A nonlinear identification method to study effective connectivity in functional MRI

Xingfeng Li [a,b,c,d,e], Guillaume Marrelec [a,b,c,d], Robert F. Hess [e], Habib Benali [a,b,c,d,*]

[a] Inserm, UPMC Univ Paris 06, UMR_S 678, Laboratoire d'Imagerie Fonctionnelle, and LINeM, GHU Pitié-Salpêtrière, 91 bd de l'Hôpital, F-75634 Paris Cedex 13, France
[b] Université de Montréal, LINeM, International Laboratory of Neuroimaging and Modelling, IUGM, 4565 chemin Queen-Mary, Montréal, Quebec, Canada H3W 1W5
[c] Université Paris 11, IFR49, DSV/I²BM/NeuroSpin, Bât 145, 91191 Gif-sur-Yvette Cedex, France
[d] Université de Montréal, CRM, CP 6128, Succursale Centre-ville, Montréal, Canada H3C 3J7
[e] McGill Vision Research, Department of Ophthalmology, McGill University, 687 Pine Avenue W. (H4-14), Montréal, Quebec, Canada H3A 1A1

ARTICLE INFO

ABSTRACT

In this paper we propose a novel approach for characterizing effective connectivity in functional magnetic resonance imaging (fMRI) data. Unlike most other methods, our approach is nonlinear and does not rely on a priori specification of a model that contains structural information of neuronal populations. Instead, it relies on a nonlinear autoregressive exogenous model and nonlinear system identification theory; the model's nonlinear connectivities are determined using a least squares method. A statistical test was developed to quantify the significance of the influence that regions exert on one another. We compared this approach with a linear method and applied it to the human visual cortex network. Results show that this method can be used to model nonlinear interaction between different regions for fMRI data.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

There has been an increasing interest in studying effective connectivity using functional magnetic resonance imaging (fMRI) analysis. Most studies have considered the question of connectivity within the range of linear models. For instance, path analysis or structural equation model (SEM) has been applied in different brain connectivity studies (McIntosh and Gonzalez-Lima, 1994; Buchel and Friston, 1997; Bullmore et al., 2000; Marrelec et al., 2008). This method is based on the assumption of linear influences between brain regions, an hypothesis that has already been challenged when modelling the dynamics of the human brain in fMRI (Friston et al., 2003).

Other methods include Granger causal modelling (GCM) (Roebroeck et al., 2005), vector multivariate autoregressive (MAR) models (Valdes-Sosa et al., 2005; Yamashita et al., 2005), and autoregressive moving average (ARMA) models (Moller et al., 2003). These methods do not rely on a priori specification of a structural model, but on the concept of Granger causality to define the existence and direction of influences between two stochastic fMRI time series. However, these methods can only cope with linear interactions between brain regions. Although nonlinear GCM was recently proposed and applied to simulated signals, its practical usefulness is yet to be demonstrated (Pereda et al., 2005). Furthermore, this method also assumes that the experimental input causing neuronal responses of the network is unknown, which is not realistic for fMRI data with a prior experimental design (Friston et al., 2003).

Besides these works, there has been a sustained effort to study brain networks based on modelling nonlinear interactions of neuronal population activity in fMRI data. For example, dynamic causal modelling (DCM) (Friston et al., 2003; Stephan et al., 2008) has been developed to model neuronal network interactions between different brain regions. The basic idea of this method is to treat the brain as an input–output system. Generally, the nonlinear model of the brain is established beforehand, and the causality is inherent in the differential equations that specify the model. Therefore, establishment of the right model is crucial to a successful study of the interaction. However, the neuronal population model in fMRI data is nonlinear (Friston et al., 2000; Friston, 2002), and the nonlinear nature of the blood-oxygen-level dependent (BOLD) hemodynamic response makes it difficult to build an exact model although there are several enhanced Balloon models available. There is still debate regarding which nonlinear model should be used for fMRI connectivity studies (Deneux and Faugeras, 2006). In addition, solving the differential equations

---

* Corresponding author. Address: UMR_S 678 Inserm/UPMC, Laboratoire d'Imagerie Fonctionnelle, Faculté de Médecine, CHU Pitié-Salpétrière, 91 Bd de l'Hôpital – F-75634 PARIS Cedex 13, France. Tel.: +33 (0) 1 53 82 84 15; fax: +33 (0) 1 53 82 84 48.

E-mail address: Habib.Benali@imed.jussieu.fr (H. Benali).

that specify the model is also a difficult task. Usually, bilinear or nonlinear methods are used to approximate the Volterra series and a Bayesian expectation-maximization (EM) algorithm is adapted to infer the equation parameters. This leads to another difficulty, namely that the solution depends upon the prior distribution selected in the Bayesian inference scheme.

The motivation for our work was to develop a general nonlinear system identification framework for the effective connectivity study of BOLD responses. We devised a scheme which identifies nonlinear connectivities using a nonlinear autoregressive exogenous model (NARX) method and provides statistics which can be used to test model interactions. First, the basic theory of NARX is introduced to fMRI analysis, where we show how the model is constructed for the investigation of nonlinear dynamic connectivity. Then, we present the least squares algorithm to identify the strength of the various connectivities. Lastly, we apply this method to the human visual system to study the brain region interactions. We found the method is effective in studying nonlinear interaction in fMRI with or without consideration of the experimental input.

2. Materials and methods

2.1. Nonlinear brain system

The physiological processes underlying the BOLD response can be modelled as a multiple-input and multiple-output (MIMO) system (Friston et al., 2000):

$$\begin{cases} \dot{x}(t) = f(x(t), u(t), \theta), \\ y(t) = g(x(t), \theta) \end{cases} \text{ and its discrete form is}$$

$$\begin{cases} x(t+1) = f(x(t), u(t), \theta), \\ y(t) = g(x(t), \theta), \end{cases} \tag{1}$$

where f and g are nonlinear functions, and θ represents the set of model parameters. $y(t)$ is the BOLD response or nonlinear brain system output, $x(t)$ is the state variable of the system, and $u(t)$ is the system input. Under some mild assumptions the discrete-time multivariate system (1) with p outputs and q inputs can be described by an autoregressive moving average with exogenous input (NARMAX) as follows (Leontaritis and Billings, 1985a,b):

$$y(t) = f_g[y(t-1), \ldots, y(t-n_y), u(t-1), \ldots, u(t-n_u), e(t-1), \ldots, e(t-n_e)] + e(t), \tag{2}$$

where $y(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_p(t) \end{bmatrix}, u(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_q(t) \end{bmatrix}, e(t) = \begin{bmatrix} e_1(t) \\ \vdots \\ e_p(t) \end{bmatrix}$, are the system output, input and noise, respectively; n_y, n_u, and n_e are the maximum lags in the output, input, and noise; $e(t)$ is a zero mean independent sequence; f_g is a new nonlinear function which can be obtained from nonlinear functions f and g. A special case of the general NARMAX model (2) is the nonlinear autoregressive with exogenous inputs (NARX) model:

$$y(t) = f_g[y(t-1), \ldots, y(t-n_y), u(t-1), \ldots, u(t-n_u)] + e(t). \tag{3}$$

By applying the regression equation, the NARMAX model (2) and NARX model (3) can be approximated as (Chen et al., 1989; Zhu and Billings, 1996; Chon et al., 1997):

$$y(t) = \sum_{m=0}^{M} a_m P_m(t) + e(t), \quad t = 0, 1, \ldots, N, \tag{4}$$

where $P_0(t) = 1$; for $M \geqslant 1$, $P_m(t) = y_1 \cdots y_i u_1 u_2 \cdots u_j$, $i \geqslant 1, j \geqslant 0$; m is the number of nonlinear terms; M is the system order; N is the total number of time point in the time series; i is the number of connected regions; j is the number of inputs. Eq. (4) denotes a general

case where both input and output terms may be present, but it should be understood that some of the P_m may contain only input or output terms and cross-products. For example, for two stationary series of N values, the inputs u_{y_1} and y_2, output y_1 of a closed-loop time-invariant nonlinear brain system can be described as (Faes et al., 2008):

$$y_1(t) = c_0 + \sum_{i=1}^{S_1} a_1(i) y_1(t-i) + \sum_{j=0}^{T_1} b_1(j) y_2(t-j)$$

$$+ \sum_{i=1}^{S_2} \sum_{j=1}^{S_2} a_2(i,j) y_1(t-i) y_1(t-j) + \sum_{i=0}^{T_2} \sum_{j=1}^{T_2} b_2(i,j) y_2$$

$$\times (t-i) y_2(t-j) + \sum_{i=1}^{S_2} \sum_{j=0}^{T_2} c_2(i,j) y_1(t-i) y_2(t-j)$$

$$+ c_1 u_{y_1}(t) + e_{y_1}(t), \tag{5}$$

where the coefficients c_0, $\{a_1(i); b_1(j); c_1\}$, and $\{a_2(i,j); b_2(i,j); c_2(i,j)\}$ denote constant (zero-th order), linear (first order), and nonlinear (second order) contributions to $y_1(t)$, respectively. u_{y_1} represents the experimental input, and e_{y_1} is the prediction error of $y_1(t)$. The model orders S_1 and S_2 are the maximum lags of the linear and nonlinear autoregressive (AR) influences, respectively, while the maximum lags for linear and nonlinear exogenous effects are determined by the model orders T_1 and T_2. The model can be represented in the matrix form:

$$Y = c_0 H_1 + H_{y_1} A_1 + H_{y_2} B_1 + H_{y_1 y_1} A_2 + H_{y_2 y_2} B_2 + H_{y_1 y_2} C_2$$

$$+ c_1 u_{y_1} + e_{y_1}, \tag{6}$$

where the vector $Y = [y_1(1), y_1(2), \ldots, y_1(N)]^T$ contains values of output series, $e_{y_1} = [e_{y_1}(1), e_{y_1}(2), \ldots, e_{y_1}(N)]$ is the prediction error series. $u_{y_1} = [u_{y_1}(1), u_{y_1}(2), \ldots, u_{y_1}(N)]$ is the experimental input time series; A_1, B_1 and C_1 are the first order vector coefficients; A_2, B_2, and C_2 are the second order vector coefficients. The matrices H_{y_1} and H_{y_2} contain the S_1 linear AR terms and the $(T_1 + 1)$ linear exogenous terms respectively:

$$H_{y_1} = \begin{bmatrix} y_1(0) & y_1(-1) & \cdots & y_1(1-S_1) \\ y_1(1) & y_1(0) & \cdots & y_1(2-S_1) \\ \vdots & \vdots & & \vdots \\ y_1(t-1) & y_1(t-2) & & y_1(t-S_1) \\ \vdots & \vdots & \cdots & \vdots \\ y_1(N-1) & y_1(N-2) & \cdots & y_1(N-S_1) \end{bmatrix},$$

$$H_{y_2} = \begin{bmatrix} y_2(1) & y_2(0) & \cdots & y_2(1-T_1) \\ y_2(2) & y_2(1) & \cdots & y_2(2-T_1) \\ \vdots & \vdots & & \vdots \\ y_2(t) & y_2(t-1) & & y_2(t-T_1) \\ \vdots & \vdots & \cdots & \vdots \\ y_2(N) & y_2(N-1) & \cdots & y_2(N-T_1), \end{bmatrix}$$

the matrix $H_{y_1 y_1}$ contains of the $S_2(S_2 + 1)/2$ quadratic AR terms given by the product of the terms of the matrix H_{y_1}. In the same way, the matrix $H_{y_2 y_2}$ contains the $(T_2 + 1)(T_2 + 2)/2$ quadratic exogenous terms, and the matrix contains the $S_2(T_2 + 1)$ cross-terms. Eq. (6) can be written as:

$$Y = W\beta + e_y, \tag{7}$$

where $W = [H_1, H_{y_1}, H_{y_2}, H_{y_1 y_1}, H_{y_2 y_2}, H_{y_1 y_2}, u_{y_1}]$, $\beta = [c_0, A_1^T, B_1^T, A_2^T, B_2^T, C_2^T, c_1]^T$. Coefficient matrix β can be estimated by least squares: $\beta = pinv(W)y$, where $pinv$ is the Moore–Penrose pseudoinverse of the matrix.

By neglecting the nonlinear terms $H_{y_1y_1}A_2 + H_{y_2y_2}B_2 + H_{y_1y_2}C_2$, experimental input u_{y_1}, and considering only the first order of AR, i.e. AR(1), this leads to:

$$Y = c_0H_1 + H_{y_1}A_1 + H_{y_2}B_1 + e_{y_1}$$

or

$$y_1(t) = c_{01} + a_{11}y_1(t-1) + a_{12}y_2(t-1) + e_1(t), \quad (8)$$
$$y_2(t) = c_{02} + a_{21}y_1(t-1) + a_{22}y_2(t-1) + e_2(t). \quad (9)$$

This is the well-known two connection linear GCM in fMRI data analysis.

2.2. Granger causality (GC) tests and directionality indices

Once the coefficients of the model are determined, Granger causality tests (Granger, 1969; Oxley and Greasley, 1998; Werner-heim, 2000) are derived based on F statistics. For simplicity and illustrative purposes, we take the nonlinear models (5) for example; the same principle can be applied for the linear system (8) and (9). The test for determining Granger-cause (GC) is (Werner-heim, 2000):

(i) y_2 is GC of y_1 if $b_1 = b_2 = c_2 = 0$ in Eq. (5) is *not* true. Given the data, we reach this conclusion if $b_1 = b_2 = c_2 = 0$ is rejected.

(ii) Similarly, y_1 Granger causes of y_2 can be investigated by reversing the input–output roles of the two series. F statistics are developed to detect significant relations (see Appendix).

Directionality indices are quantified by computation of the absolute or relative predictability improvement obtained by the NARX model compared to the nonlinear autoregressive (NAR) model (Faes et al., 2008), i.e.:

$$M_{y_1|y_2,u_1} = M_{y_1|y_1} - M_{y_1|y_1,y_2,u_1},$$

where $M_{y_1|y_1}$ represents the residual sum of square (RSS) from its own past ($T_1 = T_2 = 0$ in Eq. (5) for example), and RSS$= \frac{1}{N}\sum_{t=1}^{N}e_{y_1}(t)^2$.

$M_{y_1|y_1,y_2,u_1}$ denotes RSS for its own past and the past and present of the input series for a NARX model ($T_1 \neq 0$, $T_2 \neq 0$ in Eq. (5) for example). The relative causality index for the inputs y_2 and u_1 to the output y_1 is:

$$NM_{y_1|y_2,u_1} = \frac{M_{y_1|y_2,u_1}}{M_{y_1|y_1}}.$$

This index belongs to [0, 1]. In the same way, the causality from input y_1 and u_2 to output y_2 can be investigated by reversing the input–output roles of the two series:

$$M_{y_2|y_1,u_2} = M_{y_2|y_2} - M_{y_2|y_1,y_2,u_2}; NM_{y_2|y_1,u_2} = \frac{M_{y_2|y_1,u_2}}{M_{y_2|y_2}}.$$

Finally, the relative strength of the causal interactions from y_2 to y_1 under influence of u_1 and u_2 is calculated by the directionality index:

$$D_{y_1y_2|u_1,u_2} = \frac{NM_{y_1|y_2,u_1} - NM_{y_2|y_1,u_2}}{NM_{y_1|y_2,u_1} + NM_{y_2|y_1,u_2}}. \quad (10)$$

$D_{y_1y_2|u_1,u_2}$ ranges from -1 to 1. A negative value implies that direction of causality is from y_2 to y_1, whereas a positive value indicates that the causality if fromy_1 toy_2, and 0 means balanced bilateral interactions between y_2 and y_1.

2.3. fMRI experimental design and data analysis

Six normal subjects (age 29.8 ± 4 yrs) were used in this study (for more details regarding the protocol, see Li et al. (2007a,b)).

Studies were performed with the informed consent of the subjects and were approved by the Montreal Neurological Institute Research Ethics Committee and in accordance with the Helsinki Declaration of human rights. Briefly, a Siemens 1.5 T Magnetom scanner was used to collect both anatomical and functional images. Anatomical images were acquired using a rectangular (14.5" × 6.5") head coil (circularly polarized transmit and receive) and a T_1 weighted sequence (TR = 22 ms; TE = 10 ms; flip angle = 30°) giving 176 sagittal slices of 256×256 mm^2 image voxels. Functional scans for each subject were collected using a surface coil (circularly polarized, receive only) positioned beneath the subject's occiput. Each functional imaging session was preceded by a surface coil anatomical scan (identical to the head coil anatomical sequence, except that $80 \times 256 \times 256$ sagittal images of slice thickness 2 mm were acquired) in order to later co-register the data with the more homogeneous head-coil image. Functional scans were multislice T_2^*-weighted, gradient-echo, planar images (GE-EPI, TR = 3.0 s, TE = 51 ms, flip angle = 90°). Image volume consisted of 30 slices orthogonal to the calcarine sulcus. The field of view was 256×256 mm, the matrix size was 64×64 with a thickness of 4 mm yielding voxel sizes of $4 \times 4 \times 4$ mm.

Each retinotopic experiment (phase-encoded design, travelling square wave (Engel et al., 1997)) consisted of four acquisition runs for each eye (two eccentricity runs, two polar angle runs, two clockwise order runs, and two count-clockwise runs) each of 128 image volumes acquired at three second intervals for the left and right eye of normals. Runs were alternated between the eyes in each case while the subject was performing a task to keep awake in the scanner. The eye not being stimulated was occluded with a black patch that excluded all light from the eye. Subjects monocularly viewed a stimulus back-projected into the bore of the scanner and viewed through an angled mirror.

In addition, the middle temporal (MT) cortex or V5 cortex localizer experiment was conducted for all normal subjects. The experiment consisted of two to five acquisition runs for both eyes and area MT was localized using a block design in which a 16 Hz flickering low contrast checkerboard was compared with a static version of the same stimulus as described previously (Dumoulin et al., 2000). During the MT localizer scanning sessions, subjects binocularly viewed a stimulus back-projected into the bore of the scanner and viewed through an angled mirror. In the data preprocessing, the MT area of each subject was defined based on T statistics of BOLD response after different MT localizer runs were combined by using random effect model (Worsley et al., 2002). We localized MT in the regions which have large T value ($t = 1.96$, $P < 0.05$). We defined the common boundaries of different visual areas (from V1 to V4) by combining the retinotopic field sign map information of each subject (Engel et al., 1997; Sereno et al., 1995; Warnking et al., 2002 and references therein).

Simulation experiments were designed to validate the method. A three inputs–three outputs visual system (Fig. 1) was employed, and the system inputs and outputs were generated by computer simulation for 1000 repetitions. Gaussian noise was added to the model to evaluate the method. Statistical analysis was then carried out to study the connectivity from the simulation results.

3. Results

We applied the nonlinear identification method to a three inputs–three outputs visual system as shown in Fig. 1A (Felleman and Van Essen, 1991; Gonçalvesa and Hall, 2003; Hinrichs et al., 2006). The method was implemented using MATLAB program (computer programs are available upon request). Fig. 1B–D shows the two gamma functions used as the experimental input to the neuronal network system; typical parameters are used to
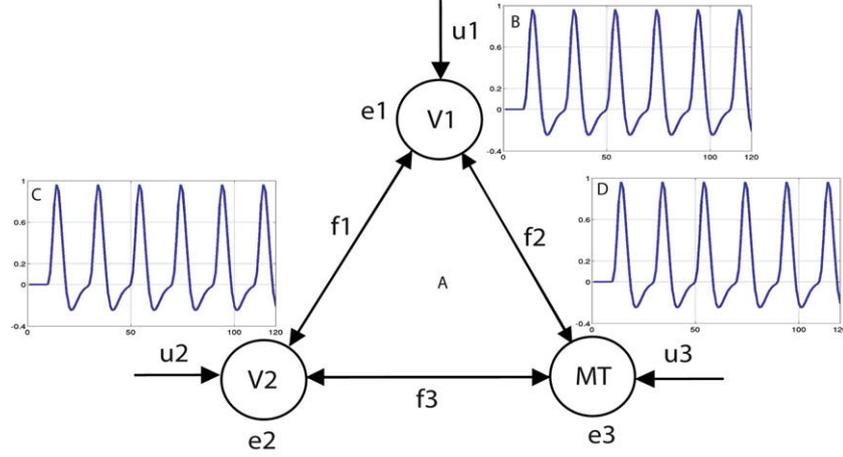
Fig. 1. Visual network system. A is the network used in the study; B–D are the experimental inputs of the system. u_1, u_2, and u_3 are inputs for V1, V2, and MT respectively, which models the indirectly neuronal system inputs derived by stimuli. f_1, f_2, and f_3 are the directional interaction functions between visual regions. e_1, e_2, and e_3 are unmeasureable noise for V1, V2, and MT respectively.

approximate the haemodynamic response function as input in visual cortex (Golver, 1999). Because a retinotopic stimulus is one kind of phase-encoded design (Engel et al., 1997), its input can be regarded as a travelling square wave. Some examples of the fMRI time series within each region are given in Fig. 2 (dotted-dash curve). Because of the combination of the BOLD response haemodynamic delay and the phase-encoded experiment design, averaging all the time series in each region would blur the response of delay which is an important factor for the effective connectivity study. We randomly sampled 96 networks from V1, V2 and MT, i.e. 96 BOLD time courses in each visual region. We used $y_1(t)$, $y_2(t)$, and $y_3(t)$ to represent the time series from V1, V2, and MT regions respectively.

Generally, the first step for nonlinear connectivity study is to build a neuronal network system as shown in Fig. 1 and get the BOLD response curves in each region as shown in Fig. 2. The next step is to select the polynomials $P_m(t)$ in Eq. (4). For the nonlinear system, the order of $P_m(t)$ is larger than 1, while for the linear system, $P_m(t)$ includes only the first order polynomials. In this study, we determine the total number of polynomials $P_m(t)$ by neglecting the 3rd and the higher order of nonlinearity, i.e. we include only 2nd nonlinearity ($M = 2$ in Eq. (4)) in the analysis, i.e. $S_1 = S_2 = 1$ and $T_1 = T_2 = 1$ in Eq. (5) of a two connection system.

3.1. Simulation experiment

Once the system, model, and the BOLD responses are determined, the model coefficients are identified by Eq. (7) for both simulation and real data experiments. We investigate the causality relations from V2 to V1, this can be done by testing the coefficients equal to zero in front of terms which contain V2. For example, for the data from Fig. 2, and 2 order of nonlinearity with experimental input; the identified connection of V1($y_1(t)$), V2($y_2(t)$), and MT($y_3(t)$) in Fig. 1 were found to be:

$$y_1(t) = -0.2050 + 0.2533y_1(t-1) + 0.2335y_2(t-1)$$
$$+ 0.1328y_3(t-1)$$
$$- 0.0465y_1(t-1)y_1(t-1) - 0.2094y_1(t-1)y_2(t-1)$$
$$- 0.0491y_1(t-1)y_3(t-1) + 0.0534y_2(t-1)y_2(t-1)$$
$$+ 0.0156y_2(t-1)y_3(t-1) + 0.0910y_3(t-1)y_3(t-1)$$
$$+ 1.0599u_1(t) + e_1(t), \quad (11)$$

$$y_2(t) = -0.1555 - 0.4603y_1(t-1) + 0.9358y_2(t-1)$$
$$- 0.0026y_3(t-1)$$
$$+ 0.0652y_1(t-1)y_1(t-1) - 0.2177y_1(t-1)y_2(t-1)$$
$$- 0.0515y_1(t-1)y_3(t-1) + 0.1122y_2(t-1)y_2(t-1)$$
$$+ 0.0645y_2(t-1)y_3(t-1) + 0.0434y_3(t-1)y_3(t-1)$$
$$+ 0.6211u_2(t) + e_2(t), \quad (12)$$

$$y_3(t) = -0.2001 + 0.1525y_1(t-1) - 0.1122y_2(t-1)$$
$$+ 0.3962y_3(t-1)$$
$$+ 0.3151y_1(t-1)y_1(t-1) - 0.3508y_1(t-1)y_2(t-1)$$
$$- 0.0826y_1(t-1)y_3(t-1) - 0.0034y_2(t-1)y_2(t-1)$$
$$- 0.0495y_2(t-1)y_3(t-1) + 0.0815y_3(t-1)y_3(t-1)$$
$$+ 0.3729u_3(t) + e_3(t). \quad (13)$$

The interaction between V1 and V2 (i.e. f_1) includes all the terms of $y_2(t)$ in Eq. (11), that is: $0.2335y_2(t-1) - 0.2094y_1(t-1)y_2 (k-1) + 0.0534y_2 (t-1)y_2 (k-1) + 0.0156y_2 (t-1)y_3(t-1)$. The directionality index is quantified from Eq. (10), $f_1 = D_{V_1V_2|MT,u_1} = 0.2527$. Similarly, we get $f_2 = D_{V_1MT|V_2,u_1} = 0.6647$, $f_3 = D_{V_2MT|V_1,u_2} = 0.4683$.

Based on the coefficients in Eqs. (11)–(13), a simulation study was conducted to validate the method. The model for the simulation study was the three inputs–three outputs nonlinear system (Fig. 1). Three time series as described in Eqs. (11)–(13) were generated according to the equations for the simulation study. The noises $e_1(t)$, $e_2(t)$, and $e_3(t)$ were produced by MATLAB function random, and it has a normal distribution with mean value of 0 and standard deviation of $\sqrt{0.2}$ (typical parameters of the noise), i.e. $e_1(t)$, $e_2(t)$, and $e_3(t) \sim N(0,0.2)$. We regenerated $y_1(t)$, $y_2(t)$, and $y_3(t)$ for 1000 repetitions after adding the Gaussian-distributed noise into the fMRI responses (Eqs. (11)–(13)). Then the method was applied to estimate the coefficients of the model for 1000 repetitions. Table 1 shows the outcome of the estimation results of Eq. (11). It was found that most of the coefficients in Eq. (11) are within the range of simulation results. This validates our method for the fMRI connectivity study.

3.2. Real data experiments

To further verify our method, we also conducted the experiments using real fMRI data. These results are given in Figs. 3–5. In Fig. 3, different methods are compared; this is done by modelling the BOLD response in Fig. 2A. The results of the linear method
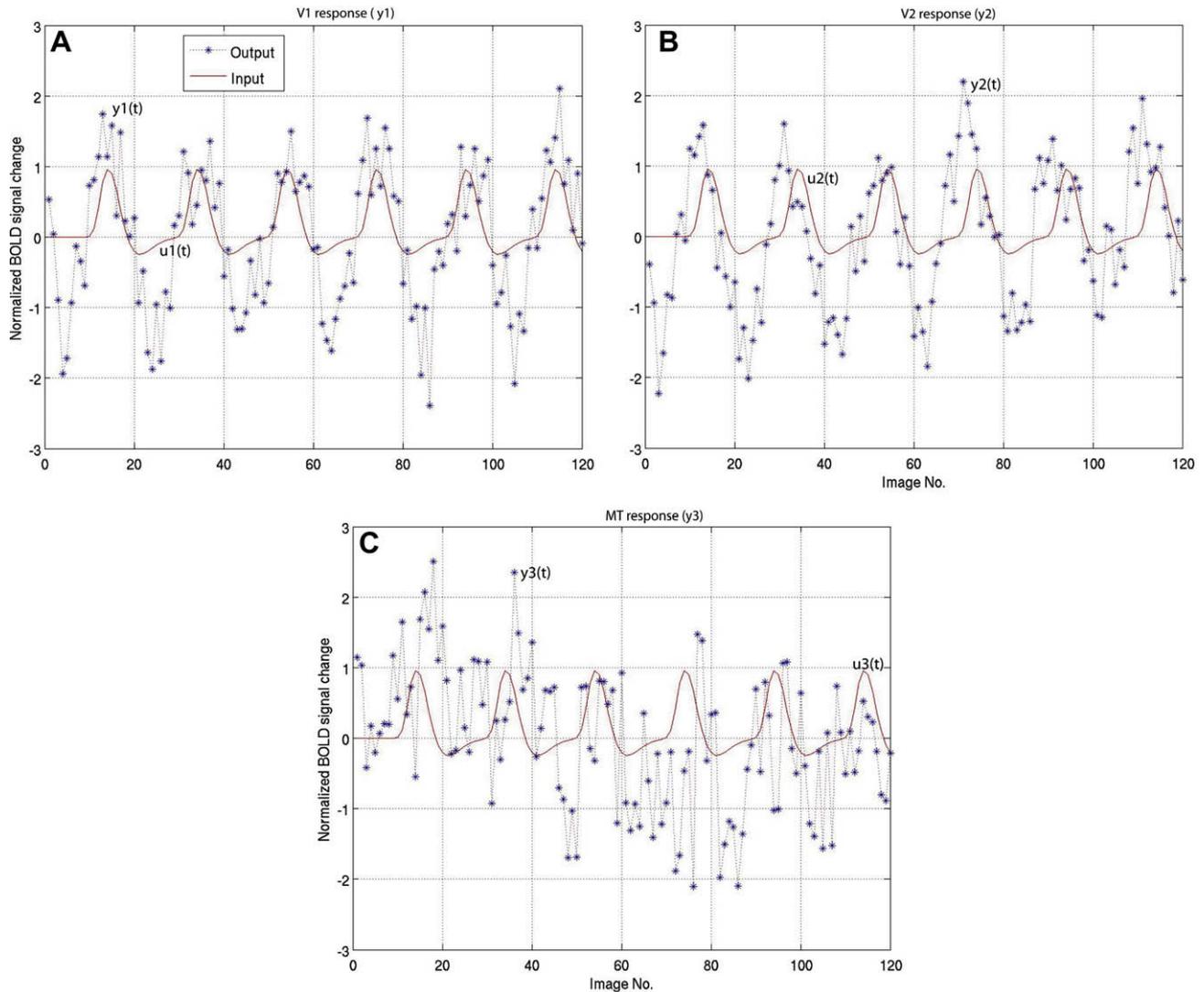
Fig. 2. Examples of typical brain response in V1 (A), V2 (B) and MT (C). A–C are from the left hemisphere of one subject. Dotted-dash curve represents the BOLD responses from V1, V2, and MT respectively (from first run of polar angle stimulus). Solid curve is the two gamma function to approximate the experimental input. Y axis is the image number; X axis is the normalized BOLD signal magnitude (in proportion).

Table 1
Simulation results for Eq. (11) with Gaussian noise $(N)(0, 0, 2)$, 1000 repetition.

	Coefficient in Eq. (11)	Mean value	Standard deviation
Constant	−0.2050	−0.1541	0.0476
$y_1(t)$	0.2533	0.2340	0.0652
$y_2(t)$	0.2335	0.2107	0.0525
$y_3(t)$	0.1328	0.0897	0.0662
$y_1(t) \times y_1(t)$	−0.0465	0.0500	0.0755
$y_1(t) \times y_2(t)$	−0.2094	−0.1881	0.1255
$y_1(t) \times y_3(t)$	−0.0491	−0.0145	0.1227
$y_2(t) \times y_2(t)$	0.0534	−3.2715e−04	0.0707
$y_2(t) \times y_3(t)$	0.0156	−0.0211	0.0953
$y_3(t) \times y_3(t)$	0.0910	0.0265	0.1011
$u(t)$	1.0599	1.1933	0.1402

are given in Fig. 3A and B, while Fig. 3C and D shows the results of the nonlinear method; Fig. 3A and C is the results without input, whereas Fig. 3B and D is the results with input. $F1$ denotes the F test for "V2 causes V1", and $F2$ represents the causality from MT to V1 for the fMRI time series in Fig. 2. In Fig. 3A–D, $F1$ and $F2 \sim F(1, 115)$, $F(1, 114)$, $F(4, 109)$, and $F(4, 108)$, respectively; RSS are also given within Fig. 3. The dotted-dash curves represent the

BOLD response in V1, and the solid curves denote the model output. From the comparison, it is obvious that nonlinear models exhibit smaller RSS (comparing Fig. 3A and C; Fig. 3B and D) when the other parameters are the same. In addition, the systems with experimental input have smaller RSS than the systems without input (comparing Fig. 3A and B; Fig. 3C and D). Finally, the influence between V1 and V2 is stronger than the influence between V1 and MT ($F1 > F2$).

Group analysis results are given in Fig. 4. The results include a total of 96 networks from all six normal subjects (from left hemisphere and right hemisphere of each fMRI run). These are fixed effects inferences and cannot be generalized to a population because we only sampled 96 networks from six normal subjects.

$F1$, $F2$, $F5$ and $F6$ denote the interaction between V1 and V2. $F3$, $F4$, $F7$, and $F8$ are the F tests for "MT influences V1". $F1$, $F3$, $F5$, and $F7$ are the F test results for the linear system, while $F2$, $F4$, $F6$, and $F8$ are the F tests for the nonlinear system. A paired t test was applied to study the difference of different methods in Fig. 4. $F1$ and $F2$ are not significantly different ($t = 1.7240$, $P < 0.05$, two tailed, degree of freedom = 190), suggesting there is not a significant difference between a linear model and nonlinear model for testing V2 influences of V1 when considering
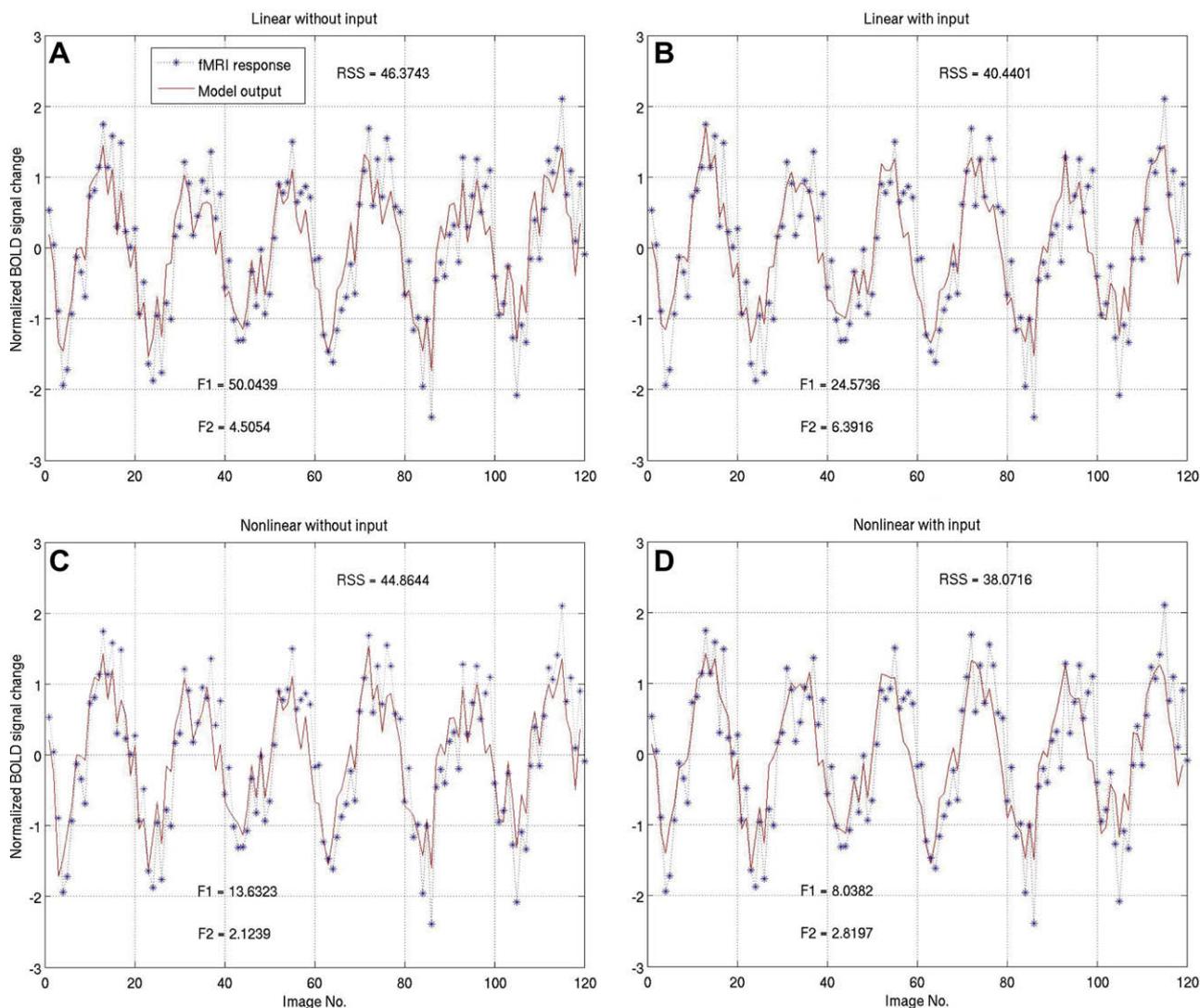
Fig. 3. Results of system identification. Dotted-dash curve represents the BOLD response in V1 (from Fig. 2A); the solid curve denotes the output results of model prediction. A and C are models without experimental input, while B and D are models with input. Solid curves in A and B are predicted from linear models, while solid curves in C and D are derived from the nonlinear models. Note that the autoregression order is 1, so the model output has 119 time points. RSS is the residual sum of square; $F1$ is F test for the influence from V2 to V1; $F2$ is the F test for the influence from MT to V1. Y axis is the image number; X axis is the normalized BOLD signal magnitude (in proportion).

experimental input. However, $F3$ (linear model with experimental input) and $F4$ (nonlinear model with experimental input) are significantly different ($t = 2.1230$, $P < 0.05$, two tailed), indicating that there is significant difference for "MT influence of V1". This is also true for $F5$ and $F6$ ($t = 2.0795$, $P < 0.05$, two tailed), suggesting significant difference for "V2 influence of V1" when no experimental input is considered. $F7$ and $F8$ are significant different ($t = 2.5016$, $P < 0.05$, two tailed), suggesting significant difference for "MT influence of V1" when no experimental input is included. Generally, F values of models without experimental input ($F5$, $F6$, $F7$, and $F8$) are higher than models with experimental inputs ($F1$, $F2$, $F3$, and $F4$), suggesting the influences are stronger if no experimental input is considered. This could be interpreted as the existence of input reduces the interactions between V1, V2, and MT. From a statistical viewpoint, this is due to nothing more than increasing of degree of freedom in the F test. This can also explain why the F value decreases if the nonlinear terms are added in the model.

By reversing the roles of V1 and V2, we investigated the influence between V2 and MT. We found that there was not a significant influence between V2 and MT, but there was a significant influence between V2 and V1. The results are given in Fig. 5 although these are fixed effects inferences.

4. Discussion

4.1. Relation with previous models

In this study, we have proposed an identification method to analyze effective connectivity in fMRI data. It is nonlinear, and can be used for modelling neuronal population systems with or without experimental input. The basic idea of this method is based on nonlinear system identification theory. It enables one to model nonlinear interactions between regions. A few previous studies of effective connectivity have modelled changes in connection strength as a function of activity in different regions (Buchel and Friston, 1997; Moller et al., 2003; Riera et al., 2004; Roebroeck et al., 2005; Yamashita et al., 2005). However, all of these studies differed in at least two ways from the approach presented here. First, previous models were essentially linear; nonlinear interactions were simply neglected or accounted for by including more
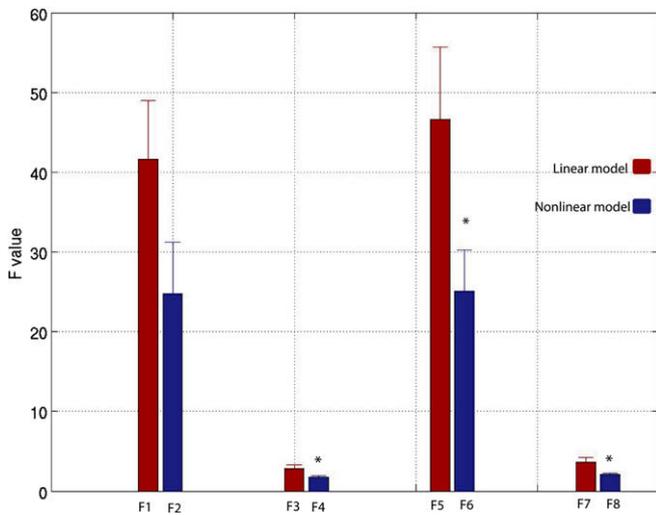
Fig. 4. Group analysis results of visual networks. $F1$, $F3$, $F5$, and $F7$ are linear models; $F2$, $F4$, $F6$, and $F8$ are nonlinear models. $F1$, $F2$, $F3$, and $F4$ are models with experimental inputs; $F5$, $F6$, $F7$, and $F8$ are models without experimental inputs. $F1$, $F2$, $F5$, and $F6$ values from V2 influences of V1; $F3$, $F4$, $F7$, and $F8$ are values derive from MT influences of V1. * denotes significant difference ($t > 1.96$, $P < 0.05$) between linear and nonlinear model.

linear regressors in the model. A second difference is that our method can take experimental input into account. This is different from GCM and ARMA models used in fMRI connectivity studies. These methods, therefore, cannot detect forwarded interaction between experimental input and fMRI response. Our method is more general; ARMA and GCM approaches can be regarded as special cases of our model. The disadvantage of including inputs is that the neuronal system is too complex to know exactly what the neuronal network input is at the neuronal level; only approximations have been used for neuronal populations in fMRI data analysis so far.

4.2. Comparison with DCM

Recently, bilinear and nonlinear DCM have been used to model the interactions in different brain regions (Friston et al., 2003; Stephan et al., 2008). Although DCM and our methods are both based on a nonlinear system identification technique, unlike DCM which builds a model before quantifying brain interactions, our method identifies the connectivity without prior structural model informa-

tion of neuronal populations. Moreover, the computation time for the connectivity estimation may differ. When calculating the model parameters, an EM algorithm is employed in DCM. In the M-step of EM (maximization step), Newton–Raphson method is often adapted, which is computationally demanding because it is iterative. Our method is a non-iteration method (using *pinv* to calculate the matrix inverse in Eq. (7)) which does not need to collect all the data and could be implemented online to study real time dynamic connectivity. If the noise terms in Eqs. (4)–(9), (11)–(13) are autocorrelated, pre-whitening the data (Eq. (4) in Worsley et al. (2002)) so that *pinv* method can be used to estimate the parameters of the model. An alternative method to calculate the matrix inverse is to employ fast orthogonal search (FOS) algorithm which is also a non-iteration method, it may allow a fast search online fMRI networks (Bagarinao et al., 2003; Li et al., 2004).

4.3. F test for GC

Instead of using RSS to qualify the influence causality (Faes et al., 2008), we developed the F test to study the causality between different brain regions. This has at least two advantages. First, it has more statistical power than RSS, because there are situation when RSS is small while the standard derivation is large. This may lead to misleading conclusions. Second, it is easy to test the significant/magnitude of the causality relation. This is achieved simply by testing whether the coefficient of interest is equal to zero or not.

4.4. Nonlinear model

Although no significant differences of F values between linear and nonlinear models were found in our data when including experimental input (Fig. 4$F1$ and $F2$), the nonlinear model is preferable for connectivity studies. This is because the nonlinear model is more accurate than the linear model in term of nonlinear system identification theory. The nonlinear model includes not only linear interactions but also nonlinear ones; this makes the model more realistic and less based on assumptions. Moreover, the nonlinear method has lower RSS in the connectivity study. It is interested to compare Fig. 3B with D and its F values within Fig. 3, we found that the F_2 (MT influence of V1) has decreased from a significant ($F(1,114) = 3.94$, $P < 0.05$) to an insignificant level in Fig. 3D. One reason for this is that the retinotopic mapping stimuli used in the study do not produce strong fMRI response in the region of MT. Another way to compare the linear model with the nonlinear model is to test linear terms against nonlinear terms within the
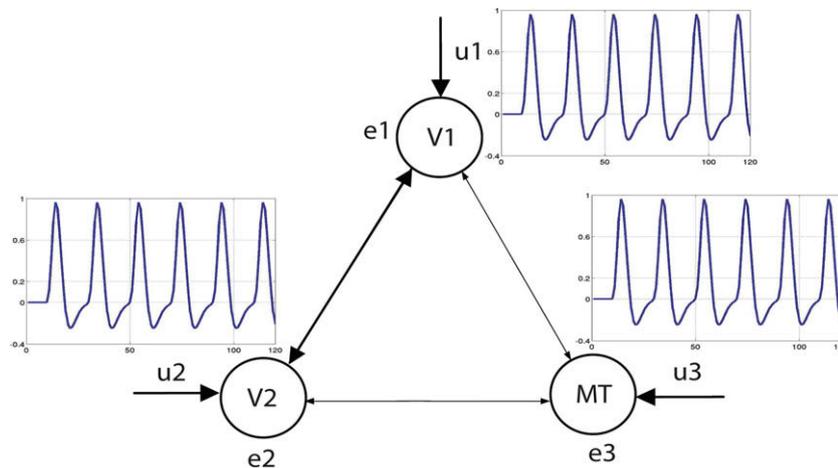

Fig. 5. Group influence results. We found stronger influence (thick line) between V1 and V2. Weaker influences (thin lines) between V1/V2 and MT were also found.

regression model. This can be done by setting the coefficients before linear or nonlinear terms to be zeros (null hypothesis) within the model, then F statistics can be employed to test the hypothesis as given in the Appendix.

For models with experimental input, it is possible to study the forward connectivity in brain networks. This can be demonstrated by the coefficient in front of the experimental input (for example in Eq. (11)). The bigger the coefficient is, the stronger the forward interaction between input and neuronal system will be. In addition, we found that when we add an experimental input into the model, the system RSS is reduced (comparing Fig. 3C with D), although it is not significant (comparing F1 and F5 in Fig. 4). This is expected, because we add more bases in the regression model for the F test, suggesting the advantage of adding the experimental input in the model.

We found a smaller RSS in the nonlinear model than in the linear model although the difference between the two models is not significant (comparing linear with nonlinear model with experimental input, $t = 1.2413$, $P < 0.05$). This could be due to the fact that the nonlinear model used in this study includes only 2nd nonlinearity, the 3rd and the higher order of nonlinearity were simply neglected in the analysis although the total number of polynomials $P_m(t)$ or nonlinearity can be determined by AIC (Akaike, 1974) in theory. These higher orders of nonlinearity could play an important role in reduction of RSS. The future work will entail an investigation of the role of higher order nonlinearity in the connectivity study.

4.5. Possible further extensions and applications

This method could be extended to EEG and MEG studies. Because the time resolution of EEG is higher than fMRI, and the advantage of our nonlinear connectivity modelling would be more evident. This is because the method could be extended to high temporal resolution signals within a long range of autoregression model and it is fast and computation less demanding to estimate the connectivity. Another possible extension would be to apply this method to a fast online large scale network using FOS algorithm to estimate the coefficients (Korenberg et al., 1988a; Korenberg, 1988b; Chen et al., 1989; Zhu and Billings, 1996; Chon et al., 1997). This is interesting especially with the increasing space and time resolution of fMRI. We have applied this method in a human visual network system. It could also be interesting to apply it to other networks such as the motor system or the auditory system. Furthermore, it could be extended to clinical studies. For example, it would be of interest to investigate any deficit to network interactions in patients with impaired cortical function.

In summary, we have proposed a new method to identify connectivity from the fMRI response. The idea and formation of the method relies on nonlinear system identification. Our method can model nonlinear interactions between different brain regions. Our method offers advantages over both present GCM and DCM. In terms of the former, our method computes both linear and nonlinear effects and can take experimental design into account. In terms of later, our method is more efficient being non-iteration and can operate without any prior structural model information of neuronal populations. We illustrated the feasibility of the method in a human visual cortex network. Results showed it can be used to model the nonlinear brain interaction for fMRI data.

Acknowledgements

Appendix. Statistical test for nonlinear connectivity analysis

From Eqs. (4) and (7), and considering the auxiliary system (Doornik, 1996; Edgerton and Shukur, 1999; Kiviet, 1986):

$$Y = W\beta + V, \tag{A.1}$$

where Y is $T \times n$, W is a $T \times K$ linear or nonlinear basis, β is $K \times n$; $V = e$ in Eq. (4) or $V = e_{y_1}$ in Eqs. (5) or (7), and $E[V] = 0$; $E[VV'] = \sigma^2$. The coefficients $\hat{\beta}$ can be obtained from least squares as in Eq. (7), where the residuals are defined by $\hat{V} = Y - W\hat{\beta}$. Testing Granger causality is equivalent to test whether the elements of β are zero. This can be done by Wald, likelihood ratio (LR), and likelihood multiplier (LM) principle (Engle, 1984). Partitioning the coefficients as $\beta = (\beta_1 : \beta_2)$ and $W = (W_1 : W_2)$ accordingly, we can write this test as:

$$H_0 : \beta_2 = 0 \text{ versus } H_1 : \beta_2 \neq 0,$$

with the maintained hypothesis given in (A.1). Defining a R^2-type measures of goodness of fit:

$$R_r^2 = 1 - \left| \hat{V}\hat{V}' \right| \left| \hat{V}_0 \hat{V}_0' \right|^{-1},$$

where $\hat{V}_0$ is the residual from regression of Y on W_1 (that is, under H_0; this is the original system), while $\hat{V}$ results from the auxiliary system (A.1), the corresponding F-approximation to the likelihood-ratio is (Doornik, 1996, Eq. (7)):

$$LMF = \frac{1 - (1 - R_r^2)^{1/r}}{(1 - R_r^2)^{1/r}} \cdot \frac{Nr - q}{np},$$

where: $r = \left(\frac{n^2p^2 - 4}{n^2 + p^2 - 5} \right)^{1/2}$, $q = \frac{1}{2}np - 1$, $N = T - k - p - \frac{1}{2}(n - p + 1)$, and k is the number of regressors in the original system (k is the column of W_1), n is the dimension of system, T is the number of observations, and $p = ns$ (s is the column of W_2). LMF has an approximate $F(np, Ns - q)$ distribution (the F-approximation is exact for fixed regressors when $p \leqslant 2$ or $n \leqslant 2$). When $n = 1$, $\frac{R^2}{1-R^2} \frac{T-k-s}{s} \sim F(s, T - k - s)$, where $R^2 = \frac{RSS_0 - RSS}{RSS_0}$. RSS_0 and RSS are the residual sum of squares (RSS) of the original and auxiliary system respectively, and $RSS = \sum_{i=1}^{T} V^2(i)$.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19, 716–723.

Bagarinao, E., Matsuo, K., Nakai, T., Sato, S., 2003. Estimation of general linear model coefficients for real-time application. NeuroImage 19, 422–429.

Buchel, C., Friston, K.J., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb. Cortex 7, 768–778.

Bullmore, E., Horwitz, B., Honey, G., Brammer, M., Williams, S., Sharma, T., 2000. How good is good enough in path analysis of fMRI data. NeuroImage 11, 289–301.

Chen, S., Billings, S.A., Luo, W., 1989. Orthogonal least squares methods and their application to non-linear system identification. Int. J. Control 50, 1873–1896.

Chon, K., Korenberg, M.J., Holstein-Rathlou, N.H., 1997. Application of fast orthogonal search to linear and nonlinear stochastic systems. Ann. Biomed. Eng. 25, 793–801.

Deneux, T., Faugeras, O., 2006. Using nonlinear models in fMRI data analysis: model selection and activation detection. NeuroImage 32, 1669–1689.

Doornik, J.A., 1996. Testing vector error autocorrelation and heteroscedasticity. Econometric Society 7th Congress, Tokyo.

Dumoulin, S., Bitter, R.G., Kabani, N.J., Baker, C.L., Goualher, G.L., Pike, G.B., Evans, A.C., 2000. A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. Cereb. Cortex 10, 454–463.

Edgerton, D., Shukur, G., 1999. Testing autocorrelation in a system perspective testing autocorrelation. Econ. Rev. 18, 343–386.

Engel, S., Glover, G.H., Wandell, B.A., 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb. Cortex 7, 181–192.

Engle, R.F., 1984. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometricsm, vol. II. North Holland, Amsterdam, pp. 775–826.

Faes, L., Nollo, G., Chon, K.H., 2008. Assessment of Granger causality by nonlinear model identification: application to short-term cardiovascular variability. Ann. Biomed. Eng. 36, 381–395.

Felleman, D., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, 1–47.

Friston, K.J., 2002. Bayesian estimation of dynamical systems: an application to fMRI. NeuroImage 16, 513–530.

Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the balloon model, Volterra Kernels, and other hemodynamics. NeuroImage 12, 466–477.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19, 1273–1302.

Golver, G., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–426.

Gonçalvesa, M.S., Hall, D.A., 2003. Connectivity analysis with structural equation modelling: an example of the effects of voxel selection. NeuroImage 20, 1455–1467.

Granger, C., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

Hinrichs, H., Heinze, H.J., Schoenfeld, M.A., 2006. Causal visual interactions as revealed by an information theoretic measure and fMRI. NeuroImage 31, 1051–1060.

Kiviet, J.F., 1986. On the rigour of some misspecification tests for modelling dynamic relationships. Rev. Econ. Stud. 53, 161–241.

Korenberg, M.J., 1988b. Identification nonlinear difference equation and functional expansion representations: the fast orthogonal algorithm. Ann. Biomed. Eng. 16, 123–142.

Korenberg, M.J., Billings, S.A., Li, Y.P., McIlroy, P.J., 1988a. Orthogonal parameter estimation algorithm for non-linear stochastic systems. Int. J. Control 48, 193–210.

Leontaritis, I., Billings, S.A., 1985a. Input–output parametric models for non-linear systems. Part 1: deterministic non-linear systems. Int. J. Control 41, 303–328.

Leontaritis, I., Billings, S.A., 1985b. Input–output parametric models for non-linear systems. Part 2: stochastic non-linear systems. Int. J. Control 41, 329–344.

Li, X., Tian, J., Wang, X., Dia, J., Ai, L., 2004. Fast orthogonal search method for modelling nonlinear hemodynamic response in fMRI. In: Amini, A., Manduca., A. (Eds.), Medical Imaging 2004: Physiology, Function, and Structure from Medical, San Diego, CA, USA. Proc. of SPIE 5369, pp. 219–226.

Li, X., Dumoulin, S.O., Mansouri, B., Hess, R.F., 2007a. Cortical deficits in human amblyopia: their regional distribution and their relationship to the contrast detection deficit. Invest. Ophthalm. Vis. Sci. 48, 1575–1591.

Li, X., Dumoulin, S.O., Mansouri, B., Hess, R.F., 2007b. The fidelity of the cortical retinotopic map in human amblyopia. Eur. J. Neurosci. 25, 1265–1277.

Marrelec, G., Bellec, P., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Benali, H., Doyon, J., 2008. Regions, systems, and the brain: hierarchical measures of functional integration in fMRI. Med. Image Anal. 12 (4), 484–496.

McIntosh, A., Gonzalez-Lima, F., 1994. Structural equation modelling and its application to network analysis in functional brain imaging. Human Brain Mapp. 2, 2–22.

Moller, E., Schack, B., Vath, N., Witte, H., 2003. Fitting of one ARMA model to multiple trials increases the time resolution of instantaneous coherence. Biol. Cybern. 89, 303–312.

Oxley, L., Greasley, D., 1998. Vector autoregression, cointegration and causality: testing for causes of the British industrial revolution. Appl. Econ. 30, 1387–1397.

Pereda, E., Quiroga, R.Q., Bhattacharya, J., 2005. Nonlinear multivariate analysis of neurophysiological signals. Prog. Neurobiol. 77, 1–37.

Riera, J., Bosch, J., Yamashita, O., Kawashima, R., Sadato, N., Okada, T., Ozaki, T., 2004. fMRI activation maps based on the NN-ARx model. NeuroImage 23, 680–697.

Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25, 230–242.

Sereno, M., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889–893.

Stephan, K., Kasper, L., Harrison, L.M., Daunizeau, J., den Ouden, H.E., Breakspear, M., Friston, K.J., 2008. Nonlinear dynamic causal models for fMRI. NeuroImage 42, 649–662.

Valdes-Sosa, P., Sanchez-Bornot, J.M., Lage-Castellanos, A., Vega-Hernandez, M., Bosch-Bayard, J., Melie-Carcia, L., Canales-Rodriguez, E., 2005. Estimating brain functional connectivity with spare multivariate autoregression. Phil. Trans. Roy. Soc. B 360, 969–981.

Warnking, J., Dojat, M., Guérie-Dugué, A., Delon-Martin, C., Olympieff, S., Richard, N., Chehikian, A., Segebarth, C., 2002. fMRI retinotopic mapping-step by step. NeuroImage 17, 1665–1683.

Wernerheim, C., 2000. Cointegration and causality in the exports-GDP nexus: the post-war evidence for Canada. Empirical Econ. 25, 111–125.

Worsley, K., Liao, C., Aston, J., Peter, V., Duncan, G.H., Morales, F., Evans, A.C., 2002. A general statistical analysis for fMRI data. NeuroImage 15, 1–15.

Yamashita, O., Sadato, N., Okada, T., Ozaki, T., 2005. Evaluating frequency-wise directed connectivity of BOLD signals applying relative power contribution with the linear multivariate time-series models. NeuroImage 25, 478–490.

Zhu, Q., Billings, S.A., 1996. Fast orthogonal identification of non-linear stochastic models and radial basis function neural networks. Int. J. Control 64, 871–886.