A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering

Andreas Galka[1,3], Okito Yamashita[2], Tohru Ozaki[2,3], Rolando Biscay[4] and Pedro Valdés-Sosa[5]

[1] *Institute of Experimental and Applied Physics, University of Kiel, 24098 Kiel, Germany*
[2] *Department of Statistical Science, The Graduate University for Advanced Studies, Minami-Azabu 4-6-7, Tokyo 106-8569, Japan*
[3] *Institute of Statistical Mathematics (ISM), Minami-Azabu 4-6-7, Tokyo 106-8569, Japan*
[4] *University of Havana, Ciudad Habana, Calle 34 No. 304-2, Cuba*
[5] *Cuban Neuroscience Center, Ave 25 No. 5202 esquina 158 Cubanacán, POB 6880, 6990, Ciudad Habana, Cuba*

July 16, 2003; revised version: January 21, 2004

Abstract

We present a new approach for estimating solutions of the dynamical inverse problem of EEG generation. In contrast to previous approaches, we reinterpret this problem as a filtering problem in a state space framework; for the purpose of its solution we propose a new extension of Kalman filtering to the case of spatiotemporal dynamics. The temporal evolution of the distributed generators of the EEG can be reconstructed at each voxel of a discretisation of the gray matter of brain. By fitting linear autoregressive models with neighbourhood interactions to EEG time series new classes of inverse solutions with improved resolution and localisation ability can be explored. For the purposes of model comparison and parameter estimation from given data we employ a likelihood maximisation approach. Both for instantaneous and dynamical inverse solutions we derive estimators of the time-dependent estimation error at each voxel. The performance of the algorithm is demonstrated by application to simulated and clinical EEG recordings. It is shown that by choosing appropriate dynamical models it becomes possible to obtain inverse solutions of considerably improved quality, as compared to the usual instantaneous inverse solutions.

1 Introduction

Recordings of electromagnetic fields emanating from human brain are well known to provide an important source of information about brain dynamics. Electrical potentials on the scalp surface are very easy to measure at a set of electrodes attached to the skin; as a result multivariate electroencephalographic (EEG) time series are obtained. With considerably higher technical effort, magnetoencephalographic (MEG) time series can also be recorded.

It is by now widely accepted that the sources of these electromagnetic fields are electrical currents within networks of neurons in the cortex and other gray matter structures

of brain; while part of this current remains confined within the dendritic trunks (*primary currents*), another part flows through the extracellular volume (*secondary currents*) [1]. In order to obtain more direct access to the dynamics governing the activity of these networks of neurons it would be desirable to have direct estimates of these sources. The estimation of these sources from recordings of EEG or MEG has recently become a subject of intense research [2, 3, 4, 5, 6, 7, 9, 13, 23, 41]; for a recent review see Baillet *et al.* [8]. In this paper we focus on the case of the EEG, but the ideas and methods to be presented remain equally valid for the MEG.

Two main classes of source models have been developed: "equivalent current dipole" approaches (also known as "parametric" methods), in which the sources are modeled by a relatively small number of focal sources at locations to be estimated from the data, and "linear distributed" approaches (also known as "imaging" or "current density reconstruction" methods), in which the sources are modeled by a dense set of dipoles distributed at fixed locations (which, in analogy to the case of magnetic resonance imaging, we shall call "voxels") throughout the head volume. Examples of parametric approaches include least-squares source estimation [9] and spatial filters, such as beamforming and multiple signal classification ("MUSIC") approaches [10]. This paper exclusively deals with the linear distributed model approach.

It is a characteristic problem of distributed source models that a large number of unknown quantities has to be estimated from a much smaller number of measurements; as a consequence of this, we are facing a problem which does not possess a unique solution, known as "inverse problem". The number of measurements given at one instant of time may be as low as 18, if the standard 10-20 system of clinical EEG recordings is employed; by increasing the number of electrodes we may eventually obtain up to a few hundred measurements, but they will fail to provide an equivalent amount of independent information due to strong correlations between adjacent electrodes. On the other hand, the number of voxels will typically be several thousand, and furthermore at each voxel site a full three-dimensional current vector has to be modelled.

In order to identify a unique solution (i.e. an "inverse solution") additional information has to be employed. So far this has been done mainly by imposing constraints on the inverse solution. Certain constraints can be obtained from neurophysiology [5]; as an example, it is reasonable to assume that only voxels within gray matter contribute substantially to the generation of the electromagnetic fields; other constraints refer to the probable direction of local current vectors at specific locations. But such constraints do not suffice to remove the ambiguity of the inverse solution.

For this purpose much more restrictive constraints are needed, such as the minimum-norm constraint suggested by Hämäläinen and Ilmoniemi [11] or the maximum-smoothness constraint suggested by Pascual-Marqui [3]. These constraints can be applied independently for each instant of time, without accessing the data measured at other instants of time, therefore we will say that the resulting inverse solutions represent solutions of the "instantaneous" inverse problem.

The idea of including data from more than a single instant of time into the estimation of inverse solutions is attractive, since more information becomes available for the solution of an ill-posed problem; consequently, there has recently been growing interest in generalising the instantaneous inverse problem to "dynamical" inverse problems and to develop algorithms for its solution [12, 6, 13, 14].

In this paper we will contribute to these efforts by developing a new interpretation of the dynamic inverse problem in its most general shape, and by proposing a new approach to its solution. In contrast to most previous work, we will not approach this problem

within a constrained least squares (or, equivalently, Bayesian) framework, but by reformulating it as a spatiotemporal state space filtering problem. So far the dynamical aspect of the available algorithms was essentially limited to imposing temporal smoothness constraints [15, 6]; from a time-domain modelling perspective such constraints correspond to the very special case of a spatially non-coupled random-walk model [14]. By appropriate generalisation our approach will permit the use of much more general predictive models in this context, such that a consistent description of the spatiotemporal dynamics of brain becomes possible.

It should be stressed that general predictive models can also be incorporated into the framework of constrained least squares, and in a companion paper to this paper we will discuss in more detail the application of this idea to the inverse problem of EEG generation [7].

As the main tool for our task we will adapt the well-known Kalman filter [16] to spatiotemporal filtering problems; it will become evident that Kalman filtering provides a natural framework for addressing the dynamical inverse problem of EEG generation. Theoretically, by employing a very high-dimensional state vector, standard Kalman filtering could deal with any spatiotemporal filtering problem, but the specific nature of the spatial dimensions (such as neighbourhood relationships between voxels) would not be properly captured, and computational expenses would soon become prohibitively large. By assuming a properly chosen structure of the dynamics and certain additional approximations the intractable high-dimensional filtering problem can be decomposed into a coupled set of tractable low-dimensional filtering problems. This adaptation can be regarded as a generalisation of standard Kalman filtering to the case of partial (space-time) differential equations.

From system theory it is known that the sufficient condition for successful application of Kalman filtering is observability of the given state space system, as represented by its state transition parameter matrix (or, in the nonlinear case, the corresponding Jacobian matrix) and its observation matrix [17, 18]. Although we will not be able to rigorously prove observability, we will discuss the application of this concept to our model and demonstrate through an explicit numerical simulation study that Kalman filtering can successfully be applied. From this simulation it will also become evident that a crucial element for the estimation of dynamical inverse solutions is given by the model according to which the dynamics of the voxel currents is assumed to evolve. If a very simple model is chosen, we will obtain solutions which offer only small improvements over solutions resulting from previous non-dynamical (i.e. instantaneous) algorithms for solving the inverse problem; if the model contains additional information about the true dynamics, much better solutions can be obtained. Such information can at least partly be obtained by choosing a model with a sufficient flexibility for adaptation to given data; this adaptation can be performed by suitable fitting of dynamical parameters and noise covariances.

We will propose to employ the maximum likelihood method for this fitting task; it will become evident that Kalman filtering is the natural tool for calculating the likelihood from EEG data. By assuming a somewhat more general viewpoint, we will address parameter estimation as a special case of model comparison, and employ appropriate statistical criteria, namely the Akaike Information Criterion (AIC) and its Bayesian variant, ABIC, for the purpose of comparing inverse solutions. Numerical examples will be shown both for simulated data and for a clinical EEG recording.

Finally, it should be mentioned that there exists already a sizable amount of published work dealing with applications of Kalman filtering to the analysis of EEG recordings, which we are unable to review in any appropriate form in this paper; and furthermore

there exist also applications of Kalman filtering, or related filtering approaches, to inverse problems, some of which also fall into the field of biomedical data analysis, like Electrical Impedance Tomography or evoked potential estimation (see e.g. [19, 20, 12]). Recently Somersalo *et al.* [14] have applied a nonlinear alternative to Kalman filtering, known as *particle filtering*, to the problem of estimating focal sources from MEG data. While providing important results and methodology, so far none of these studies has addressed the problem of reconstruction of distributed sources from EEG (or MEG) times series in the context of identification of optimal dynamical (i.e. predictive) models.

In this paper we will introduce a new practicable solution for the problem of applying Kalman filtering to very high dimensional filtering problems, as they arise in the case of spatiotemporal brain dynamics. Moreover, we will replace the largely arbitrary choice of dynamical models which can be seen in many applications of Kalman filtering, by a systematic model comparison and selection approach which provides explicit justification for the chosen model and parameters in terms of the numerical value of a statistical criterion. By this approach a wide class of very general models becomes available for data-driven modelling of brain dynamics.

2 The inverse problem of EEG generation

We start from a rectangular grid of N_v voxels covering the cortical gray matter parts of human brain; in this study inverse solutions will be confined to these voxels. In the particular discretisation which we will employ, there are $N_v = 3433$ cortical gray-matter voxels. At each voxel a local three-dimensional current vector

$$\mathbf{j}(v,t) = \big(j_x(v,t), j_y(v,t), j_z(v,t)\big)^\dagger$$

is assumed, where v is a voxel label, t denotes time, and $\dagger$ denotes matrix transposition. The column vector of all current vectors (i.e. for all gray-matter voxels) will be denoted by

$$\mathbf{J}(t) = \big(\mathbf{j}(1,t)^\dagger, \mathbf{j}(2,t)^\dagger, \ldots, \mathbf{j}(N_v,t)^\dagger\big)^\dagger \ ;$$

it represents the dynamical state variable of the entire system.

These currents are mapped to the electroencephalographic signal (EEG), which is recorded at the scalp surface. The EEG at an individual electrode shall be denoted by $y(i,t)$, where i is an electrode label; the n_c-dimensional column vector composed of all the electric potentials at all available electrodes shall be denoted by

$$\mathbf{Y}(t) = \big(y(1,t), y(2,t), \ldots, y(n_c,t)\big)^\dagger \ .$$

In this study we assume that the 10-20 system is employed; all potentials refer to average reference, although other choices are possible. Due to the choice of a reference out of the set of electrodes it is advisable to exclude one of the standard electrodes of the 10-20 system from further analysis, such that the effective dimension of $\mathbf{Y}$ becomes $n_c = 18$.

For distributed source models it is possible to approximate the mapping from $\mathbf{J}$ to $\mathbf{Y}$ by a linear function [8] whence it can be expressed as

$$\mathbf{Y}(t) = \mathsf{K}\,\mathbf{J}(t) + \boldsymbol{\epsilon}(t) \ . \tag{1}$$

Here K denotes a $n_c \times 3N_v$ transfer matrix, commonly called "lead field matrix". This matrix can approximately be calculated for a three-shell head model and given electrode

locations by the "boundary element method" [21, 3, 22, 23]. It is an essential precondition for any approach to find inverse solutions, that a reliable estimate of this matrix is available. Here we remark that typically the lead field matrix turns out to be of full rank.

It will be convenient for later use to define the individual contribution of each voxel to the vector of observations by $\mathsf{K}(v)\,\mathbf{j}(v,t)$, where $\mathsf{K}(v)$ is the $n_c \times 3$ matrix which results from extracting those three columns out of K, which are multiplied with $\mathbf{j}(v,t)$ in the process of the multiplication of K and $\mathbf{J}(t)$. From this definition, equation 1 can also be written as

$$\mathbf{Y}(t) = \sum_{v=1}^{N_v} \mathsf{K}(v)\,\mathbf{j}(v,t) + \boldsymbol{\epsilon}(t) \quad . \tag{2}$$

By $\boldsymbol{\epsilon}(t)$ we denote a vector of observational noise, which we assume to be white and Gaussian with zero mean and covariance matrix $\mathsf{C}_\epsilon = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\dagger)$. We will make the assumption that C_ϵ has the simplest possible structure, namely

$$\mathsf{C}_\epsilon = \sigma_\epsilon^2\,\mathsf{I}_{n_c} \quad , \tag{3}$$

where I_{n_c} denotes the $n_c \times n_c$ identity matrix, i.e. we assume that the observation noise is uncorrelated between all pairs of electrodes and of equal variance for all electrodes. These assumptions may be relaxed in future work.

Equation 1 is part of the standard formulation of the inverse problem of EEG [24]; in this paper we propose to interprete it as an observation equation in the framework of dynamical state-space modelling.

The inverse problem of EEG generation is given by the problem of estimating the generators $\mathbf{J}(t)$ from the observed EEG $\mathbf{Y}(t)$; this obviously constitutes an ill-posed problem since the dimension of $\mathbf{J}(t)$ is much larger than the dimension of $\mathbf{Y}(t)$. As also in the case of many other inverse problems, it is nevertheless possible to obtain approximate estimates of $\mathbf{J}(t)$. As a representative of the numerous approaches which have been proposed for this purpose we select here the "low-resolution brain electromagnetic tomography" (LORETA) algorithm, proposed by Pascual-Marqui *et al.* [3], as a starting point; a brief introduction will be given in the next section.

3 The instantaneous case

3.1 The LORETA approach

In this approach a spatial smoothness constraint is imposed on the estimate of $\mathbf{J}(t)$, which can be expressed by employing a discrete spatial Laplacian operator defined by

$$\mathsf{L} = \left(\mathsf{I}_{N_v} - \frac{1}{6}\mathsf{N}\right) \otimes \mathsf{I}_3 \quad . \tag{4}$$

Here N denotes a $N_v \times N_v$ matrix having $\mathsf{N}_{vv'} = 1$ if v' belongs to the set of neighbours of voxel v (this set shall be denoted by $\mathscr{N}(v)$), and 0 otherwise. By the symbol $\otimes$ Kronecker multiplication of matrices is denoted. The $(3i)$th row vector of L acts as a discrete differentiating operator by forming differences between the nearest neighbours of the ith voxel and ith voxel itself (with respect to the first vector component).

In the LORETA approach the inverse solution is obtained by minimizing the objective function

$$E(\mathbf{J}) = ||(\mathbf{Y} - \mathsf{K}\mathbf{J})||^2 + \lambda^2 ||\mathsf{L}\mathbf{J}||^2 \quad , \tag{5}$$

i.e. a weighted sum of the observation fitting error and of a term measuring non-smoothness by the norm of the spatial Laplacian of the state vector. $||.||$ denotes Euclidean norm. The hyperparameter λ expresses the balance between fitting of observations and the smoothness constraint; a non-zero value for λ provides regularisation for the solution [25].

Here we would like to mention that the second term in equation 5 represents a special example of a general constraint term; by appropriate choice of this term it is also possible to impose other kinds of constraints instead of spatial smoothness, such as anatomical constraints or sparseness of the inverse solution (see [8] and references cited therein).

