Model Estimation of Cerebral Hemodynamics Between Blood Flow and Volume Changes: A Data-Based Modeling Approach

Hua-Liang Wei*, Ying Zheng, Yi Pan, Daniel Coca, Liang-Min Li, J. E. W. Mayhew, and Stephen A. Billings

***Abstract*—It is well known that there is a dynamic relationship between cerebral blood flow (CBF) and cerebral blood volume (CBV). With increasing applications of functional MRI, where the blood oxygen-level-dependent signals are recorded, the understanding and accurate modeling of the hemodynamic relationship between CBF and CBV becomes increasingly important. This study presents an empirical and data-based modeling framework for model identification from CBF and CBV experimental data. It is shown that the relationship between the changes in CBF and CBV can be described using a parsimonious autoregressive with exogenous input model structure. It is observed that neither the ordinary least-squares (LS) method nor the classical total least-squares (TLS) method can produce accurate estimates from the original noisy CBF and CBV data. A regularized total least-squares (RTLS) method is thus introduced and extended to solve such an error-in-the-variables problem. Quantitative results show that the RTLS method works very well on the noisy CBF and CBV data. Finally, a combination of RTLS with a filtering method can lead to a parsimonious but very effective model that can characterize the relationship between the changes in CBF and CBV.**

***Index Terms*—Autoregressive with exogenous input model (ARX), cerebral blood flow (CBF), cerebral blood volume (CBV), parameter estimation, regularization, system identification, total least squares (TLS).**

I. Introduction

IT IS WELL known that there is a dynamic relationship between cerebral blood flow (CBF) and cerebral blood volume (CBV) [1]–[7]. With the increasing applications of positron emission tomography and functional MRI (fMRI), where the understanding of the blood-oxygen-level-dependent (BOLD) signal plays a key role, it is becoming increasing important to establish an accurate quantitative description of the dynamics relating CBF and CBV. The quantitative description of the relationship between changes in blood flow (volume flux per unit time through a tissue volume element) and blood volume was first presented by Grubb *et al.* [1], where it was suggested that the relationship between the two variants can be described using a function obeying a simple power law, i.e., $\mathrm{CBV} \propto \mathrm{CBF}^{\alpha}$, with α being a constant. This has been extensively applied when modeling the hemodynamic response to activation. However, due to the fact that this power law relationship has been derived merely based on steady-state measurements, the generalization and application to activation scenarios involving transient changes may not be valid. Buxton *et al.* [2] developed a biomechanical differential equation model, called the Balloon model, to describe how evoked changes in blood flow were transformed into a BOLD signal. Mandeville *et al.* [3] studied the relationship between the blood flow and volume changes and presented a model in terms of resistance and capacitance in the context of the standard windkessel theory. Friston *et al.* [8], [9] proposed a unified alternative representation on the basis of Volterra kernel theory, by combining system identification and model-based approaches, to describe nonlinear responses in fMRI including the modeling of the hemodynamic relationship between CBF and CBV.

Due to the complexity of the inherent neural hemodynamics for which no or very limited *a priori* information about the biophysical mechanisms (the model structure and the associated model parameters) is available, analytical or theoretical modeling approaches alone may not be adequate to obtain sufficiently reliable mathematical models to describe cerebral hemodynamics between CBF and CBV. As an alternative, empirical and data-based modeling approaches that make use of both biophysical observations and identification and information techniques provide a complementary but very powerful tool for modeling such complex systems. Regression models, including the general linear model, autoregressive with exogenous model (ARX), nonlinear regression and nonlinear network models, are among the most popular classes of representations for characterizing and understanding the dynamics of fMRI responses and related signals, see for example [10]–[17] and the references therein.

Among the existing modeling techniques, linear-in-the-parameters regression models, which include the ordinary linear regression model as a special case, are an important class of representation for signal processing and system identification. One obvious advantage of employing linear-in-the-parameters models is that they are easy to operate, because compared with nonlinear-in-the-parameters models, such models are easier to interpret physically, simpler to analyze mathematically, and

Manuscript received August 12, 2008; revised November 13, 2008. First published January 23, 2009; current version published June 10, 2009. This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC), U.K. *Asterisk indicates corresponding author.*

*H.-L. Wei is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, U.K. (e-mail: w.Hualiang@sheffield.ac.uk).

Y. Zheng, Y. Pan, and J. E. W. Mayhew are with the Department of Psychology, University of Sheffield, Sheffield S10 2TP, U.K. (e-mail: ying.zheng@sheffield.ac.uk; y.pan@sheffield.ac.uk; j.e.mayhew@sheffield.ac.uk).

D. Coca, L.-M. Li, and S. A. Billings are with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, U.K. (e-mail: d.coca@sheffield.ac.uk; l.li@sheffield.ac.uk; s.billings@sheffield.ac.uk).

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/TBME.2009.2012722

quicker to compute numerically using least-squares-based algorithms [18]. In the classical least-squares (LS) approach, which is the most commonly used method for solving linear regression problems, the measurements of the design matrix formed by the "input" variables (independent variables) are assumed to be "clean" or "noise-free" (no errors); or, the errors on the measurements of the independent variables are much smaller compared with those imposed on the "output" variables (dependent variables) and can therefore be ignored. In many cases, however, these assumptions may be unrealistic. When the classical LS approach is applied to solve linear regression problems, where these assumptions are violated, the resultant LS estimates for the associated model parameters are inevitably biased. To overcome this drawback of the ordinary LS algorithm, Golub and Van Loan [19], [20] developed an efficient numerical tool, called total LS (TLS), for solving linear regression problems, where the effects of errors on both the dependent variables and the independent variables (and thus, the design matrix) are taken into account. However, unlike the ordinary LS algorithm, where the solution can be written in a compact form, the application of the TLS algorithm involves nonlinear optimization for parameter estimation.

The central objective of this paper is to propose an empirical and data-based modeling approach that can produce an accurate but simple description of the relationship between the changes of CBF and CBV during brain activity. The associated modeling procedure involves several aspects, two of which focus on the following issues: how to determine the model structure and how to obtain accurate estimates of the model parameters by reducing the effects of the measurement errors that are inevitable in any real biomedical experiments. Following Occam's Razor (also known as the parsimonious principle) and by applying model structure detection and model validity test methods, for example, the orthogonal LS and statistical model validity test algorithms [21]–[34], the model structure can be determined effectively. As for the parameter estimation issue, as was discussed in the previous paragraph, the ordinary LS algorithm may produce biased estimates when it is directly applied to highly noisy experimental measurements. With this consideration, a regularized TLS (RTLS) method [35], implemented by using a simplex direct search optimization algorithm [36], is developed and adapted for model parameter estimation. By combining the parsimonious principle, the well-established model structure selection and model validity test methods, the regularized TLS algorithm and *a priori* information on the associated cerebral hemodynamics, parsimonious but effective models relating CBF and CBV can be obtained.

