Time-Varying System Identification Using an Ultra-Orthogonal Forward Regression and Multiwavelet Basis Functions With Applications to EEG

Yang Li, Wei-Gang Cui, Yu-Zhu Guo, Tingwen Huang, Xiao-Feng Yang, and Hua-Liang Wei

***Abstract*— A new parametric approach is proposed for nonlinear and nonstationary system identification based on a time-varying nonlinear autoregressive with exogenous input (TV-NARX) model. The TV coefficients of the TV-NARX model are expanded using multiwavelet basis functions, and the model is thus transformed into a time-invariant regression problem. An ultra-orthogonal forward regression (UOFR) algorithm aided by mutual information (MI) is designed to identify a parsimonious model structure and estimate the associated model parameters. The UOFR-MI algorithm, which uses not only the observed data themselves but also weak derivatives of the signals, is more powerful in model structure detection. The proposed approach combining the advantages of both the basis function expansion method and the UOFR-MI algorithm is proved to be capable of tracking the change of TV parameters effectively in both numerical simulations and the real EEG data.**

***Index Terms*— B-splines, EEG, mutual information (MI), parameter estimation, time-varying (TV) system identification, ultra-orthogonal forward regression (UOFR).**

I. Introduction

NONSTATIONARY signals or systems are commonly encountered in many areas of science and engineering. Nonstationary system identification is an important and challenging problem and has been drawing significant attentions [1]–[4]. An approach for processing such systems is to establish a time-varying nonlinear autoregressive with exogenous input (TV-NARX) model, which has been shown to work very well for many real-world processes [5]–[7]. A challenging task in TV system identification is to detect parsimonious model structures and estimate the associated TV coefficients based on experimental data [8], [9]. Some methods have been proposed for TV system identification, which can roughly be classified into three groups: multiple modeling [8], adaptive recursive estimation [9], and basis function expansion methods [10], [11].

In the multiple modeling strategy, the input and output data are divided into several intervals or subspaces by a time shifting window. In each of the intervals, the system is considered to be a stationary process. Thus, conventional time-invariant modeling approaches, such as autoregressive moving average with exogenous inputs (ARMAX) and nonlinear ARMAX (NARMAX), can be used to identify an individual model defined in each of the intervals or subspaces [12], [13]. However, the performance of resulted models depends on the window size, and there is no general criterion for selecting a proper window size. This limits the practical application of such approaches. In the adaptive recursive estimation scheme, model parameters are treated as random processes with certain properties [14]. The most popular algorithms for processing TV models include least-mean squares (LMS), recursive least squares (RLS), and Kalman filtering algorithms [9], [15], [16]. These conventional adaptive methods perform reasonably well when system coefficients vary slowly with time. However, when the TV coefficients change rapidly or abruptly, these approaches may fail to track the model parameters because of the slow convergence speed.

Recently, a new class of methods combining basis function expansion and linear regression approaches have been proposed for nonlinear TV system identification, where TV parameters are approximated using some predefined basis functions (e.g., wavelets) [17]–[20]. The implementation of such methods is as follows. First, a basis function expansion approach is used to transform the original TV system model to a time-invariant regression problem. Note that the initially transformed time-invariant regression model may involve a huge number of model elements or basis functions, and, therefore, could be very complicated for practical applications.

Manuscript received August 23, 2016; revised February 27, 2017 and May 18, 2017; accepted May 24, 2017. Date of publication June 22, 2017; date of current version June 21, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61671042 and Grant 61403016, in part by the Beijing Natural Science Foundation under Grant 4172037, in part by the Open Fund Project of Fujian Provincial Key Laboratory in Minjiang University under Grant MJUKF201702, in part by the Engineering and Physical Sciences Research Council, U.K., under Grant EP/I011056/1 and Platform Grant EP/H00453X/1, and in part by the Qatar National Research Fund through the National Priorities Research Program under Grant NPRP 4-1162-1-181. *(Corresponding authors: Wei-Gang Cui; Hua-Liang Wei.)*

Y. Li is with the Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing 100191, China, and also with the Open Fund Project of Fujian Provincial Key Laboratory of IPIC, Minjiang University, Fujian 350108, China (e-mail: liyang@buaa.edu.cn).

W.-G. Cui is with the Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing 100191, China (e-mail: cwg1994@buaa.edu.cn).

Y.-Z. Guo and H.-L. Wei are with the Department of Automatic Control and Systems Engineering, The University of Sheffield, S1 3DJ Sheffield, U.K. (e-mail: yuzhu.guo@sheffield.ac.uk; w.hualiang@sheffield.ac.uk).

T. Huang is with the Department of Mathematics, Texas A&M University at Qatar, Doha 23874, Qatar (e-mail: tingwen.huang@qatar.tamu.edu).

X.-F. Yang is with the School of Mechanics and Civil Engineering, China University of Mining and Technology, Beijing 100083, China (e-mail: xfyang@cumtb.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2017.2709910

So, the second step is to apply a model structure selection algorithm, e.g., the traditional orthogonal forward regression (OFR) algorithm to the transformed time-invariant regression model, to search for a parsimonious and easy-to-use model, which can be represented by a relatively small number of elements or basis functions [21], [22]. In practice, an appropriate choice of the basis functions is important to ensure the identified model performance [11]. For example, the Legendre polynomials can be used for smooth-changing parameters, and Walsh functions are appropriate for piecewise stationary TV parameters [23]. For systems with sharp or fast changing parameters, wavelets are a suitable choice [10], [22], [24]. Several wavelets-based approaches have been proposed for the implementation of a basis function expansion scheme. For example, in [19], a method combining the cardinal B-splines basis function with the LMS and OFR algorithms was proposed for effectively tracking both slow and rapid changes of TV coefficients in TV linear systems. This method was extended to nonlinear cases and obtained promising results for nonlinear TV modeling problems [25].

Although multiple wavelet basis functions combined with the classic OFR algorithm provide a general method for system identification of nonlinear TV systems, the identification of the correct model structure can still be challenging for some systems, especially when the system is either nonpersistently excited or disturbed by color noises. In these cases, the classical OFR algorithm may produce suboptimal model with spurious model terms [26], or an overparameterized model with some possible redundant or insignificant model terms [27]. Subsequently, the estimated model parameters can be biased, and thus, the resultant model may lack generalization ability.

In this paper, a new method for the identification of TV-NARX model is proposed, where TV parameters are approximated using a finite number of multiwavelet basis functions, and the model structure and parameters are determined using a more powerful ultra-OFR (UOFR) algorithm combined with mutual information (MI) [28]–[30]. Multiwavelet function-based approach has been proven efficient for tracking both the overall global trend and transient local changes in signals [10], [19], [24]. The UOFR algorithm is a recently proposed method that outperforms the conventional OFR algorithm in model structure determination in many cases. The UOFR algorithm is more efficient than the OFR in that it evaluates not only the classical-dependent relation of the desired signal on the potential explanatory variables, but also makes use of the dependent relation of the associated weak derivatives [26]. For the new ultra-least squares (ULS) loss function used in the UOFR algorithm, the MI index works more effectively for model term selection [28]. The redundant model terms confused by the traditional OFR algorithm become less significant under the new criterion and can thus be excluded from the model. An obvious advantage of the proposed procedure, which combines the multiwavelet basis function expansion approach and the UOFR-MI method, lies in its ability to track rapidly and even sharply TV processes. The proposed method is thus more suitable for parameter estimation of inherently nonstationary systems. It is also of good robustness, i.e., it can capture the TV characteristics well even when the data are contaminated with color noises. One of the main contributions of this paper is that for the first time, the newly developed UOFR-MI algorithm is introduced to the multiwavelet-based modeling framework for TV system identification. It is expected that the proposed approach can inspire further development of more powerful algorithms for nonlinear TV system identification.

The remainder of this paper is organized as follows. In Section II, the identification methodology is introduced in Sections II-A–II-C: a TV-NARX model implemented using a multiwavelets basis function expansion method in Section II-A, a ULS criterion for the basis function expanded model in Section II-B, and the UOFR-MI algorithm in Section II-C. Three numerical simulations are given to illustrate the effectiveness of the proposed method in Section III. A case study for a real EEG signal identification problem is presented in Section IV. Finally, this paper is summarized in Section V.

II. Methodology

A. TV-NARX Model Identification Using Multiwavelet Basis Functions

Many nonlinear dynamic systems can be represented by the NARMAX model [31] as

$$y(t) = f(y(t-1), \ldots, y(t-n_y), u(t-1), \ldots, u(t-n_u), e(t-1), \ldots, e(t-n_e)) + e(t) \quad (1)$$

where $y(t)$, $u(t)$, and $e(t)$ denote the system output, input, and noise sequences, with maximum lags n_y, n_u, and n_e, respectively; $f(\cdot)$ is a suitable nonlinear function, which is generally unknown for many practical modeling problems. Variant types of model structures can be used to approximate the nonlinear function $f(\cdot)$ in (1), for example, rational models, fuzzy logic-based models, neural networks, and so on. The most commonly used expression is the polynomial regressions, which has been widely used for a diverse range of nonlinear systems [25].

