boldsymbol{\zeta})\}$ with the dimensions of the parameter φ . Then, the partitioned expected Hessian is

$$\begin{aligned}\bar{\mathbf{H}}_\beta &= (\partial_\beta \text{vec}(\boldsymbol{\Upsilon}))' \boldsymbol{\Upsilon}_K^{-1} (\partial_\beta \text{vec}(\boldsymbol{\Upsilon})) \otimes \text{vec}(\boldsymbol{\Sigma})' \boldsymbol{\Sigma}_K^{-1} \text{vec}(\boldsymbol{\Sigma}) \\ \bar{\mathbf{H}}_{\beta\omega} &= (\partial_\beta \text{vec}(\boldsymbol{\Upsilon}))' \boldsymbol{\Upsilon}_K^{-1} \text{vec}(\boldsymbol{\Upsilon}) \otimes \text{vec}(\boldsymbol{\Sigma})' \boldsymbol{\Sigma}_K^{-1} (\partial_\omega \text{vec}(\boldsymbol{\Sigma})) \\ \bar{\mathbf{H}}_\omega &= \text{vec}(\boldsymbol{\Upsilon})' \boldsymbol{\Upsilon}_K^{-1} \text{vec}(\boldsymbol{\Upsilon}) \otimes (\partial_\omega \text{vec}(\boldsymbol{\Sigma}))' \boldsymbol{\Sigma}_K^{-1} (\partial_\omega \text{vec}(\boldsymbol{\Sigma})).\end{aligned}$$

Simplifying this result gives $\text{vec}(\boldsymbol{\Sigma})' \boldsymbol{\Sigma}_K^{-1} \text{vec}(\boldsymbol{\Sigma}) = \text{vec}(\boldsymbol{\Sigma}^{-1})$, $\text{vec}(\boldsymbol{\Sigma})' \boldsymbol{\Sigma}_K^{-1} \text{vec}(\boldsymbol{\Sigma}) = \text{tr}[\boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1}] = m^2$, and similarly for the other matrices with $\boldsymbol{\Upsilon}$. It then follows that:

$$\begin{aligned}\bar{\mathbf{H}}_\beta &= m^2 \text{tr}[(\partial_\beta \boldsymbol{\Upsilon})' \boldsymbol{\Upsilon}^{-1} (\partial_\beta \boldsymbol{\Upsilon}) \boldsymbol{\Upsilon}^{-1}] \\ (\bar{\mathbf{H}}_{\beta\omega})_{lj} &= \text{tr}[(\partial_\beta \boldsymbol{\Upsilon})' \boldsymbol{\Upsilon}^{-1}] \text{tr}[\boldsymbol{\Sigma}^{-1} (\partial_{\omega_j} \boldsymbol{\Sigma})] \\ (\bar{\mathbf{H}}_\omega)_{jk} &= t^2 \text{tr}[(\partial_{\omega_j} \boldsymbol{\Sigma})' \boldsymbol{\Sigma}^{-1} (\partial_{\omega_k} \boldsymbol{\Sigma}) \boldsymbol{\Sigma}^{-1}]\end{aligned}$$

where $j, k = 1, \dots, m+1$. Finally, the derivatives $(\partial_\beta \boldsymbol{\Upsilon})$ and $(\partial_{\omega_j} \boldsymbol{\Sigma})$ are required to compute the CRB. Expressed in elements of the matrices the derivatives are

$$\begin{aligned}(\partial_{\kappa_l} \sigma_{ij}) &= \delta_{il} \kappa_j \exp\left(\frac{-d_{ij}}{\alpha}\right) + \delta_{jl} \kappa_i \exp\left(\frac{-d_{ij}}{\alpha}\right) \\ (\partial_\alpha \sigma_{ij}) &= \kappa_i \kappa_j \exp\left(\frac{-d_{ij}}{\alpha}\right) \frac{d_{ij}}{\alpha^2} \\ (\partial_{\beta} v_k) &= k \beta^{k-1}\end{aligned}$$

where $l = 1, \dots, m$, and δ_{jl} is the Kronecker delta (equals 1 if and only if $j = l$ and is 0 otherwise). The CRB of $\hat{\boldsymbol{\zeta}}$ is then $\boldsymbol{\Omega} = \bar{\mathbf{H}}(\boldsymbol{\zeta})^{-1}$ with dimensions $(m+2) \times (m+2)$.