The least squares solution of the problem of minimising equation 5 is given by

$$\hat{\mathbf{J}} = (\mathsf{K}^\dagger\mathsf{K} + \lambda^2\mathsf{L}^\dagger\mathsf{L})^{-1}\mathsf{K}^\dagger\mathbf{Y} \ ; \tag{6}$$

here by $\hat{\mathbf{J}}$ the estimator of the state vector $\mathbf{J}$ is denoted. Within the framework of Bayesian inference, this solution can be interpreted as the Maximum A Posteriori (MAP) solution for the case of Gaussian distributions for the likelihood and the prior, given by [26, 7]:

$$\begin{aligned} p(\mathbf{Y}|\mathbf{J};\sigma_\epsilon^2) &\sim \mathcal{N}(\mathsf{K}\mathbf{J},\sigma_\epsilon^2) \\ p(\mathbf{J};\tau^2) &\sim \mathcal{N}(0,\tau^2(\mathsf{L}^\dagger\mathsf{L})^{-1}) \ , \end{aligned} \tag{7}$$

where we have defined $\tau = \sigma_\epsilon/\lambda$.

Note that $\hat{\mathbf{J}}$ will not depend on the reference according to which the EEG data $\mathbf{Y}$ was measured; this dependence is absorbed into the lead-field matrix. This effect represents another advantage of transforming EEG data into an estimated source current density: The notorious reference problem of EEG is completely removed by this transformation.

The matrix $\mathsf{K}^\dagger\mathsf{K}+\lambda^2\mathsf{L}^\dagger\mathsf{L}$ in equation 6 has the size $3N_v \times 3N_v \approx 10^4\times10^4$, whence actual numerical inversion is usually impracticable. The solution can be evaluated nevertheless by using the singular value decomposition of the $n_c \times 3N_v$ matrix $\mathsf{K}\mathsf{L}^{-1}$,

$$\mathsf{K}\mathsf{L}^{-1} = \mathsf{U}\mathsf{S}\mathsf{V}^\dagger \ , \tag{8}$$

where U is an orthogonal $n_c \times n_c$ matrix, S is a $n_c \times 3N_v$ matrix whose only non-zero elements are the singular values $\mathsf{S}_{ii} \equiv s_i$, $i = 1,\ldots,n_c$, and V is an orthogonal $3N_v \times 3N_v$ matrix; only the first n_c columns of V are relevant for this decomposition, and the corresponding $3N_v \times n_c$ matrix shall be denoted by $\mathsf{V}^{(1)}$. The matrix composed of the remaining $3N_v - n_c$ columns shall be denoted by $\mathsf{V}^{(2)}$. After some transformations, equation 6 becomes

$$\hat{\mathbf{J}} = \mathsf{L}^{-1}\mathsf{V}^{(1)}\,\mathrm{diag}\left(\frac{s_i}{s_i^2+\lambda^2}\right)\mathsf{U}^\dagger\mathbf{Y} \ . \tag{9}$$

Here $\mathrm{diag}(x_i)$ denotes a diagonal matrix with elements $x_1,\ldots,x_{n_c}$ on its diagonal. Numerical evaluation of this expression can be implemented very efficiently.

3.2 Estimation of the regularisation parameter λ

Since the inverse solution given by equation 9 will depend sensitively on the value of the hyperparameter λ, it should be chosen in an objective way. Various statistical criteria, such as Generalised Cross-Validation (GCV) [27], and *ad hoc* methods, such as the L-curve approach [28] have been employed for this purpose. Instead of these approaches, in this paper we have chosen to use the Akaike Bayesian Information Criterion (ABIC) [29, 30],

since we intend to directly compare inverse solutions obtained by different techniques by comparing their likelihood.

Given a time series of EEG observations $\mathbf{Y}(1), \cdots, \mathbf{Y}(N_t)$, ABIC is defined as (-2) times the type-II log-likelihood, i.e. the log-likelihood of the hyperparameters in the context of empirical Bayesian inference. In the case of the model containing unobservable variables, the type-II likelihood can be obtained by averaging the joint distribution of all variables, both observable and unobservable, over the unobservable variables, i.e. by forming the marginal distribution:

$$\mathrm{ABIC}(\sigma_\epsilon, \tau) = -2 \log L_{II}(\sigma_\epsilon, \tau) = -2 \sum_{t=1}^{N_t} \log \int p(\mathbf{Y}(t)|\mathbf{J}(t); \sigma_\epsilon^2)\, p(\mathbf{J}(t); \tau^2)\, d\mathbf{J}(t) \quad , \quad (10)$$

where $\mathbf{Y}(t)$ are the observable and $\mathbf{J}(t)$ the unobservable variables; σ_ϵ, τ are hyperparameters, and again $\tau = \sigma_\epsilon / \lambda$.

Estimators for σ_ϵ and λ can be obtained by maximising the likelihood given by equation 10; how this can be done in an efficient way will be presented elsewhere in more detail [7]. Here we give only the result:

The type-II log-likelihood $L_{II}(\sigma_\epsilon, \tau)$ itself can be shown to be

$$L_{II}(\sigma_\epsilon, \lambda) = \sum_{t=1}^{N_t} \left(\sum_{i=1}^{n_c} \log \frac{s_i^2 + \lambda^2}{\lambda^2} + n_c (1 + \log 2\pi\sigma_\epsilon^2) \right) \quad , \quad (11)$$

where the estimate of the observation noise variance σ_ϵ^2 is given by

$$\sigma_\epsilon^2 = \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{\lambda^2}{s_i^2 + \lambda^2} \bar{y}_i^2(t) \quad ; \quad (12)$$

here $\bar{y}_i(t)$ denotes the ith element of the vector $\mathsf{U}^\dagger \mathbf{Y}(t)$, where U is defined in equation 8.

As a result we obtain not only estimates for the hyperparameters, but also the possibility to calculate the ABIC value for any given inverse solution (as obtained by LORETA), i.e. an estimate for the type-II likelihood. This will enable us to compare inverse solutions obtained by different techniques, since for given data the likelihood serves as a general measure of the quality of hypotheses [31].

It should be mentioned here that despite using an improved statistical criterion, the proper choice of the regularisation parameter remains a difficult problem of the LORETA method; in practice, frequently even the order of magnitude of the appropriate value of λ is debatable and may change drastically upon seemingly insignificant changes of the data. Also for this reason there is a growing need for an alternative approach to estimating inverse solutions.

3.3 Estimation of the covariance matrix of the estimated state vector

So far the lack of an efficient method for assessing the approximate error associated with the inverse solutions obtained by LORETA has been a serious weakness of this technique; certainly it has contributed much to the widespread scepticism which the very idea of estimating solutions for the inverse problem of EEG generation is still facing. Strictly speaking, estimates of unobservable quantities without estimates of the corresponding

error have to be regarded as meaningless. For this reason we introduce a method for estimating the covariance matrix $\mathsf{C}_{\hat{\mathbf{J}}}^{(iIS)}$ of the estimated currents $\hat{\mathbf{J}}(t)$ of an inverse solution obtained by LORETA (here the superscript (iIS) denotes "instantaneous inverse solution").

The derivation of the expression for $\mathsf{C}_{\hat{\mathbf{J}}}^{(iIS)}$ is inspired by recent work by Pascual-Marqui [32]; the result is given by

$$\mathsf{C}_{\hat{\mathbf{J}}}^{(iIS)} = \tau^2 \mathsf{L}^{-1} \mathsf{V}^{(1)} \mathrm{diag}\left(\frac{s_i^2}{s_i^2+\lambda^2}\right) \mathsf{V}^{(1)\,\dagger} (\mathsf{L}^{\dagger})^{-1} \quad . \tag{13}$$

The detailed derivation is demonstrated in the Appendix.

Through this expression it becomes possible to display error estimates for each voxel (and each vector component) individually. Since a corresponding error estimate can also be computed for the new dynamical technique for estimating inverse solutions which we will introduce in this paper, another quantitative measure for the comparison of different techniques for estimating inverse solutions becomes available.

4 The dynamical case

4.1 Dynamical models of voxel currents

After discussing various aspects of the LORETA method, we shall now proceed to the formulation of the new approach for solving the dynamical inverse problem of EEG generation; it has already been mentioned that for this purpose an additional temporal smoothness constraint could be introduced into equation 5 [15, 6]. While this approach will be pursued and extended elsewhere [7], here the central concept of our approach will be a new adaptation of Kalman filtering to the case of spatiotemporal dynamics.

The Kalman filter provides the optimum tool for predicting, filtering and smoothing estimates of the state of dynamical systems which cannot be observed directly, but only through an observation equation containing observational noise [16]. As a presupposition for its application, both the equations governing the dynamics and the observation equation have to be known.

Since for the case of the dynamics of human brain no well substantiated models for the spatiotemporal dynamics are known yet, we are faced with the problem of estimating suitable models from data. Clearly this constitutes a research task of enormous complexity which reaches far beyond the scope of this paper, therefore we will only be able to explore the very first and simplest approximations to such models.

Having defined a spatial discretisation by using a finite set of voxels, it is advisable to formulate dynamical models also with temporal discretisation; for simplicity we shall regard the basic time unit of this discretisation as equal to the sampling rate of the EEG recording. The corresponding time points will be labeled by $t = 1, 2, 3, \ldots, N_t$.

In general form the dynamics of a set of N_v voxels may be described by nonlinear multivariate autoregressions given as

$$\mathbf{J}(t) = \mathcal{F}\big(\mathbf{J}(t-1), \mathbf{J}(t-2), \ldots, \mathbf{J}(t-p) \,|\, \boldsymbol{\vartheta}\big) + \boldsymbol{\eta}(t) \quad , \tag{14}$$

where p denotes the positive integer model order and $\boldsymbol{\eta}(t)$ denotes dynamical noise, which we assume to be white and Gaussian with zero mean and covariance matrix $\mathsf{C}_\eta = E(\boldsymbol{\eta}\boldsymbol{\eta}^\dagger)$.

When fitting such models to given data, $\boldsymbol{\eta}(t)$ represents a time series of innovations, i.e. components of the data which cannot be explained from the dynamics itself. It is the aim of modelling to find dynamical models which produce a Gaussian white innovation time series, such that the process of modelling can be regarded as "temporal whitening".

$\mathcal{F}$ denotes a function describing the deterministic part of the dynamics; it may depend on a vector of parameters $\boldsymbol{\vartheta}$. This function may contain considerable internal complexity and a huge number of parameters (described by $\boldsymbol{\vartheta}$), since it maps an input space of dimensionality $3N_v p$ to an output space of dimensionality $3N_v$.

The simplest non-trivial example of the class of autoregressions described by equation 14 is a linear multivariate autoregressive model of first order $\big(\mathrm{AR}(1)\big)$:

$$\mathbf{J}(t) = \mathsf{A}\,\mathbf{J}(t-1) + \boldsymbol{\eta}(t) \quad , \tag{15}$$

and here we note that the popular random-walk model is a special case of this model, with $\mathsf{A} = \mathsf{I}_{3N_v}$. The parameter matrix A is of size $(3N_v) \times (3N_v)$, which in our case is of the order of 10^8. This large number of parameters is still far too high to be estimated from real data, therefore we need additional reductions of model complexity. Also the practical application of Kalman filtering requires a simplified model structure. It is an arguably reasonable assumption that at short time scales the dynamics will be restricted to local neighbourhoods, i.e. each voxel will interact only with its direct spatial neighbours; in a rectangular grid of voxels there will 6 direct neighbours for each voxel, except for those at the boundaries of the gray-matter parts of brain. Most elements of A become zero by this proposition. The dynamical model *for each voxel* becomes

$$\mathbf{j}(v,t) = \mathsf{A}_1\,\mathbf{j}(v,t-1) + \frac{1}{6}\mathsf{B}_1 \sum_{v' \in \mathscr{N}(v)} \mathbf{j}(v',t-1) + \boldsymbol{\eta}(t) \quad , \tag{16}$$

where A_1 and B_1 now are the autoregressive parameter matrices (of size 3×3) for self-interaction and nearest-neighbour interaction, respectively, and $\mathscr{N}(v)$ denotes the set of labels of gray-matter voxels that are direct neighbours of voxel v. If furthermore we assume the absence of any interactions between the three projections of local current vectors (both within and between voxels) and total spatial homogeneity and isotropy for all pairs of neighbouring voxels (this assumption extending to the three vector projections as well), we can ultimately reduce the number of parameters to two. The dynamical model for each voxel becomes

$$\mathbf{j}(v,t) = a_1\,\mathsf{I}_3\,\mathbf{j}(v,t-1) + \frac{b_1}{6}\,\mathsf{I}_3 \sum_{v' \in \mathscr{N}(v)} \mathbf{j}(v',t-1) + \boldsymbol{\eta}(t) \quad , \tag{17}$$

where I_3 denotes the 3×3 identity matrix; now a_1 and b_1 are scalar autoregressive parameters. This model implements complete symmetry between voxels and also between projections of local currents; but clearly, this symmetry is not preserved by multiplication with the lead field matrix K. As a consequence of this, inverse solutions also will display non-symmetric behaviour with respect to voxels and projections.

The additional assumptions which are required in order to define the simplest possible non-trivial dynamical model, given by equation 17, are almost certainly without physiological justification, but nevertheless this oversimplification is necessary in order to design a tractable algorithm as a starting point. As the next step it becomes possible again to consider generalisations, such as higher model order, nonlinearities or inhomogeneities;

as a consequence it may become possible to describe more structure present in the data by the dynamical model and to relegate less power from the data to the time series of innovations $\boldsymbol{\eta}(t)$. Here the aim is that the values in $\boldsymbol{\eta}(t)$ be as small as possible, contain as little correlation as possible (i.e. being white noise) and have a distribution as close as possible to a Gaussian.

As long as only nearest-neighbour interactions are allowed and a high degree of spatial homogeneity and isotropy is maintained, the number of unknown parameters can be kept small, and the problem remains accessible for spatiotemporal Kalman filtering. By increasing the model order to p, it becomes possible to describe stochastic oscillations which may be present in the data; although this case represents a generalisation of equation 15, it can be incorporated into this equation by replacing the order-1 state vector $\mathbf{J}(t)$ by $\left(\mathbf{J}(t)^\dagger, \mathbf{J}(t-1)^\dagger, \ldots, \mathbf{J}(t-p+1)^\dagger\right)^\dagger$, such that formally the dynamics remains a linear autoregression of first order.

Equation 17 can alternatively be written as

$$\mathbf{j}(v,t) = (a_1 + b_1)\,\mathsf{I}_3\,\mathbf{j}(v,t-1) + b_1\mathsf{I}_3\left(\left(\frac{1}{6}\sum_{v'\in\mathscr{N}(v)}\mathbf{j}(v',t-1)\right) - \mathbf{j}(v,t-1)\right) + \boldsymbol{\eta}(t) \quad . \tag{18}$$

This equation shows more clearly that the dynamics at each voxel is composed of two contributions, the first representing the autoregressive dynamics of the voxel itself and the second representing small exogeneous disturbances which partly are described as pure noise and partly as the difference between the average of the states of the neighbouring voxels and the state of the voxel itself. The latter difference is the same as also used in the Hjorth source derivation [33], but here we employ it directly in the (3-dimensional) generator space (i.e. the voxel space) instead of the (2-dimensional) electrode space.

From equation 18 it can be seen that our model corresponds to a specific choice for the parameter matrix A which can be expressed as

$$\mathsf{A} = (a_1 + b_1)\,\mathsf{I}_{3N_v} - b_1\mathsf{L} \quad , \tag{19}$$

where L is the discrete spatial Laplacian operator defined in equation 4. This relation will become useful in the next section.

Finally we would like to mention that there exist the possibility of a direct generalisation of equation 18 to partial differential equation models for the description of brain dynamics in continuous time and space; recently such models have been explored by various authors [34, 35]. In the continuous case (with respect to both time and space) the second term on the rhs of equation 18 (i.e. the term corresponding to the Laplacian) becomes a second derivative with respect to space, while the two terms $\mathbf{j}(v,t) - (a_1 + b_1)\,\mathsf{I}_3\,\mathbf{j}(v,t-1)$ can be interpreted as a first derivative with respect to time; consequently, a standard diffusion equation results. If a model of order $p = 2$ is chosen, there will also be a second derivative w.r.t. time instead of the first derivative; in this case a standard wave equation results. These interpretations render the model orders $p = 1$ and $p = 2$ particularly attractive.

