Time-Varying Nonlinear Causality Detection Using Regularized Orthogonal Least Squares and Multi-Wavelets With Applications to EEG

YANG LI[1,2], MENG-YING LEI[1], YUZHU GUO[1], ZHONGYI HU[3], (Senior Member, IEEE), AND HUA-LIANG WEI[4]

[1]Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing 100191, China
[2]Open Fund Project of Fujian Provincial Key Laboratory of IPIC, Minjiang University, Fujian 350108, China
[3]Intelligent Information Systems Institute, Wenzhou University, Wenzhou 325035, China
[4]Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield S1 3DJ, U.K.

Corresponding authors: Meng-Ying Lei (lmylei@buaa.edu.cn) and Hua-Liang Wei (w.hualiang@sheffield.ac.uk)

ABSTRACT A new transient Granger causality detection method is proposed based on a time-varying parametric modeling framework, and is applied to the real EEG signals to reveal the causal information flow during motor imagery (MI) tasks. The time-varying parametric modeling approach employs a nonlinear autoregressive with external input model, whose parameters are approximated by a set of multi-wavelet basis functions. A regularized orthogonal least squares (ROLS) algorithm is then used to produce a parsimonious or sparse regression model and estimate the associated model parameters. The time-varying Granger causality between nonstationary signals can be detected accurately by making use of both the good approximation properties of multi-wavelets and the good generalization performance of the ROLS in the presence of high-level noise. Two simulation examples are presented to demonstrate the effectiveness of the proposed method for both linear and nonlinear causal detection respectively. The proposed method is then applied to real EEG signals of MI tasks. It follows that transient causal information flow over the time course between various sensorimotor related channels can be successfully revealed during the whole reaction processes. Experimental results from these case studies confirm the applicability of the proposed scheme and show its utility for the understanding of the associated neural mechanism and the potential significance for developing MI tasks based brain-computer interface systems.

INDEX TERMS Granger causality, nonlinear time-varying systems, parametric estimation, multi-wavelets, regularized orthogonal least squares (ROLS), EEG.

I. INTRODUCTION

The investigation of connectivities and dynamics of neuronal assemblies during various brain states plays a key role in understanding the underlying neurophysiological mechanisms of human brain. Recently, there is increasing interest in exploring the influence that one part of the nervous system exerts over another. One of the classical methods for extracting such influence is to determine undirected connectivity, including correlation, synchrony [1], phase coherence [2], and mutual information [3]. However, identifying the directionality of the neural interaction is essential for understanding brain behaviors. A powerful approach to describe directed causal relations of brain regions is Granger causality (GC) [4], [5], which has been proved to be useful for detecting the induced neurophysiological variations in the brain. GC has been widely used to assess causal connectivity for various brain data types such as spike trains [6], local field potentials (LFPs) [7], functional magnetic resonance imaging (fMRI) [8], electroencephalography (EEG) and event-related potentials (ERPs) [9]. Among these various neuroscience data, the high time resolution of EEG signals makes GC be applicable to provide informative causal relations, as the technique is largely dependent on calculating the correspondence of neural signals over time [9].

In conventional GC analysis, the observed time series are fitted by using time-invariant autoregressive with external input (TIVARX) models, where it is assumed that the underlying stochastic process is stationary [10]. However, due to the inherent nonstationarity of biomedical signals, the traditional GC analysis approach may not be sufficiently efficient to reveal the potential nonstationary properties in nature [11]. One popular approach to measure time-varying causality is to introduce sliding windows, which makes nonstationary signals be locally stationary. For example, Ding *et al.* [12] applied short-time windows and autoregressive with external input (ARX) model to study pairwise coherence and thereby revealed task-relevant patterns of cortical interdependence during the different cognitive task stages. The main limitation of such approaches is that the performance heavily depends on the choice of the window length in that the use of too wide windows can lead to loss of the temporal resolution, this is undesirable to real applications where the causal relationship changes rapidly over time, whereas using too narrow windows reduces the statistical reliability. A more generally applicable method may be the ARX based adaptive algorithm for time-varying Granger causality (TVGC) analysis [13], where the assumption that the signals are stationary can be removed. Nevertheless, the slow convergence speed of the recursive least-squares (RLS) may fail to track abrupt variations of the time-varying model parameters, and thus result in delay or inaccurate estimation in the causality analysis.

These time-varying linear Granger causality (TVLGC) analysis methods mentioned above can only approximate the linear causal influence between signals. Additionally, the fundamental parametric models and causal relationships of signals may be studied by stochastic nonlinear multiple time-delay systems [14]–[17]. Given that electrophysiological signals are nonlinear [18], it is essential to interpret the causal relations using nonlinear analysis methods. Recently, several nonlinear Granger causality (NGC) methods have been proposed and applied to neurophysiological signal analysis. For example, Gourévitch *et al.* [19] discussed the measures of both linear and nonlinear Granger causality and their neurophysiological applications. Their results showed that LGC sometimes produces false causal relations, whereas the performance of NGC extremely depends on the choice of model parameters. Li *et al.* [20] have presented a time-varying nonlinear autoregressive with external inputs (TVNARX) modelling framework for GC analysis in EEGs, and the classical RLS algorithm was used to estimate time-varying parameters. Their results indicated that the transient potential causality interactions can be detected from the epileptic EEG signals. However, the main deficiency of their method is that it may not be able to effectively detect rapid changing causalities due to the limitation of the slow convergence of the conventional adaptive methods. In [21], a linear and nonlinear causality detection method based on an orthogonal least squares (OLS) and TVNARX models (OLS-TVNARX) was proposed. The advantage of the OLS-TVNARX method is that time-varying causalities between signals can be detected without

constructing a complete full model. Similar to the limitation of the short sliding windowing ARX modelling approach [12], the OLS-TVNARX method chose to use some fixed window length (i.e. window length = 300) for the GC analysis of both the simulation and real EEG data, and did not suggest a good choice of window size. Clearly, the analysis performance depends on the choice of the window length. Furthermore, although the classical OLS algorithm has been proved to be an efficient method for determining parsimonious model structures [22]–[24], its performance may be affected in cases where signals are highly interrupted by noise [25].

In this paper, a new TVNARX-based parametric modelling method is proposed to detecting time-varying nonlinear Granger causality (TVNGC), where the fundamental TVNARX models are identified by employing multi-wavelet basis functions together with a robust regularized orthogonal forward regression algorithm. The proposed framework mainly includes three steps. Firstly, time-varying parameters in the TVNARX models are approximated by using a finite number of B-spline basis functions, which have excellent approximation performance for tracking both the overall global trend and transient local changes in nonstationary signals, simutaneously [26], [27]. Secondly, a sparse model structure and associated expansion model parameters are determined by a powerful regularized orthogonal least squares (ROLS) algorithm [28]–[30], which has been proven to be capable of constructing an effective parsimonious model that outperforms the traditional OLS method with improved generalization properties. The ROLS algorithm used is more useful than the conventional OLS algorithm even in the presence of severe noise since it not only uses the parsimonious principle of the OLS, but also combines the zero-order regularization criterion which the redundant model terms confused by the conventional OLS algorithm due to noise become less significant under the regularized cost function and can therefore be removed from the expansion model [28]. Time-varying nonlinear autoregressive (TVNAR) models of both univariate and bivariate systems can be exactly identified by the proposed method. Finally, in order to accurately measure time-varying transient causal interactions during the evolution of time-varying processes, a recursive computation is used to obtain the time-varying variances of the prediction errors in the sparse TV nonlinear models, and the time-varying NGC (TVNGC) can thus be calculated by the definition of the Granger causality.

The performance of the proposed approach is illustrated by using two simulation examples, and the simulation results are compared with the state-of-the-art methods including the classical RLS [13], short-windowing ARX [12] and OLS-TVNARX methods. Different performance evaluation criteria are used to measure the efficiency of the causality results, and 1~10 fold cross validation method is also applied to further verify the performance of the proposed framework. The experiment results indicate the proposed GC test scheme is more accurate and robust for detecting connectivity

IEEE Access·

Y. Li *et al.*: Time-Varying Nonlinear Causality Detection Using Regularized Orthogonal Least Squares and Multi-Wavelets

patterns in both linear and nonlinear cases even when the data are highly contaminated by noise. Furthermore, the proposed TVNGC scheme is applied to real EEG data during motor imagery (MI) tasks, where significant nonlinear dynamics and causal connectivities between signals relating to specific MI tasks have been successfully detected. The precise reaction time periods and accurate interaction strengths between different brain regions can be clearly measured, which shows the promising method for deciphering directed connectivity in EEG signals and further exploring cognitive mechanism and developing MI brain-computer interface (BCI) systems.

An obvious advantage of the proposed method is that the combination of multi-wavelet-based basis function expansion with the ROLS algorithm is applied to produce the essential sparse time-varying models with good generalization properties for the inherently nonstationary systems, even if rapidly and even sharply time-varying processes can be still tracked effectively, and transient causal information between nonstationary signals can thus be detected accurately by using the time-varying variances of the estimation errors without the assumption of stationarity and linear dependency imposed on signals. Additionally, the proposed method can capture the time-varying causalities well even when the systems are contaminated with severe noise, which is more suitable for directed interaction detection between real EEG signals. One of the main contributions in this paper is that, for the first time, the newly proposed time-varying system identification scheme is introduced to the detection of transient causal influences between time-varying systems. It is promising that the novel combination may be capable of inspiring the development of more powerful algorithms for time-varying causality detection. Furthermore, with the application to real MI EEG signals, the clearly causal flows indicated can be applied for understanding MI of related neurophysiological mechanism and further evaluating the performance improvement of MI based BCI systems.

The remainder of this paper is organized as follows. In Section II, the methodology is illustrated in three subsections: the explanation of time-varying Granger causality in II-A, the TVNARX model identification method based on multi-wavelet expansion in II-B, and the ROLS algorithm in II-C, respectively. Two simulation examples are given to show the effectiveness of the proposed method in Section III. A case study for the causality detection from real MI EEG signals is introduced in Section IV. Finally, the work is summarized in Section V. Table 1 gives a summary of description for the abbreviations used in this paper.

II. METHODOLOGY
A. GRANGER CAUSALITY
Let $X = \{x(t)\}$ and $Y = \{y(t)\}$ be two signals, with $t = 1, 2 \cdots, N$. According to the general definition of GC, if the variance of the prediction error is decreased by the inclusion of the past information of signal X for the prediction of Y, it is

TABLE 1. Description of the abbreviations used in this paper.

