This is a repository copy of *Neural activity inspired asymmetric basis function TV-NARX model for the identification of time-varying dynamic systems*.

Version: Accepted Version

Accepted Manuscript

Neural Activity Inspired Asymmetric Basis Function TV-NARX Model for the Identification of Time-Varying Dynamic Systems

Yuzhu Guo , Lipeng Wang , Yang Li , Jingjing Luo ,
Kailiang Wang , S.A. Billings , Lingzhong Guo
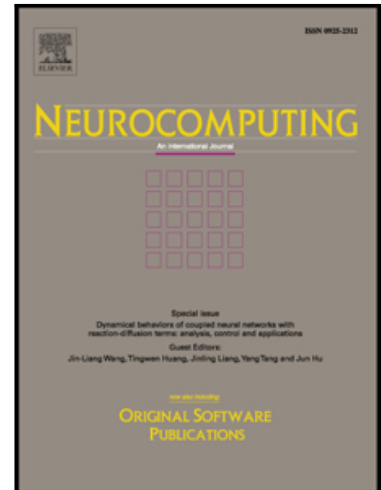
Please cite this article as: Yuzhu Guo , Lipeng Wang , Yang Li , Jingjing Luo , Kailiang Wang , S.A. Billings , Lingzhong Guo , Neural Activity Inspired Asymmetric Basis Function TV-NARX Model for the Identification of Time-Varying Dynamic Systems, *Neurocomputing* (2019), doi: https://doi.org/10.1016/j.neucom.2019.04.045

Neural Activity Inspired Asymmetric Basis Function TV-NARX Model for the Identification of Time-Varying Dynamic Systems

Yuzhu Guo, Lipeng Wang, Yang Li, Jingjing Luo, Kailiang Wang, S. A. Billings, and Lingzhong Guo

Abstract

Inspired by the unique neuronal activities, a new time-varying nonlinear autoregressive with exogenous input (TV-NARX) model is proposed for modelling nonstationary processes. The NARX nonlinear process mimics the action potential initiation and the time-varying parameters are approximated with a series of postsynaptic current like asymmetric basis functions to mimic the ion channels of the inter-neuron propagation. In the model, the time-varying parameters of the process terms are sparsely represented as the superposition of a series of asymmetric alpha basis functions in an over-complete frame. Combining the alpha basis functions with the model process terms, the system identification of the TV-NARX model from observed input and output can equivalently be treated as the system identification of a corresponding time-invariant system. The locally regularised orthogonal forward regression (LROFR) algorithm is then employed to detect the sparse model structure and estimate the associated coefficients. The excellent performance in both numerical studies and modelling of real physiological signals showed that the TV-NARX model with asymmetric basis function is more powerful and efficient in tracking both smooth trends and capturing the abrupt changes in the time-varying parameters than its symmetric counterparts.

1. Introduction

Nonlinear time-varying processes exist universally in numerous science and engineer problems, including physical, chemical, biological, physiological, neurobiological, aerospace systems, to name a few. Detection of the model structure and estimation of the associated time-varying parameters are much more challenging than counterparts in the time-invariant cases. Available methods for modelling and identifying time-varying systems include piecewise linear models, recursive least squares, least mean squares, Kalman filtering and so on. Among them, the multi-wavelet basis function based time-varying nonlinear autoregressive with exogenous input (TV-NARX) model has been proved to be powerful in representing complex nonlinear time-varying systems and identifying complex nonstationary processes in both time and frequency domains [1-4]. The new system identification methods for time-invariant systems, such as Multi-step-length gradient iterative algorithm and the Variational Bayesian Approach [5, 6] may be modified and applied to identify time-invariant systems.

The multi-wavelet basis function TV-NARX model expands time-varying parameters with multiple wavelet basis functions and the time-varying system identification can then be transferred into identifying a model with time-invariant coefficients. The orthogonal forward regression algorithms were employed to identify a parsimonious model structure and estimated the associated parameters [7].

One of the main advantages of the TV-NARX method lies in that the time-varying parameters are approximated with an over-complete basis function. Under an over-complete basis, the decomposition of a signal is not unique but this offers some advantages. Firstly, there is greater flexibility in capturing the structure in the data. Instead of a small set of general basis functions, a larger set of

more specialized basis functions can be employed such that any particular signal can be reproduced with relatively few functions producing more compact representations. Additionally, over-complete representations increase the stability of the representation in response to small perturbations of the signal. The basis function widely used to construct multi-resolution decompositions includes B-spline, Chebyshev polynomials, Legendre polynomials, wavelets, curvelets and so on [4, 8, 9]. For the diverse forms of structure that occur in natural signal it is difficult to know a priori what class of functions is most appropriate. Although over-complete bases can be more flexible in terms of how the signal is represented, there is no guarantee that the hand-selected basis vectors will well match the data structure. Ideally, we would like the basis itself to adapt to the data, so that for signal class of interest, the basis function captures the maximal amount of structure in the data.

On the one hand, the over-complete dictionary can be optimised from signal statistics using evolutionary algorithms that are of limited relevance to biology [10-12]. However, the process can be time-consuming and the results depend on the signals used in the optimisation. On the other hand, the neural system has evolved over millions of years to effectively cope with stimulations in the natural environment. Since using resources efficiently is important in the competition for survival, it is reasonable to think that the neural system has discovered efficient coding strategies for representing natural signals. Therefore, the challenge is to develop approaches for deducing computational principles relevant to biological systems.

As the fundamental units of communication between neurons, action potentials which promote the interaction of chemical or electrical signals, allow neurons to encode information by generating action potentials with a wide range of shapes, frequencies, and patterns [13]. The action potential transfers from neuron to neuron through synapses and produces postsynaptic currents (PSC), including excitatory PSC and inhibitory PSC. Neuronal communication is regulated as a balance between excitatory and inhibitory influences. Different from the action potential, the postsynaptic current is graded and can approximately be described as alpha functions [14]. Action potentials occur when the sum total of all of the excitatory and inhibitory inputs pushes the neuron's membrane potential reach the firing threshold spiking and propagating as a wave along the axon to synapses of nerve terminals. The dynamics demonstrate how changes in the membrane can constitute a signal. Similar activities have also been observed in the excitation-contraction pathway of skeletal muscle. The contraction force is controlled by a group of motion units. The generated muscle force depends on the number of motion unit involved and the fire rate, that is, the temporal spike train from motor neurons [15]. In summary, the neuron dynamics possesses the following characteristics: the connection is over-complete, one neuron can connect with many other neurons; the PSCs are of a similar alpha function shape; a nonlinear dynamic process determines how the neuron burst an action potential under the effects of EPSCs and IPSCs.

Based on the above observation, a new time-varying model inspired by the neuron activity is proposed. The new model mimics the single neuron dynamics which govern the initiation and propagation of action potential. The model is of a general nonlinear NARX structure with time varying parameters. The time varying parameters are approximated by an alpha function train to mimic the PSC based neuron interactions. In the practice, the TV-NARX model can be identified from nonstationary observations and the time varying parameter can be learned from the over-completed basis dictionary just as in reservoir computing [16].

One of the main objectives is to improve the sparse representation of the TV-NARX model. The parameters in the model can change smoothly or abruptly in real systems. Inspired by the unique shape of the post-synaptic currents, a new type of basis functions which consist of an over-complete frame is proved capable of adapting to different parameter varying. Using the novel and more powerful alpha basis function, a sparse model structure can be obtained. Compared with the traditional TV-NARX model, our new model is closely related to the biological neuronal dynamics,

which bridges the gap between mathematical models and biological system and provide a new pathway for the design of new generation spike coding neural networks. Additionally, our results give a new possible interpretation of the post-synaptic currents (PSC). Results showed that the specific PSC shape can help the neurons transmit information more efficiently, that is, transmit the same amount of information using fewer action potentials.

Unlike artificial neural networks, the new TV-NARX model is capable of giving explicit model structure and how each of the nonlinear terms and coefficients affects the system behaviours can then be analyzed, for example, frequency domain interpretation or bifurcation analysis.

The remainder of the paper is organized as follows: the biological synaptic dynamics and neuron spiking are briefly reviewed and the new TV-NARX model is proposed by generalising the mathematical model in Section 2. The identification of the TV-NARX model using the locally regularised orthogonal forward regression algorithm is discussed in Section 3. Section 4 demonstrates the system identification methods and illustrates the efficiency of the new system identification method. Conclusions are finally drawn in Section 5.

2. Neuronal dynamics and the associated PSC TV-NARX model

There are plenty of studies in both anatomy and functional mechanism of single excitable neurons, to reveal the underlining reason for their firing or the action potential (AP). As the fundamental mechanism for communication between neurons, AP transfers information predominantly through synapses, where electrical particles exchange dynamically between intracellular and extracellular environment to change the membrane potential of the post-synaptic neuron. Multiple synaptic inputs and integration mechanism are essential for the robust and precise characteristic of neuronal firing and encoding information with APs of a wide range of temporal or rate patterns [13]. Neuronal communication is exquisitely regulated as balanced between excitatory and inhibitory influences [17], and essentially it is the integration of many excitatory and inhibitory inputs with different amplitudes and phases in a nonlinear framework. Active dendritic integration as a mechanism for robust and precise grid cell firing, nonlinear dendritic processing determines angular tuning of barrel cortex neurons in vivo. Contributions of generated post-synaptic conductances and post-synaptic currents are often models as ensembles for computational implementation of the kinetics of membrane potential, such as the single-neuron paradigm of neural mass model [18].

The mechanisms of neuronal dynamics will be briefly reviewed in this section and the new TV-NARX model will be proposed by generalising the signal neuron model.

2.1 Neuronal dynamics

Action potentials are the basic currency of the brain and allow neurons to communicate with each other, computations to be performed, and information to be processed. The computation performed by single neurons can be defined as the mapping from afferent spike trains to the output spike train which is communicated to their postsynaptic targets. Single neuron dynamics essentially includes two elementary mechanisms: the synaptic dynamics based interneuron interaction and the neuron firing dynamics, namely, the initiation and propagation of action potential.

The neuronal dynamics are often described by voltage-controlled ion channels which integrate the effects of action potentials through dendrites and synapses. Ion channels are also key components in a wide variety of biological processes that involve rapid changes in cells, such as cardiac, skeletal, and smooth muscle contraction, epithelial transport of nutrients and ions, T-cell activation and pancreatic beta-cell insulin release.