4.2 Spatial whitening

Application of Kalman filtering to the full spatiotemporal model as given by equation 15 would be infeasible in terms of computational time and memory demands due to the huge

size of the parameter matrix A, i.e. if interactions between all pairs of voxels have to be considered. In the preceding section we have suggested to decompose this high-dimensional filtering problem into a collection of coupled low-dimensional local filtering problems, each centred at one individual voxel, as described by equation 16; only by this decomposition the spatiotemporal filtering problem becomes tractable. The low-dimensional systems remain coupled through neighbourhood interactions, but at each voxel these contributions are formally regarded as exogeneous variables. This decomposition approach is a central contribution of this paper.

In order to apply this decomposition to the dynamics it is also necessary that the dynamical noise covariance matrix C_η be a diagonal matrix, as assumed in equation 3 for the case of the observational noise covariance matrix C_ϵ. But in the case of C_η experience obtained from the analysis of real EEG time series has shown that such assumption will typically not be justified, rather the presence of non-vanishing instantaneous correlations at least between *neighbouring* voxels – and therefore also between the corresponding components of the dynamical noise – has to be expected.

Therefore we need an instantaneous data transformation

$$\tilde{\mathbf{J}} = \mathsf{T}\mathbf{J} \quad , \tag{20}$$

such that in the corresponding dynamical model

$$\tilde{\mathbf{J}}(t) = \tilde{\mathsf{A}}\,\tilde{\mathbf{J}}(t-1) + \tilde{\boldsymbol{\eta}}(t) \quad . \tag{21}$$

the dynamical noise covariance matrix $\mathsf{C}_{\tilde{\eta}}$ becomes diagonal (here we are furthermore assuming that all elements on the diagonal of $\mathsf{C}_{\tilde{\eta}}$ are identical, a simplification which due to the large number of voxels is necessary for practical implementation; again, this assumption may be relaxed in future work):

$$\mathsf{C}_{\tilde{\eta}} = \sigma_{\tilde{\eta}}^2\,\mathsf{I}_{3N_v} \quad . \tag{22}$$

In order to find a simple but efficient transformation we propose to extend the concept of temporal whitening to the spatial domain. A simple whitening approach in temporal domain is given by differentiating the time series; we can perform a spatial differentiating step (of second order) by applying the discrete Laplacian as defined by equation 4 to the dynamical state $\mathbf{J}$. Similar ideas for spatial whitening in the context of dynamical inverse problems have also been proposed by Baroudi *et al.* [20].

It is reasonable to assume that very fast correlations, appearing to be instantaneous with respect to the sampling time of the data, will be confined to short distances, i.e. neighbouring voxels; the Laplacian represents the easiest possible choice for removing these correlations. We note that the same transformation is employed in LORETA, but for a different purpose; nevertheless this coincidence provides useful additional interpretations, as will be shown now.

If we choose $\mathsf{T} = \mathsf{L}$, equation 21 yields

$$\mathbf{J}(t) = \mathsf{L}^{-1}\tilde{\mathsf{A}}\mathsf{L}\,\mathbf{J}(t-1) + \mathsf{L}^{-1}\tilde{\boldsymbol{\eta}}(t) \quad . \tag{23}$$

But due to equation 19 (which by definition also describes the structure of $\tilde{\mathsf{A}}$) the matrices $\tilde{\mathsf{A}}$ and L commute, such that by comparison with equation 15 we find that the choice $\mathsf{T} = \mathsf{L}$ corresponds to $\boldsymbol{\eta}(t) = \mathsf{L}^{-1}\tilde{\boldsymbol{\eta}}(t)$, and our assumption for the non-diagonal dynamical noise covariance matrix becomes

$$\mathsf{C}_\eta = \mathsf{L}^{-1}\,E(\tilde{\boldsymbol{\eta}}\tilde{\boldsymbol{\eta}}^\dagger)(\mathsf{L}^{-1})^\dagger = \sigma_{\tilde{\eta}}^2\,(\mathsf{L}^\dagger\mathsf{L})^{-1} \quad . \tag{24}$$

By comparison with LORETA in its Bayesian interpretation (see section 3.1), it can be seen, that $\sigma_{\tilde{\eta}}$ directly corresponds to the hyperparameter $\tau = \sigma_\epsilon/\lambda$, such that τ^2 can be interpreted as dynamical noise variance. This quantity is meaningful also in the case of the instantaneous inverse problem and its LORETA solution; the optimal regularisation parameter λ becomes a measure for the ratio between the observation noise variance and the dynamical noise variance, as it should be.

If other dynamical models than described by equation 17 are assumed, $\tilde{\mathsf{A}}$ and L will generally not commute. Other transformations than the plain Laplacian L may then be needed for perfect spatial whitening. But even in this case L can be expected to make $\mathsf{C}_{\tilde{\eta}}$ "more diagonal" (i.e. reduce the size of the non-diagonal elements, as compared to the diagonal elements), and for this reason we will continue to employ it as spatial whitening transformation.

We remark that the appropriate choice of $\mathsf{C}_{\tilde{\eta}}$ forms an important part of the fitting of a dynamical model; therefore refinements of our choice should be based on improvements of a suitable statistical criterion, such as AIC (see section 4.7). The Laplacian serves only as a first approximation, but in future research it should be explored, how $\mathsf{C}_{\tilde{\eta}}$ can be adapted further to specific data sets.

For the actual application of spatiotemporal Kalman filtering we will exclusively express the dynamics as $\tilde{\mathbf{J}}(t)$, i.e. using the spatially whitened version. From now on we will omit the tilde.

4.3 Spatiotemporal Kalman filtering

Given the observation equation (equation 1) and the dynamical equation (in the case of a linear first-order autoregression equation 21) we could in principle apply Kalman filtering according to its usual form; however, due to the very high dimensionality of the state variable $\mathbf{J}$ this would be infeasible. But by appropriate design of certain modifications of the standard filtering procedure it is possible to decompose the very high-dimensional filtering problem into a collection of coupled low-dimensional problems; the set of these problems is labeled by the voxel label v, i.e. it represents the spatial dimension of the problem. These modifications of the standard Kalman filter procedure are not trivial, and we will defer a detailed discussion to a later paper. Here we will only describe the main points from the viewpoint of practical implementation.

Let $\hat{\mathbf{j}}(v, t-1 \,|\, t-1)$ denote the estimate of the local current vector at voxel v at time $t-1$, i.e. the local state estimate, and $\mathsf{p}(v, t-1 \,|\, t-1)$ the corresponding estimate of the local error covariance matrix (a 3×3 matrix). The notation $\hat{\mathbf{j}}(t_1 \,|\, t_2)$ (where $t_1 \geqslant t_2$) represents the estimate of $\mathbf{j}$ at time t_1 which is based on all information having become available until (and including) time t_2. For each voxel the local state prediction is then given by

$$\hat{\mathbf{j}}(v, t \,|\, t-1) = \mathsf{A}_1 \, \hat{\mathbf{j}}(v, t-1 \,|\, t-1) + \frac{1}{6} \mathsf{B}_1 \sum_{v' \in \mathscr{N}(v)} \hat{\mathbf{j}}(v', t-1 \,|\, t-1) \quad , \tag{25}$$

and the corresponding local prediction error covariance matrix can be approximated by

$$\mathsf{p}(v, t \,|\, t-1) = \mathsf{A}_1 \, \mathsf{p}(v, t-1 \,|\, t-1) \, \mathsf{A}_1^\dagger + \sigma_\eta^2 \mathsf{I}_3 \quad . \tag{26}$$

Here we have assumed that the second term on the rhs of equation 25 behaves like an exogeneous variable, i.e. without contributing significantly to the prediction error covariance;

alternatively we may say that the covariance contribution from the neighbours is included in the term $\sigma_\eta^2 \mathsf{I}_3$, which can be done easily, since σ_η^2 results from numerical optimisation for given data and filter implementation (see section 4.7).

We remark that this approximation is not crucial for our approach to spatiotemporal Kalman filtering, and a more explicit expression for the prediction error covariance could be employed; however, for typical systems we expect that usually the second term on the rhs of equation 25 (i.e. the neighbourhood contribution) is considerably smaller than the first term (the local contribution), and therefore the resulting error should be only minor, whereas the computational expenses can be reduced noticeably by refraining from evaluating the more explicit expression at each voxel.

The local state predictions $\hat{\mathbf{j}}(v, t \,|\, t-1)$ for all voxels form the overall state prediction $\hat{\mathbf{J}}(t \,|\, t-1)$, from which the observation prediction (for all electrodes) follows as

$$\hat{\mathbf{Y}}(t \,|\, t-1) = \overline{\mathsf{K}}\, \hat{\mathbf{J}}(t \,|\, t-1) \quad . \tag{27}$$

The symbol $\overline{\mathsf{K}}$ stands for the product $\mathsf{K}\,\mathsf{L}^{-1}$; multiplication by the inverse of the Laplacian is needed due to the spatial whitening approach described in section 4.2. The actual observation at time t is $\mathbf{Y}(t)$, and the observation prediction error results as

$$\Delta\mathbf{Y}(t) = \mathbf{Y}(t) - \hat{\mathbf{Y}}(t \,|\, t-1) \quad . \tag{28}$$

We note that this multivariate time series represents the innovations of the actual observations, as opposed to the innovations of the (unobservable) system states, which have been denoted by $\boldsymbol{\eta}(t)$ in equation 14. The corresponding observation prediction error covariance matrix can be approximated by

$$\mathsf{R}(t \,|\, t-1) = \sum_{v=1}^{N_v} \overline{\mathsf{K}}(v) \mathsf{p}(v, t \,|\, t-1) \overline{\mathsf{K}}(v)^\dagger + \sigma_\epsilon^2 \mathsf{I}_{n_c} \quad . \tag{29}$$

Here the direct summation over voxels seems to provide the appropriate generalisation of the standard expression to the spatiotemporal case. The matrices $\overline{\mathsf{K}}(v)$ have been defined in section 2; again the bar refers to the fact that due to spatial whitening we have to replace K by $\mathsf{K}\,\mathsf{L}^{-1}$, before extracting the columns refering to voxel v. The Kalman gain matrix for voxel v follows as

$$\mathsf{g}(v, t) = \mathsf{p}(v, t \,|\, t-1)\, \overline{\mathsf{K}}(v)^\dagger\, \mathsf{R}(t \,|\, t-1)^{-1} \quad , \tag{30}$$

and finally the local state estimation and the corresponding local estimation error covariance matrix are given by

$$\hat{\mathbf{j}}(v, t \,|\, t) = \hat{\mathbf{j}}(v, t \,|\, t-1) + \mathsf{g}(v, t)\, \Delta\mathbf{Y}(t) \tag{31}$$

and

$$\mathsf{p}(v, t \,|\, t) = \big(\mathsf{I}_3 - \mathsf{g}(v, t) \overline{\mathsf{K}}(v)\big)\, \mathsf{p}(v, t \,|\, t-1) \quad , \tag{32}$$

respectively. For equations 29 and 32 again we make use of similar approximations as also employed in the case of equation 26.

This set of equations constitutes the new spatiotemporal Kalman filter. It should be stressed that equations 25, 26, 30, 31 and 32 are applied locally to each voxel, whereas

only equation 27 requires a large-scale multiplication of the lead-field matrix $\overline{\mathsf{K}}$ with the full $(3N_v)$-dimensional state vector $\hat{\mathbf{J}}$.

For practical application of this filter to time series data initial values $\mathbf{j}(v,1|1)$ and $\mathsf{p}(v,1|1)$ are needed. As initial state estimates $\mathbf{j}(v,1|1)$ we propose to use solutions provided by the LORETA algorithm, as presented in section 3.1; the hyperparameters are chosen by minimisation of ABIC. For this purpose we have to sacrifice the first p data points and exclude them from the Kalman filtering process. As a further refinement step, these initial values are improved by likelihood maximisation [36] in the same way as also the parameters are estimated; this will be discussed in section 4.7. Since it is impracticable to perform this optimisation directly in the $(3N_v)$-dimensional state space, we apply the optimisation to the $(n_c\,p)$-dimensional observation space of the first p observations $\mathbf{Y}(1),\ldots,\mathbf{Y}(p)$ and use the LORETA technique for mapping points in the observation space to new prospective initial states. Concerning the choice and optimisation of initial values, our implementation differs from more conventional applications of Kalman filtering, which usually do not employ this extended estimation approach.

According to our experience, the choice of initial values for $\mathsf{p}(v,1|1)$ is not critical; unity matrices can be used.

4.4 Estimation of the covariance matrix of the estimated state vector

Equation 32 provides us with covariance estimates for the reconstructed states (given by equation 31); we may use these estimates as a measure for the error of each component of the estimated state vectors. Due to the spatial whitening transformation (see section 4.2) the covariance matrix of the actual currents is given by

$$\mathsf{C}^{(dIS)}_{\mathsf{L}^{-1}\hat{\mathbf{J}}} = \mathsf{L}^{-1}\mathsf{P}(t\,|\,t)(\mathsf{L}^{\dagger})^{-1} \quad , \tag{33}$$

where the superscript (dIS) denotes "dynamical inverse solution". By $\mathsf{P}(t\,|\,t)$ the complete covariance matrix of the estimated states is denoted, i.e. a $3N_v \times 3N_v$ matrix the diagonal of which consists of the $\mathsf{p}(v,t\,|\,t)$ matrices given by equation 32 for each voxel. The remaining elements of $\mathsf{P}(t\,|\,t)$ are filled with zeros, according to our decomposition approach for this high-dimensional filtering problem.

The diagonal elements of $\mathsf{C}^{(dIS)}_{\mathsf{L}^{-1}\hat{\mathbf{J}}}$ provide time-dependent variances for each element of the complete estimated state vector $\mathsf{L}^{-1}\hat{\mathbf{J}}$ (again the pre-multiplication with L^{-1} is necessary due to spatial whitening). As also in the case of the LORETA inverse solutions we will present variances of the modulus $\hat{j}(v,t\,|\,t)$ of local vectors $\hat{\mathbf{j}}(v,t\,|\,t)$ (see equation A6 in the appendix for the appropriate expression from Gaussian error propagation).

4.5 Silent and observable sources

In the (instantaneous) inverse problem of EEG generation we are facing the apparently hopeless task to estimate a set of more than 10^4 unknown quantities from usually less than 10^2 observations. As a condition for identifying a unique solution additional constraints have to be defined and imposed, such as the smoothness constraint of LORETA. Such constraints may succeed to render the inverse problem well-posed, but at the cost of a lack of justification, e.g. in terms of physiology.

The inherent predicament of inverse problems, i.e. the very unfavourable ratio between the numbers of known and of unknown quantities, is avoided, but not solved by such

constraints. The central question remains unsolved: How can we expect that all relevant information about the spatially distributed primary current density could be reconstructed from a small number of surface measurements? In this section we will try to outline an answer to this question for the case of the dynamical inverse problem.

The problem can be expressed in quantitative form by considering the observation equation (equation 1) and the singular value decomposition of the lead field matrix (equation 8):

$$\mathbf{Y}(t) = \overline{\mathsf{K}}\,\mathbf{J}(t) + \boldsymbol{\epsilon}(t) = \mathsf{USV}^{\dagger}\,\mathbf{J}(t) + \boldsymbol{\epsilon}(t) = \mathsf{US}\,\mathbf{H}(t) + \boldsymbol{\epsilon}(t) \quad . \tag{34}$$

Here we have defined $\mathbf{H}(t) = \mathsf{V}^{\dagger}\mathbf{J}(t)$, i.e. we have applied an orthogonal transform to the state vector $\mathbf{J}(t)$. As mentioned already in section 3.1, the $n_c \times 3N_v$ matrix S is composed of a $n_c \times n_c$ diagonal matrix containing the singular values (which are all non-zero since the lead field matrix has full rank), while the remaining $(3N_v - n_c)$ columns contain only zeros. Therefore only the subspace spanned by the first n_c elements of the transformed state vector $\mathbf{H}(t)$ is mapped to the observation vector $\mathbf{Y}(t)$, while the subspace of the remaining elements is completely ignored. Components of the "true" state $\mathbf{J}(t)$ which by the orthogonal transformation are mapped into the former subspace, represent "observable sources", whereas those components which are mapped completely into this latter subspace cannot be observed, whence they are termed "silent sources" (note that such "sources" are not localised in physical space). Within the framework of the instantaneous inverse problem there exists no way to obtain information about these state components.