Abbreviation	Description
GC	Granger causality
LFP	local field potential
fMRI	functional magnetic resonance imaging
EEG	electroencephalography
ERP	event-related potential
ARX	autoregressive with external input
TIVARX	time-invariant autoregressive with external input
TVAR	time-varying autoregressive
TVARX	time-varying autoregressive with external input
TVNAR	time-varying nonlinear autoregressive
TVNARX	time-varying nonlinear autoregressive with external inputs
NARMAX	nonlinear autoregressive moving average with exogenous variable
TVGC	time-varying Granger causality
TVLGC	time-varying linear Granger causality
NGC	nonlinear Granger causality
TVNGC	time-varying nonlinear Granger causality
RLS	recursive least-squares
OLS	orthogonal least-squares
ROLS	regularized orthogonal least squares
OMP	Orthogonal Matching Pursuit
AIC	Akaike information criterion
MAE	mean absolute error
RMSE	root mean squared error
MI	motor imagery
BCI	brain-computer interface

said that X causes Y in the Granger sense. The time-invariant Granger causality from X to Y ($TIVGC_{X \to Y}$) is defined by the log ratio of the error variances from the time-invariant AR and ARX models [20], [21]:

$$TIVGC_{X \to Y} = \ln \frac{var\left(y|y^-\right)}{var\left(y|y^-, x^-\right)}$$

$$= \ln \frac{\frac{1}{N-n_y}\sum\limits_{t=1}^{N} e_1^2(t)}{\frac{1}{N-n_y-n_x}\sum\limits_{t=1}^{N} e_2^2(t)} \quad (1)$$

where y^-, x^- denote the past information of Y, X respectively, n_y and n_x are the model orders of Y and X which denote the maximum number of the associated lagged observations. Besides, $e_1(t)$ and $e_2(t)$ are the model residuals of the time-invariant univariate AR model $TIVAR(n_y)$: $y(t) = \sum\limits_{i=1}^{n_y} B_{1i}y(t-i) + e_1(t)$ and bivariate ARX model $TIVARX(n_y, n_x)$: $y(t) = \sum\limits_{i=1}^{n_x} B_{2i}y(t-i) + \sum\limits_{j_y=1}^{n_y} D_{2j}x\left(t-j_y\right) + e_2(t)$, where the former depends only on the past of Y and the latter depends on the past of both Y and X. Eq. (1) implies that Granger causality can never be negative if X causes Y, vice versa for the causality from Y to X. To evaluate the transient directed interactions between nonstationary systems, the definition of TVARX (time-varying ARX) -based TVGC is necessary [20].

Y. Li *et al.*: Time-Varying Nonlinear Causality Detection Using Regularized Orthogonal Least Squares and Multi-Wavelets

IEEE *Access*

1) TIME-VARYING LINEAR GRANGER CAUSALITY

The most commonly used models in time-varying causality test are the linear time-varying autoregressive (TVAR) and TVARX models as follows [31]:

$$x(t) = \sum_{i=1}^{n_x} a_{1i}(t) x(t-i) + \varepsilon_1(t) \quad (2)$$

$$y(t) = \sum_{i=1}^{n_y} b_{1i}(t) y(t-i) + \varepsilon_2(t) \quad (3)$$

and

$$x(t) = \sum_{i=1}^{n_x} a_{2i}(t) x(t-i) + \sum_{j_y=1}^{n_y} c_{2j_y}(t) y(t-j_y) + \varepsilon_3(t) \quad (4)$$

$$y(t) = \sum_{i=1}^{n_x} b_{2i}(t) y(t-i) + \sum_{j_y=1}^{n_y} d_{2j_y}(t) x(t-j_y) + \varepsilon_4(t) \quad (5)$$

Generally, the recursive variance computational formula is given by [20]:

$$\sigma^2(t+1) = (1-c)\sigma^2(t) + c\Delta^2(t) \quad (6)$$

where $\Delta(t)$ is the time-varying model prediction error, and $0 < c < 1$ is the recursive parameter. Set $\Delta(t)$ as the prediction error $\varepsilon_1(t)$, $\varepsilon_2(t)$, $\varepsilon_3(t)$, $\varepsilon_4(t)$ in linear models (2)-(5) respectively, then time-variant variances of the associated errors can be yielded by: $var(x|x_l^-)(t)$, $var(y|y_l^-)(t)$, $var(x|x_l^-, y_l^-)(t)$, $var(y|y_l^-, x_l^-)(t)$, where y_l^-, x_l^- indicate the set of linear terms from y^- and x^-. Consequently, the calculation of the time-varying linear Granger causalities (TVLGC) are represented by:

$$LGC_{X \to Y}(t) = \ln \frac{var(y|y_l^-)(t)}{var(y|y_l^-, x_l^-)(t)} \quad (7)$$

$$LGC_{Y \to X}(t) = \ln \frac{var(x|x_l^-)(t)}{var(x|x_l^-, y_l^-)(t)} \quad (8)$$

2) TIME-VARYING NONLINEAR GRANGER CAUSALITY

The absence of nonlinear terms in the linear TVAR and TVARX models (2)-(5) makes the models insufficient to detect nonlinear causal influences between nonstationary signals. However, there is wide evidence that the evolution of neurophysiological states is a nonlinear process [18]. Nonlinear autoregressive moving average with exogenous variable (NARMAX) models have been demonstrated to be an effective approach that can well capture nonlinear effects for various nonlinear, continuous-time and discrete-time systems [26]. The univariate time-varying nonlinear autoregressive (TVNAR) and multivariate TVNARX models are therefore appropriate for detecting nonlinear TVGC. The TVNAR(n_y) and TVNARX(n_y, n_x) model are formulated by [32] and [33]:

$$y(t) = f\left(y(t-1), \cdots, y(t-n_y)\right) + e_y(t) \quad (9)$$

$$y(t) = f\left(y(t-1), \cdots, y(t-n_y),\right.$$
$$\left. x(t-1), \cdots, x(t-n_x)\right) + e_{xy}(t) \quad (10)$$

where f is the unknown nonlinear function, the observation noise $e_y(t)$ and $e_{xy}(t)$ are assumed to be an independent zero mean noise sequence. The generally used method to approximate the unknown function $f(\cdot)$ is to employ a polynomial expression [34], and (10) can thus be represented as:

$$y(t) = \sum_{n=1}^{M} \sum_{p=0}^{n} \sum_{r_1, \cdots, r_{p+q}=1}^{R} g_{p,q}(r_1, \cdots, r_{p+q}, t)$$
$$\times \prod_{i=1}^{p} y(t-r_i) \prod_{i=p+1}^{p+q} x(t-r_i) + e_{xy}(t) \quad (11)$$

where M is the degree of the nonlinearity, with $p+q = n$, $r_i = 1, 2, \cdots, R$, $\Sigma_{r_1, \cdots, r_{p+q}=1}^{R} \equiv \Sigma_{r_1=1}^{R} \cdots \Sigma_{r_{p+q}=1}^{R}$. The vector $[g_{0,1}(1, t), \cdots, g_{0,1}(R, t), g_{1,0}(1, t), \cdots, g_{p,q}(R, \cdots, R, t)]^T$ are time-varying parameters to be estimated, where the upper index 'T' indicates the transpose of a vector or matrix. Then the time-varying nonlinear Granger causality (TVNGC) from X to Y can be expressed by [31]:

$$NGC_{X \to Y}(t) = \ln \frac{var(y|y_l^-, y_n^-)(t)}{var(y|y_l^-, y_n^-, x_l^-, x_n^-, (yx)_n^-)(t)} \quad (12)$$

where y_n^- and x_n^- are the set of nonlinear terms from past information of Y and X respectively, $(yx)_n^-$ denotes the set of nonlinear terms coupled by past information of Y and X. In addition, $var(y|y_l^-, y_n^-)(t)$ and $var(y|y_l^-, y_n^-, x_l^-, x_n^-, (yx)_n^-)(t)$ are time-varying variances of the prediction errors in TVNAR and TVNARX models by Eq. (6), where $\Delta(t)$ is the associated model estimation error, such as $e_y(t)$ in (9) for the TVNAR and $e_{xy}(t)$ in (10) for the TVNARX model, and the Y to X case is similar. Additionally, it is necessary to assess the statistical significance of the obtained causal influences between two signals. The thresholds for statistical significance are constructed from surrogate data via a permutation procedure under a null hypothesis of no interaction at the significance level $p < 10^{-6}$. This procedure contains generating 1000 permuted time series, where the permutation of the trial order can disrupt task-related interdependencies.

B. MULTI-WAVELET-BASED TVNARX MODEL IDENTIFICATION

According to wavelet theory, a square integrable scalar signal f can be approximated by the multiresolution wavelet decomposition as follows [35], [36]:

$$f(x) = \sum_{k=-\infty}^{\infty} \alpha_{j_0,k} \phi_{j_0,k}(x) + \sum_{j>j_0-1}^{\infty} \sum_{k=-\infty}^{\infty} \beta_{j,k} \psi_{j,k}(x) \quad (13)$$

where $\phi_{j_0,k}(x) = 2^{j_0/2}\phi(2^{j_0}x - k)$ and $\psi_{j,k} = 2^{j/2}\psi(2^j x - k)$ with $j_0, j, k \in Z$ (Z denotes the whole integers) are the translated and dilated version of the scaling function $\phi(x)$ and the mother wavelet $\psi(x)$, $\alpha_{j_0,k}$ and

$\beta_{j,k}$ are the wavelet decomposition coefficients. Furthermore, when the resolution scale level of the scale functions $\phi_{j_0,k}(x) = 2^{j_0/2}\phi\left(2^{j_0}x - k\right)$ is sufficiently large, namely, there exists an integer J, Eq. (13) can be reduced to $f(x) = \sum_{k=-\infty}^{\infty} \alpha_{J,k}\phi_{J,k}(x)$.

Previous studies [37], [38] show that the time-varying parameters in (11) can be well approximated using a set of multi-wavelet basis functions $\left\{\pi_\mu(t) : \mu = 1, 2, \cdots, L\right\}$. Specifically, the time-varying model (11) can be re-written by:

$$y(t) = \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{r_1,r_{p+q}=1}^{R}\sum_{\mu=1}^{L}\lambda_{p,q,\mu}(r_1,\cdots,r_{p+q})$$

$$\times\left(\pi_\mu(t)\prod_{i=1}^{p}y(t-r_i)\prod_{i=p+1}^{p+q}x(t-r_i)\right) + e_{xy}(t)$$

$$= \varphi^T(t)\theta + e_{xy}(t) \qquad (14)$$

where

$$\begin{cases}\varphi(t) = [\,\pi_\mu(t)\,\prod_{i=1}^{p}y(t-r_i)\,\prod_{i=p+1}^{p+q}x(t-r_i)]^T,\\ \theta = [\lambda_{0,1,\mu}(1),\cdots,\lambda_{0,1,\mu}(R),\lambda_{1,0,\mu}(1),\cdots,\\ \lambda_{p,q,\mu}(R,\cdots,R)]^T\end{cases} \qquad (15)$$

$\lambda_{p,q,\mu}(r_1,\cdots,r_{p+q})$ are the multi-wavelet-based expansion parameters, L is the maximum number of basis function sequences, $\varphi(t)$ and θ are the regression vector and parameter vector respectively. The TVNARX model (11) can then be transformed into time invariant model, as $\lambda_{p,q,\mu}(r_1,\cdots,r_{p+q})$ are now time-invariant.

