Article:
Li, Y., Lei, M., Cui, W. et al. (2 more authors) (2019) A parametric time frequency-conditional Granger causality method using ultra-regularized orthogonal least squares and multiwavelets for dynamic connectivity analysis in EEGs. IEEE Transactions on Biomedical Engineering. ISSN 0018-9294

https://doi.org/10.1109/TBME.2019.2906688

A Parametric Time Frequency-Conditional Granger Causality Method Using Ultra-regularized Orthogonal Least Squares and Multiwavelets for Dynamic Connectivity Analysis in EEGs

Yang Li, Mengying Lei*, Weigang Cui, Yuzhu Guo, and Hua-Liang Wei*

Abstract— Objective: This study proposes a new parametric TF-CGC (time-frequency conditional Granger causality) method for high-precision connectivity analysis over time and frequency domain in multivariate coupling nonstationary systems, and applies it to source EEG signals to reveal dynamic interaction patterns in oscillatory neocortical sensorimotor networks. Methods: The Geweke's spectral measure is combined with the TVARX (time-varying autoregressive with exogenous input) modelling approach, which uses multiwavelet-based ultra-regularized orthogonal least squares (UROLS) algorithm aided by APRESS (adjustable prediction error sum of squares), to obtain high-resolution time-varying CGC representations. The UROLS-APRESS algorithm, which adopts both the regularization technique and the ultra-least squares criterion to measure not only the signal themselves but also the weak derivatives of them, is a novel powerful method in constructing time-varying models with good generalization performance, and can accurately track smooth and fast changing causalities. The generalized measurement based on CGC decomposition is able to eliminate indirect influences in multivariate systems. Results: The proposed method is validated on two simulations and then applied to source level motor imagery (MI) EEGs, where the predicted distributions are well recovered with high TF precision, and the detected connectivity patterns of MI-EEGs are physiologically interpretable and yield new insights into the dynamical organization of oscillatory cortical networks. Conclusion: Experimental results confirm the effectiveness of the TF-CGC method in tracking rapidly varying causalities of EEG-based oscillatory networks. Significance: The novel TF-CGC method is expected to provide important information of neural mechanisms of perception and cognition.

Index Terms—EEG, time-frequency (TF) conditional Granger causality (CGC), multiwavelets, ultra-regularized orthogonal least squares (UROLS), adjustable prediction error sum of squares (APRESS), motor imagery (MI), dynamic connectivity.

Manuscript received June 28, 2018. This work was supported by the National Natural Science Foundation of China [U1809209, 61671042, 61403016], Beijing Natural Science Foundation [L182015, 4172037], Open Fund Project of Fujian Provincial Key Laboratory in Minjiang University [MJUKF201702], and Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/I011056/1 and Platform programme under Grant EP/H00453X/1, U.K. (Corresponding authors: Mengying Lei; Hua-Liang Wei.)

Yang Li is with the Department of Automation Sciences and Electrical Engineering, Beijing Advanced Innovation Center for Big Data and Brain Computing, Beijing Advanced Innovation Center for Big Date-based Precision Medicine, Beihang University, Beijing, China.

Mengying Lei, Weigang Cui, and Yuzhu Guo are with the Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing, China (e-mail: lmylei@buaa.edu.cn).

Hua-Liang Wei is with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield S1 3DJ, U.K (e-mail: w.hualiang@sheffield.ac.uk).

I. Introduction

DYNAMIC interactions within brain regions enable synchronization of neuronal oscillations, which is a suggested mechanism underlying the perceptual and cognitive functions [1]. Analyzing time-varying interaction patterns of oscillatory brain networks is a considerably important and challenging research topic in the neuroscience field [2]. The traditional measures for quantifying interdependencies among neural systems in the time and frequency domain are mainly based on correlation and coherence, respectively [3]. The two measures hold a significant share in functional network analysis, but both omit the direction information of interaction. Based on Granger causality [4], several methods, for example, directed coherence and directed transfer function (DTF) [3] have been proposed to infer directed influences between signals. These methods, however, cannot distinguish between direct and indirect causal effects in multivariate coupling systems effectively. This further boosted the development of partial directed coherence (PDC) [3, 5] and direct DTF (dDTF) [6] to deal with the drawback. In addition, mutual information [7] is another approach exploited to reveal causal dependence, and the original symmetric estimates led to a directed measure called conditional transfer entropy (TE) [7, 8].

The widely known measures discussed above previously assume the underlying signals are stationary and the interactions are constant over time, such an assumption restricts he application of these methods in some degree for dealing with time-varying neural signals especially EEGs with high nonstationarity. Recently, dynamic Granger causality (GC) analysis has emerged as a powerful technique to detect directed interactions among coupled nonstationary systems, and has been extensively investigated in neurophysiological studies [9, 10]. The key in dynamic GC detection is the identification of the time-varying autoregressive with exogenous input (TVARX) models for nonstationary signals. Several methods have been

developed for assessing dynamic GC relations in time or frequency domain [11], mainly including nonparametric method [12, 13], sliding window approach [14], adaptive multivariate estimation [10] and parametric modelling approach [15-18].

In the nonparametric GC detection method [12], the time-frequency (TF) causality analysis was based on nonparametric wavelet transforms and the performance was validated by monkey local field potentials. Nevertheless, for this method, it is difficult to select desirable initial parameters of wavelets, including the number of tapers, wavelet prototype and the resolution trade-off factor, to ensure both good time and frequency resolution simultaneously [12]; and thus the estimates may not be reliable when only a few trials of short length data sets are available. In the sliding window approach [19], the temporal functions of spectral GC can be roughly extracted by analyzing traditional time-invariant GC influences for each single window through ARX modelling algorithm. However, the time resolution of this approach is smeared and the detection performance depends on the window size, which limits its practical applicability for nonstationary systems. In the adaptive multivariate strategy, the recursive least squares (RLS) and Kalman filtering algorithms are commonly used for estimating time-varying parameters [10, 20]. These adaptive methods can detect slow varying interaction relations, but they are sensitive to noise and may fail to track rapid changing connectivity due to the deficiency of slow convergence speed [21, 22].

Compared with the above mentioned methods, the parametric approaches, based upon TVARX model identification using a basis function expansion and regression scheme, can provide better performance for dynamic GC detection [16, 21]. In such a detection framework, the basic time-varying models of signals are firstly estimated by applying a set of pre-defined basis functions with good representation properties [23-25] and running an efficient model structure determination algorithm such as the orthogonal forward regression [26, 27]; time-varying variances of model prediction errors and corresponding GCs can then be effectively calculated from the reduced refined TVARX models. For example, Li et al. employed multiwavelet basis functions with regularized orthogonal least squares (ROLS) to approximate the time-varying parameters of TVARX models, which were applied in successfully detecting both rapid and slow varying causalities between two nonstationary signals [21].

Despite the multiwavelet expansion approach with ROLS algorithm provides a general parametric method for time-varying GC detection, two defects are remained in this scheme. First, although the ROLS algorithm enables better generalization in model construction than the conventional OLS i.e. OFR (orthogonal forward regression) and works well even in the presence of severe noise [28, 29], the method may produce suboptimal model with possible spurious or insufficient model terms when the signals are not persistently exciting or contaminated by different levels of noise [30, 31]. In this case, the resulting under-fitting TVARX models might produce incorrect and low precision GC distributions. Second, this pairwise time-domain GC approach ignores frequency information which is crucial for the analysis of neurophysiological signals with abundant oscillatory content, like EEG, and it cannot distinguish direct and indirect effects among systems with more than two simultaneously acquired signals. Thus, the conventional ROLS method may fail to reveal dynamic connectivity in coupled oscillatory brain networks. Currently, because of its non-invasive nature, good temporal resolution and low cost, EEG technique is often used for studying brain activities [32-34]. However, there is still lack of high resolution time-frequency causality method for EEG-based connectivity analysis even in recent researches due to the high nonstationarity and complexity of EEG signals.

In this paper, we propose a new TF-CGC (time-frequency conditional Granger causality) method for analyzing dynamic connectivity among multivariate coupling nonstationary signals over time and frequency domain, where a powerful ultra-regularized orthogonal least squares (UROLS) algorithm is employed to measure the time-frequency conditional causalities. Specifically, the time-varying parameters in TVARX models are firstly expanded by a finite number of multiwavelet basis functions for tracking both the global trend and local changes in nonstationary signals [31, 35]. Then the UROLS algorithm, which improves the classical ROLS in using not only the residuals between the observed signals and the predicted values but also the associated weak derivatives to measure the model fitness [30], is applied to determine the parsimonious model structure and associated parameters. In the proposed UROLS algorithm, a modified cross-validation criterion named adjustable prediction error sum of squares (APRESS) is incorporated to facilitate the monitoring of the forward orthogonal search procedure and the determination of the model complexity [36, 37]. Finally, a high resolution TF-CGC representation is established by combining the accurately identified TVARX models with the statistically-explicable mathematical framework of Geweke's spectral CGC [38]. Our proposed TF-CGC method is firstly tested on two simulated nonstationary coupling systems, and then applied to source EEG data acquired from MI tasks. Experimental results demonstrate the efficiency of the proposed TF-CGC method in detecting dynamic interaction activities among nonstationary and oscillatory brain systems. A main contribution of this study is that the newly developed multiwavelet-based UROLS-APRESS algorithm is innovatively introduced to reveal dynamic connection patterns in TF domain based on the CGC decomposition. It is expected that the novel implementation of the UROLS with multiwavelets to TF-CGC analysis can provide important insights into the neural mechanisms underlying perceptual and cognitive functions, and inspire further development of more powerful approaches for dynamic connectivity analysis.

II. Methods

The classical GC is formulated based on univariate AR or bivariate ARX models. A TF-CGC decomposition method, which combines a time-varying system identification approach with Geweke's spectral CGC measure, is proposed in this work. The TF-CGC decomposition for multivariate time series is built on the TVARX modeling, thus the newly introduced nonstationary model identification method is first discussed in this section. The discussion focuses on three time series, but it can easily be extended to more than three sets of time series.

A. TVARX model identification using multiwavelets for TF-CGC analysis

Consider three stochastic processes $X=\{x(t)\}$, $Y=\{y(t)\}$ and $Z=\{z(t)\}$, with sampling index $t=1,2,\cdots,N$, where the TF-CGC relations from Y to X conditional on Z is to be evaluated. Let the joint TVARX representations of $x(t)$ and $z(t)$ be

$$\begin{aligned} x(t) &= \sum_{i=1}^{I_1} a_{11,i}(t)\,x(t-i) + \sum_{i=1}^{I_2} a_{12,i}(t)\,z(t-i) + e_1(t) \\ z(t) &= \sum_{i=1}^{I_1} a_{21,i}(t)\,x(t-i) + \sum_{i=1}^{I_2} a_{22,i}(t)\,z(t-i) + e_2(t) \end{aligned} \tag{1}$$

Denote the joint TVARX model of all three processes $x(t)$,

$y(t)$ and $z(t)$ as

$$\begin{aligned}
x(t) &= \sum_{i=1}^{I_1} b_{11,i}(t)\,x(t-i) + \sum_{i=1}^{I_2} b_{12,i}(t)\,y(t-i) + \sum_{i=1}^{I_3} b_{13,i}(t)\,z(t-i) + e_3(t) \\
y(t) &= \sum_{i=1}^{I_1} b_{21,i}(t)\,x(t-i) + \sum_{i=1}^{I_2} b_{22,i}(t)\,y(t-i) + \sum_{i=1}^{I_3} b_{23,i}(t)\,z(t-i) + e_4(t) \\
z(t) &= \sum_{i=1}^{I_1} b_{31,i}(t)\,x(t-i) + \sum_{i=1}^{I_2} b_{32,i}(t)\,y(t-i) + \sum_{i=1}^{I_3} b_{33,i}(t)\,z(t-i) + e_5(t)
\end{aligned} \tag{2}$$

where $\{a_{11,i}(t), \cdots, a_{22,i}(t), b_{11,i}(t), \cdots, b_{33,i}(t)\}$ are time-varying parameters to be estimated, and $\{e_1(t), \cdots, e_5(t)\}$ are independent and normal distributed noise sequences with zero means. An efficient solution when identifying these TVARX models is to expand the time-varying parameters onto a set of basis functions $\{\varphi_m(t): m = 1,2,\cdots,M\}$, for example, for the trivariate TVARX process with respect to signal $x(t)$ in (2)

$$\begin{aligned}
x(t) &= \sum_{i=1}^{I_1} b_{11,i}(t)\,x(t-i) + \sum_{i=1}^{I_2} b_{12,i}(t)\,y(t-i) \\
&\quad + \sum_{i=1}^{I_3} b_{13,i}(t)\,z(t-i) + e_3(t) \\
&= \sum_{n=1}^{V}\sum_{i=1}^{I_n} c_{n,i}(t)\,\kappa_{n,i} + e_3(t)
\end{aligned} \tag{3}$$

where $x(t)$, $y(t)$, $z(t)$ are the system output and input with maximum lags I_1, I_2 and I_3, respectively, $V = 3$ is the number of input variables; $c_{n,i}(t)$ (i.e. $b_{11,i}(t), b_{12,i}(t), b_{13,i}(t)$) denote the time-varying parameters, and $\kappa_{n,i}$ represent $x(t-i), y(t-i), z(t-i)$, when $n = 1,2,3$, respectively. Then the model can be expanded as

$$\begin{aligned}
x(t) &= \sum_{n=1}^{V}\sum_{i=1}^{I_n}\sum_{m=1}^{M} \alpha_{n,i,m}\varphi_m(t)\kappa_{n,i} + e_3(t) \\
&= \psi(t)^T\theta + e_3(t)
\end{aligned} \tag{4}$$

where $\alpha_{n,i,m}$ denote the time-invariant expansion parameters of basis functions $\varphi_m(t)$, M is the number of the basis sequences, $\psi(t) = [\chi_X(t), \chi_Y(t), \chi_Z(t)]^T$ is a $(I_1 + I_2 + I_3) \times M \times 1$ dimensional regression vector, in which $\chi_X(t) = [x(t-1)\phi(t)^T, x(t-2)\phi(t)^T, \cdots, x(t-I_1)\phi(t)^T]$, $\chi_Y(t) = [y(t-1)\phi(t)^T, y(t-2)\phi(t)^T, \cdots, y(t-I_2)\phi(t)^T]$, and $\chi_Z(t) = [z(t-1)\phi(t)^T, z(t-2)\phi(t)^T, \cdots, z(t-I_3)\phi(t)^T]$ with $\phi(t) = [\varphi_1(t), \varphi_2(t), \cdots, \varphi_M(t)]^T$, the expansion coefficient vector is $\theta = [\alpha_{1,1,1}, \cdots, \alpha_{1,I_1,M}, \cdots, \alpha_{V,1,1}, \cdots \alpha_{V,I_V,M}]^T$, and the upper script T represents the transpose of a vector. The initial time-varying model then becomes a time-invariant regression problem, since all $\alpha_{n,i,m}$ are now time invariant.

