

# The spatiotemporal MEG covariance matrix modeled as a sum of Kronecker products

Fetsje Bijma,\* Jan C. de Munck, and Rob M. Heethaar

Department PMT, Vrije Universiteit Medical Center, MEG Center, De Boelelaan 1118, Amsterdam, The Netherlands

Received 15 November 2004; revised 2 March 2005; accepted 4 April 2005

The single Kronecker product (KP) model for the spatiotemporal covariance of MEG residuals is extended to a sum of Kronecker products. This sum of KP is estimated such that it approximates the spatiotemporal sample covariance best in matrix norm. Contrary to the single KP, this extension allows for describing multiple, independent phenomena in the ongoing background activity. Whereas the single KP model can be interpreted by assuming that background activity is generated by randomly distributed dipoles with certain spatial and temporal characteristics, the sum model can be physiologically interpreted by assuming a composite of such processes. Taking enough terms into account, the spatiotemporal sample covariance matrix can be described exactly by this extended model.

In the estimation of the sum of KP model, it appears that the sum of the first 2 KP describes between 67% and 93%. Moreover, these first two terms describe two physiological processes in the background activity: focal, frequency-specific alpha activity, and more widespread non-frequency-specific activity. Furthermore, temporal non-stationarities due to trial-to-trial variations are not clearly visible in the first two terms, and, hence, play only a minor role in the sample covariance matrix in terms of matrix power. Considering the dipole localization, the single KP model appears to describe around 80% of the noise and seems therefore adequate. The emphasis of further improvement of localization accuracy should be on improving the source model rather than the covariance model.

© 2005 Elsevier Inc. All rights reserved.

*Keywords:* Magnetoencephalography; Spatiotemporal covariance; Kronecker product; Alpha activity; Temporal stationarity

---

## Introduction

In MEG measurements, background noise is correlated both in space and in time. Although these correlations are unknown a

priori, they are of interest for two reasons: they contain physiological information and they can be used to improve source localization (de Munck et al., 1992; Dogandziç and Nehorai, 2000; Lütkenhöner, 1998; Sekihara et al., 1994; Waldorp et al., 2002). These spatiotemporal correlations can be estimated from the measured data. The general spatiotemporal covariance matrix, however, has a large dimension, yielding two main problems, the first of which being its estimation and the second its storage. Estimation would require an unrealistically high number of measurements to achieve nonsingularity and storage would require far more memory than commonly available.

A way of resolving these two problems has been found in the parametrization of the spatiotemporal covariance matrix through a Kronecker product (KP) (Langville and Stewart, 2004; Van Loan, 2000) of a spatial and a temporal covariance matrix, reducing its dimensionality considerably (de Munck et al., 1992, 2002; Huizenga et al., 2002). The KP parametrization assumes that an arbitrary spatiotemporal correlation can be modeled as a product of a spatial and a temporal factor. These two factors are independent of each other; hence, the spatial and temporal correlations are separated from each other in the KP model. Physiologically, this model can be interpreted by assuming background noise to be generated by randomly distributed dipolar sources having amplitude functions independent of the source locations (de Munck et al., 1992).

Applications of the KP model in source localization methods have revealed that the accuracy improves when the spatiotemporal correlations instead of no or only spatial correlations are taken into account (de Munck et al., 2002; Huizenga et al., 2002). Nevertheless, there are two important shortcomings of the KP model. The first deficiency is the rigidity of the KP: the shape of the temporal cross spectrum is forced to be fixed over all channels. This is a simplification, as is illustrated by the alpha rhythm: the amount of alpha activity relative to other spontaneous activity is not equally distributed over the head. The second point of debate is trial-to-trial variations, which have been discussed in literature (Coppola et al., 1978; Duann et al., 2002; Gasser et al., 1983; Jaśkowski and Verleger, 1999; Laskaris et al., 2003; Mäkinen et al., 2005; Mocks et al., 1987; Pham et al., 1987). The

---

\* Corresponding author.

E-mail address: f.bijma@vumc.nl (F. Bijma).

Available online on ScienceDirect ([www.sciencedirect.com](http://www.sciencedirect.com)).

Signal Plus Noise (SPN) model in evoked field experiments assumes that no such variations occur in the data. This model can be formulated as

$$m_k^{i,j} = r^{i,j} + e_k^{i,j}, \quad (1)$$

where  $m_k^{i,j}$  is the measurement in trial  $k$  at the  $i$ th sensor and  $j$ th time instant,  $r^{i,j}$  is the trial-independent response to the stimulus at the  $i$ th sensor and  $j$ th time instant  $e_k^{i,j}$  is the residual (“noise”). Hence, the SPN model assumes that all trial-to-trial variations are accounted for by the and  $e_k^{i,j}$ . The response  $r^{i,j}$  is estimated by the average over trials of the measurements  $m_k^{i,j}$ . When incorrectly assumed, i.e., when  $r^{i,j}$  does depend on  $k$ , the SPN model leads to nonstationarity in the temporal covariance (de Munck et al., 2004; Truccolo et al., 2002). Nonetheless, it has been shown for Somatosensory Evoked Field data sets that both the temporal and the spatial covariance estimated under the SPN assumption can be explained by a stationary model (Bijma et al., 2003). The reason may be that the nonstationarities are suppressed by the rigidity of the KP model: if the majority of the channels show temporally stationary signals, this stationarity will dominate the temporal matrix. To overcome its shortcomings, an extension of the KP model is investigated in this study: a sum of Kronecker products.

In the sum of KP model, each term presents a combination of a spatial and a temporal pattern. Unlike the single KP model, the sum model allows for multiple temporal structures with specific spatial patterns, and can, thus, account for temporal nonstationarities in separate terms. The interpretation of the sum of KP model is analogous to that of the single KP model. Assuming that spontaneous background activity is the composite of a number of independent ongoing processes (cf. Laufs et al., 2003), each of which can be described by a random dipole model as explained above, the spatiotemporal covariance matrix becomes a sum of KP.

Although extending the single KP model to a sum of KP may seem rather straightforward, in the practical application in dipole localization it becomes quite delicate. The main problem is the inversion of a sum of Kronecker products, which, contrary to a single KP, cannot be performed by inversion of only the smaller dimensional matrices, but requires inversion in the large spatiotemporal dimension. Therefore, the emphasis in this study is primarily on estimating rather than applying the sum of KP. The aim is twofold: firstly, the estimated sum of KP contains information about the validity of the single KP model for dipole localization; secondly, from the estimated sum physiological information is assembled about the spatial and temporal features in the background activity. Hence, the goal of this paper is not to present an improved method for source localization; the accent is on investigating the spatiotemporal MEG residuals.

In the next section, first the findings and formulas of the single KP model are summarized shortly and then the estimators for the sum of KP model are derived and discussed. In the third section, the sum model is estimated for data sets of three types (VEF, SEF, AEF) and results are shown. In the final section, the results are discussed and conclusions are drawn. The technical details of the model are put together in the appendices in order to keep the text compact.

## Model

### Single KP model

In the single Kronecker product model, the covariance between two MEG residuals,  $e_k^{i,j}$  and  $e_k^{i',j'}$ , is modeled as the product of a temporal and a spatial term:

$$\varepsilon(e_k^{i,j}, e_k^{i',j'}) = X_{i,i'} T_{j,j'} \delta_{k,k'} \quad (2)$$

where  $e_k^{i,j}$  is the MEG residual measured at sensor  $i$ , time sample  $j$  in the  $k$ th trial and  $\delta_{k,k'}$  denotes the Kronecker delta function. Thus, different trials are assumed to be independent. The meaning of Eq. (2) is that the temporal covariance matrix  $T$  is fixed in space and the spatial covariance matrix  $X$  does not vary over time. In other words, space and time are not correlated.

The matrix formula for the Kronecker product model is

$$\Sigma = T \otimes X, \quad (3)$$

where  $T \in \mathbb{R}^{J \times J}$  is the temporal,  $X \in \mathbb{R}^{I \times I}$  the spatial covariance matrix and  $\Sigma \in \mathbb{R}^{IJ \times IJ}$  is the spatiotemporal covariance matrix.  $I$  denotes the number of sensors and  $J$  the number of time samples. The dimensions of these two covariance matrices are much smaller than the dimension of  $\Sigma$ , and by the structure of the Kronecker product, the computations are much less demanding (Magnus and Neudecker, 1995):

$$\Sigma^{-1} = T^{-1} \otimes X^{-1} \quad (4)$$

$$\det(\Sigma) = \det(T)^I \det(X)^J. \quad (5)$$

$X$  and  $T$  can be estimated using either the Maximum Likelihood (ML) paradigm or the Least Squares (LS) method. In the ML case, the MEG residuals are assumed to have a Gaussian distribution with the KP as the covariance matrix. The density function is maximized with respect to the matrices  $X$  and  $T$ . In the LS case, the KP is fitted to the spatiotemporal sample covariance matrix and the difference in Frobenius norm is minimized with respect to  $X$  and  $T$ . The sample covariance matrix  $\Sigma_s \in \mathbb{R}^{IJ \times IJ}$  is defined as

$$\Sigma_s = \frac{1}{K-1} \sum_{k=1}^K \text{vec}(E_k) [\text{vec}(E_k)]^t \quad (6)$$

where  $E_k \in \mathbb{R}^{I \times J}$  is the matrix containing the residuals of trial  $k$ ,

$$(E_k)_{i,j} = e_k^{i,j}, \quad (7)$$

and  $K$  is the number of trials (repetitive measurements). In both the ML and the LS case, the estimators for  $X$  and  $T$  are given by an iterative system. For the ML model, this system is (de Munck et al., 2002):

$$X_{\text{ML}} = \frac{1}{JK} \sum_{k=1}^K E_k \hat{T}_{\text{ML}}^{-1} E_k^t \quad (8)$$

$$\hat{T}_{\text{ML}} = \frac{1}{IK} \sum_{k=1}^K E_k^t \hat{X}_{\text{ML}}^{-1} E_k \quad (9)$$

and for the LS paradigm, see Appendix A.1,

$$\hat{X}_{LS} = \frac{1}{K-1} \frac{1}{\|\hat{T}_{LS}\|^2} \sum_{k=1}^K E_k \hat{T}_{LS} E_k^t \quad (10)$$

$$\hat{T}_{LS} = \frac{1}{K-1} \frac{1}{\|\hat{X}_{LS}\|^2} \sum_{k=1}^K E_k^t \hat{X}_{LS} E_k. \quad (11)$$

For the extension to the sum of KP model, the ML formulas become prohibitively complicated because the sum does not maintain the elegant structure for the inverse and the determinant as the single KP does (see Eqs. (4) and (5)). For the LS paradigm, it appears that extension is possible.