II. Data-Based Modeling Framework

A. NARX Model

It has been proved that under some mild conditions a discrete-time or discretized continuous-time dynamical system can be described by the following difference equation model [37], [38]:

$$y(n) = f(y(n-1), \ldots, y(n-p), u(n-1), \ldots, u(n-q)) + e(n) \quad (1)$$

where $u(n)$, $y(n)$, and $e(n)$ are the system input, output, and noise variables; p and q are the maximum lags in the input and output, respectively; and f is some unknown linear or nonlinear mapping. It is generally assumed that $e(n)$ is an independent identical distributed noise sequence. A commonly employed form of model (1) is the well-known nonlinear ARX (NARX) model [37]–[40] that can describe a wide range of nonlinear dynamic systems and includes several other linear and nonlinear model types, e.g., Volterra, Hammerstein, Wiener, and ARMAX models as special cases [41].

A generic form of the NARX model, with a nonlinear degree of order ℓ, is given as follows:

$$\begin{aligned} y(n) = c_0 &+ \sum_{i=1}^{d} c_i x_i(n) + \sum_{i=1}^{d}\sum_{j=1}^{d} c_{i,j} x_i(n) x_j(n) \\ &+ \cdots + \sum_{i_1=1}^{d}\sum_{i_2=1}^{d}\cdots\sum_{i_\ell=1}^{d} \\ &\times c_{i_1,i_2,\ldots,i_\ell} x_{i_1}(n) x_{i_2}(n)\cdots x_{i_\ell}(n) + e(n) \end{aligned} \quad (2)$$

where $d = p + q$ and

$$x_k(n) = \begin{cases} y(n-k), & 1 \le k \le p \\ u(n-(k-p)), & p+1 \le k \le p+q. \end{cases} \quad (3)$$

Practical applications have shown that NARX models, with a nonlinear degree of order $\ell \le 3$, can often provide satisfactory approximations for most dynamical systems. The widely used ARX model [42]–[44], as a special case of the NARX model (2), where $\ell = 1$ and $c_0 = 0$, is explicitly given by

$$y(n) = \sum_{i=1}^{p} a_i y(n-i) + \sum_{j=1}^{q} b_j u(n-j) + e(n). \quad (4)$$

The initial full linear-in-the-parameters model (2) contains a total of $M = [(p+q+\ell)!]/[\ell!(p+q)!]$ model terms, where the symbol "$(p+q)!$" indicates the factorial of the number $(p+q)$. Note that for large maximum lags p and q, the initial full model (2) may involve a great number of candidate model terms. However, experience shows that in most cases only a small number of significant model terms are necessary, and thus, should not be included in the final model to represent the underlying dynamics. Most candidate model terms are either redundant or make very little contribution to the system output and can thus be removed from the model. Several efficient model structure determination and model validity test methods have been developed over the last two decades [21]–[34].

Assume that a total of m significant model terms, denoted by $\{\phi_1(n), \phi_2(n), \ldots, \phi_m(n)\}$, have been selected from the library consisting of all the M candidate model terms. The selected m model terms can be used to form a parsimonious model

$$y(n) = \theta_1\phi_1(n) + \theta_2\phi_2(n) + \cdots + \theta_m\phi_m(n) + e(n) \quad (5)$$

where $\phi_k(n)$ denote a combination of the lagged versions of the input and output variables $u(n)$ and $y(n)$ (the constant may also be included). For example, for an NARX model with a nonlinear degree of order $\ell = 3$, $\phi_k(n)$ are then selected from the library $L = \{1\} \cup \{x_i(n) : 1 \le i \le d\} \cup \{x_i(n)x_j(n) : 1 \le i,$

$j \le d\} \cup \{x_i(n)x_j(n)x_k(n) : 1 \le i, j, k \le d\}$, where $x_k(n)$ are defined by (3). Classical linear LT type of algorithms may be applied to estimate the associated model parameters.

B. Regularized Total Least Squares

It is known that the classical LS algorithms and the standard statistical analysis for these algorithms require certain assumptions: the "input" (independent) variables, say $\phi_k(n)$ in (5), are measured without errors; or the errors imposed on the "input" variables are much smaller than those imposed on the "output" variables (dependent variables) say $y(n)$ in (5) and can therefore be ignored. For many cases, this assumption may not be satisfied, and the ordinary LS method will not work well. To solve this kind of errors-in-variables (EIV) problem, Golub and Van Loan [19], [20] developed the total TLS method. Over the last three decades, TLS methods have been successfully applied to solve a variety of EIV problems [45]–[48], including the applications to biomedical data modeling [49], [50]. In some cases, the TLS method alone, however, may not be effectively immune to the amplification effects of the noise for an ill-conditioned problem. To solve this problem, the RTLS method was proposed by combining the well-known Tikhonov regularization [51]–[54].

Taking the linear regression (linear-in-the-parameters regression) model (5) as an example, the RTLS estimate is stated as

$$\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} \left\{ \frac{||\mathbf{y} - \Phi\boldsymbol{\theta}||^2}{1 + \mu||\boldsymbol{\theta}||^2} + \lambda||\boldsymbol{\theta}||^2 \right\} \tag{6}$$

where $\mathbf{y} = [y(1), \ldots, y(N)]^T$, $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_m]^T, \Phi = [\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_m]$ with $\boldsymbol{\varphi}_k = [\phi_k(1), \ldots, \phi_m(N)]^T$ for $k = 1, 2, \ldots, m$, N is the number of available observations, and μ and λ are two adjustable parameters. Clearly, while the ordinary LS minimizes a sum of squared residuals, TLS minimizes a sum of weighted squared residuals with a penalized term formed by the square of the parameters. If $\mu = 0$, (6) reduces to the case of the Tikhonov regularization; if $\mu = 1$ and $\lambda = 0$, (6) reduces to TLS. In the present study, the adjustable parameter μ will be set to unity, i.e., $\mu = 1$, and the regulation parameter λ will be chosen by trial-and-error (see the example that follows for details).

Note that the solution to RTLS (6), with respect to the unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$, involves nonlinear optimization. Many nonlinear optimization approaches are available to solve such a nonlinear optimization problem. In this study, however, a simplex direct search optimization algorithm, proposed by Nelder and Mead [36], is applied to solve the nonlinear optimization problem. The Nelder–Mead method, first introduced by Nelder and Mead in 1965 and recently enhanced theoretically by Lagarias *et al.* [55], is a powerful direct "derivative-free" search algorithm, where neither the computation nor the approximation of derivatives or gradients are needed. The Nelder–Mead method has enjoyed enduring popularity. Of all the direct search methods, the Nelder–Mead simplex algorithm is the one most often found in numerical software packages [56].