Here we would like to argue that the situation is much different in the case of the dynamical inverse problem. If the problem of estimating unobserved quantities is treated in a state space framework, it can be easily shown that under certain circumstances information from the subspace of silent sources will propagate into the subspace of observable sources.

In order to demonstrate this effect, we now apply the orthogonal transformation $\mathbf{H}(t) = \mathsf{V}^{\dagger}\mathbf{J}(t)$ to equation 15 and obtain

$$\mathbf{H}(t) = \mathsf{V}^{\dagger}\mathsf{AV}\,\mathbf{H}(t-1) + \mathsf{V}^{\dagger}\boldsymbol{\eta}(t) = \breve{\mathsf{A}}\mathbf{H}(t-1) + \breve{\boldsymbol{\eta}}(t) \quad , \tag{35}$$

where $\breve{\mathsf{A}}$ and $\breve{\boldsymbol{\eta}}(t)$ denote the transformed transition matrix and the transformed dynamical noise vector, respectively. Let $\mathbf{H}(t) = \left(\mathbf{H}^{(1)}(t)^{\dagger}\ \ \mathbf{H}^{(2)}(t)^{\dagger}\right)^{\dagger}$, such that $\mathbf{H}^{(1)}(t)$ denotes the first n_c elements of the vector $\mathbf{H}(t)$ and $\mathbf{H}^{(2)}(t)$ the remaining elements. Then $\mathbf{H}^{(1)}(t)$ represents the subspace which is mapped to the observations, i.e. the subspace of observable sources, while $\mathbf{H}^{(2)}(t)$ represents the subspace of silent sources. The same subdivision is also applied to the transformed transition matrix $\breve{\mathsf{A}}$:

$$\begin{pmatrix} \mathbf{H}^{(1)}(t) \\ \mathbf{H}^{(2)}(t) \end{pmatrix} = \begin{pmatrix} \breve{\mathsf{A}}^{(1,1)} & \breve{\mathsf{A}}^{(1,2)} \\ \breve{\mathsf{A}}^{(2,1)} & \breve{\mathsf{A}}^{(2,2)} \end{pmatrix} \begin{pmatrix} \mathbf{H}^{(1)}(t-1) \\ \mathbf{H}^{(2)}(t-1) \end{pmatrix} + \breve{\boldsymbol{\eta}}(t) \quad , \tag{36}$$

such that $\breve{\mathsf{A}}^{(1,1)}$ and $\breve{\mathsf{A}}^{(2,2)}$ are $n_c \times n_c$ and $(3N_v - n_c) \times (3N_v - n_c)$ matrices, respectively. From this equation it can be seen that the "partial" transition matrix $\breve{\mathsf{A}}^{(1,2)}$ plays a crucial role, since it maps the subspace of silent sources to the subspace of observable sources. It is through this pathway that information from the silent sources is propagated into the observable sources, and from there into the observations.

Obviously this propagation of information will not take place if $\breve{\mathsf{A}}^{(1,2)} = \mathsf{0}$. On the contrary, in order to have information from all elements of $\mathbf{H}^{(2)}(t)$ be propagated into

$\mathbf{H}^{(1)}(t)$ each column of $\breve{\mathsf{A}}^{(1,2)}$ must contain at least one non-zero element. Now consider the special case of $\mathsf{A} = \mathsf{I}_{3N_v}$, i.e. a random-walk type dynamics without neighbour interactions. Then $\breve{\mathsf{A}} = \mathsf{V}^\dagger\mathsf{A}\mathsf{V} = \mathsf{V}^\dagger\mathsf{V} = \mathsf{I}_{3N_v}$ and consequently $\breve{\mathsf{A}}^{(1,2)} = \mathsf{0}$. This argument may seem elementary and straightforward, but given the number of applications of Kalman filtering which still apply random-walk-type dynamical models without any specific justification, it may nevertheless deserve more attention.

If, on the other hand, we choose a transition matrix with non-diagonal elements, such as given by equation 19, $\mathsf{V}^\dagger\mathsf{A}\mathsf{V}$ will not be diagonal and $\breve{\mathsf{A}}^{(1,2)} \neq \mathsf{0}$. Using the partition of V into $\mathsf{V}^{(1)}$ and $\mathsf{V}^{(2)}$, as defined in section 3.1, it can be seen that

$$\breve{\mathsf{A}}^{(1,2)} = \left(\mathsf{V}^{(1)^\dagger}\mathsf{A}\right)\mathsf{V}^{(2)} \quad . \tag{37}$$

Due to the orthogonality of V we have $\mathsf{V}^{(1)^\dagger}\mathsf{V}^{(2)} = \mathsf{0}$, i.e. the columns of $\mathsf{V}^{(1)}$ are orthogonal to the columns of $\mathsf{V}^{(2)}$; but multiplication by a non-diagonal matrix A will replace the columns of $\mathsf{V}^{(1)}$ by a set of n_c different columns which generically are no longer orthogonal to any of the columns of $\mathsf{V}^{(2)}$. Therefore we presume that *generically* all elements of $\breve{\mathsf{A}}^{(1,2)}$ will be non-zero. Consequently we expect that there will be a flow of information from all elements of $\mathbf{H}^{(2)}(t)$ into $\mathbf{H}^{(1)}(t)$.

This argument shows that only by using a dynamical model including non-vanishing neighbour interactions, state components belonging to the subspace of silent sources become accessible for reconstruction by the spatiotemporal Kalman filter.

4.6 Observability in state space models of brain dynamics

In the previous section we have given a heuristic derivation of the mechanism by which the spatiotemporal Kalman filtering approach is capable of accessing information about unobservable state components; now we would like to mention that there exists a rigorous theory addressing the question of whether for a given model of the dynamics and the observation the unobserved quantities can be reconstructed. This theory is built around the central concepts of observability and controllability [17, 18].

Assume that we are dealing with a dynamical system which evolves according to linear dynamics as described by equation 15 (with a constant transition matrix A), and that we are observing this system through an observation equation like equation 1 (with a constant observation matrix $\overline{\mathsf{K}}$). If it is possible to reconstruct the true states of the system from the observations, the pair $(\mathsf{A}, \overline{\mathsf{K}})$ is said to be "observable".

Various tests for observability of dynamical systems have been suggested. A well-known test states that the pair $(\mathsf{A}, \overline{\mathsf{K}})$ is observable, if and only if the observability matrix $\mathcal{O}$, being defined by

$$\mathcal{O} = \left[\overline{\mathsf{K}}^\dagger \;\; \mathsf{A}^\dagger\overline{\mathsf{K}}^\dagger \;\; (\mathsf{A}^\dagger)^2\overline{\mathsf{K}}^\dagger \;\; \ldots \;\; (\mathsf{A}^\dagger)^{3N_v-1}\overline{\mathsf{K}}^\dagger\right]^\dagger \quad , \tag{38}$$

has full rank, $\mathrm{rank}(\mathcal{O}) = 3N_v$ [18]. Here $3N_v$ denotes the dimension of the state vector, i.e. the number of unknown quantities. In the case of the dynamical inverse problem of EEG generation this matrix has the size $3N_v n_c \times 3N_v = 185382 \times 10299$ (when using the corresponding values of the spatial discretisation as employed in this paper), which is by far too large for numerical calculation of the rank. Kalman filtering and observability theory are usually not applied to problems of this size. For this reason, we are currently not yet able to present a rigorous proof of observability for our algorithm.

On the other hand, observability constitutes the essential precondition for the reconstruction of the unobserved states by Kalman filtering, i.e. for the identification of a unique inverse solution. In the remainder of this paper we will demonstrate by application to simulated and to real EEG data that Kalman filtering can be applied successfully to the dynamical inverse problem of EEG generation. For this reason we presume that, effectively, observability of the pair $(\mathsf{A}, \overline{\mathsf{K}})$ is given. Future research may succeed in rigorously proving observability.

4.7 Parameter estimation

The general autoregressive model described by equation 14 depends on a parameter vector $\boldsymbol{\vartheta}$; in the largely simplified model given by equation 17 we have $\boldsymbol{\vartheta} = (a_1, b_1)$, and autoregressive models of higher order have $\boldsymbol{\vartheta} = (a_1, \ldots, a_p, b_1, \ldots, b_q)$; here we are permitting the possibility of choosing $p \neq q$, i.e. choosing different autoregressive model orders for self-interaction and nearest-neighbour interaction.

Usually there will be no detailed prior knowledge about appropriate values for these parameters available. Furthermore we need estimates for the variances σ_ϵ^2 and σ_η^2, as defined by equations 3 and 24.

Estimates for these dynamical parameters and variances should be obtained preferably from actual data. This can be accomplished within the framework of spatiotemporal Kalman filtering by likelihood maximisation. So far, to the best of our knowledge, no successful applications of the principle of likelihood maximisation to the field of inverse problems have been reported; recently, Phillips *et al.* [5] have presented an approach involving restricted maximum likelihood, but their approach does not involve dynamical modelling.

Assume that an EEG time series $\mathbf{Y}(t)$ is given, where $t = 1, 2, \ldots, N_t$. At each time point the Kalman filter provides an observation prediction $\hat{\mathbf{Y}}(t)$, given by equation 27, and hence also observation innovations $\Delta\mathbf{Y}(t)$; if for these a multivariate Gaussian distribution with mean $\hat{\mathbf{Y}}(t)$ and covariance matrix $\mathsf{R}(t \,|\, t-1)$ is assumed, the logarithm of the likelihood (i.e. log-likelihood) immediately results as

$$\log L(\boldsymbol{\vartheta}, \sigma_\epsilon^2, \sigma_\eta^2) = -\frac{1}{2} \sum_{t=1}^{N_t} \Big(\log |\mathsf{R}(t \,|\, t-1)| + \Delta\mathbf{Y}(t)^\dagger \, \mathsf{R}(t \,|\, t-1)^{-1} \, \Delta\mathbf{Y}(t) + n_c \log(2\pi) \Big). \quad (39)$$

Here $|.|$ denotes matrix determinant. The log-likelihood is known to be a biased estimator of the expectation of Boltzmann entropy [31]; only a small further step is needed for the calculation of an improved unbiased estimator of (-2) times Boltzmann entropy, the well-known Akaike Information Criterion (AIC) [37]:

$$\mathrm{AIC}(\boldsymbol{\vartheta}, \sigma_\epsilon^2, \sigma_\eta^2) = -2 \log L(\boldsymbol{\vartheta}, \sigma_\epsilon^2, \sigma_\eta^2) + 2\big(\dim(\boldsymbol{\vartheta}) + 2\big) \quad , \quad (40)$$

where $\dim(\boldsymbol{\vartheta})$ denotes the number of parameters contained in the parameter vector $\boldsymbol{\vartheta}$; it is further increased by 2 due to the need to fit σ_ϵ^2 and σ_η^2 from the data.

The AIC can also be interpreted as an estimate of the distance between the estimated model and the unknown true model; the true model will remain unknown, but by comparison of the AIC values for different estimated models it still becomes possible to find the best model. Consequently, a well justified and efficient tool for obtaining estimates for unknown parameters consists of minimising the AIC; also the effect of changing the

structure of the model itself can be evaluated by monitoring the resulting change of the AIC.

As already mentioned in section 4.3, the same approach is also applied to the estimation of improved initial values for the state, $\hat{\mathbf{J}}(1|1)$, to be used by the Kalman filter.

Furthermore it is also possible to compare instantaneous inverse solutions obtained by LORETA with dynamical inverse solutions obtained by spatiotemporal Kalman filtering by directly comparing the corresponding values of ABIC (given by equations 10 and 11) and AIC. Theoretical support for this direct comparison of ordinary likelihood and type-II likelihood in modelling has been provided by Jiang [38].

5 Application to simulated EEG

5.1 Spatial discretisation

This study employs a discretisation of brain into voxels which is based on a grid of $27 \times 23 \times 27$ voxels (sagittal $\times$ axial $\times$ coronal). Out of these 16767 grid positions 8723 represent voxels actually covering the brain (and part of the surrounding tissue), out of which 3433 are regarded as gray-matter voxels belonging to the cortex; deeper brain structures, like thalamus, are not considered in this study. For the underlying brain geometry and the identification of the cortical gray-matter voxels an averaged brain model was used, which was derived from the average Probabilistic MRI Atlas produced by the Montreal Neurological Institute [39]. More details on this model can be found in [40] and references cited therein.

5.2 Design of simulation

We shall now present some results of applying the spatiotemporal Kalman filter, as presented in sections 4.3 and 4.7, to time series data. It is a well-known problem of all algorithms providing inverse solutions that it is difficult to perform meaningful evaluations of the results and the performance, since for such evaluation we would need to know the true sources.

Inverse solutions obtained from real EEG time series typically display fluctuating spatiotemporal structures, but it is usually not possible to ascertain *a posteriori* to which extent these structures describe the true brain dynamics which was present during the recording. Relative comparisons between inverse solutions can be performed by comparing the corresponding values of AIC (see equation 40), providing us with an objective criterion for model selection. For the purpose of evaluation of algorithms, it is furthermore possible to employ simulated data. If the primary currents at the gray-matter voxel sites are simulated, the corresponding EEG observations can be computed simply by multiplication with the lead field matrix, and consequently both the EEG time series and its true sources are known. But clearly these "true" sources will not represent any realistic brain dynamics, and also the underlying models for the brain and the observation will contain severe simplifications and inaccuracies. Nevertheless, for the purpose of demonstrating feasibility and potential usefulness of the proposed algorithm, and for comparing inverse solutions obtained by different algorithms, we will now design a very simple simulated brain dynamics; results for real EEG data will be shown in section 6.

A typical phenomenon of human brain dynamics is the presence of strong oscillations within local neighbourhoods, e.g. alpha activity in the visual cortex. If we regard the

"simple" dynamical model described by equation 17, where the parameters a_1 and b_1 are constant, as a device to generate simulated brain dynamics, driven by Gaussian white noise, we find that it will not produce such oscillations. In order to have oscillating behaviour in linear autoregressive models, a model order of at least $p = 2$ is needed. Alternatively the desired oscillation can be generated separately and imposed onto the brain dynamics through modulation of the system parameters. We will make use of this second alternative now.

By considering equations 15 and 19 it can be seen that the stability condition for the dynamical model described by equation 17 is approximately given by

$$|a_1 + b_1| < 1 \quad . \tag{41}$$

If we choose to keep b_1 constant and let a_1 depend explicitly on time by

$$a_1(t) = a_c\big(1 + a_s \sin(2\pi f t)\big) \tag{42}$$

and choose the parameters b_1, a_c and a_s such that a_1+b_1 will repeatedly become larger than unity, we have defined a transiently instable system. If this modulation of the parameter a_1 is confined only to those gray-matter voxels within a limited area of brain, this area will become a source of oscillations which spread out into neighbouring voxels. In this simulation we do not add dynamical noise, i.e. we are employing a linear, deterministic, explicitly time-dependent model. Alternatively the periodicity could also be generated by introducing additional state variables.