In practice, a proper selection of the basis functions is vital to ensure the identified model performance. A good suggestion is to use multiple wavelet basis functions to effectively track both rapid and slow parameter variations in time-varying processes [27]. Thus, in this work, multi-wavelet basis functions are applied to approximate the time-varying parameters in (3) as

$$c_{n,i}(t) = \sum_{r}\sum_{k\in\Gamma_r} \beta^{r}_{n,i,k}\xi^{r}_{k,j}\left(\frac{t}{N}\right) \tag{5}$$

where $\xi^r_{k,j}(\cdot)$ are wavelet basis functions, with the shift indices $k \in \Gamma_r, \Gamma_r = \{k: -r \le k \le 2^j - 1\}$ and wavelet scale j, $\beta^r_{n,i,k}$ are the associated expanded basis function parameters which are time invariant, r denotes the order of the basis functions, and the function variable t/N is normalised within $[0,1]$.

Cardinal B-splines are an important class of wavelet basis functions that simultaneously possess three remarkable properties, namely compactly supported, analytically formulated and multiresolution analysis oriented, which enable the operation of the wavelet decomposition to be more convenient [39]. Taking the cardinal B-splines as the basis function, the $\xi^r_{k,j}(\cdot)$ can be expressed by the r-th order B-spline B_r as $\xi^r_{k,j}(u) = 2^{j/2}B_r(2^j u - k)$, where j, k are the dilated and shifted versions of wavelet B_r. Generally j is chose to be 3 or a larger number in many B-splines applications [35], and a practical selection of the wavelets are $\{\xi^r_{k,j}: r = 3,4,5\}$, the detail description of B-splines properties can be found in [40]. The decomposition (5) can easily be transformed into the form of (4), where the collection $\{\varphi_m(t): m = 1,2,\cdots,M\}$ is replaced by the union of multi-B-splines families $\sum_r \sum_{k\in\Gamma_r} \xi^r_{k,j}(u)$, then the TVARX model is rewritten by

$$\begin{aligned}
x(t) &= \sum_{n=1}^{V}\sum_{i=1}^{I_n}\sum_{r}\sum_{k\in\Gamma_r} \beta^{r}_{n,i,k}\xi^{r}_{k,j}\left(\frac{t}{N}\right)\kappa_{n,i} + e_3(t) \\
&= \Psi^T(t)\delta + e_3(t)
\end{aligned} \tag{6}$$

where $\Psi^T(t)$ is the expanded term vector at time t and δ is the corresponding time-invariant parameter vector.

Equation (6) indicates that the multi-wavelet basis function expansion method converts the identification of the time-varying model (3) into a time-invariant regression problem. However, the number of candidate model terms in $\Psi^T(t)$ can be very large if the number of involved wavelet basis functions r, the wavelet scale j or the maximum lags I_1, I_2, I_3 are large; as a consequence, the initial full regression model (6) is often redundant, ill-conditioned and not ready for direct use. Thus, selecting significant terms from the pool of the expanded regressors and building a sparse model structure is highly required, and this will be introduced in the next section.

B. The UROLS algorithm for TVARX model identification in TF-CGC analysis

The identification of the TVARX model includes two steps: determining the model structure and estimating the associated parameters. In this section, a new method, referred to as ultra-regularized orthogonal least squares (UROLS), is proposed for time-varying model identification; it incorporates the following three approaches: the ultra-least squares (ULS) metric, the regularized orthogonal least squares (ROLS) algorithm, and adjustable prediction error sum of squares (APRESS).

For generic regression problems, the least squares loss function aims to achieve the best model fitting on the Lebesgue space $L^2([0,T])$, where $[0,T]$ is the time span of signals, and the model that minimizes the square of the L^2 norm is to be identified. The L^2 norm, only measures the similarity of two functions as a whole, cannot characterize the local distribution difference at each time instance, thus neglects some important information of details in shape [30]. The absence of this crucial information might lead to a model structure which cannot sufficiently represent the inherent dynamics of the data (and therefore the associated system) especially when the system is not persistently excited. It is known that most physical systems behave mainly as a low-pass filter, and are actually defined on the subspace of $L^2([0,T])$, that is, the Sobolev space $H^d([0,T]) = \{u(t) \in L^2([0,T]) | D^\upsilon u \in L^2([0,T]), \upsilon = 1,2,\cdots,d\}$, where the weak derivatives $D^\upsilon u$ up to d-th are also L^2 integrable [31]. Thus, a stricter metric defined as $u_{H^d} = \sqrt{\sum_{\upsilon=1}^{d}\|D^\upsilon u\|_2^2}$ [30], which can reveal the entire useful information of observations realized in the Sobolev space, is used in this study.

Based on the stricter norm, a new ULS criterion for identi-

fying model (6) can be expressed by

$$\begin{aligned} J_{ULS} = &\left\| x - \sum_{n=1}^{V}\sum_{i=1}^{I_n}\sum_{r}\sum_{k\in\Gamma_r}\beta^{r}_{n,i,k}u^{r}_{n,i,k}\right\|_2^2 \\ &+\sum_{\upsilon=1}^{d}\left\|\left\{\bar{x}^{\upsilon} - \sum_{n=1}^{V}\sum_{i=1}^{I_n}\sum_{r}\sum_{k\in\Gamma_r}\beta^{r}_{n,i,k}\left(\bar{u}^{r}_{n,i,k}\right)^{\upsilon}\right\}\right\|_2^2 \end{aligned} \tag{7}$$

where $u^r_{n,i,k}(t) = \xi^r_{k,j}(t/N)\,\kappa_{n,i}$ are the expanded terms, $\bar{x}^{\upsilon}$ and $\left(\bar{u}^r_{n,i,k}\right)^{\upsilon}$ represent weak derivative expressions of the signal and model terms, respectively. Given sampled data with discrete time $t = 1,2,\cdots,N$, the discrete form of $\bar{x}^{\upsilon}$ and $\left(\bar{u}^r_{n,i,k}\right)^{\upsilon}$ can be denoted as $\bar{x}^{\upsilon}(p) = \sum_{t=p}^{p+n_0} x(t)\bar{\omega}^{(\upsilon)}(t-p)$ and $\left(\bar{u}^r_{n,i,k}\right)^{\upsilon}(p) = \sum_{t=p}^{p+n_0} u^r_{n,i,k}(t)\bar{\omega}^{(\upsilon)}(t-p)$, where $\bar{\omega}^{(\upsilon)}(t)$ is the υ-th derivative of a normalized test function, n_0 is the support of the test function and $p = 1,2,\cdots,N-n_0$. This study uses the $(d+1)$-th order B-splines which have finite support and continuous d-th order derivatives as the test functions. The detailed deducing process of (7) is illustrated in Supplementary material-A.

As presented in (7), the loss function of ULS contains two parts: the first part is the standard least squares criterion which focuses on the similarity over the whole data, while the second part describes the identity of the weak derivatives which essentially emphases the unity in shape. The second part, which fully takes into account the agreement in shape of signals, makes this new criterion different to most traditional methods for model structure detection. Any detailed difference in the distribution can be characterized in the second part of the new cost function (7). Thus, the criterion J_{ULS} is a more effective metric for model identification than the conventional least squares criterion. The regression problem (6) can then be converted to solve a new ULS problem, and the matrix form can be represented as

$$X_{ULS} = \Phi_{ULS}\Theta + e \tag{8}$$

where

$$X_{ULS} = \left[x(1),\cdots,x(N),\bar{x}^1(1),\cdots,\bar{x}^d(N-n_0)\right]^T \tag{9}$$

$$\Phi_{ULS} = \begin{bmatrix} u^r_{1,i,k}(1) & \cdots & u^r_{V,i,k}(1) \\ \vdots & \cdots & \vdots \\ u^r_{1,i,k}(N) & \cdots & u^r_{V,i,k}(N) \\ \left(\bar{u}^r_{1,i,k}\right)^1(1) & \cdots & \left(\bar{u}^r_{V,i,k}\right)^1(1) \\ \vdots & \cdots & \vdots \\ \left(\bar{u}^r_{1,i,k}\right)^d(N) & \cdots & \left(\bar{u}^r_{V,i,k}\right)^d(N) \end{bmatrix} \tag{10}$$

$$\Theta = \left[\beta^r_{1,i,k},\cdots,\beta^r_{V,i,k}\right]^T \tag{11}$$

and e is the approximation error vector.

Now the TVARX model (3) is transformed into another problem of constructing model (8), which can be solved by means of a model structure detection method such as the well-known OLS (or OFR) algorithm [41, 42]. Although the OLS has proven to be an efficient procedure for model construction and refinement, the use of the parsimonious principle alone cannot entirely avoid overfitting since small-sized models constructed may still fit to the noise when the systems are highly noisy [29]. In addition, the ULS criterion may ignore the interference of overlapping information that might lead to an ill-conditioned problem during the forward regression process [43]. In order to alleviate such a dilemma, an effective zero-order ROLS ($ROLS^0$) technique combining the zero-order regularization with the OLS [28, 29] is used in this work, where a sparse model structure with good generalization performance and low computational cost can be constructed.

In (8), X_{ULS} is a vector of system outputs and Φ_{ULS} is a matrix formed by candidate terms (regressors). Denote all the candidate bases by a dictionary $D = \{\zeta_{n,i,m}: n = 1,\cdots V; i = 1,\cdots I; m = 1,\cdots M\}$, where $\zeta_{n,i,m}(t) = u^r_{n,i,k}(t)$, and the term selection procedure is to find a full dimensional subset $D_\eta = \{\zeta_{L_\gamma}: \gamma = 1,2,\cdots,\eta; L_\gamma \in \{1,2,\cdots(I_1+I_2+I_3)\times M\}\}$ $(\eta \ll (I_1+I_2+I_3)\times M)$, so that X can be approximated via a linear combination of ζ_{L_γ} as $X = \zeta_{L_1}\pi_{L_1} + \cdots + \zeta_{L_\eta}\pi_{L_\eta} + e$ or in a compact matrix form $X = \Upsilon\Pi + e$, where the regression matrix $\Upsilon = \left[\zeta_{L_1},\zeta_{L_2},\cdots,\zeta_{L_\eta}\right]$, and $\Pi = \left[\pi_{L_1},\pi_{L_2},\cdots,\pi_{L_\eta}\right]^T$ is the associated parameter vector. Assume that the matrix Φ_{ULS} is full column rank and can be orthogonally decomposed as $\Phi_{ULS} = H\Lambda$, where $H = [h_1,h_2,\cdots h_{\mathcal{M}}]$ is a matrix with $\mathcal{M}$ orthogonal columns satisfying $h_i^T h_j = 0$ when $i \neq j$, $\mathcal{M} = (I_1+I_2+I_3)\times M$, and Λ is an upper triangular matrix [28]. The model (8) can then be expressed as $X_{ULS} = H\Lambda\cdot\Theta + e = HW + e$, where $W = [w_1,w_2,\cdots,w_{\mathcal{M}}]^T$ is the orthogonal regression weight vector which satisfies the relation $W = \Lambda\cdot\Theta$.

For constructing a parsimonious model structure, here we further propose an ultra-regularized least squares (URLS) error criterion to improve the generalization capacity

$$J_{URLS} = J_{ULS} + \sum_{m=1}^{\mathcal{M}}\mu w_m^2 = e^T e + \mu W^T W \tag{12}$$

where $\mu \geq 0$ is the regularization parameter, which can be selected by adopting a Bayesian interpretation via an iterative procedure [28]. This study randomly sets an initial value μ (≥ 0), and chooses an appropriate μ value after a number of iterations. Practically the inclusion of the ULS element in the URLS metric can help improve the dynamic detection precision, and the inclusion of the regularization error helps alleviate overfitting and eliminate the overlapping information. The criterion (12) can be normalized by $X_{ULS}^T X_{ULS}$ and further simplified to obtain the zero-order regularized error reduction ratio ($RERR^0$) below [28] to determine significant terms

$$RERR^0(X,h_m) = \frac{\langle X,h_m\rangle^2}{\langle X,X\rangle\left(\langle h_m,h_m\rangle + \mu\right)} \tag{13}$$

where the symbol $\langle\cdot,\cdot\rangle$ denotes the inner product of two vectors. Furthermore, a modified cross-validation criterion named adjustable prediction error sum of squares (APRESS) is integrated into the UROLS algorithm to decide the termination of the term search process [36, 37]

$$APRESS(g) = p(g)\left[\|r_g\|^2 / N\right] \tag{14}$$

where $p(g) = 1/(1-g\varpi/N)^2$ with adjustable parameter $\varpi \geq 1$ is the penalty function, $\|r_g\|^2 = \|X\|^2 - \sum_{\varsigma=1}^{g}\frac{(r_{\varsigma-1}^T h_\varsigma)^2}{h_\varsigma^T h_\varsigma}$, $r_0 = X$ is the residual sum of squares, and $\|r_g\|^2/N$ denotes the mean-squared-errors (MSE) obtained from the associated g-term model. The term selection process is terminated when the APRESS statistic reaches the minimum at $g = \eta$, and yields a η-term model. The effect of ϖ on the results is detailed in [36, 37]. The pseudocode for the UROLS-APRESS and the detailed selection procedure is presented in Supplementary material-B.

As for the model order (lags of system variables) determination problem, a possible solution is to minimize a metric that balances the residual error caused by the model against the number of effective coefficients. In this study, the appropriate model order size is determined by minimizing the Akaike information criterion (AIC) defined as [44]

$$\mathrm{AIC}(\mathrm{i}) = \ln\left(\det\left(\Sigma_{\mathrm{i}}\right)\right) + \frac{2\mathrm{iV}^2}{\mathrm{N}} \tag{15}$$

where Σ_i is the variance of the model residuals calculated from the corresponding i-order model, and V denotes the number of input variables as before.

Finally, the selected regression matrix $\Upsilon = \left[\zeta_{L_1}, \zeta_{L_2}, \cdots, \zeta_{L_\eta}\right]$ can be orthogonally decomposed as $\Upsilon = O_\eta R_\eta$, where O_η is a matrix with η orthogonal columns and R_η is a $\eta \times \eta$ unit upper triangular matrix. The corresponding parameter vector $\Pi = \left[\pi_{L_1}, \pi_{L_2}, \cdots, \pi_{L_\eta}\right]^T$ can be calculated from the formula $R_\eta \Pi = U$, where $U = \left(O_\eta^T O_\eta\right)^{-1} R_\eta^T X$, and the time-varying coefficients in the TVARX model (3) can thus be recovered using the resultant estimates. Similar to (3), other multivariate TVARX processes expressed in (1)-(2) can also be identified by using the proposed multiwavelet-based UROLS-APRESS method.

C. The formulation of TF-CGC analysis

The proposed UROLS method can provide more accurate TVARX models for nonstationary time series with respect to $x(t)$, $y(t)$ and $z(t)$ given in (1)-(2), and this is the most considerable basis of TF-CGC analysis. The formulation of TF-CGC from Y to X conditional on Z denoted as $GC_{Y\to X|Z}(t,f)$ is provided in this section.

In (1), the initial noise terms $e_1(t)$ and $e_2(t)$ can be correlated with each other and their time-varying covariance matrix is $\boldsymbol{\Sigma}_1 = \left[\left(\Sigma_1(t)\ \Delta_1(t)\right)^T, \left(\Delta_1(t)\ \Sigma_2(t)\right)^T\right]^T$, specifically $\Sigma_1(t) = var\left(e_1(t)\right)$, $\Sigma_2(t) = var\left(e_2(t)\right)$ and $\Delta_1(t) = cov\left(e_1(t), e_2(t)\right)$ are calculated using a recursive expression $\sigma^2(t+1) = (1-\rho)\sigma^2(t) + \rho u_1(t) u_2(t)$ with $0 \le \rho < 1$ [18]. Generally, selecting $0.01 \le \rho \le 0.05$ can well balance the adaptation speed and the variance of the estimation when modelling EEG data [32]. Setting $u_1(t) = u_2(t) = e_1(t)$, $u_1(t) = u_2(t) = e_2(t)$, and $u_1(t) = e_1(t), u_2(t) = e_2(t)$, yields time-varying variances and covariance of the corresponding prediction errors $\Sigma_1(t)$, $\Sigma_2(t)$ and $\Delta_1(t)$, respectively. Define the lag operator λ to be $\lambda x(t) = x(t-1)$, then $\lambda^i x(t) = x(t-i)$ and the associated time-varying lag polynomial is $a(\lambda, t) = \sum_{i=0}^{\mathrm{I}} a_i(t)\lambda^i$, where I is the lag index of $a(t)$. Equation (1) can be rewritten as

$$\begin{pmatrix} \mathrm{a}_{11}(\lambda,\mathrm{t}) & \mathrm{a}_{12}(\lambda,\mathrm{t}) \\ \mathrm{a}_{21}(\lambda,\mathrm{t}) & \mathrm{a}_{22}(\lambda,\mathrm{t}) \end{pmatrix} \begin{pmatrix} \mathrm{x}(\mathrm{t}) \\ \mathrm{z}(\mathrm{t}) \end{pmatrix} = \begin{pmatrix} \mathrm{e}_1(\mathrm{t}) \\ \mathrm{e}_2(\mathrm{t}) \end{pmatrix} \tag{16}$$

with $a_{11,0}(t) = a_{22,0}(t) = 1$, $a_{12,0}(t) = a_{21,0}(t) = 0$. The independence of $e_1(t)$ and $e_2(t)$ is necessary for the definition of spectral domain causality [4]. Thus the normalization procedure introduced by Geweke [38] is exploited and developed to remove the correlation and further make the identification of an intrinsic part and a causal part possible in time-varying cases. The transformation consists of left-multiplying $P(t) = [(1\ 0)^T, (-\Delta_1(t)/\Sigma_1(t)\ \ 1)^T]^T$ on both sides of (16) at each time index [45], and the resulting normalized form is given as

$$\begin{pmatrix} \mathrm{A}_{11}(\lambda,\mathrm{t}) & \mathrm{A}_{12}(\lambda,\mathrm{t}) \\ \mathrm{A}_{21}(\lambda,\mathrm{t}) & \mathrm{A}_{22}(\lambda,\mathrm{t}) \end{pmatrix} \begin{pmatrix} \mathrm{x}(\mathrm{t}) \\ \mathrm{z}(\mathrm{t}) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\mathrm{t}) \\ \varepsilon_2(\mathrm{t}) \end{pmatrix} \tag{17}$$

with $A_{11,0}(t) = A_{22,0}(t) = 1$, $A_{12,0}(t) = 0$, $A_{21,0}(t)$ generally not being zero. Now $\mathrm{cov}\left(\varepsilon_1(t), \varepsilon_2(t)\right) = 0$ (the noise terms are independent), and note that $\mathrm{var}\left(\varepsilon_1(t)\right) = \Sigma_1(t)$.

In (2), the time-varying covariance matrix of the noise terms can be estimated by the recursive computation similarly as $\boldsymbol{\Sigma}_1$, and is $\boldsymbol{\Sigma}_2 = \left[\left(\Sigma_{xx}(t), \Sigma_{xy}(t), \Sigma_{xz}(t)\right)^T, \left(\Sigma_{yx}(t), \Sigma_{yy}(t), \Sigma_{yz}(t)\right)^T, \left(\Sigma_{zx}(t), \Sigma_{zy}(t), \Sigma_{zz}(t)\right)^T\right]^T$. The explicit formula of the transformation matrix $(Q(t))$ to normalize TVARX model (2) is given in Supplementary material-C. Then the associated normalized equations for (2) can be expressed as

$$\begin{pmatrix} \mathrm{B}_{11}(\lambda,\mathrm{t}) & \mathrm{B}_{12}(\lambda,\mathrm{t}) & \mathrm{B}_{13}(\lambda,\mathrm{t}) \\ \mathrm{B}_{21}(\lambda,\mathrm{t}) & \mathrm{B}_{22}(\lambda,\mathrm{t}) & \mathrm{B}_{23}(\lambda,\mathrm{t}) \\ \mathrm{B}_{31}(\lambda,\mathrm{t}) & \mathrm{B}_{32}(\lambda,\mathrm{t}) & \mathrm{B}_{33}(\lambda,\mathrm{t}) \end{pmatrix} \begin{pmatrix} \mathrm{x}(\mathrm{t}) \\ \mathrm{y}(\mathrm{t}) \\ \mathrm{z}(\mathrm{t}) \end{pmatrix} = \begin{pmatrix} \varepsilon_3(\mathrm{t}) \\ \varepsilon_4(\mathrm{t}) \\ \varepsilon_5(\mathrm{t}) \end{pmatrix} \tag{18}$$

where the noise terms are now independent to each other, and their time-varying variances are $\tilde{\Sigma}_{xx}(t)$, $\tilde{\Sigma}_{yy}(t)$ and $\tilde{\Sigma}_{zz}(t)$, respectively. According to the following crucial relations of conditional causality in the time and frequency domain derived by Geweke based on the relations of different variances [38]

$$\begin{aligned} \mathrm{GC}_{\mathrm{Y}\to\mathrm{X}|\mathrm{Z}}(\mathrm{t}) &= \mathrm{GC}_{\mathrm{Y}\varepsilon_2\to\varepsilon_1}(\mathrm{t}) \\ \mathrm{GC}_{\mathrm{Y}\to\mathrm{X}|\mathrm{Z}}(\mathrm{f}) &= \mathrm{GC}_{\mathrm{Y}\varepsilon_2\to\varepsilon_1}(\mathrm{f}) \end{aligned} \tag{19}$$

the problem of measuring the time-dependent spectral causal connectivity $GC_{Y\to X|Z}(t,f)$ can be converted to the calculation of the causal influence from $Y\varepsilon_2$ to ε_1, i.e. $GC_{Y\varepsilon_2\to\varepsilon_1}(t,f)$. In order to obtain $GC_{Y\varepsilon_2\to\varepsilon_1}(t,f)$, the variance of ε_1 is next decomposed in the time and frequency domain. Time-frequency transforming both sides of (17) leads to

$$\underbrace{\begin{pmatrix} \mathrm{A}_{11}(\mathrm{t},\mathrm{f}) & \mathrm{A}_{12}(\mathrm{t},\mathrm{f}) \\ \mathrm{A}_{21}(\mathrm{t},\mathrm{f}) & \mathrm{A}_{22}(\mathrm{t},\mathrm{f}) \end{pmatrix}}_{\mathbf{A}(\mathrm{t},\mathrm{f})} \begin{pmatrix} \mathrm{X}(\mathrm{t},\mathrm{f}) \\ \mathrm{Z}(\mathrm{t},\mathrm{f}) \end{pmatrix} = \begin{pmatrix} \mathrm{E}_1(\mathrm{t},\mathrm{f}) \\ \mathrm{E}_2(\mathrm{t},\mathrm{f}) \end{pmatrix} \tag{20}$$

where the components of the coefficient matrix $\boldsymbol{A}(t,f)$ are

$$\mathrm{A}_{11}(\mathrm{t},\mathrm{f}) = 1 - \sum_{\mathrm{i}=1}^{\mathrm{I}_1} \mathrm{A}_{11,\mathrm{i}}(\mathrm{t}) \mathrm{e}^{-\mathrm{j}_0 2\pi \mathrm{i} \mathrm{f}/\mathrm{f}_\mathrm{s}},\ \mathrm{A}_{12}(\mathrm{t},\mathrm{f}) = -\sum_{\mathrm{i}=1}^{\mathrm{I}_2} \mathrm{A}_{12,\mathrm{i}}(\mathrm{t}) \mathrm{e}^{-\mathrm{j}_0 2\pi \mathrm{i} \mathrm{f}/\mathrm{f}_\mathrm{s}}$$

$$\mathrm{A}_{21}(\mathrm{t},\mathrm{f}) = -\sum_{\mathrm{i}=1}^{\mathrm{I}_1} \mathrm{A}_{21,\mathrm{i}}(\mathrm{t}) \mathrm{e}^{-\mathrm{j}_0 2\pi \mathrm{i} \mathrm{f}/\mathrm{f}_\mathrm{s}},\ \mathrm{A}_{22}(\mathrm{t},\mathrm{f}) = 1 - \sum_{\mathrm{i}=1}^{\mathrm{I}_2} \mathrm{A}_{22,\mathrm{i}}(\mathrm{t}) \mathrm{e}^{-\mathrm{j}_0 2\pi \mathrm{i} \mathrm{f}/\mathrm{f}_\mathrm{s}}$$

with $j_0 = \sqrt{-1}$ and f_s being the sampling frequency. Similarly, calculating the time-varying spectral decomposition of (18) and representing it as

$$\underbrace{\begin{pmatrix} \mathrm{B}_{11}(\mathrm{t},\mathrm{f}) & \mathrm{B}_{12}(\mathrm{t},\mathrm{f}) & \mathrm{B}_{13}(\mathrm{t},\mathrm{f}) \\ \mathrm{B}_{21}(\mathrm{t},\mathrm{f}) & \mathrm{B}_{22}(\mathrm{t},\mathrm{f}) & \mathrm{B}_{23}(\mathrm{t},\mathrm{f}) \\ \mathrm{B}_{31}(\mathrm{t},\mathrm{f}) & \mathrm{B}_{32}(\mathrm{t},\mathrm{f}) & \mathrm{B}_{33}(\mathrm{t},\mathrm{f}) \end{pmatrix}}_{\mathbf{B}(\mathrm{t},\mathrm{f})} \begin{pmatrix} \mathrm{X}(\mathrm{t},\mathrm{f}) \\ \mathrm{Y}(\mathrm{t},\mathrm{f}) \\ \mathrm{Z}(\mathrm{t},\mathrm{f}) \end{pmatrix} = \begin{pmatrix} \mathrm{E}_3(\mathrm{t},\mathrm{f}) \\ \mathrm{E}_4(\mathrm{t},\mathrm{f}) \\ \mathrm{E}_5(\mathrm{t},\mathrm{f}) \end{pmatrix} \tag{21}$$

Recasting (20) and (21) into the transfer function format below

$$\begin{pmatrix} \mathrm{X}(\mathrm{t},\mathrm{f}) \\ \mathrm{Z}(\mathrm{t},\mathrm{f}) \end{pmatrix} = \underbrace{\begin{pmatrix} \mathrm{G}_{\mathrm{xx}}(\mathrm{t},\mathrm{f}) & \mathrm{G}_{\mathrm{xz}}(\mathrm{t},\mathrm{f}) \\ \mathrm{G}_{\mathrm{zx}}(\mathrm{t},\mathrm{f}) & \mathrm{G}_{\mathrm{zz}}(\mathrm{t},\mathrm{f}) \end{pmatrix}}_{\mathbf{G}(\mathrm{t},\mathrm{f})} \begin{pmatrix} \mathrm{E}_1(\mathrm{t},\mathrm{f}) \\ \mathrm{E}_2(\mathrm{t},\mathrm{f}) \end{pmatrix} \tag{22}$$

$$\begin{pmatrix} \mathrm{X}(\mathrm{t},\mathrm{f}) \\ \mathrm{Y}(\mathrm{t},\mathrm{f}) \\ \mathrm{Z}(\mathrm{t},\mathrm{f}) \end{pmatrix} = \underbrace{\begin{pmatrix} \mathrm{K}_{\mathrm{xx}}(\mathrm{t},\mathrm{f}) & \mathrm{K}_{\mathrm{xy}}(\mathrm{t},\mathrm{f}) & \mathrm{K}_{\mathrm{xz}}(\mathrm{t},\mathrm{f}) \\ \mathrm{K}_{\mathrm{yx}}(\mathrm{t},\mathrm{f}) & \mathrm{K}_{\mathrm{yy}}(\mathrm{t},\mathrm{f}) & \mathrm{K}_{\mathrm{yz}}(\mathrm{t},\mathrm{f}) \\ \mathrm{K}_{\mathrm{zx}}(\mathrm{t},\mathrm{f}) & \mathrm{K}_{\mathrm{zy}}(\mathrm{t},\mathrm{f}) & \mathrm{K}_{\mathrm{zz}}(\mathrm{t},\mathrm{f}) \end{pmatrix}}_{\mathbf{K}(\mathrm{t},\mathrm{f})} \begin{pmatrix} \mathrm{E}_3(\mathrm{t},\mathrm{f}) \\ \mathrm{E}_4(\mathrm{t},\mathrm{f}) \\ \mathrm{E}_5(\mathrm{t},\mathrm{f}) \end{pmatrix} \tag{23}$$

where the TF transfer function $\boldsymbol{G}(t,f)$ and $\boldsymbol{K}(t,f)$ are the inverse of the normalized coefficient matrix $\boldsymbol{A}(t,f)$ and $\boldsymbol{B}(t,f)$, that is, $\boldsymbol{G}(t,f) = \boldsymbol{A}^{-1}(t,f)$ and $\boldsymbol{K}(t,f) = \boldsymbol{B}^{-1}(t,f)$.

Assuming $X(t,f)$ and $Z(t,f)$ from (22) are identical to that from (23) [45], equations (22) and (23) are combined to yield

$$\begin{pmatrix} E_1(t,f) \\ Y(t,f) \\ E_2(t,f) \end{pmatrix} = \begin{pmatrix} G_{xx}(t,f) & 0 & G_{xz}(t,f) \\ 0 & 1 & 0 \\ G_{zx}(t,f) & 0 & G_{zz}(t,f) \end{pmatrix}^{-1} \times \begin{pmatrix} K_{xx}(t,f) & K_{xy}(t,f) & K_{xz}(t,f) \\ K_{yx}(t,f) & K_{yy}(t,f) & K_{yz}(t,f) \\ K_{zx}(t,f) & K_{zy}(t,f) & K_{zz}(t,f) \end{pmatrix} \begin{pmatrix} E_3(t,f) \\ E_4(t,f) \\ E_5(t,f) \end{pmatrix} = \underbrace{\begin{pmatrix} \Re_{xx}(t,f) & \Re_{xy}(t,f) & \Re_{xz}(t,f) \\ \Re_{yx}(t,f) & \Re_{yy}(t,f) & \Re_{yz}(t,f) \\ \Re_{zx}(t,f) & \Re_{zy}(t,f) & \Re_{zz}(t,f) \end{pmatrix}}_{\mathfrak{R}(t,f)} \begin{pmatrix} E_3(t,f) \\ E_4(t,f) \\ E_5(t,f) \end{pmatrix} \quad (24)$$

where $\mathfrak{R}(t,f) = \boldsymbol{G}^{-1}(t,f)\boldsymbol{K}(t,f)$. The time-dependent spectrum of ε_1, i.e. $E_1(t,f)$, can thus be decomposed into the following three parts based on (24) [45]

$$S_{E_1}(t,f) = \Re_{xx}(t,f)\tilde{\Sigma}_{xx}(t)\Re^*_{xx}(t,f) + \Re_{xy}(t,f)\tilde{\Sigma}_{yy}(t)\Re^*_{xy}(t,f) + \Re_{xz}(t,f)\tilde{\Sigma}_{zz}(t)\Re^*_{xz}(t,f) \quad (25)$$

where the upper script '$*$' denotes complex conjugate and transpose of a matrix, the first term can be regarded as the intrinsic power, and the remaining two terms represent the combined causal relations from Y and ε_2. Hence the causality from $Y\varepsilon_2$ to ε_1, namely the final expression for time-varying spectral causality $GC_{Y\to X|Z}(t,f)$ is

$$GC_{Y\to X|Z}(t,f) = GC_{Y\varepsilon_2\to\varepsilon_1}(t,f) = \ln\frac{\left|S_{E_1}(t,f)\right|}{\left|\Re_{xx}(t,f)\tilde{\Sigma}_{xx}(t)\Re^*_{xx}(t,f)\right|} \quad (26)$$

Note that the spectral function in (26) is a continuous function of frequency f, and can be applied to measure the spectral causality at any desired frequency from 0 up to the Nyquist frequency $f_s/2$. Generally the frequency resolution is not infinite, but relevant to the associated parameter approximations and underlying model order [22]. A hypothesis test is required to determine whether the causal interaction in the stochastic processes is significant. The thresholds for statistical significance are computed from surrogate data by a permutation procedure under a null hypothesis of no interdependence at the significance level $p < 10^{-6}$.

The new proposed method for TF-CGC decomposition can now be summarized as follows:

1) Set up the multivariate TVARX models (1) and (2), which are to be identified for TF-CGC analysis; expand all the time-varying parameters in each model using multiple B-spline basis functions and construct the corresponding time-invariant regression models (6).

2) Calculate $\{\bar{x}^{\upsilon}\}_{\upsilon=1}^{d}$ and $\left\{\left(\bar{u}^{r}_{n,i,k}\right)^{\upsilon}\right\}_{\upsilon=1}^{d}$ by modulating the output signals and expanded terms with the normalized test functions $\{\bar{\omega}^{(\upsilon)}\}_{\upsilon=1}^{d}$ and then get the ULS problem (8).

3) Select the significant term with the largest $RERR^0$ value as the first term and remove the selected expanded terms from the candidate dictionary; repeat the process and choose the ς-th term by orthogonalizing all remained expanded terms with the $\varsigma-1$ selected terms and calculating the associated $RERR^0$ value, and the term with the largest value is selected.

4) Terminating the term search process via APRESS statistic in (14), and the model orders are decided through AIC in (15).

5) Approximate the coefficients of the selected model terms, and estimate the initial time-varying parameters using formula (5), hence the essential TVARX models for TF-CGC decomposition can now be established.

6) Normalize the bivariate and trivariate TVARX models by $P(t)$ and $Q(t)$ respectively to make the noise variables independent with each other, and calculate the spectrum representation of these normalized models.

7) Achieve the calculation of TF-CGC according to (25) and (26), and the statistical GC threshold is also estimated to get the significant TF-CGC relations.

III. Simulations And Experiments

In this section, the performance of the proposed multiwavelet -based UROLS-APRESS TF-CGC approach is firstly demonstrated by using two simulation examples with various feature dimensions relative to cortical activities, and the effectiveness is compared with the state-of-the-art methods including the short windowing, adaptive RLS, OLS, UOLS and ROLS algorithms. The proposed method is further applied to real EEG signals at source-level. Specifically, the EEG source waveforms reconstructed at significant MI related cortical regions of interest (ROIs) are studied to detect oscillatory dynamic causal activities in the neocortical sensorimotor network.

A. Simulations and results

1) TF-CGC detection with various causality changes

Consider the following TVARX processes

$$\begin{aligned} x(t) &= 0.59x(t-1) - 0.2x(t-2) + a_1(t)y(t-1) + a_2(t)z(t-1) + e_x(t) \\ y(t) &= 1.58y(t-1) - 0.96y(t-2) + e_y(t) \\ z(t) &= 0.60z(t-1) - 0.91z(t-2) + e_z(t) \end{aligned} \quad (27)$$

where $e_x(t)$, $e_y(t)$, $e_z(t)$ are three independent Gaussian distributed white noises $e_x(t)\sim N(0,\sigma_x^2)$, $e_y(t)\sim N(0,\sigma_y^2)$, $e_z(t)\sim N(0,\sigma_z^2)$ with $\sigma_x^2 = 0.01$, $\sigma_y^2 = 0.01$, $\sigma_z^2 = 0.001$, respecticely. The discrete time index t is set to be equivalent to a sampling rate of 200 Hz, and each process consists of 2000 data points (i.e. $f_s = 200$ Hz, $1 \le t \le 2000$, $1/f_s \le t/f_s \le 10$ s). $a_1(t)$ and $a_2(t)$ are time-varying coupling strengths (CS) of interactions shown in Fig. 1 (a) and given as

$$a_1(t) = e^{-\frac{t/f_s}{6}}\sin(4t/f_s), \quad 1/f_s \le t/f_s \le 10\text{s}, \qquad a_2(t) = \begin{cases} 0.2t/f_s, & 1/f_s \le t/f_s \le 5\text{s}, \\ -0.2t/f_s + 2, & 5 \le t/f_s \le 10\text{s}. \end{cases} \quad (28)$$

Fig. 1(b) is the theoretical values of TF-CGC in this example, which are calculated based on model (27) with true values of $a_1(t)$ and $a_2(t)$ and the TF-CGC formulas. It is obvious that the process $x(t)$ is influenced by $y(t)$ through $a_1(t)$ with fast oscillating strength, and is also interacted by $z(t)$ with continuously increasing intensity in the first half of the process and decreasing intensity in the second half. Note that apart from the non-null causal interactions (i.e. $GC_{Y\to X|Z}(t,f)$ and $GC_{Z\to X|Y}(t,f)$), the measurements for null connections between the processes are not displayed, because no indirect causal relations are involved in this model. This example aims to demonstrate the effectiveness of the proposed method in detecting dynamic causalities with both slow and fast changes in TF domain, and the efficiency to identify indirect influences is separately illustrated in the next example.

The 3~6-th order B-splines (i.e. $\xi^r_{k,j}: r = 3,4,5,6$) with scale index $j = 4$ are used to estimate the oscillating and continuous varying parameters of the model. The output signal and all the expanded terms are modulated with the first and second order derivatives of the cubic B-splines as test functions. Then the UROLS algorithm aided by APRESS is applied to construct the parsimonious model structure and recover the associated parameters. Based upon the identified TVARX models, time-varying causal influences from $y(t)$ and $z(t)$ to $x(t)$ in TF domain are calculated by means of the proposed parametric TF-CGC method. The detected TF-CGC distributions are given

in Fig. 2(f). For comparison, the model in (27) is also estimated by using the following algorithms: the sliding short-window spectral analysis with window length of 50 samples, the RLS with forgetting factor 0.94, the OLS, UOLS and ROLS with B-splines; and the corresponding TF-CGC detection results are shown in Fig. 2(a)-(e), respectively.

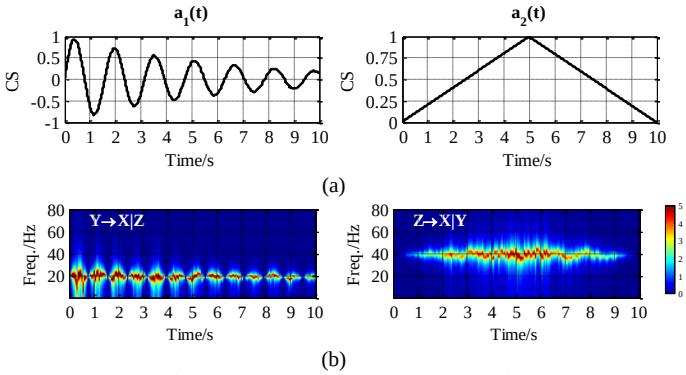


Fig. 1. Description for model (27). (a) The time courses of dynamic coupling strengths (CS). (b) The corresponding theoretical TF-CGC values.

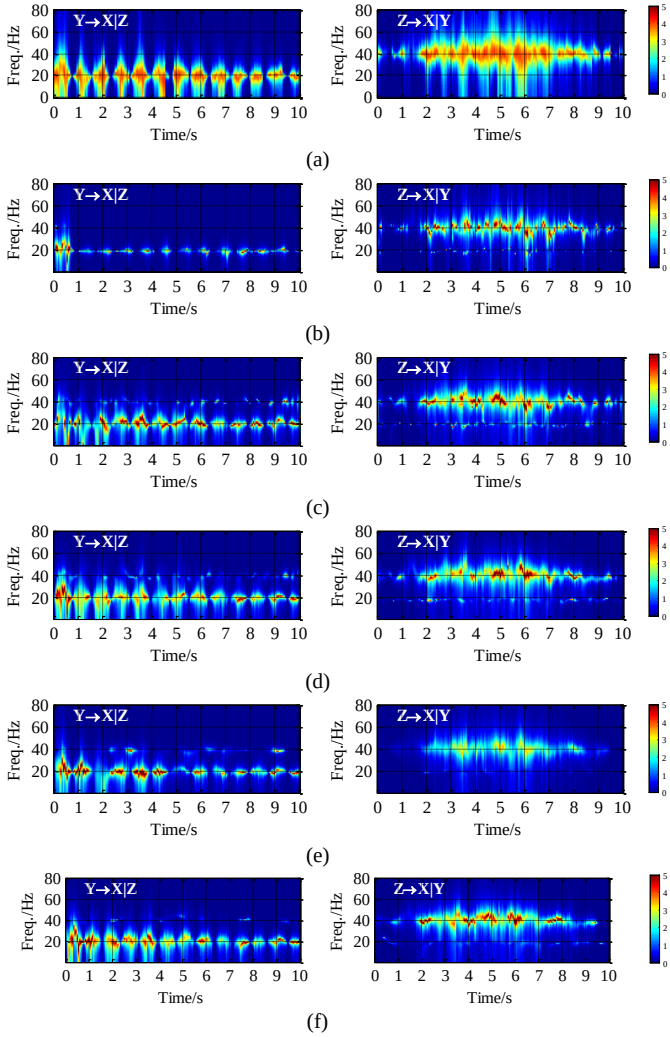


Fig. 2. The comparison of the TF-CGC results for model (27) using different methods. (a) Short-windowing. (b) RLS. (c) OLS with B-splines. (d) UOLS with B-splines. (e) ROLS with B-splines. (f) UROLS-APRESS with B-splines

.

The TF-CGC estimates by the short-window spectral method are presented in Fig. 2(a), where the causality changes along the time course can be roughly measured, but the results of the causal spectral estimation are not very well this indicates the resolution in time and frequency domain cannot be ensured simultaneously. Fig. 2(b) shows that the RLS method reflects monotonous changing interaction from $z(t)$ to $x(t)$ but fails to track oscillatory varying connectivity from $y(t)$ to $x(t)$. Fig. 2(c) indicates that the parametric TF-CGC measure using OLS with B-splines can detect the oscillating as well as ramp-shaped variations in causal influences but also produces false positive values at the wrong frequency without a desirable TF precision. The causal interactions detected by multiwavelet-based UOLS in Fig. 2(d) provides more precise TF causal results than the OLS for both the constructed changing interactions, but there still exist false positive values. Fig. 2(e) gives the causality obtained from ROLS with B-splines. The designed two types of varying influences are reflected in the results with almost no false positive values, but the causalities at some TF points are not detected and the measurements are much smaller than the theoretical values. The results shown in Fig. 2(d) and (e) shows the results produced by the UOLS and ROLS, which can be explained as a result of low specificity (over-fitting) and low sensitivity (under-fitting) in GC detection respectively. In contrast, the proposed method using UROLS with B-splines aided by APRESS (Fig. 2(f)), can better reveal the dynamic interactions containing both fast oscillating and smooth continuous causal variations at almost all time and frequency points with high temporal and spectral precision.

In order to validate the effectiveness of the proposed method quantitatively, the mean absolute error (MAE), root mean squared error (RMSE) and peak signal to noise ratio (PSNR) of the TF-CGC measurements with respect to the corresponding theoretical values are defined below

$$\mathrm{MAE} = \frac{1}{\mathrm{NF}} \sum_{t=1}^{N} \sum_{f=1}^{F} \left| \hat{\mathrm{C}}(t, f) - \mathrm{C}(t, f) \right| \tag{29}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{NF}} \sum_{t=1}^{N} \sum_{f=1}^{F} \left\| \hat{\mathrm{C}}(t, f) - \mathrm{C}(t, f) \right\|^2} \tag{30}$$

$$\mathrm{PSNR} = 20 \log_{10} \left(\mathrm{MAX} / \mathrm{RMSE} \right) \tag{31}$$

where $\hat{C}(t,f)$ is the measurements of TF-CGC $C(t,f)$ at each time and frequency point, N is the data length and F is the frequency range, MAX denotes the maximum strength of the corresponding theoretical GC distribution. The associated results are given in Table I. It is obvious that the calculated MAE, RMSE values of the proposed method are smaller than other five methods and the corresponding PSNR values are the largest one among all approaches, which statistically validate that the proposed scheme possesses better ability for tracking dynamic connectivity in both temporal and spectral domain.

TABLE I
A COMPARISON OF THE DETECTION RESULTS FOR EXAMPLE (1)

TF-CGC	Method	MAE	RMSE	PSNR
$GC_{Y\to X\|Z}(t,f)$	Short-windowing	0.3104	0.6380	22.8912
	RLS	0.3357	0.8212	20.6469
	OLS with B-splines	0.2179	0.5705	23.8114
	UOLS with B-splines	0.1959	0.4440	25.9890
	ROLS with B-splines	0.1816	0.4997	24.9615
	UROLS with B-splines	**0.1642**	**0.4269**	**26.3295**
$GC_{Z\to X\|Y}(t,f)$	Short-windowing	0.3494	0.5565	19.7380
	RLS	0.2057	0.4541	21.5172
	OLS with B-splines	0.1936	0.4209	22.1777
	UOLS with B-splines	0.1739	0.3765	23.1451
	ROLS with B-splines	0.1630	0.4016	22.5839
	UROLS with B-splines	**0.1578**	**0.3438**	**23.9337**

Note: bold values indicate the best results.

TABLE II
ANOVA F TEST FOR EXAMPLE (1)

TF-CGC	Method	Ave MAE±Std	p_1	F	p_2
$GC_{Y\to X\|Z}(t,f)$	Short-windowing	0.3033±0.0176	<0.01	105.15	<0.01
	RLS	0.3344±0.0069	<0.01		
	OLS with B-splines	0.2508±0.0192	<0.01		
	UOLS with B-splines	0.2400±0.0229	<0.01		
	ROLS with B-splines	0.2381±0.0168	0.047		
	UROLS with B-splines	**0.2263±0.0229**	/		
$GC_{Z\to X\|Y}(t,f)$	Short-windowing	0.3437±0.0218	<0.01	118.80	<0.01
	RLS	0.2845±0.0132	<0.01		
	OLS with B-splines	0.2514±0.0170	<0.01		
	UOLS with B-splines	0.2295±0.0204	0.011		
	ROLS with B-splines	0.2265±0.0254	0.037		
	UROLS with B-splines	**0.2131±0.0201**	/		

Note: p_1 refers to p-values obtained from the pairwise comparison between the proposed method and other five methods; p_2 denotes p-values obtained from the comparison between all the six methods; bold values denote the best results.

The efficiency of the proposed method shown in Table I is further confirmed by a statistical Analysis of Variance (ANOVA). Specifically, 20 different datasets, which satisfy the data size requirement for a high level reliability study [46], are generated according to model (27) with random selected noise seeds. Time-frequency causalities can be measured by all six comparing algorithms on each produced dataset, and the estimation results are statistically tested to check the significance of effectiveness of the proposed algorithm. The corresponding analysis results are given in Table II. It can be observed that the ANOVA F-test value between methods are much larger than the threshold value under 99% confidence interval ($p_2 < 0.01$), and the p-values of the proposed method vs. other methods are all less than 0.05 ($p_1 < 0.05$), indicating significant excellent causality detection power of the proposed method.

2) TF-CGC detection with indirect interactions

Consider a three-node network with nonstationary stochastic processes jointly described by the following system

$$\begin{aligned} x(t) &= 0.53x(t-1) - 0.8x(t-2) + e_x(t) \\ y(t) &= 0.53y(t-1) - 0.8y(t-2) + b_1(t)x(t-1) + e_y(t) \\ z(t) &= 0.53z(t-1) - 0.8z(t-2) + b_2(t)y(t-1) + e_z(t) \end{aligned} \quad (32)$$

where $e_x(t)$, $e_y(t)$ and $e_z(t)$ are Gaussian distributed noise with zero means and variances 0.01, $b_1(t)$ and $b_2(t)$ are time-varying coupling strengths from $x(t)$ to $y(t)$ and from $y(t)$ to $z(t)$ respectively, and the time index t is assumed to be equivalent to $f_s = 200$ Hz. Setting the coupling strengths vary according to the profiles shown in Fig. 3(a) and given by (32), 20 trials of data with each trial containing 1000 points (i.e. $1 \le t \le 1000$) are produced in this case. Fig. 3(a) also illustrates the diagram of connectivity among the simulated three nodes.

$$b_1(t) = \begin{cases} 0.5, & 1 \le t \le 500, \\ 0, & 501 \le t \le 1000. \end{cases} \quad b_2(t) = \begin{cases} 0, & 1 \le t \le 500, \\ 0.5, & 501 \le t \le 1000. \end{cases} \quad (33)$$

According to Fig. 3(a) and (33), $x(t)$ has a causal relation on $y(t)$ in the first half of the simulation time interval, and $y(t)$ drives $z(t)$ in the second half in turn, moreover the dashed arrow means that $x(t)$ has an indirect effect on $z(t)$ mediated by $y(t)$. The corresponding theoretical values of TF-CGC are shown in Fig. 3(b). In the ideal case, except the piece-wise varying immediate impacts of $x(t)$ to $y(t)$ and $y(t)$ to $z(t)$ with nonzero values, the other influences should be zero. Fig. 4(a) and (b) shows the CGC analysis results in TF domain using the short-window analysis method with a window size of 40 samples and the RLS with forgetting factor 0.90. The TF-CGC results detected by the B-splines based OLS, UOLS, ROLS, and the proposed UROLS-APRESS are displayed in Fig. 4(c)-(f), respectively.

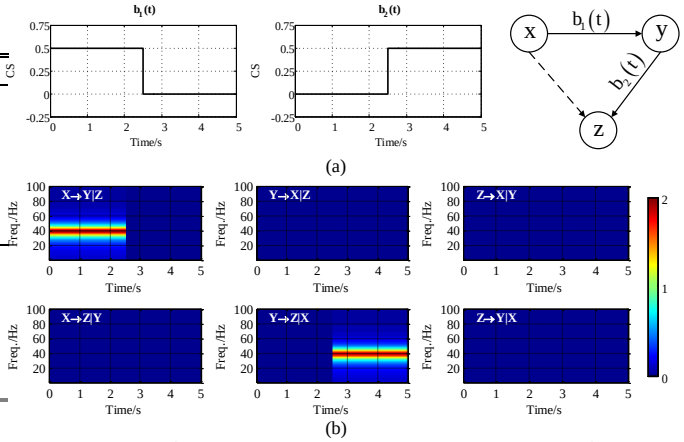
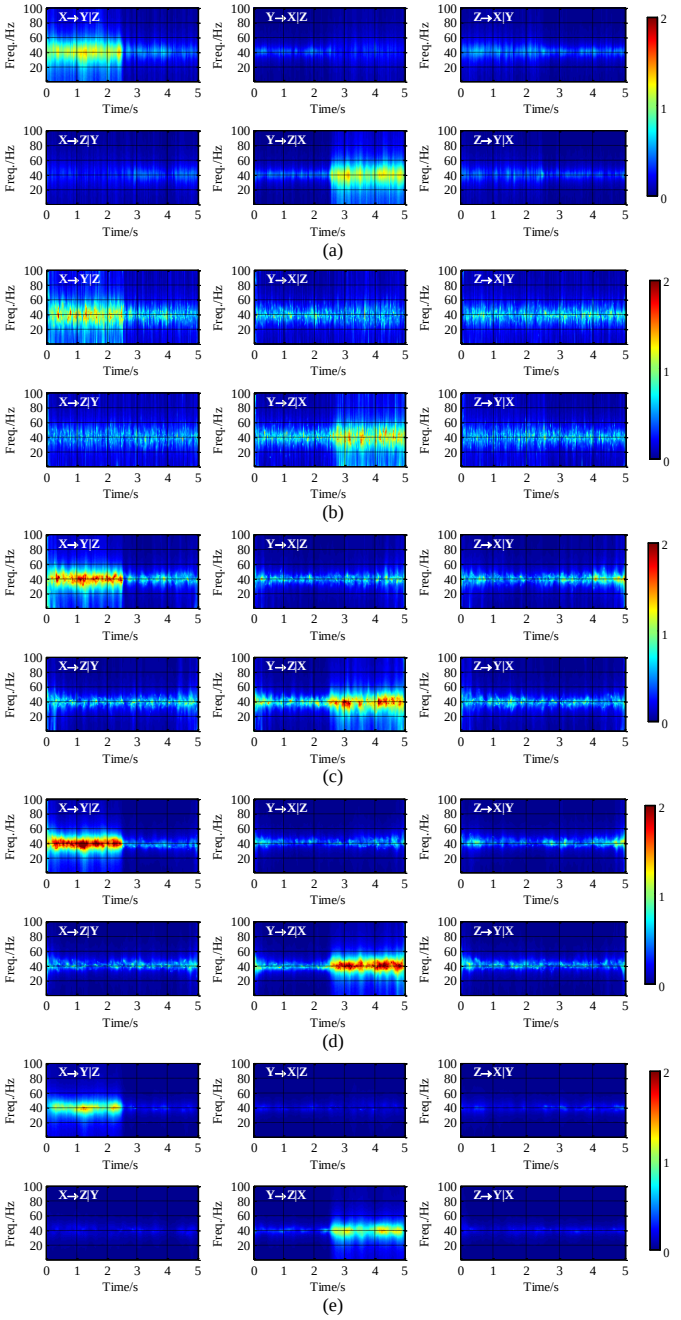
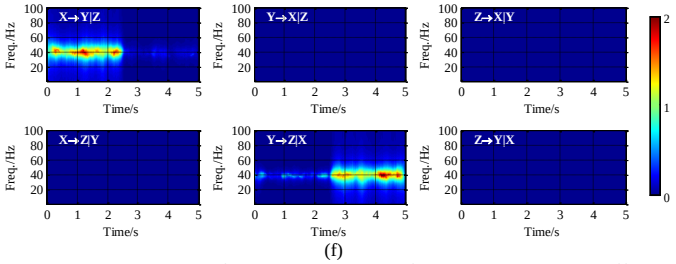


Fig. 3. Description for TVARX systems (32). (a) The time courses of dynamic coupling strengths (CS) and the diagram of interactions. (b) The corresponding theoretical TF-CGC values.





(f)

Fig. 4. The comparison of the TF-CGC results for system (32) using different methods. (a) Short-windowing. (b) RLS. (c) OLS with B-splines. (d) UOLS with B-splines. (e) ROLS with B-splines. (f) UROLS-APRESS with B-splines.

TABLE III
A COMPARISON OF THE DETECTION RESULTS FOR EXAMPLE (2)

TF-CGC	Method	MAE	RMSE	PSNR
$GC_{X\to Y\|Z}(t,f)$	Short-windowing	0.1598	0.2152	19.3630
	RLS	0.2783	0.3431	15.3117
	OLS with B-splines	0.2050	0.2744	17.2535
	UOLS with B-splines	0.1320	0.2012	19.9496
	ROLS with B-splines	0.0933	0.2016	19.9310
	UROLS with B-splines	**0.0743**	**0.1616**	**21.8529**
$GC_{X\to Z\|Y}(t,f)$	Short-windowing	0.1329	0.1649	21.6761
	RLS	0.2511	0.2996	16.4902
	OLS with B-splines	0.1917	0.2738	17.2714
	UOLS with B-splines	0.1253	0.2072	19.6922
	ROLS with B-splines	0.0437	0.0647	29.8067
	UROLS with B-splines	**0**	**0**	**∞**
$GC_{Y\to X\|Z}(t,f)$	Short-windowing	0.1108	0.1463	22.7133
	RLS	0.2402	0.3007	16.4568
	OLS with B-splines	0.1711	0.2600	17.7201
	UOLS with B-splines	0.1074	0.1852	20.6688
	ROLS with B-splines	0.0399	0.0624	30.1166
	UROLS with B-splines	**0**	**0**	**∞**
$GC_{Y\to Z\|X}(t,f)$	Short-windowing	0.1301	0.1948	19.2307
	RLS	0.2744	0.3440	15.2885
	OLS with B-splines	0.2289	0.3188	15.9510
	UOLS with B-splines	0.1600	0.2506	18.0414
	ROLS with B-splines	0.0969	0.2060	19.7442
	UROLS with B-splines	**0.0858**	**0.1853**	**20.6628**
$GC_{Z\to X\|Y}(t,f)$	Short-windowing	0.1344	0.1879	19.5412
	RLS	0.2592	0.3264	15.7450
	OLS with B-splines	0.1882	0.2881	16.8296
	UOLS with B-splines	0.1232	0.2185	19.2335
	ROLS with B-splines	0.0513	0.0833	27.6080
	UROLS with B-splines	**0**	**0**	**∞**
$GC_{Z\to Y\|X}(t,f)$	Short-windowing	0.1112	0.1580	21.0490
	RLS	0.2330	0.2928	16.6892
	OLS with B-splines	0.1688	0.2506	18.0422
	UOLS with B-splines	0.1169	0.1995	20.0212
	ROLS with B-splines	0.0386	0.0643	29.8595
	UROLS with B-splines	**0**	**0**	**∞**

The short-window spectral approach (Fig. 4(a)) gets undesirable frequency distributions with low spectrum resolution for the non-null dynamic interactions ($GC_{X\to Y|Z}(t,f)$ and $GC_{Y\to Z|X}(t,f)$), and the estimated causal values for the null influences ($GC_{X\to Z|Y}(t,f)$, $GC_{Y\to X|Z}(t,f)$, $GC_{Z\to X|Y}(t,f)$, and $GC_{Z\to Y|X}(t,f)$) are not accurate as well. The classical RLS method (Fig. 4(b)) also generates incorrect reflections of the non-zero connectivity in addition to the spurious information leakages [47] among other signal pairs predicted to be zero, which is sensitive to noise and unable to overcome the effect of mutual sources due to the limited convergence speed. This problem is partly solved by the OLS algorithm (Fig. 4(c)), but spurious interactions ascribed to model fitting errors and leakages caused by mutual sources remain apparent. The UOLS method (Fig. 4(d)), although obtains more accurate causalities than the OLS in TF domain, the over-fitting issue still cannot be well solved. The ROLS detection (Fig. 4(e)) alleviates the leakage issue to a negligible level, but the connection strengths from $x(t)$ to $y(t)$ and from $y(t)$ to $z(t)$ are much smaller than the theoretical values due to under fitting in models. By comparison, the proposed UROLS procedure (Fig. 4(f)) can exactly detect the indirect impacts with zero strengths and well reflect the piece-wise variations in causality. All these results show that the proposed approach appears to provide the most desired presentation of the transient causal connectivity.