Many application scenarios focus on revealing the deterministic input–output relationship, for such an application the NARX model, as a relatively simple and special case of the NARMAX model, is often employed without considering the MA noise terms. The NARX model has been proven to be capable of capturing nonlinear characteristics of real systems [31]. For many nonstationary systems, the TV-NARX model can often be formulated as a linear-in-the-parameter representation by expanding the nonlinear function $f(\cdot)$ using a number of time-dependent terms [25]

$$y(t) = \varphi^T(t)\theta(t) + e(t) \quad (2)$$

where $\varphi(t)$ contains monomials of lagged output and input terms, θ is the corresponding parameter vector, and $e(t)$ is a zero mean noise sequence. The power-form polynomial

representation of TV-NARX model can be written as

$$\begin{aligned} y(t) &= \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K} c_{p,q}(d_1,\ldots,d_{p+q},t) \\ &\quad\times \prod_{i=1}^{p} y(t-d_i)\prod_{i=p+1}^{p+q} u(t-d_i)+e(t) \\ &= \varphi^T(t)\theta(t)+e(t) \end{aligned} \tag{3}$$

where M is the degree of the nonlinearity, with $p+q=n$, $d_i = 1,2,\ldots K$, and $\sum_{d_1,\ldots,d_{p+q}=1}^{K} \equiv \sum_{d_1=1}^{K}\cdots\sum_{d_{p+q}=1}^{K}$. The vector $\theta(t) = [c_{0,1}(1,t),\ldots, c_{0,1}(K,t), c_{1,0}(1,t),\ldots, c_{M,0}(\overbrace{K,\ldots,K}^{M},t)]^T$ are TV parameters to be determined.

When the TV-NARX model (3) is used to represent a nonlinear TV system, the identification process involves selecting significant terms from a candidate term dictionary and estimating the corresponding parameters. Note that most algorithms developed for time-invariant model identification cannot directly be applied to a TV model identification due to the assumption that the individual model parameters are constants.

In this paper, multiwavelet basis functions are used to approximate the variation of model parameters in the TV-NARX model, and in this way, the identification of TV model is converted to solving a time-invariant regression problem, which can be solved by means of a conventional model structure detection algorithm, such as the OFR algorithm. It follows that wavelet basis functions can easily track rapid parameter variation in TV processes [19], [20], [32].

A wavelet function $\psi(x)$ is a function whose integral value is zero for the integral interval $(-\infty,+\infty)$, that is, $\int_{-\infty}^{+\infty}\psi(x)\mathrm{d}x = 0$. A shifted and dilated version (e.g., with a shift parameter l and dilation scale 2^j) of a basis function $\psi(x)$ is normally denoted as $\psi_{l,j}(x) = 2^{j/2}\psi(2^j x - l)$, $j,l \in Z$.

An arbitrary function $f(x) \in L^2(R)$ can be expressed using the basis functions as

$$f(x) = \sum_{l,j=-\infty}^{\infty} a_{l,j}\psi_{l,j}(x) \tag{4}$$

where parameters $a_{l,j}$ are called expansion coefficients.

From wavelet theory, an arbitrary function $f(x) \in L^2(R)$ can be approximated by the multiresolution wavelet decomposition as

$$f(x) = \sum_{r=1}^{\varsigma}\sum_{l\in\Gamma_r} a^r_{l,j_0}\phi^r_{l,j_0}(x) + \sum_{r=1}^{\varsigma}\sum_{j>j_0}\sum_{l\in\Gamma_r}\beta^r_{l,j}\psi^r_{l,j}(x) \tag{5}$$

where ς is the maximum number of basis sequences, $[\psi^1,\ldots,\psi^\varsigma]$ are a number of wavelets, and $[\phi^1,\ldots,\phi^\varsigma]$ are the associated scaling functions; $\psi^r_{l,j}(x) = 2^{j/2}\psi^r(2^j x - l)$ and $\phi^r_{l,j_0}(x) = 2^{j/2}\phi^r(2^{j_0}x - l)$, with the shift indices $l \in \Gamma_r$, $\Gamma_r = \{l : -r \le l \le 2^j - 1\}$ and wavelet scale j, a^r_{l,j_0} and $\beta^r_{l,j}$ are coefficients of the wavelet decomposition, and j_0 is an integer to represent the coarsest scale level. When the resolution scale level j_0 is sufficiently large, the function in (5) can be approximated by $f(x) = \sum_{r=1}^{\varsigma}\sum_{l\in\Gamma_r} a^r_{l,j}\phi^r_{l,j}(x)$.

This paper proposes using multiple wavelet basis functions to approximate TV parameters in (3) as

$$c_{p,q}(d_1,\ldots,d_{p+q},t) = \sum_r\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q})\varphi^r_{l,j}\left(\frac{t}{N}\right) \tag{6}$$

where $a^r_{p,q,l}$ represent expansion parameters, which are time invariant, $\varphi^r_{l,j}(\cdot)$ are wavelet basis functions, $t = 1,2,\ldots,N$, and N is the number of data used for model estimation. The variable $x = t/N$ is defined in [0, 1].

Cardinal B-splines are an important class of basis functions that can be used to construct multiresolution wavelet decompositions, which enable the operation of the decomposition to be more convenient [20], [33]. The first-order cardinal B-spline is the well-known Haar function defined as

$$B_1(x) = \chi_{[0,1)} = \begin{cases} 1, & x \in [0,1) \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

The explicit formulas for the second-, third-, fourth-, and fifth-order cardinal B-splines $B_2(x)$, $B_3(x)$, $B_4(x)$, and $B_5(x)$ can be found in [34]. Taking the cardinal B-splines as the basis function, $\varphi^r_{l,j}(x)$ can be expressed by the rth-order B-spline B_r as $\varphi^r_{l,j}(x) = 2^{j/2}B_r(2^j x - l)$, where j and l are the dilated and shifted versions of wavelet B_r. Some criteria to determine the value of j are given in detail [35]. Generally, $j = 4$ is an appropriate choice for many applications using cardinal B-splines. If the value of j is higher, more basis functions will be involved in approximating the TV parameters, and this may improve the resolution but would increase the computational cost. Additionally, a practical selection of the wavelets is the third-, fourth-, and fifth-order cardinal B-splines $B_3(x)$, $B_4(x)$, and $B_5(x)$, $\varphi^r_{l,j}, r = 3,4,5$, and the detailed discussion of B-splines properties can be found in [36].

By expanding TV parameters with multiwavelet basis functions, the TV-NARX model (3) becomes

$$\begin{aligned} y(t) &= \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K}\sum_r\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q}) \\ &\quad\times \varphi^r_{l,j}\left(\frac{t}{N}\right)\prod_{i=1}^{p} y(t-d_i)\prod_{i=p+1}^{p+q} u(t-d_i)+e(t) \\ &= \Psi^T(t)\eta + e(t) \end{aligned} \tag{8}$$

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

It can be seen from (8) that the basis function expansion method reduces the TV model (3) to a time-invariant linear-in-the-parameter form. However, the model may not be ready for use not only because of its complexity but also the difficulty to estimate the model parameters due to the large number of candidate expanded terms. Hence, reducing the number of expanded terms and detecting the correct model structure are vital steps in the model identification. A novel algorithm to deal with this problem will be introduced in Section II-B.

B. Ultra-Least Square Method for TV-NARX Model

The linear-in-the-parameter problem (8) can be solved using a least squares type of algorithm by minimizing the associated loss function

$$J_{\mathrm{LS}} = \left\| y - \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K}\sum_{r}\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q})x^r_{p,q,l}(d_1,\ldots,d_{p+q}) \right\|_2^2 \quad (9)$$

where $x^r_{p,q,l}(d_1,\ldots,d_{p+q})(t) = \varphi^r_{l,j}(t/N)\prod_{i=1}^{p} y(t-d_i) \times \prod_{i=p+1}^{p+q} u(t-d_i)$ is an expanded term, and $a^r_{p,q,l}(d_1,\ldots,d_{p+q})$ is the corresponding time-invariant parameter. In (8), the output and regression terms are time-dependent and defined on the Lebesgue space $L^2([0,T])$, where $[0,T]$ is the time span of signals. The task of model identification is to find an optimal model, which minimizes the square of the L^2 norm (9). In fact, the signals in (8) behave as a low-pass filter essentially and are more regular than general L^2 functions. These signals actually can be defined on some subspace of $L^2([0,T])$, for example, the Sobolev space $H^m([0,T]) = W^{m,2}([0,T])$, or more specially, $H^m([0,T]) = \{x(t) \in L^2([0,T]) \mid D^z x \in L^2([0,T]), z = 1,2,\ldots,m\}$, where D^z is the weak derivatives and L^2 is integrable. The weak derivatives $D^z x(t)$ satisfy

$$\int_{[0,T]} x(t)D^z\omega(t)dt = (-1)^z \int_{[0,T]} \omega(t)D^z x(t)dt \quad (10)$$

for any test function $\omega(t) \in C_0^\infty([0,T])$ that is smooth and has compact support on $[0,T]$. As shown in [26], models fitted by means of weak derivatives in the Sobolev space, with a number of test functions, are more effective. In this paper, test functions are used to smooth the observed signals, and it would be desirable for such functions to have a bell shape similar to a Gaussian function. In fact, the commonly used test functions are not necessarily to be infinitely differentiable [26]. The $(m+1)$th-order B-spline functions, which has finite support and continuous mth-order derivatives, will be adopted as the test functions in this paper.

The distance $\|x - \hat{x}\|_2$ defined by the L^2 Lebesgue integral only measures the differences between functions $x(t)$ and $\hat{x}(t)$ on the interval $[0,T]$. Generally, it cannot find the differences at each individual time instance. Hence, the L^2 norm commonly neglects the detail or closeness in shape and only focuses on the similarity as a whole. System identification is to discover unknown rules from the observations essentially. A particularly challenging issue in system identification is how to get the underlying dynamics or a best model when the system is disturbed by color noises or not persistently excited, and thus, many significant characteristics may be hidden in a small amount of data. Any piece of information is crucial to discovering the important rules. When using the L^2 norm as the criterion, some spurious information may not be distinguished easily [26]. The loss function defined in L^2 may not be able to exclude spurious information and effectively make use of most important information for system identification purpose. Hence, a stricter criterion, which can accurately discover more useful rules, needs to be introduced.