We define two areas in brain as centres for the generation of alpha-style oscillations, one in frontal brain and one in occipital brain. Each area is spherical and contains about 100 voxels; despite using equal radii the number is slightly different in both areas due to different content of non-gray-matter voxels. We choose the parameters a_c, a_s, b_1 and f differently for both areas: The occipital oscillation has $f = 10.65$ Hz, and the frontal oscillation has $f = 8.05$ Hz (assuming a sampling rate of 256 Hz). Careful choice of these parameters is necessary in order to obtain at least approximate global stability of the simulated dynamics. We choose for the occipital oscillation $a_c = 0.7$, $a_s = 0.75$ and $b_1 = 0.3$, and for the frontal oscillation $a_c = 0.9$, $a_s = 0.5$ and $b_1 = 0.1$. In this simulation the orientation of all vectors is the y-direction (which is the vertical direction according to the usual biometrical coordinate system for the human head); although the length of the vectors changes with time, their direction remains constant, and furthermore inversions of direction never occur. When simulating the system, an initial transient is discarded, and a multivariate time series of $N_t = 512$ points length is recorded. It represents the spatiotemporal dynamics of this simulation, i.e. for each of the $N_v = 3433$ voxels a 3-variate time series for the local current vector is recorded.

By multiplication by the lead field matrix according to equation 1 we create artificial EEG recordings from this simulated dynamics; we assume a standard recording according to the 10-20 system, average reference and a sampling rate of 256 Hz, i.e. with a length of the simulated time series of $N_t = 512$ two seconds of EEG can be represented. A small amount of Gaussian white observation noise is added to the pure EEG data (signal-to-noise ratio 100:1 in terms of standard deviations). The resulting EEG time series are shown in figure 1. As can be seen, they clearly display the two oscillations with their different frequencies, but in a quite blurred fashion. Note that due to the uniform vertical orientation of all current vectors in both hemispheres also the EEG displays complete symmetry with respect to the left (corresponding electrodes have labels with odd numbers) and the right hemisphere (even numbers). Out of the three electrodes on

the border between hemispheres, FZ, CZ and PZ, the latter has been omitted due to usage of average reference (see section 2).

The design of this simulation still contains many unrealistic elements and simplifications even beyond the intrinsic limitations of our dynamical model class, e.g. by omitting dynamical noise and assuming a uniform direction of all local vectors; also the assumption of Gaussian white observation noise may be questionable. In future work we intend to design more realistic simulations of brain dynamics and to present detailed results on the performance of different estimators of inverse solutions with respect to various design parameters.

5.3 Calculation of inverse solutions

For the EEG data shown in figure 1 we compute three inverse solutions: A "timeframe-by-timeframe" instantaneous inverse solution (using regularised LORETA), which shall be abbreviated as "*iIS*"; a dynamical inverse solution by using the spatiotemporal Kalman filter, as described in sections 4.3 and 4.7, employing the simplest possible dynamical model, as given by equation 17, which shall be abbreviated as "*dISs*"; and a dynamical inverse solution using the spatiotemporal Kalman filter, but employing the correct dynamical model, i.e. a "perfect" model, which shall be abbreviated as "*dISp*".

For the application of the Kalman filter in the case of the simplest model, four parameters $(a_1, b_1, \sigma_\epsilon^2, \sigma_\eta^2)$ have to be chosen according to the principle of maximum likelihood, i.e. by minimising equation 39. This optimisation poses no particular problems, apart from the usual problems related to nonlinear optimisation, such as local minima and high computational time consumption; it even turns out that the likelihood as a function of these parameters behaves quite smoothly. During the first part of the parameter optimisation it is advisable to allow for a transient of the Kalman filter itself to die out, before the likelihood is evaluated; only after optimising the estimate of the initial state the possibility of transients can be neglected. In our numerical simulation study maximum likelihood estimation of the unknown parameters $(a_1, b_1, \sigma_\epsilon^2, \sigma_\eta^2)$ yields the values $(0.7875, 0.2182, 4.8255, 1.4879 \times 10^{-6})$. These values are to be compared with the correct values for a_c, a_s and b_1 given in the previous subsection. Since in this simulation the model was deterministic, the dynamical noise covariance σ_η^2 should be expected to be zero; but in this case the Kalman filter was given a wrong model for the dynamics, consequently deviations of the actual observations from the predictions are interpreted as the result partly of observation noise and partly of stochastic elements in the dynamics.

In this simulation the dynamical model is ignoring two important aspects of the true dynamics, namely the fact that the autoregressive parameter a_1 is behaving differently for different groups of voxels, and that it shows explicit dependence on time for two groups of voxels. The dynamical model given by equation 17 is very primitive, and therefore it provides almost no additional information which could be used for the purpose of estimating an improved inverse solution.

As a contrast to this, we are also providing the same spatiotemporal Kalman filter with perfect knowledge about the true dynamics, i.e. not only does the filter know the correct values of the parameters as used in generating the simulated data, but also the information about the explicit time-dependence of $a_1(t)$ for the two oscillating areas and the correct assignment of the voxels to these areas are given to the filter. Nevertheless, also in this case a maximum likelihood optimisation step is employed in order to obtain estimates for $(\sigma_\epsilon^2, \sigma_\eta^2)$. The results are $(0.156\,,\, 10^{-7})$; now the estimate for the observation noise covariance is considerably smaller, compared to the case of the simplest model, while

the dynamical noise covariance almost vanishes.

5.4 Comparison of inverse solutions

The three different inverse solutions which we have obtained are given as functions of space and time, $\hat{\mathbf{j}}(v,t)$. We can compare them with the true solution (i.e. the simulated dynamics) $\mathbf{j}(v,t)$ by forming a RMS error according to

$$E = \sqrt{\frac{1}{N_v N_t} \sum_v \sum_t \left(\hat{\mathbf{j}}(v,t) - \mathbf{j}(v,t)\right)^2} \quad . \tag{43}$$

This comparison yields (for observation noise with SNR=100:1) $E = 1.4391$ for *iIS*, $E = 1.3932$ for *dISs* and $E = 0.4813$ for *dISp*. These results indicate that, compared to *iIS*, *dISs* achieves only a small improvement, if any, but using *dISp*, i.e. knowing the perfect model, a much better estimation of the currents becomes possible.

We remark that these figures deteriorate only slowly, if the amount of observation noise is increased: for SNR=100:5 we find $E = 1.4487$ (*iIS*), $E = 1.4030$ (*dISs*) and $E = 0.5459$ (*dISp*); for SNR=100:10 we find $E = 1.4782$ (*iIS*), $E = 1.4332$ (*dISs*) and $E = 0.7151$ (*dISp*). A more detailed study of the influence of observation noise on inverse solutions will be the subject of future work.

Obviously, this type of evaluation is possible only in the case of numerical simulations. A direct comparison of the three inverse solution without knowledge of the true solution can be performed by comparing the corresponding values of AIC, as given by equation 40, or ABIC, as given by equations 10 and 11. For *iIS* we obtain ABIC=110042.1, for *dISs* AIC=75549.0 and for *dISp* AIC=61936.7 . The absolute values depend on the length of the data set ($N_t = 512$ in this case), therefore they are less relevant; but by comparison we see again that both dynamical inverse solutions, as provided by Kalman filtering, represent better explanations of the observations than the instantaneous inverse solution, and that *dISp* is superior to *dISs*, as should be expected.

In figures 2 and 3 we present some graphical illustrations of the inverse solutions obtained in this simulation. Figure 2 shows the spatial distribution of true currents and inverse solutions at a fixed moment in time by displaying the maximum-intensity projection of the absolute values of the local current vectors by a gray-scale coding. Directions of vectors are not shown, since they were not allowed to vary in the simulation; but a closer analysis confirms that for all three inverse solutions the directions of most current vectors are correct and constant with respect to time.

For each case coronal, axial and sagittal projections of the spatial distribution of currents are shown. In subfigures A1, A2 and A3 the true currents from the simulated dynamics are shown. The two centres of simulated alpha activity can be seen clearly; most other areas of the brain remain inactive. The frontal centre shows a certain tendency to produce two neighbouring maxima of activity.

Subfigures B1, B2, B3 show the estimated currents according to *iIS*. It can be seen that the locations of the two main centres of activity are correctly reconstructed, but these two centres are much less focussed than in subfigures A1, A2, A3, rather does the active area spread out over most parts of the brain; in particular we see spurious structure extending into the temporal lobes. This lack of spatial resolution is a characteristic artifact of the underdetermined situation given in this inverse problem; we remark that there exists a suggestion for generating more focal inverse solutions from an instantaneous method by iterative reweighting [41].

Subfigures C1, C2, C3 show the estimated currents according to *dISs*. These results resemble to some extent those obtained by *iIS* (subfigures B1, B2, B3). This similarity, which is also found in the temporal domain (as shown in figure 3), is remarkable, since these two inverse solutions were obtained by completely different approaches, penalised Least Square (which is the essence of the LORETA approach) in the case of *iIS* and spatiotemporal Kalman filtering in the case of *dISs*. But only the estimate of the initial state for the Kalman filter was generated from *iIS*. The same phenomenon is also found for inverse solutions obtained from real EEG data sets.

Subfigures D1, D2, D3 show the estimated currents according to *dISp*. These results are much more similar to the true currents (subfigures A1, A2, A3) than to *dISs* or *iIS*: the two centres of activity are well focussed, and even a detail such as the presence of two maxima in the frontal centre is reproduced. By this result it is illustrated again that this technique has achieved a very good estimation of the sources. This success was obtained on the basis of knowing only the (simulated) EEG recording from $n_c = 18$ electrodes (as shown in figure 1) and the correct dynamical model, including the information about the time-dependency of the autoregressive parameter a_1 for certain groups of voxels. Results like this make us presume that in this particular state space filtering problem the condition of observabiblity is fulfilled.

Figure 3 shows the time series of the vertical components of true currents and inverse solutions for two selected voxels, namely a voxel in the right medial frontal gyrus, situated in the middle of the frontal centre of alpha activity (left column of subfigures), and a voxel in the left superior frontal gyrus (right column of subfigures); only half of the data points is shown, i.e. the first 256 data points. In this figure we also show error estimates according to equations 13 and 33 (plus/minus two standard deviations). Again, the letters A, B, C and D refer to true currents, *iIS*, *dISs* and *dISp*, respectively.

We have chosen to display the vertical component, since in this simulation the true current vectors were confined to the vertical direction. It should be noted that for the time domain the absolute value of local current vectors is not an appropriate quantity for representation of inverse solutions, since it does not contain information about changes of direction of the vectors; this has the effect that oscillations seem to have a doubled frequency, as compared to the vector components.

In subfigure A1 we see the strong simulated alpha oscillation of this voxel. It is reproduced by *iIS* and *dISs* (subfigures B1 and C1), but its amplitude is significantly underestimated; error intervals are larger for *iIS* than for *dISs*. In contrast to this, *dISp* (subfigure D1) reproduces the correct amplitude of this oscillation very well. During the first 0.5 seconds the transient of the Kalman filter can be clearly seen; in the remaining part of the data the oscillation of the estimated currents does not change its amplitude any more, but stays very close to the true solution.

In subfigure A2 we see the true current of a voxel which does not take part in any pronounced oscillation. The slight decrease of the current with time still is a transient behaviour resulting from the deterministic stable autoregressive dynamics, as employed in this simulation. Subfigures B2 and C2 show, that *iIS* and *dISs* incorrectly assign a spurious oscillation to this voxel, whereas *dISp* succeeds in approximately retrieving the correct dynamics: There is no trace of spurious oscillations, and the correct solution is within the error interval. This result again illustrates the much sharper localisation which can be achieved by *dISp*.

6 Application to clinical EEG

6.1 Calculation of inverse solutions

We will now estimate inverse solutions for a time series chosen from a clinical EEG recording. The data was recorded from a healthy child of 8.5 years, in awake resting state with eyes closed. Electrodes according to the standard 10-20-system were used, the sampling rate was 256 Hz, and the resolution of the AD conversion was 12bit. A time series of 2 seconds length chosen from the recording is shown in figure 4; this representation uses average reference. As can be seen from the figure, this data set displays characteristic alpha oscillations in the parietal and occipital electrodes. This particular data set was chosen merely as an example of typical clinical EEG data; in later studies it will be possible to investigate a wide range of neurological and psychiatrical diseases by this new technique for obtaining inverse solutions.

Again we compute a "timeframe-by-timeframe" instantaneous inverse solution for this data set by using LORETA and two dynamical inverse solutions by using spatiotemporal Kalman filtering. Since in this case no additional information concerning the true dynamics is available, we will use linear autoregressive models with constant coefficients, as discussed in sections 4.1 and 4.7. A first-order ($p = 1$, $q = 1$) model will be used, according to equation 17; the resulting inverse solutions shall be abbreviated as "*dISs1*". Furthermore a second-order model will be used, but only voxel self-interaction will be second order, while neighbour interaction will remain first order ($p = 2$, $q = 1$); this model results from adding an additional term $a_2 \mathbf{l}_3 \mathbf{j}(v, t-2)$ in equation 17. The resulting inverse solutions shall be abbreviated as "*dISs2*".

In the case $p = 1$, $q = 1$ maximum likelihood estimation of the unknown parameters (a_1, b_1, σ_ϵ^2, σ_η^2) yields the values (0.7875, 0.2182, 4.8255, 1.4879×10^{-6}), and for $p = 2$, $q = 1$ we find (a_1, a_2, b_1, σ_ϵ^2, σ_η^2) = (0.9923, −0.7101, 0.7993, 3.3412, 1.0507). These solutions may not yet represent global maxima, but we believe that they correspond to inverse solutions whose properties are qualitatively similar to those of the global solution.

6.2 Comparison of inverse solutions

Unlike with the simulation study of section 5, we are in this case unable to compare the inverse solutions directly with the true solution; consequently they have to be evaluated and compared by statistical criteria. For the instantaneous solution (*iIS*) we obtain ABIC=112328.4, while for the dynamical inverse solution with $p = 1$, $q = 1$ (*dISs1*) we obtain AIC=88990.1 and for $p = 2$, $q = 1$ (*dISs2*) AIC=87131.3. These values confirm again that the dynamical inverse solutions represent better explanations of the observations than the instantaneous inverse solution. We also see that increasing the model order helps to further decrease the AIC, but at the expense of a more time-consuming parameter estimation, due to the increased dimensionality of the parameter space.

The results of this state estimation problem can be further illustrated by showing in figure 5 the residuals of the data prediction, i.e. the whitened time series (compare section 4.1). The figure demonstrates the success of explaining most of the structure in the observations through the dynamical model; however, it can also be seen that some structure is remaining, especially some components obviously belonging to the alpha oscillation. Improved dynamical models will be required in order to capture also these components.

In figures 6 and 7 we again present graphical illustrations of spatial and temporal

properties of the inverse solutions. Figure 6 shows the spatial distribution of inverse solutions obtained by *iIS* (subfigures B1, B2 and B3), *dISs1* (subfigures C1, C2 and C3) and *dISs2* (subfigures D1, D2 and D3) at a fixed moment in time, again by displaying the coronal, axial and sagittal maximum-intensity projections of the absolute values of the local current vectors.

From the figure it can be seen that all three inverse solutions locate a centre of activity in the occipital region, as should be expected for EEG data displaying pronounced alpha activity in parietal and occipital electrodes; in the dynamical inverse solutions this centre has larger amplitude and appears to be better localised, compared with neighbouring brain areas. Together with the superior values of AIC this result seems to provide evidence for improved resolution and localisation abilities of the dynamical inverse solutions.

Figure 7 shows the time series of the sagittal component of the inverse solutions for two selected voxels, namely a voxel in the right cuneus (i.e. in the occipital area; left column of subfigures) and, as before, a voxel in the right medial frontal gyrus (right column of subfigures). Error estimates are also shown. The letters B, C and D refer to *iIS*, *dISs1* and *dISs2*, respectively. Among the three projections of the current vectors, for this data set the projection onto the sagittal direction showed the largest amplitudes correlated to alpha activity, therefore we have chosen to present it.

