Kernel Granger Causality Mapping Effective Connectivity on fMRI Data

Wei Liao, Daniele Marinazzo, Zhengyong Pan, Qiyong Gong, and Huafu Chen*

Abstract—Although it is accepted that linear Granger causality can reveal effective connectivity in functional magnetic resonance imaging (fMRI), the issue of detecting nonlinear connectivity has hitherto not been considered. In this paper, we address kernel Granger causality (KGC) to describe effective connectivity in simulation studies and real fMRI data of a motor imagery task. Based on the theory of reproducing kernel Hilbert spaces, KGC performs linear Granger causality in the feature space of suitable kernel functions, assuming an arbitrary degree of nonlinearity. Our results demonstrate that KGC captures effective couplings not revealed by the linear case. In addition, effective connectivity networks between the supplementary motor area (SMA) as the seed and other brain areas are obtained from KGC.

Index Terms—Effective connectivity, functional magnetic resonance imaging (fMRI), kernel Granger causality (KGC).

I. INTRODUCTION

FUNCTIONAL magnetic resonance imaging (fMRI) is widely used in neuroimaging research to detect changes in the level of activation of brain areas in response to specific tasks. In order to better understand the interactions among distributed activated brain areas in terms of functional integration, it is necessary to distinguish between two types of functional integration among different brain areas: one is functional connectivity, which investigates the correlation between measured time series [1]; the other is effective connectivity, which refers explicitly to the influence that one neuronal system exerts over another, either at a synaptic or population level, either directly or indirectly [1]. Effective connectivity research has become

increasingly important, because it can provide the possible directions of influence between time series of different brain areas, or between fMRI signals that are primarily caused by the blood oxygenation level-dependent (BOLD) effect. In recent years, many methodological advances have made possible the detection of the effective connectivity between brain areas, including covariance structural equation modeling [2], Bayesian estimation of deterministic state-space models [3], and Granger causality mapping [4]–[9].

Granger causality, proposed and formalized by Granger [10], [11], is defined as follows: one series is said to have an influence on if the prediction of made by a linear model is improved when information on is added. Exploring whether and how two time series exert dynamic causal influences on each other is a topic which is becoming increasingly important. Assuming that the hemodynamic response function (HRF) is the same throughout the entire brain, the predictive power of this technique can also be used to detect effective connectivity in the field of cognitive neuroscience [9]. In recent years, Granger causality has been used to detect effective connectivity on both electroencephalography (EEG) data [12]–[16], characterized by high temporal resolution, and fMRI data, characterized by high spatial resolution [4]–[9]. In these latter studies, vector autoregressive modeling of fMRI time series has been used in order to evaluate linear Granger causality in direct, indirect, and instantaneous relationships between a reference region and other brain regions in the time domain [4]–[6], [9] or frequency domain [7], [8].

Despite good results obtained with linear Granger causality [4]–[7], [9] in order to interpret correctly the influence that a brain region exerts over another, it is important to consider the physiological basis of the signal, which is likely to be mainly nonlinear and non stochastic [8]. Logothetis *et al.* [17], [18] suggest that fMRI BOLD response, like the neural responses, could be also a nonlinear function of stimulus contrast; fMRI experiments frequently measure the relationship between stimulus energy and the BOLD response. In general, this relationship is nonlinear-the BOLD response increases according to a compressive, nonlinear, saturating function of stimulus energy. In addition, the mapping from input stimulus sequence to measured BOLD series is recognized as being more complicated than something that can be represented by a set of HRFs that combine linearly to form the observed signal [19]. Changes in physiological states with fMRI experiment, cerebral blood flow, cerebral blood volume, and total deoxyhemoglobin content caused by neuronal activity, are examples of nonlinear functions of physiological parameters [19], [20]. Previous studies investigated the functional connectivity of fMRI data [8], [21], and study [21] performed statistical tests on different fMRI data

sets to infer the nature of the interactions still after defining nonlinear and interaction models in the framework of general linear models; [8] introduced the nonlinear effective connectivity in fMRI data revealed by linear Granger causality. However, nonlinear Granger causality has not yet been applied in fMRI data to assess effective connectivity of brain directly [22].

In this paper, aimed to investigate the nonlinear connectivity of the brain, using fMRI signals [8], we use a kernel Granger causality (KGC) method [23], [24] to explore effective connectivity during a motor imagery task in order to detect the possible directions of influence in the dominant supplementary motor areas (SMA) to other parts of the brain. Previous studies [23], [24] have exploited the properties of kernels to provide nonlinear measures of bivariate or multivariate Granger causality [23]: linear Granger causality has been reformulated, and a statistical procedure to handle overfitting in the linear case has been introduced. This formulation has been generalized to the nonlinear case by means of the kernel trick, thus obtaining a method with the following three main features: 1) KGC captures effective couplings not revealed by the linear case; 2) the nonlinearity of the regression model can be controlled by choosing the kernel function; and 3) the problem of false causalities is addressed by a selection strategy of the eigenvectors of a reduced Gram matrix whose range represents the additional features due to the second time series.

This paper is organized as follows. The proposed method for KGC is introduced in the second section. A simulation test and the fMRI experiment are then described in the third and fourth sections, respectively. Finally, a discussion and conclusions are presented.

II. KGC METHOD

To understand KGC as a generalization of the linear case, we first review the linear case briefly. The temporal dynamics of a stationary time series $\{x(t)\}_{t=1,..,N+m}$, can be described using an autoregressive model based on the past values of the time series $x_n = \sum_{j=1}^{m} \mathbf{a}_j x_{n-j} + \mathbf{e}_n$; in order to include information from a simultaneously recorded time series $\{y(t)\}_{t=1,...,N+m}$, we could consider instead a bivariate autoregressive model [4]–[6], [23], [24] which also takes into account the past values of y

$$x_n = \sum_{j=1}^{m} \mathbf{a}'_j x_{n-j} + \sum_{j=1}^{m} \mathbf{b}_j y_{n-j} + \mathbf{e}'_n$$

where the matrices $\mathbf{a}'$ and $\mathbf{b}$ are called the autoregression coefficients and the $\mathbf{e}'$ is (multivariate) white noise. The coefficients of the models are estimated using a standard least squares optimization; m is the order of the autoregressive model and is usually chosen according to Bayesian information criterion (BIC) [25] that we used in this work, or other order selection criteria (Akaike Information criterion, Hannan-Quinn criterion, and Schwartz criterion).

The concept of Granger causality [10], [11] is that y Granger-causes x if the variance of innovations of $\mathbf{e}'$ is significantly smaller than the variance of innovations of $\mathbf{e}$, as happens when the coefficients $\mathbf{b}$ are all significantly different from zero

[4]–[6], [26]. An index measuring the strength of the causal interaction is then defined as

$$\delta = 1 - \frac{\langle \mathbf{e}'^2 \rangle}{\langle \mathbf{e}^2 \rangle} \tag{1}$$

where $\langle \cdot \rangle$ denotes averaging over n (note that $\langle \mathbf{e} \rangle = \langle \mathbf{e}' \rangle = 0$). Exchanging the roles of the two series, we could evaluate the causality index in the opposite direction $(x \rightarrow y)$.

As in previous studies [23], [24], we use the following shorthand notations:

$$X_i = (x_i, \ldots, x_{i+m-1})^{\mathrm{T}}$$
$$Y_i = (y_i, \ldots, y_{i+m-1})^{\mathrm{T}}.$$

We treat these quantities as N realizations of the stochastic variables X and Y. Then, let $\mathbf{X}$ be an $m \times N$ matrix having vectors X_i as columns, and $\mathbf{Z}$ be a $2m \times N$ matrix having vectors $Z_i = \left(X_i^{\mathrm{T}}, Y_i^{\mathrm{T}}\right)^{\mathrm{T}}$ as columns. The values of α are organized in a vector $\alpha = (x_{1+m}, \ldots, x_{N+m})^{\mathrm{T}}$. Without any loss of generality we can assume that each component of X and Y has zero mean, and that vector α has zero mean and is normalized, i.e., $\alpha^{\mathrm{T}} \alpha = 1$.