In the Sobolev space $H^m([0,T])$, a stricter metric is the H^m norm that is defined as $x_{H^m} = (\sum_{z=1}^{m} \|D^z x\|_2^2)^{1/2}$. Based on this norm, a new criterion for model (8) is defined as

$$\begin{aligned} J_H = &\left\| y - \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K}\sum_{r}\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q})x^r_{p,q,l}(d_1,\ldots,d_{p+q}) \right\|_2^2 \\ &+ \sum_{z=1}^{m}\left\| \left\{ D^z y - \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K}\sum_{r}\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q})D^z x^r_{p,q,l}(d_1,\ldots,d_{p+q}) \right\} \right\|_2^2. \end{aligned} \quad (11)$$

The lost function (11) consists of two parts: the first part is the same as in the standard least squares criterion that emphases the overall agreement between two time series; the second part is the weak derivatives. It is the second part that makes it different to most conventional model detection and selection methods, which only emphasize the agreement in shape of signals. In the new lost function (11), any detailed changes in the distribution can be reflected in the second part. Therefore, the new criterion J_H is more effective for system identification. Based on the H^m norm, the objective becomes to solve a new least squares problem

$$\begin{bmatrix} y \\ D^1 y \\ \vdots \\ D^m y \end{bmatrix} = \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K}\sum_{r}\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q}) \times \begin{bmatrix} x^r_{p,q,l}(d_1,\ldots,d_{p+q}) \\ D^1 x^r_{p,q,l}(d_1,\ldots,d_{p+q}) \\ \vdots \\ D^m x^r_{p,q,l}(d_1,\ldots,d_{p+q}) \end{bmatrix}. \quad (12)$$

This is referred to as the ULS problem [26], which results in a solution that is normally different to that for (8).

Note that (11) and (12) cannot directly be applied to a data-driven system identification problem, and some work needs to be done first. For example, the weak derivatives cannot be obtained from data in a straightforward way, and the contribution of each individual component donated as $\|D^z y - \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K}\sum_{r}\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q}) D^z x^r_{p,q,l}(d_1,\ldots,d_{p+q})\|_2^2$ in (12) may be imbalanced mainly due to the different magnitudes of the weak derivatives of signals; therefore, the effect of the noise in the criterion J_H can be magnified.

The distribution of signals y and $x^r_{p,q,l}(d_1,\ldots,d_{p+q})$ will be introduced to evaluate the contribution of every unknown weak derivative. A way to fairly evaluate the contribution

of each weak derivative is using the distribution information of the target signal y and the expanded candidate model terms $x^r_{p,q,l}(d_1,\ldots,d_{p+q})$. Define the associated distribution of the signal $y(t)$ as $T_y : C_0^\infty([0,T]) \to R$, and $\langle T_y,\omega\rangle = \int_{[0,T]} y(t)\omega(t)dt$ for all $\omega \in C_0^\infty([0,T])$. The weak derivatives are defined by

$$\langle D^z T_y,\omega\rangle = (-1)^z \int_{[0,T]} y(t)\omega^{(z)}(t)dt. \tag{13}$$

Similarly, the distribution corresponding to $x^r_{p,q,l}(d_1,\ldots,d_{p+q})$ is given as

$$\left\langle T_{x^r_{p,q,l}(d_1,\ldots,d_{p+q})},\omega\right\rangle = \int_{[0,T]} x^r_{p,q,l}(d_1,\ldots,d_{p+q})\omega(t)dt. \tag{14}$$

By defining these distributions, the problem now becomes to find the relationship between the distributions T_y and $T_{x^r_{p,q,l}(d_1,\ldots,d_{p+q})}$. Therefore, the ULS problem becomes

$$\begin{bmatrix} y \\ \langle D^1 T_y,\omega\rangle \\ \vdots \\ \langle D^m T_y,\omega\rangle \end{bmatrix} = \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K}\sum_{r}\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q}) \times \begin{bmatrix} x^r_{p,q,l}(d_1,\ldots,d_{p+q}) \\ \left\langle D^1 T_{x^r_{p,q,l}(d_1,\ldots,d_{p+q})},\omega\right\rangle \\ \vdots \\ \left\langle D^m T_{x^r_{p,q,l}(d_1,\ldots,d_{p+q})},\omega\right\rangle \end{bmatrix}. \tag{15}$$

All data used to solve (15) can be obtained by computing the values of the distributions for every test function $\omega \in C_0^\infty([0,T])$. However, it should be pointed out that it may need a huge number or infinite number of functions to form a basis of $C_0^\infty([0,T])$. So the accurate evaluation of these distributions over the whole space $C_0^\infty([0,T])$ is very difficult if not possible. Hence, a tradeoff needs to be considered between the computational efficiency and the incorporation of all the distribution information in the ULS problem.

The weak derivatives of functions based on a test function $\omega(t)$ defined locally reflect the local information of these functions. When the test function is shifted along the time axis, the corresponding weak derivatives of the signal will give the corresponding local information of the function at new positions. Instead of using all test functions in the space $C_0^\infty([0,T])$, test functions $\omega(t)$ defined locally and the shifted versions $\omega(t-\tau)$ will be used in the ULS problem.

For a given test function $\omega(t)$ with a finite support on $[0,T_0]$ $(T_0 < T)$, the distribution $\langle D^z T_y,\omega(t-\tau)\rangle$ can be denoted as

$$\begin{aligned} y^z(\tau) &= \langle D^z T_y,\omega(t-\tau)\rangle \\ &= (-1)^z \int_0^T y(t)\omega^{(z)}(t-\tau)dt, \quad \tau\in[0,T-T_0] \end{aligned} \tag{16}$$

where $y^z(\tau)$ is the convolution of $y(t)$ with the derivative of the test function. Define $g(t)=\omega(-t)$, $g^{(z)}(t)$ can be viewed as a linear filter impulse response and $y^z(\tau)$ is the corresponding output of the input $y(t)$. According to the Leibniz integral rule, interchanging the order of the differentiation and the integral is permitted. Specifically, the integral has the following property:

$$\begin{aligned} \frac{d^z}{dt^z}\int_0^t y(\tau)g(t-\tau)d\tau &= \int_0^t y(\tau)\frac{\partial^z}{\partial t^z}g(t-\tau)d\tau \\ &= \int_0^t y(\tau)\frac{d^z}{dt^z}g(t-\tau)d\tau. \end{aligned} \tag{17}$$

So now the function $y^z(\tau)$ has a new physical interpretation. Similarly, the function $(\varphi^{(r)}_{j,k}x_i)^z(\tau)$ can also be represented by

$$\begin{aligned} \left(x^r_{p,q,l}(d_1,\ldots,d_{p+q})\right)^z(\tau) &= \left\langle D^z T_{x^r_{p,q,l}(d_1,\ldots,d_{p+q})},\omega(t-\tau)\right\rangle \\ &= (-1)^z\int_0^T \left(x^r_{p,q,l}(d_1,\ldots,d_{p+q})\right. \\ &\quad \left.\times(t)\omega^{(z)}(t-\tau)dt. \right. \end{aligned} \tag{18}$$

Then, the ULS problem (15) becomes

$$\begin{bmatrix} y \\ y^1 \\ \vdots \\ y^m \end{bmatrix} = \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K}\sum_{r}\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q}) \begin{bmatrix} x^r_{p,q,l} \\ \left(x^r_{p,q,l}\right)^1 \\ \vdots \\ \left(x^r_{p,q,l}\right)^m \end{bmatrix} \tag{19}$$

where y^z and $(x^r_{p,q,l})^z$ are signals defined by (16) and (18).

Another significant issue that needs to be considered in the ULS problem is that the magnitude of the difference of the derivatives may be much larger than the differences arising from the observations themselves. As a consequence, the value of the criterion could be dominated by the differences arising from the derivatives. A good criterion must be robust to the noise. Some further modifications to the test function and the associated derivatives should be made. In doing so, the test function and its derivatives are normalized as

$$\bar\omega^{(z)} = \frac{\omega^{(z)}}{\|\omega^{(z)}\|_2}, \quad z = 1,2,\ldots,m \tag{20}$$

where all signals generated by (16) have the same weight in the ULS criterion. The normalized functions will be used to replace the function $\omega^{(z)}$ to modulate signals in the ULS problem. Given a test function $\omega(t)$, the ULS criterion of problem (8) becomes

$$\begin{aligned} J_{\text{ULS}} = &\left\| y - \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K}\sum_{r}\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q})x^r_{p,q,l}(d_1,\ldots,d_{p+q})\right\|_2^2 \\ &+ \sum_{z=1}^{m}\left\| \bar y^z - \sum_{n=1}^{M}\sum_{p=0}^{n}\sum_{d_1,\ldots,d_{p+q}=1}^{K}\sum_{r}\sum_{l\in\Gamma_r} a^r_{p,q,l}(d_1,\ldots,d_{p+q})\left(\bar x^r_{p,q,l}(d_1,\ldots,d_{p+q})\right)^z\right\|_2^2 \end{aligned} \tag{21}$$

where

$$\bar{y}^z(\tau) = \int_0^t y(t)\bar{\omega}^{(z)}(t-\tau)dt \tag{22}$$

$$\left(\bar{x}^r_{p,q,l}(d_1,\ldots,d_{p+q})\right)^z = \int_0^t x^r_{p,q,l}(d_1,\ldots,d_{p+q})(t) \times \bar{\omega}^{(z)}(t-\tau)dt. \tag{23}$$

Given sampled data, the discrete-time form of the modulation process can be described as

$$\bar{y}^z(s) = \sum_{n=s}^{s+n_0} y(n)\bar{\omega}^{(z)}(n-s)$$

$$\left(\bar{x}^r_{p,q,l}(d_1,\ldots,d_{p+q})\right)^z(s) = \sum_{n=s}^{s+n_0} x^r_{p,q,l}(d_1,\ldots,d_{p+q})(n) \times \bar{\omega}^{(z)}(n-s) \tag{24}$$

where n_0 is the support of the test function and $s = 1, 2, \ldots, N - n_0$.