Synaptic conductance can be mostly described as rapid binding followed by slow unbinding of the transmitter, and alpha function [19, 20]. An approximate model of synaptic transmission is to assume that each spike evokes a change in the conductance of the postsynaptic membrane with a characteristic time course which can be described with the alpha function [14].

$$\alpha(t) = \left(\frac{t}{t_{peak}}\right)\exp\left(1 - \frac{t}{t_{peak}}\right) \tag{1}$$

Voltage-gated ion channels are represented by electrical conductance. For a train of spikes, the postsynaptic conductance change of the i-th ion channel is given by the superposition of time-shifted alpha functions:

$$g_i(t) = \sum_j \theta_{i,j} \alpha_{i,j}(t - t_{i,j}) \tag{2}$$

where $t_{i,j}$ are the spike times. The postsynaptic conductance changes are graded, weight $\theta_{i,j}$ represents the peak conductance change evoked by spike j.

Electrical input-output membrane voltage models produce a prediction for membrane output voltage as a functional electrical stimulation at the input stage. The various models in this category differ in the exact functional relationship between the input current and the output voltage and in the level of details. The membrane equation can then be described by combining the effects of each ion channel as the following ordinary differential equation

$$C\frac{dV_m}{dt} + \sum_i g_i(V_m - E_i) = 0 \tag{3}$$

where V_m is the membrane potential, C membrane capacitance. Terms $g_i(V_m - E_i)$ represent the postsynaptic current from the i-th ion channel with the postsynaptic conductance g_i. The postsynaptic conductance is represented as the superposition of the alpha function train as in (2); E_i is the reverse potential of the corresponding ion channel. The alpha functions represents the PSC from different synapses and reflect the inter-neuron interaction. The connections through synapse are highly over-complete. For example, a typical neuron in the mammalian central nervous system may receive several thousands of synaptic inputs from other neurons but not all the neurons are excited and generate an action potential.

The synaptic dynamics is followed by a spike generation process at the soma or axonal initial segment. The neuron produces an action potential when the membrane potential is over the threshold value under the effects of EPSC and IPSC. The differential equation explains the ionic mechanisms underlying the initiation and propagation of action potentials.

2.2 Generalization of the synaptic model

According to the discussion above, the neuronal model possesses following three characteristics: the connection is over-complete; the post-synaptic conductance is of a special alpha wave shape; the firing of the neuron is governed by a nonlinear dynamic process. A TV-NARX model which has all these features will be proposed for time-varying system identification by generalizing the neuronal model.

A discrete time model has computational advantages than the continuous dynamical system. The existing discrete time model for the neural dynamics include the Rulkov map [21]. However, the Rulkov map describes the spiking activity of biological neurons without considering the synaptic dynamics. Two neurons are symmetrically coupled with each other through a simplistically defined current flow [21]. Other discrete time model for excitable media includes the polynomial form models

[22, 23]. These results show that the neural excitation process can be described by nonlinear difference equations. Combining with the synaptic dynamics, a nonlinear difference equation with time-varying coefficients can be used to describe basic single neuron behaviours.

Equation (3) represents a time-varying model whose coefficients are composed of a series of alpha functions which are sparsely expanded by an over-complete basis dictionary which reflects the connections in a neuron network and the sparse expansion reflects the dynamical neuron modulation.

Discretizing the equation (3) yields a NARX model with time-varying parameters. A general form of NARX model can be given as

$$y(k) = F\left(y(k-1), \cdots, y(k-n_y), u(k-1), \cdots, u(k-n_u), e(k), \mathbf{c}(k)\right) \qquad (4)$$

where $y(k), u(k)$ denote observed and measured system output, input, and noise signals with the maximum delays n_y, n_u respectively, $e(k)$ is the model error that can often be assumed as an independent identical distributed white noise sequence with zero mean and a certain variance. $F(\cdot)$ is any nonlinear function of the output and input with time varying parameters $\mathbf{c}(k)$ [1, 7].

In order to mimic the neuronal dynamics, the time varying coefficients $\mathbf{c}(k)$ is a superposition of a series of alpha functions and the nonlinear function $F(\cdot)$ is designed to capable of describing the firing of the action potential. The obtained model originates from the neuron dynamics but of a much general form. In this study, the model will be used as a general model for system identification of nonstationary dynamic systems rather than only the single neuron behaviours.

A polynomial form TV-NARX model is often used, which is defined as

$$y(k) = \sum_{l=0}^{L} \sum_{p+q=l} \sum_{d_1, \cdots, d_p = 1}^{n_y} \sum_{d_{p+1}, \cdots, d_{p+q} = d}^{n_u} c_{p,q}(d_1, \cdots, d_{p+q}, k) \prod_{i=0}^{p} y(k-d_i) \prod_{i=p+1}^{p+q} u(k-d_i) + e(k) \qquad (5)$$

where

$$c_{p,q}(d_1, \cdots, d_{p+q}, k) = \sum_{j_{\mathbf{r}}} \theta_{j_{\mathbf{r}}} \alpha_{j_{\mathbf{r}}}(k), \quad j_{\mathbf{r}} = 1, \cdots, n_{\mathbf{r}} \qquad (6)$$

where the multiple subscript $\mathbf{r} := (p, q, d_1, \cdots, d_{p+q})$, $n_{\mathbf{r}}$ represents the number of PSC's which consist of the time varying coefficients of process term $\prod_{i=0}^{p} y(k-d_i) \prod_{i=p+1}^{p+q} u(k-d_i)$.

Model (6) gives a generalised model of a single neuron. Combining a group of models yields a new NARX Spiking Neural Network (NSNN) model which considers the spatiotemporal interactions between neurons. The NSNN model can be given as follows.

$$y^{(n)}(k) = \sum_{\mathbf{r}_n} \sum_{j_{\mathbf{r}_n}} \theta_{j_{\mathbf{r}_n}}^{(n)} \alpha_{j_{\mathbf{r}_n}}^{(n)}(k) \prod_{i=0}^{p_n} y^{(n)}(k-d_{i_n}) \prod_{i_n=p_n+1}^{p_n+q_n} u^{(n)}(k-d_{i_n}) + e^{(n)}(k) \qquad (7)$$

where the superscript represents the n-th subsystems and $\mathbf{r}_n := (p_n, q_n, d_1, \cdots, d_{p_n+q_n})$.

The use of this map to study neural networks has computational advantages because the map is easier to iterate than a continuous dynamical system. This saves memory and simplifies the computation of large neural networks. This study focuses on the system identification of the single neuron model, that is, the TV-NARX model (5) used to modelling nonstationary nonlinear processes where the alpha function train is determined using a system identification method which is discussed in the next section.

The new proposed model has two important features which are different from both biological single neuron models and general TV-NARX models. On the one hand, the new model is of a NARX model form. This enables the model to describe much wider types of nonstationary processes than biological neuron models do. On the other hand, different from general TV-NARX model, the time-varying parameters are represented as subscription of alpha basis functions. This enhances the flexibility and adaptive ability of the TV-NARX model, which will be shown in Section 4.

Remarks:

i)   Model generalised the neuron activity model. The model differs from the ion model in several ways: the new model is a discrete time model which is better suited for simulation than the differential equations;

ii) In the ion channel model, the changing of the conductions in different ion channels may not be independent. For example, both sodium and potassium channels will response to the input when a neuron action potential is injected through synapses.

iii) Model (5) may be used as a general model for the system identification of nonstationary nonlinear dynamics because of the powerful NARX model structure. The time-varying coefficients may have a different physical meaning in real applications, especially when it is used to describe the phenomena in macroscopic scales.

iv) The proposed model has a similar form as the multiple basis function TV-NARX models [1, 4]. Therefore, the system identification methods used in the references can also be used for the new model.

2.3 Smooth compactly supported alpha basis function

Compact support is an important property of wavelet functions and the associated scale function. A smooth compactly supported wavelet function has a better approximation ability. For example, an arbitrary polynomial of degree N - 1 can be written as a linear combination of integer shifts of the scaling function of a compactly supported wavelet with N vanishing moment [24]. Therefore, a smooth compactly supported basis function is expected in building a good model.

A compactly supported Beta wavelet has been proposed [25]. The associate scale function of the Beta wavelet is compactly supported and has a similar asymmetric shape as the alpha function. Thus, in the paper, we utilized the time shifted scale function, named as alpha basis function, to replace the alpha function to approximate the time-varying parameters. According to the original definition given by the author Araújo et.al, the new alpha basis function is defined as:

$$\phi(k|a,b) = \begin{cases} A\square(k-0)^{a-1}(1-k)^{b-1}, & 0 \le k \le 1 \\ 0, & otherwise \end{cases}$$

$$A = \frac{\Gamma(a+b)\square(\frac{a+b}{a+b+1})^{(a+b-1)/2}}{\Gamma(a)\Gamma(b)(a+b)^{a+b-1}}$$

(8)

where $1 \le a \le b$, $a,b \in \mathbb{N}$ are the parameters which control the shape of the alpha basis function. With the increase in the difference between the parameters a and b, the asymmetricity of the alpha basis function will increase, that is, the first half wave will be sharper and the second half flatter. $\Gamma(\cdot)$ is the generalized factorial function of Euler. According to the definition (8), the alpha basis function $\phi(k|a,b)$ only has a nonzero value in the interval $0 \le k \le 1$, so it is compactly supported, and the length of the support set is 1. An example of the alpha basis function is shown in Fig 1.
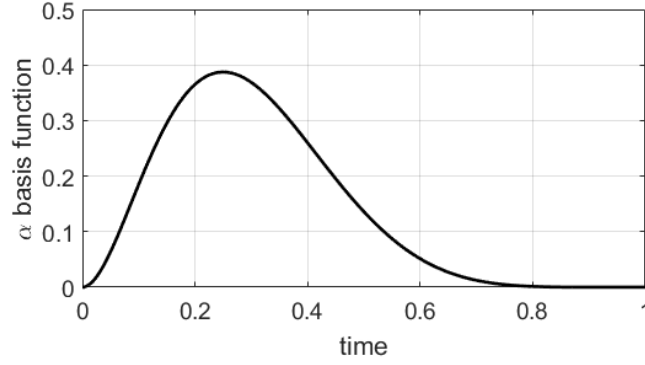
Figure 1. An example of Alpha basis function with $a = 3, b = 7$

