Joint EEG/fMRI state space model for the detection of directed interactions in human brains—a simulation study

2011 Physiol. Meas. 32 1725

(http://iopscience.iop.org/0967-3334/32/11/S01)

IOP PUBLISHING PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **32** (2011) 1725–1736 doi:10.1088/0967-3334/32/11/S01

Joint EEG/fMRI state space model for the detection of directed interactions in human brains—a simulation study

Michael Lenz[1,2,3], Mariachristina Musso[2], Yannick Linke[1,2], Oliver Tüscher[4], Jens Timmer[1,5,6], Cornelius Weiller[2] and Björn Schelter[1,6]

[1] Freiburg Center for Data Analysis and Modeling, University Freiburg, Freiburg, Germany
[2] Department of Neurology, University Medical Center of Freiburg, Freiburg, Germany
[3] Aachen Institute for Advanced Study in Computational Engineering Science, RWTH Aachen, Aachen, Germany
[4] Department of Psychiatry, University Medical Center of Freiburg, Freiburg, Germany
[5] Freiburg Institute for Advanced Studies, University Freiburg, Freiburg, Germany
[6] Institute of Physics, University Freiburg, Freiburg, Germany

E-mail: schelter@fdm.uni-freiburg.de

Received 29 April 2011, accepted for publication 18 August 2011
Published 25 October 2011
Online at stacks.iop.org/PM/32/1725

Abstract

An often addressed challenge in neuroscience research is the assignment of different tasks to specific brain regions. In many cases several brain regions are activated during a single task. Therefore, one is also interested in the temporal evolution of brain activity to infer causal relations between activated brain regions. These causal relations may be described by a directed, task specific network which consists of activated brain regions as vertices and directed edges. The edges describe the causal relations. Inference of the task specific brain network from measurements like electroencephalography (EEG) or functional magnetic resonance imaging (fMRI) is challenging, due to the low spatial resolution of the former and the low temporal resolution of the latter. Here, we present a simulation study investigating a possible combined analysis of simultaneously measured EEG and fMRI data to address the challenge specified above. A nonlinear state space model is used to distinguish between the underlying brain states and the (simulated) EEG/fMRI measurements. We make use of a modified unscented Kalman filter and a corresponding unscented smoother for the estimation of the underlying neural activity. Model parameters are estimated using an expectation-maximization algorithm, which exploits the partial linearity of our model. Inference of the brain network structure is then achieved using directed partial correlation, a measure for Granger causality. The results indicate that the convolution effect of the fMRI forward model imposes a big challenge for the parameter estimation and reduces the influence of the fMRI in combined EEG–fMRI models. It remains to be investigated

whether other models or similar combinations of other modalities such as, e.g., EEG and magnetoencephalography can increase the profit of the promising idea of combining various modalities.

Keywords: EEG–fMRI fusion, effective connectivity, Kalman filter

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Estimation of effective connectivity between different brain regions from measured signals is a demanding task, which has been addressed by several researchers. In applications, dynamic causal modeling (DCM) (Friston *et al* 2003) is often used for functional magnetic resonance imaging (fMRI) data. Another method, which was introduced for the analysis of fMRI data more recently, is directed partial correlation (dPC) (Eichler 2005, Mader *et al* 2008). The analog to the dPC in the frequency domain is partial directed coherence (Baccala *et al* 2006), which was used for inference on cortical connections using electroencephalographic (EEG) data (Supp *et al* 2007).

In fMRI data, the spatial resolution is extremely high, while the temporal resolution often does not suffice to identify causal relations between the connected brain regions. Usually it is not possible to distinguish between cause and effect. A further challenge for the analysis of fMRI data is the fact that neural activity is observed indirectly through the blood oxygen level dependence (BOLD) effect. The linkage of neural activity to the BOLD signal is nonlinear.

EEG measurements in contrast measure the neural activity more directly with a very high temporal resolution, while suffering from the ill-posed inverse problem, which is explained in the following. The measured EEG data have to be interpreted as a weighted sum of different cortical sources. The problem of deconvolving these sums, to extract the signals of single sources themselves, is called the inverse problem. The ill-posedness of the inverse problem describes the fact that there is no unique solution to it (Nunez and Srinivasan 2006). One has to deal with this problem when trying to infer causal connections between brain regions from EEG data. A further challenge in the analysis of EEG data is the presence of deep sources, resulting in a reduced EEG signal.

In order to overcome the disadvantages of the single modalities, a multimodal method to infer directed interactions is proposed here. This multimodal approach can be in general applied to many different measurements, such as EEG, magnetoencephalography (MEG), fMRI or near-infrared spectroscopy (NIRS).

This paper investigates the applicability to simulated EEG and fMRI data. A key focus is on the combination of EEG and fMRI data. In Deneux and Faugeras (2010), a similar combination of EEG and fMRI data is used to estimate source signals at a dense voxel grid over the whole brain. In contrast to this, we restrict ourselves to a few regions of interest (ROIs), but aim to also estimate effective connectivity between these brain regions.

The paper is structured as follows. Section 2 introduces a class of state space models, which we propose for the estimation of effective connectivity and gives some examples of concrete models. Section 3 details the estimation procedure. Some results on simulated data are discussed in section 4.

2. State space models for the estimation of effective connectivity

In this section, we briefly describe discrete state space models in general and propose a special class of state space models devoted to the estimation of effective connectivity. State space models are very general and can be used for many different tasks, e.g. for the decomposition of EEG signals for artifact removal (Galka *et al* 2010), where they show benefits compared to independent component analysis or factor analysis.

A discrete state space model is characterized by a Markov process $\mathbf{X}$ defining a hidden state and observations $\mathbf{Y}(t)$, $t = 1, \ldots, N$, which are conditionally independent given the state (van der Merwe and Wan 2003). It may be described by a state and an observation equation in the following way:

$$\mathbf{X}(t) = \mathbf{F}(\mathbf{X}(t-1), \mathbf{U}(t), \boldsymbol{\epsilon}(t), \boldsymbol{\Theta})$$
$$\mathbf{Y}(t) = \mathbf{G}(\mathbf{X}(t), \mathbf{V}(t), \boldsymbol{\xi}(t), \boldsymbol{\Theta}).$$

Here, the transition function $\mathbf{F}$ and the observation function $\mathbf{G}$ depend on some time-dependent external inputs $\mathbf{U}$ and $\mathbf{V}$, respectively, as well as a parameter vector $\boldsymbol{\Theta}$. Throughout this paper, the noise terms $\boldsymbol{\epsilon}$ and $\boldsymbol{\xi}$ are assumed to be independently Gaussian distributed with zero mean and the covariance matrix Q and R, respectively.

For the estimation of effective connectivity, the observations $\mathbf{Y}$ may denote any measurements from techniques such as, e.g., fMRI, EEG, MEG or NIRS. The hidden state $\mathbf{X}$ in contrast denotes quantities that are not directly available through the performed measurements, e.g. neural activity, cerebral blood flow (CBF), cerebral blood volume (CBV) or deoxyhemoglobin content.

A key step in the definition of our model is the choice of a quantity for which the connectivity is modeled. This quantity is a part of the hidden state $\mathbf{X}$ and will be denoted as $\mathbf{Z}$ in the following. In the proposed class of state space models, the transition equation for $\mathbf{Z}$ is fixed to an autoregressive process of order p (AR[p]) with a weighted additive input $\mathbf{U}$, i.e.

$$\mathbf{Z}(t) = \sum_{i=1}^{p} A_i \mathbf{Z}(t-i) + \Psi \mathbf{U}(t) + \boldsymbol{\epsilon}(t). \tag{1}$$

The parameter matrices A_i describe the connectivity between different brain regions at different time lags i, whereas the parameter matrix Ψ weights the external inputs. The transition equations of possible further states are not restricted to any special type. However, they are assumed to not incorporate additional noise terms. For simplicity, it is further assumed that there are no external inputs (apart from noise) at the measurement level, i.e. $V \equiv 0$, and that the measurement noise is additive and noise terms at different sensors are independent of each other, i.e. R is diagonal.

In summary, the proposed model for the estimation of effective connectivity is described by the following transition and observation equations:

$$\mathbf{X}(t) = \begin{pmatrix} \mathbf{Z}(t) \\ \boldsymbol{\chi}(t) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{p} A_i \mathbf{Z}(t-i) + \Psi \mathbf{U}(t) + \boldsymbol{\epsilon}(t) \\ \mathbf{F}_{\chi}(\mathbf{X}(t-1)) \end{pmatrix} \tag{2}$$

$$\mathbf{Y}(t) = \begin{pmatrix} \mathbf{Y}^{\mathrm{Mod1}}(t) \\ \mathbf{Y}^{\mathrm{Mod2}}(t) \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{G}^{\mathrm{Mod1}}(\mathbf{X}(t)) \\ \mathbf{G}^{\mathrm{Mod2}}(\mathbf{X}(t)) \\ \vdots \end{pmatrix} + \boldsymbol{\xi}(t) \tag{3}$$

$$\epsilon \sim \mathcal{N}(0, Q), \qquad \xi \sim \mathcal{N}(0, R), \tag{4}$$

where $\boldsymbol{\chi}$ denotes the additional state variables. As indicated in (3), the model can use multiple simultaneous measurements, such as, e.g., EEG and fMRI or EEG and MEG. A

critical requisite for this model is a common variable $\mathbf{Z}$ for which the connectivity is modeled among all modalities. This variable represents the main linkage between the modalities. In the following, some specific examples of the proposed model class are given to show its flexibility. We concentrate on the analysis of fMRI and EEG data, but the proposed model class can also be used to analyze MEG or NIRS data.