C. Choosing Regularization Parameter

The determination of the regularization parameter λ in RTLS (6) is a significant but difficult issue, and there is no universal criterion on how to select the parameter for general dynamical modeling problems. Some empirical or *ad hoc* methods, however, may work quite well when choosing such regularization parameters [34]. This study suggests using a trial-and-error approach and the basic idea is as follows. Let $\hat{\boldsymbol{\theta}}^{(LS)}$ be the LS estimate for the model parameter vector $\boldsymbol{\theta}$ and $\hat{\sigma}^{(\mathrm{LS})}_{\mathrm{nmse}}$ be the normalized mean-square-errors calculated from the model with the LS estimate. A rule of thumb from our experience is to initially choose a number

$$\begin{aligned}\lambda_0 &= \frac{\hat{\sigma}^{(\mathrm{LS})}_{\mathrm{nmse}}}{1 + ||\hat{\boldsymbol{\theta}}^{(\mathrm{LS})}||^2} = \frac{1}{1 + ||\hat{\boldsymbol{\theta}}^{(\mathrm{LS})}||^2} \frac{||\hat{\mathbf{e}}^{(\mathrm{LS})}||^2}{||\mathbf{y} - \bar{\mathbf{y}}||^2} \\ &= \frac{1}{1 + ||\hat{\boldsymbol{\theta}}^{(\mathrm{LS})}||^2} \frac{||\mathbf{y} - \hat{\mathbf{y}}^{(\mathrm{LS})}||^2}{||\mathbf{y} - \bar{\mathbf{y}}||^2}\end{aligned} \tag{7}$$

where $\bar{\mathbf{y}}$ is the mean of the output vector $\mathbf{y}$, $\hat{\mathbf{y}}^{(\mathrm{LS})}$ and $\hat{\mathbf{e}}^{(\mathrm{LS})}$ are the model prediction (one-step-ahead prediction) and the model residual vectors, produced by the model with LS estimate $\hat{\boldsymbol{\theta}}^{(\mathrm{LS})}$. Using the number λ_0, define a set: $\Gamma = \{\alpha_k\lambda_0 : \alpha_k = 10^{-k}\} \cup \{\beta_k\lambda_0 : \beta_k = 0.5 \times 10^{-k}\}$ for $k = 0, 1, \ldots, 5$. The regularization parameter λ will be chosen from the set Γ, where each element is set to be the candidate as the regulation parameter and the RTLS procedure is then performed. This will lead to a set of models with different RTLS estimates. The criterion for selecting the regularization parameter is to inspect the predictive capability of the resultant models. For a dynamical modeling problem, the resultant model should possess a satisfactory predicative ability in terms of model-predicted output (MPO), which is an extreme case of long-term prediction and which is the most stringent test for dynamical models. The value in Γ that produces the model with the best performance (in the sense that it minimizes the errors between the MPOs and the corresponding measurements) will be selected as the regulation parameter λ.

As an example, consider two nonlinear systems described by

$$y(n) = \sum_{i=1}^{4} a_i y(n-i) + \sum_{j=1}^{3} b_j [u(n-1)]^j \tag{8}$$

$$y(n) = \sum_{i=1}^{4} a_i y(n-i) + \sum_{j=1}^{3} b_j [u(n-j)]^2 \tag{9}$$

where the model parameter vector $\boldsymbol{\theta}^T = [a_1, \ldots, a_4, b_1, b_2, b_3] = [1.8, -2.0, 1.5, -0.5, 0.5, -0.25, -0.1]$ for both of the two models given earlier, and the input $u(n)$ was chosen to be a stochastic process

$$u(n) = w(n) - 1.96w(n-1) + 0.98w(n-2) \tag{10}$$

where $w(n)$ was a Gaussian white noise sequence with zero mean and unit variance. The models were simulated and 200 input–output data pairs were collected for both of the two models; a noise signal was then deliberately added to the data points,

TABLE I
COMPARISON OF PARAMETER ESTIMATES PRODUCED BY LS, TLS, AND RTLS, FOR MODEL GIVEN BY (8)

Parameter		a_1	a_2	a_3	a_4	b_1	b_2	b_3
True		1.8	-2.0	1.5	-0.5	0.5	-0.25	-0.1
I	LS	1.7958	-1.9938	1.4934	-0.4947	0.4942	-0.2489	-0.0982
	TLS	1.7988	-1.9981	1.4978	-0.4984	0.5015	-0.2498	-0.1034
	RTLS	1.8005	-2.0011	1.5005	-0.4991	0.4956	-0.2484	-0.0984
II	LS	1.5291	-1.5649	1.1044	-0.2943	0.4583	-0.2817	-0.1095
	TLS	1.7736	-1.9637	1.4529	-0.4650	0.4894	-0.2471	-0.0566
	RTLS	1.7919	-2.0191	1.5311	-0.5125	0.5365	-0.2553	-0.1225

Case I: SNR=40dB (for input), SNR=40dB (for output), λ =3.3704 $\times 10^{-8}$. Case II: SNR=20dB (for input), SNR=20dB (for output), λ =4.0654 $\times 10^{-6}$.

TABLE II
COMPARISON OF PARAMETER ESTIMATES PRODUCED BY LS, TLS, AND RTLS, FOR MODEL GIVEN BY (9)

Parameter		a_1	a_2	a_3	a_4	b_1	b_2	b_3
True		1.8	-2.0	1.5	-0.5	0.5	-0.25	-0.1
I	LS	1.7844	-1.9834	1.4810	-0.4918	0.5011	-0.2463	-0.0963
	TLS	1.7945	-2.0021	1.4965	-0.5022	0.5078	-0.2577	-0.0887
	RTLS	1.7962	-1.9979	1.4964	-0.4995	0.5018	-0.2515	-0.0962
II	LS	0.8128	-1.0370	0.3794	-0.0201	0.4123	0.2756	0.0058
	TLS	1.8084	-1.9874	1.5115	-0.4899	0.4295	-0.1565	-0.1721
	RTLS	1.7978	-2.0024	1.5013	-0.5063	0.4837	-0.2313	-0.0980

Case I: SNR=40dB (for input), SNR=40dB (for output), λ =2.2730 $\times 10^{-6}$. Case II: SNR=20dB (for input), SNR=20dB (for output), λ =0.0011.

and these noisy data were then used for model parameter estimation using the RTLS algorithm.

A comparison of the results produced by LS, TLS, and RTLS for the two systems given by models (8) and (9), is shown in Tables I and II, respectively. While the three methods produce almost exactly the same parameter estimates when the noise level is low (with a high SNR), for the case with high-level noise (with a low SNR), the results are significantly different: the RTLS estimates, which are slightly better than TLS, significantly outperform LS, as can obviously be observed in Table II.