In subfigures B1, C1, D1 we see for the occipital voxel a pronounced oscillation representing the alpha activity in the first half of this data set. All three inverse solutions reconstruct this oscillation at this voxel, but again we notice that *dISs1* and *dISs2* find much higher amplitude than *iIS*. More importantly, the error intervals of *dISs1* and *dISs2* are much smaller than those of *iIS*; in fact, it turns out that the error estimates for the instantaneous inverse solutions are so large that any deviation of the vector component from zero has to be regarded as non-significant. Similar remarks apply to subfigures B2, C2, D2, where *dISs1* and *dISs2* indicate the presence of a weak low-frequency wave, whereas *iIS* fails to find any structure.

It is not the aim of this paper to discuss these results from a physiological point of view, but primarily to present these new tools for obtaining improved inverse solutions and to compare them with a well-known representative of the currently available algorithms, the LORETA method. In future work it has to be investigated in detail, in which respect dynamical inverse solutions are capable of providing additional relevant information for the analysis of brain dynamics and for the diagnosis of diseases affecting the brain. In this paper our claim of superiority of dynamical inverse solutions over LORETA inverse solutions is based on the relative comparison of values of the ABIC and AIC criteria, regarded as a measure of distance between the estimated model and the unknown true model.

7 Conclusion

In this paper we have addressed the dynamical inverse problem of EEG generation, being a generalisation of the more traditional instantaneous problem of EEG generation, and we have presented a new approach for estimating solutions of this problem from actual EEG data. This approach is applicable also to MEG data and to a wide class of other inverse problems arising in the analysis of biomedical or other data. We have demonstrated how the standard Kalman filter can be adapted to the case of spatiotemporal dynamics; an essential precondition for this adaptation was the concept of spatial whitening which renders it possible to decompose an intractable high-dimensional filtering problem into a set of coupled low-dimensional problems which can be solved with moderate computational

effort. The application of Kalman filtering has the additional benefit of providing estimates of the likelihood and consequently of the Akaike Information Criterion (AIC), which serves as a well-justified tool for estimating parameters and comparing dynamical models through the maximum likelihood method. Furthermore, the Kalman filter provides error estimates for the inverse solutions almost without the need for additional computations.

We have demonstrated through numerical simulations that the quality of the inverse solutions obtained by this new dynamical approach crucially depends on the availability of appropriate models for the spatiotemporal brain dynamics. If only a very simple model is employed, the resulting inverse solution may not offer much improvement over the results provided by instantaneous techniques (e.g. LORETA). But even this is remarkable, since, as compared to the LORETA algorithm, our algorithm applies completely different numerical procedures to the data. We have explicitly shown, why non-coupled random-walk models, which are increasingly employed as "dynamical priors" in constrained least squares approaches to inverse problems, are inherently insufficient as dynamical models within a high-dimensional state space approach. Instead, we have suggested to incorporate the identification of improved predictive models as a central element into the dynamical inverse problem. The more elaborate a model is applied, the more the resulting inverse solutions will differ from the LORETA solution, and the better (in a statistical sense) it will be able to explain the observed EEG data, as measured by the improvement of the AIC value. In a numerical simulation we were able to employ a perfect model of the underlying dynamics, and we have shown that the resulting inverse solution was very similar to the true distribution of currents.

Perfect models are not available in the analysis of real EEG data, but it can be expected that by using the maximum likelihood method (or, more precisely, the method of minimising the AIC) it will be possible to gradually adapt initially simple models to given data, such that considerably improved models and consequently improved inverse solutions can be obtained. These models themselves will be highly useful for purposes of investigating brain dynamics and improving clinical diagnosis.

In this study we have chosen to employ the class of multivariate linear autoregressive models with constant parameters. The choice of this particular class anticipates stationarity of the underlying dynamical processes. This assumption is almost never fulfilled for the case of EEG time series, but may be permitted for short time intervals of one or two seconds, as used in this study. Future work will have to address this problem more thoroughly by developing dynamical model classes which are sufficiently flexible to cope with nonstationary data.

As a further advantage of the dynamical approach to estimating inverse solutions we would like to mention the possibility to calculate spatial innovation maps by forming the differences between predicted states (given by expressions such as equation 25) and estimated states (given by equation 31) for each voxel and each vector projection. These innovations describe those components of the spatiotemporal dynamics which could not be predicted by the given dynamical model, therefore they contain information either about weaknesses of the employed dynamical model, or about external forces and processes driving the dynamics. Such information cannot be obtained by instantaneous techniques. These maps may be particularly useful for localising points or areas within brain which display atypical behaviour, such as epileptic foci; they may also serve as a source of information for the analysis of the long-range connectivity structure of brain.

It has to be admitted that, despite the decomposition approach for rendering high-dimensional spatiotemporal filtering problems tractable, this technique still is much more demanding in terms of computational time consumption than most instantaneous tech-

niques. Especially the step of parameter estimation through likelihood maximisation may, depending on the model and the quality of the initial estimates for the parameters, consume considerable amounts of CPU time, since it constitutes a nonlinear optimisation task. Future work will have to design efficient ways to perform this optimisation. Furthermore, as also with many other nonlinear optimisation problems, it may not be necessary to always find the global minimum of the objective function (i.e. in our case the AIC as a function of parameters). Seemingly very different sets of parameters may belong to dynamical models with very similar behaviour. This can be seen from the parameter sets given in section 6.1.

Altogether we expect that by reinterpreting the dynamical inverse problem of EEG generation as a spatiotemporal filtering problem it will become possible to extract considerably more relevant and useful information from EEG recordings than it has been possible previously. Also for the combination of EEG data with data obtained by other techniques for recording temporally and spatially resolved information related to brain dynamics, such as MEG, fMRI and NIRS, this work holds the promise of opening up new perspectives.

Acknowledgements

T.Ozaki gratefully acknowledges support by the Japanese Society for the Promotion of Science (JSPS) through grants KIBAN 13654075 and KIBAN 15500193. A.Galka gratefully acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) through project GA 673/1-1 and by JSPS through fellowship ID No. P 03059. Clinical EEG data was kindly provided by U. Stephani and H. Muhle from the Neuropediatric Hospital of the University of Kiel, Germany.

Appendix: Estimation of the covariance matrix of the estimated state vector $\hat{\mathbf{J}}$ (obtained from LORETA)

The covariance matrices of the (assumed) "true" state vector $\mathbf{J}$ and of the corresponding EEG observation vector $\mathbf{Y}$ shall be denoted by $\mathsf{C}_{\mathbf{J}}$ and $\mathsf{C}_{\mathbf{Y}}$, respectively. From equation 1 we have

$$\mathsf{C}_{\mathbf{Y}} = \mathsf{K}\mathsf{C}_{\mathbf{J}}\mathsf{K}^{\dagger} + \mathsf{C}_{\epsilon} \quad . \tag{A1}$$

In this paper we are employing the assumptions (see equations 24 and 3)

$$\mathsf{C}_{\mathbf{J}} = \tau^2 (\mathsf{L}^{\dagger}\mathsf{L})^{-1} \tag{A2}$$

and

$$\mathsf{C}_{\epsilon} = \sigma_{\epsilon}^2 \, \mathsf{I}_{n_c} \quad . \tag{A3}$$

The estimated state vector is given by equation 6:

$$\hat{\mathbf{J}} = (\mathsf{K}^{\dagger}\mathsf{K} + \lambda^2 \mathsf{L}^{\dagger}\mathsf{L})^{-1}\mathsf{K}^{\dagger}\mathbf{Y} =: \mathcal{T}(\lambda)\,\mathbf{Y} \quad .$$

Then the covariance matrix of the estimated state vector $\hat{\mathbf{J}}$ is given as

$$\mathsf{C}_{\hat{\mathbf{J}}}^{(iIS)} = \mathcal{T}(\lambda)\,\mathsf{C}_{\mathbf{Y}}\,\mathcal{T}^{\dagger}(\lambda) \quad . \tag{A4}$$

The product $\mathsf{K}\mathsf{C}_{\mathbf{J}}\mathsf{K}^{\dagger}$ in equation A1 yields, using equations A2 and 8

$$\tau^2\mathsf{K}(\mathsf{L}^{\dagger}\mathsf{L})^{-1}\mathsf{K} = \tau^2(\mathsf{K}\mathsf{L}^{-1})(\mathsf{K}\mathsf{L}^{-1})^{\dagger} = \tau^2\mathsf{U}\mathsf{S}^2\mathsf{U}^{\dagger} \quad .$$

Using $\tau = \sigma/\lambda$ and equation A3, equation A1 becomes

$$\mathsf{C}_{\mathbf{Y}} = \tau^2(\mathsf{U}\mathsf{S}^2\mathsf{U}^{\dagger} + \lambda^2\mathsf{I}_{n_c}) \quad . \tag{A5}$$

By inserting equation 9 for $\mathcal{T}(\lambda)$ and equation A5 for $\mathsf{C}_{\mathbf{Y}}$ we finally can evaluate equation A4; by rearraging terms we obtain (using orthogonality of U)

$$\mathsf{C}_{\hat{\mathbf{J}}}^{(iIS)} = \tau^2\mathsf{L}^{-1}\mathsf{V}^{(1)}\mathrm{diag}\left(\frac{s_i^2}{s_i^2+\lambda^2}\right)\mathsf{V}^{(1)\,\dagger}(\mathsf{L}^{\dagger})^{-1} \quad ,$$

which is the desired result.

The diagonal elements of $\mathsf{C}_{\hat{\mathbf{J}}}^{(iIS)}$ provide variances for each element of the state vector $\mathbf{J}$; since in our implementation the estimates for τ and λ result from an optimisation step applied to the complete time series, these variances themselves will not depend on time, unlike the state estimates $\hat{\mathbf{J}} = \hat{\mathbf{J}}(t)$; if, however, variances of the modulus $j(v,t)$ of local vectors $\mathbf{j}(v,t)$ are computed, these will become time-dependent through Gaussian error propagation:

$$\sigma_j^2(v,t) = \frac{1}{j^2(v,t)}\left(\left(j_x(v,t)\,\sigma_{j_x}(v)\right)^2 + \left(j_y(v,t)\,\sigma_{j_y}(v)\right)^2 + \left(j_z(v,t)\,\sigma_{j_z}(v)\right)^2\right) \quad . \tag{A6}$$

Alternatively, explicitly time-dependent local variances could be obtained by repeating the optimisation on a moving window, but in this paper we have not used this option.

References

[1] P.L. Nunez. *Electrical fields of the brain*. Oxford University Press, New York, 1981.

[2] R. Greenblatt. Probabilistic reconstruction of multiple sources in the bioelectromagnetic inverse problem. *Inverse Problems*, 9:271–284, 1993.

[3] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.*, 18:49–65, 1994.

[4] R. Grave de Peralta Menendez and S. Gonzalez Andino. A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. *IEEE Trans. Biomed. Eng.*, 45:440–448, 1999.

[5] C. Phillips, M. D. Rugg, and K. J. Friston. Systematic regularization of linear inverse solutions of the EEG source localization problem. *NeuroImage*, 17:287–301, 2002.

[6] U. Schmitt, A. K. Louis, C. Wolters, and M. Vauhkonen. Efficient algorithms for the regularization of dynamic inverse problems: II. Applications. *Inverse Problems*, 18:659–676, 2002.

[7] O. Yamashita, A. Galka, T. Ozaki, R. Biscay, and P. A. Valdés-Sosa. Recursive penalized least squares solution for the dynamical inverse problem of EEG generation. *Human Brain Mapping*, 21:221–235, 2004.

[8] S. Baillet, J.C. Mosher, and R.M. Leahy. Electromagnetic brain mapping. *IEEE Sign. Proc. Mag.*, 14–30, November 2001.

[9] M. Scherg and J.S. Ebersole. Brain source imaging of focal and multifocal epileptiform EEG activity. *Neurophysiol. Clinique*, 24:51–60, 1994.

[10] J.C. Mosher, P. Lewis, and R. Leahy. Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.*, 39:541–557, 1992.

[11] M. S. Hämäläinen and R. J. Ilmoniemi. Interpreting measured magnetic fields of the brain: estimates of current distributions. Technical report TKK-F-A559, Helsinki University of Technology, Espoo, 1984.

[12] J. P. Kaipio, P. A. Karjalainen, E. Somersalo, and M. Vauhkonen. State estimation in time-varying electrical impedance tomography. *Annals N.Y. Acad. Sci.*, 873:430–439, 1999.

[13] F. Darvas, U. Schmitt, A. K. Louis, M. Fuchs, G. Knoll, and H. Buchner. Spatiotemporal current density reconstruction (stCDR) from EEG/MEG data. *Brain Topogr.*, 13:195–207, 2001.

[14] E. Somersalo, A. Voutilainen, and J. P. Kaipio. Non-stationary magnetoencephalography by Bayesian filtering of dipole models. *Phys. Rev. E*, 51:2955–2962, 1995.

[15] S. Baillet and L. Garnero. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. *IEEE Trans. Biomed. Eng.*, 44:374–385, 1997.

[16] C. K. Chui and G. Chen. *Kalman filtering: with real-time applications.* Springer Series in Information Sciences, vol. 17. Springer, Berlin, Heidelberg, New York, 1999 (3rd ed.).

[17] R. E. Kalman, P. L. Falb, and M. A. Arbib. *Topics in mathematical system theory.* International series in pure and applied mathematics. McGraw-Hill, New York, 1969.

[18] Th. Kailath. *Linear systems.* Information and system sciences series. Prentice-Hall, Englewood Cliffs, 1980.

[19] P. A. Karjalainen. *Regularization and Bayesian Methods for Evoked Potential Estimation.* Ph.D. thesis, Kuopio University Publications C. Natural and Environmental Sciences 61, Kuopio, 1997.

[20] D. Baroudi, J. Kaipio, and E. Somersalo. Dynamical electric wire tomography: a time series approach. *Inverse Problems*, 14:799–813, 1998.

[21] J. P. Ary, S. A. Klein, and D. H. Fender. Location of sources of evoked scalp potentials: corrections for skull and scalp thickness. *IEEE Trans. Biomed. Eng.*, 28:447–452, 1981.

[22] H. Buchner, G. Knoll, M. Fuchs, A. Rienäcker, R. Beckmann, M. Wagner, J. Silny, and J. Pesch. Inverse localization of electric dipole current sources in finite element models of the human head. *Electroenc. Clin. Neurophysiol.*, 102:267–278, 1997.

[23] J. J. Riera, M. E. Fuentes, P. A. Valdés, and Y. Ohárriz. EEG-distributed inverse solutions for a spherical head model. *Inverse Problems*, 14:1009–1019, 1998.

[24] R. D. Pascual-Marqui. Review of methods for solving the EEG inverse problem. *Int. J. Bioelectromagn.*, 1:75–86, 1999.

[25] B. Hofmann. *Regularization for Applied Inverse and Ill-Posed Problems*. Teubner, Leipzig, 1986.

[26] A. Tarantola. *Inverse problem theory: methods for data fitting and model parameter estimation*. Elsevier, Amsterdam, New York, 1987.

[27] G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1990.

[28] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, Englewood Cliffs, 1974.

[29] H. Akaike. Seasonal adjustment by a Bayesian modeling. *J. Time Series Anal.*, 1:1–13, 1980.

[30] H. Akaike. Likelihood and the Bayes procedure. In J. M. Bernardo, M. H. De Groot, D. U. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics*, pages 141–166. University Press, Valencia (Spain), 1980.

[31] H. Akaike. Prediction and entropy. In A. C. Atkinson and S. E. Fienberg, editors, *A celebration of statistics*, pages 1–24. Springer, Berlin, Heidelberg, New York, 1985.

[32] R. D. Pascual-Marqui. Standardized low resolution brain electromagnetic tomography (sLORETA): technical details. *Methods & Findings in Exp. & Clin. Pharmacol.*, 24D:5–12, 2002.