2.1. Application to fMRI data

Eichler (2005) and more recently Mader *et al* (2008) proposed the estimation of effective connectivity based on autoregressive processes. They modeled the fMRI measurement directly by an autoregressive process of order p (AR[p]). A straightforward extension of this model would be the explicit incorporation of measurement noise in a separate measurement equation, resulting in the following state space model:

$$\mathbf{Z}(t) = \sum_{i=1}^{p} A_i \mathbf{Z}(t-i) + \epsilon(t)$$
$$\mathbf{Y}(t) = \mathbf{Z}(t) + \boldsymbol{\xi}(t).$$

Here, $\mathbf{Y}$ is the actual fMRI measurement, while $\mathbf{Z}$ is the noise-free BOLD signal.

2.2. Application to EEG data

For the analysis of EEG data, a similar model may be used (Cheung *et al* 2008, 2009). The EEG channels are not directly associated with specific regions of the brain but represent a weighted sum of signals from different regions. To account for this fact, an additional matrix K, describing the EEG forward model, is used, which maps the cortical signals to the EEG measurements. The state space model is then described by

$$\mathbf{Z}(t) = \sum_{i=1}^{p} A_i \mathbf{Z}(t-i) + \epsilon(t) \tag{5}$$

$$\mathbf{Y}(t) = K\mathbf{Z}(t) + \boldsymbol{\xi}(t), \tag{6}$$

where $\mathbf{Z}$ describes electrical dipoles at d cortical ROIs and $\mathbf{Y}$ models the EEG signal at e EEG sensors. Consequently, the so-called lead-field matrix K is of dimension $e \times d$.

In our model, the orientation of the electrical dipoles, which are supposed to be the origin of the EEG signal, is assumed to be fixed and known. We assume a single dipole, located in the center of the ROI, to represent the EEG source of the whole ROI. This restriction may be relaxed using a cortical patch basis model (Cheung *et al* 2009, Limpiti *et al* 2006). In the simulations we use the template head model from spm8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) together with the boundary element method to calculate the EEG forward model.

For the analysis of MEG data, the same state space model can be used by replacing the EEG forward model, i.e. the matrix K, by the MEG forward model.

2.3. Combination of EEG and fMRI

To give an example of an extended model incorporating two modalities (figure 1), we consider the combination of EEG and fMRI data in a similar way as in Deneux and Faugeras (2010). The variable $\mathbf{Z}$ from the extended model above is chosen to be the mean-subtracted neural activity at d ROIs. For the fMRI part of the model, some additional state variables ($\boldsymbol{\chi}$) are introduced, namely the CBF $\mathbf{f}$, the CBV $\mathbf{v}$ and the deoxyhemoglobin content $\mathbf{q}$ normalized to

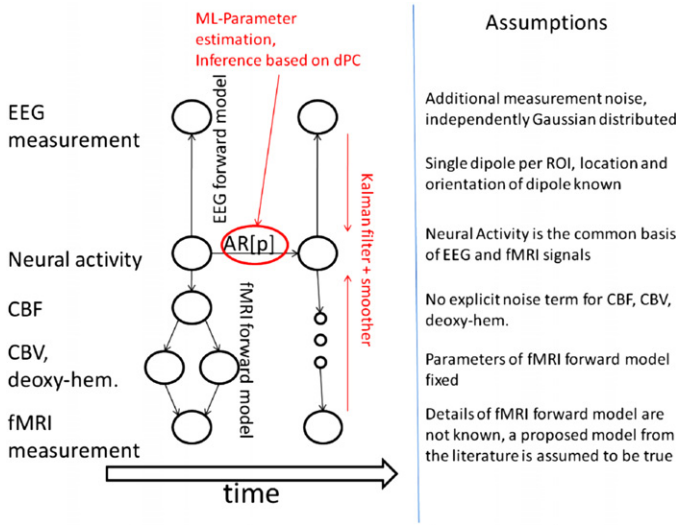

Figure 1. Sketch of the model and summary of assumptions. Neural activity is the common basis of EEG and fMRI measurements. The linkage between neural activity and the measurements is given by the respective forward models. The neural activity evolves over time according to an autoregressive process of order p (AR[p]). Red parts of the figure describe the estimation procedure consisting of the estimation of hidden variables by the Kalman filter and smoother and a maximum-likelihood (ML) parameter estimation.

baseline. The nonlinear transition function for these hemodynamic states is described by the following system of differential equations (Deneux and Faugeras 2010):

$$\begin{pmatrix} \ddot{\mathbf{f}} \\ \dot{\mathbf{v}} \\ \dot{\mathbf{q}} \end{pmatrix} = \begin{pmatrix} \epsilon \cdot (\mathbf{Z} + \mathbf{z}_0) - \kappa_s \dot{\mathbf{f}} - \kappa_f (\mathbf{f} - 1) \\ \frac{1}{\tau}(\mathbf{f} - \mathbf{v}^{\frac{1}{\alpha}}) \\ \frac{1}{\tau}\left(\mathbf{f}\frac{1-(1-E_0)^{1/\mathbf{f}}}{E_0} - \mathbf{v}^{\frac{1}{\alpha}}\frac{\mathbf{q}}{\mathbf{v}}\right) \end{pmatrix}. \tag{7}$$

Here, $\epsilon, \kappa_s, \kappa_f, \tau, \alpha$ and E_0 are physiological parameters, assumed to be known (Deneux and Faugeras 2010), and $\mathbf{z}_0$ is the mean neural activity, i.e. $\mathbf{Z} = \tilde{\mathbf{Z}} - \mathbf{z}_0$, where $\tilde{\mathbf{Z}}$ denotes the neural activity. For the transition from time t to time $t + 1$, this system of differential equations is numerically integrated with the values at time t as initial condition. Note that we assume no additive noise term for this part of the transition function. Nevertheless, the hemodynamic variables are random variables due to the dependence on the neural activity $\tilde{\mathbf{Z}}$, which is modeled with an explicit additive noise term. The observation or measurement equation for the fMRI data is eventually given by

$$\mathbf{Y}_{t+1}^{\text{fMRI}} = V_0(a_1(\mathbf{v}_t - 1) + a_2(\mathbf{q}_t - 1)) + \xi_{t+1}^{\text{fMRI}}, \tag{8}$$

where V_0, a_1 and a_2 are again known parameters as in Deneux and Faugeras (2010).

As a simplification, the electrical dipoles, representing the sources of the EEG signal, are assumed to equal the mean-subtracted neural activity. This simplification is chosen since the true connection between these quantities is, to the best of the authors' knowledge, unknown and more realistic models would further increase the complexity of the model. Thus, the EEG part of the proposed model is identical to the EEG model from the previous section (6).

3. Parameter estimation

The proposed class of state space models (2)–(4) involves several parameters, which have to be estimated to calculate effective connectivities. These include the parameter matrices A_i and Ψ of the AR process as well as the covariance matrices Q and R of the process and observation noise, respectively. Furthermore, the distribution of the state variables at time zero is required. Assuming a Gaussian distribution for the state variables, this problem reduces to the estimation of the mean μ_0 and covariance $\mathbf{Q}_0$.

The parameters can be estimated by their maximum likelihood (ML) estimates. This is done by an expectation maximization (EM) algorithm, which is similar to the EM algorithm for linear state space models (Shumway and Stoffer 2006). The EM algorithm is an iterative two-step procedure consisting of an expectation and a maximization step. These steps are iterated successively until some convergence criteria are met. The basic idea of this algorithm is to start with some initial parameters and approximate the state distribution using the Kalman filter and smoother. The approximated state distribution is used together with the observations to build a joint likelihood. This joint likelihood is then maximized in the maximization step, resulting in new parameter estimates for the next iteration.

3.1. Expectation step

The expectation step requires the calculation of the conditional state distribution, given the data. For linear state space models, this can be done by use of the linear Kalman filter and smoother, exploiting the fact that linear transformations of Gaussian random variables result again in Gaussian random variables. For nonlinear transformations, the state distribution has to be approximated (van der Merwe and Wan 2003). Two common approximations are a non-parametric approximation, e.g. by particle filters or a Gaussian approximation. For the present application, the main disadvantage of non-parametric approximations is a rather complicated ML estimation. Therefore, a Gaussian approximation is used. This reduces the calculation of the conditional state distribution to the calculation of the conditional expectation and variance. For this purpose, we make use of the unscented Kalman filter (UKF) and the corresponding unscented smoother, depicting a computational efficient second-order approximation (Julier and Uhlmann 1997). A numerically stable form of the UKF is given by the square root UKF (van der Merwe 2004). Furthermore, it is possible to constrain the sigma points to a physically realistic domain. All sigma points lying outside the valid domain are projected onto the valid domain. The resulting constrained UKF shows further improvements in the performance compared to the unconstrained UKF (Kandepu *et al* 2008, Kolas *et al* 2009). As a last modification of the usual UKF, the filter used in this paper allows for missing observations (Shumway and Stoffer 2006). This is important for the combination of two modalities with different sampling frequencies.

The unscented smoother is a straightforward implementation of the nonlinear smoother equations (Sarkka 2008) using the results from the constrained square root UKF. The ML estimation below finally requires the calculation of a lag-one covariance smoother (Shumway and Stoffer 2006), which can be approximated using the quantities from the UKF.

3.2. Maximization step

The proposed class of state space models has a particular structure, which can be exploited for the ML parameter estimation. The state equation for the variable Z (2), involving the parameters A, Ψ, and Q, is linear. For these parameters the analytical maximization as for

linear state space models (Shumway and Stoffer 2006) can be used. The jth parameter update for the parameter R is given by

$$R^{(j)} = N^{-1} \sum_{t=1}^{N} \left(Var(\mathbf{y}(t) - \hat{\mathbf{y}}(t)) + (\mathbf{y}(t) - \hat{\mathbf{y}}(t))(\mathbf{y}(t) - \hat{\mathbf{y}}(t))^T + R_{I_t}^{(j-1)} \right),$$

where $\hat{\mathbf{y}}(t) = E(G(\mathbf{X}(t))|\overline{\mathbf{y}}_N)$ and

$$\left(R_{I_t}^{(j-1)}\right)_{ii} = \begin{cases} (R^{(j-1)})_{ii}, & i \in I_t \\ 0, & \text{else} \end{cases}. \tag{9}$$

The set $I_t = \{i \in \{1, \ldots, e\} | y_i(t) \text{ is missing}\}$ is the index set of missing observations. Note that for missing observations the innovations $\mathbf{y}(t) - \hat{\mathbf{y}}(t)$ are set to zero by the proposed algorithm. Thus, for missing observations the variance of the observation error is set to the previous estimate. The off-diagonal entries of $R^{(j)}$ are set to zero due to the assumption of uncorrelated observation noise. For a nonlinear observation function G, the terms $E(G(\mathbf{X}(t))|\overline{\mathbf{y}}_N)$ and $Var(\mathbf{y}(t) - E(G(\mathbf{X}(t))|\overline{\mathbf{y}}_N))$ can be approximated by the unscented transform.

3.3. Statistical inference

For statistical inference on the connectivity, the dPC (Eichler 2005, Mader *et al* 2008) is used. The dPC $\pi_{ij}(u)$ of regions i and j at time lag u is defined as the correlation between $X_i(t)$ and $X_j(t-u)$ after removing the linear effects of $\overline{\mathbf{X}}(t)\backslash\{X_j(t-u)\}$ (Eichler 2005). A consistent estimator of the dPC is given by (Eichler 2006)

$$\hat{\pi}_{ij}(u) = \begin{cases} \dfrac{\hat{A}_{ij}(u)}{\sqrt{\hat{Q}_{ii}\hat{\tau}_{ij}(u)}}, & u > 0 \\ \dfrac{\hat{Q}_{ij}}{\sqrt{\hat{Q}_{ii}\hat{Q}_{jj}}}, & u = 0 \end{cases}, \tag{10}$$

where

$$\hat{\tau}_{ij}(u) = \hat{K}_{jj} + \sum_{v=1}^{u-1} \sum_{k,l} \hat{A}_{kj}(v)\hat{K}_{kl}\hat{A}_{lj}(v) + \frac{\hat{A}_{ij}(u)^2}{\hat{Q}_{ii}} \tag{11}$$

with $\hat{K} = \hat{Q}^{-1}$. The ML estimates of the parameters are asymptotically Gaussian distributed with zero mean and variance equal to the inverse information matrix. The resulting approximate α significance level for the absolute value of the estimated dPC ($|\hat{\pi}_{ij}(u)|$) is given by

$$\begin{cases} \Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right) \sqrt{\dfrac{1}{N} \dfrac{\widehat{VAR}(\hat{A}_{ij})}{\hat{Q}_{ii}\hat{\tau}_{ij}(u)}}, & u > 0 \\ \Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right) \sqrt{\dfrac{1}{N} \dfrac{\widehat{VAR}(\hat{Q}_{ij})}{\hat{Q}_{ii}\hat{Q}_{jj}}}, & u = 0 \end{cases}, \tag{12}$$

where $\widehat{VAR}(\hat{A}_{ij})$ is estimated by numerical evaluation of the inverse information matrix and Φ^{-1} is the inverse standard normal distribution. In this paper, the significance levels are Bonferroni corrected for multiple testing.

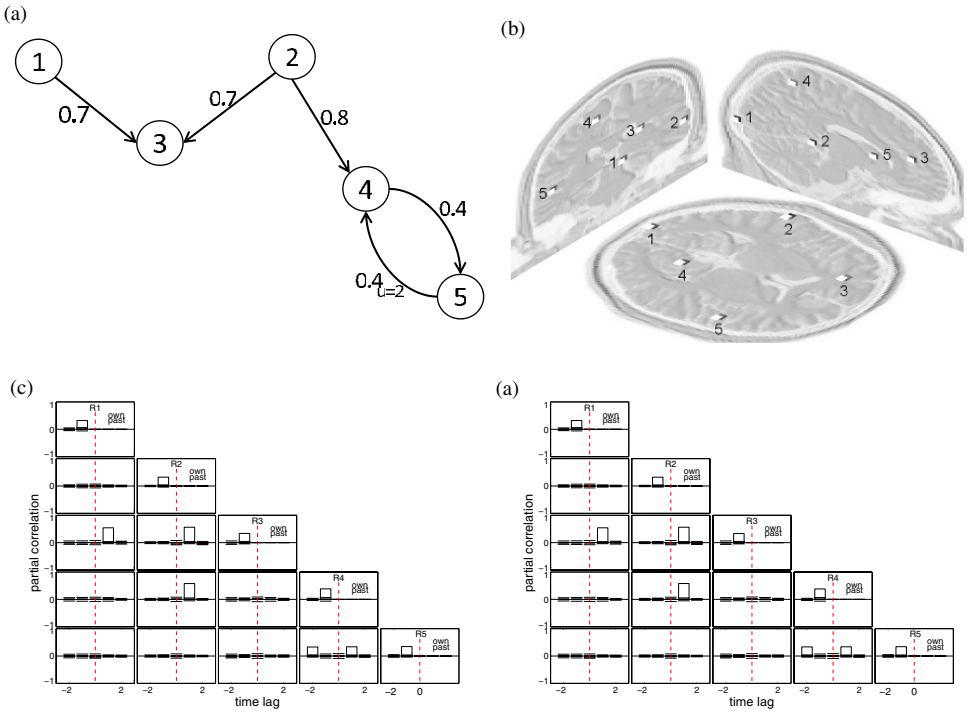

Figure 2. (a) Graphical representation of the simulated interaction structure. The arrows denote a directed interaction between two components. The numbers represent the VAR parameters with subscripts (u) indicating the time lag. Numbers without subscript denote an interaction at time lag 1. Each component has an additional influence on itself, which is not depicted. (b) Location of the regions of interest (white squares) for the simulation example without deep source. The regions are projected on the depicted slices to allow the visualization of all regions in a single graphic. All regions are close to the surface of the head. (c) Estimated dPC values and corresponding significance levels for the estimation without deep sources based on EEG data alone. Negative time lags in the coordinate system (i, j) denote an influence from the component i on the component j. Positive time lags show an influence in the opposite direction. At time lag zero instantaneous interactions are visualized. The significance levels are Bonferroni corrected to achieve a global significance level of 0.05. (d) Same as (c) but for the estimation based on EEG and fMRI data.

4. Results

In this section, the accuracy of the procedure for estimating effective connectivity is tested based on simulated data. For all simulation examples presented in this paper, a five-dimensional autoregressive process of order 2 is used to model the connectivity between five different brain regions. The parameter matrices of the AR[2] are specified as follows (see also figure 2(a)):

$$A(1) = \begin{pmatrix} 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0.4 & 0 & 0 & 0 \\ 0.7 & 0.7 & 0.4 & 0 & 0 \\ 0 & 0.8 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0.4 & 0.4 \end{pmatrix} \qquad A(2) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

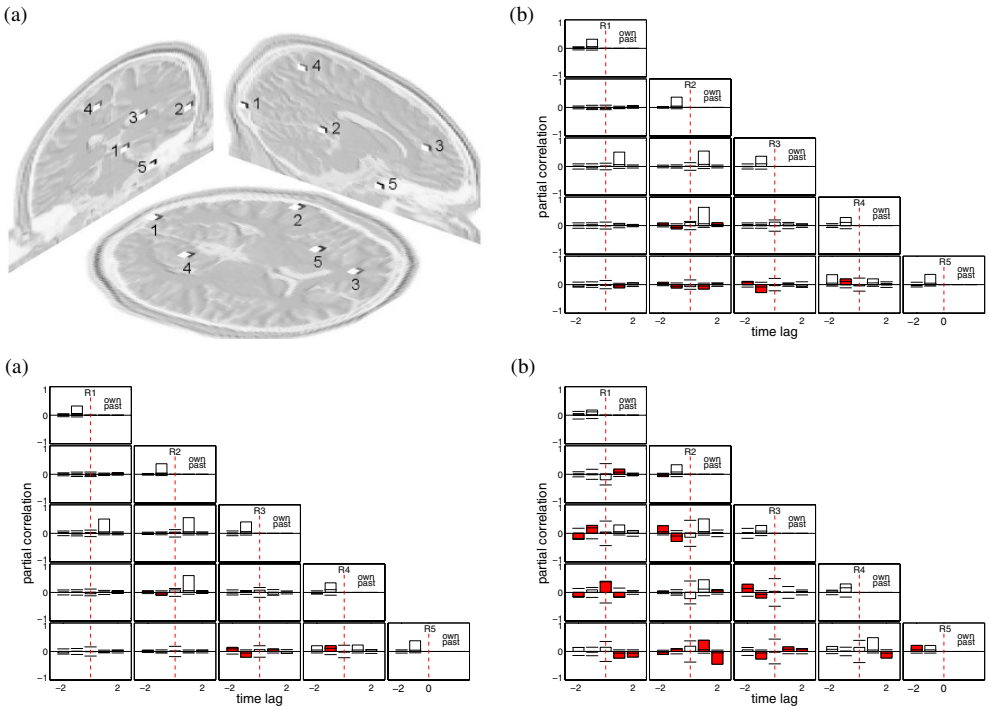

Figure 3. (a) Location of the regions of interest (white squares) in the simulated example with a deep source. The regions are projected on the depicted slices to allow the visualization of all regions in a single graphic. Region 5 depicts a deep source. All other regions are close to the surface of the head. (b) Estimated dPC values and corresponding significance levels for the estimation with a deep source based on EEG data alone. Negative time lags in the coordinate system (i, j) denote an influence from the component i on the component j. Positive time lags show an influence in the opposite direction. At time lag zero instantaneous interactions are visualized. The significance levels are Bonferroni corrected to achieve a global significance level of 0.05. False positive estimates are marked red. (c) Same as (b) but for the estimation based on EEG and fMRI data. (d) The same for the estimation based on fMRI data alone.

The covariance matrix Q equals the unity matrix. The covariance matrix R is set to

$$R = \begin{pmatrix} \alpha^{\mathrm{Mod1}} I_{e_{\mathrm{Mod1}}} & 0 & \cdots \\ 0 & \alpha^{\mathrm{Mod2}} I_{e_{\mathrm{Mod2}}} & \cdots \\ \vdots & \cdots & \ddots \end{pmatrix},$$

where $\alpha^{\mathrm{Mod}i}$ is a modality specific constant and $I_{e_{\mathrm{Mod}i}}$ is the unity matrix of dimension equal to the dimension of $Y^{\mathrm{Mod}i}(t)$. For each simulation, $N = 2000$ timesteps are simulated. For the estimation procedure, the order of the autoregressive process was set to the true order of 2.

The focus of this paper is the combination of EEG and fMRI data. Therefore, data are simulated according to the combined model based on EEG and fMRI data (section 2.3). The connectivity between brain regions is estimated in three different estimations, based solely on EEG data, solely on fMRI data as well as on both EEG and fMRI data. Furthermore, two different scenarios are compared differing in the location of the brain regions. In the first scenario, the brain regions are located close to the surface of the head and distant from each

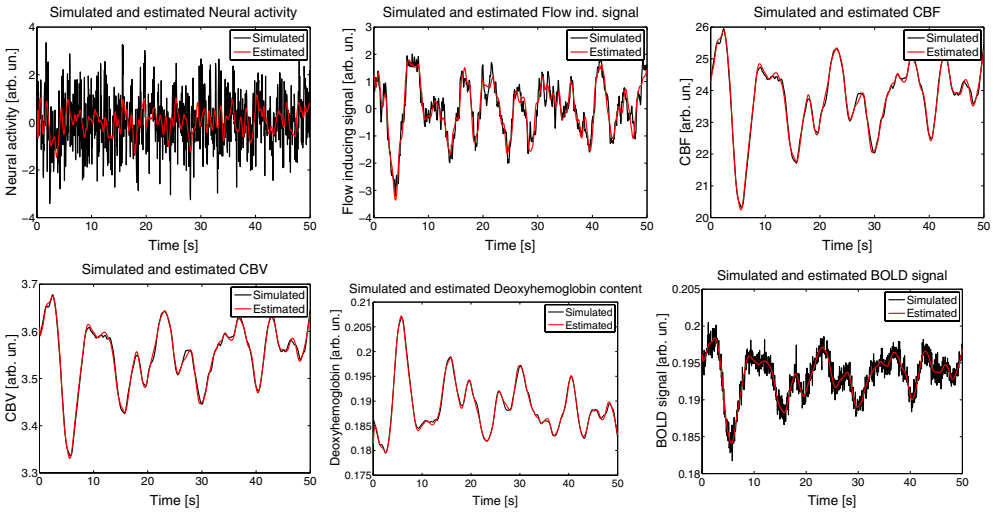

Figure 4. Simulated and estimated states and BOLD signal for the first region. The estimation was performed using fMRI data alone. The black curves depict the simulated values, and the red curves represent the estimated values. The simulated BOLD signal is depicted with additional measurement noise (bottom-right graphic). The estimated signal represents the BOLD signal without measurement noise. The term 'flow inducing signal' describes the derivative of the cerebral blood flow (CBF).

other (figure 2(b)). The second scenario is more challenging, in that one region is located deep in the brain (figure 3(a)). This results in a comparably weak EEG signal corresponding to this region.

For the first scenario, the results are accurate for the estimation based on EEG data alone as well as for the estimation based on EEG and fMRI data (figure 2(c) and (d)). The second scenario results in some false positive estimates for the estimation based on EEG data alone (figure 3(b)). The strongest false positives are those involving the deep region 5. For the estimation based on EEG and fMRI data the results are similar (figure 3(c)).

The third estimation is based on the fMRI data alone. Note that this estimation is not based on the linear fMRI model described in section 2.1, but on the nonlinear model used in the combined EEG/fMRI approach. The resulting dPC estimates show large deviations from the correct values (figure 3(d)). A large number of false positive test results are present. One possible reason for the incorrect dPC estimates is the temporal convolution inherent in the hemodynamic model. This may also explain the high uncertainty represented through the large significance levels in figure 3(d).

For a more detailed investigation of the effects of the hemodynamic model, the simulated and estimated states and observations for the estimation based on fMRI data are compared in the following (figure 4). The estimated CBF, CBV and deoxyhemoglobin content show only small differences to the corresponding simulated values (figure 4). The simulated and estimated BOLD signals are also very similar. Note that the simulated BOLD signal in figure 4 is shown with additional measurement noise. The estimated flow inducing signal, i.e. the derivative of the CBF, shows larger differences to the corresponding simulated values. Even larger differences are present between the estimated and simulated neural activity. The estimated neural activity appears to be a smoothed version of the simulated neural activity.

One reason for the inaccurately estimated neural activity may be the convolution effect of the hemodynamic model. It appears as if the convolution effect mainly affects the transformation from the neural activity to the CBF. This can be deduced since the simulated CBF is considerably smoother than the simulated neural activity, whereas the smoothness of the curves stays similar for the other transformations (figure 4). The estimation of the states based on the data is accurate as long as the simulated curve is sufficiently smooth. This observation, combined with the previous observation, confirms the hypothesis that the temporal convolution negatively affects the estimation results.