Cardinal B-splines are an important class of basis functions that can form multiresolution wavelet decompositions [39]. They are compactly supported and can be analytically formulated in an explicit form, this unique property makes the operation of the multiresolution decomposition (13) much more convenient. From the recursive definition of cardinal B-spline functions [40]:

$$B_m(x) = \frac{x}{m-1}B_{m-1}(x) + \frac{m-x}{m-1}B_{m-1}(x-1), \quad m \geq 2 \qquad (16)$$

where $B_1(x) = 1$ with $x \in [0,1)$, the *mth* order B-spline $B_m(x)$ is defined on $[0,m]$. Taking the cardinal B-splines as the basis function, the $\phi_{j,k}(x)$ can be expressed as $\phi_{j,k}(x) = 2^{j/2}B_m\left(2^j x - k\right)$, where the dilation and translation indices j and k should satisfy $0 \leq 2^j x - k \leq m$. Assume that the function $f(x)$ to be estimated with decompositions (13) is defined within $[0,1]$, for any given dilation index j, the translation index k are restricted to the collection $\Gamma_m = \left\{k : -m \leq k \leq 2^j - 1\right\}$. Generally a practical selection of the order m are 2, 3, 4, 5, and the detail discussion of B-splines properties can be found in [41] and [42]. Additionally, $j \geq 3$ is an appropriate choice for most applications using cardinal

B-splines. The higher the value of j, the more basis functions are used and thus the resolution improves, but this will also introduce more parameters and increase the computational cost. Some determination criteria of the proper j are discussed in detail in [25].

Hence time-varying coefficients $g_{p,q}(r_1,\cdots,r_{p+q},t)$ in Eq. (11) can be estimated by using a combination of B-splines basis functions from the families $\phi_k^m(x) = 2^{j/2}B_m\left(2^j x - k\right)$ with $k \in \Gamma_m, m = 2 \sim 5$, which can be expressed as follows:

$$g_{p,q}(r_1, r_2\cdots,r_{p+q},t)$$

$$= \sum_m\sum_{k\in\Gamma_m}\lambda_{p,q,k}^m(r_1, r_2\cdots,r_{p+q})\phi_k^m\left(\frac{t}{N}\right) \qquad (17)$$

where N is the number of sample observations for $t = 1, 2, \cdots, N$. The decomposition (17) can easily be transformed into the form of (14), where the union of the families $\left\{\sum_m\phi_k^m(t) : k \in \Gamma_m, m = 2 \sim 5\right\}$ replace the set $\left\{\pi_\mu(t) : \mu = 1, 2, \cdots, L\right\}$.

Although the regression terms are usually sparsely distributed in the associated space, the number of candidate regression terms in the initial full regression Eq. (14) may be very large, which makes the problem be ill-posed. In addition, with the parsimonious principle, the ill-posed problem can be avoided and the model constructed can be achieved the generalization performance. Recently, the OLS algorithm has been proved to be an efficient method for constructing parsimonious model structures. However, the parsimonious structures alone may not be sufficient to eliminate overfitting and guarantee good generalization performance if modeling data are highly interrupted by noise [30]. Therefore, it is a crucial procedure to construct a parsimonious or sparse model structure with good generalization performance and approximate the associated parameters in basis function expansion based time-varying system identification. The ROLS algorithm used in this study to solve these problems will be introduced in the following section.

C. REGULARIZED ORTHOGONAL FORWARD REGRESSION

The orthogonal least squares (OLS) type of algorithms have proven very efficient to deal with model term selection problems [22]–[24], [43]. However, the error criterion used in the OLS algorithm is the total squared error, which may lead to overfitting especially when observable data are highly noisy [23], [44]. To solve this issue, the ROLS algorithm [28]–[30] based on the zero-order regularization with the OLS algorithm is employed to construct a more generalized procedure for constructing sparse model structure. Collecting (14) for $t = 1, 2, \cdots, N$ together can get the associated compact matrix form $Y = \Phi\Theta + E$, where $Y = [y(1), y(2),\ldots,y(N)]^T$, $\Phi = [\varphi(1), \varphi(2),\ldots,\varphi(N)]^T$ is a $N \times H$ dimensional regression matrix, H is the number of all the candidate terms, $\Theta = [\chi_1, \chi_2,\ldots,\chi_H]^T$ is the parameter vector to be estimated, and $E = [e_{xy}(1), e_{xy}(2),\ldots,e_{xy}(N)]^T$,

respectively. Assume the regression matrix Φ is of a full column rank and the procedure of the associated orthogonal transformation can be expressed by:

$$\Phi = \underbrace{[w_1 \cdots w_H]}_{W} \underbrace{\begin{bmatrix} 1 & \delta_{1,2} & \cdots & \delta_{1,H} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \delta_{H-1,H} \\ 0 & \cdots & 0 & 1 \end{bmatrix}}_{A} \quad (18)$$

where W is an $N \times H$ matrix with orthogonal columns satisfying $w_{ii}^T w_{jj} = 0$ when $ii \neq jj$, A is an $H \times H$ unit upper triangular matrix, so that the regressor matrix form can be denoted as $Y = (\Phi A^{-1})(A\Theta) + E = WP + E$, where $P = [\rho_1, \rho_2, \ldots, \rho_H]^T$ is an auxiliary parameter vector.

The key procedure to obtain high computational efficiency and the accuracy of the ROLS scheme is the zero-order regularized cost function defined as $F = E^T E + \tau P^T P$ [28], where $\tau \geq 0$ is the regularization parameter, then the unknown parameters ρ_{ii} can be estimated by $\rho_{ii} = \langle Y, w_{ii} \rangle / (\langle w_{ii}, w_{ii} \rangle + \tau)$, $ii = 1, 2, \ldots, H$, where the symbol $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors. The parameter vector Θ can thus be determined by using $A\Theta = P$ with solved A and P, and the regularized error reduction ratio ($rerr_{ii}$) introduced by w_{ii} can be derived as:

$$rerr_{ii} = \frac{(\langle w_{ii}, w_{ii} \rangle + \tau) \rho_i^2}{\langle Y, Y \rangle} = \frac{\langle Y, w_{ii} \rangle^2}{\langle Y, Y \rangle (\langle w_{ii}, w_{ii} \rangle + \tau)} \quad (19)$$

Each selected procedure is chosen to decrease maximally the regularized squared error F, namely the term gives the largest $rerr_{ii}$ at each iteration ii is chosen, and significant regressors can be selected in a forward regression process. Choose an error tolerance $\zeta : 0 < \zeta < 1$, and the selection will be terminated at $N_T th$ step when $1 - \sum_{i=1}^{N_T} rerr_i < \zeta$ is satisfied.

The determination of the optimal value of τ depends on the intrinsic property of the underlying system and the selection of proper basis functions. Previous studies have proven that the estimation performance of the multi-wavelets-based model may be not sensitive to the precise value of τ [28]. An elegant method to determine the regularization parameter is to apply a Bayesian interpretation to the ROLS algorithm which results in the following iterative procedure for calculating τ [28]:

$$\begin{cases} \eta = \sum_{ii=1}^{N_T} \frac{w_{ii}^T w_{ii}}{w_{ii}^T w_{ii} + \tau} \\ \tau = \frac{\eta}{N-\eta} \frac{E^T E}{P^T P} \end{cases} \quad (20)$$

Let τ_0 be an initial value of τ, for example $\tau_0 = 1$, a satisfied τ can be found after a few iterations. This new forward selection algorithm is capable of constructing an accurate parsimonious model structure with improved generalization property and efficient operation performance as the similar computational requirement of the OLS algorithm. Applying the ROLS algorithm to Eq. (14), an optimal subset of significant regressors can be selected and the model parameters can

be estimated effectively. The original time-varying parameters $g_{p,q}(r_1, r_2 \cdots, r_{p+q}, t)$ in the TVNARX model (11) can then be recovered by using those resultant estimates. Take the Gram-Schmidt algorithm as the orthogonalization method, the procedure for the detection of sparse model structure through the ROLS scheme can be implemented in a stepwise manner given below.

Input:
  Output signal $Y = [y(1), y(2), \ldots, y(N)]^T$;
  Candidate terms $\Phi = \{\varphi_h : h = 1, 2, \ldots, H\}$;
  Predesigned threshold $\Xi < 10^{-10}$.
Step 1. Set $\mathfrak{I}_1 = \{1, 2, \ldots, H\}$;
  for $h = 1$ to H
    $w_h = \varphi_h$; $\rho_h = \langle Y, w_h \rangle / (\langle w_h, w_h \rangle + \tau)$;
    $rerr_h = (\langle w_h, w_h \rangle + \tau) \rho_h^2 / \langle Y, Y \rangle$;
  end for
  $\hbar_1 = \underset{h \in \mathfrak{I}_1}{\arg\max} \{rerr_h\}$;
  $w_1^1 = w_{\hbar_1}$; $\rho_1^1 = \langle Y, w_1^1 \rangle / (\langle w_1^1, w_1^1 \rangle + \tau)$;
Step υ. $\upsilon \geq 2$:
  for $\upsilon = 2$ to H
    $\mathfrak{I}_\upsilon = \mathfrak{I}_{\upsilon-1} \setminus \{\hbar_{\upsilon-1}\}$;
    for all $h \in \mathfrak{I}_\upsilon$
      $w_h = \varphi_h - \sum_{\ell=1}^{\upsilon-1} \left(\varphi_h^T w_\ell^1 / w_\ell^{1^T} w_\ell^1 \right) w_\ell^1$;
      $rerr_h = \langle Y, w_h \rangle^2 / (\langle Y, Y \rangle (\langle w_h, w_h \rangle + \tau))$;
    end for
    $\mathfrak{I}_\upsilon = \mathfrak{I}_\upsilon \setminus \left\{ \underset{h \in \mathfrak{I}_\upsilon}{\arg} \left(w_h^T w_h < \Xi \right) \right\}$;
    $\hbar_\upsilon = \underset{h \in \mathfrak{I}_\upsilon}{\arg\max} \{rerr_h\}$;
    $w_\upsilon^1 = w_{\hbar_\upsilon}$; $\rho_\upsilon^1 = \langle Y, w_\upsilon^1 \rangle / (\langle w_\upsilon^1, w_\upsilon^1 \rangle + \tau)$;
  end for
Output:
  Selected model terms $\Phi^1 = \left[\varphi_{\hbar_1}, \varphi_{\hbar_2}, \ldots, \varphi_{\hbar_{N_T}} \right]$.

As to the issue of the model order determination, a possible solution is to minimize a criterion that balances the variance accounted for by the model against the number of coefficients to be estimated. In this work, the correct model order size is determined by the Akaike information criterion (AIC) [45]:

$$AIC(n_y) = \ln\left(\det\left(\Sigma_{n_y} \right) \right) + \frac{2 n_y n_{var}^2}{N} \quad (21)$$

where Σ_{n_y} is the variance of the model residuals calculated from the associated $n_y th$ order model, and n_{var} is the number of the variables.

In conclusion, the new proposed framework for time-varying Granger causality detection can be summarized as follows:

(1) Set up the linear or nonlinear univariate TVAR (such as Eqs. (2)-(3) for linear case and (9) for nonlinear case) and multivariate TVARX (such as Eqs. (4)-(5) for linear and (10) for nonlinear case) models for the nonstationary procedures in the observed input-output systems.