III. DATA MODELING BETWEEN NORMALIZED CHANGES IN CBF AND CBV

A. Datasets

The datasets presented here were reworked from [57]. The experimental procedures for concurrent measurements of blood flow and volume changes were described in more detail in [4] and [6]. These are briefly reviewed here. The animals used were hooded Lister rats weighing between 300 and 400 g, anesthetized with urethane (1.25 g/kg, intraperitoneal injection), and atropine (0.4 mL/kg, subcutaneous injection). The whisker barrel of a rat was located using single-wavelength ($\sim$590 nm) illumination before the slit spectrograph mounted on the camera was placed over the center of the barrel region. A laser-Doppler flowmeter (LDF) probe (Perimed, Stockholm, Sweden: fiber separation 0.25 mm) was then sited over the barrel region ($<$1 mm from the skull surface) to measure CBF. Spectroscopic data were analyzed using a pathlength scaling algorithm to provide the CBV time series [58].

The CBF time series was sampled at 30 Hz, while the CBV was sampled at 7.5 Hz. For convenience of data modeling, the

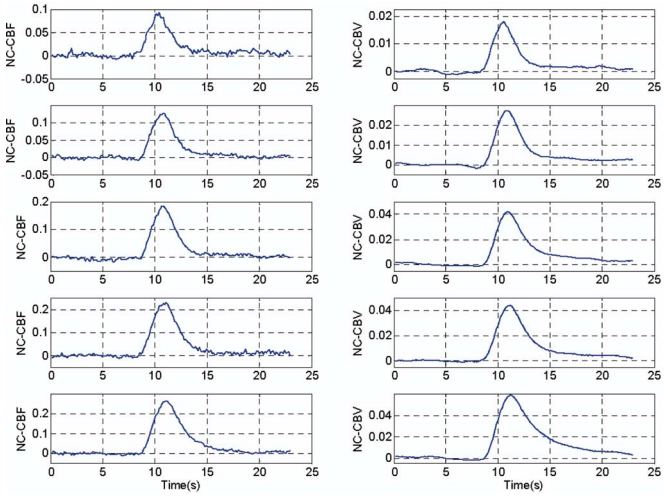


Fig. 1. Measurements of normalized changes in CBF and CBV. From (top) to (bottom), the plots are for the cases of 1, 2, 3, 4, and 5 Hz. ("NC-CBF" and "NC-CBV" indicate the normalized changes in CBF and CBV, respectively.)

CBF data were then downsampled at 7.5 Hz. Brief 2-s stimuli of 1, 2, 3, 4, and 5 Hz were randomly interleaved and applied within a single experimental run with stimulus intensity of 1.2 mA and a pulse width of 0.3 ms. Thirty trials were obtained for each stimulus condition with each trial lasting 23 s (with stimulus onset at 8 s), and an intertrial interval of 25 s to avoid hemodynamic refractory period.

In the following, the measurement of the normalized changes in CBF is treated as the input, denoted by $u(n)$, and the measurement of the normalized changes in CBV is treated as the output, denoted by $y(n)$. The objective is to learn, from available measurements of the five cases (1, 2, 3, 4, and 5 Hz), a common model structure that is suitable for describing the hemodynamics (measured by the normalized changes in CBF and CBV). Here, the normalized changes in CBF and CBV are defined as $\Delta\text{CBF}/\text{CBF}_0$ and $\Delta\text{CBV}/\text{CBV}_0$, respectively, where the prefix symbol "Δ" indicates the associated relative changes and the subscript "0" the relevant baseline values. A total of 172 input–output data points are involved in each of the datasets for the five cases, and these are shown in Fig. 1.

B. Model Identification

A model term and variable selection algorithm [31] was performed over each of the five datasets, and the significant model variables were determined to be $x_1(t) = y(t-1)$, $x_2(t) = y(t-2)$, $x_3(t) = u(t)$, $x_4(t) = u(t-1)$, and $x_5(t) = u(t-2)$. Three types of NARX models were considered: 1) an ARX model given by (4); 2) an NARX model with a nonlinear degree of order $\ell = 2$; and 3) an NARX model with a nonlinear degree of order $\ell = 3$. The initial full models of all the three types were formed using the five selected significant model variables. Each of the three initial full models was then used to generate a parsimonious model that fits all the five datasets, and this was implemented by using a common model structure selection algorithm [34], [59] over the five datasets. The model identification algorithm presented in [34] and [59] is briefly

TABLE III
PARAMETER ESTIMATES PRODUCED BY LS, TLS, AND RTLS, FOR CBF AND CBV MODELING PROBLEM

Parameter		a_1	a_2	b_0	b_1	b_2	NMSE(MPO)
I	LS	1.4819	-0.6017	0.0063	0.0494	-0.0232	0.0269
	TLS	2.0100	-0.9631	-0.0152	0.0397	-0.0358	(*)
	RTLS	**0.2814**	**0.1411**	**0.0529**	**0.0460**	**0.0196**	**0.0132**
II	LS	1.6368	-0.6544	0.0139	0.0293	-0.0400	0.1083
	TLS	1.9449	-0.9183	-0.0050	0.0298	-0.0325	(*)
	RTLS	**0.3243**	**0.2813**	**0.0757**	**0.0353**	**-0.0188**	**0.0354**
III	LS	1.8200	-0.8247	-0.0051	0.0450	-0.0424	7.3215
	TLS	1.9977	-0.9765	-0.0186	0.0477	-0.0363	(*)
	RTLS	**0.6073**	**0.2664**	**0.0833**	**0.0268**	**-0.0751**	**0.0417**
IV	LS	1.7724	-0.7736	-0.0084	0.0495	-0.0415	0.9283
	TLS	2.0908	-1.0461	-0.0266	0.0541	-0.0382	(*)
	RTLS	**0.4889**	**0.2847**	**0.0598**	**0.0329**	**-0.0440**	**0.0235**
V	LS	1.7673	-0.7627	-0.0032	0.0534	-0.0519	7.1514
	TLS	1.9558	-0.9386	-0.0199	0.0574	-0.0428	(*)
	RTLS	**0.7642**	**0.1665**	**0.0843**	**0.0318**	**-0.0970**	**0.0163**

I: for the data set of 1Hz, $\lambda = 7.3165 \times 10^{-5}$. II: for the data set of 2Hz, $\lambda = 3.1252 \times 10^{-4}$.
III: for the data set of 3Hz, $\lambda = 1.1499 \times 10^{-4}$. IV: for the data set of 4Hz, $\lambda = 1.5689 \times 10^{-4}$.
V: for the data set of 5Hz, $\lambda = 5.9245 \times 10^{-5}$.
MSE(MPO): the normalized mean-square error was calculated for the associated model predicted output (that is different from short-term predictions).
(*): the model produced by TLS is instable and the associated MPO is divergent;