Now, for each $i = 1, \ldots, N$, the vectors $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_N)^{\mathrm{T}}$ and $\tilde{\mathbf{x}}' = (\tilde{x}'_1, \ldots, \tilde{x}'_N)^{\mathrm{T}}$ are the estimated values by linear regression, in two cases. It is easy to show that $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}'$ have the following geometrical interpretation. Let $H \subseteq \Re^N$ be the range of the $N \times N$ matrix $\mathbf{K} = \mathbf{X}^{\mathrm{T}} \mathbf{X}$. Then $\tilde{\mathbf{x}}$ is the projection of α on H. In other words, calling P the projector on the space H, we have $\tilde{\mathbf{x}} = P\alpha$. Let us define $\mathbf{u} = \alpha - P\alpha$. Analogously $\tilde{x}' = P'\alpha$, P' being the projector on the $2m$-dimensional space $H' \subseteq \Re^N$, equal to the range of the matrix $\mathbf{K}' = \mathbf{Z}^{\mathrm{T}} \mathbf{Z}$. We now note that $H \subseteq H'$; hence, we may decompose H' as follows: $H' = H \oplus H^\perp$, where $H^\perp$ is the space of all vectors of H' orthogonal to all vectors of H. $H^\perp$ corresponds to the additional features due to the inclusion of y variables. Calling $P^\perp$ the projector on , we can write linear Granger causality index as

$$\delta = \frac{\|p^\perp u\|^2}{1 - \tilde{\mathbf{x}}^{\mathrm{T}} \tilde{\mathbf{x}}}. \tag{2}$$

Linear Granger causality is usually assessed according to well-known test statistics; see, e.g., [27]. Instead of assessing the presence (or not) of causality by means of a single statistical test, and in view of the nonlinear extension, we introduce a causality index which by construction is not affected by overfitting. We note that $H^\perp$ is the range of the matrix $\tilde{\mathbf{K}} = \mathbf{K}' - \mathbf{K}'P - \mathbf{P}(\mathbf{K}' - \mathbf{K}'P) = \mathbf{K}' - \mathbf{PK}' - \mathbf{K}'\mathbf{P} + \mathbf{PK}'\mathbf{P}$.

It follows that $H^\perp$ is spanned by the set of eigenvectors, with nonvanishing eigenvalues, of $\tilde{\mathbf{K}}$. Calling these eigenvectors, we have

$$\delta = \sum_{i=1}^{m} r_i^2 \tag{3}$$

where r_i is the Pearson correlation coefficient of u and t_i. We evaluate the probability p_i that r_i is due to chance using Fisher's r-to-z transformation [28] that can convert the r_i value to the z_i value with zero mean, unit variance, Gaussian distributions under the null hypothesis of no correlation. In order to avoid

false causalities and compensate the threshold of significance for multiple comparison, we use the false discovery rate (FDR) correction [29] to select the eigenvectors t_i, correlating to y, with the expected fraction of false positives equal to 0.05. Then we calculate a new causality index by summing only the $\{r_{i'}\}$ that pass the FDR test, thus obtaining, what we call, a filtered linear Granger causality index

$$\delta_F(y \to x) = \sum_{i'} r_{i'}^2. \tag{4}$$

It is assumed that δ_F measures the causality $y \to x$. By exchanging the roles of the two time series, we may evaluate the causality index $\delta_F(x \to y)$.

The nonlinear Granger causality derived from the linear case has been described completely in previous studies [23], [24], and is based on the theory of reproducing kernel Hilbert spaces. Given a kernel function K, with spectral representation $K(X, X') = \sum_a \lambda_a \phi_a(X) \phi_a(X')$, we consider H, the range of the $N \times N$ Gram matrix K with elements $K_{ij} = k(X_i, X_j)$.

While using both X and Y to predict α, we evaluate the Gram matrix $\mathbf{K}'$ with elements $K'_{ij} = k(Z_i, Z_j)$. The regression values now form the vector $\tilde{x}'$ equal to the projection of α on $\tilde{x}'$, i.e., the range of $\mathbf{K}'$. In the following we consider two choices for the kernel: the inhomogeneous polynomial (IP) of integer order p, and the Gaussian kernel.

A. IP Kernel

The IP kernel [23], [24] of integer order p is $K_p(X, X') = (1 + X^T X')^P$. Along the same lines as described for the linear case, we construct the kernel Granger causality taking into account only the eigenvectors of $\tilde{\mathbf{K}}$ that pass the FDR test

$$\delta_F^K = \sum_{i'} r_{i'}^2. \tag{5}$$

B. Gaussian Kernel

The Gaussian kernel [23], [24] is defined as

$$K_\sigma(X, X') = \exp\left(-\frac{(X - X')^T (X - X')}{2\sigma^2}\right) \tag{6}$$

and depends on the width σ, which controls the complexity of the model: the dimension of the range of the Gram matrix decreases as σ increases. As in previous cases, we may consider H, the range of the Gram matrix $\mathbf{K}$ and H', the range of $\mathbf{K}'$, but in this case the condition $H \subseteq H'$ does not necessarily hold. Therefore, some differences in the approach are necessary. We call L the m_1-dimensional span of the eigenvectors of $\mathbf{K}$, the eigenvalues of which are not smaller than $\mu \lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{K}$ and μ is a small number (we use 10^{-6}). We evaluate $\tilde{x} = P\alpha$, where P is the projector on L. After evaluating the Gram matrix $\mathbf{K}'$, the following matrix is considered:

$$\mathbf{K}^* = \sum_{i=1}^{m_2} \rho_i \mathbf{w}_i \mathbf{w}_i^T \tag{7}$$

where $\{w\}$ are the eigenvectors of $\mathbf{K}'$, and the sum is over the eigenvalues $\{\rho_i\}$ not smaller than μ times the largest eigenvalue of $\mathbf{K}'$. Then we evaluate $\hat{\mathbf{K}} = \mathbf{K}^* - \mathbf{P}\mathbf{K}^* - \mathbf{K}^*\mathbf{P} + \mathbf{P}\mathbf{K}^*\mathbf{P}$, and denote by $P^\perp$, the projector onto the m_3-dimensional range of $\tilde{\mathbf{K}}$. Note that the condition $m_2 = m_1 + m_3$ may not be strictly satisfied in this case. The kernel Granger causality index for the Gaussian kernel is then constructed as in the previous case [see (5)].

C. Multivariate Kernel Causality

Another important extension of Granger's original definition of causality is the consideration of the multivariate case: for three or more simultaneous time series, the causal relation between any two of the series may be direct, may be mediated by a third one, or may be a combination of both (see also the conditional causality proposed by Geweke [30]). To this end, a previous study [24] has handled the multivariate case in more detail. Here, we denote

$$X(c)_i = (x(c)_i, \ldots, x(c)_{i+m-1})^T$$

for $c = 1, \ldots, M$ and $i = 1, \ldots, N$. In order to the causality $\{x(a)\} \to \{x(b)\}$, we define, for $i = 1, \ldots, N$

$$Z_i = \left(X(1)_i^T, \ldots, X(a)_i^T, \ldots, X(M)_i^T\right)^T$$

containing all the input variables, and

$$X_i = \left(X(1)_i^T, \ldots, X(M)_i^T\right)^T$$

containing all the input variables but those related to $\{x(a)\}$. Gram matrices $\mathbf{K}$ and $\mathbf{K}'$ are then evaluated: $K_{ij} = k(X_i, X_j)$ and $K'_{ij} = k(Z_i, Z_j)$. The target vector is now $\alpha = (x(b)_{1+m}, \ldots, x(b)_{N+m})^T$. Also in this case, both for IP kernel or the Gaussian one, we calculate the causality index as in (5); in this case, all the available variables are taken into account.

III. SIMULATION STUDY

A. Nonlinear Model Data

In line with previous linear Granger causality studies on fMRI data [4]–[6], [9], the BOLD fMRI signal can be approximately obtained from a low-pass filtered and subsampled version of local filed potentials (LFPs) based on the conception of the high correlation between LFPs and BOLD fMRI signal [17], [18]. We considered three time series obtained from each of the following coupled autoregressive processes [23], [24]:

$$\begin{aligned}
x_1(t) &= (1 - e)\left(1 - ax_1^2(t-1)\right) \\
&\quad + e\left(1 - ax_2^2(t-1)\right) + s\tau_1(t) \\
x_2(t) &= 1 - ax_2^2(t-1) + s\tau_2(t) \\
x_3(t) &= (1 - e)\left(1 - ax_3^2(t-1)\right) \\
&\quad + e\left(1 - ax_1^2(t-1)\right) + s\tau_3(t) \tag{8}
\end{aligned}$$

where $a = 1.8$, $s = 0.02$, $e = 0.2$, and the τ's are unit variance Gaussian noise terms. The causal relationships implemented in these equations are $2 \to 1$ and $1 \to 3$. Analyzing segments of
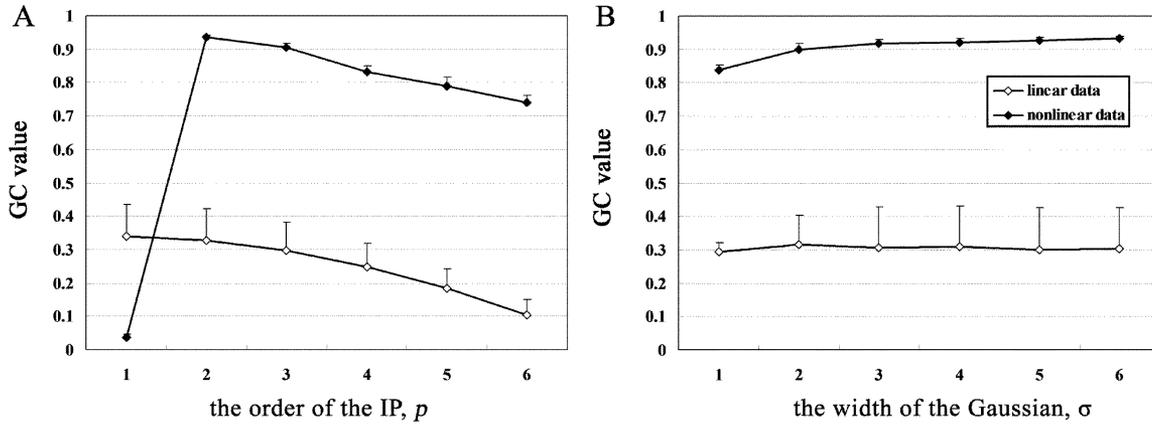
Fig. 1. GC value with different IP order and Gaussian width for the nonlinear (8) and linear (9) time series. The causal index of $1 \rightarrow 3$ both for the nonlinear and linear model data with $m = 1$ chosen by the BIC. (a) The IP kernel with various values of p. (b) Gaussian kernel with various values of σ. Vertical bars indicate estimated standard errors.

length $N = 1000$, we use the IP kernel with various values of p and the Gaussian kernel with various values of σ to show how the two parameters affect the causality index. Then we evaluate the bivariate causality for all pairs of maps within the 100 times simulation with a random τ and some fixed p and σ to verify the validity of KGC.

B. Linear Model Data

Here we compare the performance of KGC introduced here and Geweke's (linear) Granger causality analysis, that is Geweke's dependence measure $F_{x,y}$ [26], which from now on we will abbreviate with GC, that was widely used in fMRI data [4]–[9] for linear data, we also generated three linear time series, similar to those described by Baccala and Sameshima [31]

$$x_1(t) = 0.95\sqrt{2}x_1(t-1) - 0.9025x_1(t-1) + 0.5\tau_1(t),$$
$$x_2(t) = 0.8x_1(t-1) + 0.5\tau_2(t),$$
$$x_3(t) = -0.7x_1(t-1) + 0.5\tau_3(t) \qquad (9)$$

where the τ's are unit variance Gaussian noise terms. The causal paths modeled in these equations are $1 \rightarrow 2$ and $1 \rightarrow 3$. We also evaluate the bivariate causality for all pairs of maps within 100 times simulation using GC [4]–[6], [9].

C. Results

The results of the IP kernel with various values of p and the Gaussian kernel with various values of σ are shown in Fig. 1. For the 100 times simulation, the optimal order of the model, m, is chosen by the BIC criterion [25]. The optimal order is always equal to 1 for linear and almost always equal to 1 (more than 90% of the times) for the nonlinear model, so we stick to this values of m for all the simulations. Fig. 1(a) shows that when $p = 1$ (which corresponds to GC), the causality index of $1 \rightarrow 3$ of the nonlinear model data (solid diamond) is very low. However, it is significantly higher (up to about 0.7) with a p value between 2 and 6. On the other hand, the causality index of $1 \rightarrow 3$ of the linear model data (clear diamond) is relatively higher (up

to about 0.35) when $p = 1$. However, it is relatively lower with a p value between 2 and 6. This results suggest that when $p = 1$ KGC method also can effectively evaluate the Granger causality for the linear time series; when the $P \geq 2$ this KGC method properly evaluated the Granger causality for the nonlinear time series.

For Gaussian kernel Granger causality method, Fig. 1(b) shows that with $\sigma = 1$ to 6, the causality index (up to about 0.8) is consistent with the given influence assumption for nonlinear model data.

Based on the results of the simulation study using different values of p and σ, we then select a fixed value of $p = 2$ for the IP kernel and a fixed value of $\sigma = 6$ for Gaussian kernel for each pair of simulated data including both the linear and nonlinear model data mentioned above. To compare the validity of detecting causality, the results of the $p = 1$ (linear Granger causality) are also evaluated. The six pairwise causality indices computed by each method for each time series are displayed in Fig. 2. Fig. 2(a) shows that IP kernel, $p = 1$ (linear Granger causality) can evaluate the given influences $1 \rightarrow 2$ and $1 \rightarrow 3$ [(9)], which are also revealed by the $p = 2$ and Gaussian kernel method. However, the Fig. 2(b) clearly shows that both the IP kernel, $p = 2$, and Gaussian kernel method can evaluate the given influences $2 \rightarrow 1$ and $1 \rightarrow 3$ [(8)], which are not revealed by the $p = 1$. Also, both the IP kernel, $p = 2$, and Gaussian kernel methods reveal the influence $2 \rightarrow 3$, for which the kernel Granger causality index is significant lower than in the influences $2 \rightarrow 1$ and $1 \rightarrow 3$. Although the causal relationships of $2 \rightarrow 1$ and $1 \rightarrow 3$ are defined in the simulated data, indeed, the influence $2 \rightarrow 3$ is actually mediated by 1. Furthermore, the lower indirect influence value of $2 \rightarrow 3$ is revealed by KGC.

It is well known [15] that a pairwise evaluation for multivariate data has the limits that one cannot discern whether the influence between two time series is direct or is mediated by the other [30]. Fig. 3 shows that the influences $2 \rightarrow 1$ and $1 \rightarrow 3$ in both bivariate and multivariate analysis. The bivariate analysis revealed also a slight causal influence $2 \rightarrow 3$, while the multivariate analysis labelled it as non significant (the causality index was equal to zero).
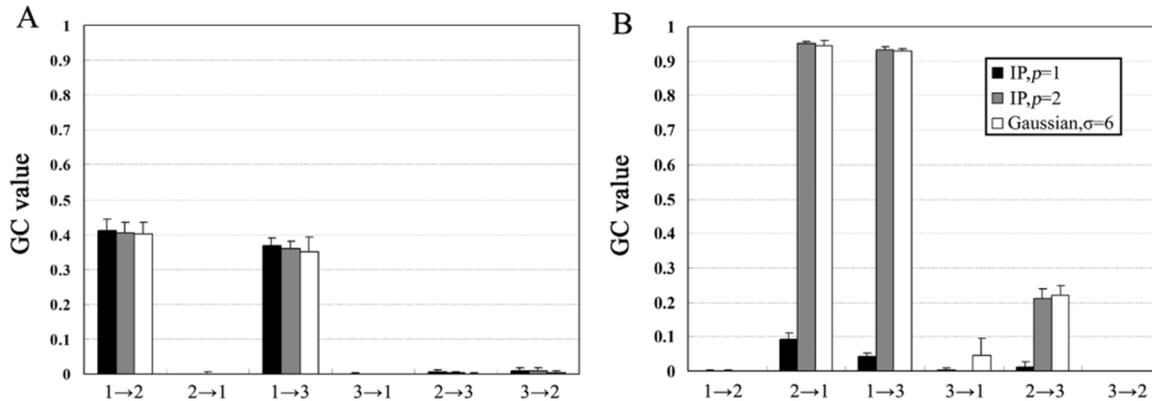
Fig. 2. Comparison of the three GC models with simulated data including the nonlinear (8) and linear (9) time series. (a) The causal relationships between all pairs of linear time series are evaluated along with $m = 1$ chosen by the BIC. (b) The causal relationships between all pairs of linear time series are evaluated along with $m = 1$ chosen by the BIC. Vertical bars indicate estimated standard errors.
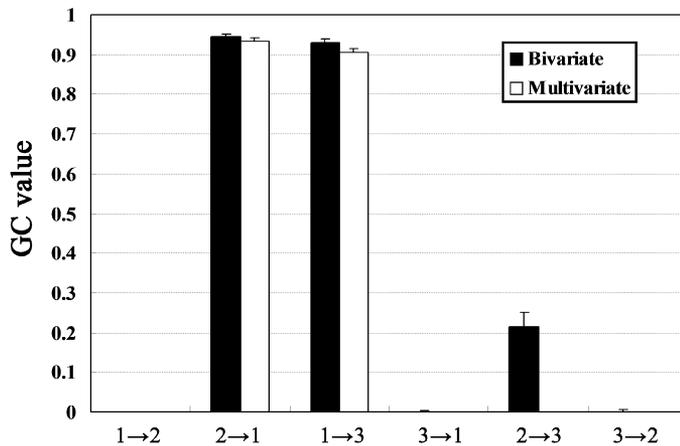

Fig. 3. Comparison of the bivariate and multivariate GC model. The causal relationships between all pairs of voxels compared between the bivariate and multivariate analysis. Here $m = 1$ is chosen by the BIC, the IP kernel is used with $p = 2$. Vertical bars indicate estimated standard errors.

IV. FMRI DATA TEST

A. Experimental Paradigm and Data Acquisition

Ten subjects perform a left hand motor imagery task. After a stimulus lasting for 4 s, subjects try to imagine that they are tapping their fingers in the same rhythm as defined by the stimulus. Subjects continue imagining the finger tapping for 10 s until a cue stimulus informs them to actually perform what they have imagined. The actual finger tapping lasts for 6 s. After that, the subjects rest for 10 s. All visual stimuli are presented on a computer monitor. Each run includes 10 trials and each trial lasts for 30 s (4 s to present the sequence, 10 s for the imagery, 6 s for the tapping, and 10 s to rest). In total, each run contains 155 volumes. Functional imaging data was gathered on a 3.0 T scanner, GE-Signa (Huaxi MR Research Center, Chengdu, China) with a $T2*$-weighted the gradient echo, echo-planar-imaging (EPI) sequence parameters as follows: 30 transverse slices, $\mathrm{TR} = 2000\,\mathrm{ms}$, $\mathrm{TE} = 30\,\mathrm{ms}$, $\mathrm{FOV} = 24\,\mathrm{cm}$, matrix $= 64 \times 64$, voxel size $= 3.75 \times 3.75 \times 5\,\mathrm{mm}^3$ (without gap), flip angle $= 90°$. A structural image was also acquired in the interval during the first and second run (volume size $512 \times 512 \times 156$ with a voxel size of $0.47 \times 0.47 \times 1\,\mathrm{mm}^3$).

B. Data-Processing Procedure

The experimental data were preprocessed using statistical parametric mapping software (SPM2, http://www.fil.ion.ucl.uk/spm). To allow for magnetization equilibrium, the first five images with no stimulus were discarded. The remaining 150 images were first corrected for the acquisition time delay among different slices, and then the images were realigned to the first volume for head-motion correction. Second, a high-pass filter with a cutoff of 1/128 Hz was used to remove low-frequency noise. Third, the data were spatially normalized into standard stereotaxic space at $2 \times 2 \times 2\,\mathrm{mm}^3$, using Montreal Neurological Institute EPI template of SPM2 software. The data were then spatially smoothed to each brain's 3-D volume by convolution with an isotropic Gaussian kernel ($\mathrm{FWHM} = 8\,\mathrm{mm}$) to increase the MR signal to noise ratio. Fourth, linear trends were removed from the data to eliminate the effect of gross signal drifts, which could be caused by scanner instabilities and/or a gross physiological change in the subject. Finally, in this paper, the region of interest (ROI) was defined as a sphere within the activated area located in the supplementary motor area (SMA), centered at the voxel of the highest statistical value and a radius of 10 mm [6].

C. Effective Connectivity Analysis

In this paper, we only paid attention to the effective connectivity of the motor imagery condition; therefore, the other three phases (sequence informing, motor execution, and resting) in the time series of fMRI BOLD signals for each subject were excluded by regression. Several other preprocessing steps for fMRI time series were implemented to reduce spurious variances. First, regression of six head motion parameters was performed. That is, each regional mean time series was further corrected for the effect of head movement by regression on the translations and rotations estimated in the image realignment procedure. The residuals of these regressions constituted the set of regional mean time series used for further analyses. Second,

TABLE I
SUMMARY OF THE RESULTS OF GROUP ANALYSIS FOR IP $(p = 2)$ KGC

Significantly connected area	Hem	BA	MNI Coordinates			Volume (cm³)	Max. values
			X	Y	Z		
the regions influenced by the seed (SMA) ($X \to Y$)							
Cerebellum	L		-24	-86	-32	0.31	0.0481
SMarg	L	40/42	-56	-32	24	0.30	0.0455
CMA	L	23/24	0	-4	44	3.18	0.0950
	R	23/24	2	-4	44	1.93	0.0863
Postcentral	L	1/2/3/4	-12	-38	82	0.25	0.0512
	R	1/2/3/4/5	28	-48	72	5.23	0.0735
Precentral	R	3/4/6	46	2	38	1.78	0.0546
IPL	R	39/40	46	-36	56	0.31	0.0456
SPL	L	5/7/40	-28	-56	70	2.14	0.0643
	R	2/5/7/40	26	-48	72	3.62	0.0702
SMA	L	6/23/24/32	0	2	46	1.75	0.082
	R	6/23/24/32	0	-2	46	1.02	0.082
Paracentral lobule	L	3/4/5	-2	-38	74	1.33	0.0634
	R	4/5	2	-38	74	1.74	0.0599
Putamen	L	48	-26	6	0	0.93	0.0504
	R	34/48	30	4	-10	0.34	0.044
the regions influencing the seed ($Y \to X$)							
Cerebellum	L		-12	-84	-34	1.34	0.0466
SMarg	L	2/3/40/41/48	-52	-40	28	0.77	0.0398
CMA	L	23/24/32	0	8	32	0.21	0.0354
	R	23/24/32	2	6	30	0.17	0.0361
Postcentral	L	1/2/3/4	-52	-8	44	3.88	0.0528
	R	3/4	58	-6	38	0.52	0.0408
Precentral	L	3/4/6	-20	-20	76	2.09	0.0644
	R	4/6	12	-20	80	0.53	0.0458
IPL	L	39/40	-48	-56	40	1.18	0.0382
SPL	L	2/7/40	-24	-64	58	1.54	0.0414
	R	5/7	18	-68	64	0.77	0.0385
SMA	L	6	0	-10	56	0.23	0.0388
	R	4/6	12	-16	74	1.31	0.0412
Paracentral lobule	R	4/5/6	10	-22	80	0.94	0.0514

Abbreviations: BA=Brodmann's area, Hem=hemisphere, L=left, R=right, CMA=cingulate motor area, IPL=inferior parietal lobule, SMA=supplementary motor area, SPL=superior parietal lobule, SMarg=SuparaMarginal.

the time series in blood vessels and the cerebrospinal fluid was regressed.

For the effective connectivity analysis, a seed x was first defined as the average signal of all the activated voxels within such a ROI. Then, the seed x was chosen to accomplish effective connectivity assessment of the SMA against the rest of the voxels in whole brain separately for each subject using the bivariate IP and Gaussian KGC and GC [4]–[6], [9] method. And, the optimal order for the fMRI data is also used selected by BIC like the simulations. Moreover, the optimal order is almost always equal to 1 for real fMRI data. Subsequently, in order to extend inference of individual statistical analyses to the general population from which the subjects were drawn, a group analysis was performed by averaging the ten maps of $F_{x \to y}$ across ten subjects to get the averaged $\overline{F}_{x \to y}$ map for $F_{x \to y}$, as well as for $F_{y \to x}$, in line with our previous paper [6]. Since the distribution of the indices was unknown, to assess the significance of Granger causality at the group-level, we used the framework of a parametric estimation based on bootstrap methodology [32] for both GC and KGC. This approach was widely employed in the analysis of fMRI studies [5], [6], [9]. The major steps were as follows. 1) $\overline{F}_{x \to yi}$, which has the same size of $\overline{F}_{x \to y}$, was obtained from randomly sampling $\overline{F}_{x \to y}$ for N times (N was the size of $\overline{F}_{x \to y}$) with repetition.

2) Step 1) was repeated for a large number B of times, so that $\overline{F}_{x \to y2}, \ldots, \overline{F}_{x \to yB}$ were calculated. $B = 1000$ was used in our study. In general, the value of $B = 200 \sim 5000$ is sufficient [32]. 3) $\overline{F}_{x \to y1}, \overline{F}_{x \to y2}, \overline{F}_{x \to y3}, \ldots, \overline{F}_{x \to yB}$ were pooled together to acquire the null distribution of $F_{x \to y}$, which is reasonable to approximation to the Gaussian distribution under the large samples, so that we gained the p value for $F_{x \to y}$ of each voxels.

When performing simultaneous statistical tests for all voxels, we assessed the significance thresholds by correcting for multiple comparisons using the FDR [29], so that an expected proportion of false positives among all tests was rejected by the null hypothesis $F_{x \to y} = 0$. This was favorable for dealing with multiple comparison problems while retaining considerable power in the detection of effects and adapting to the noise level in the data. The FDR corrected thresholds corresponding to the synchronization difference index for the term being nonnegative $\overline{F}_{x \to y}$ at an accepted FDR level q were acquired from the set of p values obtained by the bootstrap methodology over all voxels. These p values obtained at a given difference term were then regarded as the proportion of larger values in the null distribution. Subsequently, the FDR corrected threshold was achieved from the p values as follows: Given the ordered collection of p values, let r be the largest i for which $p[i] \leq (i/N) \times (q/(c(N)))$, then
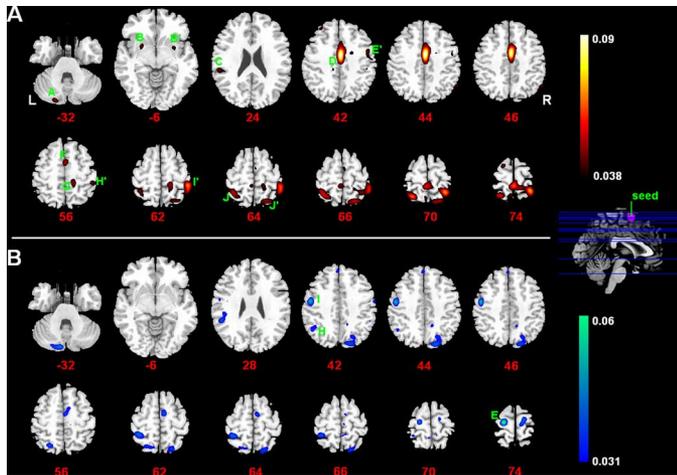
Fig. 4. Kernel Granger causality map of the regions influenced by the seed (SMA) (a) and the regions influencing the seed (SMA) (b) for real fMRI data using IP kernel with $p = 2$. (L: left, R: right. A: cerebellum; B: left putamen; B′: right putamen; C: suparaMarginal; D: cingulate motor area; E: left precentral; E′: right precentral; F: supplementary motor area; G: paracentral lobule; H: left inferior parietal lobule; H′: right inferior parietal lobule; I: left postcentral; I′prime: right postcentral; J: left superior parietal lobule; J′: right superior parietal lobule).
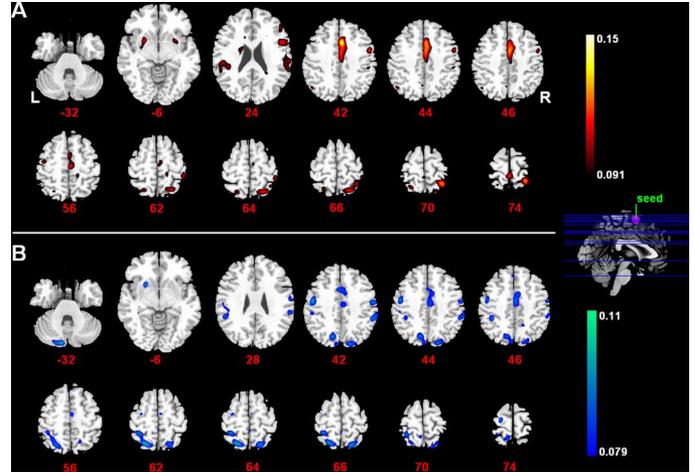

Fig. 5. Kernel Granger causality map of the regions influenced by the seed (SMA) (a) and the regions influencing the seed (SMA) (b) for real fMRI data using Gaussian kernel with $\sigma = 6$.

the threshold was set at the value corresponding to the p value $p[r]$. Where $c(N)$ was a constant determined by assumptions on the joint distribution of p values over all voxels. In this paper, it was set to $c(N) = \ln(N) + r$, where $r = 0.5772$ is Euler's constant, which applied to any distribution of p values over voxels [29]. More details for this approach could be found in a previous fMRI study [5]. It is worth to remark that the FDR correction was applied *a posteriori* only for GC, because FDR correction is already implemented in KGC when we select the eigenvectors t_i which are correlated with y [see (4)].

