Imperial College Press
www.icpress.co.uk

SPATIO-TEMPORAL CORRELATIONS FROM fMRI TIME SERIES BASED ON THE NN-ARx MODEL

J. BOSCH-BAYARD[*,†], J. RIERA-DIAZ[‡,§§], R. BISCAY-LIRIO[§],
K. F. K. WONG[†], A. GALKA[¶], O. YAMASHITA[‖],
N. SADATO[**], R. KAWASHIMA[‡], E. AUBERT-VAZQUEZ[*],
R. RODRIGUEZ-ROJAS[††], P. VALDES-SOSA[*],
F. MIWAKEICHI[‡‡] and T. OZAKI[†]

[*]Cuban Neuroscience Center, Havana, Cuba

[†]The Institute of Statistical Mathematics, Tokyo, Japan

[‡]IDAC, Tohoku University, Sendai, Japan

[§]DEUV-CIMFAV, Universidad de Valparaiso, Chile

[¶]Institute of Experimental and Applied Physics
University of Kiel, Germany
Department of Neurology, University of Kiel, Germany

[‖]ATR Brain Information Communication
Research Laboratory Group, Kyoto, Japan

[**]National Institute for Physiological Science, Okazaki, Japan

[††]Center for Neurological Restoration, Havana, Cuba

[‡‡]Department of Statistical Modeling
The Institute of Statistical Mathematics, Japan
[‡]http://www.idac.tohoku.ac.jp/

For the purpose of statistical characterization of the spatio-temporal correlation structure of brain functioning from high-dimensional fMRI time series, we introduce an innovation approach. This is based on whitening the data by the Nearest-Neighbors AutoRegressive model with external inputs (NN-ARx). Correlations between the resulting innovations are an extension of the usual correlations, in which mean-correction is carried out by the dynamic NN-ARx model instead of the static, standard linear model for fMRI time series. Measures of dependencies between regions are defined by summarizing correlations among innovations at several time lags over pairs of voxels. Such summarization does not involve averaging the data over each region, which prevents loss of information in case of non-homogeneous regions. Statistical tests based on these measures are elaborated, which allow for assessing the correlation structure in search of connectivity. Results of application of the NN-ARx approach to fMRI data recorded in visual stimuli experiments are shown. Finally, a number of issues related with its potential and limitations are commented.

Keywords: fMRI; time series; NN-ARx; causality; AIC; connectivity; whitening; innovations.

[§§]Corresponding author.

382   *Bosch-Bayard et al.*

1. Introduction

The method of functional Magnetic Resonance Imaging (fMRI) using Blood Oxygen Level-Dependent (BOLD) contrast has high resolution to localize hemodynamic activation in space. On this ground, a variety of statistical methods have been proposed to detect activation from fMRI data, which have been widely applied in experimental studies where the subjects are involved in the performance of sensory, motor or cognitive tasks [10–12].

In recent years, there have been increasing interest in using fMRI time series as a source of relevant information for understanding the interactions between brain regions under task-related activation. However, a number of problems arose when attempting to accomplish this task. Some of the most critical ones are the following:

(a) The dimensionality problem that comes from the high spatial resolution of fMRI images. As a solution, special reference voxels or regions of interest are chosen, and the data within each region is summarized by its average or first principal component [27].

(b) The aliasing problem has not been paid much attention to in the past, although it may easily cause confusion in the study of brain connectivity. The concept of brain connectivity refers to dynamic activity at the level of neural events, while fMRI data represents slowly changing hemodynamic BOLD signals measured with a sampling time of a few seconds, which only indirectly reflects the underlying fast neural dynamics on the millisecond scale. This may be properly overcome only by introducing special assumptions about the dynamics [42].

(c) The elucidation of brain connectivity from fMRI time series also faces the fundamental problem of the limits involved in inferring causal relations from empirical data. In general, correlation patterns of the output of a system cannot be conclusively interpreted in terms of causal relations between its components without specific information coming from previous theoretical knowledge (appropriate model assumptions) or designed experimental manipulations (particularly, through the input to the system). This methodological issue has been well-known for many years in some applied fields such as statistical experimental design and system identification, and it has been recently made conceptually clearer in terms of causal graphs [36]. Its relevance for the analysis of brain connectivity has begun to be explicitly dealt with by several authors. In particular, the acknowledged terminologies of functional and effective connectivity [9, 21] have contributed to draw attention to the distinction between observed and causal dependencies in connectivity analysis. Difficulties related to observationally equivalent causal models and spurious causality due to unobserved variables have been pointed [7]. The importance of elaborating models that incorporate specific knowledge about causal relations at the neural level have also been emphasized [14, 42, 47]. Furthermore, new possibilities in experimental manipulation of the functioning brain to look for causal interactions are

being initiated, e.g., the "perturb-and-measure" approach based on combining transcranial magnetic stimulation with fMRI [35].

Several approaches have been proposed for studying brain connectivity from fMRI time series. They have different advantages and limitations in dealing with the varied problems involved in this task, as those just described. Standard connectivity analysis by means of correlations [26, 44, 50] has shortcomings such that it is not dynamic, thus throwing away temporal information. In particular, only zero lag, instantaneous correlations are usually taken into consideration. Furthermore, in order to compute the correlations, mean-correction of the data (which involves the estimation of the hemodynamic response function) is carried out on the basis of the standard linear model for fMRI time series [12]. This model is considered to be static as its stochastic input (i.e., the error or perturbation term of the model) at a given instant does not affect the future of its output signal. Another non-dynamical approach is structural equation modeling [3, 30]. Its peculiar advantage is allowing for the representation of causal influences between brain regions, but their graph must be pre-specified by the analyst. Thus, it is not very useful when such knowledge is unavailable. In contrast, vector autoregressive modeling [5, 18, 20, 27, 49, 52] is an approach that does not assume previous knowledge of the connectivity graph, and explicitly makes use of temporal information. In this framework, the assessment of causal influences is carried out in terms of (temporal) predictability; namely, directed measures of influence related to Granger causality or Akaike's relative power contribution. However, until now this approach is not fully dynamic in respect to correction by the mean function of the data. Indeed, the latter is estimated and removed according to the static, standard linear model for fMRI data, as part of a preprocessing step disentangled from the autoregressive model. Finally, another recent approach that has been suggested is dynamic causal modeling [14, 47]. This is based on models composed of (i) a dynamic model for the activities of a set of neural masses in response to controlled external stimulation (with structural model parameters coding causal interactions between the neural masses) and (ii) a dynamic model for the biophysical translation of the neural activities into recorded signals. Regrettably, specification and implementation of these models require intensive computation and substantial neurophysiologic knowledge (which is specific for each task-related experiment and may be unavailable). These challenges must be faced to make it feasible for use with large systems of neural masses and diverse experimental settings.

In the present paper, we confine ourselves to the problem of statistical characterization of the spatio-temporal correlation structure of fMRI data in the original full-dimensional time series. We introduce a new practical procedure to be employed in studies of brain connectivity that is based on Nearest-Neighbors AutoRegressive modeling with external input, which subsequently will be referred to as NN-ARx approach.

The NN-ARx model was introduced in [41] as a discrete-time linear approximation to a stochastic extension [40] of the deterministic continuous-time hemodynamic

384   *Bosch-Bayard et al.*

model described in [13]. It is therefore a dynamic generalization of the standard linear model for fMRI time series [12]. It includes not only additive terms corresponding to a deterministic trend, a linear filter of the stimulus signal and a perturbation noise (likewise the standard linear model) but further autoregressive terms that reflect contributions at each voxel from its own past and the past of its neighboring voxels.

The coefficients of the NN-ARx model explain local spatio-temporal dependencies in the data. They allow us to transform the observed time series into innovations (or residuals) that have zero mean and a correlation structure in which dependencies between neighboring voxels are removed while dependencies between non-neighboring voxels are retained. This is the starting point of the NN-ARx approach that we introduce in the present paper. On this basis, we make use of the correlations between pairs of innovations at voxels in different brain regions to construct measures that summarize interregional influences. Such a summarization is carried out without taking the average of the data within each region, thus preventing loss of information in non-homogeneous regions. Both instantaneous and lagged correlations are taken into consideration in order to retain dynamic information on the direction of connections. Furthermore, we introduce statistical tests to detect significant interregional connections through these measures by resampling methods.

As this new approach computes correlations on the basis of mean-correction of the fMRI signals by a dynamic model that comprises the static, usual linear model as a particular instance, it may be deemed as a dynamic generalization of the standard connectivity analysis by means of correlations [26, 44]. Also likewise the conventional correlation analysis, detected connections should be primarily thought of as functional connections (i.e., observed dependencies), and possible interpretations of them as also effective (i.e., causal) connections should remain tentative though plausible while waiting for further experimental corroboration and neurophysiologic understanding. Note that the stipulated approach has several distinctive features that include mean-correction by the dynamic NN-ARx model and computation of measures of interregional influences that does not involve averaging the data within regions.

The introduced method was applied to two data sets: a group of subjects performing a visual task, and a group of blind subjects during a tactile discrimination task. The results obtained show the feasibility and flexibility of the method, and are consistent with previous reports describing linking pathways involved in these tasks.

The organization of the paper is as follows: The NN-ARx approach is described in Sec. 2. Section 3 presents and discusses results of the application of the NN-ARx approach to the data recorded from the subjects under two task-related experimental settings. Finally, in Sec. 4 various issues related to the potential and limitations of the introduced approach, as well as possible extensions, are explored.

2. Method

2.1. *NN-ARx modeling and estimation*

fMRI data consists of high dimensional spatio-temporal measurements of the hemo-dynamic activity in the brain. The data may be represented by a vector time series y_t whose components y_t^v are the magnitudes of the recorded signals at time $t = 1, \ldots, T$ in the voxel with spatial location $v = (i, j, k)$.