Although the alpha basis function is not a wavelet because it doesn't integrate to zero. However, the scale function is more flexible as basis function to fit a curve because different wavelet functions can easily be constructed by calculating the difference between two scale functions with time delay. Wavelets which are generated from alpha basis functions with different time delay is shown in Fig 2. Actually, this mimics the supposition of EPSC and IPSC in intracellular potentials. Zheng et al. showed that local field potential (LFP) can be modelled as the combination of EPSC and IPSC [17].
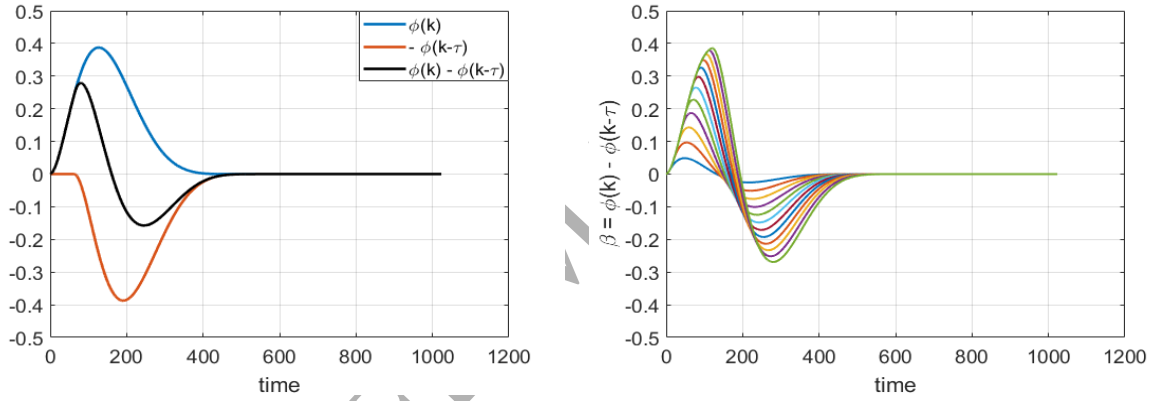

Figure 2. Generation of Beta wavelet with $\beta = \alpha_1 - \alpha_2$

An over-complete frame $\left\{\phi(k, \mu, \sigma | a, b)\right\}_{\mu, \sigma}$ can be constructed by varying the translation and

scale parameters μ and σ of the scale function, where $\phi(k, \mu, \sigma | a, b) = \phi(\frac{k - \mu}{\sigma}, | a, b)$ The

corresponding discrete time version wavelet function is $\phi(k, m, \sigma | a, b) = \phi(\frac{k - m}{\sigma}, | a, b)$, $k, m \in \mathbb{Z}$.

The time-varying coefficients can then be expanded under the frame as

$$c_{p,q}(d_1, \cdots, d_{p+q}, k) = \sum_{j_{\mathbf{r}}} \theta_{j_{\mathbf{r}}} \phi\left(\frac{k - m_{j_{\mathbf{r}}}}{\sigma_{j_{\mathbf{r}}}} \middle| a_{j_{\mathbf{r}}}, b_{j_{\mathbf{r}}}\right) \tag{9}$$

where the multiple subscript $\mathbf{r} := \left(p, q, d_1, \cdots, d_{p+q}\right)$.

Substituting (9) for the time varying coefficients in model (5) yields the new TV-NARX model

$$y(k) = \sum_{\mathbf{r}} \sum_{j_{\mathbf{r}}} \theta_{j_{\mathbf{r}}} \phi\left(\frac{k - m_{j_{\mathbf{r}}}}{\sigma_{j_{\mathbf{r}}}} \middle| a_{j_{\mathbf{r}}}, b_{j_{\mathbf{r}}}\right) \prod_{i=0}^{p} y(k - d_i) \prod_{i=p+1}^{p+q} u(k - d_i) + e(k) \tag{10}$$

3. Identification of the new TV-NARX model

The identification of the proposed model (10) from observations involves three related processes: the detection of the polynomial process terms, determination of the wavelet basis function set which consists the time-varying parameters, and estimation of the associated parameters.

Denote the term on the right hand side of the equation as

$$\psi_{j_{\mathbf{r}}} = \phi\left(\left.\frac{k - m_{j_{\mathbf{r}}}}{\sigma_{j_{\mathbf{r}}}}\right| a_{j_{\mathbf{r}}}, b_{j_{\mathbf{r}}}\right) \prod_{i=0}^{p} y(k-d_i) \prod_{i=p+1}^{p+q} u(k-d_i).$$

Model (10) becomes

$$y(k) = \sum_{\mathbf{r}} \sum_{j_{\mathbf{r}}} \theta_{j_{\mathbf{r}}} \psi_{j_{\mathbf{r}}} + e(k) \tag{11}$$

Denote the number of process term $\prod_{i=0}^{p} y(k-d_i) \prod_{i=p+1}^{p+q} u(k-d_i)$ as $R := \#\{\mathbf{r}\}$, and each time varying coefficient consists of n_r alpha basis functions for $r = 1, \cdots, R$. The model can be rearranged as

$$y(k) = \sum_{r=1}^{R} \sum_{j_r=1}^{n_r} \theta_{j_r} \psi_{j_r} + e(k) = \sum_{j=1}^{n_1+n_2+\cdots,n_R} \theta_j \psi_j + e(k) \tag{12}$$

Hence, the system identification of the TV-NARX model (10) becomes the detection of the model terms $\{\psi_j\}_{j=1}^{n_1+\cdots+n_R}$ and estimation of the time invariant parameter θ_j's. The problem reduces to the identification of a time invariant model (12).

The problem that identifies the non-stationary nonlinear system represented by the formula (10) involves selecting the most significant terms from a pre-defined candidate dictionary to build a model which is sufficient to describe the observed system behaviours and estimating corresponding parameters depended on a certain model structure. Note that some excellent algorithms, such as orthogonal forward regression (OFR) algorithm and its variants, could have been applied to solve this issue if the parameters are time-invariant. Combining the alpha basis functions with the process term transfers the time-varying model into a model with time invariant coefficients. The principle and steps of standard OFR algorithm is briefly introduced in Section 3.1 and the pseudocode of OFR algorithm is given in Appendix A.

The system identification process can then be summarised as:

a) Construct the dictionary of the process terms $\left\{\prod_{i=0}^{p} y_n(k-d_i) \prod_{i=p+1}^{p+q} u_n(k-d_i)\right\}_{p,q,d_i}$, and denote the cardinality of the set as N_p;

b) Construct the over-complete alpha basis function basis $\left\{\prod_{i=0}^{p} y_n(k-d_i) \prod_{i=p+1}^{p+q} u_n(k-d_i)\right\}_{p,q,d_i}$

$\otimes \{\phi(k,m,\sigma|a,b)\}_{m,\sigma,a,b}$, and denote the cardinality of the set as N_b;

8

c) Combine the two dictionaries by the Kronecker product $\otimes \{\phi(k,m,\sigma|a,b)\}_{m,\sigma,a,b} \frac{1}{2}$ to produce the associated time invariant system term dictionary $\{\psi_j\}_{j=1}^{N_p \times N_b}$, where $\otimes$ denote the Kronecker product, for example, $\{a_1, a_2, a_3\} \otimes \{b_1, b_2\} = \{a_1 b_1, a_1 b_2, a_2 b_1, a_2 b_2, a_3 b_1, a_3 b_2\}$.

d) Using the OFR algorithm to select significant terms from the dictionary $\{\psi_j\}_{j=1}^{N_p \times N_b}$ and estimate the associated parameters;

e) Collect all the alpha basis functions with the same process term to reproduce the time varying parameters as the weighted sum of the alpha basis functions;

f) Validate the obtained model.

3.1 LROFR algorithms for TV-NARX model identification

Once the time-varying nonlinear model (5) is reduced to the associated time-invariant linear-in-the-parameter form (12) by expanding the time-varying coefficients with alpha basis functions, the system identification methods for time invariant system can be applied. Since both the structures of the NARX model and the spike train of the alpha basis functions consisting the parameters are not known a prior, an over-complete term dictionary is constructed as the candidates for the model term selection. However, the detection of the model structure is entwined with the estimation of the associated parameter. Tedious trial-and-error processes for a parsimonious model structure needs to re-estimate the parameters for each trial. The combinations can be enormous and the process is computationally infeasible.

The OFR algorithm family which decouples the model structure detection and the parameter estimation by orthogonalizing the model terms and selecting model terms stepwise, has successfully used for the identification of different kinds of model. Based on the Error Reduction Ratio (ERR) significance criterion, a parsimonious model can be constructed in an efficient model selection process [26]. OFR with the assistance of local regularization, that is, the LROFR algorithm may further enhance the capacity for model selection to produce a sparser model with good generalization performance [27]. The LROFR algorithm will be used for the identification of the proposed model. The LROFR algorithm is briefly reviewed as follows.

Collecting the observation and the alpha basis function, model (12) can be written in a matrix form as:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{\theta} + \mathbf{e} \tag{13}$$

where $\mathbf{y}$ is the output vector, $\mathbf{\Phi}$ is the regression matrix and $\mathbf{\theta}$ is associated parameter vector to be determined.

In model (13), $\mathbf{y}$ denotes output data. The column vectors in $\mathbf{\Phi} = [\mathbf{\Phi}_1, \mathbf{\Phi}_2 \cdots, \mathbf{\Phi}_{n_m}]$ are candidate dictionary consisted of delay of $y(k)$, $u(k)$ and alpha basis function according to (10)-(12). Parameter vector $\theta = [\theta_1, \theta_2, \cdots, \theta_{n_m}]^T$. Let an orthogonal decomposition of the regression matrix $\mathbf{\Phi}$:

$$\mathbf{\Phi} = WA \tag{14}$$

where $W = [w_1, \cdots, w_{n_m}]$ denotes an orthonormal matrix by column, A is an upper triangular matrix with unit diagonal elements:

9

$$A = \begin{pmatrix} 1 & a_{1,2} & \cdots & a_{1,n_m} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n_m-1,n_m} \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