Another reason for the smooth estimated neural activity may be the following. The Kalman filter uses information up to the current instant of time. The BOLD response to the neural activity, however, is time delayed. Therefore, the fMRI data provide hardly any information for the Kalman filter. Most information is incorporated through the Kalman smoother. The Kalman smoother, in turn, usually provides smooth estimates (van der Merwe and Wan 2003). Thus, the delayed hemodynamic response in combination with the method used for the estimation of the neural activity may constitute a further reason for the inaccurate estimates.

5. Conclusion

The analysis of multimodal signals is a key challenge in nowadays research. Here, we present an idea based on state space models that due to its flexibility enables the analysis of various signals recorded from different modalities. As revealed by the simulation study, the fMRI poses the biggest challenge due to the BOLD signal. To what extent this limitation can be overcome is to be analyzed in further studies. The possible combination of other modalities such as, e.g., EEG and MEG using the described framework may also be very promising and has to be investigated in future studies. The advantage of a combination of EEG and MEG lies in the fact that both measurements have a high temporal resolution while still providing complementary information.

In summary, we have presented a rather general framework that enables the analysis of various types of signals. It is thereby not only possible to estimate the underlying activity of interest but also its Granger-causal interdependence structure. The estimation of effective connectivity, thus, becomes possible by combining the strengths of the various signal types.

Acknowledgments

This work was supported by the German Science Foundation (Ti315/4- 2) and the Excellence Initiative of the German Federal and State Governments. BS is indebted to the Landesstiftung Baden-Württemberg for the financial support of this research project by the Elite programme for Postdocs. OT was supported by BMBF grant no 01GW0730.

References

Baccala L A, Takahashi D Y and Sameshima K 2006 *Handbook of Time Series Analysis* ed B Schelter, M Winterhalder and J Timmer (New York: Wiley-VCH) pp 411–36

Cheung B L P, Riedner B, Tononi G and Van Veen B D 2009 State-space multivariate autoregressive models for estimation of cortical connectivity from EEG *31st Annu. Int. Conf. IEEE EMBS (Minneapolis, MN, USA, 2–6 Sept)* pp 61–4

Cheung B L P and Van Veen B D 2008 Estimation of cortical multivariate autoregressive models for EEG/MEG using an expectation-maximization algorithm *5th IEEE Int. Symp. on Biomedical Imaging: Macro to Nano* pp 1235–8

Deneux T and Faugeras O 2010 EEG–fMRI fusion of paradigm-free activity using Kalman filtering *Neural Comput*. **22** 906–48

Eichler M 2005 A graphical approach for evaluating effective connectivity in neural systems *Phil. Trans. Biol. Sci*. **360** 953–67

Eichler M 2006 *Handbook of Time Series Analysis* ed B Schelter, M Winterhalder and J Timmer (New York: Wiley-VCH) pp 335–72

Friston K J, Harrison L and Penny W 2003 Dynamic causal modelling *NeuroImage* **19** 1273–302

Galka A, Wong K F K, Ozaki T, Muhle H, Stephani U and Siniatchkin M 2010 Decomposition of neurological multivariate time series by state space modelling *Bull. Math. Biol*. **73** 285–324

Julier S J and Uhlmann J K 1997 A new extension of the Kalman filter to nonlinear systems *Int. Symp. Aerospace/Defense Sensing, Simulation and Controls* vol 3068 pp 182–93

Kandepu R, Foss B and Imsland L 2008 Applying the unscented Kalman filter for nonlinear state estimation *J. Process Control* **18** 753–68

Kolas S, Foss B A and Schei T S 2009 Constrained nonlinear state estimation based on the UKF approach *Comput. Chem. Eng*. **33** 1386–401

Limpiti T, Van Veen B D and Wakai R T 2006 Cortical patch basis model for spatially extended neural activity *IEEE Trans. Biomed. Eng*. **53** 1740–54

Mader W, Feess D, Lange R, Saur D, Glauche V, Weiller C, Timmer J and Schelter B 2008 On the detection of direct directed information flow in fMRI *IEEE J. Sel. Top. Signal Process*. **2** 965–74

Nunez P L and Srinivasan R 2006 *Electric Fields of the Brain: The Neurophysics of EEG* (Oxford: Oxford University Press)

Sarkka S 2008 Unscented Rauch–Tung–Striebel smoother *IEEE Trans. Autom. Control* **53** 845–9

Shumway R H and Stoffer D S 2006 *Time Series Analysis and Its Applications: With R Examples* (Berlin: Springer)

Supp G G, Schlögl A, Trujillo-Barreto N, Müller M M and Gruber T 2007 Directed cortical information flow during human object recognition: analyzing induced EEG gamma-band responses in brain's source space *PLOS ONE* **2** e684

van der Merwe R 2004 Sigma-point Kalman filters for probability inference in dynamic state-space models *PhD Thesis* Oregon Health and Science University

van der Merwe R and Wan E 2003 Sigma-point Kalman filters for probabilistic inference in dynamic state-space models *Proc. Workshop on Advances in Machine Learning*