[33] B. Hjorth. Source derivation simplifies topographical EEG interpretation. *Americ. J. EEG Techn.*, 20:121–132, 1980.

[34] P. A. Robinson, C. J. Rennie, and J.J. Wright. Propagation and stability of waves of electrical activity in the cerebral cortex. *Phys. Rev. E*, 56:826–840, 1997.

[35] V. K. Jirsa, K. J. Jantzen, A. Fuchs, and J. A. S. Kelso. Spatiotemporal forward solution of the MEG using network modeling. *IEEE Trans. Med. Imag.*, 21:493–504, 2002.

[36] T. Ozaki, J. C. Jimenez, and V. Haggan-Ozaki. The role of the likelihood function in the estimation of chaos models. *J. Time Series Analysis*, 21:363–387, 2000.

[37] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Autom. Contr.*, 19:716–723, 1974.

[38] Xing-Qi Jiang. *Bayesian methods for modeling, identification and estimation of stochastic systems*. Ph.D. thesis, Department of Statistical Science, Graduate University for Advanced Studies, Tokyo, 1992.

[39] J. C. Mazziotta, A. Toga, A. C. Evans, P. Fox, and J. Lancaster. A probabilistic atlas of the human brain: theory and rationale for its development. *NeuroImage*, 2:89–101, 1995.

[40] J. Bosch-Bayard, P. A. Valdés-Sosa, T. Virues-Alba, E. Aubert-Vázquez, E. Roy John, T. Harmony, J. Riera-Díaz, and N. Trujillo-Barreto. 3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA). *Clin. Electroenc.*, 32:47–61, 2001.

[41] I. F. Gorodnitsky, J. S. Georg, and B. D. Rao. Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm. *Electroenc. Clin. Neurophysiol.*, 95:231–251, 1995.

Figure captions

Fig. 1: Simulated EEG recording for 18 standard electrodes according to the 10-20 system (PZ has been omitted); electrode abbreviations are given on the vertical axis. The EEG potential is measured in arbitrary units versus average reference of 19 electrodes (including PZ), time is measured in seconds, assuming a sampling rate of the simulated dynamics of 256 Hz.

Fig. 2: Gray-scale coded representation of maximum-intensity projection of the three-dimensional field of absolute values of local current vectors for the gray-matter voxels of a model brain at a fixed point in time, using coronal projection (left column), axial projection (middle column) and sagittal projection (right column). Subfigures A1, A2, A3 show the original current vectors used in the simulation; subfigures B1, B2, B3 show the estimated current vectors according to the instantaneous inverse solution (LORETA), subfigures C1, C2, C3 show the estimated current vectors according to the dynamical inverse solution using the simplest dynamical model, and subfigures D1, D2, D3 show the estimated current vectors according to the dynamical inverse solution using the perfect dynamical model.

Fig. 3: Vertical (axial) component of local current vectors for a voxel in right medial frontal gyrus (left column of subfigures) and for a voxel in left superior frontal gyrus (right column of subfigures) versus time, based on inverse solutions obtained from the simulated EEG recording shown in figure 1. Subfigures A1, A2 show values from the original current vectors used in the simulation, subfigures B1, B2 show results from estimated current vectors according to the instantaneous inverse solution (LORETA), subfigures C1, C2 show results from estimated current vectors according to the dynamical inverse solution using the simplest dynamical model, and subfigures D1, D2 show results from estimated current vectors according to the dynamical inverse solution using the perfect dynamical model. Thick lines represent the estimates, while thin lines represent error intervals (95% confidence profiles). Note the different scale on the vertical axis for subfigures of left and right columns.

Fig. 4: Clinical EEG recording from a healthy child (8.5 years, awake, eyes closed) for 18 standard electrodes according to the 10-20 system (PZ has been omitted); electrode abbreviations are given on the vertical axis. The EEG potential is measured in arbitrary units versus average reference of 19 electrodes (including PZ), time is measured in seconds, the sampling rate is 256 Hz.

Fig. 5: Observation prediction errors for the EEG recording shown in figure 4, generated by a dynamical inverse solution based on a second-order autoregressive model. The scale on the vertical axis has been enlarged by a factor of 2.7, as compared to figure 4.

Fig. 6: Gray-scale coded representation of maximum-intensity projection of the three-dimensional field of absolute values of local current vectors for the gray-matter voxels of a model brain at a fixed point in time, using coronal projection (left column), axial projection (middle column) and sagittal projection (right column), based on inverse so-

lutions obtained from the EEG recording shown in figure 4. Subfigures B1, B2, B3 show the estimated current vectors according to the instantaneous inverse solution (LORETA), subfigures C1, C2, C3 show the estimated current vectors according to the dynamical inverse solution using a first-order autoregressive model, and subfigures D1, D2, D3 show the estimated current vectors according to the dynamical inverse solution using a second-order autoregressive model.

Fig. 7: Vertical (axial) component of local current vectors for a voxel in right cuneus (left column of subfigures) and for a voxel in right medial frontal gyrus (right column of subfigures) versus time, based on inverse solutions obtained from the EEG recording shown in figure 4. Subfigures B1, B2 show results from estimated current vectors according to the instantaneous inverse solution (LORETA), subfigures C1, C2 show results from estimated current vectors according to the dynamical inverse solution using a first-order autoregressive model, and subfigures D1, D2 show results from the estimated current vectors according to the dynamical inverse solution using a second-order autoregressive model. Thick lines represent the estimates, while thin lines represent error intervals (95% confidence profiles).

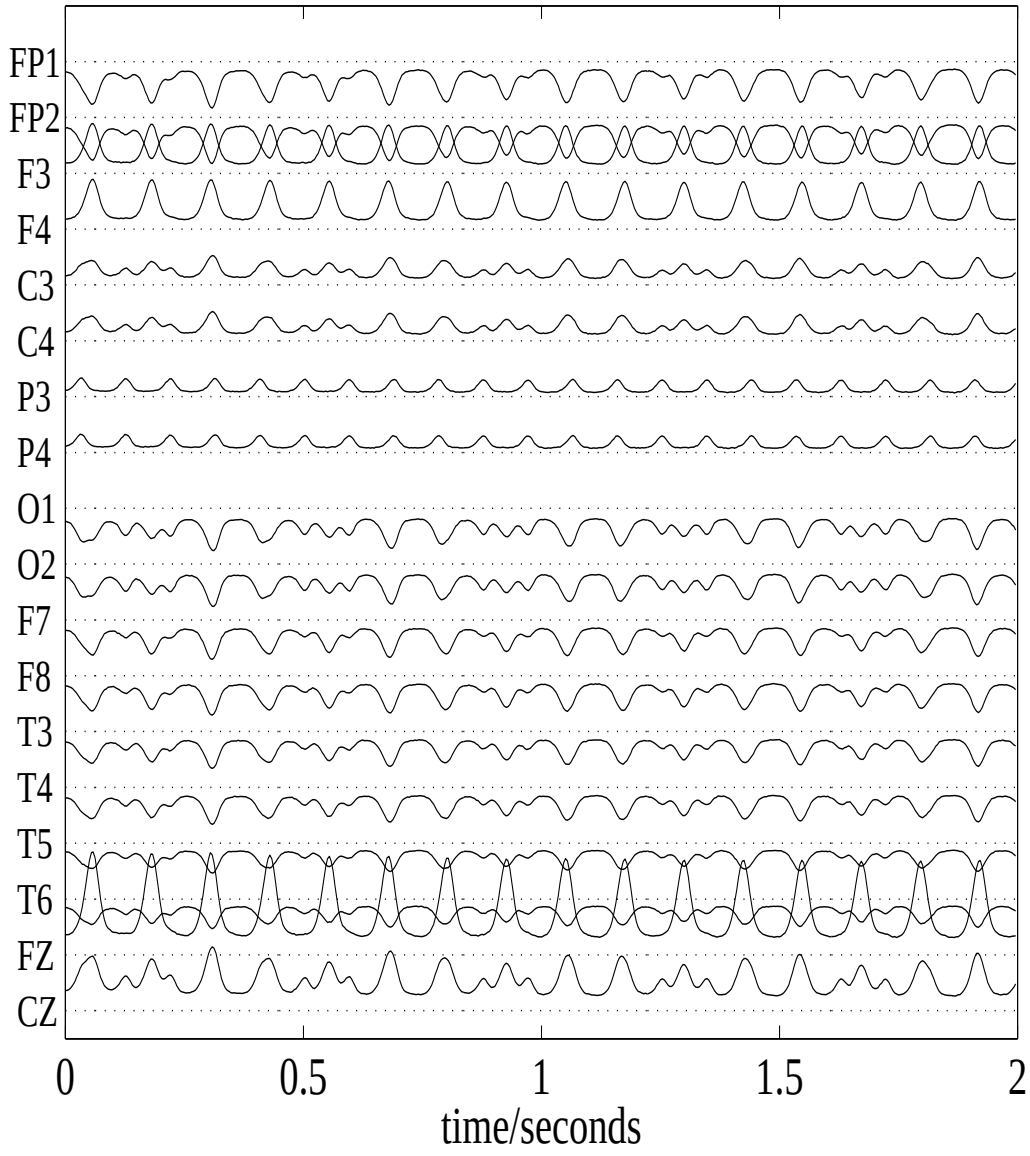
FP1
FP2
F3
F4
C3
C4
P3
P4
O1
O2
F7
F8
T3
T4
T5
T6
FZ
CZ
0
0.5
1
1.5
2
time/seconds

Figure 1:

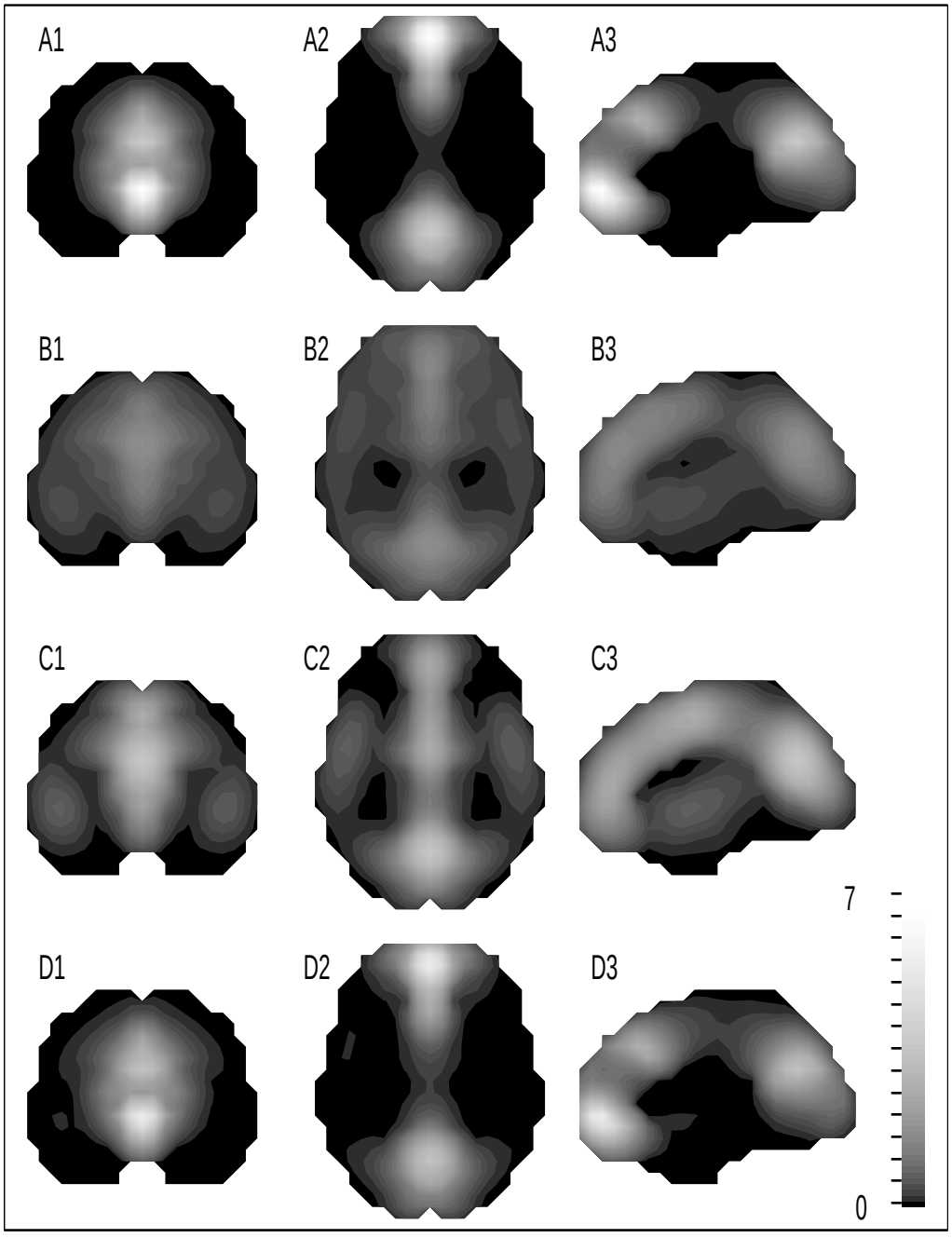

Figure 2:

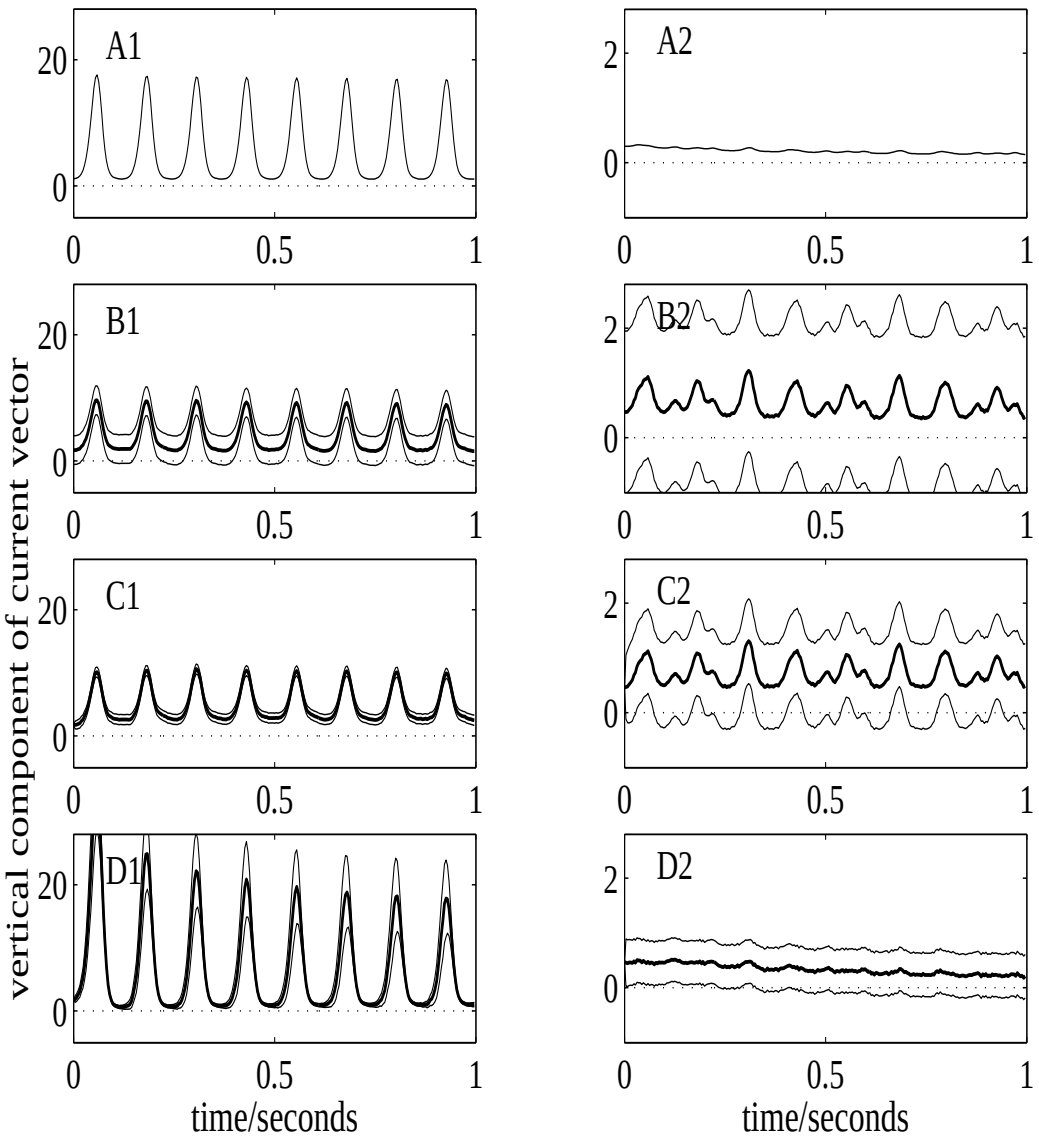
A1
A2
B1
B2
C1
C2
D1
D2
20
0
2
0
0.5
1
vertical component of current vector
time/seconds
time/seconds