The NN-ARx model for fMRI data [41] has the general form

$$y_t^v = \mu_t^v + \sum_{k=1}^{p} \phi_k^v y_{t-k}^v + X^v \xi_{t-\Delta}^v + \sum_{k=0}^{r} \theta_k^v s_{t-k-d} + \varepsilon_t^v, \tag{1}$$

which is composed of the following terms:

(a) $\mu_t^v = \sum_{k=0}^{\delta} \gamma_k^v t^k$ is a polynomial ***trend term*** intended to gather the trend or potential drift of the recorded signal due to factors unrelated to brain activity (such as instability in the scanner, motion artifacts, slow variations of blood pressure, etc.). Indeed, Gössl *et al.* [19] have shown that polynomial models are a simple approach that achieves results comparable with more sophisticated ones for removing such nuisance sources of variation in the data;

(b) $\sum_{k=1}^{p} \phi_k^v y_{t-k}^v$ is the ***autoregressive term*** that describes the hemodynamics linear model at the voxel v as function of its own past activity;

(c) $X^v \xi_{t-\Delta}^v$ is a linear ***neighborhood term*** that represents the contribution of past activities of the neighboring voxels to the activity at voxel v at time t (i.e., y_t^v); more precisely, $\xi_{t-\Delta}^v = \{y_{t-\Delta}^{v'}, v' \in \Omega_v\}$ is the column vector of past activities of voxels in the region Ω_v that contains the nearest neighbors of voxel v, and $X^v = \{\chi_v^{v'}, v' \in \Omega_v\}$ is a row vector of (time-independent) coefficients that summarizes the anisotropic properties of the local vascular correlations;

(d) $\sum_{k=0}^{r} \theta_k^v s_{t-k-d}$ is the ***stimulus term*** consisting of a stimulus sequence weighted by neuronal activity effectiveness; and

(e) ε_t^v is a Gaussian ***perturbation term*** with zero mean and variance σ_v^2, which represents the stochastic input due to fluctuations of the neural activity around its mean value at each instant.

Here, d and Δ are time lags that reflect delays of the effects of the stimu-lus and of the neighboring voxels, respectively. The parameters of the model are $\{\phi_k^v, X^v, \theta_h^v, \sigma_v, \gamma_l^v\}$ where $k = 1, 2, \ldots, p; h = 1, 2, \ldots, r$ and $l = 1, 2, \ldots, \delta$. The model selection consists of determining both the model orders p, r, δ and the delays d, Δ. These magnitudes, related to the complexity of the dynamics, are called global parameters and are found in the vector $\Lambda = (p, r, \delta, d, \Delta)$. Details on the determina-tion of these parameters are given in Sec. 2.5 below.

From a more empirical modeling view, NN-ARx may also be motivated as a simple linear model accounting for the local spatial and temporal dependencies of the data, thus allowing the transformation of the data into innovations. Given the

past values $y_{t-1}, y_{t-2}, \ldots$ of the recorded signals at all voxels and the past values $S_{t-d}, S_{t-d-1}, \ldots$ of the stimulus signal, the innovation or prediction error at a voxel v is defined by

$$\zeta_t^v = y_t^v - E(y_{t-1}^v / y_{t-1}, y_{t-2}, \ldots, S_{t-d}, S_{t-d-1,\ldots}),$$

where $E(./.)$ denotes conditional expectation. By construction, ζ_t^v is a white (i.e., uncorrelated) temporal sequence that is also uncorrelated with the past data $y_{t-1}, y_{t-2}, \ldots$. A linear approximation to these innovations is provided by the residuals of the NN-ARx model,

$$\varepsilon_t^v = y_t^v - \left(\mu_t^v + \sum_{k=1}^{p} \phi_k^v y_{t-k}^v + X^v \xi_{t-\Delta}^v + \sum_{k=0}^{r} \theta_k^v s_{t-k-d} \right),$$

in which the conditional mean is approximated by

$$E[y_t^v / y_{t-1}^v, \ldots, y_{t-p}^v, \xi_{t-\Delta}^{(v)}, S_{t-d}, S_{t-d-1,\ldots}, S_{t-d-r}]$$

$$= y_{t/t-1}^v = \mu_t^v + \sum_{k=1}^{p} \phi_k^v y_{t-k}^v + X^v \xi_{t-\Delta}^v + \sum_{k=0}^{r} \theta_k^v s_{t-k-d}.$$

In other words, according to this model, the one-step-ahead prediction for a voxel at position v is given by the addition of (a) a trend component, (b) a linear combination of the own past of the voxel, (c) a linear combination of the past of its nearest neighboring voxels, and (d) a linear combination of the past of the stimulus.

Note that this model reduces to the standard linear model for fMRI time series, as described in Firston *et al.* [12], if the terms (b) and (c) are disregarded. Furthermore, the sequence of parameters θ_k^v is conventionally called the hemodynamic response function (HRF) to the stimulus s_t. Thus, the NN-ARx model is an extension of the standard linear model resulting from the inclusion of further dynamic terms, namely (b) and (c).

Model parameters $\{\phi_k^v, X^v, \theta_k^v, \sigma_v, \gamma_k^v\}$ at each voxel are estimated by the least squares method. Model selection is carried out by minimizing the Akaike Information Criterion (AIC) [2, 45].

In order to further simplify the local spatial correlation structure of the residuals or innovations ε_t^v it is convenient to apply a Laplacian pre-filtering; i.e., to transform the data by the (discrete) Laplacian operator L, defined by

$$y_t^v = L x_t^v = x_t^v - \frac{1}{G} \sum_{v' \in \Omega_v} x_t^{v'},$$

for any random field x_t^v, where G is the number of neighboring voxels in Ω_v. Whitening by Laplacian filtering represents a simple but useful way of characterizing the spatially homogeneous instantaneous correlation between the noises of neighboring voxels, and was used for the first time by Galka *et al.* [16] and Yamashita *et al.* [51] for the purpose of estimating dynamical inverse solutions from EEG time series [1]. In particular, spatial correlations due to harmonic random components of the data are removed by multiplying by the Laplacian because, by definition, the Laplacian

of any harmonic function is zero [25]. In the present context, it allows for whitening spatial correlations unfavorably introduced by the usual application of Gaussian kernels in the fMRI analysis. By fitting the NN-ARx model both with and without applying the Laplacian operator to the same fMRI data set, and comparing the resulting values of the AIC, it can be confirmed that employing the Laplacian provides a superior modeling. An additional advantage of Laplacian pre-filtering, as a consequence of its whitening effect, is the improved efficiency of the least squares estimators of the model parameters.

More details about the biophysical basis of the NN-ARx model and the procedures involved in the estimation of its parameters, as well as validations with simulated and real data sets, may be found in Riera *et al.* [41].

2.2. *NN-ARx-based voxel-wise correlations*

Conventional correlations between voxels in fMRI analysis [26, 44] are correlations between the residuals resulting from mean-correction according to the standard linear model for fMRI time series [12]. The immediate extension of this procedure to the NN-ARx model leads to the computation of the correlation between the two voxels v and w as the correlation between the residuals obtained by removing the conditional mean according to the NN-ARx model, i.e., the correlation between the innovations or prediction errors ζ_t^v and ζ_t^w:

$$R_0^{v,w} = corr(\zeta_t^v, \zeta_t^w).$$

As mentioned above, correction by the conditional mean $y_{t/t-1}^v$ tends to eliminate local dependencies of each voxel with respect to its own past and the past of its neighbors, but retains dependencies on non-neighboring voxels. The Laplacian operator, applied as a preprocessing step, was also designed to remove instantaneous spatial correlations only between neighboring voxels, but not between distant voxels. Hence, the correlations between innovations of distant voxels contain information on the structure of influences between brain regions. This will be used to construct measures of interregional dependencies in the next subsection.

Note also that the zero lag, instantaneous correlation $R_0^{v,w}$ plausibly reflects underlying interactions from past to present activities at the neural level. Indeed, the neuronal electrical signals transfer much faster than hemodynamic activities and may reach remote voxels within a millisecond, i.e., almost instantaneous with respect to the typical time scale of the hemodynamics. Furthermore, two distant voxels with a special neuronal connection may simultaneously receive a large signal input coming from a common source, whereby the prediction errors of these two voxels become strongly correlated.

In order to capture additional information on the temporal arrow of the dependencies between two voxels v and w we also define directed or lagged correlations between innovations, i.e.,

$$R_l^{v,w} = corr(\zeta_{t+l}^v, \zeta_t^w),$$

where the time lag l can be negative or positive. For $l < 0$, $R_l^{v,w}$ represents the correlation of the past activity (at time $t + l < t$) at the voxel v and the activity at the present time t at voxel w. This is therefore a directed correlation that reflects the observed influence of the past at the voxel v, and on the present at voxel w. Similar interpretations have the lagged correlations $R_l^{v,w}$ for $l > 0$, just by interchanging past and present at the voxels.

The NN-ARx-based correlations $R_l^{v,w}$ can be estimated by the sample correlation between the time series residuals ζ_{t+l}^v and ζ_t^w ($t = 1, \ldots, T$) obtained from fitting the NN-ARx models at the voxels v and w.

2.3. *Measures of dependencies between regions*

Correlations between past and present data at different regions offer a tool easy to compute and interpret for describing directed interregional influences. The usual way of determining interregional correlations is based on summarizing the recorded signal within each region by its average or first eigenvariate [44]. However, depending on the size and localization of the regions, and on the experimental setting, each region may not represent a functionally homogeneous neural mass but may involve several activities with each one interacting with different activities in subsets of other regions. In such a case, the summarization by averaging or first eigenvariate could fail to suitably represent each region. This difficulty can be overcome by successive refinements of the definition of the regions into smaller, localized ones. We, here, introduce an alternative approach that does not require such detailed refinements.

Let M regions of interest in the brain be specified: $\Pi^V, V = 1, \ldots, M$. Consider an arbitrary pair of such regions, say Π^V and Π^W. For each time lag $l = -L, \ldots, -1, 0, 1, \ldots, L$, ($L$ being the maximum lag to be considered), let $R_l^{v,w}$ be the correlation between the prediction errors ζ_t^v and ζ_t^w, as defined above, where v ranges over the voxels in Π^V, and w over the voxels in Π^W. With the intention of capturing both negative and positive correlations, we define the squared correlations,

$$S_l^{v,w} = (R_l^{v,w})^2.$$

In order to summarize the dependencies between voxels in the two regions Π^V and Π^W at a given lag l, a number of measures can be considered on the basis of the squared correlations $S_l^{v,w}$, such as their mean or median. A robust summarization is given by the magnitude:

$U_l^{V,W} = $ upper 90th percentile of the values of $S_l^{v,w}$ over $v \in \Pi^V$ and $w \in \Pi^W$.

Likewise the mean, $U_l^{V,W}$ takes a large value if most of the squared correlations $S_l^{v,w}$ are high, i.e., if practically any voxel in one region is highly correlated with any voxel in the other region. However, an advantage of the measure $U_l^{V,W}$ is that it also shows a high value when there are strong interactions between the regions with some subsets being uncorrelated. This allows for capturing dependencies between regions even in cases in which they are not functionally homogeneous. A further robust

benefit of this measure is that it is not affected by a small fraction of extremely high correlations that may appear when the segmentation procedure used to demarcate the regions results in a number of misclassified voxels.

Further summarization over time lags can be obtained by averaging the measures $U_l^{V,W}$ over positive and negative lags:

$$U_+^{V,W} = \sqrt{\frac{1}{L+1} \sum_{l=0}^{L} U_l^{V,W}},$$

$$U_-^{V,W} = \sqrt{\frac{1}{L+1} \sum_{l=-L}^{0} U_l^{V,W}}.$$

That is, $U_+^{V,W}$ reflects correlations between the past activity in the region Π^V and the future activity in the region Π^W over different lags, while $U_-^{V,W}$ summarizes correlations between the past of Π^W and the future of Π^V. The square roots in the previous formulae are used to facilitate interpretation, since they bring back the measures from squares to the correlation scale.

Note that, by definition, $U_-^{V,W} = U_+^{W,V}$, hence the analysis can be reduced to the measures $U_+^{V,W}$.

Due to the fact that the sampling time in fMRI recordings is large (normally about 3 seconds), it is not necessary to use high lag orders for the purpose of fMRI analysis. In practice, we set the maximum lag to be $L = 1$.

A simple simulation is used here to illustrate that this approach preserves correlations between regions much better than simply computing correlations between averages over regions. Consider three regions R1, R2 and R3, each one composed of 25 voxels. The voxels in regions R1 and R3 are divided in three subsets: R1a, R1b, R1c, R3a, R3b and R3c (see also the last column of the first row in Fig. 1). The time series for voxels in R2, R1c, R3a, R3b and R3c are independent standard Gaussian white noises, while the time series at voxels in regions R1a and R1b are generated according to the model:

$$X_{R1a}(t) = 0.75 * X_{R3a}(t-1) + \omega_a(t),$$
$$X_{R1b}(t) = -0.75 * X_{R3b}(t-1) + \omega_b(t),$$

where $X_-(t)$ represents the time series at time t for voxels in the specified subset, and $\omega_a(t)$ and $\omega_b(t)$ are independent standard Gaussian white noises. Thus, there are high correlations between R1a and the past of R3a, as well between R1b and the past of R3b, but with opposing signs. As can be seen in the first row of Fig. 1, sample averages within regions show very low (non-significant) correlations between them. On the contrary, the alternative approach of measuring dependencies between regions introduced in this paper correctly captures the high correlation between R1 and R3 with lag -1.
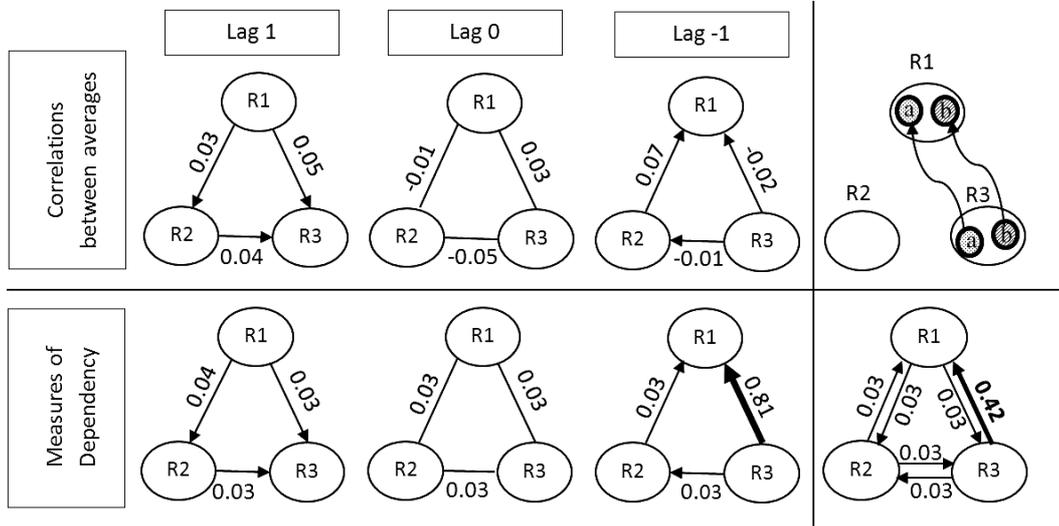
390   *Bosch-Bayard et al.*

Fig. 1.   Correlation results for simulated time series at three regions R1, R2 and R3, according to a model that induces high dependencies with different signs between subsets of R1 and R3 (see left upper corner of this figure, and also text for details). First row: correlations between the averages of the regions for the time lag 1, 0 and -1. Second row: measures of dependencies $U_l^{V,W}$ between the regions for lags 1, 0 and -1 (first three columns), and also the measures $U_+^{V,W}$ and $U_-^{V,W}$ (last column). Note that the latter shows high dependency between regions R1 and R3 for Lag -1 ($U_{-1}^{R1,R3} = 0.81$), and also $U_-^{R1,R3} = 0.42$, while standard correlations between averages (first row) fail to capture such a strong dependency between R1 and R3.

It is worth noting that the introduced measures of dependencies can be computed on the basis of the prediction errors, not only of the NN-ARx model but also of any model for fMRI time series.

2.4. *Statistical tests for interregional dependency graphs*

The measures of directed interregional dependencies $U_+^{V,W}$ over pairs of different regions $(V, W = 1, \ldots, M)$ determine a dependency or connectivity graph. The nodes of this graph are the given regions, and there is a directed edge or arrow from the region Π^V towards the region Π^W if the influence of the past of Π^V on the future of Π^W, as measured by $U_+^{V,W}$, is different from zero.

In practice, only estimates $\widehat{U}_+^{V,W}$ of the measures $U_+^{V,W}$ are available, thus there is a need for constructing statistical tests to detect significant connections. For simplicity, denote $U_+^{V,W}$ by $U(V, W)$. We will first describe a test to detect non-null connections ($U(V, W) \neq 0$) for the one sample problem corresponding to the data provided by one subject. Since the measures of interest are quite complex and their distributions are unknown, a resampling or parametric bootstrap approach, which entails intensive simulations, will be followed (see Bullmore *et al.* [4] and references therein for a wide variety of other simulation-based statistical tests for EEG and fMRI data analysis). Indeed, the main idea in constructing resampling tests is to

simulate data sets under the null hypothesis. In the present case, this is the hypothesis of zero correlations between voxels in different regions.

The main steps of the testing procedure can be described as follows:

(a) Fit the NN-ARx model to the original data. From this, obtain estimates $\{\widehat{\phi}_k^v, \widehat{X}^v, \widehat{\theta}_k^v, \widehat{\sigma}_v, \widehat{\gamma}_k^v\}$ of the model parameters, estimates $\widehat{\Sigma}^V$ of the covariance matrix of the vector of residuals $\zeta_t^V = (\zeta_t^v : v \in \Pi^V)$ at each region Π^V, and estimates $\widehat{U}(V,W)$ of the connectivity measure for each pair of regions. Also, compute the maximum $\widehat{U}$ of the quantities $\widehat{U}(V,W)$ for all pairs of different regions. (For practical reasons, if the regions are very large, the estimates of the high dimensional matrices $\widehat{\Sigma}^V$ are based on the singular value decomposition of the sample covariance matrix of the vector of residuals, retaining only a number of principal components.)

(b) Repeat for $b = 1, \ldots, B$ (number of bootstrap simulations, $B > 100$) the following steps:

  (i) Generate independent Gaussian vectors $\zeta_t^{V,b} = (\zeta_t^{v,b} : v \in \Pi^V)$ with zero means and covariance matrices $\widehat{\Sigma}^V$, for $t = 1, \ldots, T$, $V = 1, \ldots, M$.

  (ii) Using the values of the model parameters $\{\widehat{\phi}_k^v, \widehat{X}^v, \widehat{\theta}_k^v, \widehat{\sigma}_v, \widehat{\gamma}_k^v\}$ obtained in step (a) and the bootstrap time series residuals $\zeta_t^{v,b}$ as model errors, compute (according to the corresponding NN-ARx model) a bootstrap data set $y_t^{v,b}$.

  (iii) Repeat the computations of the step (a) but using the bootstrap sample $y_t^{v,b}$ as data. Denote the resulting value the maximum measure by $\widehat{U}^b$.

(c) Compute the upper 95% percentile $C_{0.95}$ of the values $\widehat{U}^b, b = 1, \ldots, B$. This threshold is then used for testing the null hypothesis of non-connectivity between pairs of regions. That is, the hypothesis of lack of connection from a region Π^V towards a region Π^W, i.e. $U(V,W) = 0$, is rejected at the 5% significance level if the observed value $\widehat{U}(V,W)$ is greater than $C_{0.95}$.

A similar resampling procedure can be used for testing connectivity in one sample problem corresponding to a group of subjects. Specifically, the only necessary modification is carrying out the step (a) above for each subject, and then averaging the resulting measures of dependencies over the number of subjects.

A similar test may also be elaborated for two sample problems, in which the aim is to compare the connectivity patterns of two groups of subjects or one group of subjects under two different experimental conditions.

The testing procedure just described may also be applied to other measures $U(V,W)$ of dependency between regions; for instance, the measure $U_l^{V,W}$ that summarizes the correlations between regions for a particular lag l.

Also note that the main computational burden of such tests based on simulations is the repeated generation of random normal variables, because other computations involved are just correlations.

392   *Bosch-Bayard et al.*

2.5.   *Algorithmic summary of the main steps in connectivity analysis based on the NN-ARx model*

The diagram in Fig. 2 summarizes the sequence of steps that we followed for data processing, analysis and visualization in exploring connectivity on the basis of the correlation structure revealed by the NN-ARx model.

2.5.1.   *Data preprocessing*

Before statistical modeling of the fMRI time series, the individual fMRI images were realigned in order to remove movement-related artifacts, and the slice timing was adjusted to that of the middle slice. The images were then smoothed.

Taking into consideration that the NN-ARx is a dynamic model, we try to preserve the dynamics of the data by applying the least possible amount of preprocessing that may affect it. We avoid preprocessing procedures such as normalization or co-registration to the anatomical image. Instead, to allow for the possibility of doing inter-subject statistics or referring the results to a standard brain, we normalized the results (model parameters and innovations for all voxels) to the standard space provided by the Montreal Neurological Institute [8].

2.5.2.   *NN-ARx model selection and fitting*

Model selection consists in the estimation of the global model parameters Λ, namely the model orders and delays p, r, δ, d, Δ (see Sec. 2.1). It is a voxelwise task that was
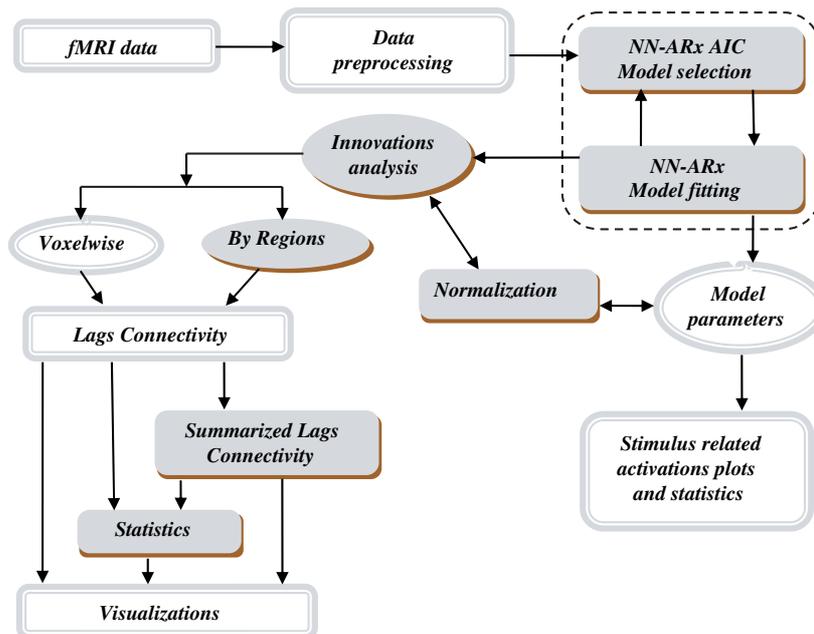

Fig. 2.   Sequence of processing steps for connectivity analysis based on the NN-ARx model.

accomplished by minimizing the AIC with respect to these parameters over specified ranges. The latter is set according to the user expertise; but a short range (say 1 to 4) is usually appropriate in practice. For the purpose of present work and based on our previous experiences, we fixed the delay of the neighboring voxels Δ to 1, thus, obviating any estimation of this parameter.

2.5.3. *Interregional connectivity graphs*

While voxelwise connectivity analysis will be the subject of future works, in the present paper we focus on interregional connectivity analysis. For this purpose, we divided the brain into regions using segmentation of the gray matter of the brain provided by the Montreal Neurological Institute. Specifically, we adopted the one that divides the brain in 71 regions (see details in Appendix A).

Lagged correlations based on the fitted NN-ARx model were computed, as well as their summarization through the directed measures of connectivity between regions, as described in Sec. 2.3. Then, statistically significant connections were detected by the tests introduced in Sec. 2.4.

Finally, 3D graphic tools were used to visualize the connectivity graph. They aid the analyst to explore the topology and strength of the correlation structure, and to acquire a global view of its neural distribution.

3. Results and Discussion

We analyzed two data sets: first, a group of subjects performing a simple visual task; second, a group of blind subjects during a tactile discrimination task. The experimental settings are described below. For reference, the first data set is the same one that (for other purposes) was studied in Riera *et al.* [41].

For each experimental data set we followed the processing steps described in Sec. 2.5. That is, for connectivity analysis we used the measurement of interregional connectivity $U_+^{V,W}$ as defined in Sec. 2.3, which summarizes lagged correlations between regions on the basis of NN-ARx modeling. The brain was divided into 71 regions using segmentation of the gray matter of the brain provided by the Montreal Neurological Institute. The connectivity measures all 71 regions, resulting in a 71×71 connectivity matrix for each subject. For each group of subjects, we then applied the statistical test described in Sec. 2.4.

3.1. *Visual experiment*

Visual paradigm: A 3-T scanner (VP, General Electric, Milwaukee, WI) was used in this experiment to collect the visual stimulus data. Ten volunteers (5 males and 5 females) aged 25–43 years were used in the visual paradigm, consisting of 3 blocks of 30 seconds checkerboard visual stimulus and 30 seconds of control condition (starting from task condition). During the task condition, the checkerboard

was intermittently presented at a frequency of 8 Hz. Tight but comfortable foam padding was placed around the subject's head to minimize head movement.

fMRI parameters: Inter-scan interval TR = 3 seconds. Each volume consisted of 36 slices from the bottom to the top of the head, with a voxel size of 3.44 × 3.44 mm^2 in plane, a slice thickness of 3.5 mm and a 0.5 mm gap covering the whole brain. T2*-weighted, gradient echo, echo planar imaging (EPI) sequences (TE = 30 milliseconds, FOV = 22 cm).

Parameters of scanner for anatomical reference: T2*-weighted, 2D-fast spin echo sequence (with parameters of FA = 90 degree, TR = 6000 milliseconds and TE = 70 milliseconds) consisting of 112 trans-axial slices, with slice thickness 1.5 mm, and pixel size was 0.859 × 0.859 mm^2.

Figure 3 shows some of the plots and statistics used to assess model goodness of fit for different models (rows of this Figure): the model with only the stimulus term (A); the model with stimulus and AR terms (B); and the complete NN-ARx model with stimulus, autoregressive and neighborhood terms (C). It can be observed that for no model, the Gaussianity of the prediction errors is rejected but AIC values and the degree of whiteness of residuals notably improve as more terms are included.
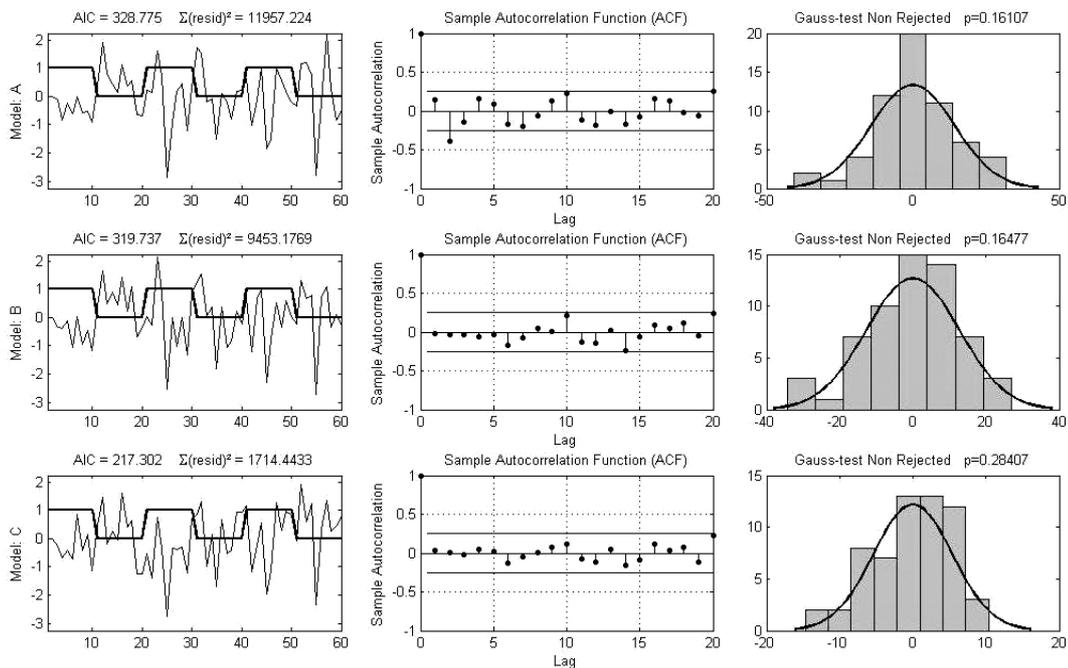
Fig. 3.   Plots and statistics used to assess goodness of fit for different models for a selected voxel at the calcarine sulcus. Columns from left to right: first, plots of the stimulus (bold step curve) and prediction errors signals, and values of the AIC, sum of residual squares; second, sample autocorrelation; third, histograms and Bera-Jarque test for Gaussianity. Rows, from upper to bottom: first, model with only the stimulus term (A); second, model with stimulus and autoregressive terms (B); third, complete NN-ARx model with stimulus, autoregressive and neighborhood terms (C). Note that the goodness of fit as measured by AIC improves as more terms are included.