The regression model (14) can alternatively be expressed as:

$$y = Wg + e \tag{15}$$

where $g = [g_1, \cdots, g_{n_m}]^T$ is the orthogonal regression weight vector satisfy:

$$g = A\theta \tag{16}$$

The set Φ is commonly superfluous and over-complete, so we need to choose some important terms and reckon the coefficients adapted to terms chosen. Follow [27], in order to select the l th regressor from all n_m candidates, what we should do first at l th stage is setting a small positive number T_z to specify the zero threshold and automatically avoid any ill-conditioning or singular problem. That is to say, for $l \le j \le n_m$, if $(\Phi_j^{(l-1)})^T \Phi_j^{(l-1)} < T_z$, the j th candidate is not considered. Then calculate:

$$g_l^{(j)} = (\Phi_j^{(l-1)})^T y^{(l-1)} \Big/ \left((\Phi_j^{(l-1)})^T \Phi_j^{(l-1)} + \lambda_j \right) \tag{17}$$

$$[\text{err}]_l^{(j)} = (g_l^{(j)})^2 \left((\Phi_j^{(l-1)})^T \Phi_j^{(l-1)} + \lambda_j \right) \Big/ y^T y \tag{18}$$

where λ denotes the regularization coefficient and err presents the error reduction ratio which is defined in [26], Next find:

$$[err]_l = [err]_l^{j_l} = \max\{[err]_l^{(j)}, l \le j \le n_m\} \tag{19}$$

and swap the j_l th and l th column each other in $\Phi^{(l-1)}$, A and λ respectively. At last, perform the orthogonalization and computation as indicated below to obtain the l th row of A as well as update $\Phi^{(l)}, g_l, y^{(l)}$, specially $\Phi_j^{(0)} = \Phi_j, y^{(0)} = y, 1 \le j \le n_m$:

$$\left. \begin{aligned} w_l &= \Phi_l^{(l-1)} \\ a_{lj} &= w_l^T \Phi_j^{(l-1)} \Big/ w_l^T w_l \\ \Phi_j^{(l)} &= \Phi_j^{(l-1)} - a_{lj} w_l \end{aligned} \right\} \quad 1 \le l \le n_m - 1, \quad l+1 \le j \le n_m \tag{20}$$

$$\left. \begin{aligned} g_l &= w_l^T y^{(l-1)} \Big/ (w_l^T w_l + \lambda_l), \\ y^{(l)} &= y^{(l-1)} - g_l w_l, \end{aligned} \right\} \quad 1 \le l \le n_m \tag{21}$$

Once the stop criterion (22) is reached, the selection will be stopped with a subset model which is composed of n_s significant regressors.

$$1 - \sum_{l=1}^{n_s} [err]_l < \rho \tag{22}$$

where $0 \le \rho \le 1$ is a chosen tolerance. It should be noted that the selection of ρ is very significant. In the paper, we applied the optimal number of regression terms determined by the APRESS criterion [2] to obtain an optimal threshold ρ based on formula (22) indirectly. Specifically, the model selection will stop when the model structure is the best. After that, we used the optimal threshold

determined by APRESS to stop the model selection. The parameter $\boldsymbol{\theta}$ can then be derived by rearranging equation (16).

The process demonstrated above actually is a standard OFR method if $\lambda = 0$, if not, λ can be renewed by two steps:

Step 1: assign the same small enough positive value such as 0.001 to all λ and a tolerant number, and use the procedure described in (17)-(22) to select a subset model with n_s terms.

Step 2: renew λ using (23) with $n_m = n_s$ as follow. If λ remains sufficiently unchanged in two successive iterations or a pre-set maximum iteration number is reached, stop; otherwise, go to Step 1.

$$
\left.\begin{array}{l}
\lambda_i^{new} = \dfrac{\gamma_i}{N - \gamma} \dfrac{e^T e}{g_i^2}, \\[2ex]
\gamma_i = w_i^T w_i \big/ (w_i^T w_i + \lambda_i^{old}), \\[2ex]
\gamma = \displaystyle\sum_{i=1}^{n_m} \gamma_i
\end{array}\right\} \quad 1 \le i \le n_m
\tag{23}
$$

Combined with wavelet basis function decomposition, the system identification method for the new TV-NARX model can be summarized as follow:

Step 1: The output and input data of TV-NARX model should be expressed as model (4) first.

Step 2: Complete the candidate dictionary definition and expand the time-varying coefficients with alpha basis function according to (10)-(12), finally get the model (13) to be identified.

Step 3: Initialize the λ and select the significant n_s terms and corresponding parameters to represent the system to be decided employing (14)-(23).

4. Numerical studies and application to real nonstationary signal

In this section, a numerical experiment is first employed to demonstrate the advantages of asymmetric basis function in fitting arbitrary changing signals over their symmetric counterparts. Linear and nonlinear numerical examples and real applications in modelling of physiological signals were used to demonstrate the efficiency of the new TV-NARX model and the associated system identification algorithms. In the numerical studies, a modified 10-fold cross validation method and the model prediction output of the system were applied to validate the obtained model and the performance of the proposed model is compared with state-of-the-art time-varying system identification methods. The model sparsity and prediction error of the time-varying coefficients are used to evaluate the new time-varying system identification method. Modelling of real EMG signal showed that the new model is powerful to characterize real physiological signals and can also provide spike train as the extra information for explaining muscle activation. The application of the new TV-NARX method in the clinical detection of freezing of gait of Parkinson's disease showed that the new method provides an explicit model structure and can be used for further model based analysis. This is the advantage of the new method over traditional neural networks which only provide the input-output description.

4.1 Example 1: A numerical example

In the part, we performed a numerical experiment to prove that the asymmetric alpha basis function inspired by the neural dynamic has a better performance than the symmetric basis function in fitting randomly changing signals. We constructed a set of basis functions with different asymmetricities using the formula (8) with the two parameters a and b ranging from 1 to 70 to control the

asymmetricity of the basis. Shifting each basis function yielded a term dictionary for signal fitting, where all the terms in the dictionary are of same shape/asymmetricity but different phases. One hundred uniform random signals were generated and the empirical mode decomposition (EMD) was applied to smoothen the signals. The denoised signals were used as the signal set for evaluating the descriptive ability of the basis functions with different asymmetricities. For computational simplicity, we used the OFR rather than the LROFR algorithm to fit the signals. For each dictionary, fixed the model size were used and the average R squared value [28] over the 100 signals were used to evaluate the efficiency of the basis function in tracking the sequence. The results are shown in Fig 3.

The results showed that the models with asymmetric terms had better performance than the symmetric ones which are located along $a=b$ in the plots and the best model performance achieved along the two ridges in Fig 3(a). It is easy to check that the basis functions with parameters along each of the ridges have similar asymmetricity. In other words, the models with greater R-squared value will use less terms to achieve the same accuracy, namely, a sparser model structure.

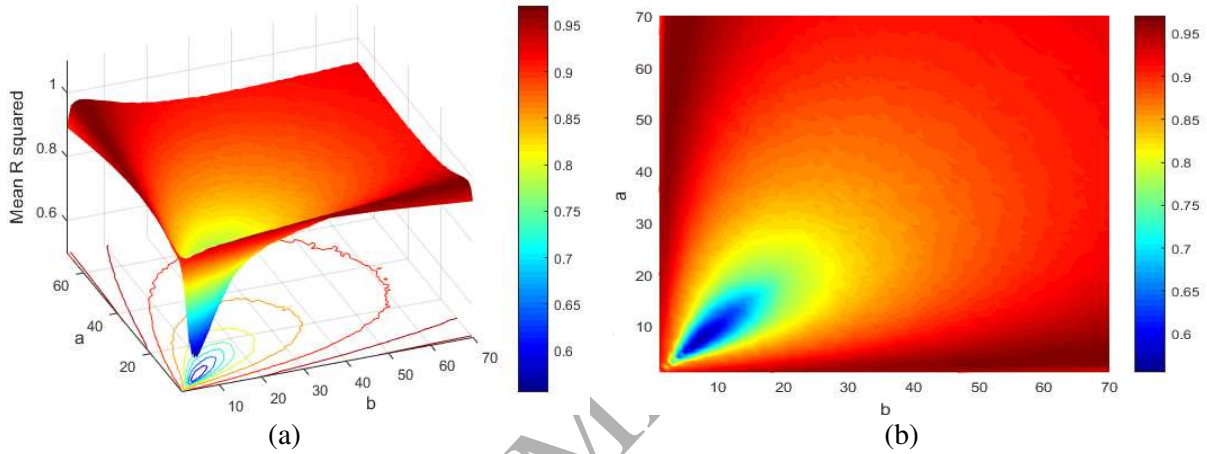(a)   (b)

Figure 3. (a). The mean R squared at the combination of a and b ranging from 1 to 70, where the ridge represents the maximum of mean R squared; (b). The contour plot of the mean R squared at the combination of a and b ranging from 1 to 70, where the different color represents the corresponding value of the mean R squared.

The experimental results indicate that asymmetric basis functions have a better descriptive ability in capturing the signal features with fewer model terms than the symmetric basis functions do. It is straightforward to conclude that a sparse model structure can be obtained when an appropriate asymmetric basis function is used to replace the symmetric basis function to fitting the time-varying parameters in a TV-NARX model.

4.2 Example 2: A linear time-varying system

Consider the system as follows:

$$\begin{cases} v(k) = a_1(k)y(k-1) + a_2(k)y(k-2) + b_1(k)u(k-1) + b_2(k)u(k-2) \\ y(k) = v(k) + \dfrac{1}{1 - 0.8z^{-1}}e(k) \end{cases} \tag{24}$$

where the input $u(k)$ is a Pseudo-Random Binary Sequence (PRBS); $e(k)$ is Gaussian distributed noise with zero mean and variance 0.001, namely, with a 20dB SNR. The time-varying parameters are defined as:

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

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

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

$$b_2(k) = 0.6, 1 \le k \le N,$$

(25)

respectively, where $N = 2048$ is the number of samples.

In order to identify a sparse model structure, an over-complete dictionary which consists of time-shifted alpha basis function is constructed. Assuming the support length of the basic alpha basis function $\phi(k, 0, \sigma | a, b)$, then the dictionary is defined as $D_\alpha = \left\{ \phi(k, m, \sigma | a, b) \right\}_{m=-(s_\alpha - 1)}^{(s_\alpha - 1)}$, $k = 1, \cdots, N$ is s_α, where, for simplicity, all the basis functions are set to have the same shape parameters $a = 3$, $b = 7$, and scale σ. The scale was optimised in the identification by comparing the model performance under different scale parameters. This dictionary composes of an over-complete frame because there are a total number of $N + s_\alpha - 1 > N$ vectors in the dictionary. Dictionaries with more basis functions with different scales and shape parameters can be constructed. However, in this and following examples we restricted the dictionary to the simplest setting to illustrate the powerful descriptive ability of the new asymmetric basis function. The dictionary of process terms is constructed as $D_p = \left\{ y(k - d) \right\}_{d=1}^{4}$. Therefore, the dictionary for the associated time invariant problem will be $D = \left\{ \phi(k, m, \sigma | a, b) y(k - d) \right\}_{d,m}$, $1 \le d \le 4$, $-s_\alpha + 1 \le m \le s_\alpha - 1$.

Collecting the data and system (24) is identified using both OFR and LROFR algorithms. Results are shown in Fig 4 (c) and (d). The model performance is evaluated using the mean absolute error of the reconstructed time varying parameters, which is defined as

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

(26)

where $c(k)$ and $\hat{c}(k)$ represent the original and identified model parameters, respectively.

In order to illustrate the efficiency of the new proposed model. The results are compared with those of state-of-the-art methods [1, 4]. System (24) was also identified using multiple wavelet TVNAX model where the coefficients were approximated using symmetric B-spline wavelet basis function. The model dictionaries were organized in two different ways. In the first case, the wavelet basis functions constructed by time-shifting a basis 4th order B-spline wavelet as we did in the new proposed method. The results for this kind of model is shown in Fig 4 (a). Secondly, a multiple discrete B-spline wavelet basis was constructed including 3rd, 4th, and 5th B-spline wavelets with different scales. For more details, please refer to [4]. The results of this kind of model are shown in Fig 4(b). The four methods are summarized in Tab 1 and a comprehensive comparison of these four methods under different SNR levels are given in Tab 2.

Table 1. Summary of four TVNARX identification methods

Methods	Basis functions	Algorithm	Model complexity*

Method1	Time shifted 4th order B-spline wavelet functions	OFR	172
Method2	Multiple B-spline wavelet basis functions	OFR	133
Method3	Time shifted asymmetric alpha basis functions	OFR	124
Method4	**Time shifted asymmetric alpha basis functions**	**LROFR**	**115**

* The model complexity was determined by the APRESS criterion [2].

Results show that the asymmetric alpha basis function characterizes both sharp and smooth changes in the time-varying coefficients better than the symmetric B-spline wavelets even when the multi-wavelets basis with multi-order and scales were used. Fig 4(c) shows that OFR algorithm estimates may have undesired peaks when the coefficients change sharply while this problem has successfully been solved by introducing the local regularisations in Fig 4(d).

(a) OFR with Time-shifted 4th order B-spline wavelet functions

(b) OFR with Multiple B-spline wavelet basis functions

(c) OFR Time shifted asymmetric alpha basis functions

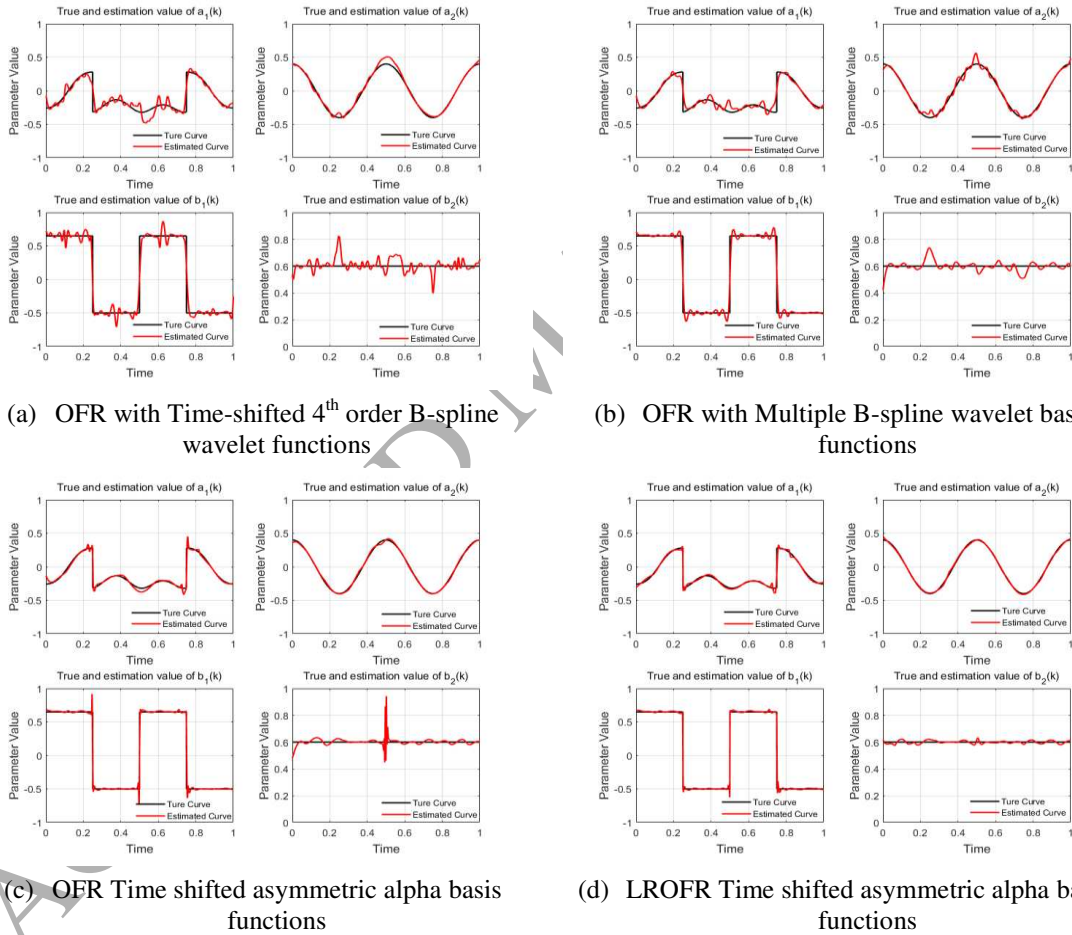(d) LROFR Time shifted asymmetric alpha basis functions

Figure 4. Identification results of TVARX defined by employing different methods. The bold black line indicates the true value of TV parameters, and the red line denotes the approximation value of the parameters.

In order to validate the robustness and effectiveness of the new method, the four system identification approaches are compared under three different noise level, that is, SNR equals 20dB, 15dB, and 10dB. The simulations were run 100 times for each case with random noise. The mean MAE value over the 100 simulations is shown in Tab 2. The hyper-parameters, such as the model complexity and the wavelet scales, in each method, were optimised to produce the best performance.

14

It can be observed that the proposed method with asymmetric alpha basis functions and LROFR algorithm is superior to the other methods using fewer terms and single wavelet scale. It is reasonable to believe that the new method may characterise the time-varying process better when multi-scale alpha basis function is used.

Model validation is essential and has been widely used for evaluating the efficiency of models or system identification algorithms. However, there are limited results on the model validation of nonstationary models. Many validation methods cannot directly be adopted for nonstationary cases. We conducted a modified 10-fold cross validation to validate the obtained models using the proposed methods. Specifically, 10% of the data were randomly sampled as the test data. The R squared and root mean square error (RMSE) defined in (27) of the system output were then computed and compared with those for the 90% training. Ten rounds modified cross validation were performed and the average of the result was shown in Tab 3. Results show that not an obvious difference in the two evaluation metrics between the training and test sets. In order to further validate the model performance, the model prediction output under 20dB noise was shown in Fig 5, where only the initial values of the time output and the identified time-varying coefficients were used to reproduce the system output. Both the cross-validation and model prediction output indicate excellent prediction power and efficiency of the obtained model.

$$RMSE\big(y(k)\big) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\big|y(k) - \hat{y}(k)\big|^2} \tag{27}$$

where $y(k)$ and $\hat{y}(k)$ represent the original and identified model output, respectively. N denotes the length of the observations.

Table 2. A comparison of methods performance for system (24)
with SNR=10, 15, 20dB

Approach	SNR(dB)	MAE of estimated parameters			
		$a_1(k)$	$a_2(k)$	$b_1(k)$	$b_2(k)$
OFR with Time shifted 4th order B-spline wavelet functions	10	0.0627	0.0412	0.0590	0.0429
	15	0.0475	0.0305	0.0449	0.0315
	20	0.0401	0.0248	0.0469	0.0258
OFR with Multiple B-spline wavelet basis functions	10	0.0574	0.0377	0.0471	0.0384
	15	0.0462	0.0287	0.0394	0.0293
	20	0.0404	0.0239	0.0361	0.0242
OFR Time shifted asymmetric alpha basis functions	10	0.0450	0.0292	0.0331	0.0310
	15	0.0283	0.0181	0.0207	0.0191
	20	0.0192	0.0120	0.0144	0.0131
LROFR Time shifted asymmetric alpha basis functions	10	**0.0439**	**0.0283**	**0.0303**	**0.0302**
	15	**0.0274**	**0.0174**	**0.0179**	**0.0174**
	20	**0.0177**	**0.0111**	**0.0116**	**0.0116**

where the bold font represents the best results.

Table 3. Average model prediction error by cross validation
for Example 2

SNR(dB)	Train dataset		Test dataset	
	R squared	RMSE	R squared	RMSE

10	0.9570	0.1118	0.9346	0.1324
15	0.9843	0.0659	0.9731	0.0832
20	0.9946	0.0386	0.9794	0.0716

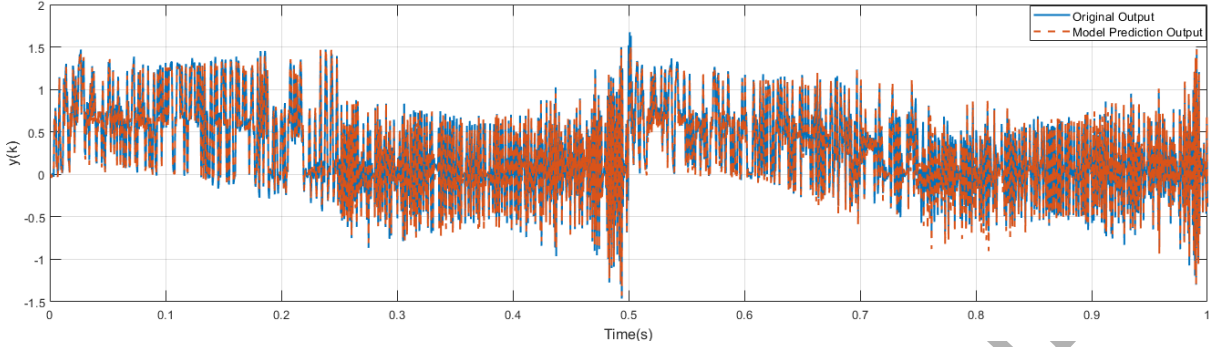Figure 5. The origin output and model prediction output under SNR = 20 dB in linear Example 2.

4.3 Example 3: A nonlinear time-varying system

In this example, a time-varying system with both input and output nonlinearities are identified to illustrate the efficiency of the proposed new method for identifying general nonlinear systems.