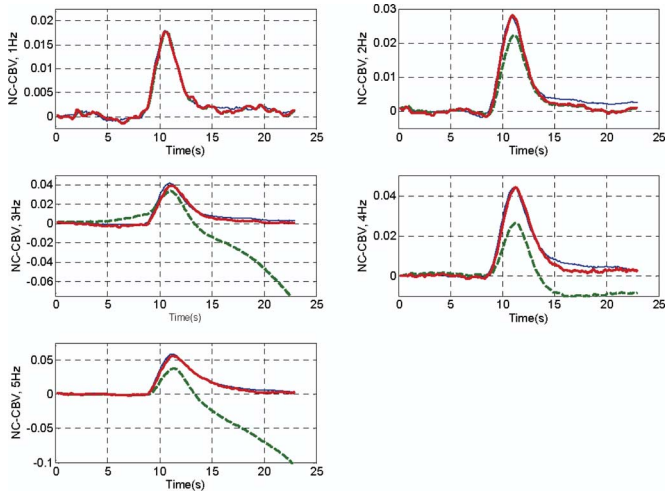


Fig. 2. Comparisons of the MPOs from the LS and RTLS-related models (given in Table III) and the associated measurements, for the five cases of 1, 2, 3, 4, and 5 Hz. In each figure, the thin solid line indicates the measurement, the thick solid line indicates the MPO produced by the RTLS-related model, and the thick dashed line indicates the MPO produced by the LS-related model.

described in the Appendix. By comparing the resultant model performance and by following the parsimonious principle, the common model structure that fits all the five datasets was determined to be

$$y(n) = a_1 y(n-1) + a_2 y(n-2) + b_0 u(n) + b_1 u(n-1) + b_2 u(n-2) \quad (11)$$

where the estimates of the coefficients a_i $(i = 1, 2)$ and b_j $(j = 0, 1, 2)$ using LS, TLS, and RTLS, for the five cases of 1, 2, 3, 4, and 5 Hz, are shown in Table III.

By setting the first two observations of CBV (the output) as the initial condition and by using the observations of the CBF as the inputs, the models produced by the LS and RTLS algorithms were simulated. The associated MPOs are shown in Fig. 2. Note that MPOs here are different from short-term (or multistep) ahead predictions and are a far more severe test than the often-used one-step ahead (OSA) predictions since the latter can often look good even for very poor models.

From Table III and Fig. 2, the following conclusions can be drawn: 1) for the given real datasets, where the associated measurements of changes in CBF and CBV may be contaminated by noise, the ordinary LS method does not work well; neither does the TLS method; 2) the RTLS method significantly outperforms both the LS and TLS methods, in that it can very effectively handle the errors-in-variables problems here; 3) the proposed empirical choice of the regularization parameter λ given by (7) works very well for the RTLS algorithm.

C. Data Filtering

From the results presented in previous sections, there are significant differences between the LS and RTLS estimates. This implies that there may be some noise in the associated input and output observations, because if the measurements are "clean,"

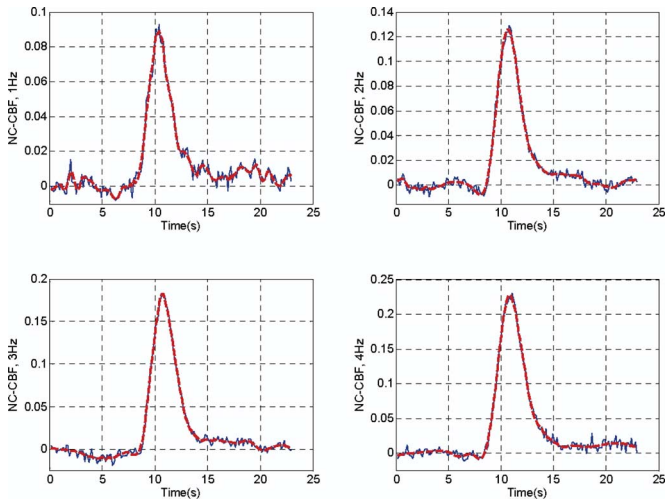


Fig. 3. Comparisons between the original CBF data and the filtered data for the cases of 1, 2, 3, and 4 Hz. The thin solid lines indicate the measurements, and the thick dashed lines indicate the filtered data.

then the LS and RTLS algorithms should in theory produce almost exactly the same parameter estimates. By visually inspecting the measurements of changes in CBF shown in Fig. 1, it can be observed that the CBF data are quite noisy compared with the CBV data. As a trial-and-error approach, the CBF data were then filtered by using wavelet filtering methods. As will be illustrated, filtering the original CBF data is useful for further improving the identified model performance.

The original CBF data in all the five datasets were filtered with Daubechies' wavelets [60]. The filtered data, along with the relevant original data for the first four cases of 1, 2, 3, and 4 Hz, are shown in Fig. 3 (the 5-Hz case was omitted here to save space). Using the filtered CBF data as the input and the associated CBV data (unfiltered) as the output, the coefficients

TABLE IV
RTLS ESTIMATES FOR CBF AND CBV MODELING PROBLEM, WHERE FILTERED CBF DATA PLAY AS INPUT AND ORIGINAL CBV DATA AS OUTPUT

a_1	a_2	b_0	b_1	b_2	NMSE (MPO)
0.2948	0.1929	0.0901	0.0134	0.0023	0.0133
1.0085	-0.0273	0.0093	0.1775	-0.1816	0.0157
0.9872	-0.0164	0.1024	-0.0018	-0.0910	0.0086
0.8348	0.1180	0.0258	0.1395	-0.1540	0.0059
0.6617	0.3001	0.1153	0.0308	-0.1343	0.0027

Model parameter for all the five cases was estimated using the RTLS algorithm.

MSE(MPO): MPO was calculated by simulating the associated model where the original CBF data (unfiltered) was set to be the input, and the MPO was then compared with the original real measurement.

I: for the data set of 1Hz, $\lambda = 5.2565 \times 10^{-5}$. II: for the data set of 2Hz, $\lambda = 1.0357 \times 10^{-6}$.
III: for the data set of 3Hz, $\lambda = 2.2285 \times 10^{-6}$. IV: for the data set of 4Hz, $\lambda = 2.8804 \times 10^{-6}$.
V: for the data set of 5Hz, $\lambda = 5.9841 \times 10^{-6}$.