The (5%) significance threshold for the interregional connectivity measures was calculated using the bootstrap technique, as described in Sec. 2.4.

The complete NN-ARx model is the one selected according to the AIC. As a reference, Fig. 4 shows the results of the connectivity analysis for the three models (A)–(C).

As shown in Fig. 4(C) and Table 1 in Appendix B, our method detects significant connections between visual striate and extrastriate cortex, corresponding with occipital pole, inferior occipital, medial occipito temporal and lingual gyri for the nomenclature used in this work. These results are in agreement with functional neuroimaging studies on visual attention [22]. Striate regions involve primary visual areas while extrastriate cortex is activated during focused attention [22, 29]. In turn, our results show those areas interconnected to frontal regions through frontoparietal and temporal areas, respectively. Within this framework our study extends cited reports describing linking pathways, strengths and directions.

For comparison, we have also included in Fig. 4 results obtained for simpler models, i.e., disregarding some terms of the NN-ARx model. It can be observed that for simpler models, not including the neighborhood term (A and B in Fig. 4) a number of connections appear in areas not related with the visual task. We also show results obtained using maximum correlation (i.e., the 1.00 percentile) instead of the 0.90 percentile (see E in Fig. 4). This seems to be less robust, leading to patterns with many connections that are difficult to interpret in terms of the neurophysiological knowledge about visual processing. Finally, in part D of Fig. 4 we show the effect of averaging over each region on the connectivity analysis. It is evident that all the meaningful connections in the visual area are lost except a few connections between homologous regions.

3.2. *Braille tactile discrimination task*

As another illustrative example, the fMRI data from a group of 6 blind subjects during a Braille tactile discrimination task [43] was analyzed.

During a Braille tactile discrimination task, a session consisted of six task and six rest periods that alternate each 30 seconds in duration. During the task period, Braille stimuli were presented passively to the subject's right index finger for 3 seconds every 6 seconds. The subject's left hand was placed on a button box connected to a microcomputer for recording the response. The subject responded by pushing a button with his left index finger if the pair-wise characters were the same, or with their middle finger if the characters were different. A 30 second rest condition followed, in which the subject pushed buttons with his left index and middle finger alternately. The cue for a response was a touch to the subject's left toe, given every 6 seconds by the examiner. The comparison of images collected during the discrimination task with those during rest periods allows for correction for the

396   *Bosch-Bayard et al.*

(A)

(B)

(C)

(D)

(E)

(F)
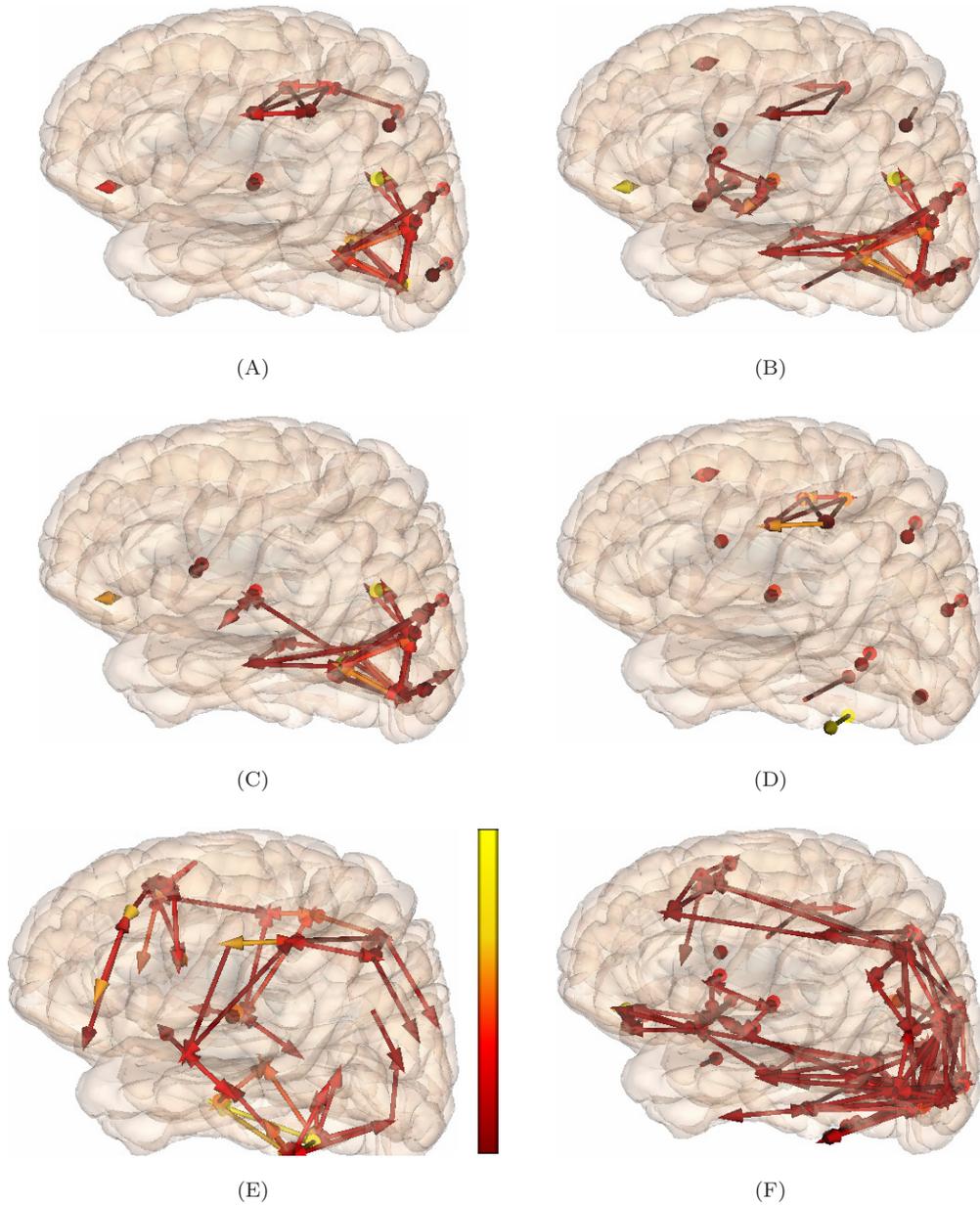
Fig. 4.   Significant connections obtained for the visual experiment (A–E) and tactile discrimination task (F). Measures of dependencies are obtained under model with only the stimulus term (A), with stimulus and autoregressive terms (B), and with all the terms of the NN-ARx (C, E and F). The measures are computed by using either the 0.90 percentile (in A, B, C and F) or maximum correlation (in E). For comparison, the standard correlations based on averaging voxels over each region are also shown in D. Note that averaging leads to miss all dependencies in the occipital area (D); the use of the maximum correlation seems to be less robust leading to patterns with many connections and difficult to support in terms of neurophysiological knowledge (E); and the complete NN-ARx model concentrates on long-range connections in the occipital area.

effects of the cue and response movement. A total of 30 pairs of Braille characters were presented, half of which were different and half of which were the same.

MRI parameters: A time-course series of 126 volumes was acquired using T2*-weighted, gradient echo, echo planar imaging (EPI) sequences with a 3.0 Tesla MRimager (VP, General Electric, Milwaukee, WI). Each volume consisted of 36 slices, with a slice thickness of 3.5 mm and a 0.5-mm gap, to include the entire cerebral and cerebellar cortex. The time-interval between two successive acquisitions of the same image was 3000 ms, and echo time was 30 ms. The field of view (FOV) was 22 cm. The in-plane matrix size was 64×64 pixels with a pixel dimension of 3.44×3.44 mm. For anatomical reference, T2*-weighted fast spin echo images were obtained with location variables identical to those of the EPIs. In addition, high-resolution whole-brain MRIs were obtained with a conventional T2*-weighted, fast spin echo sequence. A total of 112 transaxial images were obtained. The inplane matrix size was 256×256, and slice thickness was 1.5 mm, and pixel size was 0.859×0.859 mm.

The (5%) significance threshold was obtained by the bootstrap technique. The results obtained from the analysis of this data show that the present approach captures the task-related activity. The effective connectivity showed here (see Fig. 4(F) and Table 2 in Appendix B) represents the intrinsic (not task-related) connectivity that reflects the deafferentation-related plastic change.

The connections do not appear at random but are grouped into different isolated sub-networks: (i) thalamic; (ii) occipital; (iii) from the cingulate to the middle frontal orbital gyrus; (iv) from the supra parietal lobes to the postcentral gyrus; and (v) between the left and right precuneus.