Similar to the previous example, the MAE, RMSE and PSNR of the TF-CGC estimates are presented in Table III. Obviously, the proposed scheme has a better measuring performance for both abruptly changing direct impacts and indirect influences compared with other five approaches, indicating the advantage of the proposed TF-CGC method in detecting time-varying causality changes for the coupled nonstationary systems.

Table IV gives the ANOVA results of the proposed TF-CGC method and other five compared methods on the 20 single trial datasets. The smaller detection errors and significant p-values indicate that the proposed algorithm can achieve better detection performance than other five methods; this statistically demonstrates that the proposed algorithm can be an effective tool for multivariate time-varying directed connection analysis.

To more explicitly demonstrate the TF resolution and pre-cision of the proposed method in detecting both direct and indirect connectivity, the time and frequency domain causality functions are compared with the mutual information and conditional TE (time domain, see Fig. 5), Geweke spectral causality and dDTF (frequency domain, see Fig. 6) respec-tively. It is obvious that the proposed method outperforms the mutual information (no direction) and TE in terms of higher temporal resolution and more accurate causal strengths. The novel method is also superior to the Geweke spectral method and dDTF in that the proposed method completely eliminates indirect influences and more precise spectrum connection re-lations. These results further validate the efficiency of our proposed method for analyzing dynamic direct and indirect causal interactions in TF domain.

TABLE IV
ANOVA F TEST FOR EXAMPLE (2)

TF-CGC	Method	Ave MAE±Std	p_1	F	p_2
$GC_{X\to Y\|Z}(t,f)$	Short-windowing	0.2121±0.0178	<0.01	470.05	<0.01
	RLS	0.3359±0.0113	<0.01		
	OLS with B-splines	0.2576±0.0284	<0.01		
	UOLS with B-splines	0.1830±0.0192	<0.01		
	ROLS with B-splines	0.1250±0.0098	0.025		
	UROLS with B-splines	**0.1125±0.0098**	/		
$GC_{X\to Z\|Y}(t,f)$	Short-windowing	0.1329±0.0134	<0.01	400.50	<0.01
	RLS	0.2511±0.0214	<0.01		
	OLS with B-splines	0.1917±0.0345	<0.01		
	UOLS with B-splines	0.1253±0.0258	<0.01		
	ROLS with B-splines	0.0437±0.0081	<0.01		
	UROLS with B-splines	**0**	/		
$GC_{Y\to X\|Z}(t,f)$	Short-windowing	0.1108±0.0138	<0.01	328.87	<0.01
	RLS	0.2402±0.0284	<0.01		
	OLS with B-splines	0.1711±0.0351	<0.01		
	UOLS with B-splines	0.1074±0.0197	<0.01		
	ROLS with B-splines	0.0399±0.0067	<0.01		
	UROLS with B-splines	**0**	/		
$GC_{Y\to Z\|X}(t,f)$	Short-windowing	0.1822±0.0106	<0.01	504.65	<0.01
	RLS	0.3308±0.0164	<0.01		
	OLS with B-splines	0.2830±0.0244	<0.01		
	UOLS with B-splines	0.2070±0.0207	<0.01		
	ROLS with B-splines	0.1308±0.0093	0.023		
	UROLS with B-splines	**0.1186±0.0137**	/		
$GC_{Z\to X\|Y}(t,f)$	Short-windowing	0.1344±0.0158	<0.01	437.40	<0.01
	RLS	0.2592±0.0189	<0.01		
	OLS with B-splines	0.1882±0.0357	<0.01		
	UOLS with B-splines	0.1232±0.0202	<0.01		
	ROLS with B-splines	0.0513±0.0089	<0.01		
	UROLS with B-splines	**0**	/		
$GC_{Z\to Y\|X}(t,f)$	Short-windowing	0.1112±0.0107	<0.01	382.44	<0.01
	RLS	0.2330±0.0149	<0.01		
	OLS with B-splines	0.1688±0.0361	<0.01		
	UOLS with B-splines	0.1169±0.0233	<0.01		
	ROLS with B-splines	0.0386±0.0079	<0.01		
	UROLS with B-splines	**0**	/		

Note: bold values indicate the best results

B. Applications to MI-EEG data at source-level

1) EEG data preprocessing

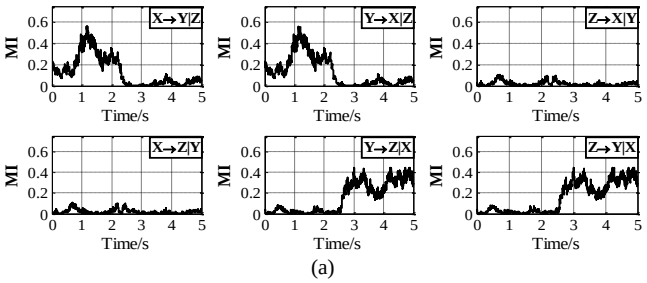
To illustrate applicability of the proposed method for real EEG data connectivity analysis, the proposed method is applied to real MI-EEG source signals. The EEG dataset used in this work is available publicly from the Physiobank Motor/ Mental Imagery database [48], which consists of 109 subjects performing different MI tasks while 64-channel EEGs were recorded based 10-10 systems, sampled at 160 Hz. The blocks that subjects imagined movements of left- and right-hand are selected. Subjects performed a total of 45 trials and imagined one of the two tasks for a duration of 4 s in chosen blocks.

The EEGs were notch filtered to remove 60 Hz AC-line noises. For each trial, the mean of the pre-stimulus samples with duration of 2 s are subtracted for baseline correction, and the stimulus-triggered ensemble average is removed to mitigate the effect of inter-trial variations and the nonstationarity embodied in the mean [49]. Three electrodes (T9, T10 and Iz) are discarded in the following source analysis, since they are spatial outliers relative to the other 61 electrodes which cover the scalp in an approximate uniformly distributed manner [50].

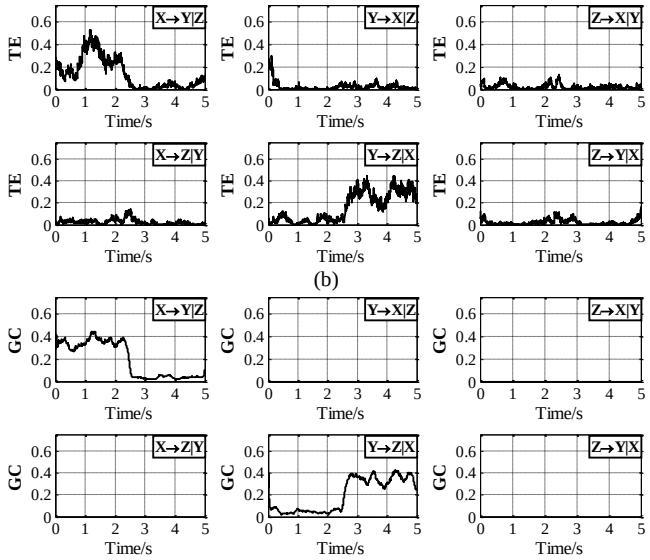
2) EEG source reconstruction

EEG-sources are firstly reconstructed based on event-related potentials (ERPs), and the TF-CGC decomposition is then performed on the estimated single-trial source waveforms to find the directed connectivity patterns in the neocortical sensorimotor networks. The preprocessed 61-channel EEG data from each participant are respectively averaged across trials to arrive at ERPs for left- and right-hand MI of 109 subjects. The 109 ERPs of two MI conditions are used in the exact low resolution electromagnetic tomography (eLORETA) to reconstruct EEG sources on the cortical surface [50]. The computations for inverse solution in eLORETA are implemented in a realistic head model based on the MNI152 (Montreal Neurological Institute) template, with the three -dimensional solution space restricted to cortical gray matter, as determined by the probabilistic Talairach (TAL) atlas. An entire 6239 cortical gray matter voxels with 5 mm spatial resolution constitute the solution space. EEG-source reconstruction at the whole brain level (all 6239 cortical voxels) is calculated, and a voxel by voxel comparison between left- and right-hand MI conditions is performed to determine the source information consisting the cortical sensorimotor network.

According to the results of the statistical comparison, cortical activities at five different sites are significant (with extreme $p < 0.05$ in the t-test), and are selected as the sources. Thus ROIs (see Table V in Supplementary material-D), which formed by the single voxel nearest to the location of these significant areas, are marked as the network nodes for the next connectivity analysis. It shows that the source level network contains four active nodes symmetrically distributed on the left and right sides of the primary somatosensory cortex (BA3.L1, BA3.L2, BA3.R1, BA3.R2) and a limbic area node (BA 23). Note that the influence from limbic system on motor behavior are widespread which can range from the beginning of action to the motivational pace of motor output [51]. Thus, the inclusion of these nodes in the causal network is reasonable and necessary for this MI study.

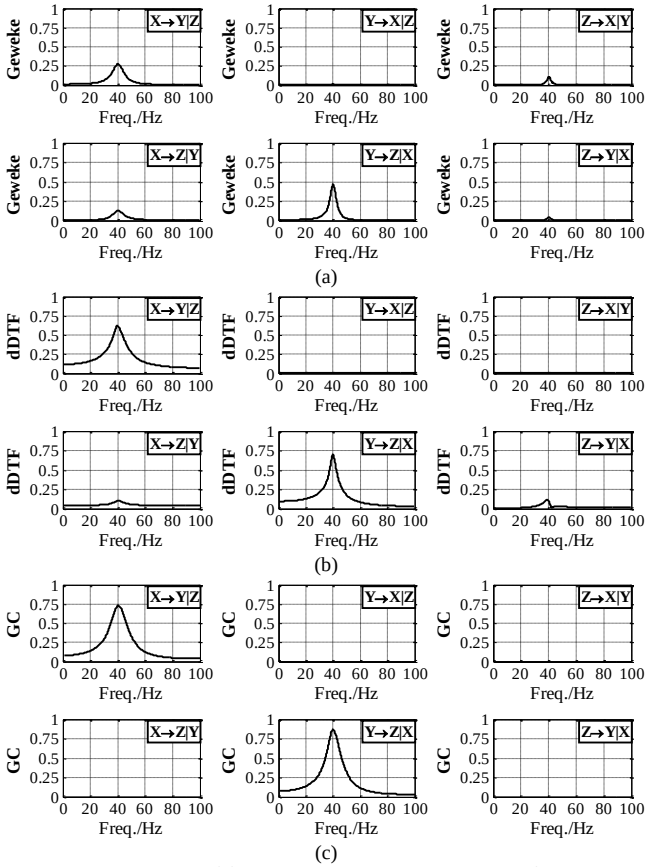


(a)



(b) (c)

Fig. 5. The comparison of time domain causality results for system (32) using different methods. (a) Mutual information. (b) Conditional TE. (c) UROLS-APRESS TF-CGC, obtained by spectral averaging of the corresponding frequency range values.



(a) (b) (c)

Fig. 6. The comparison of frequency domain causality results for system (32) using different methods. (a) Geweke spectral method. (b) dDTF. (c) UROLS-APRESS TF-CGC, obtained by temporal averaging of the corresponding time-varying values.

For convenience, denote ROI 1 as Limbic Lobe or BA23, ROI 2 as Parietal Lobe.L1 or BA3.L1, ROI 3 as Parietal Lobe.R1 or BA3.R1, ROI 4 as Parietal Lobe.L2 or BA3.L2, and ROI 5 as Parietal Lobe.R2 or BA3.R2 in the following analysis. Time series of each trial electric neuronal activity at these five ROIs are estimated with eLORETA from the single-trial EEG data. Based on the obtained single-trial source signal wave-

forms, the TF causal activities under different MI conditions are then analyzed by the TF-CGC method.

3) TF-CGC analysis of MI-EEG source signals

TF-CGCs between the source waveforms of five ROIs for left- and right-hand MI responses are all computed by the proposed method. The TVARX models for EEG source signals from five ROIs are first constructed. Then the 3∼6-th order B-splines with a scale index $j = 4$ are adopted to expand the time-varying parameters, and the cubic B-spline basis functions are employed as the modulate function. For each trial source data, the optimal model order can be determined by minimizing the AIC criterion in (15) with the range of $1 \le i \le 15$ [32]. Similar to [21], the selected model orders are no more than 8 in this study. After identifying the TVARX models applying the proposed UROLS- APRESS algorithm, the TF-CGCs among the five ROIs composed sensorimotor network can be detected by (28) and (29).