#### Sum KP model

As stated in the Introduction, extending the single KP to a sum of KP allows for a more general spatiotemporal covariance structure. The sum model is expressed as

$$\Sigma = \sum_{n=1}^N T_n \otimes X_n \quad (12)$$

and the corresponding LS cost function is

$$C_{LS} = \|\Sigma_s - \sum_{n=1}^N T_n \otimes X_n\|^2 \quad (13)$$

where  $\|\cdot\|^2$  denotes the Frobenius norm. In order to minimize the cost function, the algorithm presented by Van Loan to find the best sum of KP approximation to a given matrix is used (Van Loan, 2000). The matrix elements in Eqs. (12) and (13) are rearranged according to Van Loan's shuffle operator  $S: \mathbb{R}^{I \times J \times I \times J} \rightarrow \mathbb{R}^{I^2 \times J^2}$  such that Eq. (12) is transformed into

$$S(\Sigma) = \sum_{n=1}^N \text{vec}(X_n) \text{vec}(T_n)^t, \quad (14)$$

and the cost function in Eq. (13) becomes

$$C_{LS} = \|S(\Sigma_s) - \sum_{n=1}^N \text{vec}(X_n) \text{vec}(T_n)^t\|^2. \quad (15)$$

From Appendix A.2, containing more details about  $S$ , the formula for the shuffled sample covariance matrix follows:

$$S(\Sigma_s) = \frac{1}{K-1} \sum_{k=1}^K E_k \otimes E_k. \quad (16)$$

Note that the dimension of this shuffled matrix is  $I^2 \times J^2$ , which is in general not square.

LS estimators for  $X_n$  and  $T_n$ ,  $n = 1, \dots, N$ , are obtained by minimizing  $C_{LS}$  in Eq. (13). This minimization is equivalent to finding the best rank  $N$  approximation of  $S(\Sigma_s)$ , which can be obtained from the Singular Value Decomposition (SVD) of  $S(\Sigma_s)$ . Write the SVD of  $S(\Sigma_s)$  (Golub and Van Loan, 1990; Magnus and Neudecker, 1995)

$$S(\Sigma_s) = U \Delta V^t, \quad (17)$$

where  $U \in \mathbb{R}^{I^2 \times I^2}$  and  $V \in \mathbb{R}^{J^2 \times J^2}$  are orthogonal matrices and  $\Delta = (\Delta_0, \mathbf{0}) \in \mathbb{R}^{I^2 \times J^2}$  and  $\Delta_0 \in \mathbb{R}^{I^2 \times I^2}$  is diagonal. Here, it is

assumed that  $I \leq J$ . In case  $I > J$ , the expression for  $\Delta$  becomes  $\begin{pmatrix} \Delta_0 \\ \mathbf{0} \end{pmatrix}$ , with  $\Delta_0 \in \mathbb{R}^{J^2 \times J^2}$ . The best rank  $N$  approximation of  $S(\Sigma_s)$ , for  $N \leq \min(I, J)$ , is now given by

$$\sum_{n=1}^N U_n \sigma_n V_n^t \quad (18)$$

where  $U_n (V_n)$  denotes the  $n$ th column of  $U (V)$  and  $\sigma_n = \Delta_{n,n}$ , the  $(n, n)$ th entry of  $\Delta_0$ . Hence, the estimators for  $\text{vec}(X_n)$  and  $\text{vec}(T_n)$  are given by

$$\text{vec}(\hat{X}_n) = U_n \quad (19)$$

$$\text{vec}(\hat{T}_n) = \sigma_n V_n \quad (20)$$

for  $1 \leq n \leq N$ . Note that these estimators are not unique: e.g., multiplying Eq. (19) and dividing Eq. (20) by the same constant yields an equivalent solution. Throughout this section, the normalization as in Eqs. (19) and (20) is used. It follows from Eq. (18) that the entire sample covariance matrix can be described by a sum of KP, when  $N$  is taken equal to  $\min(I, J)$ . Furthermore, from the  $\sigma_n$ , the distribution of explained matrix power over the KP terms is obtained:

$$\text{rel pow } n\text{th term} = \frac{\sigma_n^2}{\|S(\Sigma_s)\|^2} \times 100\% = \frac{\sigma_n^2}{\|\Sigma_s\|^2} \times 100\%. \quad (21)$$

Despite the straightforward application of Van Loan's method, the estimators in Eqs. (19) and (20) are not convenient in practice, due to the dimensionality of the desired SVD. Therefore, alternative estimators are deduced below.

The alternative way of estimating the terms  $(T_n, X_n)$  in the KP sum uses Lagrange multipliers (Marsden and Tromba, 1988). To find the best rank  $N$  approximation, it suffices to first find the best rank 1 approximation and successively find all the subsequent terms one after another. The initial term,  $n = 1$ , corresponds to the best rank one approximation of the sample covariance matrix and is estimated by the system of Eqs. (10) and (11). As the higher-order terms are estimated one by one, for any  $p$  satisfying  $1 < p \leq N$ , the first  $p - 1$  terms,  $(T_n, X_n)$  for  $n = 1, \dots, p - 1$ , will have been estimated at the instant of estimation of the  $p$ th term. In other words, the best rank  $p - 1$  approximation of  $S(\Sigma_s)$  is known and the best rank  $p$  approximation has to be estimated. This step is explained in Appendix A.3 from which expressions for  $\hat{X}_p$  and  $\hat{T}_p$  follow:

$$\begin{aligned} \hat{X}_p &= \frac{1}{(K-1)\|\hat{T}_p\|^2} \\ &\times \sum_{k=1}^K \left[ E_k \hat{T}_p E_k^t - \sum_{n=1}^{p-1} \text{tr}[\hat{X}_n E_k \hat{T}_p E_k^t] \hat{X}_n \right] \text{ for } 1 < p \leq N \end{aligned} \quad (22)$$

$$\begin{aligned} \hat{T}_p &= \frac{1}{(K-1)} \\ &\times \sum_{k=1}^K \left[ E_k^t \hat{X}_p E_k - \sum_{n=1}^{p-1} \frac{1}{\|\hat{T}_n\|_F^2} \text{tr}[\hat{T}_n E_k^t \hat{X}_p E_k] \hat{T}_n \right] \text{ for } \\ &1 < p \leq N. \end{aligned} \quad (23)$$

The starting value for this iterative system and for the system in Eqs. (10) and (11) is to set  $T_p = \mathbf{I}_J$  and start with updating  $X_p$ . The iteration stops when the relative difference in matrix power between an estimate and the next estimate is less than  $10^{-12}$ , i.e., when  $\frac{\|A^s - A^{s+1}\|_F^2}{\|A^s\|_F^2} < 10^{-12}$  for both  $A = T_p$  and  $A = X_p$ , where  $A^s$  indicates the  $s$ th estimate in the iteration. To verify the iterative estimators in Eqs. (22) and (23), the SVD estimation method in Eqs. (19) and (20) was applied to small data sets ( $I = 30, J = 30$ ). For these small data sets, the first four iteratively estimated KP terms were compared to the estimated terms from the SVD. This comparison showed that both methods yielded identical matrices.

Note that, although the cost function  $C_{LS}$  in Eq. (13) is expressed in terms of the spatiotemporal sample covariance matrix  $\Sigma_s$ , the solution in the iterative system of Eqs. (22) and (23) does not require the storage of this huge matrix  $\Sigma_s$  in memory. In order to find the solution, apart from the recorded data, one only needs to store the terms  $(X_n, T_n)$  in memory, which is of order  $(I^2 + J^2)$ .

### Rewriting the sum of 2 Kronecker products

The terms in the sum of KP are estimated under the constraint of orthogonal “vecced” matrices, see Eqs. (55) and (56) of Appendix A.3. In order to interpret the matrices  $(X_n, T_n)$  in each term as covariance matrices of the underlying physiological processes, these matrices should be positive definite. However, the orthogonality constraint forces the higher-order terms to be *indefinite* matrices. This can be explained by the following reasoning. The first term consists of two positive definite matrices,  $T_1$  and  $X_1$ , representing the best rank 1 approximation of  $\mathcal{S}(\Sigma_s)$ . Therefore, there exists a nonsingular matrix  $W_1 \in \mathbb{R}^{J \times J}$  such that  $T_1 = W_1 W_1^t > 0$ . To show that any higher-order temporal matrix  $T_n$  for  $n > 1$  must be indefinite, it is demonstrated below that the assumption of  $T_n$  being positive (negative) definite leads to a contradiction. Assuming  $T_n$  to be positive definite implies that there exists a matrix  $W_n \in \mathbb{R}^{J \times J}$  such that  $T_n = W_n W_n^t > 0$ . Substituting  $W_1$  and  $W_n$ , the orthogonality constraint can be written as

$$\begin{aligned} 0 &= \text{vec}(T_1)\text{vec}(T_n)^t = \text{tr}(T_1 T_n) = \text{tr}(W_1 W_1^t W_n W_n^t) \\ &= \text{tr}(W_1^t W_n W_n^t W_1) = \|W_1^t W_n\|^2 > 0 \end{aligned} \quad (24)$$

which is a contradiction. For  $T_n$  negative definite, a similar contradiction can be derived. Hence,  $T_n$  must be indefinite. The same holds true for all the higher-order spatial matrices  $X_n$ . The Kronecker product of two indefinite matrices,  $A$  and  $B$ , is again indefinite, because the eigenvalues of  $A \otimes B$  are given by  $\lambda_i \mu_j$ , all possible combinations of  $\lambda_i$  an eigenvalue of  $A$ , and  $\mu_j$  an eigenvalue of  $B$  (Van Loan, 2000). Consequently, the higher-order terms in the sum,  $T_n \otimes X_n$  for  $n > 1$ , are indefinite. In sum, this implies that the higher-order terms cannot be interpreted as physiologically meaningful covariance matrices. Therefore, the estimated sum is converted to an interpretable sum of KP. For this conversion, the freedom of a best rank  $N$  approximation is exploited.

In general, the freedom in the shuffled sum of  $N$  Kronecker products in Eq. (14) can be exhibited by a nonsingular matrix  $H \in \mathbb{R}^{N \times N}$ :

$$\begin{aligned} \mathcal{S}(\Sigma) &= \sum_{n=1}^N \text{vec}(X_n)\text{vec}(T_n)^t \\ &= (\text{vec}(X_1) \cdots \text{vec}(X_N)) \begin{pmatrix} \text{vec}(T_1)^t \\ \vdots \\ \text{vec}(T_N)^t \end{pmatrix} \\ &= [(\text{vec}(X_1) \cdots \text{vec}(X_N))H] \begin{bmatrix} H^{-1} \begin{pmatrix} \text{vec}(T_1)^t \\ \vdots \\ \text{vec}(T_N)^t \end{pmatrix} \end{bmatrix} \\ &= (\text{vec}(\tilde{X}_1) \cdots \text{vec}(\tilde{X}_N)) \begin{pmatrix} \text{vec}(\tilde{T}_1)^t \\ \vdots \\ \text{vec}(\tilde{T}_N)^t \end{pmatrix}. \end{aligned} \quad (25)$$

In this rewritten expression, the  $\tilde{X}_n$  ( $\tilde{T}_n$ ) matrices are linear combinations of the  $X_n$  ( $T_n$ ) matrices and are thus symmetric. Furthermore, note that this rewriting does not damage the KP structure:  $\sum_s = \sum_n T_n \otimes X_n = \sum_n \tilde{T}_n \otimes X_n$ . To convert the estimated sum to an interpretable sum of KP, one should try to find a matrix  $H$  such that the  $(\tilde{T}_n, \tilde{X}_n)$  are positive (semi-)definite for  $n = 1, \dots, N$ . The remainder of this section concentrates on the special case  $N = 2$ .

For  $N = 2$ , the matrix  $H$  becomes a nonsingular  $(2 \times 2)$  matrix and can be written as

$$H = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (26)$$

with  $ad - bc = 1$  such that  $H^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ . In practice, it appears that the best “orthogonal” sum of two KP is not always positive definite, but usually contains some small negative eigenvalues. Although these values are very small, this indicates that the orthogonally estimated sum of two KP is indefinite. Consequently, it is not possible to rewrite this sum of two KP as a sum of two positive definite KP, because a positive definite sum cannot equal an indefinite expression. Therefore, we seek a matrix  $H$  that is optimal in a slightly different way. Given the estimated sum,  $T_1 \otimes X_1 + T_2 \otimes X_2$ , the shuffled version of the rewritten sum in terms of  $a, b, c$ , and  $d$  is

$$\begin{aligned} &(\text{vec}(X_1) \text{vec}(X_2)) \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} \text{vec}(T_1)^t \\ \text{vec}(T_2)^t \end{pmatrix} \\ &= (a \text{vec}(X_1) + c \text{vec}(X_2) \ b \text{vec}(X_1) + d \text{vec}(X_2)) \\ &\quad \times \begin{pmatrix} d \text{vec}(T_1)^t - b \text{vec}(T_2)^t \\ -c \text{vec}(T_1)^t + a \text{vec}(T_2)^t \end{pmatrix}. \end{aligned}$$

Hence,

$$X_1 = a \tilde{X}_1 + c X_2 \quad (27)$$

$$X_1 = b \tilde{X}_2 + d X_2 \quad (28)$$

$$T_1 = d \tilde{T}_1 - b T_2 \quad (29)$$

$$T_2 = -c T_1 + a \tilde{T}_2. \quad (30)$$

As noticed above, these matrices cannot be all four positive definite. These matrices are uniquely decomposed into a symmetric positive part and a symmetric negative part, the latter of which might be zero for some but not all of the four matrices. This can be expressed as:

$$\tilde{X}_1 = U_{\tilde{X}_1} A_{\tilde{X}_1}^+ U_{\tilde{X}_1}^t - V_{\tilde{X}_1} A_{\tilde{X}_1}^- V_{\tilde{X}_1}^t \quad (31)$$

$$\tilde{X}_2 = U_{\tilde{X}_2} A_{\tilde{X}_2}^+ U_{\tilde{X}_2}^t - V_{\tilde{X}_2} A_{\tilde{X}_2}^- V_{\tilde{X}_2}^t \quad (32)$$

$$\tilde{T}_1 = U_{\tilde{T}_1} A_{\tilde{T}_1}^+ U_{\tilde{T}_1}^t - V_{\tilde{T}_1} A_{\tilde{T}_1}^- V_{\tilde{T}_1}^t \quad (33)$$

$$\tilde{T}_2 = U_{\tilde{T}_2} A_{\tilde{T}_2}^+ U_{\tilde{T}_2}^t - V_{\tilde{T}_2} A_{\tilde{T}_2}^- V_{\tilde{T}_2}^t \quad (34)$$

where all  $U_{\#}$  and  $V_{\#}$  are matrices with orthogonal columns, and the  $A_{\#}^+$  and  $A_{\#}^-$  are positive diagonal matrices with descending entries along the diagonal. Now  $a$ ,  $b$ ,  $c$ , and  $d$  are estimated such that the matrix power corresponding to the negative eigenvalues of the four matrices, i.e., the power of the  $A_{\#}^-$  matrices, is minimum. Then, the  $A_{\#}^-$  matrices are set to zero, such that the final rewritten sum, denoted by  $\tilde{T}_1 \otimes \tilde{X}_1 + \tilde{T}_2 \otimes \tilde{X}_2$ , only contains positive (semi-)definite matrices:

$$\tilde{X}_1 = U_{\tilde{X}_1} A_{\tilde{X}_1}^+ U_{\tilde{X}_1}^t \quad (35)$$

$$\tilde{X}_2 = U_{\tilde{X}_2} A_{\tilde{X}_2}^+ U_{\tilde{X}_2}^t \quad (36)$$

$$\tilde{T}_1 = U_{\tilde{T}_1} A_{\tilde{T}_1}^+ U_{\tilde{T}_1}^t \quad (37)$$

$$\tilde{T}_2 = U_{\tilde{T}_2} A_{\tilde{T}_2}^+ U_{\tilde{T}_2}^t \quad (38)$$

Summarizing, the cost function used to find  $a$ ,  $b$ ,  $c$ , and  $d$  is

$$C(a, b, c, d) = \frac{\|\Sigma_s - \tilde{T}_1 \otimes \tilde{X}_1 - \tilde{T}_2 \otimes \tilde{X}_2\|^2}{\|\Sigma_s\|^2} \quad (39)$$

and the relative matrix power of the sample covariance matrix explained by the rewritten sum,  $\tilde{T}_1 \otimes \tilde{X}_1 + \tilde{T}_2 \otimes \tilde{X}_2$ , is defined as

$$\begin{aligned} \text{rel pow rewr sum} \\ = \left( 1 - \frac{\|\Sigma_s - \tilde{T}_1 \otimes \tilde{X}_1 - \tilde{T}_2 \otimes \tilde{X}_2\|^2}{\|\Sigma_s\|^2} \right) \\ \times 100\%. \end{aligned} \quad (40)$$

Compared to the contribution of the orthogonally estimated terms in Eq. (21), the relative contribution of the two rewritten terms,  $\tilde{T}_1 \otimes \tilde{X}_1$  and  $\tilde{T}_2 \otimes \tilde{X}_2$ , is less well-defined because the terms are not orthogonal anymore. To compute the contribution of the two terms, the veced KP terms are considered as elements in  $\mathbb{R}^{I^2 J^2}$ . In Fig. 1, this embedding is illustrated; the first term is represented by vector  $v_1$  and the second by  $v_2$ . The sum is drawn as vector  $v$ . The relative contribution of  $v_1$  ( $v_2$ ) to  $v$  is the length  $w_1 = v_1 \cdot \frac{v}{\|v\|}$  ( $w_2 = v_2 \cdot \frac{v}{\|v\|}$ ) divided by the length of  $v$ :

$$\frac{w_1}{\|v\|} = \frac{v_1 \cdot v}{\|v\|^2} \quad (41)$$

$$\frac{w_2}{\|v\|} = \frac{v_2 \cdot v}{\|v\|^2} \quad (42)$$

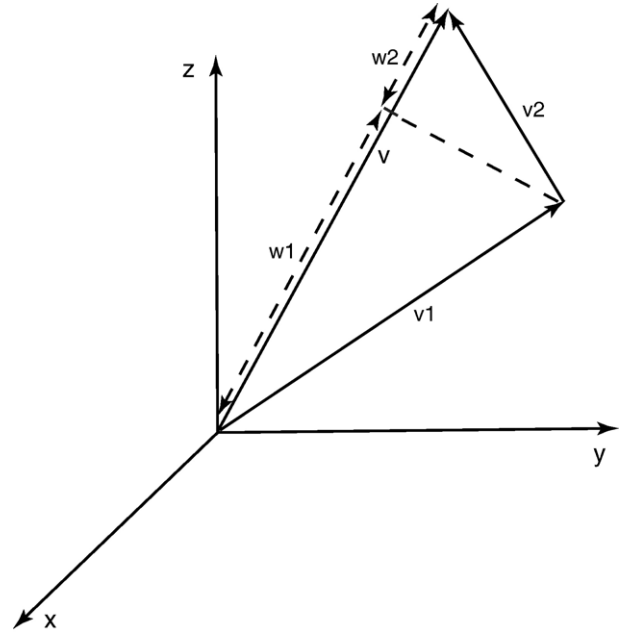


Fig. 1. Illustration of the computation of the contribution of the rewritten terms. Vector  $v_1$  ( $v_2$ ) presents the first (second) rewritten term and the vector  $v$  corresponds to the rewritten sum. The contribution of  $v_1$  ( $v_2$ ) to the length of  $v$  is the length  $w_1$  ( $w_2$ ).

Here the  $a \cdot b$  denotes the inner product of vectors  $a$  and  $b$ . Note that the sum of the contributions of the vectors equals the contribution of the sum of the vectors, that is 1. This principle is applied to  $v_n = \text{vec}(\tilde{T}_n \otimes \tilde{X}_n)$ , for  $n = 1, 2$ . The relative explained power of the sum, Eq. (40), is split into two parts, proportional to the relative contributions of the two terms, yielding

$$\text{rel pownthrewr term} : = \frac{\frac{\|\tilde{T}_n\|^2 \|\tilde{X}_n\|^2 + \text{tr}(\tilde{T}_1 \tilde{T}_2) \text{tr}(\tilde{X}_1 \tilde{X}_2)}{\|\tilde{T}_1 \otimes \tilde{X}_1 + \tilde{T}_2 \otimes \tilde{X}_2\|^2}}{\frac{\|\Sigma_s\|^2 - \|\Sigma_s - \tilde{T}_1 \otimes \tilde{X}_1 - \tilde{T}_2 \otimes \tilde{X}_2\|^2}{\|\Sigma_s\|^2}} \quad (43)$$

for  $n = 1, 2$ . The computation of this power distribution requires the computation of  $\|\Sigma_s\|^2$ . It appeared that, compared to the estimation of the orthogonal terms in Eqs. (22) and (23), the calculation of  $\|\Sigma_s\|^2$  requires considerably more computation time.

## Results

The sum of KP model was applied to evoked response MEG data sets of three different kinds: Somatosensory Evoked Field (SEF) (4 subjects), Visual Evoked Field (VEF) (3 subjects), and Auditory Evoked Field (AEF) data (3 subjects). First, for each data set considered, the average signal over trials was subtracted to obtain the MEG residuals and an offset correction over one alpha period was applied. The alpha period was obtained from the frequency spectrum of the raw data. The offset correction over one alpha period is optimal to reduce the introduction of nonstationarities in the temporal covariance due to alpha background activity, as explained in Bijma et al. (2003) (single KP model) and Appendix A.4 (sum KP model). After the first KP was estimated using Eqs. (10) and (11), the second KP was found from the iterative system in Eqs. (22) and (23). Then, the



relative matrix power explained by the first two KP terms was calculated according to Eq. (21). To find the optimal values for  $a$ ,  $b$ ,  $c$ , and  $d$ , a global search was performed for a spatiotemporally downsampled data set. This reduced data set was obtained by downsampling both in space and in time to approximately 30 time samples and 30 sensors. The so-obtained optimal values for  $a$ ,  $b$ ,  $c$ , and  $d$  were used to rewrite the original data set. Finally, the relative matrix power explained by the rewritten terms was computed according to Eq. (43).