Consider the following system

$$\begin{cases} v(k) = c_{1,0}(1,k)y(k-1) + c_{2,0}(2,2,k)y(k-2)^2 + c_{0,1}(1,1,k)u(k-1)^2 + c_{0,2}(2,k)u(k-2) \\ y(k) = v(k) + \dfrac{1}{1-0.5z^{-1}}e(k) \end{cases} \tag{28}$$

where the input is a Gaussian random sequence and $u(k) \sim N(0,1)$, $1 \leq k \leq N$; $e(k)$ is Gaussian distributed with variance 0.075^2 and the SNR is 20dB. The time varying parameters are defined as:

$$c_{1,0}(1,k) = \begin{cases} 0.1, 1 \leq k \leq N/2, \\ 0.2, N/2+1 \leq k \leq N, \end{cases}$$

$$c_{2,0}(2,2,k) = -0.05 + 0.02\cos(10\pi k/N), 1 \leq k \leq N,$$

$$c_{0,1}(1,1,k) = \begin{cases} 0.8, 1 \leq k \leq 2N/5, \\ 1, 2N/5+1 \leq k \leq 7N/10, \\ 0.6, 7N/10+1 \leq k \leq N, \end{cases} \tag{29}$$

$$c_{0,2}(2,k) = \begin{cases} -0.6, 1 \leq k \leq N/5, \\ -0.8, N/5+1 \leq k \leq 4N/5, \\ -0.5, 4N/5+1 \leq k \leq N, \end{cases}$$

respectively.

Repeat the same process as in Example 1. The reconstructed time-varying parameters using the four system identification methods are shown in Fig 6. The model sizes used in the four different methods were 150, 120, 114 and 94, respectively. It can be observed that the new method with asymmetric alpha basis functions and LROFR produced the best results using the least model terms.

A comprehensive comparison of the four system identification under different SNR level is shown in Tab 4. Results show that the new asymmetric wavelet TV-NARX model with LROFR algorithm can produce the best estimation in all the four time varying parameters.

Table 4: A comparison of methods performance for TVNARX model (28)
with SNR=10, 15 ,20dB

| Approach	SNR(dB)	MAE of estimated parameters			
		$c_{1,0}(1,k)$	$c_{1,0}(2,2,k)$	$c_{0,1}(1,1,k)$	$c_{0,2}(2,k)$
OFR with Time shifted 4th order B-spline wavelet functions	10	0.0393	0.0165	0.0426	0.0638
	15	0.0282	0.0984	0.0275	0.0377
	20	0.0175	0.0054	0.0181	0.0228
OFR with Multiple B-spline wavelet basis functions	10	0.0492	0.0126	0.0381	0.0644
	15	0.0313	0.0075	0.0268	0.0396
	20	0.0188	0.0046	0.0198	0.0250
OFR Time shifted asymmetric alpha basis functions	10	0.0368	0.0140	0.0382	0.0585
	15	0.0249	0.0081	0.0230	0.0359
	20	0.0142	0.0044	0.0129	0.0198
LROFR Time shifted asymmetric alpha basis functions	10	**0.0334**	**0.0120**	**0.0318**	**0.0522**
	15	**0.0232**	**0.0070**	**0.0203**	**0.0325**
	20	**0.0127**	**0.0039**	**0.0117**	**0.0182**

where the bold line represents the better results.
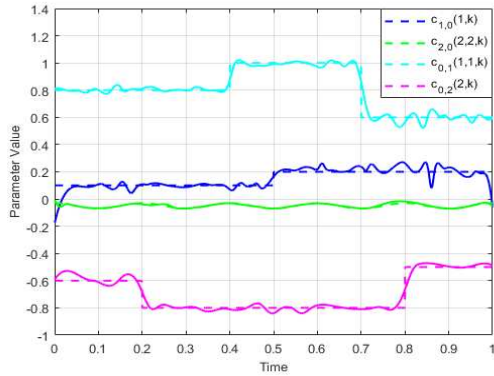
(a) OFR with Time-shifted 4th order B-spline wavelet functions
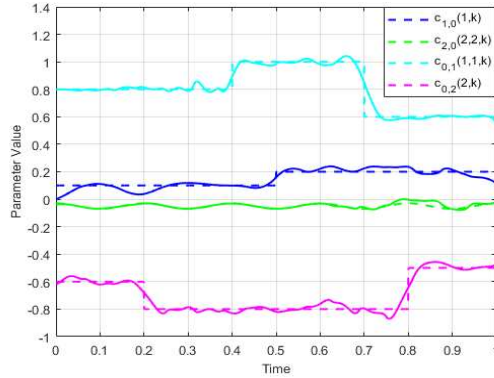
(b) OFR with Multiple B-spline wavelet basis functions
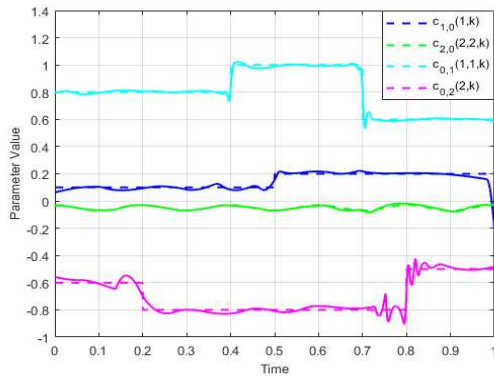
(c) OFR Time shifted asymmetric alpha basis functions
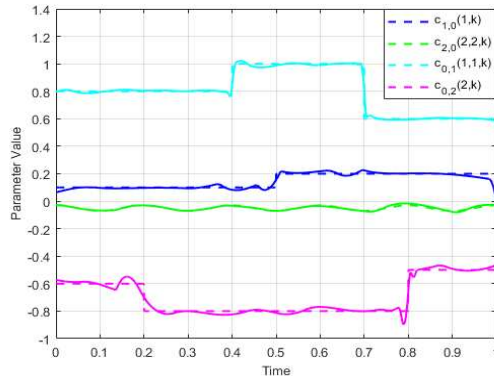
(d) LROFR Time shifted asymmetric alpha basis functions

Figure 6. Identification results of TVNARX defined by (28) using four approaches. The dotted line with different color indicates the true value of TV parameters and the real line with the corresponding color denotes the approximation value of the parameters.

The same model validation processes as Example 2 have been performed and the results were shown in Tab 5 and Fig 7. It can clearly be observed that the proposed model has excellent performance in both identifying the model structure and capturing the changes in parameters.

Table 5. Average model prediction error by cross validation for Example 3

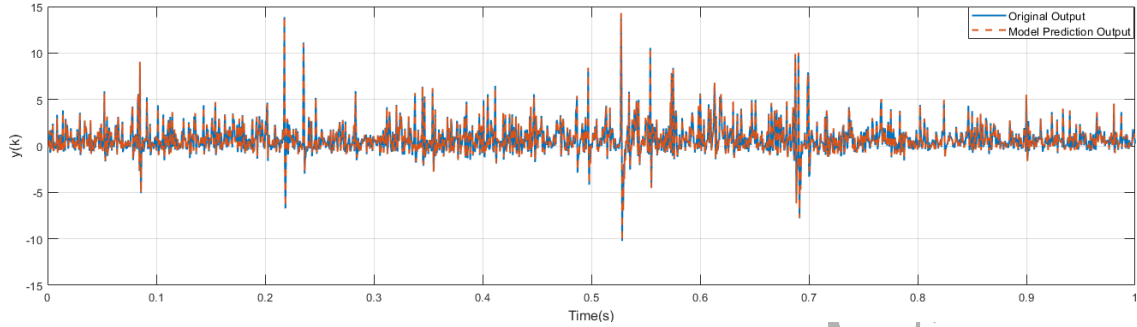| SNR(dB)	Train dataset		Test dataset	
	MAE	RMSE	MAE	RMSE
10	0.9211	0.4284	0.9041	0.4750
15	0.9696	0.2410	0.9616	0.2737
20	0.9924	0.1364	0.9901	0.1529

Figure 7. The origin output and model prediction output under SNR = 20 dB in nonlinear Example 3.

Both linear and nonlinear examples showed that the proposed TV-NARX model which inspired by neuron dynamics is more powerful in the characterisation of the time-varying processes than its counterparts with symmetric wavelet basis functions. This, to some extent, evidenced the adaptive functions of the alpha basis function shape PSCs in neuron dynamics.

4.4 Example 4: Application to real nonstationary EMG signal identification

Previous numerical studies show the efficiency of the new system identification method. In this part, the TV-NARX model and the associated algorithm will be used to investigate real time-varying nonlinear EMG signals. It can be seen that the changes in the signal frequency will be reflected in the time-varying coefficients and some extra information of the spike train can be obtained for the study of skeletal muscle activation.

It is known that the muscle contraction force is controlled by a group motion units. The generated muscle force depends on the activation of the involved motion units, that is, the temporal spike train of motor neurons [15]. The study about how muscle activation is related to the EMG measurement is crucial for evaluating the correctness of neuro-musculoskeletal models [29, 30].

The EMG data were recorded from the left leg anterior muscle of a patient with Parkinson's disease at Beijing Tiantan Hospital, China. The patient was at rest. Written informed consent was signed and procedures were approved by the Beijing Tiantan Hospital Ethics Committee. A 10s data is randomly selected from a total of 5 minutes of recording with a sampling rate of 500 Hz. Considering the range of tremor frequency in Parkinson's disease is usually 6-8 Hz, the origin EMG signal was low-pass filtered with a cutoff frequency 15 Hz. The data are shown in Fig 8 as the blue solid line.
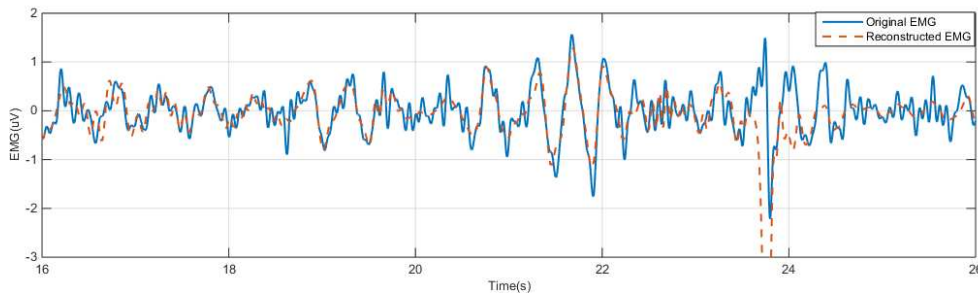

18

Figure 8. A comparison of the real EMG data and reconstructed data established by TV-NARX model (31), where the bold black line denotes the original EMG signal while the red line is the fitting line.

The initial guess of the TV-NARX model structure was chosen as

$$y(k) = \sum_{i=1}^{4} c_{1,0}(i,k)y(k-i) + \sum_{i=1}^{4}\sum_{j=1}^{4} c_{2,0}(i,j,k)y(k-i)y(k-j) + c_{0,0}(k) + e(k) \qquad (30)$$

including 15 candidates process terms. The asymmetric alpha basis functions were used to fit the time-varying parameters. The LROFR algorithm is adopted to select the significant terms. Results showed that only linear terms are significant in the description of the EMG dynamics and the obtained model is given as (31).

$$y(k) = \sum_{i=1}^{4} c_{1,0}(i,k)y(k-i) + e(k) \qquad (31)$$

A total number of 350 alpha basis functions were used to fit the time-varying parameters. The reconstruction of the parameters are shown in Fig 9 and the model prediction output is represented by the red broken line in Fig 8. The corresponding spike train which leads to the parameter fluctuations can also be obtained from the model, which is shown in Fig 9 represented by the pulses. These spike trains provide extra information which may relate to the activation of the motor units which contribute to the muscle force.
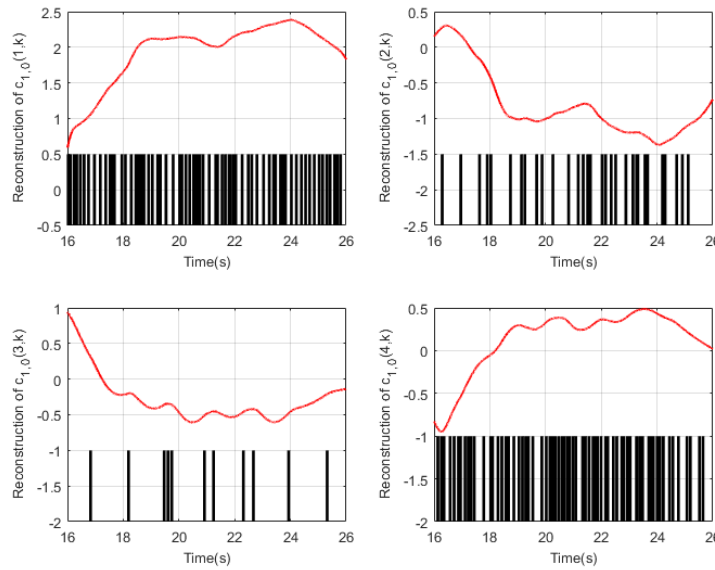

Figure 9. The time-varying coefficients in the model (31) identified using the new proposed method. The red curves show the time-varying parameters synthesised using the alpha basis function trains which are presented as the black pulses.

4.5 Example 5: Application in the detection of freezing of gait in Parkinson's disease

In this part, the proposed TV-NARX model with asymmetric alpha basis function expansion was used to modelling the nonstationary gait data of Parkinson's disease where the accelerations in Daphnet Freezing of Gait Dataset were used [31]. A typical 4.5s vertical accelerometion at the ankle including both normal locomotion and freezing of gait were low pass filtered and the preprocessed data are shown in Fig 10(a).

By selecting the significant terms using the LROLR algorithm, a second order linear time varing model of form (32) was obtained

$$y(k) = \sum_{i=1}^{2} c_{1,0}(i,k)y(k-i) + e(k) \tag{32}$$

A total number of 89 alpha basis functions were selected to fit the two time-varying parameters and the reconstructed parameters was shown in Fig 10(b).

Unlike purely phenomenological models, TV-NARX models can provide an explicit model structure and can be used for further model based analysis. In this example, the obtained TV-NARX model was then used to estimate the time varying time-frequency spectrum of the gait data according to formula (33):

$$P_{TDS}(k,f) = \delta_e^2 \left| \frac{1}{1 - \sum_{i=1}^{2} c_{1,0}(i,k)e^{-j2\pi if/f_s}} \right|^2 \tag{33}$$

where $P_{TDS}(k,f)$ is the time-frequency spectral estimation value at continuous independent variables time k and frequency f, $\hat{c}_{1,0}(i,k)$ is the estimation of model parameters $c_{1,0}(i,k)$, δ_e^2 is the variance of model residual, f_s = 64Hz is the sampling frequency and $j = \sqrt{-1}$ denotes the imaginary part of a complex number.

Results have shown that the model based method is capable of providing a time-frequency spectrum with high resolution in both time and frequency domain [32]. The obtained time-frequency spectrum which characterises the variation of frequency components over time is shown in Fig 10(c).

A biomarker, namely, Freeze Index (FI) [33] for the detection of FOG can then be defined based on the time-frequency spectrum as the ratio of energy in freezing band (3-8Hz) and normal locomotion band (0.5-3Hz). The calculated FI using formula (34) is shown in Fig 10 (d).

$$FI(k) = \frac{\int_3^8 P_{TDS}(k,f)df}{\int_{0.5}^3 P_{TDS}(k,f)df} \tag{34}$$

The results in Fig 10(b,c,d) showed that the FI can accurately detect the occurrence of the FOG. The trend of parameters and time-frequency spectrum is consistent with the occurrence of FOG episodes in Fig 10(b,c). Based on the new proposed method a convenient framework can be constructed for discriminating FOG events easily and accurately.

(a)

(b)

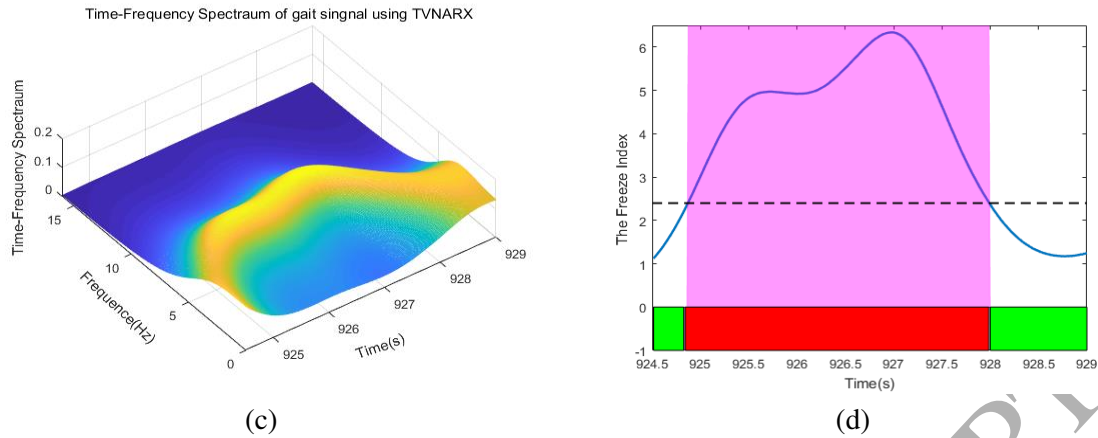Time-Frequency Spectrum of gait singnal using TVNARX

(c)

(d)
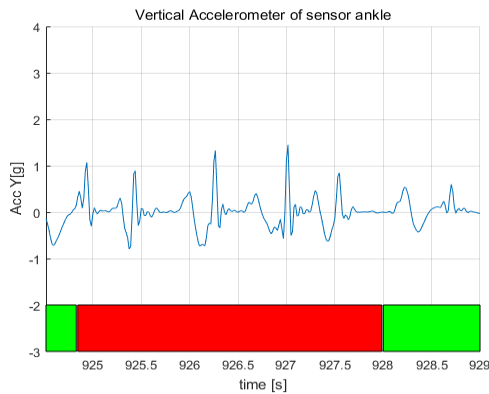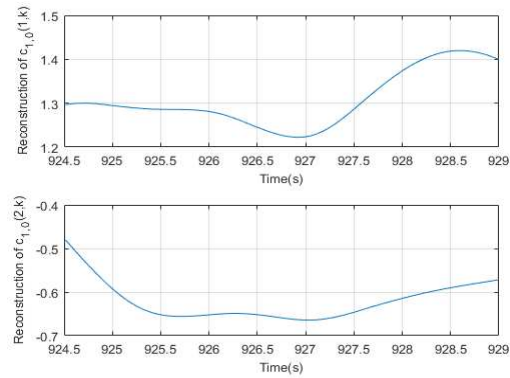
Figure 10. (a). The extracted vertical accelerometer data with a 10s window from subject 02, and the color bar at the bottom shows the ground truth labeled by experts, where red color denotes freeze of gait episodes and green no freeze of gait periods, respectively; (b). The reconstructed time-varying coefficients; (c). The time-frequency spectrum identified using the new proposed method; (d). The FI obtained by the new time-frequency spectrum. The black horizontal broken lines in the FI represented a proper threshold that discriminates the FOG events and normal movement. The magenta shadowed area demonstrated the FOG episodes identified by FI calculation method.

5. Conclusions and discussion

In this paper, we addressed the system identification of nonstationary signals with the TV-NARX model. Inspired by the neuronal dynamics, a new type TV-NARX model has been proposed by generalised the neuronal models. A new over-complete asymmetric basis function dictionary has proposed to sparsely represent time varying coefficients. Simulations show that the PSC shape basis can efficiently characterise the time varying systems with fast and slow changing coefficients even under low SNR. This may explain the unique time course of synaptic dynamics and the adaptive ability of the neural system to varying environments. Combining with the OFR algorithms, a general system identification method which is applicable to complex time varying dynamics is proposed.

The new TV-NARX model with the alpha basis function surpasses the one with symmetric basis functions in performance. This can be interpreted by the unique shape of the asymmetric alpha basis which is composed of a steep rising part and a relatively flat falling part so that it can characterise both sharp and smooth changes in a signal. Additionally, the decomposition process of time-varying data using asymmetric alpha basis functions is alike to the formation process of biological signals such as LFP through superposition of excitatory and inhibitory PSCs. This paper bridged the mathematical model and nonstationary biological processes, may be helpful to understand the neural and cognitive processes underlying the generation of biological signals, such as EEG, EMG.

The new proposed model can be used as the basic blocks for advanced spiking artificial neural networks which considers the spatiotemporal neural interactions. The time varying coefficients describes the effects of action potentials from other neurons and the output can be considered as the membrane potential. The NARX nonlinear process characterising the nonlinear firing process and the threshold can be described when the NARX dynamics is well designed. Connecting a group of the TV-NARX model, a new spiking neural network model NSNN can be constructed and this will be studied   in   future   work.