The ULS problem can now be written as

$$Y_{\text{ULS}} = \Phi_{\text{ULS}}\Theta \tag{25}$$

where

$$Y_{\text{ULS}} = [y(1),\ldots,y(N),\bar{y}^1(1),\ldots\bar{y}^1(N-n_0),\ldots, \times\ \bar{y}^m(1),\ldots,\bar{y}^m(N-n_0)]^T \tag{26}$$

$$\Phi_{\text{ULS}} = \begin{bmatrix} x^r_{0,1,l}(d_1)(1) & \cdots & x^r_{M,0,l}(d_1,\ldots,d_M)(1) \\ \vdots & \cdots & \vdots \\ x^r_{0,1,l}(d_1)(N) & \cdots & x^r_{M,0,l}(d_1,\ldots,d_M)(N) \\ \left(\bar{x}^r_{0,1,l}(d_1)\right)^1(1) & \cdots & \left(\bar{x}^r_{M,0,l}(d_1,\ldots,d_M)\right)^1(1) \\ \vdots & \cdots & \vdots \\ \left(\bar{x}^r_{0,1,l}(d_1)\right)^m(N) & \cdots & \left(\bar{x}^r_{M,0,l}(d_1,\ldots,d_M)\right)^m(N) \end{bmatrix} \tag{27}$$

$$\Theta = \left[a^r_{0,1,l}(d_1),\ldots,a^r_{M,0,l}(d_1,\cdots,d_M)\right]^{\text{T}}. \tag{28}$$

C. UOFR-MI Algorithm for TV-NARX Model Identification

The identification of the TV-NARX model mainly consists of two key steps: model structure detection (i.e., model reduction and refinement) and parameter estimation. In this paper, we follow [28]–[30] to use an MI to measure nonlinear correlation between two signals (i.e., the system output and the individual candidate model terms), and the MI-based measure is incorporated to a UOFR procedure to lead to a new method called the UOFR with MI (UOFR-MI), which is described in the following in detail.

Given two random discrete variables $x \in \chi$ and $y \in \Upsilon$, with marginal probability mass functions $p(x)$, $p(y)$ and a joint probability mass function $p(x, y)$, the MI $I(x, y)$ is defined as

$$I(x,y) = E\left\{\log\left(\frac{p(x,y)}{p(x)p(y)}\right)\right\} = \sum_{x\in\chi}\sum_{y\in\Upsilon} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right). \tag{29}$$

The MI $I(x, y)$ is the reduction in the uncertainty of y due to some knowledge of x and vice versa [28], [29]. It can be used to measure the amount of the information one variable shares with another.

Consider the identification problem of the model (25), where Y_{ULS} is the system output vector and Φ_{ULS} is a matrix whose columns are the candidate regressor vectors, denoted as $D = \{\phi_1, \phi_2, \ldots, \phi_L\}$. Note that the candidate set D is usually redundant. The modeling problem is to select significant regressors $D_h = \{\pi_1, \pi_2, \ldots, \pi_h\}(h \le L)$ from the library D, so that the output y can be satisfactorily approximated by the linear combination of $\pi_1, \pi_2, \ldots, \pi_h$ as $y = \theta_1\pi_1 + \cdots + \theta_h\pi_h + e$ or the matrix form $y = \Pi\Theta + e$, where the matrix $\Pi = [\pi_1, \pi_2, \ldots, \pi_h]$ is of a full column rank, $\Theta = [\theta_1, \ldots, \theta_h]^T$ is a parameter vector, and e is the approximation error, respectively.

Following [28] and [29], and denoting $r_0 = y$, the selection procedure begins to calculate $\rho_1 = \{\arg\max_{1\le j\le L}\{I(r_0, \phi_j)\}$, where $I(\cdot,\cdot)$ is the MI. The first selected regressor can be determined as $\pi_1 = \phi_{\rho_1}$. Select it as the first associated orthogonal basis $q_1 = \phi_{\rho_1}$, and set $r_1 = r_0 - (r_0^T q_1/q_1^T q_1)q_1$.

In general, the λth selected regressor can be chosen as follows. At the $(\lambda-1)$th step, let $D_{\lambda-1} = \{\pi_1, \pi_2, \ldots, \pi_{\lambda-1}\}$ be the set of regressors selected so far, and it can be transformed into orthogonal bases $q_1, q_2, \ldots, q_{\lambda-1}$ via an orthogonal transformation. The next selected regressor can be determined via

$$q_j^{(\lambda)} = \phi_j - \sum_{k=1}^{\lambda-1}\frac{\phi_j^T q_k}{q_k^T q_k}q_k \tag{30}$$

$$\rho_\lambda = \arg\max_{j\ne\rho_k, 1\le k\le\lambda-1}\left\{I\left(r_{\lambda-1}, q_j^{(\lambda)}\right)\right\} \tag{31}$$

where $\phi_j \in D - D_{\lambda-1}$, and $r_{\lambda-1}$ is the residual vector obtained from the $(\lambda-1)$th step. The λth selected regressor is chosen as $\pi_\lambda = \phi_{\rho_\lambda}$. The residual vector r_λ is defined as

$$r_\lambda = r_{\lambda-1} - \frac{\left(r_{\lambda-1}^T q_\lambda\right)^2}{q_\lambda^T q_\lambda}. \tag{32}$$

All significant regressors can be selected step by step, one at a time. Since vectors r_λ and q_λ are orthogonal, so

$$\|r_\lambda\|^2 = \|r_{\lambda-1}\|^2 - \frac{\left(r_{\lambda-1}^T q_\lambda\right)^2}{q_\lambda^T q_\lambda}. \tag{33}$$

By summing (32) and (33), yields

$$y = \sum_{\lambda=1}^{n}\frac{r_{\lambda-1}^T q_\lambda}{q_\lambda^T q_\lambda}q_\lambda + r_n \tag{34}$$

$$\|r_n\|^2 = \|y\|^2 - \sum_{\lambda=1}^{n}\frac{\left(r_{\lambda-1}^T q_\lambda\right)^2}{q_\lambda^T q_\lambda}. \tag{35}$$

The residual sum of squares $\|r_n\|^2$ can be used to form the model selection criteria. Furthermore, the Akaike Information Criterion (AIC) will be adopted to terminate the UOFR-MI procedure when an appropriate number of model terms are included in the model [37]. A popular version of AIC can be written as

$$\text{AIC}(n) = N\log[\|r_n\|^2/N] + 2n. \tag{36}$$

If the AIC reaches the minimum at $n = h$, the term selection procedure can be terminated, and resulting in an h-term model. The selected regression matrix $P_h = [\pi_1, \pi_2, \ldots, \pi_h]$ will be full rank in columns and it can be orthogonally decomposed as $P = Q_h R_h$, where R_h is an $h \times h$ unit upper triangular matrix and Q_h is an $N \times h$ matrix with orthogonal columns $q_1, q_2, \ldots, q_h$. The corresponding unknown coefficients vector $\Theta = [\theta_1, \theta_2, \ldots, \theta_h]$ in the original nonorthogonalized space can be calculated from the formula $R_h\Theta = G$, where $G = (Q_h^T Q_h)^{-1} R_h^T Y$. In addition, some measures can be applied to avoid selecting model terms, which are strongly correlated. At the λth step, a subset $D_{\lambda-1}$ containing $\lambda-1$ orthogonal bases has been selected. If ϕ_j is strongly correlated with some bases in $D_{\lambda-1}$, it can then be shown that $(q_j^{(\lambda)})^T q_j^{(\lambda)} = 0$ [13], [28], [29]. So the proposed algorithm can automatically discard the candidate basis ϕ_j if $(q_j^{(\lambda)})^T q_j^{(\lambda)} < \delta$, where δ is a predetermined positive number that is closed to 0, for example, $\delta = 10^{-r}$ and $r \geq 10$. In this way, any ill-conditioning or severe multicollinearity can be avoided [13].

The new proposed algorithm for TV-NARX model identification can now be summarized as follows.

1) Set up the TV-NARX model (3) to be identified and expand all TV coefficients of nonlinear terms by using B-spline basis functions, and construct problem (8).
2) Construct a test function $\omega(t)$ and calculate its derivatives $\omega^{(1)}(t), \omega^{(2)}(t), \ldots, \omega^{(m)}(t)$. Normalize the derivatives according to (20) and get the normalized modulating function $\bar{\omega}^{(z)}$.
3) Calculate $\bar{y}^z$s and $(\bar{x}^r_{p,q,l}(d_1, \ldots, d_{p+q}))^z$s by modulating the dependent variables and expanded terms using the normalized function $\bar{\omega}^{(z)}$ according to (24) and then obtain the new problem (25).
4) Calculate the MI between the output vector and expanded terms, and select the significant term with the largest value of the MI as the first term and then remove the expanded terms, which have been selected from the dictionary.
5) At the λth step, orthogonalize all expanded terms still in the dictionary with the $\lambda - 1$ selected terms. Calculate the MI with the residual vector $r_{\lambda-1}$ and select the term with the largest value as the λth selected term.
6) Determine the number of significant model terms using, for example, the AIC or other criterions.
7) Estimate coefficients of the selected model terms, and achieve the estimation of TV parameters using (6).