D. IP Kernel Results

Table I shows a summary of the group analysis results for IP kernel with $p = 2$ along with order $m = 1$ for the effective connectivity between the seed (SMA, BA 6) and other brain regions during the left-hand motor imagery. The coordinates of the center of the significantly interacting areas, along with the cluster size, are characterized in detail.

Fig. 4 shows two influence measures, the $F_{x \to y}$ and $F_{y \to x}$ maps (FDR corrected, $p = 0.01$) using the IP kernel with $p = 2$. In the map of $F_{x \to y}$ (Fig. 4(a)), the bilateral superior parietal lobule (SPL, BA 5,7,40), cingulate motor area (CMA, BA 23,24), postcentral (BA 1,2,3,4), paracentral lobule (BA 4,5) and putamen (BA 48); the contralateral inferior parietal lobule (IPL, BA 39,40) and precentral (BA 3,4,6) and the ipsilateral SuparaMarginal (SMarg, BA 40,42) and cerebellum were extensively influenced by the seed (SMA, BA 6). The map of $F_{x \to y}$ [Fig. 4(b)] shows the regions directly influencing the seed. Besides the bilateral SPL, CMA, SMA, postcentral and precentral, the influence of the following regions can be clearly observed: the contralateral paracentral lobule and the ipsilateral SMarg and cerebellum.

E. Gaussian Kernel Results

Fig. 5 also shows two influence measures, the $F_{x \to y}$ and $F_{y \to x}$ maps (FDR corrected, $p = 0.01$) using Gaussian kernel with $\sigma = 6$. In the map of $F_{x \to y}$ [Fig. 5(a)], the bilateral SPL, SMarg, CMA, SMA, postcentral, precentral, paracentral lobule, putamen and the superior temporal lobe (STL, BA 38,42); the contralateral cerebellum were extensively influenced by the seed. The map of $F_{y \to x}$ [Fig. 5(b)] shows the regions directly influencing the seed. Besides the bilateral SPL, SMarg, CMA, SMA, postcentral and precentral, the influence of the following regions can be clearly observed: the contralateral paracentral lobule and the ipsilateral putamen and cerebellum.

F. Comparisons of the Causality Results Obtained With KGC and GC

We compare the causal connectivity results showing the $F_{x \to y}$ and $F_{y \to x}$ maps (FDR corrected, $p = 0.01$), obtained from KGC (IP kernel with $p = 2$), and KC method. The brain regions which were involved in causal influences revealed by both methods are listed in Table II. The coordinates of the center of the significantly interacting areas, along with the cluster size, are characterized in detail. In Fig. 6 we display three Granger causality sub-maps: brain regions found only by KGC (shown in red), brain regions found only by GC (shown in green), and common brain regions (the regions found by KGC intersecting the regions found by GC) (shown in yellow). In the map of the regions influenced by the seed (SMA) [Fig. 6(a)], we find that the significant causal influences from the seed to the bilateral CMA, SMA, SPL, paracentral lobule, putamen, and the contralateral postcentral were found by KGC only. Significant influences from the seed to the bilateral IPL and SMA were detected only with GC. Regarding the regions influencing the seed (SMA) [Fig. 6(b)], KGC detected strong influences from the bilateral SPL and the contralateral postcentral and precentral. GC only found that the part of SMA extensively influenced the seed.

TABLE II
CAUSAL INFLUENCES FOUND BY BOTH KGC AND GC

Significantly connected area	Hem	BA	MNI Coordinates			Volume (cm³)	Max. values
			X	Y	Z		
the regions influenced by the seed (SMA) ($X \rightarrow Y$)							
Cerebellum			0	-36	-16	0.12	0.0539
Postcentral	R	1/2/3/4/5	28	-48	72	1.55	0.0735
SPL	R	2/5/7/40	26	-48	72	1.23	0.0702
SMA	L	6	0	2	54	0.27	0.0568
	R	6	2	0	56	0.12	0.0471
the regions influencing the seed ($Y \rightarrow X$)							
Cerebellum	L		-10	-86	-32	0.21	0.0415
SMA	L	6	0	-10	56	0.19	0.0388
	R	4/6	4	-6	60	0.59	0.0404

Abbreviations: BA=Brodmann's area, Hem=hemisphere, L=left, R=right, SMA=supplementary motor area, SPL=superior parietal lobule.
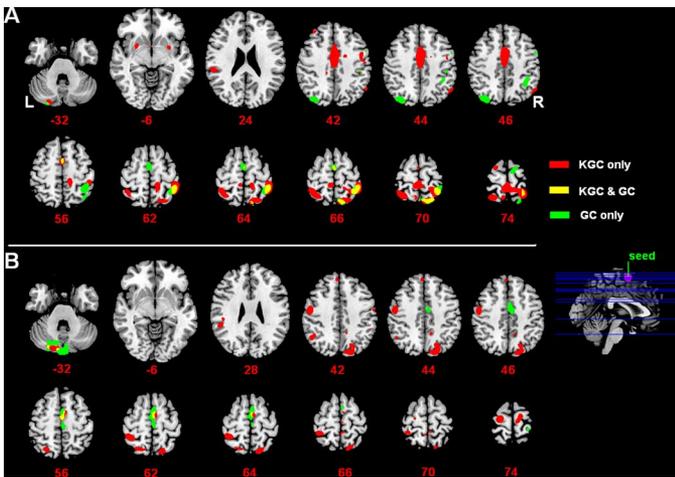

Fig. 6. Granger causality map of the regions influenced by the seed (SMA) (a) and the regions influencing the seed (SMA) (b) for real fMRI data. Three Granger causality submaps: brain regions found only by KGC (shown in red), brain regions found only by GC (shown in green), and common brain regions (the regions found by KGC intersecting the regions found by GC) (shown in yellow).

V. DISCUSSION

A. Kernel Granger Causality

We have used a recently proposed approach to Granger causality, based on kernel methods [23], to analyze effective connectivity. In this paper, in line with the definition of nonlinear kernel Granger causality, our approach allows the analysis of the possible directions of influence between time series of different brain areas on fMRI data [6]. The linear Granger causality has been widely used to map directed influence over the brain using fMRI data [4]–[7], [9] and EEG data [12]–[16]. A hemodynamic response has been associated with changes in physiological states, including cerebral blood flow, cerebral blood volume and total deoxyhemoglobin content, caused by neuronal activity, all of which are nonlinear functions of physiological parameters [19], [20]. Linear models of the estimation of causal pathways in fMRI data are necessarily limited in their ability to capture the characteristics of an intrinsically nonlinear signal. To our knowledge, therefore, this is the first study that proposed kernel Granger causality

characterizing the effective connectivity on fMRI data. In the simulation study, our approach that including the bivariate and multivariate analysis measures Granger causality of time series, assuming an arbitrary degree of nonlinearity, while controlling overfitting, and thus avoids the problem of false causalities (Fig. 3). More importantly, our approach can efficiently detect causal relationships under the assumption of nonlinearity of fMRI data.

B. Spurious Causalities

On the other hand, it is important to eliminate spurious causalities [11] that may appear when time series are both influenced by the other external sources that are not taken into account from the Granger causality analysis. To detect causal processes correctly, it is important to take into account a possible time lag between the measurements of the two time series. A recent study based on partial Granger causality [16] has tackled this issue, revealing the underlying interactions among elements in a network in the presence of exogenous inputs and latent variables.