As the power density of EEG roughly follows power-law decay as frequency increases [52], spectral causality are typically small in magnitude at higher frequencies. Also considering that the reaction for MI tasks generally happens in the previous seconds after the stimulation, the significant TF-CGCs in 0-40 Hz during 0-2 s (with the stimulus time as 0 s) for left-hand MI are shown in Fig. 7(b), with dashed boxes indexed by 1 and 2 to outline the influences from the regions located in left hemisphere to those in right hemisphere and from right to left respectively, and the performance is compared with the short window method with duration of 200 samples (Fig. 7(a)). The corresponding results for right-hand MI are given in Fig. 8.

From Fig. 7(a) and Fig. 8(a), the short window method can roughly detect the general causal trends between left and right regions under left- and right-hand MI tasks [53], i.e. causalities from the contralateral areas to the ipsilateral are relatively more apparent than those in the opposite direction (mainly see relations in dashed box 1 and 2). However, the temporal and spectral resolutions are not desired, and specific interaction patterns cannot be clearly observed. Additionally, the optimal window size levels obtained are notably high and spread across the whole TF plane with emphasis on low frequency components (lower than alpha band), which may be associated with the residual mutual contents of the MI-EEG source signals [47, 54]. By contrast, the proposed UROLS method can provide more explicit TF causal distributions with a high TF resolution and negligible mutual components in low frequency range.

Specifically, Fig. 7(b) displays the following observations for left-hand MI with the proposed method: (i) the conditional causal influences from right regions to left (dashed box 2) are stronger than that from left to right (dashed box 1) especially after around 0.5 s; (ii) the enhancement of causal relations over the ipsilateral areas ($GC_{BA3.L1 \rightleftarrows BA3.L2}(t,f)$) and the blocking of interactions over the contralateral scalp ($GC_{BA3.R1 \rightleftarrows BA3.R2}(t,f)$) are detected along with the timeframe; and (iii) the causalities out of Limbic Lobe are more obvious than interactions input to it, and Limbic Lobe exerts greater causal influences on left regions (ipsilateral, i.e. BA3.L1 and BA3.L2) than right areas (contralateral, i.e. BA3.R1 and BA3.R2). In addition, the corresponding causal patterns for right-hand conditions (Fig. 8(b)) show that: the causalities from left areas to right (dashed box 1) are more significant than that from right to left (dashed box 2) after approximately 0.25 s; the ipsilateral increase and contralateral decrease are also reflected in strong $GC_{BA3.R1 \rightleftarrows BA3.R2}(t,f)$ and small $GC_{BA3.L1 \rightleftarrows BA3.L2}(t,f)$; and Limbic Lobe outputs more evident causal influences to the ipsilateral sites (BA3.R1 and BA3.R2) than contralateral regions (BA3.L1 and BA3.L2).

Given that alpha (8-14 Hz) and beta (14-30 Hz) bands are the most studied frequency bands when investigating the oscillatory cortical activity during motor operations [54], the dynamic causal interactions in 8-30 Hz are particularly studied to estimate network connectivity patterns during MI tasks. By setting the five ROIs as the sensorimotor network nodes, the net causal flows are computed using the formula $CF_{node} = \sum_{i_c=1}^{N_n}\left(G_{node \rightarrow i_c} - G_{i_c \rightarrow node}\right)$, where N_n is the total number of nodes in a network and G is the 8-30 Hz integrated Granger causality, with self-causality assumed to be zero [52]. The positive CF denotes the net outgoing causal information flow away from the node (causal source), and the negative CF refers to the net incoming flow towards the node (causal sink). The results are presented in Fig. 9(a) and (b) for left- and right-hand MI respectively; besides, Fig. 10(a) and (b) give the CGCs averaged across 8-30 Hz plotted as functions of time for bidirectional influences between left and right regions, aiming to access a high-resolution time response.

Fig. 9(a) shows that in left-hand MI, the ipsilateral regions (BA3.L1 and BA3.L2) function mostly as targets whereas the contralateral sites (BA3.R1 and BA3.R2) become dominant sources with no significant changes over time and frequency. For right-hand MI (Fig. 9(c)), BA3.R1 and BA3.R2 located at ipsilateral areas function predominantly as targets, and BA3.L1 and BA3.L2 in contralateral areas function as sources. Additionally, Limbic Lobe is the prominent source in both left- and right-hand MI conditions within the mainly entire timeframe and frequency range. From Fig. 10 (a) and (b) in, the decrease of band averaged CGC from the ipsilateral regions to contralateral regions and the increase of CGC in the opposite directions are observed in both MI tasks, where the differences between the bidirectional influences enhanced apparently during 0.25-0.45 s in left-hand conditions, while the corresponding discrepancies occurred in 0-0.25 s for right-hand MI. The dynamic brain network representations for TF-CGC are given in Supplementary material-E, and the significant high-resolution causal patterns obtained via the proposed method are concluded in Supplementary material-F.

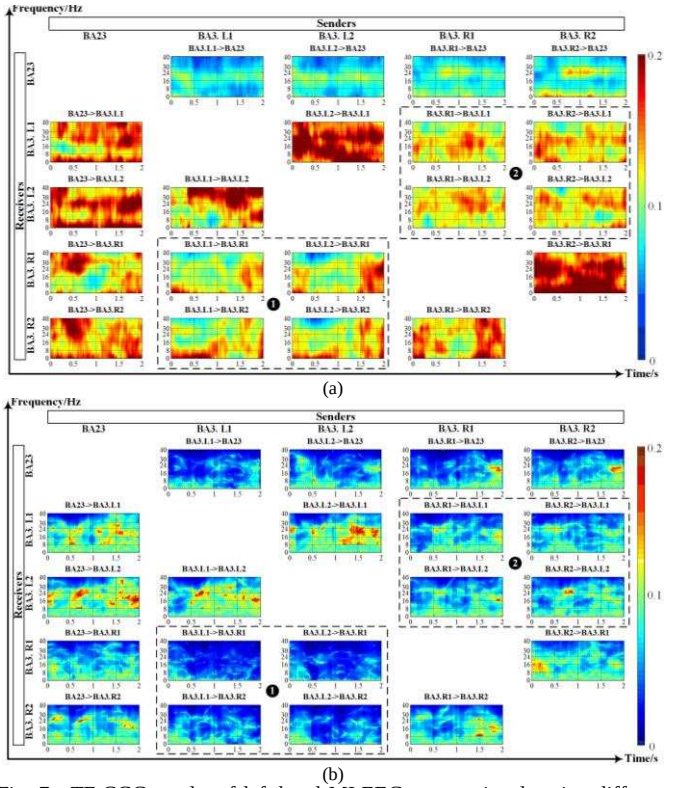


Fig. 7. TF-CGC results of left-hand MI-EEG source signals using different methods. (a) Short-windowing. (b) The proposed UROLS-APRESS.

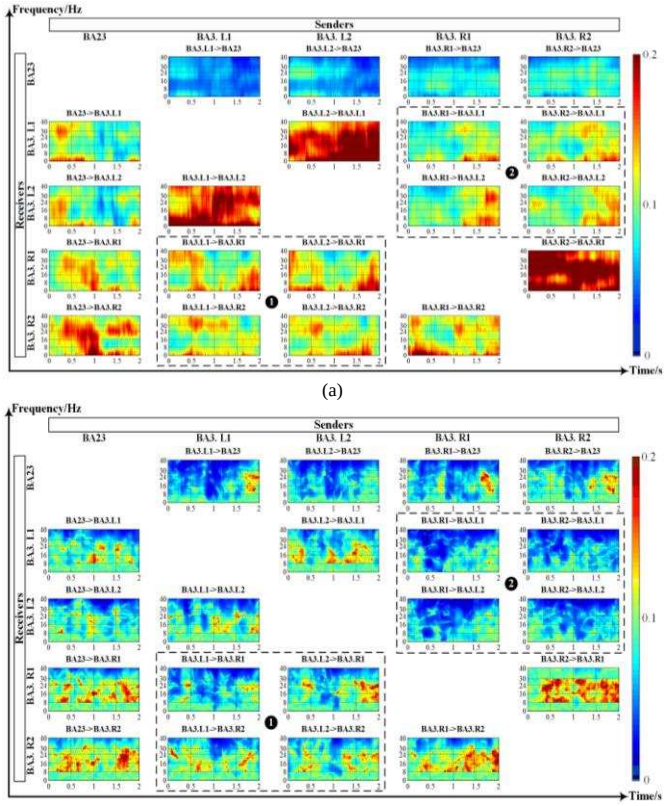


Fig. 8. TF-CGC results of right-hand MI-EEG source signals using different methods. (a) Short-windowing. (b) The proposed UROLS-APRESS.

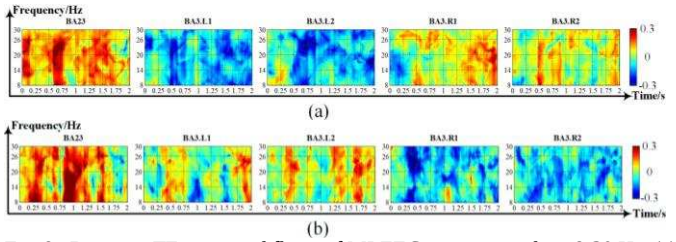


Fig. 9. Dynamic TF net causal flows of MI-EEG source signals in 8-30 Hz. (a) left-hand MI. (b) right-hand MI.

IV. Discussion

A. Significance of results

The key findings of our causality analysis are summarized as follows: (i) the ultra-least squares and APRESS criterion for TVARX model structure construction improve the precision and resolution of dynamic TF-CGC analysis; (ii) the zero-order regularization increases the ability for nonstationary causality detection; (iii) the parametric conditional causal measure based on accurately identified time-varying models can effectively differentiate between direct and indirect interactions; (iv) the precise independent estimation of TVARX models for source EEG signals can alleviate the distorting effect of volume conduction within multichannel EEGs. The proposed method outperforms the conventional short window spectral method, the adaptive RLS, and the other parametric causality approaches based on OLS, UOLS and ROLS in both simulations and real source EEG data analysis. The detection results in time and frequency domain are also superior to the Geweke spectral causality, dDTF, mutual information and TE with a higher time-frequency resolution and precision for both direct and indirect influences. An advantage of the proposed algorithm over the sliding short-time window or adaptive recursive approaches is that it does not need to assume window lengths or stochastic model types for underlying TVARX models, and the dynamic

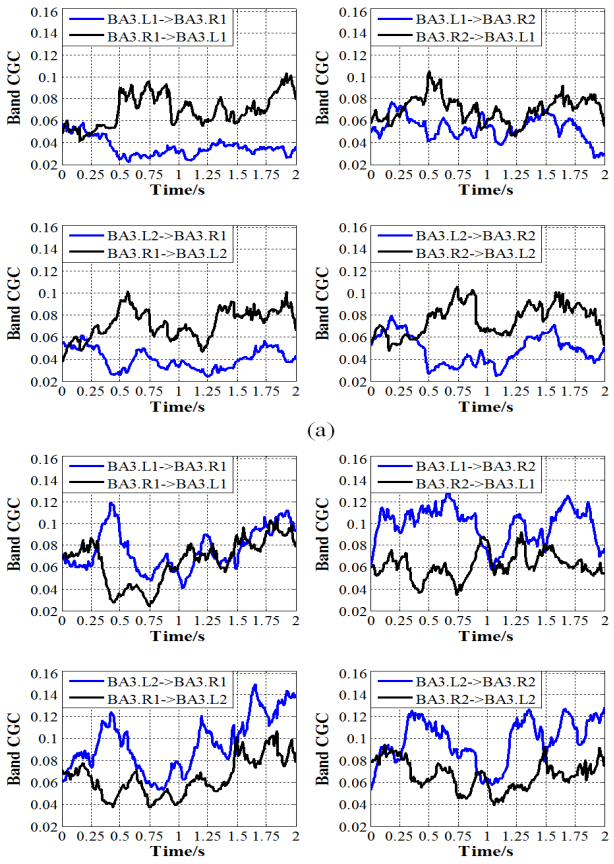


Fig. 10. Averaged time-varying CGCs of MI- EEG source signals in 8-30 Hz. (a) left-hand MI. (b) right-hand MI.

causalities can be calculated by just specifying a number of basis functions that can be used to estimate TVARX models. Furthermore, in comparison with similar existing functional series expansion methods (i.e. multiwavelet-based OLS, UOLS, ROLS), the proposed approach can produce more accurate causal patterns by making use of regularized extra information characterized by the regularization and weak derivatives of nonstationary signals.

B. Efficacy of the ULS metric

As demonstrated in simulations (see panel (d) UOLS vs. (c) OLS, and (f) UROLS vs. (e) ROLS in Fig. 2 and Fig. 4), the ultra-least squares metric for dynamic causality analysis improves the TF accuracy in detecting various time-varying interaction processes. The metric is more efficient than conventional OLS in that it evaluates not only the classical dependent relation of the desired signal on the potential explanatory variables, but also takes advantage of the dependent relation of the associated weak derivatives. The inclusion of weak derivatives, which considers the relative relations between data points of signals (emphasizes the agreement in signal shape [30]), can construct more accurate TVARX models for time-varying systems and further enhance the efficiency to track rapidly and sharply changing TF causalities especially when the signals are not persistently excited. The proposed method is thus more suitable for high-resolution time-frequency connectivity analysis of inherently nonstationary coupled systems.

C. Efficacy of regularization

From the results of simulations (see panel (e) ROLS vs. (c) OLS, and (f) UROLS vs. (d) UOLS in Fig. 2 and Fig. 4), the regularization operation increases the noise-resistibility and robustness of the method for causality detection. Incorporating the zero-order regularization into the parsimonious principle of the OFR, i.e. the regularized ERR criterion, can reflect the potential overlapped information in candidate regressors that ul-

timately helps to precisely identify the TVARX processes with improved generalization performance. Therefore, the obtained causality distributions are of good robustness, enabling the regularized method to capture dynamic causal relations well even when the signals data are contaminated with high noise.

D. Efficacy of the APRESS criterion

For the novel UROLS algorithm used in the TF-CGC analysis for TVARX model identification, the APRESS criterion works more effectively for model term selection. This newly incorporated criterion modifies generalized cross-validation criterion LOO (leave-one -out) by using an adjustable penalty function, thus is more applicable in model determination for nonstationary systems. The redundant model terms confused by the traditional OLS type algorithms become less significant under the new criterion and can then be excluded from the original regressor model. This enables the proposed method to exactly distinguish indirect interactions and obtain completely true null values over the whole TF plane for them.

E. The effect of volume conduction

As illustrated in the TF-CGC results of MI-EEG source signals (Figs. 7-8), the proposed UROLS-APRESS method has the potential to attenuate spurious correlations caused by the volume conduction effects in causality analysis of EEG data. The effect of volume conduction is a significant challenge in connectivity analysis for scalp EEG where a given brain source is often reflected in several EEG signals, and consequently, their similarity may be falsely perceived as connectivity by the analysis procedures [55]. Two primary approaches have been suggested to deal with this problem: one is to analyze EEG connectivity at the source level [56]; and the other is to identify information in the correlation structure that is unlikely to be explained by common sources, such as the partial coherence and phase lag index methods [55]. Therefore, based on these approaches, this paper first calculates single-trial source signals from scalp EEGs, and then estimates accurate time-varying EEG models using the proposed multiwavelet-based UROLS-APRESS algorithm, which can efficiently avoid spurious coupling relations due to similarity of signals. Finally source-level CGCs are measured precisely using the accurate EEG models, and thus the volume conduction issue can be well solved. Specifically, compared to the short window plots (Fig. 7(a), Fig. 8(a)) which show strikingly high levels towards both higher and lower frequency components, the experiment results by the proposed method (Fig. 7(b), Fig. 8(b)) show the dominant contents at around 8-30 Hz, which indicates the incorrect spectral components which mainly represent the volume conduction effects [47] are effectively reduced through the accurate independent approximation for source data. In conclusion, our proposed method can mitigate the volume conduction impact by source level conditional causality analysis with high-precision TVARX models for nonstationary EEG signals.

F. Limitations and future directions

Though achieving good dynamic TF causality detection performance, the proposed UROLS-APRESS method still possesses two major limitations. The first is its heavy computational load, which may be much higher than traditional dynamic connectivity analysis methods. The main reason is that, when identifying each underlying TVARX model, the multiwavelet-based ultra-regularized algorithm involves the iterative selection procedure from a number of expanded candidate terms, and the number of simultaneous equations to be calculated are also increased. Another main limitation is that the TF-CGC method is developed only based on the linear TVARX models and the nonlinear TF connectivity among nonstationary systems is not considered in this early stage. Similar to [57], our previous work has also studied time-varying nonlinear Granger causality detection method which was applied that to MI-EEG signals in the time domain [21]. However, it should be noted that the dynamic nonlinear causality analysis in time-frequency domain can be much more complicated than that only in the time domain, especially for the nonlinear ARX (NARX) model based analysis. Our future work would focus on improving the performance of the proposed method from the following two aspects. One is to extend the TVARX model to a nonlinear case (e.g. TV-NARX model), and another is to develop and adapt fast algorithms for model term selection which can either reduce the computational load or significantly improve the interative model selection procedure.

V. Conclusion

A new parametric TF-CGC method is proposed for multivariate time-varying connectivity analysis in TF domain, where the UROLS-APRESS with multiwavelets is employed in generalized spectral CGC measure to achieve a high-resolution causality detection. Analyses on the simulation data show that the proposed approach can well detect both rapidly varying direct causalities and indirect effects among coupling systems over time and frequency. For real source MI-EEG data, the obtained connectivity patterns are physiologically and anatomically interpretable, and yield important insights into the dynamical organization of 8-30 Hz cortical activities. An obvious advantage of the proposed method lies in its ability to track fast changing causal influences and eliminate indirect effects caused by mutual sources; these are mainly attributed to the use of UROLS-APRESS algorithm. The novel real application of the TF-CGC analysis to EEG signals can provide quantified and more detailed information of the underlying dynamic activities in oscillatory networks. Thus, the procedure that, how oscillating networks coordinate activity between neocortical regions mediating sensory processing to arrive at motor perceptual decisions, can be better understood through this study.

References

[1] G. Buzsáki and A. Draguhn, "Neuronal oscillations in cortical networks," *science,* vol. 304, pp. 1926-1929, 2004.

[2] F. Varela *et al.*, "The brainweb: phase synchronization and large-scale integration," *Nature reviews neuroscience,* vol. 2, p. 229, 2001.

[3] M. Winterhalder *et al.*, "Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems," *Signal processing,* vol. 85, pp. 2137-2160, 2005.

[4] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society,* pp. 424-438, 1969.

[5] E. Siggiridou, V. K. Kimiskidis, and D. Kugiumtzis, "Dimension reduction of frequency-based direct Granger causality measures on short time series," *Journal of neuroscience methods,* vol. 289, pp. 64-74, 2017.

[6] A. Korzeniewska, M. Mańczak, M. Kamiński, K. J. Blinowska, and S. Kasicki, "Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method," *Journal of neuroscience methods,* vol. 125, pp. 195-207, 2003.

[7] C. Siettos and J. Starke, "Multiscale modeling of brain dynamics: from single neurons and networks to mathematical tools," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine,* vol. 8, pp. 438-458, 2016.

[8] C. J. Stam, "Nonlinear dynamical analysis of EEG and MEG: review of an emerging field," *Clinical neurophysiology,* vol. 116, pp. 2266-2301, 2005.

[9] J. Toppi, L. Astolfi, G. R. Poudel, C. R. Innes, F. Babiloni, and R. D. Jones, "Time-varying effective connectivity of the cortical neuroelectric activity associated with behavioural microsleeps," *Neuroimage,* vol. 124, pp. 421-432, 2016.

[10] T. Schäck, M. Muma, M. Feng, C. Guan, and A. M. Zoubir, "Robust Nonlinear Causality Analysis of Nonstationary Multivariate Physiological Time Series," *IEEE Transactions on Biomedical Engineering,* vol. 65, pp. 1213-1225, 2018.

[11] C. Jin, H. Jia, P. Lanka, D. Rangaprakash, L. Li, T. Liu, X. Hu, and G. Deshpande, "Dynamic brain connectivity is a better predictor of PTSD than static connectivity," *Human brain mapping,* vol. 38, pp. 4479-4496, 2017.

[12] M. Dhamala, G. Rangarajan, and M. Ding, "Analyzing information flow in brain networks with nonparametric Granger causality," *Neuroimage,* vol.

41, pp. 354-362, 2008.
[13] M. Dhamala, H. Liang, S. L. Bressler, and M. Ding, "Granger-Geweke causality: Estimation and interpretation," *NeuroImage,* 2018.
[14] N. Leonardi, J. Richiardi, M. Gschwind, S. Simioni, J.-M. Annoni, M. Schluep, P. Vuilleumier, and D. Van De Ville, "Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest," *NeuroImage,* vol. 83, pp. 937-950, 2013.
[15] J. R. Sato, E. A. Junior, D. Y. Takahashi, M. de Maria Felix, M. J. Brammer, and P. A. Morettin, "A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality," *Neuroimage,* vol. 31, pp. 187-196, 2006.
[16] 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 Transactions on Biomedical Engineering,* vol. 60, pp. 3141-3148, 2013.
[17] F. He, H.-L. Wei, S. A. Billings, and P. G. Sarrigiannis, "A nonlinear generalization of spectral granger causality," *IEEE Transactions on Biomedical Engineering,* vol. 61, pp. 1693-1701, 2014.
[18] Y. Li, H. L. Wei, S. A. Billings, and X. F. Liao, "Time-varying linear and nonlinear parametric model for Granger causality analysis," *Physical Review E,* vol. 85, Apr 10 2012.
[19] Y. Zhang, Y. Chen, S. L. Bressler, and M. Ding, "Response preparation and inhibition: the role of the cortical sensorimotor beta rhythm," *Neuroscience,* vol. 156, pp. 238-246, 2008.
[20] M. F. Pagnotta and G. Plomp, "Time-varying MVAR algorithms for directed connectivity analysis: Critical comparison in simulations and benchmark EEG data," *PloS one,* vol. 13, p. e0198846, 2018.
[21] Y. Li, M.-Y. Lei, Y. Guo, Z. Hu, and H.-L. Wei, "Time-varying nonlinear causality detection using regularized orthogonal least squares and multi-wavelets with applications to EEG," *IEEE Access,* vol. 6, pp. 17826-17840, 2018.
[22] Y. Li, Q. Liu, S.-R. Tan, and R. H. M. Chan, "High-resolution time-frequency analysis of EEG signals using multiscale radial basis functions," *Neurocomputing,* vol. 195, pp. 96-103, Jun 26 2016.
[23] S. A. Billings, H.-L. Wei, and M. A. Balikhin, "Generalized multiscale radial basis function networks," *Neural Networks,* vol. 20, pp. 1081-1094, 2007.
[24] 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," *International Journal of Systems Science,* vol. 48, pp. 2658-2668, 2017.
[25] Y. Li, W.-G. Cui, H. Huang, Y.-Z. Guo, K. Li, and T. Tan, "Epileptic seizure detection in EEG signals using sparse multiscale radial basis function networks and the Fisher vector approach," *Knowledge-Based Systems,* vol. 164, pp. 96-106, 2019.
[26] S. A. Billings and H.-L. Wei, "A new class of wavelet networks for nonlinear system identification," *IEEE Transactions on neural networks,* vol. 16, pp. 862-874, 2005.
[27] Y. Li, H.-l. Wei, and S. A. Billings, "Identification of time-varying systems using multi-wavelet basis functions," *IEEE Transactions on Control Systems Technology,* vol. 19, pp. 656-663, 2011.
[28] S. Chen, E. Chng, and K. Alkadhimi, "Regularized orthogonal least squares algorithm for constructing radial basis function networks," *International Journal of Control,* vol. 64, pp. 829-837, 1996.
[29] K. L. Lee and S. A. Billings, "Time series prediction using support vector machines, the orthogonal and the regularized orthogonal least-squares algorithms," *International Journal of Systems Science,* vol. 33, pp. 811-821, 2002.
[30] Y. Guo, L. Guo, S. Billings, and H.-L. Wei, "Ultra-orthogonal forward regression algorithms for the identification of non-linear dynamic systems," *Neurocomputing,* vol. 173, pp. 715-723, 2016.
[31] 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 transactions on neural networks and learning systems,* 2017.
[32] W. Hesse, E. Möller, M. Arnold, and B. Schack, "The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies," *Journal of neuroscience methods,* vol. 124, pp. 27-44, 2003.
[33] 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, pp. 99-120, Jan 2016.
[34] A. K. Seth, A. B. Barrett, and L. Barnett, "Granger causality analysis in neuroscience and neuroimaging," *Journal of Neuroscience,* vol. 35, pp. 3293-3297, 2015.
[35] 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, 2016.
[36] B. Zhang and S. Billings, "Volterra series truncation and kernel estimation of nonlinear systems in the frequency domain," *Mechanical Systems and Signal Processing,* vol. 84, pp. 39-57, 2017.
[37] S. A. Billings and H. L. Wei, "An adaptive orthogonal search algorithm for model subset selection and non-linear system identification," *International Journal Of Control,* vol. 81, pp. 714-724, May 2008.
[38] J. F. Geweke, "Measures of conditional linear dependence and feedback between time series," *Journal of the American Statistical Association,* vol. 79, pp. 907-915, 1984.
[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 Transactions on Neural Networks and Learning Systems,* 2016.
[40] C. K. Chui and J.-z. Wang, "On compactly supported spline wavelets and a duality principle," *Transactions of the American Mathematical Society,* vol. 330, pp. 903-915, 1992.
[41] S. Billings, S. Chen, and M. Korenberg, "Identification of MIMO non-linear systems using a forward-regression orthogonal estimator," *International journal of control,* vol. 49, pp. 2157-2189, 1989.
[42] 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," *Journal of Geophysical Research: Space Physics,* vol. 116, 2011.
[43] S. Chen, "Local regularization assisted orthogonal least squares regression," *Neurocomputing,* vol. 69, pp. 559-585, 2006.
[44] H. Akaike, "A new look at the statistical model identification," *IEEE transactions on automatic control,* vol. 19, pp. 716-723, 1974.
[45] Y. H. Chen, S. L. Bressler, and M. Z. Ding, "Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data," *Journal Of Neuroscience Methods,* vol. 150, pp. 228-237, Jan 30 2006.
[46] S. Walter, M. Eliasziw, and A. Donner, "Sample size and optimal designs for reliability studies," *Statistics in medicine,* vol. 17, pp. 101-110, 1998.
[47] A. H. Omidvarnia, G. Azemi, B. Boashash, J. M. O'Toole, P. B. Colditz, and S. Vanhatalo, "Measuring time-varying information flow in scalp EEG signals: orthogonalized partial directed coherence," *IEEE Trans. Biomed. Engineering,* vol. 61, pp. 680-693, 2014.
[48] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet," *Circulation,* vol. 101, pp. e215-e220, 2000.
[49] 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 cybernetics,* vol. 83, pp. 35-45, 2000.
[50] R. D. Pascual-Marqui, R. J. Biscay, J. Bosch-Bayard, D. Lehmann, K. Kochi, T. Kinoshita, N. Yamada, and N. Sadato, "Assessing direct paths of intracortical causal information flow of oscillatory activity with the isolated effective coherence (iCoh)," *Frontiers In Human Neuroscience,* vol. 8, Jun 20 2014.
[51] L. Heimer and G. W. Van Hoesen, "The limbic lobe and its output channels: Implications for emotional functions and adaptive behavior," *Neuroscience And Biobehavioral Reviews,* vol. 30, pp. 126-147, 2006.
[52] B. M. Adhikari, K. Sathian, C. M. Epstein, B. Lamichhane, and M. Dhamala, "Oscillatory activity in neocortical networks during tactile discrimination near the limit of spatial acuity," *Neuroimage,* vol. 91, pp. 300-310, 2014.
[53] C. Grefkes, S. B. Eickhoff, D. A. Nowak, M. Dafotakis, and G. R. Fink, "Dynamic intra- and interhemispheric interactions during unilateral and bilateral hand movements assessed with fMRI and DCM," *Neuroimage,* vol. 41, pp. 1382-1394, Jul 15 2008.
[54] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE,* vol. 89, pp. 1123-1134, 2001.
[55] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Human brain mapping,* vol. 28, pp. 1178-1193, 2007.
[56] M. J. Brookes *et al.*, "Investigating the electrophysiological basis of resting state networks using magnetoencephalography," *Proceedings of the National Academy of Sciences,* p. 201112685, 2011.
[57] D. Marinazzo *et al.*, "Nonlinear connectivity by Granger causality," *Neuroimage,* vol. 58, pp. 330-338, 2011.