III. Simulation Examples

Identification of nonlinear TV systems essentially involves the determination of the model structure and the estimation of the corresponding TV parameters [38]. A correct model structure is important for not only reproducing the system behaviors but also revealing the underlying mechanisms. In this section, three numerical simulations are used to illustrate the effectiveness of the new UOFR-MI algorithm in both model structure detection and TV parameter tracking. Specifically, the first example is to illustrate the performance of the UOFR-MI algorithm for the detection of the TV model structure. Then, we test the proposed UOFR-MI, using B-splines as the building blocks (basis functions) to approximate TV parameters of two simulated nonstationary systems, and compared with two conventional methods: RLS and OFR methods. The data in the second example are generated from a known simulated linear TV system. It is demonstrated that the new proposed procedure can identify the system structure accurately and track the true piecewise varying parameter variations very well. Furthermore, the third example involves the modeling problem of a second-order discrete-time TV system with the second-order nonlinearity. The objective here is to illustrate the capability of the proposed procedure for fast tracking and adaptively capturing the abrupt changes of piecewise varying parameters.

A. Example 1: Detection of the Time-Varying System Structure

Consider the nonpersistently excited TV system as

$$\begin{cases} w(t) = a_1(t)y(t-1) + a_2(t)y^2(t-2) \\ \qquad + b_1(t)u^2(t-1) + b_2(t)u(t-2) \\ y(t) = w(t) + \dfrac{1}{1-0.5z^{-1}}e(t) \end{cases} \tag{37}$$

where $e(t) \sim N(0, \sigma^2)$ with $\sigma = 0.05$, and the input signal is generated by an AR process

$$u(t) = \frac{0.25}{1 - 1.6z^{-1} + 0.64z^{-2}}v(t) \tag{38}$$

where $v(t)$ is a Gaussian distributed noise $v(k) \sim N(0, 1)$. TV parameters are given as

$$\begin{aligned} a_1(t) &= \begin{cases} 0.1, & 1 \leq t \leq 500 \\ 0.2, & 501 \leq t \leq 1000 \end{cases} \\ a_2(t) &= -0.05 + 0.02\cos(10\pi t/1000), \quad 1 \leq t \leq 1000 \\ b_1(t) &= \begin{cases} 0.8, & 1 \leq t \leq 400 \\ 1, & 401 \leq t \leq 700 \\ 0.6, & 701 \leq t \leq 1000 \end{cases} \\ b_2(t) &= \begin{cases} -0.6, & 1 \leq t \leq 200 \\ -0.8, & 201 \leq t \leq 800 \\ -0.5, & 801 \leq t \leq 1000. \end{cases} \end{aligned} \tag{39}$$

For system (37), the candidate model inputs are chosen to be the following delayed input and output variables:$\{y(t-1), y(t-2), y(t-3), y(t-4), u(t-1), u(t-2), u(t-3), u(t-4)\}$. Totally, 45 candidate model terms are included in the term dictionary (the nonlinear degree of the polynomial model is chosen to be 2). The third-, fourth-, and fifth-order B-spline functions are used to expand TV parameters. The identification results produced by the classic OFR algorithm are presented in Table I where the correct model terms are denoted in bold. Notice that the OFR algorithm selects a redundant term $y^2(t-1)$. It is also not straightforward to reasonably explain why the correct terms $y^2(t-2)$, $u(t-2)$ are less significant than the redundant term. This probably could be explained that the traditional OFR algorithm may converge to the suboptimal solution when the system is not persistently excited.

TABLE I
SELECTED TERMS BY THE STANDARD OFR ALGORITHM FOR EXAMPLE 1

No.	Terms	$ERR_i \times 100\%$
1	$\sum \varphi^r_{l,j}(\boldsymbol{t/N}) \times \boldsymbol{u^2(t-1)}$	38.3638
2	$\sum \varphi^r_{l,j}(\boldsymbol{t/N}) \times \boldsymbol{y(t-1)}$	28.7975
3	$\sum \varphi^r_{l,j}(t/N) \times y^2(t-1)$	24.1366
4	$\sum \varphi^r_{l,j}(\boldsymbol{t/N}) \times \boldsymbol{y^2(t-2)}$	3.1795
5	$\sum \varphi^r_{l,j}(\boldsymbol{t/N}) \times \boldsymbol{u(t-2)}$	2.3389

Note: terms in bold indicate the correct model terms.

TABLE II
SELECTED TERMS BY THE UOFR-MI ALGORITHM FOR EXAMPLE 1

No.	Terms	$ERR_i \times 100\%$
1	$\sum \varphi^r_{l,j}(\boldsymbol{t/N}) \times \boldsymbol{u^2(t-1)}$	77.9702
2	$\sum \varphi^r_{l,j}(\boldsymbol{t/N}) \times \boldsymbol{y(t-1)}$	7.6505
3	$\sum \varphi^r_{l,j}(\boldsymbol{t/N}) \times \boldsymbol{y^2(t-2)}$	7.1419
4	$\sum \varphi^r_{l,j}(\boldsymbol{t/N}) \times \boldsymbol{u(t-2)}$	4.8754

Note: terms in bold indicate the correct model terms.

The UOFR-MI algorithm is also used to identify the model structure. The output signal and candidate terms are modulated by the first- and second-order derivatives of the cubic B-spline test function. The model identified by the UOFR-MI algorithm is given in Table II. It is obvious that all the system model terms are correctly selected. This indicates that under the ULS criterion, the significance of redundant terms is reduced, and does not appear in the identified model. So the UOFR-MI algorithm correctly detects the correct structure of the TV system.

B. Example 2: Linear Time-Varying Parameter Estimation

Consider the system

$$\begin{cases} w(t) = a_1(t)y(t-1) + a_2(t)y(t-2) \\ \qquad\quad + b_1(t)u(t-1) + b_2(t)u(t-2) \\ y(t) = w(t) + \dfrac{1}{1-0.8z^{-1}}e(t) \end{cases} \tag{40}$$

where $e(t)$ is a Gaussian distributed noise with a variance 0.1, and TV parameters are given as

$$a_1(t) = \begin{cases} 0.32\cos(1.5-\cos(4\pi t/N+\pi)), & 1 \le t \le N/4 \\ 0.32\cos(3-\cos(4\pi t/N+\pi/2)), & N/4+1 \le t \le 3N/4 \\ 0.32\cos(1.5-\cos(4\pi t/N+\pi)), & 3N/4+1 \le t \le N \end{cases}$$

$$a_2(t) = 0.4\cos(4\pi t/N), \quad 1 \le t \le N$$

$$b_1(t) = \begin{cases} 0.65, & 1 \le t \le N/4 \\ -0.5, & N/4+1 \le t \le N/2 \\ 0.65, & N/2+1 \le t \le 3N/4 \\ -0.5, & /4+1 \le t \le N \end{cases}$$

$$b_2(t) = 0.6, \quad 1 \le t \le N \tag{41}$$

where $N = 512$ is the length of data. The input $u(t)$ is a pseudorandom binary sequence.

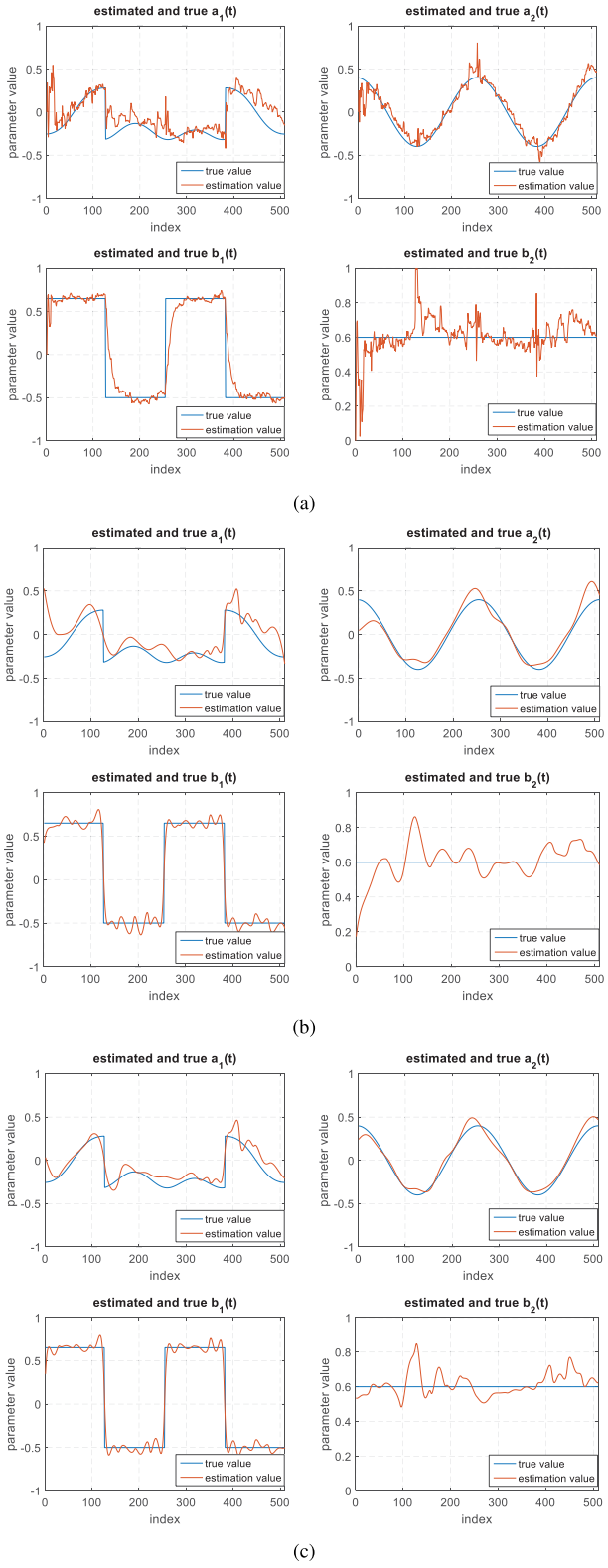


Fig. 1. Identification results of the TVARX (40) using different approaches. Blue curves indicate the true value of TV parameters, and red curves represent the estimates of the parameters. (a) RLS algorithm with forgetting factor $\mu = 0.9$. (b) OFR method with B-splines. (c) UOFR-MI method with B-splines.

The third-, fourth-, and fifth-order cardinal B-splines ($\varphi^r_{l,j}, r = 3, 4, 5$) with scale index $j = 4$ are adopted to approximate piecewise varying coefficients. The output

TABLE III
COMPARISON OF THE ESTIMATION RESULTS FOR EXAMPLE 2 (SNR = 10, 15, AND 20 dB)

Approach	Estimated coefficients	10dB			15dB			20dB		
		MAE	RMSE	Std	MAE	RMSE	Std	MAE	RMSE	Std
RLS ($\mu = 0.90$)	$a_1(t)$	0.2574	2.0201	0.1963	0.1397	1.2336	0.4170	0.0984	1.9865	0.1397
	$a_2(t)$	0.1067	0.3713	0.2024	0.1007	0.4079	0.2309	0.0801	0.3533	0.1892
	$b_1(t)$	0.1160	1.2509	0.1662	0.0902	1.1909	0.1573	0.0747	1.1702	0.0916
	$b_2(t)$	0.1532	0.3290	0.1956	0.0927	0.3039	0.1813	0.0622	0.1502	0.0827
OFR with B-splines	$a_1(t)$	0.2524	1.8889	0.1646	0.1216	1.1085	0.1028	0.0913	1.0267	0.0810
	$a_2(t)$	0.1035	0.2564	0.1431	0.0651	0.1749	0.1149	0.0456	0.1463	0.0827
	$b_1(t)$	0.1362	1.2864	0.1684	0.0849	0.8595	0.0662	0.0501	1.0067	0.0562
	$b_2(t)$	0.1499	0.3159	0.1873	0.0774	0.1499	0.0899	0.0603	0.1322	0.0772
UOFR-MI with B-splines	$\boldsymbol{a_1(t)}$	**0.1563**	**1.4998**	**0.1269**	**0.0842**	**0.0925**	**0.0725**	**0.0712**	**0.9009**	**0.0781**
	$\boldsymbol{a_2(t)}$	**0.0622**	**0.1687**	**0.0953**	**0.0580**	**0.0929**	**0.1118**	**0.0435**	**0.1416**	**0.0816**
	$\boldsymbol{b_1(t)}$	**0.1009**	**1.1802**	**0.1323**	**0.0596**	**0.0521**	**0.0646**	**0.0307**	**0.4143**	**0.0426**
	$\boldsymbol{b_2(t)}$	**0.1044**	**0.2223**	**0.1335**	**0.0603**	**0.0559**	**0.0665**	**0.0436**	**0.0982**	**0.0580**

where bold values indicate the best results.

data and all candidate expanded terms are modulated by the first- and second-order derivatives of the cubic B-spline basis functions. The UOFR-MI algorithm is then applied to select significant model terms from candidate terms expanded by the B-spline functions and estimate the corresponding coefficients. The constructed model parameters are shown in Fig. 1(c). For comparison, the standard RLS algorithm with forgetting factor $\mu = 0.90$ and the conventional OFR algorithm with B-splines are also applied to identify the TV system, and the results of parameter estimates are given in Fig. 1(a) and (b), respectively.

Fig. 1(a) shows that the RLS algorithm obtains smooth estimates but fails to track rapidly varying piecewise parameters. Fig. 1(b) presents the results of parameter estimates using the OFR algorithm with B-splines. The estimates can track abrupt variations in TV parameters well but the estimation of the constant model parameter $b_2(t)$ is not good. In contrast, the UOFR-MI approach can better track the variation of the parameters and more effectively capture major features of different waveforms: the constant value, smooth changes, and abrupt changes.

In order to verity the effectiveness of the proposed method, three noise data (corrupted by color noises) are generated with different signal-to-noise ratios (SNRs) of 20, 15, and 10 dB, respectively. The value of the mean absolute error (MAE), normalized root mean squared error (RMSE), and the standard deviation (Std) of the TV parameter estimates for these three cases is given in Table III. It is obvious that the calculated MAE, RMSE, and Std of the proposed method are smaller than other two methods; this statistically confirms that the proposed procedure is more adaptive and possesses better ability for tracking the TV parameters even with color noise contamination. Particularly, it is worth noting that the superiority of the proposed method is more significant when the noise level increases. Note that, in this paper, MAE and RMSE are defined as

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}|\hat{c}(t) - c(t)| \tag{42}$$

TABLE IV
AVERAGE PREDICTION ERRORS BY CROSS VALIDATION FOR EXAMPLE 2

Approach	10dB		15dB		20dB	
	RMSE	Std	RMSE	Std	RMSE	Std
RLS ($\mu = 0.90$)	1.8504	0.4114	1.6787	0.3210	1.5054	0.2966
OFR with B-splines	0.8630	0.2935	0.7079	0.2076	0.6862	0.1395
UOFR-MI with B-splines	**0.8081**	**0.2454**	**0.6987**	**0.1670**	**0.4925**	**0.1050**

where bold values indicate the best results.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\frac{[\hat{c}(t) - c(t)]^2}{c(t)^2}} \tag{43}$$

where $\hat{c}(t)$ represents the estimates of coefficients $c(t)$ in the TV-NARX model and N is the length of the observations.

The efficiency of the proposed algorithm can be further evaluated by the 10-fold cross validation method. Specifically, the testing and training data subsets are randomly sampled from the B-splines-based expansion model. TV parameters are estimated by the three compared algorithms on the training data, and the prediction results from different methods are tested by the testing data subset. Totally, five rounds 10-fold cross validation is performed and the average of the test results is shown in Table IV. It can be observed that the prediction error from the proposed algorithm is smaller than the other two parameter estimation methods, indicating excellent prediction power of the proposed method.

C. Example 3: Nonlinear Time-Varying Parameter Estimation

To test a more severe situation, the data in the third example are produced from a TV-NARX model as follows:

$$\begin{cases} w(k) = c_{1,0}(1,k)y(k-1) + c_{1,0}(2,k)y(k-2) \\ \qquad + c_{0,1}(1,k)u(k-1) + c_{0,2}(2,2,k)u(k-2)^2 \\ y(k) = w(k) + \dfrac{1}{1-0.5z^{-1}}e(t) \end{cases} \tag{44}$$

TABLE V
COMPARISON OF THE ESTIMATION RESULTS FOR EXAMPLE 3 (SNR = 10, 15, AND 20 dB)

Approach	Estimated coefficients	10dB			15dB			20dB		
		MAE	RMSE	Std	MAE	RMSE	Std	MAE	RMSE	Std
RLS($(\mu = 0.95)$)	$c_{1,0}(1)$	0.1245	0.3584	0.1758	0.1346	0.5212	0.4007	0.1002	0.3205	0.1634
	$c_{1,0}(2)$	0.0203	0.2763	0.0258	0.0230	0.5108	0.0733	0.0158	0.3137	0.0311
	$c_{0,1}(1)$	0.0922	0.4048	0.1216	0.0722	0.2483	0.3189	0.0583	0.2913	0.0807
	$c_{0,2}(2,2)$	0.0533	0.4687	0.1147	0.0322	0.3054	0.0822	0.0277	0.1509	0.0628
OFR with B-splines	$c_{1,0}(1)$	0.1125	0.2215	0.1183	0.0939	0.2318	0.1133	0.0466	0.1207	0.0684
	$c_{1,0}(2)$	0.0302	0.3718	0.0366	0.0231	0.3117	0.0309	0.0154	0.2019	0.0202
	$c_{0,1}(1)$	0.1099	0.5161	0.1515	0.0676	0.3068	0.0905	0.0319	0.1384	0.0408
	$c_{0,2}(2,2)$	0.0599	0.3067	0.0760	0.0293	0.1712	0.0449	0.0236	0.1249	0.0355
UOFR-MI with B-splines	$\boldsymbol{c_{1,0}(1)}$	**0.0774**	**0.1747**	**0.0963**	**0.0790**	**0.2060**	**0.1019**	**0.0431**	**0.1125**	**0.0612**
	$\boldsymbol{c_{1,0}(2)}$	**0.0193**	**0.2225**	**0.0223**	**0.0183**	**0.2880**	**0.0285**	**0.0109**	**0.1525**	**0.0152**
	$\boldsymbol{c_{0,1}(1)}$	**0.0776**	**0.3596**	**0.1059**	**0.0465**	**0.2088**	**0.0626**	**0.0299**	**0.1298**	**0.0389**
	$\boldsymbol{c_{0,2}(2,2)}$	**0.0335**	**0.1912**	**0.0433**	**0.0253**	**0.1580**	**0.0382**	**0.0161**	**0.0905**	**0.0283**

where bold values indicate the best results.

where the piecewise varying parameters are given as

$$c_{1,0}(1,k) = \begin{cases} 1, & 0 \le k\Delta t < 0.2 \text{ s} \\ 0.4, & 0.2 \le k\Delta t < 0.4 \text{ s} \\ 0.8, & 0.4 \le k\Delta t \le 1 \text{ s} \end{cases}$$

$$c_{1,0}(2,k) = -0.3$$

$$c_{0,1}(1,k) = 0.1$$

$$c_{0,2}(2,2,k) = \begin{cases} 0.2, & 0 \le k\Delta t < 0.6 \text{ s} \\ 0.5, & 0.6 \le k\Delta t < 0.8 \text{ s} \\ 0.3, & 0.8 \le k\Delta t \le 1 \text{ s}. \end{cases} \quad (45)$$

The input $u(t)$ is a Gaussian random sequence with variance 1. $e(k)$ is Gaussian distributed with variance 0.06^2, that is, $e(k) \sim N(0, 0.06^2)$. The sampling time Δt is 0.0025 s, which indicates $f_s = 400$ Hz.

Fig. 2(a) shows the estimation results of TV coefficient using the RLS algorithm with forgetting factor $\mu = 0.95$. It shows that the results cannot track the abrupt changes in the TV parameters due to the limited convergence speed. Fig. 2(b) shows the estimation results of the OFR method with B-splines. The performance of the proposed algorithm is slightly better than that of the OFR (with B-splines) but much better than that of the RLS in terms of the overall variance of the parameter estimates. Similar to the previous example, the MAE, RMSE, and Std of TV parameter estimates are presented in Table V. Compared with other two methods, it is obvious that the proposed procedure has a better tracking performance even with a high level of noise, namely, a low SNR, which indicates the proposed UOFR-MI with B-splines method can be an effective tool for analyzing noise contaminated signals.

Table VI shows the cross validation results of the proposed UOFR-MI with B-splines method and other two compared methods, i.e., the RLS algorithm and OFR with B-splines method. The smaller prediction errors indicate that the proposed algorithm can achieve better predictive performance than other two methods; this demonstrates the advantage of the proposed algorithm in model structure detection and tracking system changes for the TV system.

TABLE VI
AVERAGE PREDICTION ERRORS BY CROSS VALIDATION FOR EXAMPLE 3

Approach	10dB		15dB		20dB	
	RMSE	Std	RMSE	Std	RMSE	Std
RLS ($\mu = 0.95$)	1.7073	0.2183	1.2051	0.1583	1.3406	0.1760
OFR with B-splines	0.8545	0.2147	0.6786	0.1194	0.6363	0.0754
UOFR-MI with B-splines	**0.7507**	**0.1434**	**0.5133**	**0.0877**	**0.5907**	**0.0564**

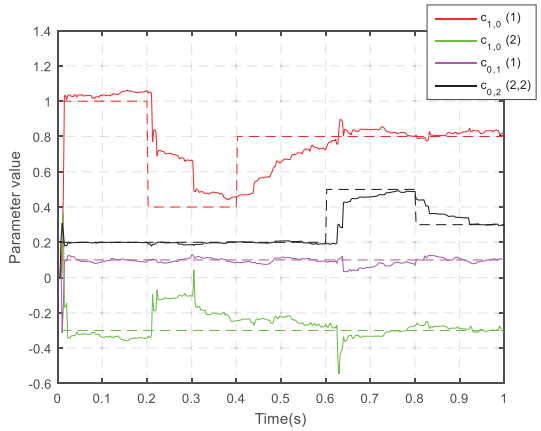
where bold values indicate the best results.

IV. APPLICATION TO REAL DATA: EEG DATA MODELING AND ANALYSIS

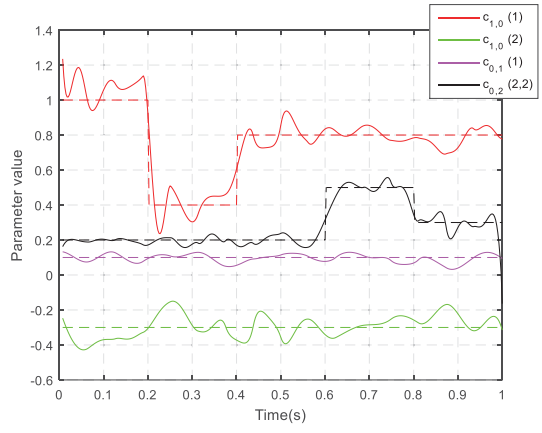
The UOFR-MI has been successfully used for the identification of TV systems in the previous simulations. In this section, the proposed method is applied to a real EEG recording to illustrate its applicability and efficiency for real data modeling problem. The EEG data set used here is available publicly from the University of Bonn [39]. Fig. 3 shows a snapshot of the EEG signal in a time scale of 6 s, with a sampling rate of 173.61 Hz. The initial candidate NARX model structure for the EEG signal is chosen to be

$$y(t) = c_0 + \sum_{i=1}^{10} c_1(i)y(t-i) + \sum_{j=1}^{10} c_2(i,j)y(t-i) \times y(t-j) + e(t). \quad (46)$$

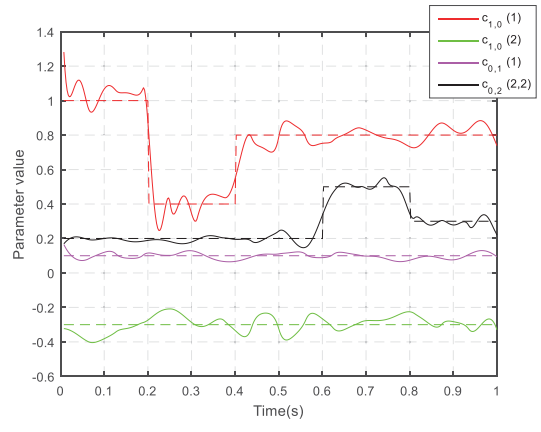
The initial model involves a total of 66 candidate model terms. The UOFR-MI algorithm is applied to identify a compact TV-NARX model from the initial candidate model. According to the AIC criterion, a model structure with five significant model terms is preferred and the model is as



(a)



(b)



(c)

Fig. 2. Identification results of the TV-NARX (44) using different approaches. The true TV parameters are shown as dashed lines, and the estimates are solid lines. (a) RLS algorithm with forgetting factor $\mu = 0.95$. (b) OFR method with B-splines. (c) UOFR-MI method with B-splines.

follows:

$$y(t) = \sum_{i=1}^{3} c_1(i,t)y(t-i) + c_2(1,1,t)y^2(t-1) + c_2(3,3,t)y^2(t-3) + e(t). \quad (47)$$

The third-, fourth-, and fifth-order cardinal B-splines with a scale $j = 4$ are adopted to expand TV parameters, and the

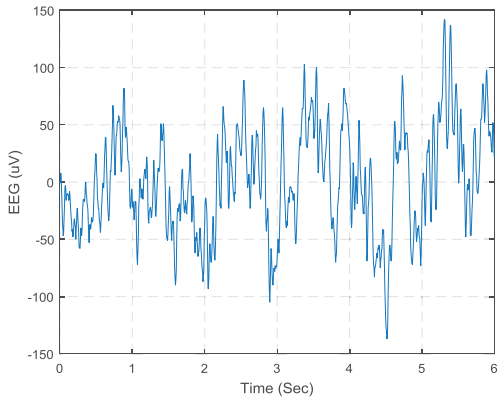


Fig. 3. EEG signal recorded during 6 s with a sampling rate of 173.61 Hz.

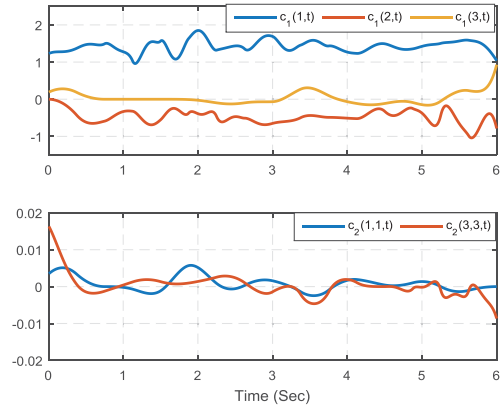


Fig. 4. TV coefficient estimates of nonlinear constructed model (47) for the EEG recordings.

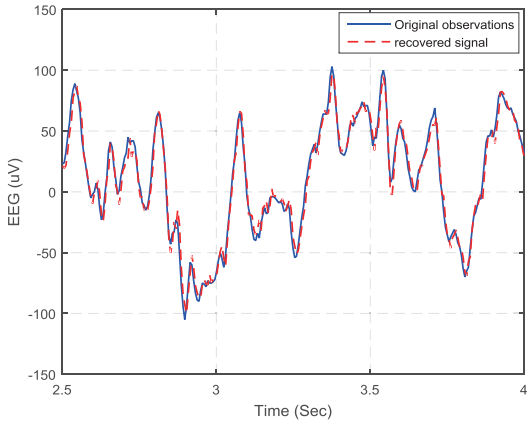


Fig. 5. Comparison of the recovered signal from the identified TV-NARX model (47) and the original EEG observations. Blue line indicates the EEG observations and the red dashed line indicates the recovered signal from the TV-NARX model. For a clear visualization, only the date points in the period from 2.5 to 4 s are displayed.

cubic B-spline basis functions are employed as the modulate function. The estimates of five TV coefficients are presented in Fig. 4. The recovered signal, calculated from the TV-NARX model (47), and associated original EEG recordings are shown in Fig. 5.

TABLE VII
AVERAGE PREDICTION ERRORS BY CROSS VALIDATION FOR EEG RECORDINGS

Method	RMSE	Std
RLS (μ=0.95)	0.9749	8.9631
OFR with B-splines	0.4743	7.4936
UOFR-MI with B-splines	**0.3749**	**7.4712**

where bold values indicate the best results.

Additionally, five rounds of 10-fold cross validation are conducted to evaluate the effectiveness of the proposed UOFR-MI with B-splines method. In this paper, the testing data are randomly sampled from the multiwavelet basis function expansion model. The average prediction error, measured by RMSE and Std, is given in Table VII. Clearly, in comparison with the models given by the RLS method and the OFR with B-splines method, the TV-NARX model established by the UOFR-MI algorithm is more flexible and adaptable for capturing the sharp variations of the TV EEG signal.