With the asymptotic distribution $N_{m+2}(\boldsymbol{\zeta}, \psi^2 \boldsymbol{\Omega})$ of $\hat{\boldsymbol{\zeta}}$, a vector $\boldsymbol{\zeta}_*$ can be generated with the predefined variance level ψ^2 . This vector contains the parameters for the matrix $\boldsymbol{\Upsilon}(\beta_*) \otimes \boldsymbol{\Sigma}(\omega_*)$ with created sampling errors.

REFERENCES

- [1] J. C. de Munck, "The estimation of time varying dipoles on the basis of evoked potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 77, pp. 156–160, 1990.
- [2] H. M. Huizenga and P. C. M. Molenaar, "Equivalent source estimation of scalp potential fields contaminated by heteroscedastic and correlated noise," *Brain Topogr.*, vol. 8, pp. 13–33, 1995.
- [3] T. Yamazaki, B. W. Van Dijk, and H. Spekreijse, "The accuracy of localizing equivalent dipoles and spatio-temporal correlations of background EEG," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 9, pp. 1114–1121, Sep. 1998.
- [4] L. J. Waldorp, H. M. Huizenga, R. P. P. P. Grasman, B. E. Böcker, J. C. de Munck, and P. C. M. Molenaar, "Model selection in electromagnetic source analysis with an application to VEFs," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, pp. 1121–1129, Mar. 2002.
- [5] 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*, vol. 20, pp. 26–49, 1986.
- [6] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Signal Process.*, vol. SP-33, pp. 387–392, 1985.
- [7] J. C. Mosher, P. S. Lewis, and R. M. Leahy, "Multiple dipole modeling and localization from spatio-temporal MEG data," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 6, pp. 541–557, Jun. 1992.
- [8] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Cao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
- [9] 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 Trans. Biomed. Eng.*, vol. 49, no. 6, pp. 533–539, Jun. 2002.

- [10] 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 Trans. Signal Process.*, vol. 50, no. 7, pp. 1565–1572, Jul. 2002.
- [11] J. Geweke, "Simulation methods for model criticism and robustness analysis," in *Bayesian Statistics*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Eds. Oxford, U.K.: Oxford Univ. Press, 1998, pp. 275–299.
- [12] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Economics, Revised Edition*. Chichester, U.K.: Wiley, 1999.
- [13] C. Chatfield, *The Analysis of Time Series: An Introduction*. London, U.K.: Chapman & Hall, 1989.
- [14] H. M. Huizenga and P. C. M. Molenaar, "Ordinary least squares dipole localization is influenced by the reference," *Electroencephalogr. Clin. Neurophysiol.*, vol. 99, pp. 562–567, 1996.
- [15] 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 Trans. Biomed. Eng.*, vol. 48, no. 6, pp. 737–741, Jun. 2001.
- [16] H. M. Huizenga and P. C. M. Molenaar, "Estimating and testing the sources of evoked potentials in the brain," *Multivariate Behavioral Res.*, vol. 29, pp. 237–267, 1994.
- [17] M. Scherg, "Functional imaging and localization of electromagnetic brain activity," *Brain Topogr.*, vol. 5, pp. 103–111, 1992.
- [18] H. Akaike, "Information and an extension of the maximum likelihood principle," *Proc. 2nd Int. Symp. Information Theory, Suppl. Prob. Control Inf. Theory*, pp. 267–281, 1973.
- [19] G. C. Chow, "A comparison of information and posterior probability criteria for model selection," *J. Econometrics*, vol. 16, pp. 21–33, 1981.
- [20] H. M. Huizenga, D. J. Heslenfeld, and P. C. M. Molenaar, "Optimal measurement conditions for spatiotemporal EEG/MEG source analysis," *Psychometrika*, vol. 67, pp. 299–313, 2002.
- [21] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*. Toronto, ON, Canada: Wiley, 1989.
- [22] H. M. Huizenga, T. L. van Zuijlen, D. J. Heslenfeld, and P. C. M. Molenaar, "Simultaneous MEG and EEG source analysis," *Phys. Med. Biol.*, vol. 46, pp. 1737–1751, 2001.
- [23] B. Lütkenhöner, "Dipole separability in a neuromagnetic source analysis," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 5, pp. 572–581, May 1998.
- [24] A. W. van der Vaart, *Asymptotic Statistics*. New York: Cambridge Univ. Press, 1998.
- [25] K. G. Jöreskog, "Analysis of covariance structures," *Scandinavian Journal of Statistics*, vol. 8, pp. 65–95, 1981.
- [26] M. W. Browne, "Generalized least squares estimators in the analysis of covariance structures," *South African Statistical J.*, vol. 8, pp. 1–24, 1974.



Lourens J. Waldorp (S'00–A'03–M'04) received the masters degree in methodological psychology in 1998 and the Ph.D. degree in 2004 both from the University of Amsterdam, Amsterdam, The Netherlands.

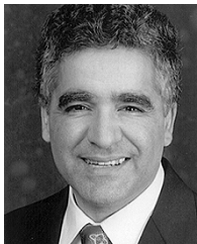
He currently holds a postdoctoral position at both the University of Amsterdam and the University of Maastricht, Maastricht, the Netherlands. His research interests include mathematical statistics, statistics in psychophysiological experiments, signal processing, and neuroscience in general.



Hilde M. Huizenga was born in 1965. She received the MD degree in psychology from the University of Groningen, The Netherlands, in 1990, and the Ph.D. degree (*cum laude*) in psychology from the University of Amsterdam, Amsterdam, The Netherlands, in 1995.

In 1996–2001, she was a Postdoctoral Fellow, currently she is Assistant Professor at the University of Amsterdam. Her main research interest is statistical analysis of neuroscientific data. In particular, non-linear regression and covariance structure analysis

of EEG/MEG sources and their interactions.



Arye Nehorai (S'80–M'83–SM'90–F'94) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion, Haifa, Israel, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

After graduation he worked as a Research Engineer for Systems Control Technology, Inc., in Palo Alto, CA. From 1985 to 1989, he was Assistant Professor and from 1989 to 1995 he was Associate Professor with the Department of Electrical Engineering at Yale University, New Haven, CT. In 1995,

he joined the Department of Electrical Engineering and Computer Science at The University of Illinois at Chicago (UIC), as a Full Professor. From 2000 to 2001, he was Chair of the department's Electrical and Computer Engineering (ECE) Division, which is now a new department. In 2001, he was named University Scholar of the University of Illinois. He holds a joint professorship with the ECE and Bioengineering Departments at UIC. His research interests are in signal processing, communications, and biomedicine.

Dr. Nehorai is Vice President–Publications and Chair of the Publications Board of the IEEE Signal Processing Society. He is also a member of the Board of Governors and of the Executive Committee of this Society. He was Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from January 2000 to December 2002, and is currently a Member of the Editorial Board of *Signal Processing*, the *IEEE Signal Processing Magazine*, and *The Journal of the Franklin Institute*. He is the founder and Guest Editor of the special columns on Leadership Reflections in the *IEEE Signal Processing Magazine*. He has previously been an Associate Editor of IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, IEEE JOURNAL OF OCEANIC ENGINEERING, and *Circuits, Systems, and Signal Processing*. He served as Chairman of the Connecticut IEEE Signal Processing Chapter from 1986 to 1995, and a Founding Member, Vice-Chair, and later Chair of the IEEE Signal Processing Society's Technical Committee on Sensor Array and Multichannel (SAM) Processing from 1998 to 2002. He was the co-General Chair of the First and Second *IEEE SAM Signal Processing Workshops* held in 2000 and 2002. He was co-recipient, with P. Stoica, of the 1989 IEEE Signal Processing Society's Senior Award for Best Paper, and co-author of the 2003 Young Author Best Paper Award of this Society, with A. Dogandzic. He received the Faculty Research Award from the UIC College of Engineering in 1999 and was Adviser of the UIC Outstanding Ph.D. Thesis Award in 2001. He was elected Distinguished Lecturer of the IEEE Signal Processing Society for the term 2004 to 2005. He has been a Fellow of the Royal Statistical Society since 1996.

Raoul P. P. Grasman (S'00–A'03–M'04) was born in 1973. He received a degree in experimental psychology (*cum laude*) and the Ph.D. degree from the University of Amsterdam, Amsterdam, The Netherlands, in 1998 and 2004, respectively.

He currently holds a postdoctoral position at the University of Amsterdam. His research interests concern the methodology of cognitive neuroscience and experimental psychology research.



Peter C. M. Molenaar was born in 1946. His doctoral dissertation was about multidimensional signal analysis.

He is Research Director of several programs and Department Head of the Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands. His current research interests include signal analysis and applied nonlinear dynamics.