(2) For each model to be identified, expand all the time-varying coefficients in the linear or nonlinear

models by multiple B-spline basis functions, and construct the Eq. (14).

(3) Select significant regressors from the expanded candidate terms by using ROLS algorithm described above.

(4) Determine the number of proper model orders for time-varying models by AIC criterion (21).

(5) Approximate associated coefficients of the selected model terms, and recovery the original time-varying parameters by Eq. (17), thus time-variant prediction errors for each univariate and multivariate time-varying autoregressive models can be obtained.

(6) Calculate time-varying variances of the prediction errors by Eq. (6), and achieve the measure of time-varying Granger causalities based on the GC definition in Eq. (7) and Eq. (12).

III. SIMULATION EXAMPLES

This section presents two simulation examples to illustrate and verify the performance of the proposed TVGC detection method. The proposed method and three other methods (i.e. RLS [13], short-windowing ARX [12] and OLS-TVNARX [21]) are applied to the same simulation data, and the results are compared.

A. TIME-VARYING LINEAR GRANGER CAUSALITY

For the TVLGC test, consider two TVARX (2, 2) models:

$$
\begin{aligned}
x(t) &= a_{2,1}(t)\, x(t-1) + a_{2,2}(t)\, x(t-2) \\
&\quad + c_{2,1}(t)\, y(t-1) + c_{2,2}(t)\, y(t-2) + \xi_1(t) \\
y(t) &= b_{2,1}(t)\, y(t-1) + b_{2,2}(t)\, y(t-2) \\
&\quad + d_{2,1}(t)\, x(t-1) + d_{2,2}(t)\, x(t-2) + \xi_2(t)
\end{aligned}
\tag{22}
$$

where

$$
a_{2,1}(t) = \begin{cases} -0.6, & 1 \le t < 400, \\ 0.3, & 400 \le t \le 1000, \end{cases}
$$

$$
a_{2,2}(t) = 0.1, \quad 1 \le t \le 1000,
$$

$$
b_{2,1}(t) = \begin{cases} 0.3, & 1 \le t < 400, \\ -0.6, & 400 \le t \le 1000, \end{cases}
$$

$$
b_{2,2}(t) = 0.1, \quad 1 \le t \le 1000,
$$

$$
c_{2,1}(t) = \begin{cases} 0, & 1 \le t < 200, \\ 0.6, & 200 \le t \le 380, \\ 0, & 380 < t \le 1000, \end{cases}
$$

$$
c_{2,2}(t) = \begin{cases} 0, & 1 \le t < 200, \\ 0.5, & 200 \le t \le 380, \\ 0, & 380 < t \le 1000, \end{cases}
$$

$$
d_{2,1}(t) = \begin{cases} 0, & 1 \le t < 700, \\ 0.6, & 700 \le t \le 1000, \end{cases}
$$

$$
d_{2,2}(t) = \begin{cases} 0, & 1 \le t < 700, \\ 0.5, & 700 \le t \le 1000. \end{cases}
\tag{23}
$$

and ξ_1, ξ_2 are Gaussian white noise processes with zero means and variances:

$$
var(\xi_1) = \begin{cases} 0.9, & 1 \le t < 600, \\ 2.0, & 600 \le t \le 1000, \end{cases}
$$

$$
var(\xi_2) = \begin{cases} 2.0, & 1 \le t < 600, \\ 0.9, & 600 \le t \le 1000. \end{cases}
\tag{24}
$$

It is known that causality relation between x and y is given as follows: 1) from sample point 200 to sample point 380, the signal y causes the signal x; 2) starting with the sample point 700, the signal x causes the signal y; 3) for the first 199 sample points and the sample points between 381 and 699, there is no causal relation between x and y. A second order TVAR model and a TVARX (2, 2) model for the signal y are estimated to calculate the TVLGC from x to y, and the same procedure is implemented for the case from y to x. The two models defined by (22) are simulated and total of 1000 data points are recorded. In order to verify the effectiveness and robustness of the proposed method, three different levels of white Gaussian noise (WGN), with signal-to-noise ratio (SNR) being 20 dB, 10 dB, and 5 dB, respectively, are added to the original simulation data.

The B-spline basis functions selected from $\{\phi_k^m : m = 3, 4, 5\}$ with the scale index $j = 3$ are employed to approximate the time-varying parameters. The ROLS algorithm is then applied to select significant regressors from a large number of candidate terms and estimate the associated parameters. Furthermore, the TVGCs are calculated by using both (6) and (7). The TVGC results from the proposed method are shown in Fig. 1 under three different noise levels, where the bold font indicates the proposed method. For comparison, the classical RLS algorithm with a forgetting factor $\rho = 0.98$, the ARX-based short sliding windowing algorithm and the conventional OLS-TVARX method, are also used to measure the TVGCs, and the associated results are given in Fig. 1.

For Fig.1, the classical RLS algorithm based approach fails to faithfully track piece-wise changes in the directed dependencies due to the slow convergence of the algorithm even under less severe noise level (SNR = 20dB). For the results of the TVGCs using the sliding window method, this method may be insufficient to guarantee high time resolution and track accuracy simultaneously because its efficiency is heavily dependent on the choice of the sliding window size. The OLS-TVARX method can detect abrupt time-varying causalities while it is susceptible to background noises, and fluctuations and estimation error can obviously be observed in the detection plot especially when the data are severely contaminated by noise, such as the case of SNR = 5 dB. Particularly, the TVLGCs measured using the proposed method, where the expected influence of y on x from sample point 200 to sample point 380 is confirmed by the positive values of $LGC_{y \to x}(t)$ (black solid lines), and the opposite influence of x on y starting at the sample point 700 is identified by the positive values of $LGC_{x \to y}(t)$ (blue solid lines). The values of the GC test for both $LGC_{y \to x}(t)$ and $LGC_{x \to y}(t)$

Y. Li *et al.*: Time-Varying Nonlinear Causality Detection Using Regularized Orthogonal Least Squares and Multi-Wavelets
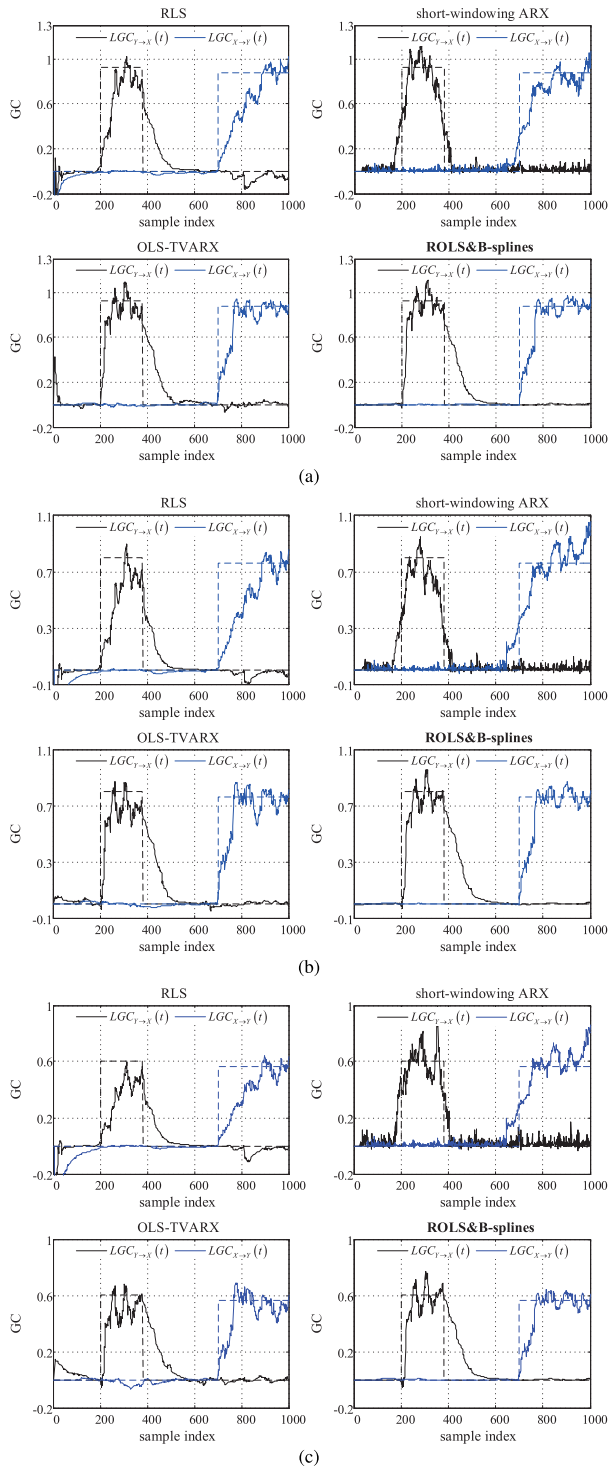
IEEE *Access*

FIGURE 1. TVLGC detection results from simulation example A in (22) using different methods under three noise cases. (a) SNR = 20dB, (b) SNR = 10dB, (c) SNR = 5dB, where TVLGCs $LGC_{Y \to X}(t)$ and $LGC_{X \to Y}(t)$ are black and blue solid lines, and the dashed lines is the associated time-invariant Granger causalities $TIVGC_{Y \to X}$ and $TIVGC_{X \to Y}$, respectively.

are nearly zero within the sample index intervals $1 \leq t < 200$ and $380 < t < 700$, which indicates that there is no causal interaction between two signals during these sample index period. Furthermore, time-varying causalities

change slightly around the estimations of the associated time-invariant Granger causality (TIVGC) (black and blue dashed lines) within the stationary period $200 \leq t \leq 380$ and $700 \leq t \leq 1000$. In comparison with three conventional methods, the proposed ROLS with B-splines approach can better track the variations of the causalities and more accurately capture different patterns of changes in the time-varying causality: the constant value, smooth changes and abrupt changes, even in the presence of different levels of noise.

In order to quantitatively evaluate the effectiveness of the proposed method, the mean absolute error (MAE) and root mean squared error (RMSE) of the TVGC estimates with respect to the associated time-invariant values are calculated for three SNR cases: 20 dB, 10 dB and 5 dB, respectively, and the comparison results are shown in Table 2. It is obvious that the values of MAE and RMSE by the proposed approach are the smallest ones among four methods for three noise cases mentioned above. These results statistically confirm the superiority of the proposed multi-wavelets-based ROLS method for detecting time-varying causality in the presence of noise. The MAE and RMSE in this study are defined as:

$$MAE = \frac{1}{N} \sum_{k=1}^{N} |\hat{G}(k) - G(k)| \qquad (25)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^{N} \left\| \frac{\hat{G}(k) - G(k)}{G(k)} \right\|^2} \qquad (26)$$

where $\hat{G}(k)$ represents the estimates of TVGC $G(k)$, and N is the length of data.