This problem is particularly important in fMRI time-series: A certain area A can cause a response in another area B through neuronal activity, but the hemodynamic responses in A might appear much later than those in B. This would imply that B Granger-caused A. This is not an issue for EEG data [12]–[16], because in this case one does not have to consider regional variations in the HRF. For our kernel Granger causality analysis, we had to assume that the HRF is exactly the same everywhere in the brain [6]. This raises the hypothesis that interregional variability of the HRF might generate spurious inferences of connectivity [7]. However, the impact of variable hemodynamic delays between regions is tolerable in practice with a TR of up to 2 s, which is quite realistic in many fMRI studies [7]. In this work, like in previous studies [4]–[6], [9], we performed a whole brain voxels-wise analysis using only bivariate KGC. Indeed, the presence of much more variables (voxels) than observations (the length of fMRI time series) makes whole brain voxels-wise approach a mathematically ill-posed problem for multivariate KGC analysis. The number of variables would be greatly reduced using ROIs rather than the whole brain data [7],

[8]. Our opinion is that there are some limits in ROI-based method, where the selection of those ROI in advance must imply a prior knowledge of the problem that could lead to the lost of potentially interesting brain areas. On the other hand, using bivariate analysis in this study meant that our assessment of effective connectivity could possibly ignore the influences of other areas when assessing the coupling between the reference region and any particular voxel. A subsequent analysis, that we plan to further pursue, could consist of two successive steps: first, a whole brain voxels-wise bivariate KGC will be performed, in order to individuate clusters of voxels which seem to belong to the same causal flow, and then multivariate KGC on time series representative of these clusters will be performed to gain insight on the directions of influence through the whole brain.

C. Effective Connectivity Network Between the SMA and Other Brain Regions

In this study, we mainly focused on the effective connectivity of the motor imagery condition; the time series of fMRI BOLD signals in those three phases (sequence informing, motor execution, and resting) for each subject were excluded by regression [33]. So we consider the remaining signal after regression, the contribution from them should be reasonably small. Therefore, it should be reasonable to regard the identified connectivity networks as the effective connectivity of the motor imagery condition.

Our results showed a forward and backward effective connectivity loop between the SMA and SPL (Fig. 4 and Table I). This conjunction is consistent with previous neurophysiological studies, which showed matching patterns of neuronal activity within the frontal and parietal cortices [34]. As in studies of monkeys, the human frontal and parietal areas are organized as a tightly coupled functional system with highly specific connections between their respective anatomical subdivisions [34]. Moreover, the parietal cortex has been found to play a role in generating mental movement representation supported by neuropsychological examination [35] and in sensorimotor mapping of temporal and spatial relations. Also, the connectivity between the SMA and SPL in our study could be verified by the study of Rizzolatti *et al.* [36], in which area F3 of the monkey brain, which is homologous to the SMA of humans, was found to be strongly connected with the SPL.

In addition, a forward and backward effective connectivity loop was also detected from the SMA to the postcentral that is mostly considered as the primary somatosensory cortex (S1) [Fig. 4(a) and Table I]. These results corroborate that the S1 is influenced by the SMA, as shown in a previous motor imagery fMRI study [37]. As during a motor task, SMA activity occurs before S1 activity [38]. However, several motor imagery studies considered that the S1 only participated in motor execution. A loop of effective connection was also found within the SMA and bilateral postcentral cortex during the left motor imagery. These findings are demonstrated by the study of Matsumoto *et al.* [39], which detected bi-directional connectivity between the SMA and S1.

D. Different Effective Connectivity Results Between KGC and GC

GC [4]–[6], [9] is based on linear vector autoregression models. Nonetheless, there is evidence that the hemodynamic processes associated with changes in physiological states, including cerebral blood flow, cerebral blood volume and total deoxyhemoglobin content are nonlinear [19]–[22]. For instance, it is known that the putamen plays an important role in motor imagery using fMRI [34], [35]; in our study we found strong influences from the seed to the bilateral putamen only using KGC method. These influences have not been detected by GC. However, the most recent data suggest that performance of a sequence is mediated by a striatal-cortical circuit that includes motor cortical regions (SMA), cerebellar dentate, thalamus, and putamen [40]. Our causality connectivity analysis between those two brain regions also reveals this circuit when imaging a sequence. Although KGC will be able to identify additional regions not identified by GC, the neurobiological basis for those nonlinear directions is not explained well. One possible explanation may be that many complex systems, including the brain, are complex, nonlinear, and operate far from equilibrium [41]. That is why KGC are maybe appropriate to model it than linear methods to characterize the nonlinear causal influence in the brain. GC only found that the bilateral IPL and SMA were extensively influenced by the seed [Fig. 6(a)], and that the part of SMA extensively influenced the seed [Fig. 6(b)]. In our study we find some remarkable differences between causal influences evaluated with KGC and GC. GC found that some parts of the SMA influenced the seed [slices 44–56 Fig. 6(b)] while KGC found that some parts of the SMA in the same slice location are influenced by the seed [Fig. 6(a)]: here we could assume that the causal influences could be mainly linear in one direction and mainly nonlinear in the other. Furthermore we can assume that noise levels and significance tests play here an important role: as noise increases, and for a given time series length, nonlinear influences become more difficult to detect, and their statistical significance decreases more rapidly than for linear influences [23]. We do not state which method is better using the fMRI data to characterize the effective networks in motor imagery. Rather, KGC method captures effective couplings not revealed by GC. More extensive biologically meaning explanations of effective connectivity of supplementary motor areas during motor imagery using the conventional GC will be reported elsewhere (Chen *et al.* [42]).

E. Limitations and Future Works

Several limitations are also important to mention. It is important that the effective connectivity between neural processes and the observed BOLD signal is a convolution of these processes by a HRF [4], [5], [17], [18]. Logothetis *et al.* [17] suggested that, like the neural responses, fMRI BOLD response was also found to be a nonlinear function of stimulus contrast [43], [44]; however, a linear systems analysis on the fMRI responses predicted a linear relationship between the BOLD and neural activity [43]. So, many of the more recent effective connectivity approaches are based on stochastic or deterministic dynamic models, and

capable of capturing temporal structure. Volterra series representation [45] characterizes interactions in a nonlinear convolution model relating multiple inputs to a single output. Thus dynamic nonlinear influences on a single region can be characterized. So, more realistic (physiologically) simulations will be needed to explore in future works (Balloon, Windkessel, or Volterra models) to characterize interactions in a nonlinear convolution model relating multiple inputs to a single output.

An additional development of Granger's causality idea is remarkable. Natural time series, including ones from economics and neurobiology, contain oscillatory aspects in specific frequency bands [46]. Thus, frequency domain Granger causality [7], [8] is desirable to have a spectral representation of causal influence. Major progress in this causal direction has been made by Geweke [26], [30]. According to Geweke's variance decomposition at a particular frequency, total power includes an intrinsic power and a casual power. Time domain Granger causality is the integral of the frequency domain causality over the entire frequency range. In this sense, the time domain Granger causality, which was used in the current study, may result from strong spectral peak(s). Future work, however, will be also needed to explore whether there is a power spectrum peak for a significant direction, which would provide more insights on the effective connectivity.

VI. CONCLUSION

We have proposed kernel Granger causality to detect effective connectivity between different brain areas using fMRI data. The validity of this algorithm has been also tested in a simulation study. The results show that our method can identify pathways of causal influence between nonlinear time series and thus could provide new insights on the effective connectivity between brain regions.

REFERENCES

[1] K. J. Friston, C. D. Frith, and R. S. Frackowiak, "Time-dependent changes in effective connectivity measured with PET," *Hum. Brain Mapp.*, vol. 1, pp. 69–79, 1993.

[2] C. Buchel and K. J. Friston, "Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI," *Cereb. Cortex*, vol. 7, pp. 768–778, 1997.

[3] L. Harrison and W. Penny, "Dynamic causal modeling," *NeuroImage*, vol. 19, pp. 1273–1302, 2003.