Acknowledgments

Appendix A

For the linear-in-the-parameter system:

$$y(k) = \sum_{i=1}^{M} p_i(k)\theta_i + e(k), \quad i = 1, \cdots N \tag{A.1}$$

where $y(k)$ is the output data or the term to regress upon, the $p_i(k)$ denotes functions of past inputs, outputs and noise. $e(k)$ is the modelling error, and θ_i is the time-invariant coefficients to be estimated. M and N are the size of candidate terms and the length of sample data, respectively. The formula (A.1) can be rewritten in a matrix form as:

$$Y = P\Theta + \Xi \tag{A.2}$$

where $P = [p_1 \ p_2 \ \cdots \ p_M]$ is the regression matrix. For the system, the standard OFR algorithm for the detection of the model structure and the estimation of the associated parameters, is listed below:

Step 1. Set $I_1 = \{1, 2, \cdots, M\}$; $\sigma = Y^T Y$;

    for $i = 1$ to M

        $w_i = p_i$;

        $ERR_i = (\langle Y, w_i \rangle^2 / \sigma \langle w_i, w_i \rangle) \times 100\%$;

        $a_{11} = 1$;

    end for

    $l_1 = \arg \max_{i \in I_1} \{ERR_i\}$;

    $w_1^0 = w_{l_1}$; $g_1^0 = \langle Y, w_1^0 \rangle / \langle w_1^0, w_1^0 \rangle$;

Step j. Set $j \geq 2$ and $\tau = 10^{-10}$;

    for $j = 2$ to M

        $I_j = I_{j-1} \backslash \{l_{j-1}\}$;

        for all $i \in I_j$

$$w_i = p_i - \sum_{k=1}^{j-1} (\langle p_i, w_k^0 \rangle / \langle w_k^0, w_k^0 \rangle) w_k^0 \ ;$$

$$ERR_i = (\langle Y, w_i \rangle^2 / \sigma \langle w_i, w_i \rangle) \times 100\% \ ;$$

        end for (end loop for i)

        $J_j = \{\arg_{i \in I_j} (w_i^T w_i) < \tau\}$; $I_j = I_j \backslash J_j$;

        $l_j = \arg \max_{i \in I_j} \{ERR_i\}$;

        $w_j^0 = w_{l_j}$; $g_j^0 = \langle Y, w_j^0 \rangle / \langle w_j^0, w_j^0 \rangle$;

        $a_{jj} = 1$;

for $k = 1$ to $j - 1$

$$a_{kj} = \langle p_{lj}, w_k^0 \rangle / \langle w_k^0, w_k^0 \rangle \ ;$$

end for (end loop for k)

end for (end loop for j)

References

[1] Y. Li, H. Wei, S.A. Billings, Identification of Time-Varying Systems Using Multi-Wavelet Basis Functions, IEEE Transactions on Control Systems Technology, 19 (2011) 656-663.

[2] S.A. Billings, H.L. Wei, An adaptive orthogonal search algorithm for model subset selection and non-linear system identification, International Journal of Control, 81 (2008) 714-724.

[3] Y. Li, M. Lei, Y. Guo, Z. Hu, H. Wei, Time-Varying Nonlinear Causality Detection Using Regularized Orthogonal Least Squares and Multi-Wavelets With Applications to EEG, IEEE Access, 6 (2018) 17826-17840.

[4] Y. Li, W. Cui, Y. Guo, T. Huang, X. Yang, H. 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, 29 (2018) 2960-2972.

[5] J. Chen, B. Huang, F. Ding, Y. Gu, Variational Bayesian approach for ARX systems with missing observations and varying time-delays, Automatica, 94 (2018) 194-204.

[6] J. Chen, F. Ding, Y. Liu, Q. Zhu, Multi-step-length gradient iterative algorithm for equation-error type models, Systems & Control Letters, 115 (2018) 15-21.

[7] S.A. Billings, Nonlinear System Identification : NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains (John Wiley & Sons Ltd, Hoboken, New Jersey, 2013).

[8] S. Xu, Y. Li, T. Huang, R. Chan, A Sparse Multiwavelet-Based Generalized Laguerre–Volterra Model for Identifying Time-Varying Neural Dynamics from Spiking Activities, Entropy, 19 (2017) 425.

[9] F. He, H.-L. Wei, S.A. Billings, Identification and frequency domain analysis of non-stationary and nonlinear systems using time-varying NARMAX models, International Journal of Systems Science, 46 (2013) 2087-2100.

[10] B.A. Olshausen, Highly overcomplete sparse coding, IS&T/SPIE Electronic Imaging, (SPIE2013), pp. 9.

[11] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature, 381 (1996) 607.

[12] M. Aharon, M. Elad, A. Bruckstein, K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, IEEE Transactions on Signal Processing, 54 (2006) 4311-4322.

[13] B.P. Bean, The action potential in mammalian central neurons, Nature Reviews Neuroscience, 8 (2007) 451.

[14] W. Rall, Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input, Journal of neurophysiology, 30 (1967) 1138-1168.

[15] J.C. Calderón, P. Bolaños, C. Caputo, The excitation–contraction coupling mechanism in skeletal muscle, Biophysical Reviews, 6 (2014) 133-160.

[16] B. Schrauwen, D. Verstraeten, J.V. Campenhout, An Overview of Reservoir Computing: Theory, Applications and Implementations, Proceedings of the 15th European Symposium on Artificial Neural Networks. p. 471-4822007), pp. 471-482.

[17] Y. Zheng, J.J. Luo, S. Harris, A. Kennerley, J. Berwick, S.A. Billings, J. Mayhew, Balanced excitation and inhibition: model based analysis of local field potentials, Neuroimage, 63 (2012) 81-94.

[18] H.R. Wilson, J.D. Cowan, A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue, Kybernetik, 13 (1973) 55-80.

[19] M.A. Wilson, J.M. Bower, The simulation of large-scale neural networks, in: K. Christof, S. Idan (Eds.) Methods in neuronal modeling, (MIT Press, 1989), pp. 291-333.

[20] A. Destexhe, Z.F. Mainen, T.J. Sejnowski, Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism, Journal of Computational Neuroscience, 1 (1994) 195-230.

[21] N.F. Rulkov, Modeling of spiking-bursting neural behavior using two-dimensional map, Physical Review E, 65 (2002) 041922.

[22] Y. Guo, Y. Zhao, D. Coca, S.A. Billings, A simple scalar coupled map lattice model for excitable media, International Journal of Bifurcation and Chaos, 21 (2011) 3277-3292.

[23] Y. Guo, Y. Zhao, S.A. Billings, D. Coca, R.I. Ristic, L.L. Dematos, Identification of excitable media using a scalar coupled map lattice model, International Journal of Bifurcation and Chaos, 20 (2010) 2137-2150.

[24] D.F. Walnut, Smooth, Compactly Supported Wavelets, in: D.F. Walnut (Ed.) An Introduction to Wavelet Analysis, (Birkhäuser Boston, Boston, MA, 2004), pp. 249-285.

[25] G. A. A. Araújo, H. M. de Oliveira, G. , Compactly supported one-cyclic wavelets derived from Beta distributions, Journal of Communication and Information Systems, 20 (2005) 105-111.

[26] S.A. Billings, S. Chen, M.J. Korenberg, Identification of MIMO non-linear systems using a forward-regression orthogonal estimator, International Journal of Control, 49 (1989) 2157-2189.

[27] S. Chen, Local regularization assisted orthogonal least squares regression, Neurocomputing, 69 (2006) 559-585.

[28] L.T. Skovgaard, Applied regression analysis. 3rd edn. N. R. Draper and H. Smith, Wiley, New York, 1998. No. of pages: xvii+706. Price: £45. ISBN 0-471-17082-8, Statistics in Medicine, 19 (2000) 3136-3139.

[29] T.S. Buchanan, D.G. Lloyd, K. Manal, T.F. Besier, Neuromusculoskeletal Modeling: Estimation of Muscle Forces and Joint Moments and Movements From Measurements of Neural Command, Journal of applied biomechanics, 20 (2004) 367-395.

[30] D.G. Thelen, F.C. Anderson, Using computed muscle control to generate forward dynamic simulations of human walking from experimental data, Journal of biomechanics, 39 (2006) 1107-1115.

[31] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J.M. Hausdorff, N. Giladi, G. Troster, Wearable Assistant for Parkinson's Disease Patients With the Freezing of Gait Symptom, IEEE Transactions on Information Technology in Biomedicine, 14 (2010) 436-446.

[32] Y. Li, M.-L. Luo, K. Li, A multiwavelet-based time-varying model identification approach for time–frequency analysis of EEG signals, Neurocomputing, 193 (2016) 106-114.

[33] S.T. Moore, H.G. MacDougall, W.G. Ondo, Ambulatory monitoring of freezing of gait in Parkinson's disease, Journal of neuroscience methods, 167 (2008) 340-348.
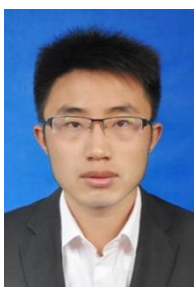
Yuzhu Guo
Department of Automation Sciences and Electrical Engineering, Beihang University
Beijing, China
Email: yuzhuguo@buaa.edu.cn.

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

Lipeng Wang
Department of Automation Sciences and Electrical Engineering, Beihang University
Beijing, China
Email: wlptjwx@buaa.edu.cn.

Lipeng Wang received his bachelor degree from Tianjin University of Technology in 2017. Now he is a graduate student in Department of Automation Science and Electrical Engineering at Beihang University, Beijing, China. His main research interests include time series analysis, machine learning, NARMAX methods, brain-inspired intelligence and computing, and bioinformatics.

Yang Li
Department of Automation Sciences and Electrical Engineering, Beihang University

Beijing, China
Email: liyang@buaa.edu.cn.

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.

Jingjing Luo
Academy of Engineering and Technology Handan Campus, Fudan University
Shanghai, China
Email: luojingjing@fudan.edu.

Jingjing Luo received the BEng degree from Zhejiang University, Hangzhou, China, and the Ph.D. degree from The University of Sheffield, Sheffield, U.K., in 2014. She is currently a Research Fellow in Academy of Engineering and Technology at Fudan University, Shanghai, China. Her current research interests include biomedical signal processing, neurorehabilitation, human machine interaction and intelligent medical system.

Kailiang Wang
Department of functional neurosurgery, Beijing Neurosurgical Institute, Capital Medical University
Beijing, China
Email: wkl_20080113@126.com

Kailiang Wang，as a M.D. student in Beijing neurosurgical institute，majored at functional neurosurgery，especially in the field of deep brain stimulation for movement disorders.

S. A. Billings
Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK
Email: s.billings@sheffield.ac.uk

Stephen A. Billings is a professor in the Department of Automatic Control and Systems Engineering, University of Sheffield, UK and leads the Signal Processing and Complex Systems research group. His research interests include system identification and information proces- sing for nonlinear systems, narmax methods, model validation, prediction, spectral analysis, adaptive systems, nonlinear systems analysis and design, neural networks, wavelets, fractals, machine vision, cellular automata, spatio-temporal systems, fMRI and optical imagery of the brain, synthetic biology and related fields.

Lingzhong Guo
Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK
Email: l.guo@sheffield.ac.uk

Lingzhong Guo received both the BSc and MSc degrees in mathematics in China, and the PhD degree in Bristol Robotic Laboratory, the University of the West of England, UK. He is currently a lecturer in the Department of Automatic Control and Systems Engineering, the University of Sheffield, UK. His research interests include identification and analysis of nonlinear dynamic systems with applications in biomedical system.