Figure 3:

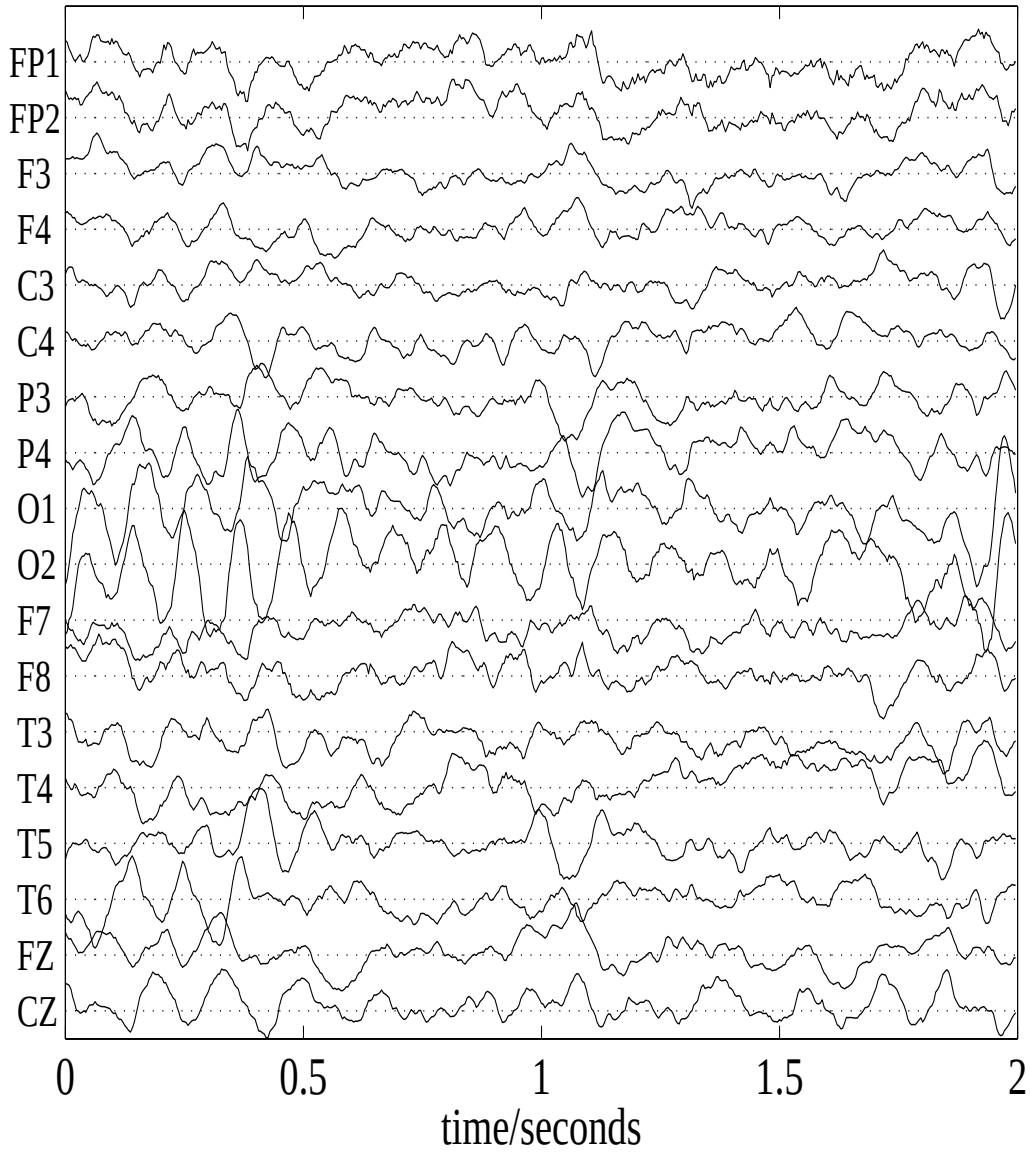
FP1
FP2
F3
F4
C3
C4
P3
P4
O1
O2
F7
F8
T3
T4
T5
T6
FZ
CZ
0
0.5
1
1.5
2
time/seconds

Figure 4:

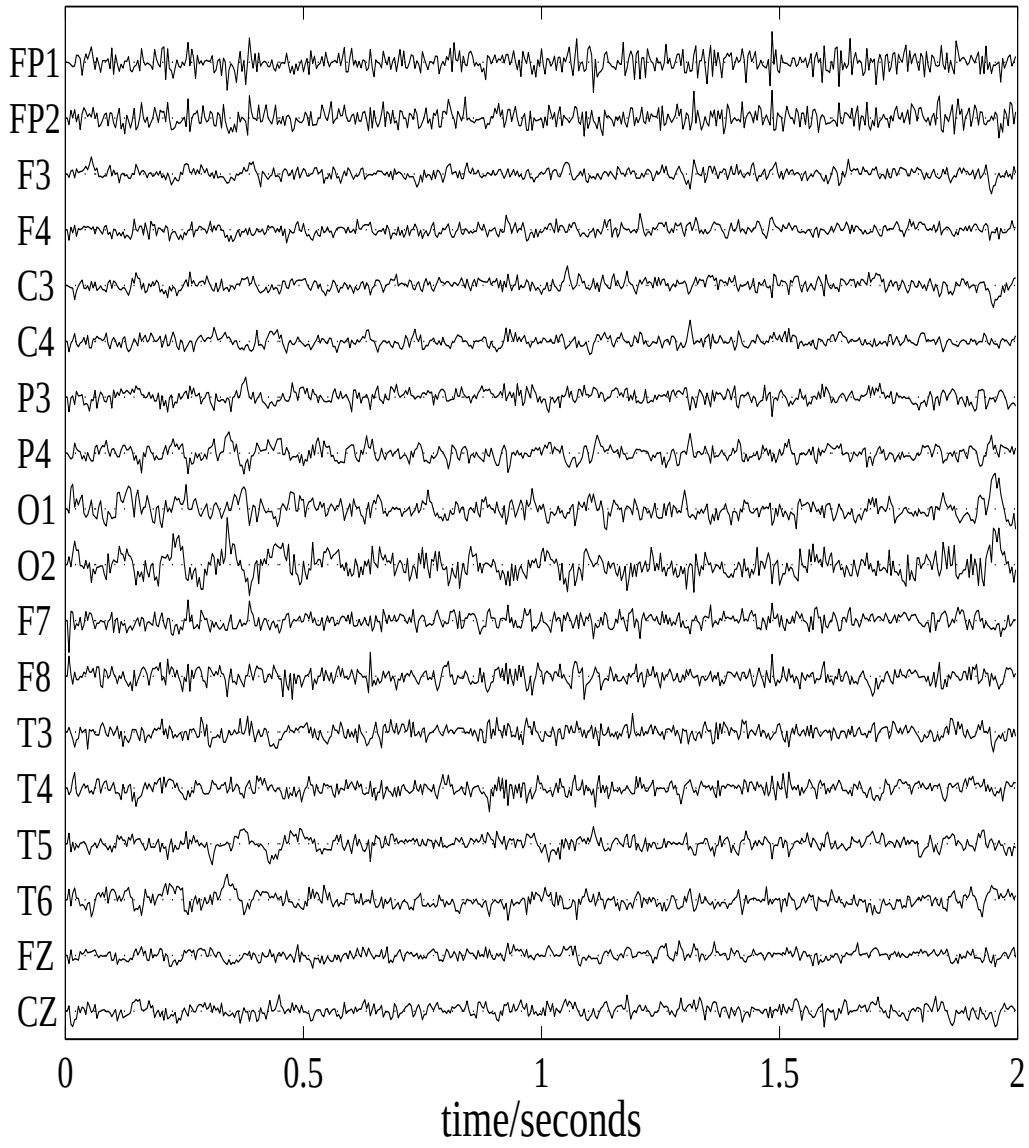
FP1
FP2
F3
F4
C3
C4
P3
P4
O1
O2
F7
F8
T3
T4
T5
T6
FZ
CZ
0
0.5
1
1.5
2
time/seconds

Figure 5:

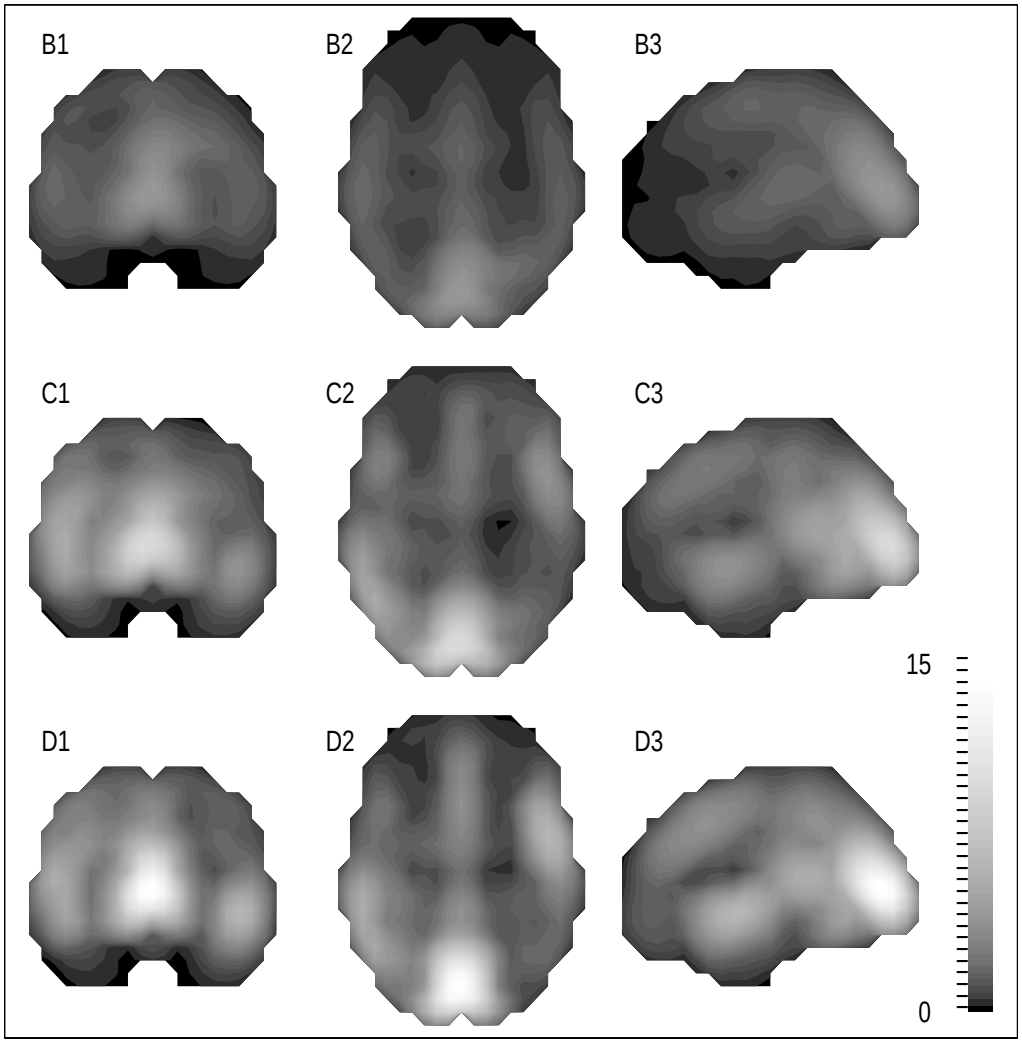

Figure 6:

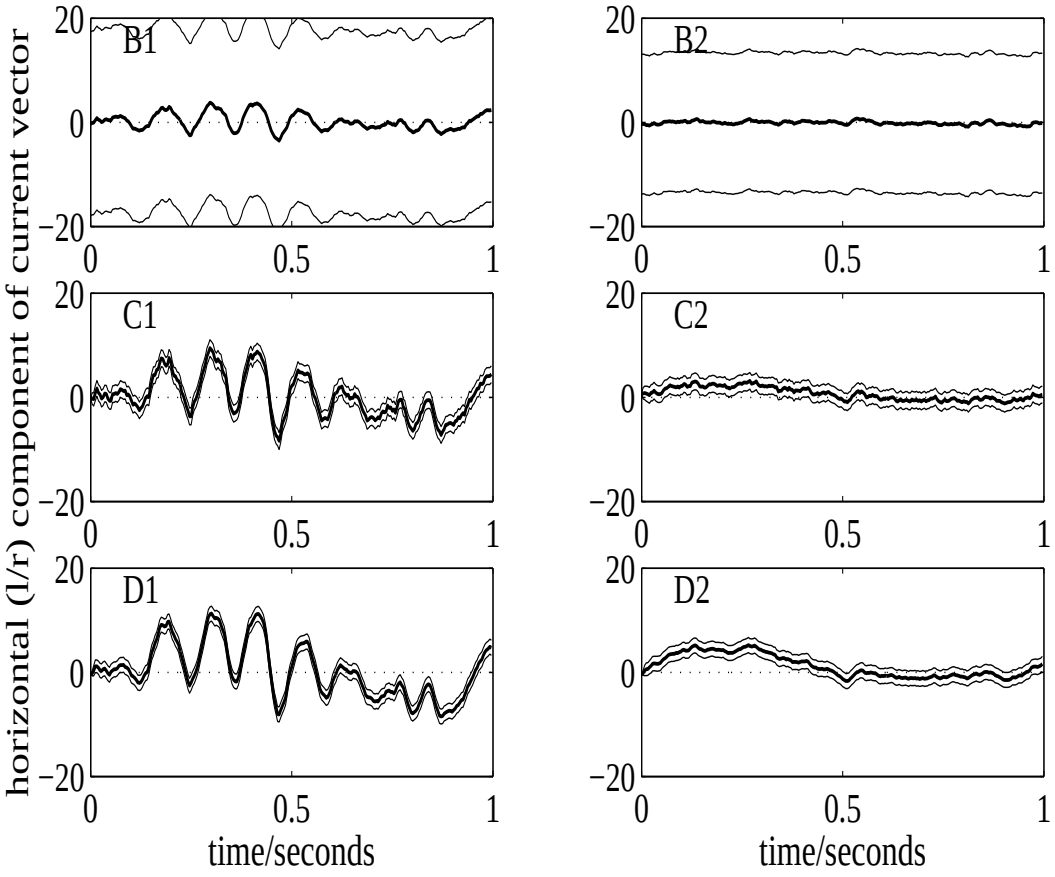
B1
B2
C1
C2
D1
D2
20
0
−20
0
0.5
1
horizontal (l/r) component of current vector
time/seconds

Figure 7: