Modeling spatiotemporal covariance for magnetoencephalography or electroencephalography source analysis

Sergey M. Plis,* J. S. George, S. C. Jun, J. Paré-Blagoev, D. M. Ranken, C. C. Wood, and D. M. Schmidt
Applied Modern Physics Group, Los Alamos National Laboratory, MS-D454, Los Alamos, New Mexico 87545, USA
(Received 6 July 2006; revised manuscript received 1 November 2006; published 30 January 2007)

We propose a new model to approximate spatiotemporal noise covariance for use in neural electromagnetic source analysis, which better captures temporal variability in background activity. As with other existing formalisms, our model employs a Kronecker product of matrices representing temporal and spatial covariance. In our model, spatial components are allowed to have differing temporal covariances. Variability is represented as a series of Kronecker products of spatial component covariances and corresponding temporal covariances. Unlike previous attempts to model covariance through a sum of Kronecker products, our model is designed to have a computationally manageable inverse. Despite increased descriptive power, inversion of the model is fast, making it useful in source analysis. We have explored two versions of the model. One is estimated based on the assumption that spatial components of background noise have uncorrelated time courses. Another version, which gives closer approximation, is based on the assumption that time courses are statistically independent. The accuracy of the structural approximation is compared to an existing model, based on a single Kronecker product, using both Frobenius norm of the difference between spatiotemporal sample covariance and a model, and scatter plots. Performance of ours and previous models is compared in source analysis of a large number of single dipole problems with simulated time courses and with background from authentic magnetoencephalography data.

DOI: 10.1103/PhysRevE.75.011928 PACS number(s): 87.57.Ra, 87.80.Tq, 02.30.Zz

I. INTRODUCTION

The physical and physiological consequences of the correlated activity of substantial populations of neurons can be detected with noninvasive measurement techniques, including electroencephalography (EEG) and magnetoencephalography (MEG). These macroscopic electrophysiological techniques can resolve the time course of neural population activation with millisecond temporal resolution. Neural electromagnetic (NEM) responses are governed by the same physical processes that give rise to electric and magnetic fields in other systems: vector currents established by potential differences along cellular structures give rise to an electric field aligned with the current and an orthogonal magnetic field that encircles the current element. Because many different sensors typically detect signal contributions from a given source, data sets often contain identifiable patterns of spatial covariance associated with sources of interest as well as background processes. Because neural activation typically proceeds with a characteristic time course, spatial covariance components often exhibit structured temporal covariance and correlation.

The unique strengths of neural electromagnetic methods stem from their capacity to define the dynamics of neural population activity. Even a single electrode pasted to the scalp may disclose a complex temporal wave-form consisting of a series of peaks and valleys. The first 50 years of work with EEG involved little quantitative effort to localize the sources of observed topographies in the surface potential data. Inspection or simple quantification of temporal wave-form features served as the basis of diagnostic procedures in clinical neurology as well as experimental studies of cognitive processing. The development of MEG and the recognition that many observed field distributions could be explained by a simple forward model lead to advances in procedures that have subsequently been applied to EEG data. Basic and clinical neuroscience are very motivated to identify the anatomical sources of observed functional activity, as evident in the explosion of interest in functional magnetic resonance imaging (fMRI). Suitable geometric models of neural sources, coupled with physical "forward models" describing the relationships between sources, detectors, and the tissue medium, and adequate optimization strategies, enable useful localization of neural electromagnetic sources, in spite of the ill-posed, ambiguous nature of the inverse problem. Even if the objective of analysis is to describe the dynamics of neural activation, this is most effective in the context of an adequate model of the underlying neural sources.

Ongoing spontaneous activity recorded at the surface of the human head using MEG or EEG, typically is characterized by regions of relatively large amplitude oscillatory patterns that vary as a function of position on the head and state of the subject. The signals associated with responses to individual stimuli or other punctuate cognitive or control processes are typically much smaller and require specialized experimental paradigms and signal processing techniques to pull the signals out of the noise. In order to enhance the consistent aspects of the neural response while suppressing the contribution of other physiological processes or environmental noise, most investigators employ sensory-evoked response or event-related paradigms, averaging temporal sequences time locked to the stimulus or a behavioral response.

The central limit theorem lends support to the common assumption that the averaged background data are Gaussian distributed, even though the distribution of a single trial

*E-mail address: pliz@lanl.gov

 ©2007 The American Physical Society

background may not be Gaussian. The log likelihood function is a common mathematical expression quantifying the likelihood that a given model (e.g., of neural current) could have produced the measured data. For Gaussian, zero-mean-averaged background noise, the log likelihood function is given by

$$-\frac{1}{2}\sum_{ktk't'}\left[b_{kt}-\int \mathcal{L}_k(x)j(x,t)dx\right]\mathcal{C}^{-1}_{kt;k't'} \times\left[b_{k't'}-\int \mathcal{L}_{k'}(x')j(x',t')dx'\right]. \quad (1)$$

Here, b_{kt} are the averaged measurements (the data being analyzed) at sensor k and time t; $j(x,t)$ is the neural (source) current distribution over space x and time t; $\mathcal{L}_k(x)$ is the forward or lead field for sensor k, i.e., the linear operator that connects source currents to predicted measurements in the absence of noise. $\mathcal{C}$ is the covariance of the averaged background activity, which describes second-order statistical properties of the MEG/EEG data in the absence of sources. This activity mainly consists of signals created by brain function, sensor noise, and external noise that was not adequately shielded or canceled. Typically, such cumulative activity is not of interest in inverse analysis, e.g., ERP analysis, and is considered noise. In this paper, we interchangeably use terms noise and background to refer to the ongoing activity of no interest.

Several classes of techniques have been used for neural electromagnetic source localization [1–6], but most use the likelihood formulation in some way. Source localization methods such as these can be applied to account for a field or potential distribution observed at an instant in time. However, spatiotemporal techniques that attempt to account for the full data set using a limited number of sources, each with an associated time course, have proven to be much more powerful. For example, such methods allow the resolution of sources with overlapping field or potential distributions and overlapping activation time courses, which are ambiguous in a single spatial map. These methods operate by implicitly considering the structure in the spatiotemporal covariance of the signals. As a consequence, covariance that exists in the data set but is not explicitly modeled by component sources may lead to mislocalization or biased estimates of source time courses [7]. Despite ample evidence that the background is correlated over space and time, the covariance is commonly taken to be diagonal [8–10]. In order to address this problem, we have undertaken the development of methods to characterize the data covariance due to correlated noise or background physiological activity.

The sample covariance of the averaged background data is related to the sample covariance of the single-trial background data by a simple expression: $\mathcal{C}=\frac{1}{M}\hat{\mathcal{C}}$. Here, M is the number of trials being averaged and $\hat{\mathcal{C}}$ is the sample covariance of the single-trial background data. This relation assumes the trials are independent draws from the single-trial distribution but does not assume a particular form for this distribution. The task of estimating the average background covariance may be accomplished by estimating the single-trial background covariance and scaling it by the number of trials in the average.

Given that MEG/EEG is measured in M trials, on L sensors and in C time samples, let $\mathbf{E}_m$ be the $L\times C$ single-trial noise matrix at trial m. In this case, the conventional way to estimate the full covariance matrix of dimension $N=LC$ for the averaged noise is by

$$\mathcal{C}=\frac{1}{M(M-1)}\sum_{m=1}^{M}[\mathrm{vec}(\mathbf{E}_m)-\mathrm{vec}(\bar{\mathbf{E}})][\mathrm{vec}(\mathbf{E}_m)-\mathrm{vec}(\bar{\mathbf{E}})]^T, \quad (2)$$

$$\bar{\mathbf{E}}=\frac{1}{M}\sum_{m=1}^{M}\mathbf{E}_m, \quad (3)$$

where vec($\mathbf{E}$) is all the columns of $\mathbf{E}$ stacked in a vector. In order to simplify notation in this paper, we use M both for the number of available noise samples and for the number of stimuli over which averaging is done. Note that, in the general case, these numbers may not be the same.

There are a number of reasons why this estimate of the full covariance is difficult to use and why an approximation is needed. First, even for modern multisensor detectors, sufficient experimental data are rarely available to adequately estimate the large number of parameters present in the full covariance matrix. For example, for 35 time samples and 121 channels and considering the fact that the covariance matrix is symmetric, 8 969 730 parameters should be determined $((121\times35)[(121\times35)+1]/2)$. This far exceeds the amount of data available. Second, because the spatiotemporal noise covariance matrix is so large, a tremendous amount of memory is required for its storage. Third, this full covariance is almost impossible to handle in the likelihood formulation, since the computation time of calculating the inverse still renders the task infeasible in most interesting cases, even if it was possible to estimate the covariance with the given data. A naive algorithm for matrix inversion takes $O(N^3)$ time, where N is the dimensionality of the matrix. Though there are some improvements over this result for large matrices [11,12], the problem is overwhelming for typical computing equipment and for interesting values of N. To summarize, it is almost always impossible to estimate the full spatiotemporal covariance due to lack of data; in those cases when the estimation is possible, the inversion is computationally hard, and in any case, the amount of storage required is high. Because of all these difficulties with the estimation of the full spatiotemporal covariance, an accurate approximation is needed. In addition to addressing the three problems mentioned above, a good approximation should capture as much of the structure in the noise as possible and reduce the errors in inverse solutions.

In this paper, we describe two spatiotemporal covariance models tailored for use in localization algorithms (Sec. III). The first model is a multipair Kronecker product (KP) approximation based on an orthogonal spatial basis, and the second one is an extended, more general variant based on an independent spatial basis.

Two common models are described (Sec. II). The first is the widely used diagonal approximation, which has no spatial or temporal correlation and whose diagonal elements consist of sensor noise variances. The second model is a KP approximation of a temporal covariance and a spatial covariance under the assumption that spatiotemporal covariance is separable [13]. Motivation for the need of a more complex model is given.

In order to compare how well our models approximate the full spatiotemporal sample covariance, we used criteria based on the Frobenius norm and scatter plots. These criteria applied to the single Kronecker product model and our models help to order models by how structurally close they are to the modeled matrix.

Since the goal of better noise characterization by covariance modeling is to improve results of inverse algorithms, we have chosen to test and compare our models to some of the existing ones by using them in algorithms for source localization. Performance of the approximations in reconstructing dipole locations and time courses is tested using a large number of simulated single-dipole data sets constructed from empirical data for background noise and simulated dipole sources covering a wide range of locations and orientations (Sec. IV). We note that many of the most serious problems with source localization arise from the presence of multiple nearby sources with overlapping time courses. However, a systematic investigation of such issues is beyond the objective and scope of the work reported here.

II. SOME OF THE COMMON MODELS

It is common when solving the inverse problem to model covariance as a diagonal or even as an identity matrix. In the diagonal case, the approximated full spatiotemporal covariance is expressed as $\mathcal{C} \propto \mathbf{T} \otimes \mathbf{S}$, where $\mathbf{T}$ is the temporal covariance that is taken to be the identity; $\mathbf{S}$ is the diagonal spatial covariance with elements of the diagonal being sensor variances; and $\otimes$ is the Kronecker product. This model is easy to estimate. It has only L parameters, where L is the number of sensors. At the same time, this simple model is better than not using any covariance estimate at all as in the ordinary least-squares (OLS) approach.

In order to improve localization results achievable with the very simple diagonal model, more complicated models were sought. De Munck *et al.* [13] proposed a model based on the assumption that the spatiotemporal covariance of the background in MEG and EEG is separable. This assumption naturally leads to the use of KP in the model formulation:

$$\mathcal{C} \approx \mathbf{T} \otimes \mathbf{S} = \begin{bmatrix} \mathbf{t}_{11}\mathbf{S} & \mathbf{t}_{12}\mathbf{S} & \cdots & \mathbf{t}_{1C}\mathbf{S} \\ \mathbf{t}_{21}\mathbf{S} & \mathbf{t}_{22}\mathbf{S} & \cdots & \mathbf{t}_{2C}\mathbf{S} \\ \vdots & \vdots & & \vdots \\ \mathbf{t}_{C1}\mathbf{S} & \mathbf{t}_{C2}\mathbf{S} & \cdots & \mathbf{t}_{CC}\mathbf{S} \end{bmatrix}. \tag{4}$$

Temporal covariance is a $C \times C$ matrix $\mathbf{T}$ and spatial covariance is an $L \times L$ matrix $\mathbf{S}$, where C is the number of time samples and L is the number of sensors. In [13], the temporal covariance matrix, $\mathbf{T}$ is normalized to 1.

In addition to being more descriptive than the diagonal model, the KP model has an easy inverse. This makes it useful in source localization. As demonstrated in [13] the analysis results improve considerably using the KP model compared to results when using the diagonal model.

However, because it employs a single spatial and temporal covariance pair, the simplest KP model can only describe a very limited range of covariance. In order to reduce this limitation, others have suggested the use of a sum of KPs of temporal and spatial covariances as an approximation to the full spatiotemporal covariance matrix [14]. This allows the description of several background processes each with their own spatial and temporal patterns. Though shown potentially to be useful, this model is not well suited for source analyses since its inversion is computationally demanding.

In this paper, we introduce and compare models that are conceptually between the computationally convenient but restrictive KP model of [13] and the very general sum of KPs model of [14], which is not well suited for source analysis. In Sec. III, we present two models that allow more temporal variability than the KP model does. These models are useful for source analysis because they have a computationally manageable inverse in contrast to the sum of KP model. But they are not as general in their descriptive power.

III. SERIES KP MODELS

In the Kronecker product model described above, all spatial components of the background have the same temporal covariance structure. This statement describes the main assumption of the KP model. In order to motivate multipair models, we rewrite the KP model in the following way. First, perform a spectral decomposition of the spatial covariance:

$$\mathbf{T} \otimes \mathbf{S} = \mathbf{T} \otimes \sum_{l=1}^{L} \sigma^l \boldsymbol{\mathcal{S}}^l, \tag{5}$$

where $\boldsymbol{\mathcal{S}}^l = v_l v_l^T$ is an orthonormal basis component represented as a rank 1 matrix, σ^l is an eigenvalue, and v_l is an eigenvector. This form is one of the conventional ways of writing a spectral representation of a matrix. Using the identity $\mathbf{A} \otimes (\mathbf{B}+\mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}$, Eq. (5) can be represented as

$$\sum_{l=1}^{L} \mathbf{T} \otimes \sigma^l \boldsymbol{\mathcal{S}}^l = \sum_{l=1}^{L} \sigma^l \mathbf{T} \otimes \boldsymbol{\mathcal{S}}^l. \tag{6}$$

The left-hand side of Eq. (6) makes it obvious that each spatial component has the same temporal covariance. The contribution of each temporal covariance is weighted by the variance of the corresponding orthonormal spatial component, as seen from the right-hand side of (6). In the final sum, such weighting of temporal covariances makes no difference because they all are the same.

In a more realistic case, several noise generators having distinct spatial patterns (or belonging to separate spatial components) may also have different and independent temporal structure. Furthermore, some background sources of noise that are external to the head can also be captured by such a model.

A. Orthogonal basis multipair model

We propose an alternative KP model of the spatiotemporal noise covariance that is a *series* of orthonormal spatial components $\boldsymbol{\mathcal{S}}^l$ of the background data and their corresponding temporal covariance matrices $\boldsymbol{\mathcal{T}}^l$expressed as

$$\mathcal{C} \approx \sum_{l=1}^{L} \boldsymbol{\mathcal{T}}^l \otimes \boldsymbol{\mathcal{S}}^l. \tag{7}$$

The model is built on the following assumptions: A1, Spatiotemporal noise is generated by L spatially orthogonal generators that do not change their location during the period of interest; A2, each spatial component defined in signal space has a time course uncorrelated with those of other components; and A3, superposition of time courses of each component measured at the sensors has a Gaussian distribution with zero mean.

Let us assume for now that we are given the orthonormal components $\boldsymbol{\mathcal{S}}^l = v_l v_l^T$. (We will discuss a way to obtain those later in the section.) As in (5), the orthonormal basis spans the whole sensor space having dimension L. The next task is to estimate temporal covariances $\boldsymbol{\mathcal{T}}^l$of each component. This can be done due to the assumption A3 that the background is Gaussian and using the maximal likelihood estimation as demonstrated in Appendix A. Assume that we have M single-trial noise samples $\mathbf{E}_m$– a $L \times C$ matrix with L number of sensors and C number of time points. Now we can summarize orthogonal basis multipair model in the following way:

$$\mathcal{C} \approx \tilde{\mathcal{C}} \equiv \sum_{l=1}^{L} \boldsymbol{\mathcal{T}}^l \otimes \boldsymbol{\mathcal{S}}^l,$$

$$\boldsymbol{\mathcal{T}}^l = \frac{1}{M^2} \sum_{m=1}^{M} \mathbf{E}_m^T \boldsymbol{\mathcal{S}}^l \mathbf{E}_m,$$

$$\boldsymbol{\mathcal{S}}^l = v_l v_l^T. \tag{8}$$

Expression (8) represents a model for the averaged noise; hence, the normalizing factor in the estimate of $\boldsymbol{\mathcal{T}}^l$is squared.

Another important feature expected of a covariance model to make it useful in source analysis is a computationally manageable inverse. The inversion of this multipair model can be expressed by

$$\tilde{\mathcal{C}}^{-1} = \sum_{l=1}^{L} (\boldsymbol{\mathcal{T}}^l)^{-1} \otimes \boldsymbol{\mathcal{S}}^l \tag{9}$$

(see Appendix B for details). This inversion is easily manageable.

Although this model has far fewer free parameters than the full spatiotemporal covariance estimate (2), the number of parameters is larger than that of KP model (though not by a great amount). The increase in parameters provides an increase of expressive power beyond that of the single-pair model (4). Furthermore, it is free from assumptions of identical temporal character of the noise over all spatial locations and, hence, may capture more information about the structure of the noise.

We derived this model without any assumptions about how spatial orthogonal components were obtained. That means that, in principle, any set of such components should keep the validity of the derivation and retain the invertibility of the model. However, since we assume that the background has the distribution close to the Gaussian, then it is natural to use singular value decomposition (SVD) to obtain orthogonal spatial components. The result of SVD [which performs principal component analysis (PCA) in our case] are spatial components whose time courses have a cross correlation of zero. It is used in this work to estimate the spatially orthogonal components of the orthogonal basis multipair model. SVD of the noise data collected in matrix $\mathbf{A}$ by stacking M single-trial spatiotemporal samples $\mathbf{E}_m$ looks like $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$. $\mathbf{U}$ is a $CM \times L$ orthogonal matrix; $\boldsymbol{\Sigma}$ is a $L \times L$ diagonal matrix with singular values of $\mathbf{A}$ λ_l as diagonal elements; and $\mathbf{V}$ is an $L \times L$ orthogonal matrix of spatial components. Each row of $\mathbf{V}^T$ is a spatial component v_l that is used to form orthogonal spatial components $\boldsymbol{\mathcal{S}}^l$ of our model (7).

Very often SVD is used for dimensionality reduction when only the most significant singular values are accounted for and all other values are neglected [15]. In this application, we do not use SVD for this purpose. Dimensionality reduction, i.e., approximation of the full covariance, is already performed in (7) based on the stated assumptions. The sole purpose of SVD in estimation of this model is to find spatially orthogonal components and their corresponding time courses.

B. Independent basis multipair model

The model introduced in Sec. III B has greater expressive power than the models presented previously. Nevertheless, its reliance on PCA lets it find those noise generators whose time courses are uncorrelated by placing an additional orthogonality constraint on spatial components. However, there is no guarantee that background activity components will be orthogonal. A better physical model would be to assume that noise generators produce time courses statistically independent from one another. In addition to better describing different processes generating background (e.g., alpha rhythm, unfiltered sensor noise, environmental perturbations), it removes the orthogonality constraint. Thus, instead of finding spatially orthogonal and temporally uncorrelated components, we derive components that are temporally independent.

Let us assume that we are given independent spatial components of the background. Denote each independent component as w_l, its corresponding spatial matrix as $\boldsymbol{\mathcal{R}}^l = w_l w_l^T$, and the matrix in which the lth row is the w_l vector as $\mathbf{W}^{-1}$. In this framework, the only assumption needed is independence of each w_l from the rest of the components. Estimation of temporal covariances of each independent spatial component can be performed in the Maximum likelihood framework (see Appendix A),

$$\mathcal{C} \approx \tilde{\mathcal{C}} \equiv \sum_{l=1}^{L} \boldsymbol{\mathcal{T}}^l \otimes \boldsymbol{\mathcal{R}}^l$$

$$\mathcal{T}^l = \frac{1}{M^2}\sum_{m=1}^{M} \mathbf{E}_m^T \mathbf{W}\mathbf{W}^T \mathcal{R}^l \mathbf{W}\mathbf{W}^T \mathbf{E}_m$$

$$\mathcal{R}^l = w_l w_l^T. \tag{10}$$

Here, w_l stands for the lth row vector of $\mathbf{W}^{-1}$ as previously described and M^2 is due to the averaged noise modeling.

The model is more general compared to the one using orthogonal bases from Sec. III A. It can be constructed in terms of any spatially independent basis set, which also includes orthogonal sets. At the same time, its inversion is still manageable and can be expressed as (see Appendix B for details)

$$\tilde{\mathcal{C}}^{-1} = \sum_{l=1}^{L} (\mathcal{T}^l)^{-1} \otimes [\mathbf{W}\mathbf{W}^T \mathcal{R}^l \mathbf{W}\mathbf{W}^T]. \tag{11}$$

This makes the model useful in source analysis.

The above reasoning applies to any statistically independent components of the measured signal. It is crucial to adopt an independent component analysis (ICA) algorithm well suited to the neuroimaging signals. Different ICA algorithms have been applied to MEG [16–20] and EEG [21–23] data. Among them, the most widely used are second-order blind identification (SOBI) [24], Infomax [25], and fICA [26]. The SOBI algorithm was used in [17,18], and a solid motivation for applying it to MEG data is given there. Following this motivation, we chose the SOBI algorithm for this research. In this paper, we have used a SOBI algorithm implementation provided by the ICALAB toolbox [27] with the default settings.

Our goal in this paper is not to evaluate performance of different ICA algorithms in the suggested framework. Rather, here we chose one algorithm to demonstrate the feasibility and applicability of ICA for finding independent spatial components $\mathcal{R}^l$ for modeling background noise. The procedure we used can be described as follows.

Assume that we have M times single-trial noise data and that an ICA algorithm is applied to this data constructed in a matrix $\mathbf{A}$,

$$\begin{pmatrix} \mathbf{E}_1^T \\ \vdots \\ \mathbf{E}_M^T \end{pmatrix} = \mathbf{A} = \begin{pmatrix} \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_M \end{pmatrix} \mathbf{W}^{-1}. \tag{12}$$

In Eq. (12), $\mathbf{W}$ is a $L \times L$ unmixing matrix obtained as a result of an ICA algorithm, the $\mathbf{U}_m$ form a $CM \times L$ matrix of independent components $\mathbf{U}$. This $\mathbf{W}$ matrix is the matrix used in (10) and (11).

Although ICA is widely used for dimensionality reduction, we note that in this application no dimensionality reduction is performed based on ICA; the technique is used exclusively as a discovery tool.

IV. COMPARING PERFORMANCE OF COVARIANCE MODELS

The empirical MEG data and anatomical information used in this work were acquired in the following experiment:

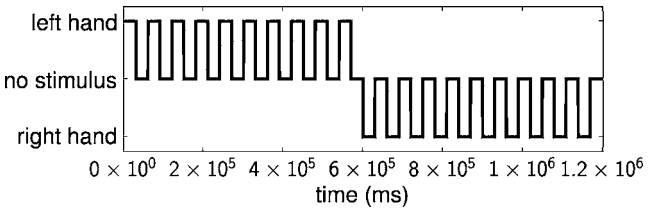

FIG. 1. The manner in which stimuli were applied in the experiment. The x axis shows time samples, and y axis shows stimulus types for the right-hand stimulus, no stimulus, and left-hand stimulus. Within each stimulus block, there is a series of ≈60 stimuli at random ISI.

Electrical stimulation of the median nerve (at the wrist) was applied at the motor twitch threshold, using a block design of 30 s on, 30 s off for a total of ten blocks for each of eight runs. The stimulus alternated across runs, with four runs each of left-side stimulation and of right-side stimulation. The ISI (interstimulus interval) was randomized between 0.25 and 0.75 s (Fig. 1). Since there is no stimulus for long periods, this design provides a large sample of brain noise data, which is useful in other noise studies. Data were collected on a 4D Neuroimaging Neuromag-122 whole-head gradiometer system with 122 channels [28]. The experiment used a male subject, age 38, sampling rate was set to 1000 Hz. In this paper, data from sensor 51 was not used since its output had too many artifacts, leaving 121 usable channels.

This somatosensory data set contained high-frequency transient signals resulting from the electrical stimulus. These transients are distorted when filtered with a linear filter, due to ringing effects and adversely affect nearby data, including the expected early response at about 20 ms poststimulus. To avoid this and still be able to remove low-frequency drifts, a median filter was used as follows. First, the signal was filtered with a median filter of window size set to obtain a 1 Hz low-pass filter (1000 samples, in our case). Second, the result from the previous step was subtracted from initial measurements to obtain a 1 Hz high-pass filtered signal. The large 60 Hz noise and harmonics in the data were reduced but not removed by replacing the points in the power spectrum in the data near 60 Hz and harmonics with values that interpolated between adjacent power spectrum points.

Many different approaches can be used to evaluate how well various models approximate the full spatiotemporal covariance. Typically, some norm of the difference between true sample covariance and its approximating model is calculated, for instance, the Frobenius norm. Kullback-Leibler divergence [29] is also a popular measure. A scatter plot can be a good "visual" tool for comparison as described, for example, in [30].

Measures such as the above-mentioned Frobenius norm give a cumulative quantitative basis for comparing how well a model approximates the sample covariance. In order to compute the Frobenius norm, the sample spatiotemporal covariance needs to be estimated. The experimental setup used to obtain the data for this work provided plenty of background data not available in a typical experiment. This background comes from periods free from stimulus. Abundance

TABLE I. Frobenius norm of the difference between sample spatiaotemporal covariance and three models: single KP model, orthogonal basis multipair model, and independent basis multipair model.

Model	Single KP	orthogonal	independent
Difference (%)	81.5	27.2	17.8

of the background data allowed us to reliably estimate sample spatiotemporal covariance for 121 sensors and 35 time points (35 ms). Using this estimate of the sample covariance $\tilde{\mathcal{C}}$, we were able to calculate a normalized Frobenius norm of the difference between the modeled covariance and each of the models according to expression (13), where $\|\cdot\|$ denotes Frobenius norm.

$$\frac{\|\tilde{\mathcal{C}} - \mathcal{M}\|}{\|\tilde{\mathcal{C}}\|} 100\,\% \,. \tag{13}$$

Results of this comparison for the single KP model, orthogonal basis multipair model, and independent basis multipair model are summarized in Table I. We have also looked at Kullback-Leibler divergence using a smaller subset of sensors and time points in order to provide invertibility of the full sample covariance needed for calculations. The behavior was very similar to the one in Table I.

Using a cumulative measure to compare models is useful, but different cumulative measures can give different and even contradictory results. In order to provide an additional reference point of model comparison, we used scatter plots. Scatter plots contain more information about structural differences and similarities of the models. In the case when matrices are identical, all points should fall on the 45° line between the x- and y-axes. Deviation from this behavior demonstrates differences between models at values of different magnitude. Each of the models compared to the full spatiotemporal sample covariance using Frobenius norm was compared using a scatter plot of a model across the full covariance as shown in Fig. 2.

Scatter plots allow us to compare two matrices A and B element by element (a_{ij} with b_{ij}) and give an idea of dissimilarities in their structure. One can conclude from Fig. 2, that independent basis multipair model of Fig. 2(c) is structurally the closest to the spatiotemporal sample covariance when comparing it to scatter plots in Figs. 2(a) and 2(b). This means that in terms of modeling the spatiotemporal sample covariance it gives the best approximation. The conclusion is also supported by the Frobenius norm results of Table I.

The ultimate goal of modeling full spatiotemporal covariance is to achieve better source analysis, including estimation of source location and time course. Thus, evaluation of model performance in a dipole-based source analysis is employed as an evaluation metric in this study. A large number of single dipole data sets were analyzed using different background noise models. Each data set was constructed using empirical whole head MEG background data together with simulated dipole sources.

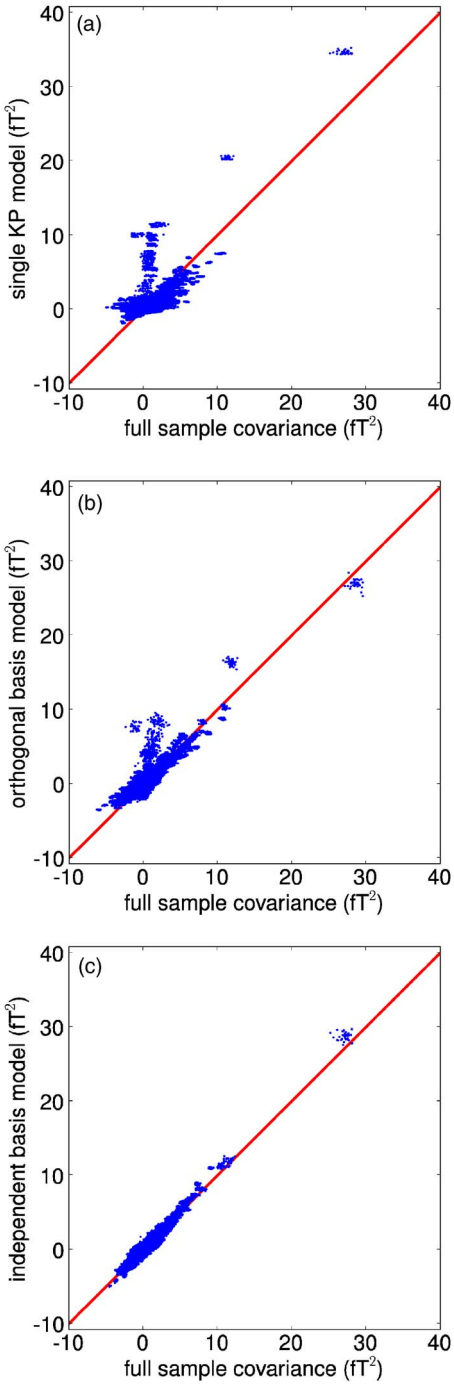

FIG. 2. (Color online) Scatter plot of covariance models across full spatiotemporal sample covariance: (a) single KP model, (b) orthogonal basis multipair model, and (c) independent basis multipair model.

Noise samples of 35 ms latency were extracted from the prestimulus area of right-side stimulation no earlier than 300 ms after the previous stimulus was applied. Since filtering can affect temporal covariance it is important to estimate covariance after filtering has been applied. All covariance models in this section were estimated using this continuous background data, scaled appropriately to obtain covariances of the noise averaged over 602 samples. All the covariance

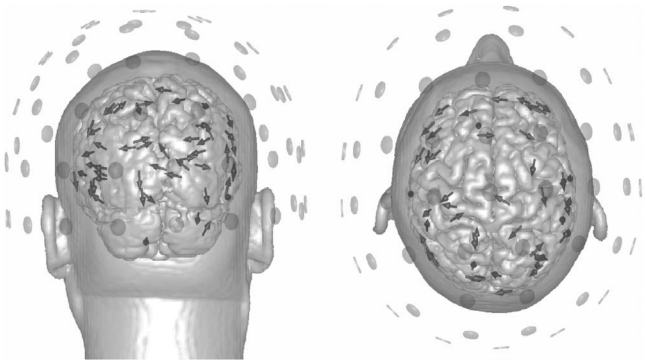

FIG. 3. Dipoles randomly scattered over the cortex. These dipoles give fairly uniform coverage of the cortex and with different orientations create many possible realistic sources.

models yielded similar sensor variances, differences were only observed in the correlation structure.

Continuous noise samples were combined in a way that allowed for averaging over 602 independent samples. This approach supplied six different average noise data sets. Different signal-to-noise ratios (SNR) for single dipole problems were obtained by scaling these noise samples before combining them with simulated data. The measure of SNR used in this paper was constructed by squaring all values of the signal vector and then adding them together (inner product of the signal vector) and dividing this value by squared and summed values of the noise vector; a square root was taken from the obtained ratio.

The locations and orientations of 50 dipoles were drawn at random from the gray-matter voxels that had been tagged from the subject's anatomical MRI used in the empirical MEG experiment. In order to mitigate the effect of depth vs strength that would complicate the interpretation of results, the area from which random locations were drawn was constrained to be further than 5 cm from the head center (Fig. 3).

Each single dipole problem consisting of 121 sensor values over 35 ms (121×35 matrix) was constructed in the following way: (i) For each dipole in the set of 50, a sinusoidal time course was used; (ii) this dipole was projected to the sensor space using the Sarvas spherical head model [31]; and (iii) one of the six noise sample sets was added to the simulated signal. Noise was appropriately scaled in advance according to the intended SNR.

The total number of single dipole problems run through the inverse solution routine was 2400. This number combines 50 dipoles with 6 noise kinds for each and 8 SNR values (0.3, 0.4, 0.5, 0.7, 1.0, 1.5, 2.0, 3.0). Figure 4 demonstrates how different the noise samples were and what diversity was introduced by changing SNR.

For each data set, we estimated the location, orientation, and time course of a single dipole that maximized the likelihood, using the different background covariance models. In these trials, the location and orientation of the dipole did not vary over time. Initial attempts to solve for these parameters using an optimization algorithm were plagued by local minima problems, which is common in dipole inverse algorithms [32,33]. The degree of these local minima errors confounded the errors between background covariance models. To mitigate this confound, we employed a sampling algorithm using Markov chain Monte Carlo (MCMC) [3,34,35], which sampled the location and orientation parameters from the likelihood using the maximum likelihood time-course values for a given set of location and orientation parameters. From these samples, we then calculated the mean values of the parameters and used this as our estimate of the maximum likelihood result. This reduced the local minima problems but did not eliminate them. To further reduce the local minima effects, we ran multiple MCMC samplings for each data set and chose the results that had the highest likelihood. We are confident that the results from this set of procedures primarily reflect the errors associated with the different background covariance models.

There are 300 results for each SNR value for each model. Figure 5 shows a histogram of location errors for the diago-

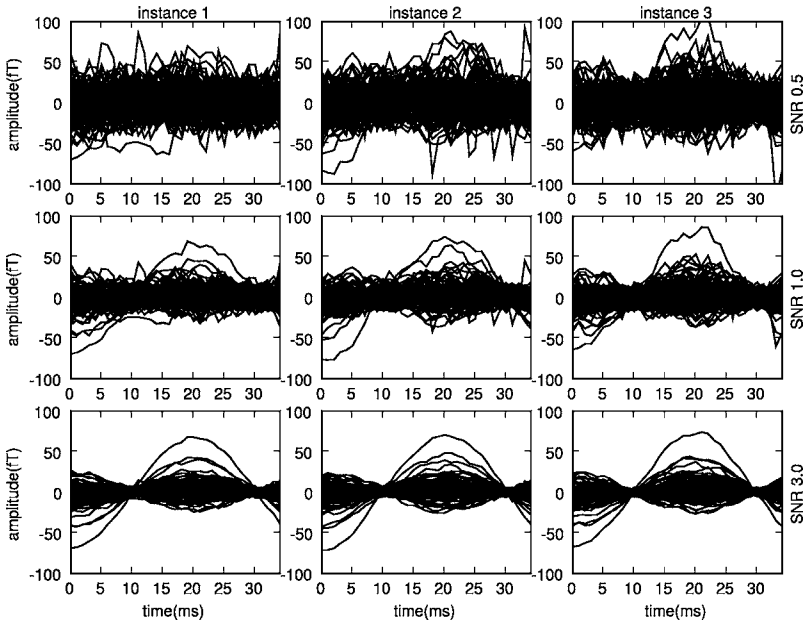

FIG. 4. Three noise sample sets and three signal-to-noise ratios for one of the dipoles. Columns have the same signal-to-noise ratio, and rows have the same noise instance added to the simulated dipole.

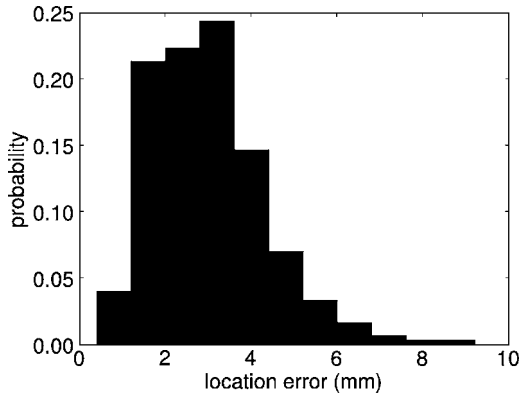

FIG. 5. Location error histogram for the case of using the diagonal covariance approximation at SNR 1.0.

nal model with SNR of 1.0. The shape of this distribution (non-Gaussian with large tails), is typical for all of the models and SNR values. From each of these distributions, we calculated the point on the error axis below which 90% of the probability mass is concentrated. Figures 6(a) and 6(b) show these 90% error plots for location and time courses as a function of SNR for each background covariance model. Here, the location error was calculated as the root-mean-squared error, and the time-course error was calculated as the root of the squared error averaged over time. The two multipair models had the lowest errors, followed by the KP model with slightly higher errors. Finally, the diagonal model had by far the largest errors.

V. DISCUSSION

This work presents a model of spatiotemporal sample covariance, which is a significantly closer approximation than the models suggested previously. By better modeling temporal variability in the background, it demonstrates improvements in source time-course estimation and localization. Improvements in estimating the dynamics of sources is expected to be even better in complex situations with greater variability in the temporal behavior.

A spatiotemporal covariance model based on a sum of KP proposed in [14] is interesting as a noise exploration tool, but it is not well suited for inverse analysis because of its computationally intensive inversion. Our work targeted the development of practical models for analysis purposes that are located between the KP model [13] and the sum of KP model in the hierarchy of descriptive power. A simplification made in our models that made them different from the model of [14] is making spatial covariances rank 1. This made our models less general but, importantly, allowed fast inversion, which made them practical.

NEM analysis methods that are based on averaging assume that the structure of the response of interest is static from trial to trial and that background processes are unaffected by stimulation. Both of these assumptions may be violated in physiological data.

Single-pass analytical methodologies are increasingly applied to the analysis of neural electromagnetic data. Frequency decomposition techniques have been used to explore the putative role of oscillatory activity in certain perceptual processes. Spatial filtering techniques (based on a computed or assumed source model) can be applied to the data to estimate the activation time course of the source. Techniques such as minimum variance beam-forming, synthetic aperture magnetometry, magnetic field tomography, and some forms of ICA are most effectively applied to analysis of single trial data. However, these methods suffer from the fact that noise and background activity may greatly exceed the signals of interest.

Our studies of source localization did not disclose large advantages of ICA relative to the orthogonal basis model. The problem is that both PCA and ICA analysis did a good job for this instantiation of noise. Bases on the norm metric and scatter plots, even a perfect sample covariance model would make no significant improvement in source localization in this case. Previous experience shows that noise causes the greatest problems for source localization algorithms

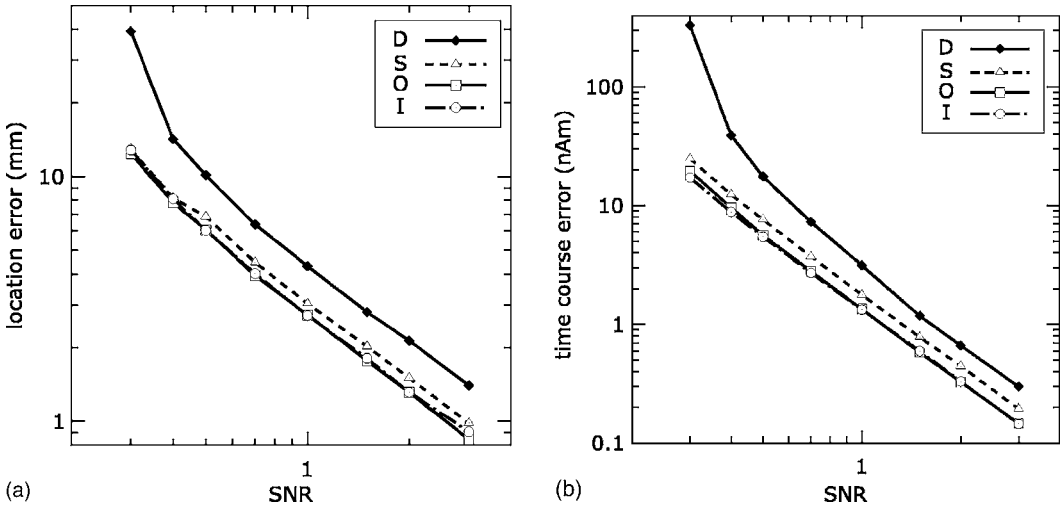

FIG. 6. Combined 90% location error measure (a) and combined 90% time-course errors (b) for a series of single dipole problems: D, diagonal covariance model; S, single KP model [Eq. (4)]; O, orthogonal basis model [Eq. (8)]; and I, independent basis model [Eq. (10)].

when noise components look like signal: i.e., with dipolelike spatial distributions that overlap spatially with target signals and have correlated time courses (cf. [13,36,37]). We could redesign our simulated source model to better match features in the noise in order to make baseline source localization performance worse, but this effectively rigs the test. We have little doubt that in authentic data signal and noise are more similar in structure and, thus, more likely to interact. This suggests that the relative advantage due to the use of noise covariance models will be greater.

It is important to note that, even though we observed limited performance advantages of our models for location estimation at low SNR, our models were always superior in time-course estimation. If the models are evaluated on the basis of combined spatial and temporal performance, then multipair models always performed better. In single-trial analysis, localization performance might not be of as much of an importance as estimation of the time-course dynamical features. For this type of analysis, dynamical features of the background can play a much greater role due to phase locking and similar phenomena. Even small consistent improvements of estimating noise temporal behavior may be of great importance in such a setting.

Figure 7 shows contour plots of spatial components from the orthogonal basis series model and their corresponding temporal covariances. There are clear differences among the temporal covariances. However, if one takes out the components that are either sensor dominated with very short temporal correlations (B2, B6 and D6) and the large 60% powerline component (B1), the remaining components, which are presumably dominated by background brain activity, have similar temporal correlation structure. It will be interesting to see if in future work this feature is observed across multiple subjects and over a more extended temporal range.

The assumption that each noise generator contributing to the measured background retains a consistent spatial distribution is supported by observations of stability of brain function in the resting state [38]. Spatial diversity of the background signal can be observed, for example, as nonuniform spatial distribution of the α rhythm over the cortex. Moreover, the background may not have many generators with distinct independent time courses. In the data examined here, the temporal covariances across components were relatively similar in the multipair models. This is a possible explanation of why the multipair models yield relatively little improvement over the KP model in these studies.

Despite almost identical performance of the multipair models, we suggest that, in terms of generality, the model based on independent basis should be preferred. Structural comparisons using Frobenius norm and scatter plots also demonstrate that the independent basis multipair model is closer to the full spatiotemporal sample covariance. Its power is not only in the increase of degrees of freedom and, consequently, the ability to describe a bigger class of possible spatial components, but also in the discriminative criterion. The main criterion used in PCA is minimization of the reconstruction error [39]. As a result, this produces spatially uncorrelated components. This may not be a good model of the MEG/EEG background, especially that due to the background brain activity. It seems more reasonable to assume that different noise generators are independent. In that case, the independent basis model is complex enough to describe the underlying process, and a correct choice of algorithm (for example an ICA algorithm) can estimate independent generators.

Considering the observation that not all spatial components may have distinct temporal structure, a further generalization is needed. The multipair model can have an additional parameter for the number of significant components. In this case, only significant components will contribute to the final covariance with their respective distinct time courses. Other components would be suppressed. Spatial covariances of components may have different rank depending on areas they cover. Such reasoning goes in accordance with noise generator models, that state that different regions (or resting networks) acting as background generators have their own temporal structure [40,41]. Unfortunately, an approach to achieve this generalization is not obvious since special care should be taken to retain inversion of the model.

Multipair models can, in principle, be directly applied to the analysis of EEG data. Though the topic needs further study, it is expected that the different structure of EEG would not affect the multipair models as much as the possible added complexity of the noise. We expect that in the case of the heterogeneous noise sources expected in EEG signals, multipair models should show even better performance due to their increased expressive power and ability to incorporate information about many noise components.

Different models that would cover the range between the diagonal and the KP model can be developed. For such models, additional study is needed to investigate how spatial models alone or temporal models alone would compare in performance. These truncated models can be useful in the case when only a small amount of data is available for estimation. But when the single-pair model can be estimated it subsumes all lower-order models and should not perform worse than these. As we show in this paper, multipair models subsume the KP model and, thus, should not perform worse. This was demonstrated by the increased accuracy in localization and especially time-course estimation. In the case when sufficient estimation data are available and the additional computational load of multipair models does not make a significant difference, we suggest using multipair models to increase accuracy. We believe that even in the worst case, multipair models should perform as well or better than the one-pair model.

Many of the methods employed for source-based analysis of neural electromagnetic data have their roots in methods developed in other areas of physics and engineering, e.g., MUSIC [2] and beam-forming [4] approaches are based on techniques used for signal extraction from phased array antennas. In turn, methodological development and enhancements for biomedical application have been used in other disciplines of physical science. We anticipate that the methods described here for characterizing the spatiotemporal covariance structure in neural electromagnetic data will find application in other physics applications, such as geophysical inverse problems, mapping of lightening strikes, and other noise-limited analysis problems that depend on the integration data from an array of discrete sensors.

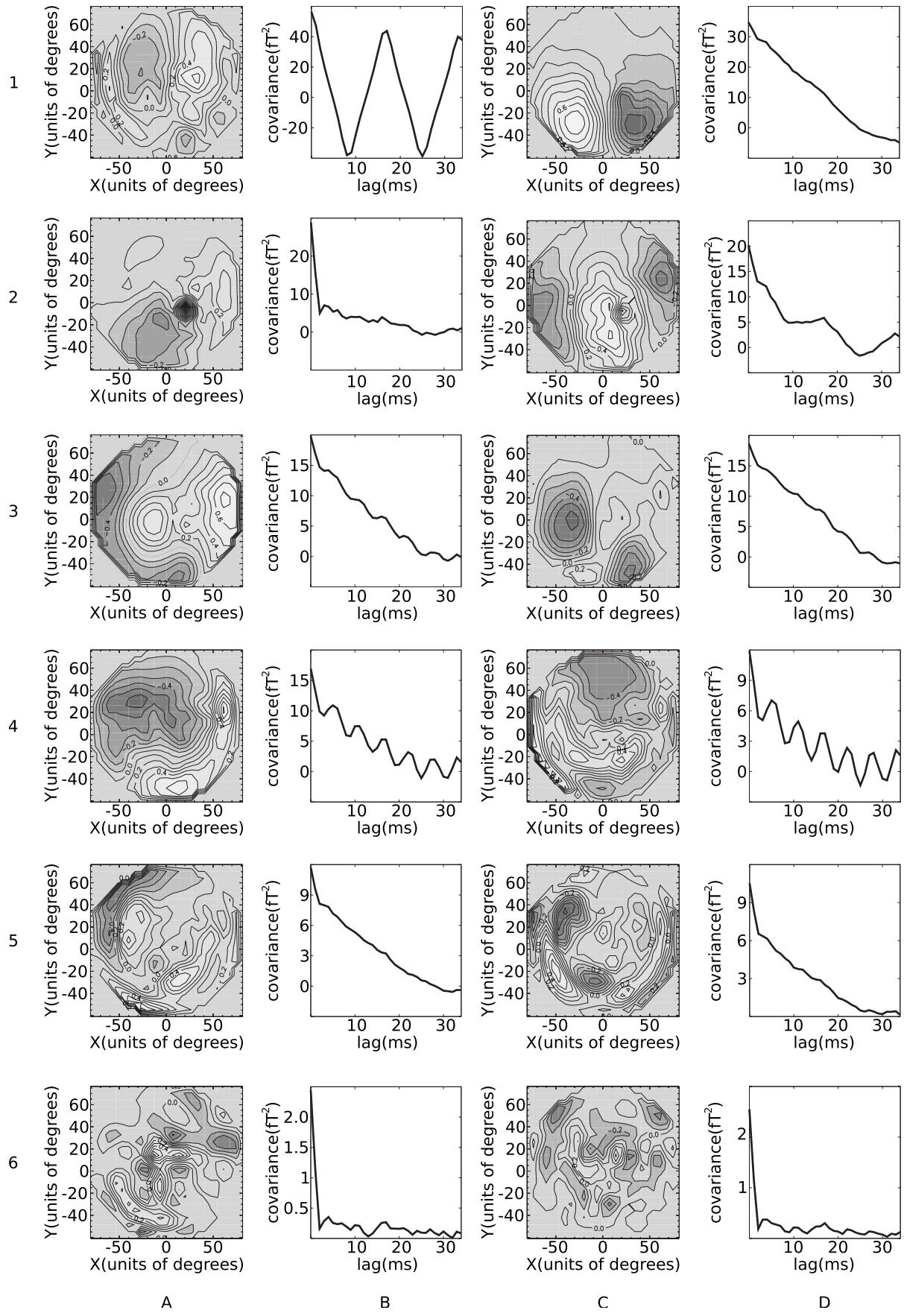

FIG. 7. Contour plots of the first ten most significant spatial components with corresponding temporal covariance functions and of two of the least significant spatial components.

VI. CONCLUSIONS

In this paper, we present two multipair spatiotemporal covariance models that better capture temporal variability in background activity. A significant improvement over previous attempts to use sum of KP is that our models produce computationally manageable inverse. The increased ability to describe temporal structure improved the detection of temporal dynamics of the signal and localization results. The accuracy of the structural approximation was compared using both a norm metric and scatter plots. Models were compared on the basis of their performance in a localization algorithm. Performance statistics were gathered from inverse solutions to a large number of single dipole problems with simulated sources and background from real MEG data. In terms of localization error and detection of signal dynamics, the orthogonal basis and independent basis multipair models demonstrated the best performance.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. 2 R01 EB000310-05 and the Mental Illness and Neuroscience Discovery (MIND) Institute. We thank John Mosher and Robert Kraus for fruitful discussions and help with noise filtering. We thank Barak Pearlmutter for advice about ICA algorithms. We thank Elaine Best for help in preprocessing empirical data using MEGAN (http://www.lanl.gov/p/p21/megan.shtml). Figure 3 showing the dipoles on the cortex was generated using MRIVIEW [33].

APPENDIX A: DERIVATION OF MULTIPAIR MODELS

In the following derivation, we use determinants of the multipair models. Calculation of these determinants is omitted for brevity. The inverse of the orthogonal basis multipair model obtained in (9) is utilized in the following. Using these result, the log likelihood for the Gaussian probability density function (PDF) for the orthogonal basis multipair model is expressed by

$$\mathcal{L} = \text{const} - \frac{M}{2}\sum_{l=1}^{L} \ln(|\boldsymbol{\mathcal{T}}^l|) - \frac{1}{2}\,\text{tr}\left[\sum_{m=1}^{M}\sum_{l=1}^{L} \mathbf{E}_m^T \boldsymbol{\mathcal{S}}^l \mathbf{E}_m (\boldsymbol{\mathcal{T}}^l)^{-1}\right] \tag{A1}$$

Differentiating with respect to $\boldsymbol{\mathcal{T}}^l$results in

$$\begin{aligned} d\mathcal{L} &= -\frac{M}{2}\,\text{tr}[(\boldsymbol{\mathcal{T}}^l)^{-1} d\boldsymbol{\mathcal{T}}^l] + \frac{1}{2}\,\text{tr}\left[\sum_{m=1}^{M} \mathbf{E}_m^T \boldsymbol{\mathcal{S}}^l \mathbf{E}_m (\boldsymbol{\mathcal{T}}^l)^{-1} d\boldsymbol{\mathcal{T}}^l (\boldsymbol{\mathcal{T}}^l)^{-1}\right] \\ &= -\frac{M}{2}\,\text{tr}[(\boldsymbol{\mathcal{T}}^l)^{-1} d\boldsymbol{\mathcal{T}}^l] + \frac{1}{2}\,\text{tr}\left[\sum_{m=1}^{M} (\boldsymbol{\mathcal{T}}^l)^{-1} \mathbf{E}_m^T \boldsymbol{\mathcal{S}}^l \mathbf{E}_m (\boldsymbol{\mathcal{T}}^l)^{-1} d\boldsymbol{\mathcal{T}}^l\right] \\ &= -\frac{M}{2}\,\text{tr}\left\{\left[(\boldsymbol{\mathcal{T}}^l)^{-1} - \frac{1}{M}\sum_{m=1}^{M} (\boldsymbol{\mathcal{T}}^l)^{-1} \mathbf{E}_m^T \boldsymbol{\mathcal{S}}^l \mathbf{E}_m (\boldsymbol{\mathcal{T}}^l)^{-1}\right] d\boldsymbol{\mathcal{T}}^l\right\}. \end{aligned} \tag{A2}$$

Setting the above derivative to zero in order to find the maximum,

$$-\frac{M}{2}\,\text{tr}\left\{\left[(\boldsymbol{\mathcal{T}}^l)^{-1} - (\boldsymbol{\mathcal{T}}^l)^{-1}\left(\frac{1}{M}\sum_{m=1}^{M} \mathbf{E}_m^T \boldsymbol{\mathcal{S}}^l \mathbf{E}_m\right)(\boldsymbol{\mathcal{T}}^l)^{-1}\right]\right\} = 0 \tag{A3}$$

$$(\boldsymbol{\mathcal{T}}^l)^{-1} - (\boldsymbol{\mathcal{T}}^l)^{-1}\left(\frac{1}{M}\sum_{m=1}^{M} \mathbf{E}_m^T \boldsymbol{\mathcal{S}}^l \mathbf{E}_m\right)(\boldsymbol{\mathcal{T}}^l)^{-1} = 0 \tag{A4}$$

$$(\boldsymbol{\mathcal{T}}^l)^{-1}\left(\frac{1}{M}\sum_{m=1}^{M} \mathbf{E}_m^T \boldsymbol{\mathcal{S}}^l \mathbf{E}_m\right)(\boldsymbol{\mathcal{T}}^l)^{-1} = (\boldsymbol{\mathcal{T}}^l)^{-1}, \tag{A5}$$

and multiplying both sides of Eq. (A5) by $(\boldsymbol{\mathcal{T}}^l)^{-1}$from the left and right, we obtain the final result,

$$\boldsymbol{\mathcal{T}}^l = \frac{1}{M}\sum_{m=1}^{M} \mathbf{E}_m^T \boldsymbol{\mathcal{S}}^l \mathbf{E}_m. \tag{A6}$$

Next, we estimate temporal covariance for the independent basis multipair model. The inverse of this model is taken as in (11). The log likelihood in this case,

$$\mathcal{L} = \text{const} - \frac{M}{2}\left[2C\ln(|\mathbf{W}^{-1}|) + \sum_{l=1}^{L}\ln(|\boldsymbol{\mathcal{T}}^l|)\right] - \frac{1}{2}\,\text{tr}\left[\sum_{m=1}^{M}\sum_{l=1}^{L} \mathbf{E}_m^T \mathbf{W}\mathbf{W}^T \boldsymbol{\mathcal{R}}^l \mathbf{W}\mathbf{W}^T \mathbf{E}_m (\boldsymbol{\mathcal{T}}^l)^{-1}\right] \tag{A7}$$

Following the same path for derivation as for the orthogonal basis model case, we end up with

$$\boldsymbol{\mathcal{T}}^l = \frac{1}{M}\sum_{m=1}^{M} \mathbf{E}_m^T \mathbf{W}\mathbf{W}^T \boldsymbol{\mathcal{R}}^l \mathbf{W}\mathbf{W}^T \mathbf{E}_m. \tag{A8}$$

APPENDIX B: INVERSION OF THE MULTIPAIR MODELS

We need the following identity for derivations of this section:

$$(\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \otimes \mathbf{B}_2) = (\mathbf{A}_1\mathbf{A}_2 \otimes \mathbf{B}_1\mathbf{B}_2). \tag{B1}$$

First, let us prove that the inverse for the orthogonal basis multipair model is as in (9). By the orthogonality of $\boldsymbol{\mathcal{S}}^l$ ($\boldsymbol{\mathcal{S}}^l\boldsymbol{\mathcal{S}}^{l'}=0$ if $l \neq l'$),

$$\tilde{\mathcal{C}}\tilde{\mathcal{C}}^{-1} = \left(\sum_{l=1}^{L} \lambda_l^2 \boldsymbol{\mathcal{T}}^l \otimes \boldsymbol{\mathcal{S}}^l\right)\left[\sum_{l=1}^{L} \lambda_l^{-2} (\boldsymbol{\mathcal{T}}^l)^{-1} \otimes \boldsymbol{\mathcal{S}}^l\right] = \sum_{l=1}^{L} \mathbf{I} \otimes (\boldsymbol{\mathcal{S}}^l)^2. \tag{B2}$$

By two properties of $\boldsymbol{\mathcal{S}}^l$: (i) $(\boldsymbol{\mathcal{S}}^l)^2 = \boldsymbol{\mathcal{S}}^l$; (ii) $\Sigma_l \boldsymbol{\mathcal{S}}^l = \mathbf{I}$ due to orthogonality of $\mathbf{V}$, we obtain the desired one,

$$\tilde{\mathcal{C}}\tilde{\mathcal{C}}^{-1} = \sum_{l=1}^{L} \mathbf{I} \otimes (\boldsymbol{\mathcal{S}}^l)^2 = \mathbf{I} \otimes \sum_{l=1}^{L} \boldsymbol{\mathcal{S}}^l = \mathbf{I} \otimes \mathbf{I} = \mathbf{I}. \tag{B3}$$

Properties of $\boldsymbol{\mathcal{S}}^l$ are proved as follows:
Property (i):

$$(\boldsymbol{\mathcal{S}}^l)^2 = (v_l v_l^T)(v_l v_l^T) = v_l (v_l^T v_l) v_l^T = v_l v_l^T \ (v_l^T v_l = 1 \text{ by orthogonality of } \mathbf{V}) = \boldsymbol{\mathcal{S}}^l \tag{B4}$$

Property (ii): $\mathbf{I} = \mathbf{V}\mathbf{V}^T = \Sigma_l v_l v_l^T$.

Now we show that the inverse of the independent components multipair model is as in (11). Using the identity (B1),

$$\left(\sum_{l=1}^{L} \boldsymbol{\mathcal{T}}^l \otimes \boldsymbol{\mathcal{R}}^l\right)\left(\sum_{l=1}^{L} (\boldsymbol{\mathcal{T}}^l)^{-1} \otimes [\mathbf{W}\mathbf{W}^T \boldsymbol{\mathcal{R}}^l \mathbf{W}\mathbf{W}^T]\right) = \sum_{l,l'=1}^{L} \boldsymbol{\mathcal{T}}^l (\boldsymbol{\mathcal{T}}^{l'})^{-1} \otimes [\boldsymbol{\mathcal{R}}^l \mathbf{W}\mathbf{W}^T \boldsymbol{\mathcal{R}}^{l'} \mathbf{W}\mathbf{W}^T]. \tag{B5}$$

According to the definition of $\boldsymbol{\mathcal{R}}_l$,

$$\boldsymbol{\mathcal{R}}_l = w_l w_l^T = (\mathbf{W}^{-1})^T e_l e_l^T \mathbf{W}^{-1}. \tag{B6}$$

Here $\{e_l | l = 1, \cdots, L\}$ are orthonormal canonical bases vectors. By the orthogonality of e_l and substituting (B6) into (B5), we obtain

$$\begin{aligned} \boldsymbol{\mathcal{R}}^l \mathbf{W}\mathbf{W}^T \boldsymbol{\mathcal{R}}^{l'} \mathbf{W}\mathbf{W}^T &= (\mathbf{W}^{-1})^T e_l e_l^T \mathbf{W}^{-1} \mathbf{W}\mathbf{W}^T (\mathbf{W}^{-1})^T e_{l'} e_{l'}^T \mathbf{W}^{-1} \mathbf{W}\mathbf{W}^T \\ &= (\mathbf{W}^{-1})^T e_l e_l^T e_{l'} e_{l'}^T \mathbf{W}^T \\ &= \delta(l,l') (\mathbf{W}^{-1})^T e_l e_l^T \mathbf{W}^T. \end{aligned} \tag{B7}$$

Finally, we obtain the desired result,

$$\begin{aligned} &\left(\sum_{l=1}^{L} \boldsymbol{\mathcal{T}}^l \otimes \boldsymbol{\mathcal{R}}^l\right)\left(\sum_{l=1}^{L} (\boldsymbol{\mathcal{T}}^l)^{-1} \otimes [\mathbf{W}\mathbf{W}^T \boldsymbol{\mathcal{R}}^l \mathbf{W}\mathbf{W}^T]\right) \\ &= \mathbf{I} \otimes (\mathbf{W}^{-1})^T \left(\sum_l e_l e_l^T\right) \mathbf{W}^T \\ &= \mathbf{I} \otimes (\mathbf{W}^{-1})^T \mathbf{I} \mathbf{W}^T \\ &= \mathbf{I}. \end{aligned} \tag{B8}$$

[1] M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. Lounasmaa, Rev. Mod. Phys. **65**, 413 (1993).
[2] J. C. Mosher, P. S. Lewis, and R. M. Leahy, IEEE Trans. Biomed. Eng. **39**, 541 (1992).
[3] D. M. Schmidt, J. S. George, and C. C. Wood, Hum. Brain Mapp **7**, 195 (1999).
[4] B. D. VanVeen, W. vanDrongelen, M. Yuchtman, and A. Suzuki, IEEE Trans. Biomed. Eng. **44**, 867 (1997).
[5] M. Hämäläinen and R. Ilmoniemi, Med. Biol. Eng. Comput. **32**, 35 (1994).
[6] J. W. Phillips, R. M. Leahy, and J. C. Mosher, IEEE Trans. Med. Imaging **16**, 338 (1997).
[7] S. C. Jun, B. A. Pearlmutter, and G. Nolte, Phys. Med. Biol. **47**, 2547 (2002).
[8] S. Kuriki, F. Takeuchi, and T. Kobayashi, Electroencephalogr. Clin. Neurophysiol. **92**, 56 (1994).
[9] H. Huizenga and P. Molenaar, Electroencephalogr. Clin. Neurophysiol. **99**, 562 (1996).
[10] B. Lutkenhoner, Electroencephalogr. Clin. Neurophysiol. **106**, 322 (1998).
[11] V. Strassen, Numer. Math. **13**, 354 (1969).
[12] D. Coppersmith and S. Winograd, J. Symb. Comput. **9**, 251 (1990).
[13] J. C. de Munck, H. M. Huizenga, L. J. Waldorp, and R. Heethaar, IEEE Trans. Signal Process. **50**, 1565 (2002).
[14] F. Bijma, J. C. de Munck, and R. M. Heethaar, Neuroimage **27**, 402 (2005), ISSN 1053-8119.
[15] L. K. Saul, K. Q. Weinberger, J. H. Ham, F. Sha, and D. D. Lee, *Spectral Methods for Dimensionality Reduction*, in Semisupervised Learning, edited by B. Schoelkopf, O. Chapelle. A. Zien (MIT Press, Cambridge, MA, 2006).
[16] R. Vigario, J. Sarela, V. Jousmaki, M. Hamalainen, and E. Oja, IEEE Trans. Biomed. Eng. **47**, 589 (2000).
[17] A. Tang, B. Pearlmutter, N. Malaszenko, D. Phung, and B. Reeb, Neural Comput. **14**, 1827 (2002).
[18] A. Tang, B. Pearlmutter, N. Malaszenko, and D. Phung, Neuroimage **17**, 1773 (2002).
[19] A. Ziehe, K. Muller, G. Nolte, B. Mackert, and G. Curio, J. Cell Physiol. **47**, 75 (2000).
[20] J. Cao, N. Murata, S. Amari, A. Cichocki, T. Takeda, H. Endo, and N. Harada, IEICE Trans. Fundamentals **E83-A**, 1757 (2000).
[21] S. Makeig, T. Jung, A. Bell, D. Ghahremani, and T. Sejnowski, Proc. Natl. Acad. Sci. U.S.A. **94**, 10979 (1997).
[22] S. Makeig, M. Westerfield, T. Jung, J. Covington, J. Townsend, T. Sejnowski, and E. Courchesne, J. Neurosci. **19**, 2665 (1999).
[23] T. Jung, S. Makeig, C. Humphries, T. Lee, M. McKeown, V. Iragui, and T. Sejnowski, Prog. Surf. Sci. **37**, 163 (2000).
[24] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and É.

Moulines, in *Proceedings of the International Conference on Digital Signal Processing*, Cyprus (University of Cyprus Press, Nicosia, 1993), pp. 346–351.
[25] A. J. Bell and T. J. Sejnowski, Neural Comput. **7**, 1129 (1995).
[26] A. Hyvärinen and E. Oja, Neural Comput. **9**, 1483 (1997).
[27] A. Cichocki, S. Amari, and K. Siwek, http://www.bsp.brain.riken.jp/ICALAB
[28] A. Ahonen, M. Hamalainen, M. Kajola, J. Knuutila, P. Laine, O. Lounasmaa, L. Parkkonen, J. Simola, and C. Tesche, Phys. Status Solidi A **T49**, 198 (1993).
[29] S. Kullback and R. A. Leibler, Ann. Math. Stat. **22**, 79 (1951).
[30] H. M. Huizenga, J. C. De Munck, L. J. Waldorp, and R. P. Grasman, IEEE Trans. Biomed. Eng. **50**, 533 (2002).
[31] J. Sarvas, Phys. Med. Biol. **32**, 11 (1987).
[32] M. Huang, C. J. Aine, S. Supek, E. Best, D. Ranken, and E. R. Flynn, Electroencephalogr. Clin. Neurophysiol. **108**, 32 (1998).
[33] D. Ranken, E. Best, J. Stephen, D. Schmidt, J. George, C. C. Wood, and M. X. Huang, in *Biomag 2002 Proceedings* (VDE Verlag, Jena, 2002), pp. 785–787, http://www.lanl.gov/p/p21/mriview.shtml
[34] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (Chapman & Hall, London, 1995).
[35] S. C. Jun, J. S. George, J. Pare-Blagoev, S. Plis, D. M. Ranken, D. M. Schmidt, and C. C. Wood, Neuroimage **28**, 84 (2005).
[36] S. Supek and C. J. Aine, Hum. Brain Mapp **5**, 139 (1997).
[37] S. Supek and C. J. Aine, Hum. Brain Mapp **5**, 154 (1997).
[38] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman, Proc. Natl. Acad. Sci. U.S.A. **98**, 676 (2001).
[39] J. J. Gerbrands, Pattern Recogn. **14**, 375 (1980), ISSN 0031-3203.
[40] H. Laufs, K. Krakow, P. Sterzer, E. Eger, A. Beyerle, A. Salek-Haddadi, and A. Kleinschmidt, Proc. Natl. Acad. Sci. U.S.A. **100**, 11053 (2003).
[41] J. C. de Munck, P. C. M. Vijn, and F. H. L. da Silva, IEEE Trans. Biomed. Eng. **39**, 791 (1992).