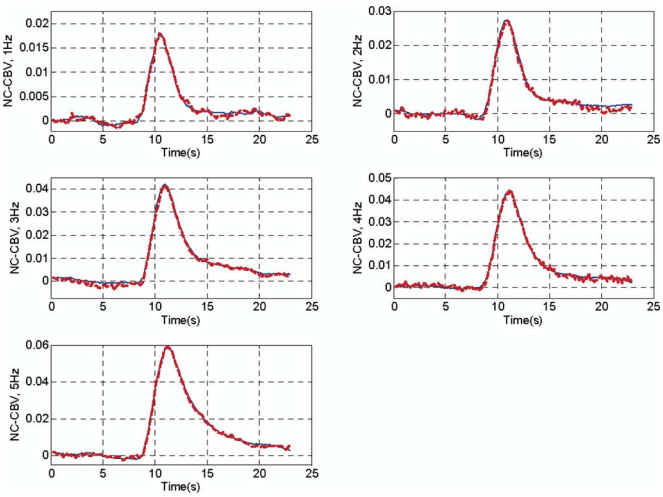


Fig. 4. Comparisons between the measurements and the associated MPOs produced by the models estimated from the filtered CBF data using the proposed RTLS algorithm. The thin solid lines indicate the measurements, and the thick dashed lines indicate the associated MPOs.

of the ARX model of the form (11) were then reestimated using the RTLS method, and the associated parameter estimates are shown in Table IV.

The five models given in Table IV were simulated; for each case, the original CBF data (unfiltered) was used as the model input, and the associated MPO was then compared with the original CBV data. Comparisons between the MPOs and the corresponding original observations are shown in Fig. 4. From Table IV and Fig. 4, it is quite clear that models estimated from the filtered CBF data are much better than those from the original CBF data. While Fig. 4 only provides some visual perception, the values of the normalized MSE listed in Tables IV and III give a quantitative comparison. As can be noticed, the normalized MSE for the MPO given in Table IV is much smaller than that given in Table III, for all the five cases.

It should be stressed that we generally would not recommend arbitrarily filtering measured data as the initial step for data preprocessing, in particular, for application cases where data are used for typical nonlinear dynamic systems modeling and identification, because this may mask the underlying nonlinear dynamics. Filtering out some frequencies would imply that these are only linear effects, whereas it may be that these are the direct result of nonlinear behavior. For example, a frequency component at f hertz could also be caused by a nonlinear intermodulation effect $f_1 - f_2$, where the difference between f_1 and f_2 is f hertz [61]–[63]. However, if there is evidence that a carefully chosen filter, together with an efficient modeling algorithm, can produce much better model estimation, like the modeling practice here, then filtering may be introduced to improve the model performance.

D. Continuous-Time Model

In some cases, it may be desirable to identify continuous-time models. From linear systems and signal processing theory, the linear discrete-time model (11) can easily be converted into a continuous model. First, the discrete-time (z-domain) transfer function of (11) is given by

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - a_1 z^{-1} - a_2 z^{-2}}. \tag{12}$$

By applying the well-known Tustin transform (also called the bilinear transform), i.e., by letting

$$z = e^{sT_s} \approx \frac{1 + sT_s/2}{1 - sT_s/2} \tag{13}$$

where T_s is the sampling interval, the z-domain transfer function can then be converted into the s-domain. Taking the case of 5 Hz as an example, where $T_s = 2/15s$, and the associated coefficients are listed in Table IV, the s-domain transfer function is given by

$$G_5(s) = \frac{-0.03654s^2 + 5.501s + 1.965}{s^2 + 28.65s + 6.316}. \tag{14}$$

The transfer function (14) can further be converted into a differential equation model

$$\frac{d^2y}{dt^2} + 28.65\frac{dy}{dt} + 6.316y(t) = -0.03654\frac{d^2u}{dt^2} + 5.501\frac{du}{dt} + 1.965u(t). \tag{15}$$

Driven by the input (measurement of changes in CBF) in the associated dataset of the 5 Hz case, the continuous-time model (14) was simulated by using an extrapolation method in the Runge–Kutta family of ordinary differential equation solvers provided by MATLAB in the ordinary differential equation toolbox, and a comparison of the output produced by the continuous-time model (14) and that by the discrete-time model given in Table IV is shown in Fig. 5. Note that there is a discrepancy between the outputs produced by the continuous-time and discrete-time models here, as the bilinear transform defined by (13), like many other model transforms, is not perfect. The maximum relative error, defined by $R_e = \max(|y_c - y_d|)/\max(|y_d|)$, where y_c and y_d are the outputs that were, respectively, produced by the identified continuous-time and discrete-time models, was calculated to be $R_e = 0.425/0.593 \approx 7.17\%$. While such an error level here can be considered acceptable, some more complex

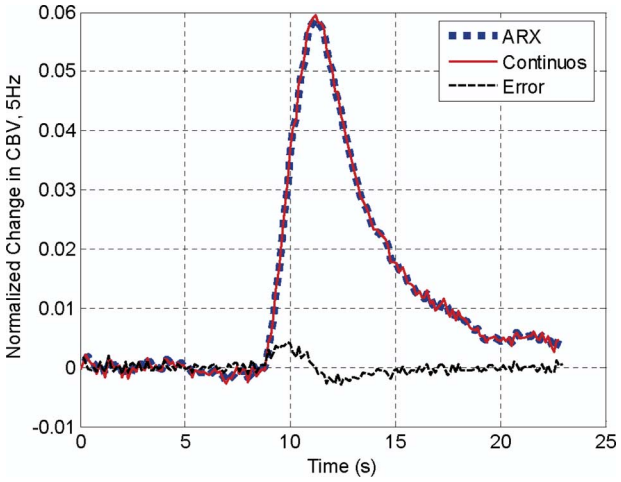


Fig. 5. Comparison of the output produced by the continuous-time model (14) and that by the discrete-time model given in Table IV, for the 5 Hz case.

and accurate transformations may still be needed to reduce the discrepancy of the outputs produced by discrete-time and the resultant continuous-time models.

IV. Discussion and Conclusion

Although the relationship between the normalized CBF and CBV is well known to be nonlinear in general, this paper demonstrated that it is possible to describe the dynamic relationship between the normalized *changes* in CBF and CBV by a set of linear differential equations, under certain experimental conditions and within certain ranges of physiological parameters. The data used in the paper were collected from anesthetized animals under electrical stimulation. Further work will be needed to examine if this linear relationship holds under other anesthetic conditions, e.g., for awake animals, and under other experimental conditions, e.g., under hypercapnia or hypocapnia.

The focus of the paper has been on the development of a data-based modeling approach for the identification of models that can be used to describe the dynamical relationship of changes in CBF and CBV during neural activity. This is a complicated blackbox system where the true model structure is unknown, and thus, needs to be identified from available experimental data. The central task of data-based modeling of such a structure-unknown system involves several aspects including model variable selection, model structure specification and detection, parameter estimation, and model validation. In this study, a NARX, which has been widely used for nonlinear system identification, was chosen as the initial candidate model structure. Compared with many other model structures, for example, typical neural networks, the NARX model structure possesses several advantages, some of which are as follows.

1) A wide range of nonlinear systems can be described using the NARX model.
2) Over the last two decades, the NARX model has been systematically studied and a series of excellent algorithms have been developed for the identification of such models. This means that model structure detection and parameter estimation for such a model can be performed speedily and efficiently using existing algorithms.
3) The NARX model is transparent, and thus, can easily be related back to the underlying system.
4) Algorithms that exist can directly map the NARX model and continuous-time ODE models into the frequency domain [61]–[63]; this allows the user to reveal the explicit link from the time-domain model parameters to the frequency-domain properties.

For the model parameter estimation problem, it has been illustrated that neither the ordinary LS method nor the classical TLS method can produce reliable estimates from the available CBF and CBV data, which were contaminated by noise. The RTLS method, however, works very well when applied to the error-in-variables problem here. Note that the application of RTLS involves nonlinear optimization and the need to estimate the value of the regularization parameter. The Nelder–Mead simplex direct search optimization algorithm was introduced to solve the RTLS equation (6), where the initial value of the unknown parameters was chosen to be the LS estimates. While Nelder–Mead algorithm, coupled with the rule of thumb for choosing the regularization parameter given in (7), can work very well for model parameter estimation, there still exists space to further optimize the choice of the regularization parameter, as well as the initial value for the free parameters to be optimized.

It can be believed that the basic ideas and algorithms developed in this paper can be directly applied or extended to other biomedical modeling problems, where limited *a priori* information is available and the true model structure of the underlying dynamics is unknown.

Appendix

Multiple Orthogonal Search Algorithm for Model Selection

Assume that a total of K experiments have been carried out on the same system and K different datasets have accordingly been obtained. Also, assume that there exists a common model structure that can best fit all the K datasets. Denote the input and the output sequence for the kth experiment by $\{u_k(t)\}_{t=1}^{N_k}$ and $\{y_k(t)\}_{t=1}^{N_k}$, respectively, for $k = 1, \ldots, K$. The kth predictor vector is thus given by

$$\mathbf{x}_k(t) = [x_{k,1}(t), \ldots x_{k,d}(t)]^T = [y_k(t-1), \ldots, y_k(t-n_y), u_k(t-1), \cdots, u_k(t-n_u)]^T.$$

It is assumed that all the K datasets can be represented using a common model structure, with a different parameter set, deduced from the initial candidate regression model given as

$$y_k(t) = \sum_{m=1}^{M} \theta_{k,m}\phi_m(\mathbf{x}_k(t)) + e_k(t) = \sum_{m=1}^{M} \theta_{k,m}\phi_{k,m}(t) + e_k(t). \quad (16)$$

This can be expressed using a compact matrix form

$$\mathbf{y}_k = \Phi_k \boldsymbol{\theta}_k + \mathbf{e}_k \tag{17}$$

where $\mathbf{y}_k = [y_k(1), \ldots, y_k(N_k)]^T$, $\boldsymbol{\theta}_k = [\theta_{k,1}, \ldots, \theta_{k,M}]^T$, $\mathbf{e}_k = [e_k(1), \ldots, e_k(N_k)]^T$, and $\Phi_k = [\boldsymbol{\varphi}_{k,1}, \ldots, \boldsymbol{\varphi}_{k,M}]$ with $\boldsymbol{\varphi}_{k,m} = [\phi_{k,m}(1), \ldots, \phi_{k,m}(N_k)]^T$ for $k = 1, \ldots, K$ and $m = 1, \ldots, M$.

For large lags n_y and n_u, the regression model (16) often involves a large number of candidate model terms, even if the nonlinear degree ℓ is not very high, say $\ell = 2$ or $\ell = 3$. Experience has shown that an initial candidate model with a large number of candidate model terms can often be drastically reduced to a simple sparse model that only includes a few number of effectively selected significant model terms. A simple sparse model is usually desirable for practical applications including system analysis, design, control, and prediction.

Let $I = \{1, 2, \ldots, M\}$. Denote by $D = \{\phi_m : m \in I\}$ the dictionary of candidate model terms for an initially chosen candidate common model structure that fits to all the K regression models. For the kth dataset, the dictionary D can be used to form a dual dictionary $V_k = \{\boldsymbol{\varphi}_{k,m} : m \in I\}$, where the mth candidate basis vector $\boldsymbol{\varphi}_{k,m}$ is formed by the mth candidate model term $\phi_m \in D$, in the sense that $\boldsymbol{\varphi}_{k,m} = [\phi_m(\mathbf{x}_k(1)), \ldots, \phi_m(\mathbf{x}_k(N_i))]^T$ $(k = 1, \ldots, K)$. The common model structure selection problem is equivalent to finding, from I, a subset of indices, $I_n = \{i_m : m = 1, \ldots, n, i_m \in I\}$ where $n \leq M$, so that $\mathbf{y}_i$ $(i = 1, \ldots, K)$ can be approximated using a linear combination of $\boldsymbol{\varphi}_{i_1}, \boldsymbol{\varphi}_{i_2}, \ldots, \boldsymbol{\varphi}_{i_n}$ as

$$\mathbf{y}_i = \theta_{i,1}\boldsymbol{\varphi}_{i,i_1} + \cdots + \theta_{i,n}\boldsymbol{\varphi}_{i,i_n} + \mathbf{e}_i. \tag{18}$$

A squared correlation coefficient can be used to measure the dependency between two associated random vectors [33], [64]. The squared correlation coefficient between two vectors $\mathbf{x}$ and $\mathbf{y}$ of size N is defined as

$$C(\mathbf{x}, \mathbf{y}) = \frac{(\mathbf{x}^T\mathbf{y})^2}{(\mathbf{x}^T\mathbf{x})(\mathbf{y}^T\mathbf{y})} = \frac{(\sum_{i=1}^N x_i y_i)^2}{\sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i^2}. \tag{19}$$

The squared correlation coefficient is closely related to the error reduction ratio (ERR) criterion defined in the orthogonal LS algorithm for model structure selection [26].

Let $\mathbf{r}_{k,0} = \mathbf{y}_k$ $(k = 1, 2, \ldots, K)$. For $k = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, M$, calculate $c_1(k, j) = C(\mathbf{y}_k, \boldsymbol{\varphi}_{k,j})$, and define

$$s_1 = \arg \max_{1 \leq j \leq M} \left\{ \frac{1}{K} \sum_{k=1}^K c_1(k, j) \right\}. \tag{20}$$

The first significant common model term can then be selected as the s_1th element, ϕ_{s_1}, in the dictionary D. Accordingly, the first significant basis vector for the kth regression model is thus $\boldsymbol{\alpha}_{k,1} = \boldsymbol{\varphi}_{k,s_1}$, and the first associated orthogonal basis vector can then be chosen as $\mathbf{q}_{k,1} = \boldsymbol{\varphi}_{k,s_1}$. The model residual for the kth regression model, related to the first step search, is given as

$$\mathbf{r}_{k,1} = \mathbf{r}_{k,0} - \frac{\mathbf{r}_{k,0}^T\mathbf{q}_{k,1}}{\mathbf{q}_{k,1}^T\mathbf{q}_{k,1}}\mathbf{q}_{k,1}. \tag{21}$$

Notice that $c_1(k, s_1)$ can be viewed as the ERR that is introduced by including the first basis vector $\boldsymbol{\alpha}_{k,1} = \boldsymbol{\varphi}_{k,s_1}$ into the kth regression model. The criterion (20), by maximizing the sum of the ERR values relative to all the K datasets, guarantees that the variation of the outputs in all the K datasets can be explained by including the model term ϕ_{s_1}, with the highest percentage, compared with selecting any other candidate model term $\phi \in D = \{\phi_m : m \in I\}$. The first average error reduction ratio (AERR) is given by $\text{AERR}(1) = (1/K)\sum_{k=1}^K c_1(k, s_1)$.