Additionally, the performance of the proposed scheme can be further evaluated by the cross validation with different folds. Specifically, the testing and training data subset are randomly selected from the generated 1000 data points. TVAR and TVARX models can be identified by four compared algorithms on the training data, and the causality measurement results from different approaches can be tested by the testing data subset. The MAE of test results for 1~10 fold are given in Fig.2. Estimation errors in Fig. 2 by the proposed ROLS with multiple B-splines method are smaller than other three causality prediction algorithms for all testing folds, indicating excellent causal detection power of the proposed framework. Particularly, it is worth noting that the superiority of the proposed method is clearer when the noise level increases. These results demonstrate that the proposed approach takes the advantages of the good local approximation performance of B-splines and the excellent generalization property of the ROLS algorithm, and thus enables to track rapid variations in time-varying causalities effectively, especially when data are contaminated by severe noise.

B. TIME-VARYING NONLINEAR GRANGER CAUSALITY
To further test the performance of the proposed approach for nonlinear causality detection, the following TVNARX model

IEEE *Access*

Y. Li *et al.*: Time-Varying Nonlinear Causality Detection Using Regularized Orthogonal Least Squares and Multi-Wavelets

TABLE 2. A performance comparison of the causality test using four different methods with three SNR cases for example1 A.

| Method	Direction of TVLGC	20dB		10dB		5dB	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
RLS	$LGC_{y \to x}(t)\ (200 \le t \le 380)$	0.1098	0.2202	0.0924	0.1853	0.0846	0.1688
	$LGC_{x \to y}(t)\ (700 \le t \le 1000)$	0.1024	0.2098	0.108	0.2183	0.0971	0.2115
Short-windowing ARX	$LGC_{y \to x}(t)\ (200 \le t \le 380)$	0.0932	0.2182	0.0863	0.1984	0.0774	0.1626
	$LGC_{x \to y}(t)\ (700 \le t \le 1000)$	0.0814	0.1873	0.0719	0.1557	0.0530	0.1114
OLS-TVARX	$LGC_{y \to x}(t)\ (200 \le t \le 380)$	0.0883	0.1903	0.0837	0.1797	0.0780	0.1502
	$LGC_{x \to y}(t)\ (700 \le t \le 1000)$	0.0504	0.1417	0.0466	0.1188	0.0430	0.0965
ROLS with B-splines	$LGC_{y \to x}(t)\ (200 \le t \le 380)$	**0.0790**	**0.1871**	**0.0738**	**0.1782**	**0.0617**	**0.1489**
	$LGC_{x \to y}(t)\ (700 \le t \le 1000)$	**0.0465**	**0.1382**	**0.0407**	**0.1159**	**0.0320**	**0.0914**

Note: bold values indicate the best results.
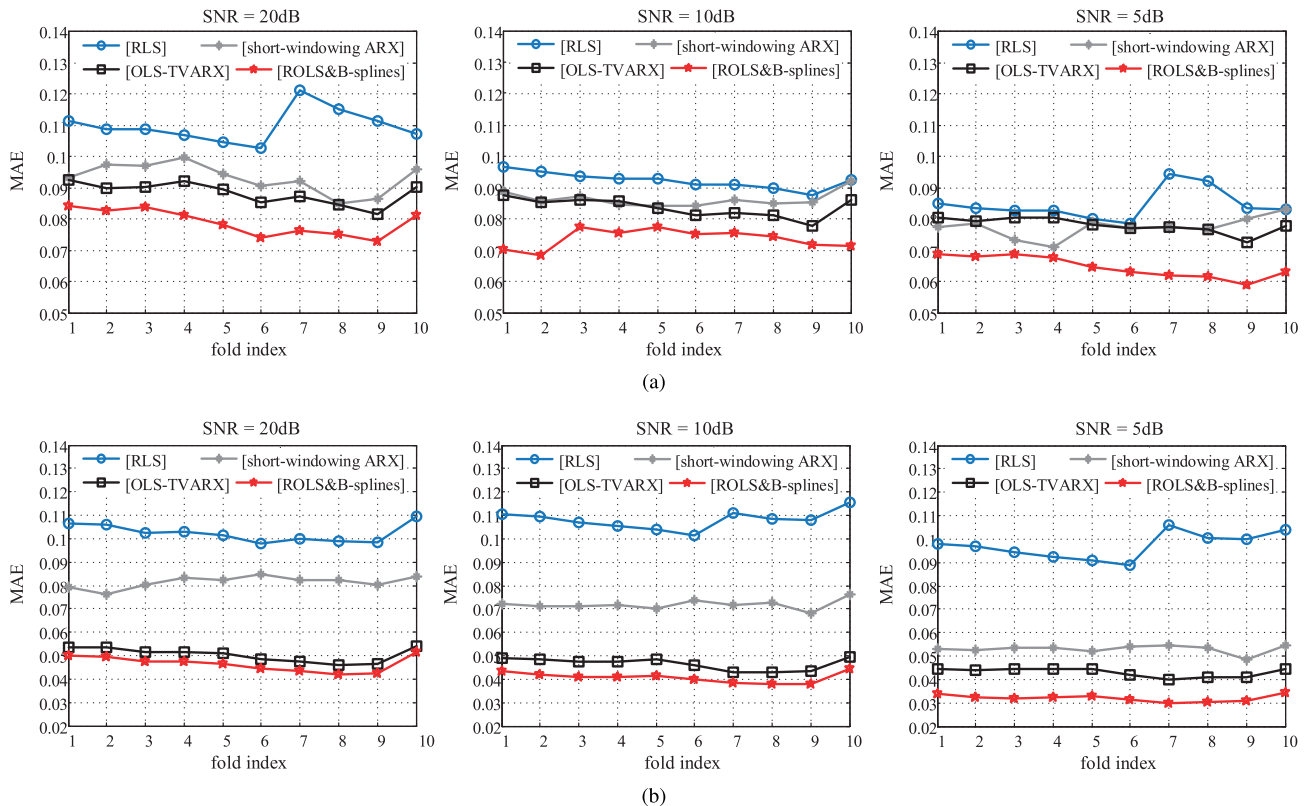
FIGURE 2. MAE of fold index of cross validation using four compared methods under three noise cases for example A. (a) $LGC_{Y \to X}(t)$, (b) $LGC_{X \to Y}(t)$.

is used to generate simulation data:

$$y(t) = h_{1,1}(t)\, y(t-1) + h_{1,2}(t)\, x(t-1)$$
$$+ h_{2,1}(t)\, y^2(t-1) + h_{2,2}(t)\, x^2(t-1) + e(t) \quad (27)$$

where $x(t)$ is a random sequence uniformly distributed in $[-1, 1]$, $e(t)$ is a Gaussian white noise sequence with zero mean and variance 0.04. The time-varying parameters are given below:

$$h_{1,1}(t) = \begin{cases} 0, & 1 \le t \le 400, \\ -0.5, & 400 < t \le 1000, \end{cases}$$

$$h_{1,2}(t) = \begin{cases} 0, & 1 \le t \le 300, \\ -0.8, & 300 < t \le 700, \\ -0.5, & 700 < t \le 1000, \end{cases}$$

$$h_{2,2}(t) = \begin{cases} 0, & 1 \le t \le 300, \\ -0.5, & 300 < t \le 700, \\ 0, & 700 < t \le 1000. \end{cases}$$

$$h_{2,1}(t) = \begin{cases} 0, & 1 \le t \le 400, \\ 0.2, & 400 < t \le 1000, \end{cases} \quad (28)$$

This means that the coefficient $h_{1,2}(t)$ determines the linear causal influence from x to y during the period of

Y. Li *et al.*: Time-Varying Nonlinear Causality Detection Using Regularized Orthogonal Least Squares and Multi-Wavelets

IEEE *Access*

[301, 700], and the coefficient $h_{2,2}(t)$ decides the nonlinear causal influence from x to y during the same period. During the period of [701, 1000], there is no nonlinear causal influence from x to y. As in Example 1, three different levels of WGN, with SNR of 30 dB, 20 dB, and 10 dB, respectively, are added to the original simulation data.

The TVNAR and TVNARX models, with a nonlinear degree $\kappa = 2$, are constructed using the following four time-varying parametric methods: 1) RLS with a forgetting factor $\rho = 0.97$, 2) ARX with highly overlapped short sliding windows, 3) OLS-TVNARX, and 4) the proposed ROLS with multi-wavelets. The B-spline basis functions selected from $\{\phi_k^m : m = 3, 4, 5, 6\}$, with the scale index $j = 4$, are employed to approximate the time-varying parameters. The associated TVNGC results using the above four methods under three noise cases are shown in Fig. 3.

In Fig. 3, it is obvious that the TVNGC detection results by the proposed method outperform the other three methods including the RLS, short-window ARX, OLS-TVNARX, where the bold font indicates the proposed method. Specifically, the RLS method from the first subgraph of panels (a)-(c)) is unable to rapidly detect the abrupt changes in nonlinear GC at the sample indices 300 and 700 under these three noise conditions due to the deficiency of the slow convergence. The second subgraph of panels (a)-(c) show the results measured by short-windowing method, which give lagged and inaccurate detection results for the nonlinear causality in comparison with the proposed approach particularly under severely noise case (SNR = 10 dB). From the third subgraph of panels (a)-(c) calculated by OLS-TVNARX, quite similar result as the proposed scheme is obtained for the case of SNR = 30 dB, while with the level of WGN increasing, the proposed approach performs better than the OLS-TVNARX method especially when SNR = 10dB.

The MAE and RMSE of the estimated TVNGC by the four methods are calculated using the associated time-invariant causal index values as a reference, and the results are shown in Table 3, where the statistic values confirm better tracking ability of the proposed method for both linear and nonlinear causal detection under all three noise conditions. Similar to the previous linear example, the cross validation results with 1~10 fold by the proposed multi-wavelets-based ROLS method and other three compared approaches are presented in Fig. 4. The smaller testing errors obtained have been proved that the proposed scheme can achieve better causality prediction efficiency than other three methods especially with a high level of noise, *i.e.*, SNR = 10 dB. These results in Figs. 3-4 and the statistical comparisons (Table 3) demonstrate that the proposed method can be an effective tool for analyzing GC of nonstationary signals even severely contaminated by noise such as real electrophysiological signals.

IV. APPLICATION TO MOTOR IMAGERY EEG SIGNALS
A. DATASET OVERVIEW
In this section, the proposed GC detection scheme is applied to analyze time-varying directed interactions between motor
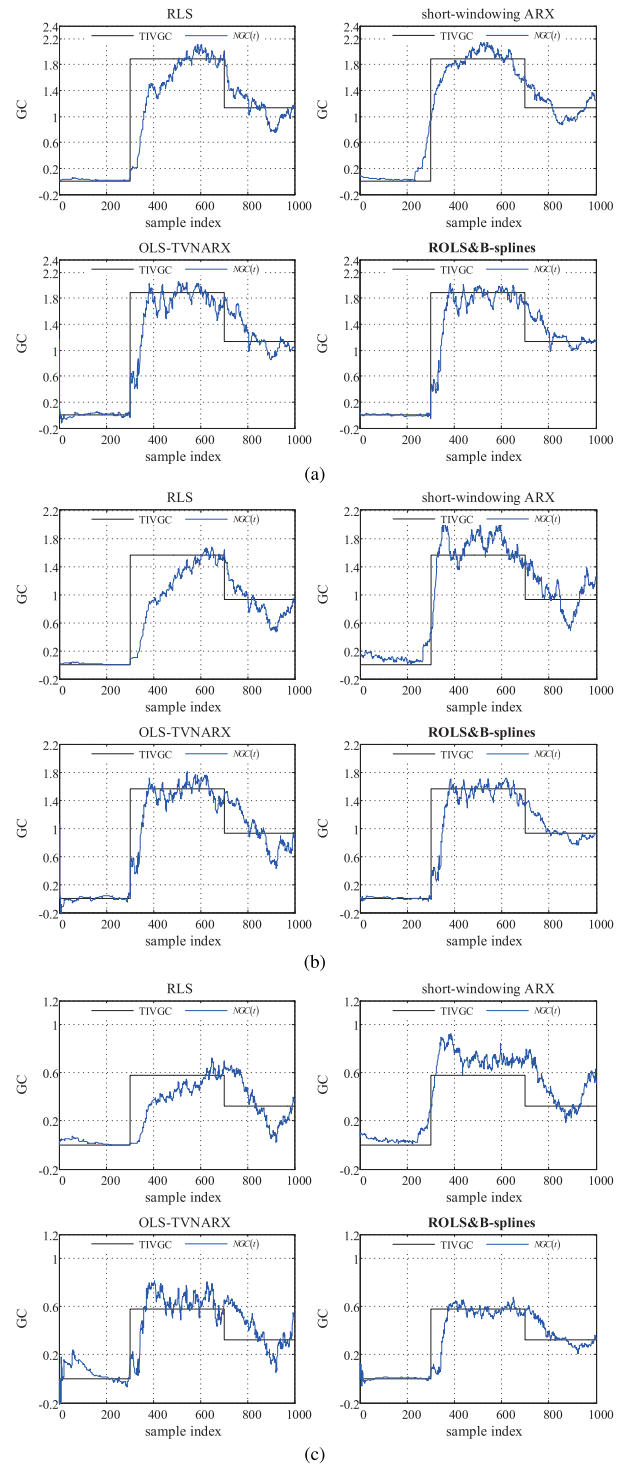

FIGURE 3. TVNGC detection results for model (27) using different methods under three noise cases. (a) SNR = 30dB, (b) SNR = 20dB, (c) SNR = 10dB, where TVNGCs are shown as blue solid lines, and the black lines indicate the associated time-invariant Granger causalities (TIVGCs), respectively.

imagery (MI) EEG signals. Here MI indicates the imagination of a particular motor action without any actual execution of limbs, which is showed promising effectiveness in various research fields including neuroscience and
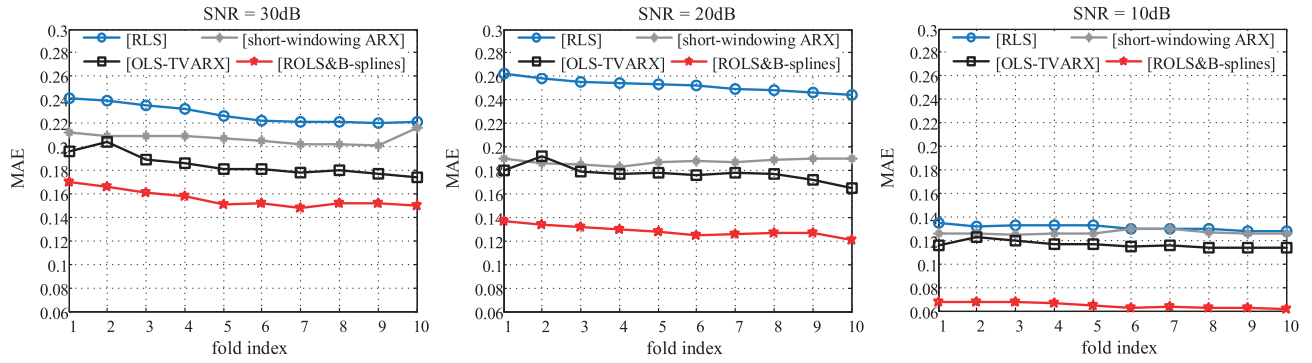
FIGURE 4. MAE of fold index of cross validation for example B using four different methods under three noise cases.

TABLE 3. A performance comparison of the GC test for example B.

Method	30 dB		20 dB		10 dB	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
RLS	0.228	0.4313	0.2524	0.4316	0.1310	0.1915
Short-windowing ARX	0.2073	0.3816	0.1875	0.2973	0.1266	0.1608
OLS-TVNARX	0.1846	0.3425	0.1773	0.2995	0.1164	0.1614
ROLS with B-splines	**0.1550**	**0.3316**	**0.1283**	**0.2785**	**0.0646**	**0.1283**

Note: bold values indicate the best results.

rehabilitation [46], [47]. Specifically, MI is the most commonly used experimental paradigm in brain-computer interface (BCI) system, which has a significant practical importance and provides a potential communication between the human brain and the computer [48]–[50]. Recent investigations based on EEG report the existence of the directional connectivity of motor-related areas during MI tasks [51]–[53]. The EEG dataset used in this study is available publicly from PhysioNet [54], created by the BCI2000 instrumentation system [55]. The EEG signals were recorded from 109 healthy subjects during different MI tasks , consisting of 64-channel data measured by the international 10-10 system [56], sampled at 160 Hz. Specifically, three runs where the subjects imagined movement of left hand and right hand are selected in this study, and totally 21 trials with each duration of 4s are included.

It has been proven that the neural activity related to the hand movement imagery is almost exclusively contained within channels C3, C4, and Cz [57]. Hence C3 and C4 channels are selected as an example for time-varying causality study. Consider that the MI task is performed within the time period 0~4s, EEG epochs of 6s duration, 1s before and 5s after the stimulus are prepared for the GC analysis. In order to mitigate the effect of the nonstationarity embodied in the mean, inter-trial variations, and the ensemble average, the point-by-point is removed from each trial along with dividing by the ensemble standard deviation [12]. The pre-processed average event-related potentials (AERPs) of channels C3 and C4 recorded from one subject during left and right hand MI
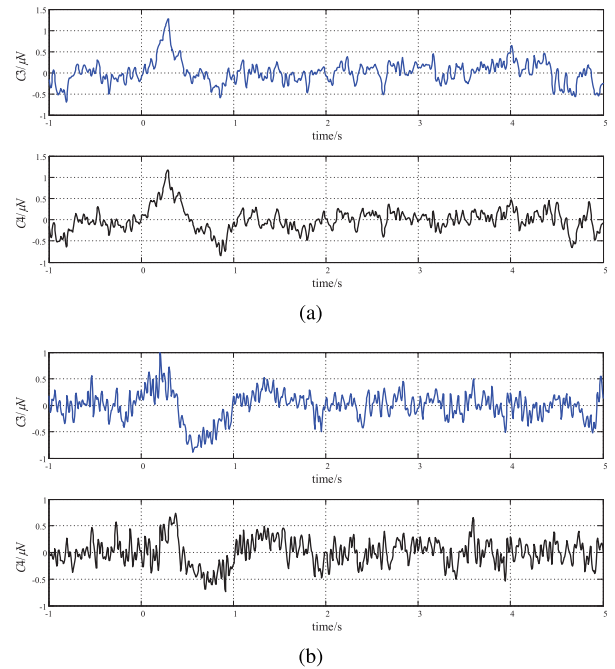

FIGURE 5. The AERPs of C3 and C4 during MI: (a) left hand; (b) right hand.

tasks are displayed in Fig. 5 (a) and (b).

B. TVNGC ANALYSIS OF MI EEG SIGNALS
Both TVNAR and TVNARX models with a nonlinear degree $\kappa = 2$ are used to represent the potential causal relations between channel C3 and C4 during left and right hand imagery tasks. The initial $TVNARX (n_y, n_x)$ model for EEG signals is given below:

$$y(t) = \varpi_0 + \sum_{i=1}^{n_y} \varpi_1(i) y(t-i) + \sum_{j=1}^{n_x} \varpi_2(j) x(t-j)$$
$$+ \sum_{i_1=1}^{n_y} \sum_{i_2=1}^{n_y} \varpi_3(i_1, i_2) y(t-i_1) y(t-i_2)$$
$$+ \sum_{j_1=1}^{n_x} \sum_{j_2=1}^{n_x} \varpi_4(j_1, j_2) x(t-j_1) x(t-j_2)$$

Y. Li *et al.*: Time-Varying Nonlinear Causality Detection Using Regularized Orthogonal Least Squares and Multi-Wavelets
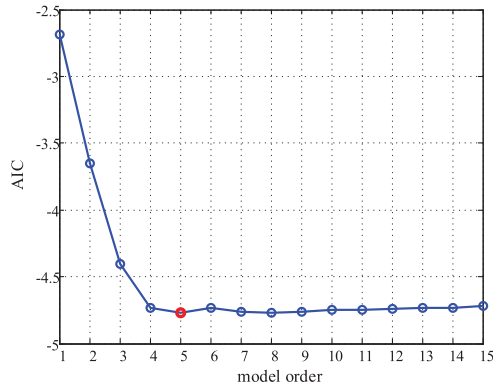
IEEE *Access*

FIGURE 6. A typical example of the proper mode size determined by Eq. (21) for one trial.

$$+ \sum_{i=1}^{n_y} \sum_{j=1}^{n_x} \varpi_5 (i, j) \, y \, (t - i) \, x \, (t - j) \tag{29}$$

The third, fourth, fifth and sixth order B-splines with the scale index $j = 4$ are employed to construct the TVNAR and TVNARX models. For each trial, the optimal model order can be determined by minimizing the AIC criterion in Eq. (21) with the range of $1 \leq n_y \leq 15$ [13]. Fig. 6 shows a typical example of the order selection process for one trial using (21), and the optimal model order is equal to 5. Similarly, the optimal model order of all trials can be calculated. For example, the results of 21 trials sampled from channel C3 while one subject performing left and right hand MI activities are shown in Fig. 7. Based on the constructed TVNAR and TVNARX models by the proposed multi-wavelets-based ROLS method, the TVNGCs between channels C3 and C4 in both directions can be further evaluated by Eqs. (6) and (11).

Fig. 8(a) is the time-varying nonlinear causality results between left hand MI EEG signals shown in Fig. 5(a), where blue curve represents causality from C3 to C4 and black curve describes that from C4 to C3, and the dotted line denotes the corresponding significance threshold. For the right hand MI signals shown in Fig. 5(b), the causal relations between them are given in Fig. 9(a), where the permutation threshold is also represented as the black dotted line. In addition, following the causal flow defined in [58], the time-varying causal flows between channels C3 and C4 under different MI tasks are also calculated for a better understanding of causal connectivity from the aspect of graph theoretical analysis. The associated topographical maps of the causal flows within MI period 0-4s are presented in Fig. 8(b) and Fig. 9(b), which give a spatiotemporal representation of the time-varying GC, and thus make the changing process of the quantified causalities to be visual and intuitive.

From Fig. 8, the strength of nonlinear interaction from C4 to C3 ($NGC_{C4 \rightarrow C3} (t)$) is larger than that from C3 to C4 ($NGC_{C3 \rightarrow C4} (t)$) during left hand imagery within the period of 1.5~2.9s. Just as the topographical maps shown in Fig. 8(b), when performing left hand imagery, Channel C4 exerts strong influence on C3 over the time interval
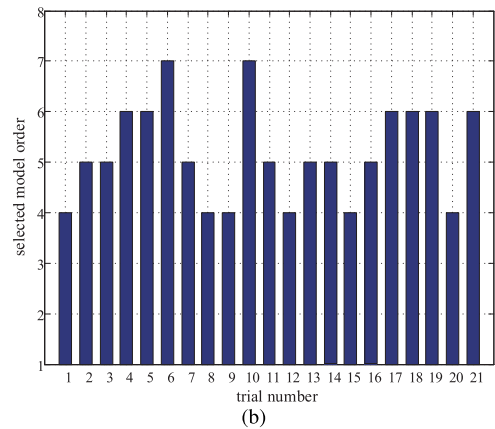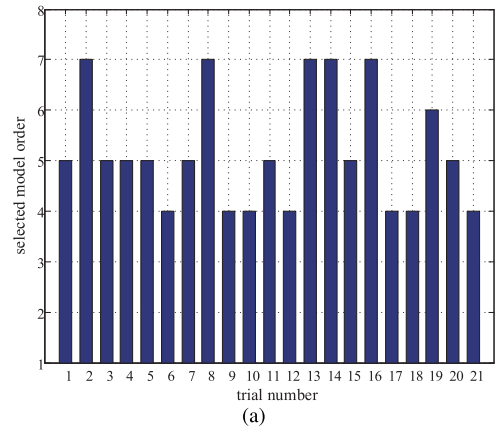

FIGURE 7. The results of the optimal model order for 21 trials sampled from channel C3 during MI: (a) left hand; (b) right hand.
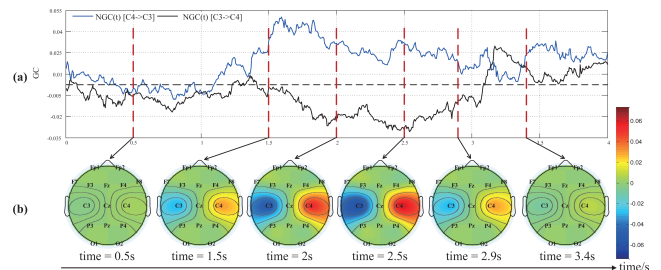

FIGURE 8. (a) Time-varying nonlinear causalities between C3 and C4 during left hand MI (blue curve: GC from C4 to C3, black curve: GC from C3 to C4), and the significance threshold is represented by black dotted line. (b) The associated topographical maps of causal flows.

[1.5s, 2.9s], and the associated causal flow is positive, hence channel C4 can be treated as a causal source with respect to C3. In contrast, the information flow from C3 to C4 is negative and thus C3 is regarded as a causal sink [58]. On the contrary of the case in Fig. 8, the values of $NGC_{C3 \rightarrow C4} (t)$ in Fig. 9 is significantly larger than $NGC_{C4 \rightarrow C3} (t)$ for right hand imagery over the time interval [1.0s, 3.2s], and the associated topographical maps in Fig. 9(b) indicate that channel C3 is the causal source and channel C4 is the causal sink in this time period under right hand MI task.
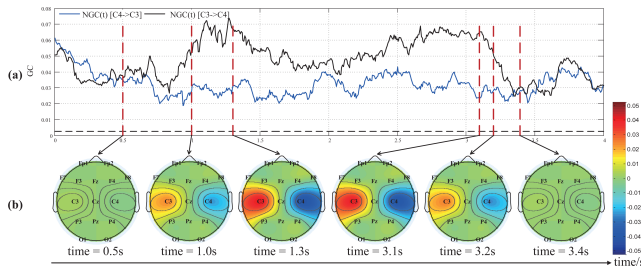
FIGURE 9. (a) Time-varying nonlinear causalities between C3 and C4 during right hand MI (blue curve: GC from C4 to C3, black curve: GC from C3 to C4), and the significance threshold is represented by black dotted line. (b) The associated topographical maps of causal flows.

Figs. 8-9 present that the transient changes of the nonlinear GC between C3 and C4 under MI tasks can be clearly measured by employing the newly introduced TVNGC testing method. Specifically, an obvious nonlinear causality from C4 to C3 for the imagination of left hand and a strong nonlinear directional connectivity from C3 to C4 during right hand imagery are detected. These nonlinear results are consistent with the recent studies reported in [51] and [52], and can better reflect the neural connectivity variations between inherent nonlinear EEG signals induced during MI tasks because of the nonlinearity of the fundamental models used in the proposed approach. Additionally, the precise time periods of the interaction for MI can be well determined, such as [1.5s, 2.9s] and [1.0s, 3.2s] for left and right hand MI, respectively, and the instantaneous dynamical processes of causalities between different brain regions over the whole MI tasks with 4s can be clearly revealed, which demonstrates the applicability of the high time resolution causal relations obtained by the proposed framework.

V. CONCLUSION

A new TVNGC detection method has been proposed based on a parametric modelling framework, where the associated time-dependent parameters are approximated by a set of multi-wavelet basis functions so that the initial time-varying model can be re-formulated to a time-invariant linear-in-the-parameters form. The ROLS algorithm is further applied to reduce the linear-in-the-parameters model and the resultant coefficients are then used to recover the original time-varying parameters. Three case studies have been carried out to illustrate the performance of the proposed method, these include two simulation examples with known causal relations and an application to real EEG signals during MI tasks. The simulation examples show that the proposed approach can effectively detect time-varying linear and nonlinear causal interactions, and its overall performance outperforms the other three methods in the presence of high-level noise. For real MI EEG signals, strong directional connectivities during left and right hand imagery tasks have been observed, which demonstrates that the proposed procedure is more powerful in detecting fast-changing causalities between two nonstationary biomedical signals.

Note that the proposed causal detection framework is suitable for causality analysis between time-varying bivariate systems, while the direct causal interaction among three or more simultaneous time series and spectral causal representation are not discussed in this early stage, which may fail to reveal essential potential connectivities of the whole brain EEG signals. In order to further improve the applicability of the proposed method, the causal detection framework will be further extended to multivariate cases and spectrum representation evaluated by using multi-channel EEG recordings. Another main limitation of the proposed approach is its heavy computation load, which may be much higher than existing adaptive detection methods, this is mainly caused by the calculation and selection procedure of a number of expansion terms considered for each basic model. We intend to improve the efficiency of the time-varying GC analysis to reduce the computation time by applying other sparse representation algorithms like least absolute shrinkage and selection operator (Lasso) or Orthogonal Matching Pursuit (OMP). These results will be published in our future work.
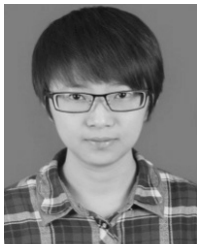
REFERENCES

[1] S. Palva and J. M. Palva, "Discovering oscillatory interaction networks with M/EEG: Challenges and breakthroughs," *Trends Cognit. Sci.*, vol. 16, no. 4, pp. 219–230, 2012.

[2] B. S. W. Ng, N. K. Logothetis, and C. Kayser, "EEG phase patterns reflect the selectivity of neural firing," *Cerebral Cortex*, vol. 23, no. 2, pp. 389–398, 2012.

[3] U. Melia et al., "Mutual information measures applied to EEG signals for sleepiness characterization," *Med. Eng. Phys.*, vol. 37, no. 3, pp. 297–308, 2015.

[4] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[5] A. K. Seth, A. B. Barrett, and L. Barnett, "Granger causality analysis in neuroscience and neuroimaging," *J. Neuroscience*, vol. 35, no. 8, pp. 3293–3297, 2015.

[6] M. Hu, W. Li, and H. Liang, "A copula-based Granger causality measure for the analysis of neural spike train data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, Jan. 2015, doi: 10.1109/TCBB.2014.2388311.

[7] M. V. Sysoeva, E. Sitnikova, I. V. Sysoev, B. P. Bezruchko, and G. van Luijtelaar, "Application of adaptive nonlinear Granger causality: Disclosing network changes before and after absence seizure onset in a genetic rat model," *J. Neurosci. Methods*, vol. 226, pp. 33–41, Apr. 2014.

[8] A. K. Seth, P. Chorley, and L. C. Barnett, "Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling," *NeuroImage*, vol. 65, pp. 540–555, Jan. 2013.

[9] V. Youssofzadeh, G. Prasad, M. Naeem, and K. Wong-Lin, "Temporal information of directed causal connectivity in multi-trial ERP data using partial Granger causality," *Neuroinformatics*, vol. 14, no. 1, pp. 99–120, 2016.

[10] Q. Luo et al., "Spatio-temporal Granger causality: A new framework," *NeuroImage*, vol. 79, pp. 241–263, Oct. 2013.

[11] X. Wang, Y. Chen, and M. Ding, "Estimating Granger causality after stimulus onset: A cautionary note," *NeuroImage*, vol. 41, no. 3, pp. 767–776, Jul. 2008.

[12] M. Ding, S. L. Bressler, W. Yang, and H. Liang, "Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: Data preprocessing, model validation, and variability assessment," *Biological*, vol. 83, no. 1, pp. 35–45, 2000.

[13] W. Hesse, E. Möller, M. Arnold, and B. Schack, "The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies," *J. Neurosci. Methods*, vol. 124, no. 1, pp. 27–44, 2003.

Y. Li *et al.*: Time-Varying Nonlinear Causality Detection Using Regularized Orthogonal Least Squares and Multi-Wavelets

IEEE *Access*

[14] H. Wang, W. Sun, and P. X. Liu, "Adaptive intelligent control of nonaffine nonlinear time-delay systems with dynamic uncertainties," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 7, pp. 1474–1485, Jul. 2017.

[15] H. Wang, P. X. Liu, and P. Shi, "Observer-based fuzzy adaptive output-feedback control of stochastic nonlinear multiple time-delay systems," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2568–2578, Sep. 2017.

[16] X. D. Zhao, P. Shi, X. L. Zheng, and J. H. Zhang, "Intelligent tracking control for a class of uncertain high-order nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 9, pp. 1976–1982, Sep. 2016.

[17] X. Zhao, P. Shi, and X. Zheng, "Fuzzy adaptive control design and discretization for a class of nonlinear uncertain systems," *IEEE Trans. Cybern.*, vol. 46, no. 6, pp. 1476–1483, Jun. 2016.

[18] E. Başar, *Chaos in Brain Function: Containing Original Chapters by E. Başar and T. H. Bullock and Topical Articles Reprinted From the Springer Series in Brain Dynamics*. Berlin, Germany: Springer, 2012.

[19] B. Gourévitch, R. Le Bouquin-Jeannès, and G. Faucon, "Linear and non-linear causality between signals: Methods, examples and neurophysiological applications," *Biol. Cybern.*, vol. 95, no. 4, pp. 349–369, 2006.

[20] Y. Li, H.-L. Wei, S. A. Billings, and X.-F. Liao, "Time-varying linear and nonlinear parametric model for Granger causality analysis," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 85, no. 4, p. 041906, 2012.

[21] Y. Zhao, S. A. Billings, H.-L. Wei, and P. G. Sarrigiannis, "A parametric method to measure time-varying linear and nonlinear causality with applications to EEG data," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3141–3148, Nov. 2013.

[22] S. A. Billings, S. Chen, and M. J. Korenberg, "Identification of MIMO non-linear systems using a forward-regression orthogonal estimator," *Int. J. Control*, vol. 49, no. 6, pp. 2157–2189, 1989.

[23] R. J. Boynton, M. A. Balikhin, S. A. Billings, H. L. Wei, and N. Ganushkina, "Using the NARMAX OLS-ERR algorithm to obtain the most influential coupling functions that affect the evolution of the magnetosphere," *J. Geophys. Res. Space Phys.*, vol. 116, no. A5, 2011.

[24] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[25] H. L. Wei and S. A. Billings, "Identification of time-varying systems using multiresolution wavelet models," *Int. J. Syst. Sci.*, vol. 33, no. 15, pp. 1217–1228, 2002.

[26] Y. Li, W.-G. Cui, Y.-Z. Guo, T. Huang, X.-F. Yang, and H.-L. Wei, "Time-varying system identification using an ultra-orthogonal forward regression and multiwavelet basis functions with applications to EEG," *IEEE Trans. Neural Netw. Learn. Syst.*, Jun. 2017, doi: 10.1109/TNNLS.2017.2709910.

[27] Y. Li, W. Cui, M. Luo, K. Li, and L. Wang, "Epileptic seizure detection based on time-frequency images of EEG signals using Gaussian mixture model and gray level co-occurrence matrix features," *Int. J. Neural Syst.*, Jan. 2018, doi: 10.1142/S012906571850003X.

[28] S. Chen, E. S. Chng, and K. Alkadhimi, "Regularized orthogonal least squares algorithm for constructing radial basis function networks," *Int. J. Control*, vol. 64, no. 5, pp. 829–837, 1996.

[29] S. Chen, X. Hong, and C. J. Harris, "Sparse kernel regression modeling using combined locally regularized orthogonal least squares and D-optimality experimental design," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 1029–1036, Jun. 2003.

[30] K. L. Lee and S. A. Billings, "Time series prediction using support vector machines, the orthogonal and the regularized orthogonal least-squares algorithms," *Int. J. Syst. Sci.*, vol. 33, no. 10, pp. 811–821, 2002.

[31] Y. Zhao, S. A. Billings, H. Wei, F. He, and P. G. Sarrigiannis, "A new NARX-based Granger linear and nonlinear casual influence detection method with applications to EEG data," *J. Neurosci. Methods*, vol. 212, no. 1, pp. 79–86, 2013.

[32] H.-L. Wei and S. A. Billings, "Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information," *Int. J. Model., Identificat. Control*, vol. 3, no. 4, pp. 341–356, 2008.

[33] A. M. Marshall *et al.*, "Quantifying heterogeneous responses of fish community size structure using novel combined statistical techniques," *Global Change Biol.*, vol. 22, no. 5, pp. 1755–1768, 2016.

[34] Y. Li, H.-L. Wei, S. A. Billings, and P. G. Sarrigiannis, "Identification of nonlinear time-varying systems using an online sliding-window and common model structure selection (CMSS) approach with applications to EEG," *Int. J. Syst. Sci.*, vol. 47, no. 11, pp. 2671–2681, 2016.

[35] C. K. Chui, *An Introduction to Wavelets*. New York, NY, USA: Elsevier, 2016.

[36] Y. Li, W.-G. Cui, M.-L. Luo, K. Li, and L. Wang, "High-resolution time-frequency representation of EEG data using multi-scale wavelets," *Int. J. Syst. Sci.*, vol. 48, no. 12, pp. 2658–2668, 2017.

[37] H.-L. Wei, S. A. Billings, and J. J. Liu, "Time-varying parametric modelling and time-dependent spectral characterisation with applications to EEG signals using multiwavelets," *Int. J. Model., Identificat. Control*, vol. 9, no. 3, pp. 215–224, 2010.

[38] Y. Li, H.-L. Wei, and S. A. Billings, "Identification of time-varying systems using multi-wavelet basis functions," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 3, pp. 656–663, May 2011.

[39] S. Chen, X. Hong, E. F. Khalaf, F. E. Alsaadi, and C. J. Harris, "Comparative performance of complex-valued B-spline and polynomial models applied to iterative frequency-domain decision feedback equalization of Hammerstein channels," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2872–2884, Dec. 2016.

[40] H. L. Wei and S. A. Billings, "An efficient nonlinear cardinal b-spline model for high tide forecasts at the Venice lagoon," *Nonlinear Processes Geophys.*, vol. 13, no. 5, pp. 577–584, 2006.

[41] C. K. Chui and J.-Z. Wang, "On compactly supported spline wavelets and a duality principle," *Trans. Amer. Math. Soc.*, vol. 330, no. 2, pp. 903–915, Apr. 1992.

[42] Y. Li, M.-L. Luo, and K. Li, "A multiwavelet-based time-varying model identification approach for time–frequency analysis of EEG signals," *Neurocomputing*, vol. 193, pp. 106–114, Jun. 2016.

[43] S. Chen, S. A. Billings, and P. M. Grant, "Non-linear system identification using neural networks," *Int. J. Control*, vol. 51, no. 6, pp. 1191–1214, 1990.

[44] G. R. Bigg *et al.*, "A century of variation in the dependence of greenland iceberg calving on ice sheet surface mass balance and regional climate change," in *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 470, no. 2166, p. 20130662, 2014.

[45] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[46] S. R. Soekadar, N. Birbaumer, M. W. Slutzky, and L. G. Cohen, "Brain–machine interfaces in neurorehabilitation of stroke," *Neurobiol. Disease*, vol. 83, pp. 172–179, Nov. 2015.

[47] M. Takahashi *et al.*, "Event related desynchronization-modulated functional electrical stimulation system for stroke rehabilitation: A feasibility study," *J. Neuroeng. Rehabil.*, vol. 9, no. 1, p. 56, 2012.

[48] M. Gongora, C. Peressutti, S. Machado, S. Teixeira, B. Velasques, and P. Ribeiro, "Progress and prospects in neurorehabilitation: Clinical applications of stem cells and brain-computer interface for spinal cord lesions," *Neurol. Sci.*, vol. 34, no. 4, pp. 427–433, 2013.

[49] J. Long, Y. Li, H. Wang, T. Yu, J. Pan, and F. Li, "A hybrid brain computer interface to control the direction and speed of a simulated or real wheelchair," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 5, pp. 720–729, Sep. 2012.

[50] R. Zhang *et al.*, "Efficient resting-state EEG network facilitates motor imagery performance," *J. Neural Eng.*, vol. 12, no. 6, p. 066024, 2015.

[51] S. Hu, H. Wang, J. Zhang, W. Kong, Y. Cao, and R. Kozma, "Comparison analysis: Granger causality and new causality and their applications to motor imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 7, pp. 1429–1444, Jul. 2016.

[52] D. Rathee, H. Cecotti, and G. Prasad, "Estimation of effective fronto-parietal connectivity during motor imagery using partial Granger causality analysis," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2055–2062.

[53] T. Liu *et al.*, "Cortical dynamic causality network for auditory-motor tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 8, pp. 1092–1099, Aug. 2017.

[54] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[55] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.

[56] E. Niedermeyer and F. L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Baltimore, MD, USA: Williams & Wilkins, 2005.

IEEE *Access*

Y. Li *et al.*: Time-Varying Nonlinear Causality Detection Using Regularized Orthogonal Least Squares and Multi-Wavelets

[57] J. Wagner, T. Solis-Escalante, P. Grieshofer, C. Neuper, and G. Müller-Putz, and R. Scherer, "Level of participation in robotic-assisted treadmill walking modulates midline sensorimotor EEG rhythms in able-bodied subjects," *NeuroImage*, vol. 63, no. 3, pp. 1203–1211, 2012.

[58] A. K. Seth, "Causal connectivity of evolved neural networks during behavior," *Netw. Comput. Neural Syst.*, vol. 16, no. 1, pp. 35–54, 2005.

YUZHU GUO received the B.Sc. and M.Sc. degrees from the Beijing Institute of Technology, Beijing, China, and the Ph.D. degree in automatic control and systems engineering from The University of Sheffield, U.K., in 2009. He is currently an Associate Professor with the School of Automation Science and Electrical Engineering, Beihang University. His research interests include system identification and information processing for nonlinear systems, narmax methods, nonlinear spectral analysis, spatio-temporal systems, multi-scale modeling of neuro-musculo-skeletal systems, and bioinformatics.

YANG LI received the Ph.D. degree in automatic control and systems engineering from Sheffield University in 2011. After one year of Post-Doctoral Research with the Department of Computer and Biomedical Engineering, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. He joined Beihang University as an Associate Professor with the Department of Automation Sciences and Electrical Engineering in 2013. His main research interests include system identification and modeling for complex nonlinear processes: NARMAX methodology and applications, nonlinear and nonstationary signal processing, intelligent computation and data mining, parameter estimation and model optimization, and sparse representation.

ZHONGYI HU (SM'16) received the B.S. degree in automation from the Beijing Institute of Mechanical Industry in 1995 and the M.S. and Ph.D. degrees in control engineering and intelligent traffic engineering from the Wuhan University of Technology, in 2006 and 2015, respectively. He is currently an Associate Professor with Wenzhou University. His current research interests include intelligent systems, target tracking, image processing and information management. He is a Senior Member of CCF.

HUA-LIANG WEI received the Ph.D. degree from the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, U.K., in 2004.

He is currently a Senior Lecturer with the Department of Automatic Control and Systems Engineering, The University of Sheffield. His current research interests include system identification and data analytics for complex systems, data driven modeling and data mining, NARMAX methodology and its applications, statistical digital signal processing, machine learning and neural networks, spatio-temporal system modeling, neuro-wavelet models for learning, nonstationary (time varying) process modeling, forecasting of complex dynamic processes, generalized regression analysis, linear and nonlinear optimization, multidisciplinary applications in medicine and biomedicine, medical informatics, space weather, and environmental sciences.

MENG-YING LEI was born in Xiantao, Hubei, China, in 1995. She received the bachelor's degree from the Department of Automation, Huazhong Agricultural University, Wuhan, China, in 2016. She is currently pursuing the master's degree with the Department of Automation Science and Electrical Engineering, Beihang University. Her research interests include signal processing, machine learning, system identification, and time-frequency domain analysis.

• • •