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

[5] A. Roebroeck, E. Formisano, and R. Goebel, "Mapping directed influence over the brain using Granger causality and fMRI," *NeuroImage*, vol. 25, pp. 230–242, 2005.

[6] Q. Gao, H. Chen, and Q. Gong, "Evaluation of the effective connectivity of the dominant primary motor cortex during bimanual movement using Granger causality," *Neurosci. Lett.*, vol. 443, pp. 1–6, 2008.

[7] G. Deshpande, X. Hu, R. Stilla, and K. Sathian, "Effective connectivity during haptic perception: A study using Granger causality analysis of functional magnetic resonance imaging data," *NeuroImage*, vol. 40, pp. 1807–1814, 2008.

[8] G. Deshpande, S. LaConte, S. Peltier, and X. Hu, "," in . : , "Connectivity analysis of human functional MRI data: From linear to nonlinear and static to dynamic," *Lecture Notes Comput. Sci.*, vol. 4091, pp. 17–24, 2006.

[9] J. Upadhyay, A. Silver, T. A. Knaus, K. A. Lindgren, D. S. Kim, and H. Tager-Flusberg, "Effective and structural connectivity in the human auditory cortex," *J. Neurosci.*, vol. 28, pp. 3341–3349, 2008.

[10] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424–438, 1969.

[11] W. J. Granger, "Testing for causality: A personal viewpoint," *J. Econ. Dyn. Contr.*, vol. 2, pp. 329–352, 1980.

[12] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality," in *Proc. Nat. Acad. Sci.*, 2004, vol. 101, pp. 9849–9854.

[13] X. Wang, Y. Chen, and M. Ding, "Estimating Granger causality after stimulus onset: A cautionary note," *NeuroImage*, vol. 41, pp. 767–776, 2008.

[14] M. Dhamala, G. Rangarajan, and M. Ding, "Estimating Granger causality from Fourier and wavelet transforms of time series data," *Phys. Rev. Lett.*, vol. 100, pp. 018701–018701, 2008.

[15] M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biol. Cybern.*, vol. 85, pp. 145–157, 2001.

[16] S. Guo, A. K. Seth, K. M. Kendrick, C. Zhou, and J. Feng, "Partial Granger causality-eliminating exogenous inputs and latent variables," *J. Neurosci. Meth.*, vol. 172, pp. 79–93, 2008.

[17] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fMRI signal," *Nature*, vol. 412, pp. 150–157, 2001.

[18] N. K. Logothetis and B. A. Wandell, "Interpreting the bold signal," *Annu. Rev. Physiol.*, vol. 66, pp. 735–769, 2004.

[19] L. A. Johnston, E. Duff, I. Mareels, and G. F. Egan, "Nonlinear estimation of the bold signal," *NeuroImage*, vol. 40, pp. 504–514, 2008.

[20] R. B. Buxton, K. Uludag, D. J. Dubowitz, and T. T. Liu, "Modeling the hemodynamic response to brain activation," *NeuroImage*, vol. 23, pp. S220–S233, 2004.

[21] P. J. Lahaye, J. B. Poline, G. Flandin, S. Dodel, and L. Garnero, "Functional connectivity: Studying nonlinear, delayed interactions between bold signals," *NeuroImage*, vol. 20, pp. 962–974, 2003.

[22] K. J. Friston, "Causal modelling and brain connectivity in functional magnetic resonance imaging," *PLoS Biol.*, vol. 7, pp. e1000033–e1000033, 2009.

[23] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear Granger causality," *Phys. Rev. Lett.*, vol. 100, pp. 144103–144103, 2008.

[24] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel-Granger causality and the analysis of dynamical networks," *Phys. Rev. E*, vol. 77, pp. 056215–056215, 2008.

[25] A. D. R. McQuarrie and C. L. Tsai, *Regression and time series model selection*. Singapore: World Scientific.

[26] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *J. Am. Stat. Assoc.*, vol. 77, pp. 304–313, 1982.

[27] K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Phys. Rep.*, vol. 441, pp. 1–46, 2007.

[28] J. H. Zar, *Biostatistical Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[29] C. R. Genovese, N. A. Lazar, and T. Nichols, "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *NeuroImage*, vol. 15, pp. 1273–1302, 2002.

[30] J. Geweke, "Measures of conditional linear dependence and feedback between time series," *J. Am. Stat. Assoc.*, vol. 79, pp. 907–915, 1984.

[31] L. Baccala and K. Sameshima, "Partial directed coherence: A new conception neural structure determination," *Biol. Cybern.*, vol. 84, pp. 463–474, 2001.

[32] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman Hall, 1993.

[33] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle, "The human brain is intrinsically organized into dynamic, anticorrelated functional networks," in *Proc. Nat. Acad. Sci.*, 2005, vol. 102, pp. 9673–9678.

[34] E. Gerardin, A. Sirigu, S. Lehericy, J. Poline, B. Gaymard, C. Marsault, Y. Agid, and D. L. Bihan, "Partially overlapping neural networks for real and imagined hand movements," *Cereb. Cortex*, vol. 10, pp. 1093–1104, 2000.

[35] T. Hanakawa, I. Immisch, K. Toma, M. A. Dimyan, P. V. Gelderen, and M. Hallett, "Functional properties of brain areas associated with motor execution and imagery," *J. Neurophysiol.*, vol. 83, pp. 989–1002, 2003.

[36] G. Rizzolatti, G. Luppino, and M. Matteli, "The organization of the cortical motor system: New concepts," *Electroencephalogr. Clin. Neurophysiol.*, vol. 106, pp. 283–296, 1998.

[37] A. Solodkin, P. Hlustik, E. E. Chen, and S. L. Small, "Fine modulation in network activation during motor execution and motor imagery," *Cereb. Cortex*, vol. 14, pp. 1246–1255, 2004.

[38] B. P. Rogers, J. D. Carew, and M. E. Meyerand, "Hemispheric asymmetry in supplementary motor area connectivity during unilateral finger movements," *NeuroImage*, vol. 22, pp. 855–859, 2004.

[39] R. Matsumoto, D. R. Nair, E. LaPresto, W. Bingaman, H. Shibasaki, and H. O. Lüders, "Functional connectivity in human cortical motor system: A cortico-cortical evoked potential study," *Brain*, vol. 130, pp. 181–197, 2007.

[40] A. Floyer-Lea and P. Matthews, "Changing brain networks for visuomotor control with increased movement automaticity," *J. Neurophysiol.*, vol. 92, pp. 2405–2412, 2004.

[41] W. Klonowski, "From conformons to human brains: An informal overview of nonlinear dynamics and its applications in biomedicine," *Nonlinear Biomed. Phys.*, vol. 1, no. 1, pp. 5–5, 2007.

[42] H. Chen, Q. Yang, W. Liao, and Q. GongS. Shen, "Evaluation of the effective connectivity of supplementary motor areas during motor imagery using Granger causality mapping," *Neuroimage*, vol. 47, pp. 1844–1853, 2009.

[43] G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger, "Linear systems analysis of functional magnetic resonance imaging in human V1," *J. Neurosci.*, vol. 16, pp. 4207–4221, 1996.

[44] G. M. Boynton, J. B. Demb, G. H. Glover, and D. J. Heeger, "Neuronal basis of contrast discrimination," *Vision. Res.*, vol. 39, pp. 257–269, 1999.

[45] K. J. Friston and C. Buchel, "Attentional modulation of effective connectivity from V2 to V5/MT in humans," in *Proc. Nat. Acad. Sci.*, 2000, vol. 97, pp. 7591–7596.

[46] M. Ding, Y. Chen, and S. Bressler, "Granger causality: Basic theory and application to neuroscience," in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, B. Schelter, M. Winterhalder, and J. Timmer, Eds. Berlin, Germany: Wiley-VCH, 2006, pp. 437–459.