When the threshold is decreased to the significant level of 5%, connections between the sub networks appear, as well as projections from the supra parietal lobe and the postcentral gyrus to the precentral and the middle frontal gyrus.

4. Some Final Comments

Furthermore, we would like to comment on some issues concerning the advantages, limitations, variants and extensions of the NN-ARx approach introduced in the present paper. They are relevant for both practical implementation and future research.

Computational burden

The bootstrap statistical tests introduced in the present work for detecting inter-regional connections require intensive computation, as a consequence of repeated generation of random variables. This drawback is shared with other resampling methods that have been suggested for EEG and fMRI analysis [4]. They allow for great flexibility in constructing complex test statistics and have wide validity under minimal probability assumptions but at the expense of considerable computer effort.

An open problem is the elaboration of asymptotic approximations to such tests that obviate the need for expensive simulations, at least in situations in which the amount of data is sufficiently large.

Nonlinear extensions of NN-ARx modeling

It has been shown that there exists some nonlinear dynamical structure in BOLD signals [13, 40]. On the other hand, Lahaye *et al.* [27] reported that nonlinearity in fMRI time series is weak. Our simulation study also implies that linear modeling approaches are quite robust in detecting instantaneous correlations, even though the time series might be generated by a nonlinear system. However, this does not mean that the ability of detecting instantaneous correlations cannot be further improved by employing nonlinear dynamical models. The generalization of the NN-ARx model methodology into the realm of nonlinear models, e.g., by generalizing the AR parameters to become state-dependent [34, 38] may be an interesting topic for future research.

Physiological interpretation of lagged correlations

fMRI time series, far from being completely random, show obvious patterns of temporal and spatial correlation structure. A plausible physiological interpretation of instantaneous correlations between distant voxels could be that spatio-temporal hemodynamic activities triggered by fast neural dynamics seem to occur almost simultaneously when recorded with a low sampling frequency. But we also find lagged correlations in fMRI time series, e.g., over a lag of one sampling period. It will be very difficult to provide meaningful interpretations for such correlations, unless additional physiological information relating to the specific experimental setup (e.g., cognitive task, sensory stimulation, etc.) can be taken into account.

Innovation approach for dynamic causal models

The approach introduced in the present paper strongly relies on extracting relevant information from the prediction errors or innovations associated to a dynamic model. The importance of innovations are worth emphasizing. First, they are not a nuisance, but a useful source of dynamical information. Second, under general conditions they have a simple distribution, thus allowing for computing the likelihood function of the model as a general basis for statistical inference (i.e., for model selection, parameter estimation, filtering, etc.). We will comment more on the theoretical framework that supports the innovation approach.

A useful way to express the correlation structure of a time series data is to consider the prediction of the series. We aim at removing temporal dependencies on the past and to whiten the observed time series y_t into an independent series by subtracting the predicted value from the realized (and observed) data value y_t:

$$\zeta_t = y_t - E[y_t/y_{t-1}, y_{t-2}, \ldots].$$

FA 1

To perform this whitening step we need to find a suitable dynamical model providing the prediction of the time series by using past observations. This concept has led to the well-known innovation approach developed by N. Wiener in the 1930s. In this framework the prediction errors (i.e., the innovations) are not a nuisance, but a useful source of information that contains the key to explain the dynamical processes in which we are interested.

The concepts described thus far are supported by a very strong mathematical theorem given by Levy ([28]; see also theorem 41 in Protter [39]) which states that "for any continuous–time Markov process y_t the corresponding innovations can be represented, under mild conditions, as the sum of two white noise processes, namely a Gaussian noise process and a Poisson noise process". This theorem is a stronger version of the well-known theorem for Markov diffusion processes [6, 23], according to which "any dynamical process can be represented by a differential equation driven by Gaussian white noise, if the process is Markov and continuous (i.e., without any discontinuous jump)". The case of additional observation noise has been treated by Frost and Kailath [15]. Consequently, we expect that, under the assumption of continuous dynamics, the time series of resulting innovations, for an optimal predictor, will be uncorrelated (in fact, independent) and Gaussian, even if, due to possible nonlinearities in dynamics, the process is non-Gaussian distributed. This theorem implies that, if we employ a properly chosen model for the dynamics, the prediction errors will be distributed as Gaussian white noise. Then the log-likelihood function for the time series may be calculated using the standard Gaussian likelihood, even though the original observed time series may have displayed nonlinearities and non-Gaussian distribution).

The methodology based on Levy's theorem has been employed in time series analysis since the early 1990s [17,32–34]. In neurosciences it has been applied to the identification of dynamical causal models, such as the Zetterberg model for EEG time series [48] and Balloon model for fMRI time series [40].

400   *Bosch-Bayard et al.*

Appendix A

71 regions gray matter segmentation defined by the MNI

1 = Mid-FrontoOrb-Gy-R

2 = Mid-Front-Gy-R

3 = Insula-R

4 = Precentral-Gy-R

5 = Lat-FrontoOrb-Gy-R

6 = Cingulate-R

7 = Mid-Front-Gy-L

8 = Sup-Front-Gy-R

9 = Globus-pallidus-R

10 = Globus-pallidus-L

11 = Putamen-L

12 = Inf-Front-Gy-L

13 = Putamen-R

14 = Parahippocampal-Gy-L

15 = Angular-Gy-R

16 = Brain-stem

17 = Subthalamic-Nuc-R

18 = Accumbens-R

19 = Uncus-R

20 = Cingulate-region-L

21 = Precuneus-R

22 = Subthalamic-Nuc-L

23 = Hippocampal-R

24 = Inf-Occip-Gy-L

25 = Sup-Occip-Gy-R

26 = Caudate-Nuc-L

27 = Supramarginal-Gy-L

28 = Mid-Front-Gy-L

29 = Sup-Pariet-Lob-L

30 = Caudate-Nuc-R

31 = Cuneus-L

32 = Precuneus-L

33 = Supramarginal-Gy-R

34 = Sup-Temp-Gy-L

35 = Uncus-L

36 = Mid-Occip-Gy-R

37 = Mid-Temp-Gy-L

38 = Cerebellum-L

39 = Lingual-Gy-L

40 = Sup-Front-Gy-L

41 = Accumbens-L

42 = Postcentral-Gy-L

43 = Inf-Front-Gy-R

44 = Cerebellum-R

45 = Precentral-Gy-L

46 = Mid-front-Orb-Gy-L

47 = Sup-Pariet-Lob-R

48 = Lat-Front-Orb-Gy-L

49 = Inf-Occip-Gy-R

50 = Sup-Occip-Gy-L

51 = Lat-OccipTemp-Gy-R

52 = Hippocampal-L

53 = Thalamus-L

54 = Insula-L

55 = Postcentral-Gy-R

56 = Lingual-Gy-R

57 = Mid-Front-Gy-R

58 = Mid-OccipTemp-Gy-L

59 = Parahippocampal-Gy-R

60 = Mid-Temp-Gy-R

61 = Occip-pole-R

62 = Inf-Temp-Gy-R

63 = Sup-Temp-Gy-R

64 = Mid-Occip-Gy-L

65 = Angular-Gy-L

66 = Inf-Temp-Gy-L

67 = Mid-OccipTemp-Gy-R

68 = Cuneus-R

69 = Lat-OccipTemp-Gy-L

70 = Thalamus-R

71 = Occip-pole-L

Appendix B

Table 1.   Values of the significant correlations between the brain regions for the group of subjects under the visual task. Sorted in descending order.

Region 1	Region 2	→	←	Region 1	Region 2	→	←
LingGR	LingGL	0.54	0.53	MidOccTmpGR	CunR	0.45	0.44
PreCunR	PreCunL	0.51	0.51	ThalL	SubThaNucL	0.45	
MidOccTmpGL	MidOccTmpGR	0.51	0.51	SupOccGL	SupOccGR	0.44	0.44
CunL	CunR	0.5	0.49	PreCunL	CunL	0.44	
MidfrontOrbGL	MidFroOrbGR	0.49		PreCunR	CunR	0.44	
MidFroOrbGR	MidfrontOrbGL	0.49		CunR	PreCunL	0.43	
LingGL	MidOccTmpGL	0.48	0.46	ParaHipGR	ParaHipGL	0.43	0.43
MidOccTmpGR	CunL	0.48	0.48	CunL	LingGL	0.43	
LingGR	MidOccTmpGL	0.48	0.45	MidOccTmpGL	LatOccTmpGR	0.43	
MidOccTmpGL	CunL	0.48	0.47	CunR	LingGL	0.43	
LingGR	MidOccTmpGR	0.47	0.44	ParaHipGL	MidOccTmpGR	0.43	
LingGL	CunR	0.47		ParaHipGR	MidOccTmpGR	0.43	
LingGR	CunR	0.47		LingGR	LatOccTmpGR	0.43	
LingGL	MidOccTmpGR	0.46	0.44	CunL	SupOccGL	0.43	0.42
LingGL	CunL	0.46		CunR	LingGR	0.42	
ThalL	ThalR	0.46	0.46	ParaHipGL	CunL	0.42	
PreCunL	CunR	0.46					
LingGR	CunL	0.46					
MidOccTmpGL	CunR	0.45	0.44				

Table 2.   Values of the significant correlations between the brain regions for the group of subjects during the tactile discrimination task. Sorted in descending order.

Region 1	Region 2	→	←	Region 1	Region 2	→	←
MidFroOrbGR	MidfrontOrbGL	0.46	0.46	LingGL	CunL	0.34	
LingGR	LingGL	0.44	0.44	MidOccTmpGR	CunR	0.34	
OccpoleR	LingGR	0.41		OccpoleL	MidOccTmpGL	0.34	
OccpoleR	InfOccGR	0.41	0.41	CunR	CunL	0.34	0.34
LingGR	LatOccTmpGR	0.4	0.4	OccpoleL	SupOccGR	0.34	0.34
OccpoleL	LingGL	0.4	0.4	MidOccGL	AnguGL	0.34	
InfOccGR	LingGR	0.4	0.39	MidFroGL	SupFroGR	0.34	
PreCunL	PreCunR	0.4	0.4	CerebR	LingGR	0.34	
OccpoleR	LingGL	0.4	0.39	CerebR	LingGL	0.34	
LingGR	OccpoleR	0.4		LatOccTmpGR	OccpoleL	0.34	
InfOccGR	LatOccTmpGR	0.4	0.38	PutL	PutR	0.34	0.34
OccpoleL	OccpoleR	0.4	0.39	SupParLobR	LingGL	0.34	
InfOccGR	LingGL	0.4	0.38	MidTmpGL	InfTmpGL	0.34	0.34
LingGL	MidOccTmpGL	0.39	0.39	InfOccGR	InfTmpGR	0.34	0.33
AccumL	AccumR	0.39	0.37	OccpoleR	MidOccTmpGR	0.34	
SupParLobL	SupParLobR	0.39	0.39	SupParLobL	AnguGR	0.34	
LingGL	LatOccTmpGR	0.38	0.37	MidOccTmpGL	OccpoleL	0.34	
LingGR	OccpoleL	0.38	0.38	MidTmpGL	InfOccGL	0.34	
LingGL	LatOccTmpGL	0.38	0.37	OccpoleR	CunR	0.34	
CaudNucL	CaudNucR	0.38	0.37	InfOccGR	CerebL	0.34	
LingGR	MidOccTmpGR	0.38	0.37	SupParLobL	PreCunL	0.34	0.33
OccpoleL	InfOccGL	0.38	0.38	CaudNucL	LingGL	0.34	

Table 2.   (*Continued*)

Region 1	Region 2	→	←	Region 1	Region 2	→	←
LingGR	LatOccTmpGL	0.38	0.37	OccpoleR	MidOccTmpGL	0.34	
ThalL	ThalR	0.38	0.37	MidFroGR	InfFroGR	0.34	0.33
MidOccTmpGR	MidOccTmpGL	0.38	0.38	InsR	InsL	0.34	
InfOccGR	OccpoleL	0.37	0.37	ThalL	LingGR	0.33	0.33
OccpoleR	InfOccGL	0.37	0.37	PosCenGL	SupParLobR	0.33	
ThalL	SubThaNucL	0.37		InfOccGR	CunR	0.33	
SupOccGR	SupOccGL	0.37	0.37	AnguGL	SupParLobL	0.33	
LingGR	MidOccTmpGL	0.37	0.36	SupParLobR	AnguGR	0.33	
LingGR	InfOccGL	0.37	0.36	LingGR	MidOccGL	0.33	
MidOccGL	InfOccGL	0.37	0.36	SupFroGL	MidFroGL	0.33	0.33
OccpoleR	LatOccTmpGR	0.37	0.36	SupParLobR	LingGR	0.33	
InfOccGL	LingGL	0.37	0.37	CaudNucL	SubThaNucL	0.33	
SupFroGL	SupFroGR	0.37	0.37	SupParLobL	PreCenGR	0.33	
LingGR	CerebL	0.37	0.35	LingGL	SubThaNucL	0.33	
AccumL	PutL	0.36	0.36	CaudNucL	LingGR	0.33	0.33
SupParLobL	SupOccGL	0.36	0.34	PreCunL	SupOccGL	0.33	
InfOccGR	InfOccGL	0.36	0.36	SupFroGL	MidFroGL	0.33	
SupOccGR	LingGR	0.36	0.36	MidFroGL	MidFroGL	0.33	0.33
LingGR	CunR	0.36	0.34	LatOccTmpGR	MidOccTmpGL	0.33	0.33
LingGR	SupOccGL	0.36	0.36	SupParLobR	MidFroGL	0.33	
InfOccGL	SupOccGR	0.36	0.36	PreCenGL	PreCenGR	0.33	0.33
MidOccGL	SupOccGL	0.36	0.36	MidOccGL	MidOccGR	0.33	0.33
OccpoleR	SupOccGR	0.36	0.36	MidFroGL	SupParLobL	0.33	
InfOccGL	LatOccTmpGL	0.36	0.36	CunR	MidOccTmpGR	0.33	
MidFroGL	MidFroGR	0.36	0.35	PosCenGL	MidFroGL	0.33	0.33
LingGL	MidOccTmpGR	0.36	0.35	SupParLobR	PreCunL	0.33	
InfOccGR	LatOccTmpGL	0.36	0.35	CaudNucL	PutR	0.33	
ParaHipGR	MidOccTmpGR	0.36	0.35	CaudNucL	PutL	0.33	
LatOccTmpGL	LatOccTmpGR	0.36	0.36	MidOccTmpGL	LatOccTmpGL	0.33	0.33
InfOccGL	SupOccGL	0.36	0.36	PosCenGL	PreCenGL	0.33	0.33
SupParLobR	SupOccGR	0.36	0.34	CunR	LingGL	0.33	
MidOccGL	SupOccGR	0.36	0.35	SupParLobL	PosCenGR	0.33	
GlobPalL	SubThaNucL	0.36	0.35	SubThaNucL	ThalL	0.33	
SubThaNucL	SubThaNucR	0.36	0.35	OccpoleL	CunR	0.33	
LingGL	CerebL	0.36	0.35	SupParLobL	LingGR	0.33	
OccpoleL	LatOccTmpGL	0.35	0.35	AccumL	MidOccTmpGR	0.33	
SupOccGL	LingGL	0.35	0.35	OccpoleR	CerebL	0.33	
CingRegL	CingR	0.35	0.35	InfOccGR	CerebR	0.33	
OccpoleR	SupOccGL	0.35	0.35	MidFroGL	MidFroGR	0.33	0.33
PosCenGR	PreCenGR	0.35	0.35	ThalL	MidOccTmpGL	0.33	
InfOccGR	SupOccGR	0.35	0.35	ParaHipGL	InfTmpGL	0.33	
LatFroOrbGL	LatFroOrbGR	0.35	0.35	OccpoleR	MidOccGL	0.33	
LingGL	CunR	0.35		ParaHipGR	MidOccTmpGL	0.33	0.33
LingGR	SubThaNucL	0.35		SupFroGR	PreCenGR	0.33	0.33
LingGR	CerebR	0.35		ParaHipGL	AccumL	0.33	
LatOccTmpGL	InfTmpGL	0.35	0.35	PosCenGL	PreCenGR	0.33	
SupFroGR	MidFroGR	0.35	0.34	CaudNucR	PutR	0.33	
SupParLobL	AnguGL	0.35		MidFroGL	MidFroGR	0.33	
ParaHipGL	ParaHipGR	0.35	0.34	MidOccTmpGL	CunL	0.33	

Spatio-Temporal Correlations from fMRI Time Series Based on the NN-ARx Model   403

Table 2.   (*Continued*)

Region 1	Region 2	→	←	Region 1	Region 2	→	←
SupParLobL	PosCenGL	0.35	0.34	ParaHipGL	MidOccTmpGR	0.33	
SupOccGL	OccpoleL	0.35	0.35	InfOccGL	LatOccTmpGR	0.33	
OccpoleR	LatOccTmpGL	0.35	0.34	SupParLobR	PreCenGR	0.33	
SupParLobR	SupOccGL	0.35		InfTmpGR	LatOccTmpGR	0.33	
LatOccTmpGR	MidOccTmpGR	0.35	0.34	ThalL	LingGL	0.33	
OccpoleL	LatOccTmpGR	0.34		InfOccGR	SupOccGL	0.33	0.33
SupParLobL	MidFroGL	0.34		SupParLobR	PosCenGR	0.33	
CaudNucL	ThalL	0.34	0.34	MidOccGL	LingGL	0.33	
InsR	AccumL	0.34		SupFroGR	MidFroGR	0.33	
SupOccGR	LingGL	0.34	0.34	InfOccGR	MidOccTmpGR	0.33	
SupParLobL	SupOccGR	0.34		SupOccGL	SupParLobR	0.33	
SupParLobL	LingGL	0.34		SupOccGL	MidOccTmpGL	0.33	
MidOccGL	OccpoleL	0.34	0.34	ParaHipGL	MidOccTmpGL	0.33	
InfOccGL	MidTmpGL	0.34		CunL	LingGL	0.33	
AnguGL	MidOccGL	0.34		InfOccGR	MidOccTmpGL	0.33	
LatFroOrbGR	MidFroOrbGR	0.34	0.34	CerebR	CerebL	0.33	
SupParLobL	MidFroGL	0.34		SupOccGR	SupParLobL	0.33	
MidFroGL	PreCenGR	0.34	0.34	MidFroGL	SupParLobL	0.33	
SupParLobL	MidOccGL	0.34		PreCenGR	MidFroGR	0.33	
LingGL	CerebR	0.34		ThalL	SubThaNucR	0.33	
SupParLobR	PosCenGL	0.34					
SupFroGR	MidFroGL	0.34					

Acknowledgments

References

[1] Alvarez-Amador A, Valdes-Sosa P, Marqui-Pascual RD, Galan-Garcia L, Biscay-Lirio R, Bosch-Bayard J, On the structure of EEG development, *Electroencephal Clin Neurophysio* **73**:10–19, 1989.

[2] Akaike H, Prediction and entropy, in Atkinson AC, Feinberg SE (eds.), *A Celebration of Statistics*, Springer, Berlin, Heidelberg, New York, pp. 1–24, 1985.

[3] Buechel C, Friston KJ, Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modeling and fMRI, *Cereb Cortex* **7**:768–778, 1997.

[4] Bullmore ET, Suckling J, Brammer MJ, In praise of tedious permutation, in Moore M (ed.), *Spatial Statistics: Methodological Aspects and Applications*, Springer-Verlag, New York, pp. 83–200, 2001.

404   *Bosch-Bayard et al.*

[5] Dodel S, Hellmann JM, Geisel T, Functional connectivity by cross-correlation clustering, *Neurocomputing*, pp. 44–46, 1065–1077, 2002.

[6] Doob JL, *Stochastic Processes*, Wiley, New York, 1953.

[7] Eichler M, A graphical approach for evaluating effective connectivity in neural systems, *Philos Trans R Soc B Biol Sci* **360**:953–967, 2005.

[8] Evans AC, Kamber M, Collins DL, MacDonald D, An MRI-based probalistic atlas of neuroanatomy, in Shorvon SD (ed.), *Magnetic Resonance Scanning and Epilepsy*, Plenum Press, New York, pp. 263–274, 1994.

[9] Friston KJ, Functional and effective connectivity in neuroimaging: A synthesis, *Hum Brain Mapp* **2**:56–78, 1994.

[10] Friston KJ, Jezzard P, Turner R, Analysis of functional MRI time series, *Hum Brain Mapp* **1**:153–171, 1994.

[11] Friston KJ, Statistical parametric mapping: Ontology and current issues, *J Cereb Blood Flow Metab* **15**:361–370, 1995.

[12] Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith C, Frackowiak RSJ, Statistical parametric maps in functional imaging: A general linear approach, *Hum Brain Mapp* **2**:189–210, 1995.

[13] Friston KJ, Mechelli A, Turner R, Price CJ, Nonlinear responses in fMRI: The balloon model, Volterra kernels and other hemodynamics, *Neuroimage* **12**:466–477, 2000.

[14] Friston KJ, Harrison L, Penny W, Dynamic causal modeling, *Neuroimage* **19**:1273–1302, 2003.

[15] Frost PA, Kailath T, An innovation approach to least squares estimation — Part III: Nonlinear estimation in white Gaussian noise, *IEEE Trans Automat Contr* **16**:217–226, 1971.

[16] Galka A, Yamashita O, Ozaki T, Biscay R, Valdes-Sosa P, A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering, *Neuroimage* **23**:435–453, 2004.

[17] Galka A, Ozaki T, Bosch-Bayard J, Yamashita O, Whitening as a tool for estimating mutual information in spatiotemporal data sets, *J Stat Phys* **124**(5):1275–1315, 2006.

[18] Goebel R, Roebroeck A, Kim D, Formisano E, Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping, *Magn Reson Imaging* **21**:1251–1261, 2003.

[19] Gössl C, Auer DP, Fahrmeir L, Dynamics models in fMRI, *Magn Reson Med* **43**:72–81, 2000.

[20] Harrison L, Penny WD, Friston KJ, Multivariate autoregressive modeling of fMRI time series, *Neuroimage* **19**:1477–1491, 2003.

[21] Horwitz B, The elusive concept of brain connectivity, *Neuroimage* **19**:466–470, 2003.

[22] Iacoboni M, Attention and Sensorimotor Integration, in Toga AW, Mazziota J (eds.), *Brain Mapping. The Systems*, AP, pp. 331–363.

[23] Ito K, On stochastic differential equations, *Mem Am Math Soc* **4**:1–51, 1951.

[24] Kailath T, A view of three decades of linear filtering theory, *IEEE Trans Inf Theory* **20**:145–181, 1974.

[25] Kellogg OD, *Foundations of Potential Theory*, Springer-Verlag, Berlin, Heidelberg, New York, 1967.

[26] Koch MA, Norris DG, Hund-Georgiadis M, An investigation of functional and anatomical connectivity using magnetic resonance imaging, *Neuroimage* **16**:241–250, 2002.

[27] Lahaye PJ, Poline JB, Flandin G, Dodel S, Garnero L, Functional connectivity: Studying nonlinear delayed integrations between BOLD signals, *Neuroimage* **20**:962–974, 2003.

[28] Levy P, Sur une classe de courbes de l'espace de Hilbert et sur une equationintegrale non lineaire, *Ann Sci Ecole Norm Sup* **73**:121–156, 1956.

[29] Macaluso E, Frith C, Interhemispheric differences in extrastriate areas during visuo-spatial selective attention, *Neuroimage* **12**:485–494, 2000.

[30] McIntosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitz B, Network analysis of cortical visual pathways mapped with PET, *J Neurosci* **14**:655–666, 1994.

[31] McIntosh AR, Gonzalez-Lima F, Structural equation modeling and its application to network analysis in functional brain imaging, *Hum Brain Mapp* **2**:2–22, 1994.

[32] Ozaki T, Identification of nonlinearities and non-Gaussianities in time series, in Brillinger DR, Caines P, Geweke J, Parzen E, Rosenblatt E, Taqqu MS (eds.), *New Directions in Time Series Analysis, Part I*, IMA Volumes in Mathematics and its Applications, Volume 45, Springer, pp. 227–264, 1992.

[33] Ozaki T, Iino M, An innovation approach to non-Gaussian time series analysis, *J. Appl Probab* **38A**:78–92, 2000.

[34] Ozaki T, Jimenez JC, Peng H, Haggan-Ozaki V, The innovation approach to the identification of nonlinear causal models in time series analysis, in Brillinger DR, Robinson EA, Schoenberg FP (eds.), *Time Series Analysis and Applications to Geophysical Systems*, IMA Volumes in Mathematics and its Applications, Volume 139, Springer, pp. 195–226, 2004.

[35] Paus T, Inferring causality in brain images: A perturbation approach, *Philos Trans R Soc Lond B Biol Sci* **360**:1109–1114, 2005.

[36] Pearl J, *Causality: Models, Reasoning, and Inference*, Cambridge University Press, New York, 2000.

[37] Penny WD, Friston KJ, Classical and Bayesian inference, in Frackowiak RSJ, Friston KJ, Frith C, Dolan R, Friston KJ, Price CJ, Zeki S, Ashburner J, Penny WD (eds.), *Human Brain Function*, 2nd ed., Academic Press, 2003.

[38] Priestley MB, *Nonlinear and Non-Stationary Time Series Analysis*, Academic Press, San Diego, 1988.

[39] Protter P, *Stochastic Integration and Differential Equations*, Springer, Berlin, 1990.

[40] Riera-Diaz JJ, Watanabe J, Kazuki I, Naoki M, Aubert E, Ozaki T, Kawashima R, A state-space model of the hemodynamic approach: Nonlinear filtering of BOLD signals, *Neuroimage* **21**:547–567, 2004a.

[41] Riera-Diaz JJ, Bosch-Bayard J, Yamashita O, Kawashima R, Sadato N, Okada T, Ozaki T, fMRI activation maps based on the NN-ARx model, *Neuroimage* **23**(2):680–697, 2004b.

[42] Riera-Diaz JJ, Aubert E, Iwata K, Kawashima R, Wan X, Ozaki T, Fusing EEG and fMRI based on a bottom-up model: Inferring activation and effective connectivity in neural masses, *Philos Trans R Soc Lond B Biol Sci* **360**:1025–1041, 1457, 2005.

[43] Sadato N, Okada T, Honda M, Yonekura Y, Critical period for cross-modal plasticity in blind humans: A functional MRI study, *Neuroimage* **16**:389–400, 2002.

406   *Bosch-Bayard et al.*

[44] Salvador R, Suckling J, Coleman MR, Pickard JD, Menon D, Bullmore E, Neurophys-
       iological architecture of functional magnetic resonance images of human brain, *Cereb
       Cortex* **15**:1332–1342, 2005.
[45] Shibata R, Asymptotically efficient selection of order of the model for estimating param-
       eters of a linear process, *Ann Stat* **8**:147–164, 1980.
[46] Spirtes P, Glymour C, Scheines R, *Causation, Prediction, and Search*, 2nd ed., MIT
       Press, New York, 2000.
[47] Stephan KE, Friston KJ, Models of Effective Connectivity in Neural Systems, in *Hand-
       book of Complex Systems*, Springer Berlin/Heidelberg, pp. 303–327.
[48] Valdes PA, Jimenez JC, Riera J, Biscay R, Ozaki T, Nonlinear EEG analysis based on
       a neural mass model, *Biol Cybern* **81**:415–424, 1999.
[49] Valdés-Sosa PA, Sanchez-Bornot JM, Lage-Castellanos A, Vega-Hernandez M, Bosch-
       Bayard J, Melie-García L, Canales-Rodriguez E, Estimating brain functional connec-
       tivity with sparse multivariate autoregression, *Philos Trans R Soc Lond B Biol Sci*
       360:969–981:1457, 2005.
[50] Worsley KJ, Chen J-I, Lerch J, Evans AC, Comparing connectivity via thresholding
       correlations and SVD, *Philos Trans R Soc Lond B Biol Sci* **360**:913–920, 2005.
[51] Yamashita O, Galka A, Ozaki T, Biscay R, Valdes-Sosa P, Recursive penalized least
       squares solution for dynamical inverse problems of EEG generation, *Hum Brain Mapp*
       21:221–235, 2004.
[52] Yamashita O, Sadato N, Okada T, Ozaki T, Evaluating directed connectivity of BOLD
       signals applying the multivariate autoregressive model and relative power contribution,
       Neuroimage **25**:478–490, 2005.