Table 1 presents the resulting power distributions of both the orthogonally estimated and the rewritten sum of two KP as well as the values for  $a$ ,  $b$ ,  $c$ , and  $d$  for all data sets in this study. This table shows that the first KP describes between 62% and 91% power of the sample covariance matrix, and the second term between 1% and 12%. Rewriting the first two terms into two positive semi-definite terms only reduces the total amount of explained matrix power by a negligible amount (less than 1.4%). As can be seen from Table 1, the power distribution over the rewritten terms varies over subjects and data types. These values are directly related to the varying values for  $a$ ,  $b$ ,  $c$ , and  $d$ . In the global search over  $a$ ,  $b$ ,  $c$ , and  $d$  for the downsampled data sets, it appeared that the cost function contains several local minima that are very close to the global minimum in cost value. Apparently, rewriting the sum is not very sensitive to (small) changes in  $a$ ,  $b$ ,  $c$ , and  $d$ .

Table 2 contains the positivity percentages of the temporal and spatial matrices for all data sets. By positivity percentages, the relative matrix power explained by the positive eigenvalues of matrix is meant. The first orthogonally estimated term is always positive definite, and therefore the entries of the first two columns, X01 and T01, all equal 100%. The third and fourth columns contain the positivity percentages of the matrices in the second KP of the orthogonally estimated sum. Clearly, these matrices are far from positive definite, and can be even mainly negative (e.g., subject 5 VEF). After rewriting, all four matrices (XR1, TR1, XR2, and TR2) are well-nigh or completely positive definite. Subject 5 VEF is the only case showing positivity percentages below 95%. This table reveals the effect of rewriting in terms of interpretability: before rewriting, the second term does not possess interpretability as a covariance matrix, while after rewriting, both terms can be interpreted as covariance matrices. Namely, the rewritten matrices are slightly singular – their small

negative parts were set to zero – and this singularity can be interpreted as a light linear dependency among the signals.

For one data set of each kind (SEF, VEF, AEF), the results are illustrated here. In Figs. 2–4 the temporal matrices of the three data sets are shown. The visualization of the temporal covariance matrices is through plotting the entries of the matrix in color. In order to plot a temporal covariance matrix, the entries are scaled such that the entry that is largest in absolute value equals 1 or  $-1$ . The color scale used for these covariance plots is presented in Fig. 5. As with usual printing of matrices, all entries are arranged in a square, and instead of values, corresponding colors are plotted. Nonstationarities and oscillations in the temporal domain can now easily be detected: a stationary temporal covariance matrix has a constant value (color) along its (sub-)diagonals and oscillations in the covariance are reflected by a line pattern parallel to the diagonal. Figs. 2 and 4 show in panels (A) and (B) (the orthogonal terms) a clear oscillation in the covariance. The frequency of this oscillation is approximately 10 Hz; hence, this oscillation shows the alpha activity in the background noise. As expected, these matrices are not purely oscillatory, that is, more noise features are present besides alpha activity. In rewriting the terms, the different noise characteristics of the orthogonally estimated terms are rearranged such that the cost function in Eq. (39) is minimum. It appears that after rewriting, the alpha activity is contained mainly in the second term, while most of the remaining activity is gathered in the first rewritten term. Note that the cost function in Eq. (39) is not frequency specific. For the VEF data set, presented in Fig. 3, alpha oscillations in the orthogonal terms are much smaller and are mainly visible in the second term, panel (B). Nonetheless, after rewriting, no alpha activity is visible in the first term, whereas the second term mainly consists of alpha activity.

Regarding the nonstationarities, Figs. 2–4 show predominantly stationary temporal matrices. However, some nonstationary patterns can be detected; that is, the color along the (sub-)diagonal varies somewhat. The main nonstationarity that can be seen from the color plots is the increase along the diagonal and the sub-diagonals. This is the common consequence of the offset correction over the pre-stimulus interval, which artificially pulls the (co)variance over that time window towards zero (Bijma et al., 2003). Further nonstationarities generally occur in the second orthogonal component (panel (B)), and in one or both rewritten terms. These nonstationarities may be caused by beta activity for which the

Table 1  
The power distribution in the orthogonally estimated sum and in the rewritten sum of 2 KP for all data sets in the study

S	TW	BPF	O1	O2	0	$a$	$b$	$c$	$d$	ADD	AL	RS	DIF
1SEF	574	None	84.79	3.10	87.89	1.30	-0.20	1.43	0.55	59.46	27.47	86.93	0.96
2SEF	574	None	71.82	11.98	83.80	1.00	-0.70	0.79	0.45	38.63	44.79	83.42	0.38
3SEF	574	None	77.66	5.80	83.46	0.90	-0.55	1.00	0.50	37.18	45.05	82.23	1.23
4SEF	574	None	86.63	1.16	87.79	1.30	-0.15	1.03	0.65	72.88	14.76	87.64	0.15
5VEF	480	0–50	86.36	2.71	89.07	0.85	-0.20	0.96	0.95	68.13	19.59	87.72	1.34
6VEF	480	0–50	90.77	1.58	92.35	1.15	-0.10	0.80	0.80	83.41	8.92	92.33	0.02
7VEF	480	0–50	90.67	2.25	92.92	0.95	-0.10	0.98	0.95	81.96	10.99	92.90	0.02
8AEF	480	None	62.12	5.16	67.28	0.60	-0.20	2.15	0.95	37.51	29.71	67.22	0.06
9AEF	480	None	77.10	10.35	87.45	1.10	-0.05	0.20	0.90	76.00	11.25	87.26	0.19
10AEF	480	None	78.25	7.02	85.28	1.00	-0.15	0.67	0.90	70.87	14.36	85.23	0.05

The first column (S) denotes the subject and the kind of data (SEF/VEF/AEF), TW indicates the length of the time window analyzed in milliseconds. Used band pass filtering is stated in the BPF column. O1 (O2) denotes the relative matrix power of the sample covariance matrix in the first (second) orthogonally estimated term and 0 is the sum of O1 and O2. The columns ADD and AL denote the relative matrix power explained by the rewritten terms, ADD indicates the widespread background activity, and AL the alpha activity. RS is the sum of ADD and AL. DIF is the difference between 0 and RS. The given values for  $a$ ,  $b$ ,  $c$ , and  $d$  are the values used for rewriting the sum.

Table 2

The positivity percentages of the matrices in both terms in the orthogonally estimated sum and in the rewritten sum

S	X01	T01	X02	T02	XR1	TR1	XR2	TR2
1SEF	100.00	100.00	51.97	74.50	100.00	98.38	100.00	99.75
2SEF	100.00	100.00	52.31	71.40	99.77	99.07	100.00	100.00
3SEF	100.00	100.00	64.49	84.89	98.43	97.12	100.00	100.00
4SEF	100.00	100.00	39.91	66.61	100.00	100.00	97.70	99.77
5VEF	100.00	100.00	38.29	20.98	100.00	100.00	94.36	91.38
6VEF	100.00	100.00	67.93	87.02	100.00	100.00	99.68	99.80
7VEF	100.00	100.00	78.28	91.02	100.00	99.97	99.98	100.00
8AEF	100.00	100.00	54.61	63.48	100.00	99.82	99.99	99.99
9AEF	100.00	100.00	91.51	96.39	100.00	100.00	99.00	99.14
10AEF	100.00	100.00	70.68	86.82	100.00	100.00	99.67	99.85

The positivity percentage of a matrix equals the relative matrix power that is accounted for by the positive eigenvalues of that matrix. The first column indicates the subject. Columns X01 (X02) and T01 (T02) show the percentages of the spatial and temporal matrices in the first (second) term of the orthogonally estimated sum of 2 KP. All entries in the X01 and T01 column equal 100% because the first orthogonal term is positive definite. The positivity percentages after rewriting are given in the columns XR1 (XR2) and TR1 (TR2) for the spatial and temporal matrices in both rewritten terms.

offset correction window is not optimal. The relative matrix power corresponding to these nonstationarities is very small in comparison to that of the alpha activity. Namely, this feature only occurs in

the second orthogonal term, and the power in this second orthogonal term is much smaller than the power of the first. Moreover, the second orthogonal term is still mainly stationary.

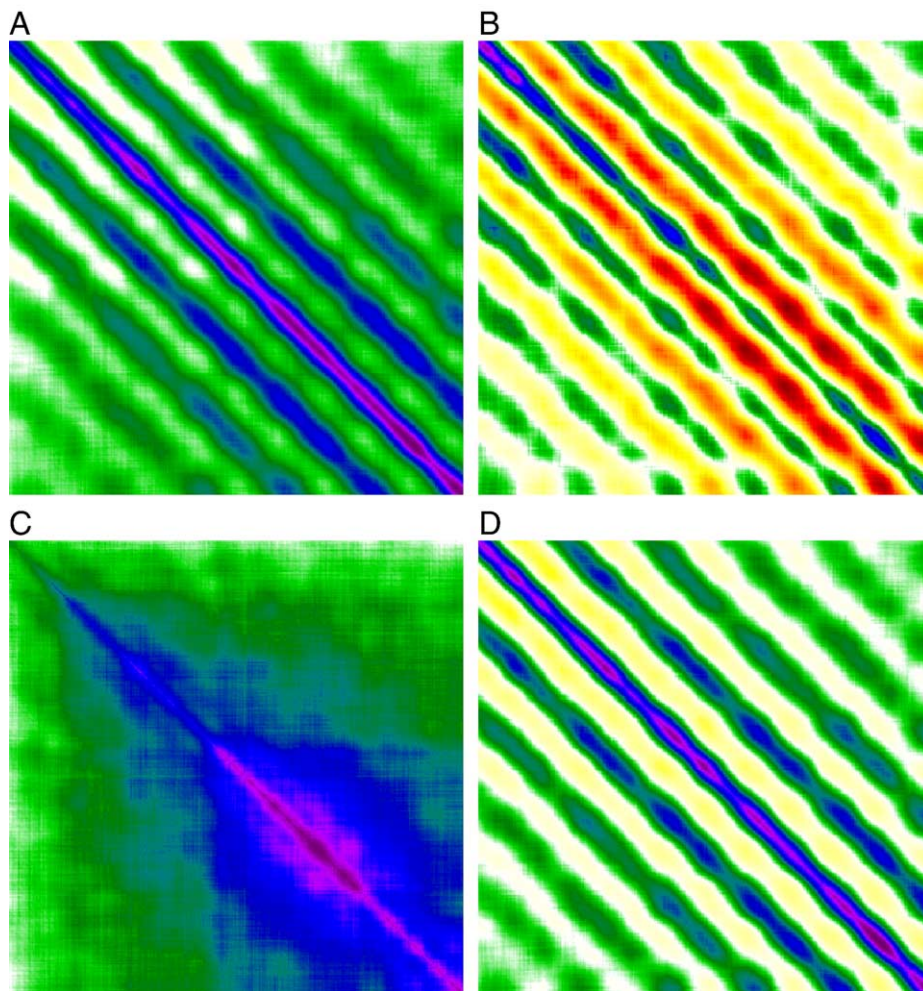


Fig. 2. Estimated temporal matrices in the 2SEF data set. Frames A and B show the first two temporal matrices of the orthogonally estimated terms, frames C and D show the first two temporal matrices of the rewritten sum of two matrices. The time scale is 574 ms by 574 ms. The entries of the matrices are plotted in color. The percentages show the relative matrix power of the sample covariance matrix explained by the KP term. (A) First term in orthogonal sum (71.8%). (B) Second term in orthogonal sum (12.0%). (C) First term in rewritten sum (38.6%). (D) Second term in rewritten sum (44.8%).

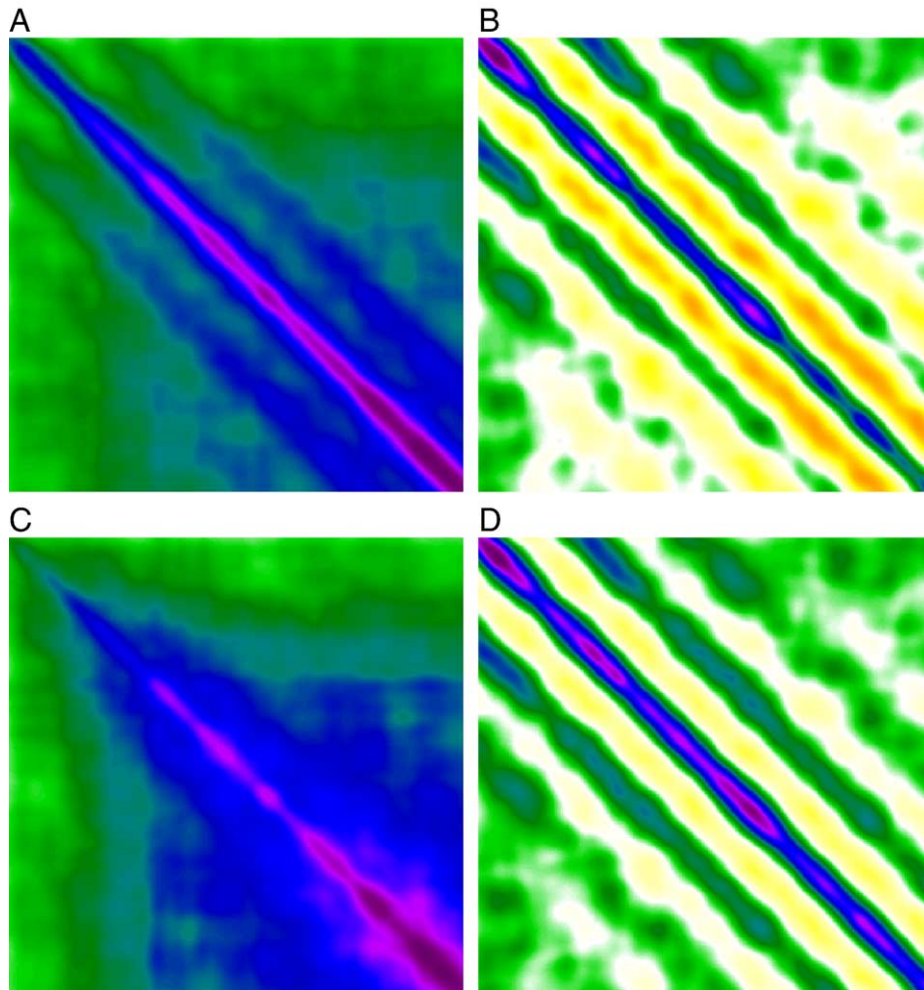


Fig. 3. Estimated temporal matrices in the 7VEF data set. Frames A and B show the first two temporal matrices of the orthogonally estimated terms, frames C and D show the first two temporal matrices of the rewritten sum of two matrices. The time scale is 478 ms by 478 ms. The entries of the matrices are plotted in color. The percentages show the relative matrix power of the sample covariance matrix explained by the KP term. (A) First term in orthogonal sum (90.7%). (B) Second term in orthogonal sum (2.2%). (C) First term in rewritten sum (81.9%). (D) Second term in rewritten sum (11.0%).

The rewritten spatial variances of these three data sets are presented in Figs. 6–8. The spatial covariance matrices are visualized by projecting the variance (the diagonal of the matrix) in color scale on the MEG helmet. For all data sets, the second term, corresponding to the alpha pattern in the temporal matrix, shows a focal highlighted area in the parieto-occipital area. The spatial distribution of the first term is more widespread, though tends to be more in the temporal region.

In sum, minimizing the cost function in Eq. (39) yields two rewritten KP terms, each of which describes a distinct process in the background activity. The first rewritten term describes a rather widespread, not frequency-specific process, while the second term describes the focal alpha activity with its characteristic 10 Hz frequency.

## Discussion

The sum of Kronecker products provides a general model for the spatiotemporal covariance matrix of MEG residuals. Different terms in the sum can describe different, independent phenomena in

the ongoing background activity, each of which has its own temporal and spatial characteristics. These processes can be interpreted as generated by randomly distributed dipoles with a certain spatial and temporal distribution. This way, the sum model solves the rigidity drawback of the single KP model. Theoretically, when enough terms are taken into account, the sum describes the sample covariance matrix exactly.

The first aim of this study is the validation of the single KP model for dipole localization in terms of accuracy. In practice, it occurred that the first KP term describes roughly between 62% and 91% of the sample covariance matrix and the second between 1% and 12%, whereas the sum of 2 KP explains between 67% and 93%. The higher the order of the term, the smaller the amount of explained power. Therefore, taking into account more than 1 KP term in the localization is not expected to yield a major improvement. Namely, the common practice to neglect all the correlations, i.e., both in space and in time, yields an acceptable accuracy at high signal-to-noise ratio. This accuracy is enhanced by taking into account the spatial correlations only, and a further improvement is achieved when the spatiotemporal covariance, the first KP term, is taken into account (de Munck et al., 2002).



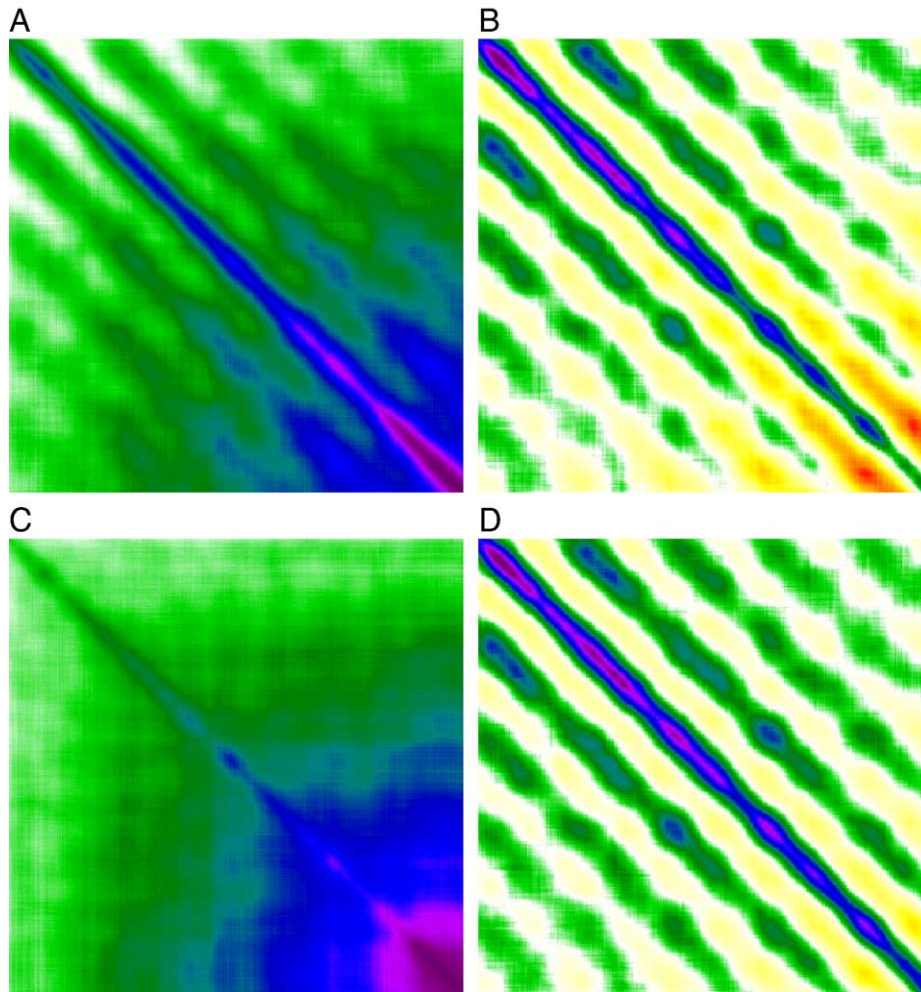


Fig. 4. Estimated temporal matrices in the 10AEF data set. Frames A and B show the first two temporal matrices of the orthogonally estimated terms, frames C and D show the first two temporal matrices of the rewritten sum of two matrices. The time scale is 318 ms by 318 ms. The entries of the matrices are plotted in color. The percentages show the relative matrix power of the sample covariance matrix explained by the KP term. (A) First term in orthogonal sum (78.3%). (B) Second term in orthogonal sum (7.0%). (C) First term in rewritten sum (70.9%). (D) Second term in rewritten sum (14.4%).

Considering the matrix power explained already by the first term, the second and higher-order terms are not expected to enhance the localization accuracy considerably (see Table 1). To improve the source localization further, the emphasis should be on improving the source model, which is likely to be more beneficial. This indicates that the existing covariance model for source localization, the single KP, is adequate.

Nevertheless, the estimated sum of KP contains interesting physiological information, which is the second goal of the present study. There are two aspects regarding this aim: the separation between alpha activity and the remainder, and nonstationarities. For all subjects, the two terms of the rewritten sum show one term

corresponding to alpha activity and the other to additional noise. It is emphasized that this separation comes forth by minimizing the cost function in Eq. (39) and is not caused by an a priori, frequency-specific constraint. The alpha term is characterized by frequency-specific (10 Hz) temporal features and a focal parieto-occipital pattern in space. The additional term shows more widespread characteristics, both in space and in time: there is no frequency-specific character in time and the spatial distribution is widespread, though seems to be enhanced in the temporal region. The power distribution over the rewritten terms in Table 1 suggests that the VEF data sets contain less alpha background activity than the SEF and AEF data. This can be interpreted in line with the discussion in the literature about whether the visual stimulus resets the phase of the spontaneous alpha rhythm (e.g., Brandt, 1997; Klimesch et al., 2004; Kolev et al., 1997; Makeig et al., 2002). If this is true, subtracting the average includes subtracting a major amount of alpha activity and less alpha activity will remain in the MEG residuals. Nevertheless, for all subjects, including the VEF subjects, separation between the alpha activity and the additional activity is striking, although the entireness of the separation varies slightly over subjects. This separation can be interpreted in line

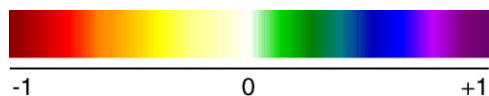


Fig. 5. The color scale used in Figs. 2–4. White indicates zero, purple is positive, red is negative.

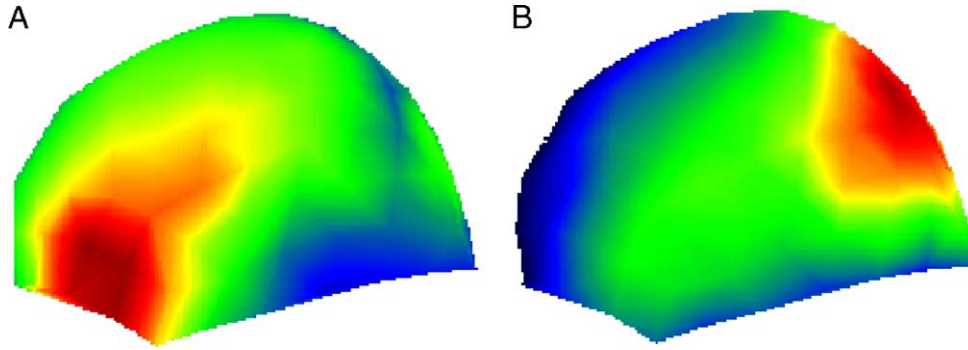


Fig. 6. Rewritten spatial variances in the 2SEF data set. The diagonal entries of the spatial covariance matrix, the variances, are plotted in color on the MEG helmet. Only the left side of the helmet is shown, the right side is similar. (A) First term in rewritten sum. (B) Second term in rewritten sum.

with the Poisson Modulated Alpha Model (Bijma et al., 2003), which models the background activity in the temporal domain as a sum of alpha activity and additional random activity.

In fact, the underlying noise model of the sum model consisting of  $N$  Kronecker products, is a model that describes the noise as the sum of  $N$  independent random dipole processes (Grasman, 2004):

$$e_k^{i,j} = \sum_{n=1}^N \sum_{p=1}^{P_n} \Psi(\zeta^{k,n})_{ip} S_{pj}^{k,n}. \quad (44)$$

Each of these processes consists of  $P_n$  randomly distributed dipoles with source parameters  $\zeta^{k,n} = (\zeta_1^{k,n}, \dots, \zeta_{P_n}^{k,n})$  and forward fields  $\Psi(\zeta^{k,n})$ . The corresponding source time function of each dipole  $p$ ,  $S_{pj}^{k,n}$ , is assumed to be independent of the location and orientation in  $\zeta_p^{k,n}$ , and of the source time functions at other locations  $p'$ . This assumption leads to a Kronecker product structure in the spatiotemporal covariance of the  $n$ th process,  $T_n \otimes X_n$ , as shown in de Munck et al. (1992). The composite of  $N$  such independent processes then has a sum of  $N$  Kronecker products as covariance structure, because the sum of independent Gaussian variables again has the normal distribution. Each of these processes presents an ongoing process in the brain, for example, alpha activity, pathological theta activity, or beta waves. The entire sum in Eq. (44) may be interpreted as the so-called resting state network (Greicius et al., 2003; Raichle et al., 2001). The present study, separating only two processes in the background noise, detects an alpha process and an additional process which is likely

to be a collection of many small processes. Ideally, one would wish to estimate all parameters of this noise model, e.g., the numbers  $N$  and  $P_n$ , the distribution of the source parameters  $\zeta^{k,n}$  and the  $N$  Kronecker products. In practice, though, this estimation is hampered by the need for a nonconventional amount of computer power and the large number of (nonlinear) parameters. Therefore, simplifying assumptions need to be made in the practical estimation, e.g., fitting the sum of two KP to the sample covariance matrix as in the present study.

The second aspect of physiological information is about the temporal nonstationarities. As stated in the Introduction, the temporal covariance matrices will contain nonstationarities when the SPN model is assumed incorrectly. However, in the presented results, the temporal matrices are mainly stationary. The most apparent nonstationarity is the increase along the (sub-)diagonals, caused by the offset correction. Furthermore, along the diagonal, one can see small oscillations, reflected in the colored bands having varying width. After rewriting the sum, these small oscillations are mainly visible in the alpha term (Figs. 2–4). A possible explanation for these oscillations again lies in the offset correction: for oscillatory background activity with the beta frequency, the offset correction window of one alpha period is not optimal and will introduce a small oscillation along the diagonal and sub-diagonals (Bijma et al., 2003). Further nonstationarities that indicate real trial-to-trial variabilities are very limited in the color plots. Despite the minor role of these real nonstationarities in terms of matrix power, the present study does not disapprove their existence.

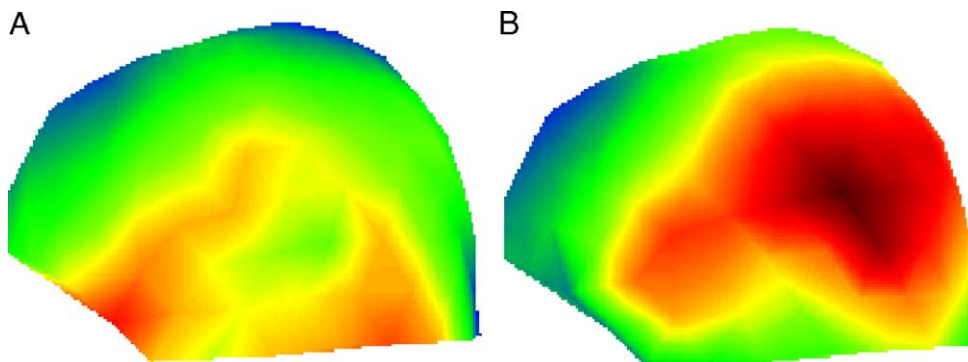


Fig. 7. Rewritten spatial variances in the 7VEF data set. The diagonal entries of spatial covariance matrix, the variances, are plotted in color on the MEG helmet. Only left side of the helmet is shown, the right side is similar. (A) First term in rewritten sum. (B) Second term in rewritten sum.

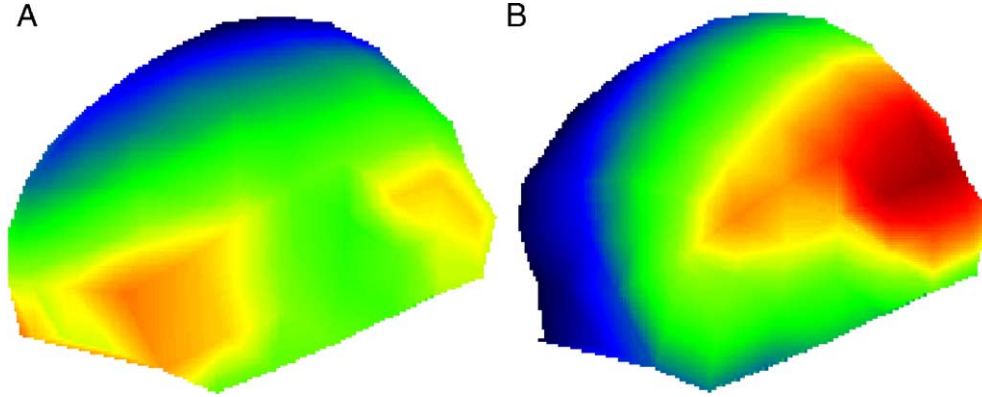


Fig. 8. Rewritten spatial variances in the 10AEF data set. The diagonal entries of the spatial covariance matrix, the variances, are plotted in color on the MEG helmet. Only the right side of the helmet is shown, the left side is similar. (A) First term in rewritten sum. (B) Second term in rewritten sum. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In sum, it appears from the distribution of explained matrix power that it is unlikely that dipole localization will improve considerably by adding more terms to the single KP noise model. Secondly, rewriting the sum of two KP suggests that the noise consists of a focal parieto-occipital alpha part and a more widespread noise part. And finally, nonstationarities due to real trial-to-trial variations in this study appeared to be very limited in terms of matrix power. In a forthcoming study, the possibilities of modeling the generators of these two noise parts will be investigated and an explanation for the small nonstationarities will be sought.

## Appendix A

### A.1. Estimating a single Kronecker product

In this appendix, expressions for the LS estimators of  $X$  and  $T$  in the single KP model are derived. The LS cost function can be written as:

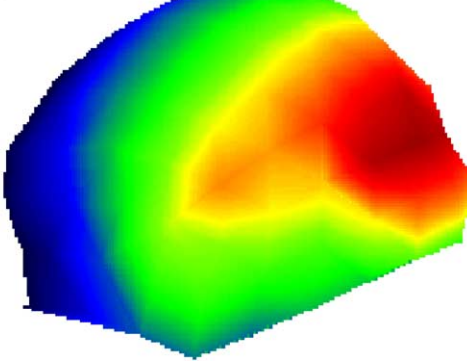
$$\begin{aligned} \|\Sigma_s - T \otimes X\|^2 &= \text{tr}(\Sigma_s \Sigma_s') + \text{tr}(T^2) \text{tr}(X^2) \\ &\quad - 2 \frac{1}{K-1} \sum_{k=1}^K \text{tr}(E_k T E_k' X). \end{aligned} \quad (45)$$

The optimal  $X$  and  $T$  are found by differentiating Eq. (45) and subsequently equating the first derivative to zero. Differentiation with respect to matrices is performed according to the rules derived in Magnus and Neudecker (1995). For  $X$ , this yields

$$\begin{aligned} d_X \|\Sigma_s - T \otimes X\|^2 &= 2 \text{tr}(T^2) \text{tr}(X dX) - 2 \frac{1}{K-1} \sum_{k=1}^K \text{tr}(E_k T E_k' dX) \\ &= 2 \text{tr} \left( \left[ \text{tr}(T)^2 X - \frac{1}{K-1} \sum_{k=1}^K E_k T E_k' \right] dX \right) = 0 \\ \Rightarrow \text{tr}(T)^2 X - \frac{1}{K-1} \sum_{k=1}^K E_k T E_k' &= 0 \end{aligned} \quad (46)$$

Rewriting this equation yields Eq. (10) as LS estimator for  $X$  and a similar derivation yields Eq. (11) for  $T$ .

B



### A.2. Van Loan's shuffle operator

The shuffle operator  $\mathcal{S}$  is defined by

$$\text{vec} \circ \mathcal{S} = (I_J \otimes K_{J,I} \otimes I_I) \circ \text{vec}, \quad (47)$$

where the composition operator  $\circ$  is defined as  $(L_1 \circ L_2) A = L_1(L_2(A))$  and  $K_{p,q} : \mathbb{R}^1 \times p^q \rightarrow \mathbb{R}^1 \times qp$  is the general commutation matrix (Magnus and Neudecker, 1995):

$$K_{p,q}(\text{vec}(A)) = \text{vec}(A^t) \quad (48)$$

for any matrix  $A \in \mathbb{R}^{p \times q}$ . To see Eq. (47), the following equality from (Magnus and Neudecker, 1995) is used

$$(I_n \otimes K_{m,q} \otimes I_p) \text{vec}(A \otimes B) = \text{vec}(A) \otimes \text{vec}(B). \quad (49)$$

for any  $A \in \mathbb{R}^{m \times n}$  and any  $B \in \mathbb{R}^{p \times q}$ . Applying this equality with  $A = T$  and  $B = X$  and using

$$\text{vec}(\text{vec}(X_n) \text{vec}(T_n)^t) = \text{vec}(T_n) \otimes \text{vec}(X_n) \quad (50)$$

one arrives at Eq. (47). Applying Eq. (47) to Eq. (6) and successively applying Eqs. (50) and (49) yields the formula for the shuffled sample covariance,  $\mathcal{S}(\Sigma_s)$ , in Eq. (16).

### A.3. The LS estimators for the higher-order terms in the sum of KP

In this appendix, the LS estimators for the higher-order terms in the sum of KP model are derived. The  $p$ th order term is estimated after the terms  $1, \dots, p-1$  have been estimated. Estimation of the  $p$ th term is by differentiation of the cost function in Eq. (13) and applying Lagrange multiplication (Marsden and Tromba, 1988). Abbreviate

$$\mathbf{x}_n = \text{vec}(X_n) \quad (51)$$

$$\mathbf{t}_n = \text{vec}(T_n). \quad (52)$$

Suppose the first  $p-1$  terms have been estimated, where the normalization is chosen such that  $\|\mathbf{x}_n\|^2 = 1$  for  $n = 1, \dots, p-1$ .

The cost function for the  $p$ th term is then

$$\begin{aligned} \left\| \left[ \mathcal{S}(\Sigma_s) - \sum_{n=1}^{p-1} \mathbf{x}_n \mathbf{t}_n^t \right] - \mathbf{x}_p \mathbf{t}_p^t \right\|^2 &= \|\mathcal{S}_{p-1} - \mathbf{x}_p \mathbf{t}_p^t\|^2 \\ &= \text{tr} \left( \mathcal{S}_{p-1}^t \mathcal{S}_{p-1} - 2\mathcal{S}_{p-1}^t \mathbf{x}_p \mathbf{t}_p^t + \mathbf{t}_p \mathbf{x}_p^t \mathbf{x}_p \mathbf{t}_p^t \right), \end{aligned} \quad (53)$$

where

$$\mathcal{S}_{p-1} = \mathcal{S}(\Sigma_s) - \sum_{n=1}^{p-1} \mathbf{x}_n \mathbf{t}_n^t. \quad (54)$$

In other words, the  $p$ th term is the best rank-one approximation to  $\mathcal{S}_{p-1}$ . Considering Eqs. (19) and (20), this minimization is subject to

$$\mathbf{x}_n^t \mathbf{x}_p = 0 \text{ for } 1 \leq n \leq p-1 \quad (55)$$

$$\mathbf{t}_n^t \mathbf{t}_p = 0 \text{ for } 1 \leq n \leq p-1 \quad (56)$$

The derivative of Eq. (53) with respect to  $\mathbf{x}_p$  is given by

$$\begin{aligned} &\text{tr} \left( -2\mathcal{S}_{p-1}^t d\mathbf{x}_p \mathbf{t}_p^t + \mathbf{t}_p d\mathbf{x}_p^t \mathbf{x}_p \mathbf{t}_p^t + \mathbf{t}_p \mathbf{x}_p^t d\mathbf{x}_p \mathbf{t}_p^t \right) \\ &= 2\text{tr} \left( \left( -\mathbf{t}_p^t \mathcal{S}_{p-1}^t + \mathbf{t}_p^t \mathbf{x}_p \mathbf{x}_p^t \right) d\mathbf{x}_p \right) \end{aligned} \quad (57)$$

and, similarly, with respect to  $\mathbf{t}_p$

$$2\text{tr} \left( \left( -\mathbf{x}_p^t \mathcal{S}_{p-1} + \mathbf{x}_p^t \mathbf{x}_p \mathbf{t}_p^t \right) d\mathbf{t}_p \right). \quad (58)$$

The derivatives of the constraints in Eqs. (55) and (56) are

$$\mathbf{x}_n^t d\mathbf{x}_p = 0 \text{ for } 1 \leq n \leq p-1 \quad (59)$$

$$\mathbf{t}_n^t d\mathbf{t}_p = 0 \text{ for } 1 \leq n \leq p-1 \quad (60)$$

The method of Lagrange multiplication now yields the following system of equations

$$\mathbf{x}_n^t \mathbf{x}_p = 0 \in \mathbb{R} \text{ for } 1 \leq n \leq p-1 \quad (61)$$

$$\mathbf{t}_n^t \mathbf{t}_p = 0 \in \mathbb{R} \text{ for } 1 \leq n \leq p-1 \quad (62)$$

$$-\mathcal{S}_{p-1} \mathbf{t}_p + \mathbf{t}_p^t \mathbf{x}_p \mathbf{x}_p^t \mathbf{t}_p + \sum_{n=1}^{p-1} \lambda_n \mathbf{x}_n = 0 \in \mathbb{R}^{I^2 \times 1} \quad (63)$$

$$-\mathcal{S}_{p-1}^t \mathbf{x}_p + \mathbf{x}_p^t \mathbf{x}_p \mathbf{t}_p + \sum_{n=1}^{p-1} \mu_n \mathbf{t}_n = 0 \in \mathbb{R}^{I^2 \times 1} \quad (64)$$

which has to be solved for  $\mathbf{x}_p$ ,  $\mathbf{t}_p$ ,  $\lambda_1, \dots, \lambda_{p-1}$  and  $\mu_1, \dots, \mu_{p-1}$ . The solution for  $\lambda_m$ ,  $m = 1, \dots, p-1$ , follows from the inner product of  $\mathbf{x}_m$  and the vectors in Eq. (63):

$$\begin{aligned} &-\mathbf{x}_m^t \mathcal{S}_{p-1} \mathbf{t}_p + \mathbf{x}_m^t \mathbf{t}_p^t \mathbf{x}_p \mathbf{x}_p^t \mathbf{t}_p + \sum_{n=1}^{p-1} \lambda_n \mathbf{x}_m^t \mathbf{x}_n = 0 \\ &\Rightarrow \lambda_m = \frac{1}{\mathbf{x}_m^t \mathbf{x}_m} \mathbf{x}_m^t \mathcal{S}_{p-1} \mathbf{t}_p = \mathbf{x}_m^t \mathcal{S}_{p-1} \mathbf{t}_p. \end{aligned} \quad (65)$$

Likewise, solving the inner product of  $\mathbf{t}_m$  and the vectors in Eq. (64) for  $\mu_m$  yields

$$\mu_m = \frac{\mathbf{t}_m^t \mathcal{S}_{p-1}^t \mathbf{x}_p}{\mathbf{t}_m^t \mathbf{t}_m}. \quad (66)$$

Substituting Eq. (65) into Eq. (63) and using Eqs. (54) and (62) yields

$$\begin{aligned} 0 &= -\mathcal{S}_{p-1} \mathbf{t}_p + \mathbf{t}_p^t \mathbf{x}_p \mathbf{x}_p^t \mathbf{t}_p + \sum_{n=1}^{p-1} \mathbf{x}_n^t \mathcal{S}_{p-1} \mathbf{t}_p \mathbf{x}_n \\ &= -\mathcal{S}_{p-1} \mathbf{t}_p + \mathbf{t}_p^t \mathbf{x}_p \mathbf{x}_p^t \mathbf{t}_p + \left( \sum_{n=1}^{p-1} \mathbf{x}_n \mathbf{x}_n^t \right) \mathcal{S}_{p-1} \mathbf{t}_p \\ &\Rightarrow = \frac{1}{\mathbf{t}_p^t \mathbf{t}_p} \left[ I_{I^2} - \sum_{n=1}^{p-1} \mathbf{x}_n \mathbf{x}_n^t \right] \mathcal{S}_{p-1} \mathbf{t}_p \\ &= \frac{1}{\mathbf{t}_p^t \mathbf{t}_p} \left[ I_{I^2} - \sum_{n=1}^{p-1} \mathbf{x}_n \mathbf{x}_n^t \right] \mathcal{S}(\Sigma_s) \mathbf{t}_p \end{aligned} \quad (67)$$

where the last simplification follows from the constraints in Eq. (56). Similarly, Eq. (66) substituted into Eq. (64) together with Eqs. (54), (55), and (61) yield

$$\begin{aligned} 0 &= -\mathcal{S}_{p-1}^t \mathbf{x}_p + \mathbf{x}_p^t \mathbf{x}_p \mathbf{t}_p + \sum_{n=1}^{p-1} \frac{\mathbf{t}_n^t \mathcal{S}_{p-1}^t \mathbf{x}_p}{\mathbf{t}_n^t \mathbf{t}_n} \mathbf{t}_n \\ &= \mathcal{S}_{p-1}^t \mathbf{x}_p + \mathbf{x}_p^t \mathbf{x}_p \mathbf{t}_p + \left( \sum_{n=1}^{p-1} \frac{\mathbf{t}_n \mathbf{t}_n^t}{\mathbf{t}_n^t \mathbf{t}_n} \right) \mathcal{S}_{p-1}^t \mathbf{x}_p \\ &\Rightarrow \mathbf{t}_p = \frac{1}{\mathbf{x}_p^t \mathbf{x}_p} \left[ I_{I^2} - \sum_{n=1}^{p-1} \frac{\mathbf{t}_n \mathbf{t}_n^t}{\mathbf{t}_n^t \mathbf{t}_n} \right] \mathcal{S}_{p-1}^t \mathbf{x}_p \\ &= \frac{1}{\mathbf{x}_p^t \mathbf{x}_p} \left[ I_{I^2} - \sum_{n=1}^{p-1} \frac{\mathbf{t}_n \mathbf{t}_n^t}{\mathbf{t}_n^t \mathbf{t}_n} \right] (\mathcal{S}(\Sigma_s))^t \mathbf{x}_p. \end{aligned} \quad (68)$$

Eqs. (67) and (68) are solved iteratively and the normalization is chosen such that  $\|\mathbf{x}_p\|^2 = 1$ .

Note that a closed form expression for the estimators can be obtained by substituting Eq. (68) in Eq. (67):

$$\mathbf{x}_p = \frac{1}{\mathbf{t}_p^t \mathbf{t}_p} \left[ I_{I^2} - \sum_{n=1}^{p-1} \mathbf{x}_n \mathbf{x}_n^t \right] \mathcal{S}(\Sigma_s) \left[ I_{I^2} - \sum_{n=1}^{p-1} \frac{\mathbf{t}_n \mathbf{t}_n^t}{\mathbf{t}_n^t \mathbf{t}_n} \right] (\mathcal{S}(\Sigma_s))^t \mathbf{x}_p, \quad (69)$$

showing that  $\mathbf{x}_p$  is an eigenvector of the  $I^2 \times I^2$  matrix

$$\left[ I_{I^2} - \sum_{n=1}^{p-1} \mathbf{x}_n \mathbf{x}_n^t \right] \mathcal{S}(\Sigma_s) \left[ I_{I^2} - \sum_{n=1}^{p-1} \frac{\mathbf{t}_n \mathbf{t}_n^t}{\mathbf{t}_n^t \mathbf{t}_n} \right] (\mathcal{S}(\Sigma_s))^t, \quad (70)$$

corresponding to the largest eigenvalue in order to minimize Eq. (53). Computing eigenvalues and eigenvectors of this  $I^2 \times I^2$  matrix is, like computing the SVD of  $\mathcal{S}(\Sigma_s)$ , in practice not convenient.



Finally, expressions for the iterative estimators for the matrices  $X_p$  and  $T_p$  are derived from Eqs. (16), (67), and (68).

$$\begin{aligned}
\text{vec}(X_p) &= \frac{1}{(K-1)\|T_p\|_F^2} \sum_{k=1}^K \left[ I_{p^2} - \sum_{n=1}^{p-1} \text{vec}(X_n)(\text{vec}(X_n))^t \right] \\
&\quad \times (E_k \otimes E_k) \text{vec}(T_p) \\
&= \frac{1}{(K-1)\|T_p\|_F^2} \sum_{k=1}^K \left[ I_{p^2} - \sum_{n=1}^{p-1} \text{vec}(X_n)(\text{vec}(X_n))^t \right] \\
&\quad \times \text{vec}[E_k T_p E_k^t] \\
&= \frac{1}{(K-1)\|T_p\|_F^2} \sum_{k=1}^K \left[ \text{vec}[E_k T_p E_k^t] \right. \\
&\quad \left. - \sum_{n=1}^{p-1} \text{vec}(X_n)(\text{vec}(X_n))^t \text{vec}[E_k T_p E_k^t] \right] \\
&= \frac{1}{(K-1)\|T_p\|_F^2} \sum_{k=1}^K \left[ \text{vec}[E_k T_p E_k^t] \right. \\
&\quad \left. - \sum_{n=1}^{p-1} \text{tr}[X_n E_k T_p E_k^t] \text{vec}(X_n) \right] \quad (71)
\end{aligned}$$

This yields Eq. (22) and a similar calculation yields Eq. (23).

#### A.4. Offset correction in case of multiple KP

The expression for the offset corrected residual is

$$\check{e}_k^{i,j} = e_k^{i,j} - \frac{1}{L} \sum_{l=1}^L e_k^{i,j_l} \quad (72)$$

where  $(j_1, \dots, j_L)$  denotes the time window over which the correction is performed. In Bijma et al. (2003), an expression is derived for the temporal covariance matrix of the corrected residuals, denoted here by  $T^c$ , for the single KP model ( $\Sigma = T \otimes X$ ) in terms of the covariance matrix of the uncorrected residuals,  $T$ :

$$\begin{aligned}
T_{j,j'}^c &= T_{j,j'} - \frac{1}{L} \sum_{l=1}^L T_{j_l,j'} - \frac{1}{L} \sum_{m=1}^L T_{j,j_m} \\
&\quad + \frac{1}{L^2} \sum_{l=1}^L \sum_{m=1}^L T_{j_l,j_m}. \quad (73)
\end{aligned}$$

For the sum of KP model,  $\Sigma = \sum T_n \otimes X_n$ , the covariance between two uncorrected residuals,  $e_k^{i,j}$  and  $e_{k'}^{i',j'}$ , is

$$\varepsilon(e_k^{i,j}, e_{k'}^{i',j'}) = \delta_{k,k'} \sum_{n=1}^N (X_n)_{i,i'} (T_n)_{j,j'}. \quad (74)$$

Applying the offset correction of Eq. (72) to the residuals in Eq. (74), one can derive the covariance matrix of the corrected residuals in the case of the sum of KP. Straightforward calculating

shows that each temporal matrix in this sum is altered according to Eq. (73).

## References

- Bijma, F., de Munck, J.C., Huizenga, H.M., Heethaar, R.M., 2003. A mathematical approach to the temporal stationarity of background noise in MEG/EEG measurements. *NeuroImage* 20 (1), 233–243.
- Brandt, M.E., 1997. Visual and auditory evoked phase resting of the alpha EEG. *Int. J. Psychophysiol.* 26, 285–298.
- Coppola, R., Tabor, R., Buchsbaum, M.S., 1978. Signal to noise ratio and response variability measurements in single trial evoked potentials. *Electroencephalogr. Clin. Neurophysiol.* 44, 214–222.
- de Munck, J.C., Vijn, P.C.M., Lopes da Silva, F.H., 1992. A random dipole model for spontaneous brain activity. *IEEE Trans. Biomed. Eng.* 39 (8), 791–804.
- de Munck, J.C., Huizenga, H.M., Waldorp, L.J., Heethaar, R.M., 2002. Estimating stationary dipoles from MEG/EEG data contaminated with spatially and temporally correlated background noise. *IEEE Trans. Signal Process.* 50 (7), 1565–1572.
- de Munck, J.C., Bijma, F., Gaura, P., Sieluzycycki, C., Branco, M.I., Heethaar, R.M., 2004. A maximum likelihood estimator for trial to trial variations in noisy MEG/EEG data sets. *IEEE Trans. Biomed. Eng.* 51 (12), 2123–2128.
- Duann, J.R., Jung, T.P., Kuo, W.J., Yeh, T.C., Makeig, S., Hsieh, J.C., Sejnowski, T.J., 2002. Single-trial variability in event-related BOLD signals. *NeuroImage* 15 (4), 823–835.
- Dogandžić, A., Nehorai, A., 2000. Estimated evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays. *IEEE Trans. Signal Process.* 48 (1), 13–25.
- Gasser, T., Mocks, J., Verleger, R., 1983. SELAVCO: a method to deal with trial-to-trial variability of evoked potentials. *Electroencephalogr. Clin. Neurophysiol.* 55, 717–723.
- Golub, H., Van Loan, C.F., 1990. *Matrix Computations*, 2nd edition John Hopkins Univ. Press, Baltimore.
- Grasman, R.P.P.P., 2004. Sensor array signal processing and the neuro-electromagnetic inverse problem in functional connectivity analysis of the brain Ph.D. thesis, University of Amsterdam.
- Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci.* 100 (1), 253–258.
- Huizenga, H.M., de Munck, J.C., Waldorp, L.J., Grasman, R.P.P.P., 2002. Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model. *IEEE Trans. Biomed. Eng.* 49 (6), 533–539.
- Jaśkowski, P., Verleger, R., 1999. Amplitude and latencies of single-trial ERP's estimated by a maximum-likelihood method. *IEEE Trans. Biomed. Eng.* 46, 987–993.
- Klimesch, W., Schack, B., Schabus, M., Doppelmayr, M., Gruber, W., Sauseng, P., 2004. Phase-locked alpha and theta oscillations generate the P1 and N1 complex and are related to memory performance. *Cogn. Brain Res.* 19 (3), 302–316.
- Kolev, V., Yordanova, J., Shtirman, M., Bařar, E., 1997. Event-related alpha oscillations in task processing. *Clin. Neurophysiol.* 110, 1784–1792.
- Langville, A.N., Stewart, W.J., 2004. The Kronecker product and stochastic automata networks. *J. Comput. Appl. Math.* 167, 429–447.
- Laskaris, N.A., Liu, L.C., Ionnides, A.A., 2003. Single-trial variability in early visual neuromagnetic responses: an explorative study based on the regional activation contributing to the N70m peak. *NeuroImage* 20 (2), 765–783.
- Laufs, H., Krakow, K., Sterzer, P., Eger, E., Beyerle, A., Salek-Haddadi, A., Kleinschmidt, A., 2003. Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proc. Natl. Acad. Sci.* 100 (19), 11053–11058.

- Lütkenhöner, B., 1998. Dipole source localization by means of maximum likelihood estimation: I. Theory and simulations. *Electroencephalogr. Clin. Neurophysiol.* 106, 314–321.
- Magnus, J.R., Neudecker, H., 1995. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised edition. John Wiley and Sons, Chichester.
- Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J., 2002. Dynamic brain sources of visual evoked responses. *Science* 295, 690–694.
- Makinen, V., Tiitinen, H., May, P., 2005. Auditory event-related responses are generated independently of ongoing brain activity. *NeuroImage* 24 (4), 961–968.
- Marsden, J.E., Tromba, A.J., 1988. *Vector Calculus*. Freeman and Company, New York.
- Mocks, J., Gasser, T., Puan, P.D., Köhler, W., 1987. Trial-to-trial variability of single potentials: methodological concepts and results. *Int. J. Neurosci.* 33, 25–32.
- Pham, D.T., Mocks, J., Köhler, W., Gasser, T., 1987. Variable latencies of noisy signals: estimation and testing in brain potential data. *Biometrika* 74, 525–533.
- Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. *Proc. Natl. Acad. Sci.* 98 (2), 676–682.
- Sekihara, K., Takeuchi, F., Kuriki, S., Koizumi, H., 1994. Reduction of brain noise influence in evoked neuromagnetic source localization using noise spatial correlation. *Phys. Med. Biol.* 39, 937–946.
- Truccolo, W.A., Ding, M., Knuth, K.H., Nakamura, R., Bressler, L., 2002. Trial-to-trial variability of cortical evoked responses: implications for the analysis of functional connectivity. *Clin. Neurophysiol.* 113, 206–226.
- Van Loan, C.F., 2000. The ubiquitous Kronecker product. *J. Comput. Appl. Math.* 123, 85–100.
- Waldorp, L.J., Huizenga, H.M., Dolan, C.V., Molenaar, P.C.M., 2002. Estimated generalized least squares electromagnetic source analysis based on a parametric noise covariance model. *IEEE Trans. Biomed. Eng.* 48 (6), 737–741.