V. CONCLUSION

A new UOFR-MI algorithm combined with the basis function expansion framework has been proposed, where time-dependent parameters were approximated by a set of multiwavelet basis functions. Three numerical case studies were performed to test the performance of the proposed method. In these case studies, several different types of TV parameters were considered, including both smooth and abrupt changes. Several criteria were used to measure the performance of the estimated parameters, and 10-fold cross validation method was also employed to further verify the efficiency of the constructed models. The proposed algorithm shows a good ability and flexibility in capturing time-varying properties of interest. The results of simulation examples indicate that the proposed method can determine the model structure effectively and give more accurate estimates for either rapid or smooth piecewise varying parameters. Furthermore, the parameter estimation and prediction results on EEG recordings show that the proposed procedure is more powerful in tracking fast changes of TV biomedical signals.

An advantage of the proposed method over traditional methods, for example, Kalman filtering and adaptive recursive estimation approaches, is that it does not need to assume some stochastic model types for TV coefficients themselves, but only need to specify a set of basis functions that can be used to approximate the change of TV coefficients. In comparison with similar existing functional series expansion methods, the proposed method can produce more efficient model structure by making use of extra information characterized by the weak derivatives of relevant signals. The main drawback of the proposed method is its heavy computational load, which may be much higher than the existing functional series expansion methods; this is because the ULS problem (12) is different from the traditional least squares approach and involves the calculation of a number of simultaneous equations. This computational issue, however, becomes less critical when a high performance PC is available.

REFERENCES

[1] M. D. Spiridonakos and S. D. Fassois, "Non-stationary random vibration modelling and analysis via functional series time-dependent ARMA (FS-TARMA) models—A critical survey," *Mech. Syst. Signal Process.*, vol. 47, nos. 1–2, pp. 175–224, Aug. 2014.

[2] W. C. Su, C. Y. Liu, and C. S. Huang, "Identification of instantaneous modal parameter of time-varying systems via a wavelet-based approach and its application," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 29, no. 4, pp. 279–298, Apr. 2014.

[3] P. Dewilde and A.-J. Van der Veen, *Time-Varying Systems and Computations*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1998.

[4] L. Sheng, Z. Wang, L. Zou, and F. E. Alsaadi, "Event-based H_∞ state estimation for time-varying stochastic dynamical networks with state- and disturbance-dependent noises," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2016.2580601.

[5] O. Nelles, *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Berlin, Germany: Springer, 2001.

[6] F. He, S. A. Billings, H.-L. Wei, P. G. Sarrigiannis, and Y. Zhao, "Spectral analysis for nonstationary and nonlinear systems: A discrete-time-model-based approach," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2233–2241, Aug. 2013.

[7] J. R. A. Solares and H.-L. Wei, "Nonlinear model structure detection and parameter estimation using a novel bagging method based on distance correlation metric," *Nonlinear Dyn.*, vol. 82, no. 1, pp. 201–215, Oct. 2015.

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

[9] J. C. M. Bermudez and N. J. Bershad, "Transient and tracking performance analysis of the quantized LMS algorithm for time-varying system identification," *IEEE Trans. Signal Process.*, vol. 44, no. 8, pp. 1990–1997, Aug. 1996.

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

[11] 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. 2016.

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

[13] H. L. Wei, S. A. Billings, and J. Liu, "Term and variable selection for non-linear system identification," *Int. J. Control*, vol. 77, no. 1, pp. 86–110, 2004.

[14] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1411–1423, Nov. 2006.

[15] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification—A survey," *Automatica*, vol. 26, no. 1, pp. 7–21, Jan. 1990.

[16] C.-H. Loh, C.-Y. Lin, and C.-C. Huang, "Time domain identification of frames under earthquake loadings," *J. Eng. Mech.*, vol. 126, no. 7, pp. 693–703, Jul. 2000.

[17] M. Niedzwiecki and T. Klaput, "Fast recursive basis function estimators for identification of time-varying processes," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1925–1934, Aug. 2002.

[18] K. H. Chon, H. Zhao, R. Zou, and K. Ju, "Multiple time-varying dynamic analysis using multiple sets of basis functions," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 5, pp. 956–960, May 2005.

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

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

[21] Y. Zheng, Z. Lin, and D. B. H. Tay, "Time-varying parametric system multiresolution identification by wavelets," *Int. J. Syst. Sci.*, vol. 32, no. 6, pp. 775–793, 2001.

[22] S. A. Billings and H.-L. Wei, "A new class of wavelet networks for nonlinear system identification," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 862–874, Jul. 2005.

[23] R. Zou, H. Wang, and K. H. Chon, "A robust time-varying identification algorithm using basis functions," *Ann. Biomed. Eng.*, vol. 31, no. 7, pp. 840–853, Jul. 2003.
[24] Y. Li, H.-L. Wei, S. A. Billings, and P. Sarrigiannis, "Time-varying model identification for time–frequency feature extraction from EEG data," *J. Neurosci. Methods*, vol. 196, no. 1, pp. 151–158, Mar. 2011.
[25] F. He, H.-L. Wei, and S. A. Billings, "Identification and frequency domain analysis of non-stationary and nonlinear systems using time-varying NARMAX models," *Int. J. Syst. Sci.*, vol. 46, no. 11, pp. 2087–2100, 2015.
[26] Y. Guo, L. Z. Guo, S. A. Billings, and H.-L. Wei, "Ultra-orthogonal forward regression algorithms for the identification of non-linear dynamic systems," *Neurocomputing*, vol. 173, pp. 715–723, Jan. 2016.
[27] L. Piroddi and W. Spinelli, "An identification algorithm for polynomial NARX models based on simulation error minimization," *Int. J. Control*, vol. 76, no. 17, pp. 1767–1781, 2003.
[28] H.-L. Wei and S. A. Billings, "Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information," *Int. J. Model., Identificat. Control*, vol. 3, no. 4, pp. 341–356, 2008.
[29] S. A. Billings and H.-L. Wei, "Sparse model identification using a forward orthogonal regression algorithm aided by mutual information," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 306–310, Jan. 2007.
[30] S. Wang, H.-L. Wei, D. Coca, and S. A. Billings, "Model term selection for spatio-temporal system identification using mutual information," *Int. J. Syst. Sci.*, vol. 44, no. 2, pp. 223–231, 2013.
[31] S. A. Billings, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. Hoboken, NJ, USA: Wiley, 2013.
[32] U. R. Acharya, S. V. Sree, P. C. A. Ang, R. Yanti, and J. S. Suri, "Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals," *Int. J. neural Syst.*, vol. 22, no. 2, p. 1250002, Apr. 2012.
[33] S. Chen, X. Hong, E. F. Khalaf, F. E. Alsaadi, and C. J. Harris, "Comparative performance of complex-valued B-spline and polynomial models applied to iterative frequency-domain decision feedback equalization of Hammerstein channels," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2016.2609001.
[34] H.-L. Wei and S. A. Billings, "An efficient nonlinear cardinal b-spline model for high tide forecasts at the Venice lagoon," *Nonlinear Process. Geophys.*, vol. 13, no. 5, pp. 577–584, 2006.
[35] H. L. Wei and S. Billings, "Identification of time-varying systems using multiresolution wavelet models," *Int. J. Syst. Sci.*, vol. 33, no. 15, pp. 1217–1228, 2002.
[36] C. K. Chui and J.-Z. Wang, "On compactly supported spline wavelets and a duality principle," *Trans. Amer. Math. Soc.*, vol. 330, no. 2, pp. 903–915, Apr. 1992.
[37] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.
[38] Y. Guo, L. Z. Guo, S. A. Billings, and H.-L. Wei, "An iterative orthogonal forward regression algorithm," *Int. J. Syst. Sci.*, vol. 46, no. 5, pp. 776–789, 2015.
[39] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 6, p. 061907, 2001.

Yang Li received the Ph.D. degree in automatic control and systems engineering from The University of Sheffield, Sheffield, U.K., in 2011.

He did post-doctoral research with the Department of Computer and Biomedical Engineering, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, for one year. In 2013, he joined the Department of Automation Sciences and Electrical Engineering, Beihang University, Beijing, China, as an Associate Professor. His current research interests include system identification and modeling for complex nonlinear processes: NARMAX methodology and applications, nonlinear and nonstationary signal processing, intelligent computation and data mining, parameter estimation and model optimization, and sparse representation.

Wei-Gang Cui was born in Shaanxi, China, in 1994. He received the bachelor's degree in mathematics from Beihang University, Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree with the Department of Automation Science and Electrical Engineering.

His current research interests include signal processing, system identification, and time–frequency domain analysis.

Yu-Zhu Guo received the B.Sc. and M.Sc. degrees from the Beijing Institute of Technology, Beijing, China, and the Ph.D. degree in automatic control and systems engineering from The University of Sheffield, Sheffield, U.K., in 2009.

He is currently a researcher at The University of Sheffield, UK, and Associate Professor with the School of Automation Science and Electrical Engineering, Beihang University, Beijing. His current research interests include system identification and information processing for nonlinear systems, NARMAX methods, nonlinear spectral analysis, spatio-temporal systems, multiscale modeling of neuro-musculo-skeletal systems, and bioinformatics.

Tingwen Huang (SM'12) received the B.S. degree from Southwest University, Chongqing, China, in 1990; the M.S. degree from Sichuan University, Chengdu, China, in 1993; and the Ph.D. degree from Texas A&M University, College Station, TX, USA, in 2002.

He has expertise in chaotic dynamical systems, neural networks, optimization and control, and traveling wave phenomena. He is currently a Professor with Texas A&M University at Qatar, Doha, Qatar.

Dr. Huang currently serves as an Editorial Board Member for two international journals, i.e., the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS and *Cognitive Computation*.

Xiao-Feng Yang is currently an Associate Professor with the Department of Mechanics and Architectural Engineering, China University of Mining and Technology, Beijing, China. He has authored over 20 international journal papers and holds five patents. His current research interests include the new energy and safety engineering.

Hua-Liang Wei received the Ph.D. degree in the Department of Automatic Control and Systems Engineering, the University of Sheffield, UK, in 2004.

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