In general, the mth significant model term ϕ_{s_m} can be chosen as follows. Assume that at the $(m-1)$th step, $(m-1)$ significant model terms, $\phi_1, \phi_2, \ldots, \phi_{m-1}$, have been selected. Let $\boldsymbol{\alpha}_{k,1}, \ldots, \boldsymbol{\alpha}_{k,m-1}$ be the associated basis vectors for the kth regression model, and assume that the $(m-1)$ selected bases have been transformed into a new group of orthogonal bases $\mathbf{q}_{k,1}, \ldots, \mathbf{q}_{k,m-1}$ via some orthogonal transformation. Let

$$\mathbf{p}_{k,j}^{(m)} = \boldsymbol{\varphi}_{k,j} - \sum_{s=1}^{m-1} \frac{\boldsymbol{\varphi}_{k,j}^T\mathbf{q}_{k,s}}{\mathbf{q}_{k,s}^T\mathbf{q}_{k,s}}\mathbf{q}_{k,s}, \qquad j \in J_m \tag{22}$$

where $J_m = \{j : 1 \leq j \leq M, j \neq s_t, 1 \leq t \leq m-1\}$. For $k = 1, \ldots, K$ and $j \in J_m$, calculate $c_m(k, j) = C(\mathbf{y}_k, \mathbf{p}_{k,j}^{(m)})$, and define

$$s_m = \arg \max_{j \in J_m} \left\{ \frac{1}{K} \sum_{k=1}^K c_m(k, j) \right\}. \tag{23}$$

The mth significant common model term is thus selected as the s_mth element, ϕ_{s_m}, in the dictionary D. Accordingly, the mth significant basis vector for the kth regression model is thus $\boldsymbol{\alpha}_{k,m} = \boldsymbol{\varphi}_{k,s_m}$, and the associated orthogonal basis vector can then be chosen as $\mathbf{q}_{k,m} = \mathbf{p}_{k,s_m}^{(m)}$. The model residual for the kth regression model, related to the mth step search, is given as

$$\mathbf{r}_{k,m} = \mathbf{r}_{k,m-1} - \frac{\mathbf{r}_{k,m-1}^T\mathbf{q}_{k,m}}{\mathbf{q}_{k,m}^T\mathbf{q}_{k,m}}\mathbf{q}_{k,m}. \tag{24}$$

Subsequent significant bases can be selected in the same way step by step, one model term at a time. Once the first $(m-1)$ basis vectors $\boldsymbol{\alpha}_{k,1}, \ldots, \boldsymbol{\alpha}_{k,m-1}$ (respectively, the associated orthogonalized vectors $\mathbf{q}_{k,1}, \ldots, \mathbf{q}_{k,m-1}$) have been determined, then these $(m-1)$ bases together with the mth vector $\boldsymbol{\alpha}_{k,m} = \boldsymbol{\varphi}_{k,s_m}$ (respectively, the orthogonalized vector $\mathbf{q}_{k,m} = \mathbf{p}_{k,s_m}^{(m)}$), can explain the variation in the outputs of the K datasets with a higher percentage than by including any other candidate vectors. The quantity $\text{AERR}(m) = (1/K)\sum_{k=1}^K c_m(k, s_m)$ is referred to as the mth AERR. While this step-by-step forward selection algorithm is a nonexhaustive search method and may not always produce the perfect global optimal solution, it can usually produce satisfactory and nearly optimal results for most real world problems

From (24), the vectors $\mathbf{r}_{k,m}$ and $\mathbf{q}_{k,m}$ are orthogonal, and hence

$$||\mathbf{r}_{k,m}||^2 = ||\mathbf{r}_{k,m-1}||^2 - \frac{(\mathbf{r}_{k,m-1}^T\mathbf{q}_{k,m})^2}{\mathbf{q}_{k,m}^T\mathbf{q}_{k,m}}. \tag{25}$$

By, respectively, summing (24) and (25) for m from 1 to n, yields

$$\mathbf{y}_k = \sum_{m=1}^{n} \frac{\mathbf{r}_{k,m-1}^T \mathbf{q}_{k,m}}{\mathbf{q}_{k,m}^T \mathbf{q}_{k,m}} \mathbf{q}_{k,m} + \mathbf{r}_{k,n} \tag{26}$$

$$\begin{aligned} ||\mathbf{r}_{k,n}||^2 &= ||\mathbf{r}_{k,n-1}||^2 - \frac{(\mathbf{r}_{k,n-1}^T \mathbf{q}_{k,n})^2}{\mathbf{q}_{k,n}^T \mathbf{q}_{k,n}} \\ &= ||\mathbf{y}_k||^2 - \sum_{m=1}^{n} \frac{(\mathbf{r}_{k,m-1}^T \mathbf{q}_{k,m})^2}{\mathbf{q}_{k,m}^T \mathbf{q}_{k,m}}. \end{aligned} \tag{27}$$

From (26) and (27), the model residual $\mathbf{r}_{k,n}$ can be used to form a criterion for model selection and the search procedure will be terminated when the norm $||\mathbf{r}_{k,n}||^2$ satisfies some specified conditions. In the present study, an approximate minimum description length (AMDL) criterion presented in [65] and [66] is used to determine the model size. For the case of $K = 1$, AMDL is defined as

$$\text{AMDL}(n) = 0.5 \log_2[\text{MSE}(n)] + \frac{1.5n \log_2 N}{N} \tag{28}$$

where $\text{MSE} = ||\mathbf{r}_n||^2/N$ is the mean-square-error associated with model of n terms, N is the length of the associated training dataset, n is the number of model terms, and $\mathbf{r}_n$ is the associated model residual. Other criteria [32], [64] can also be used to replace (28) to monitor the orthogonal search procedure.

The present study uses the following average AMDL as the criterion to determine the number of common model terms:

$$\text{AAMDL}(n) = \frac{1}{K} \sum_{k=1}^{K} \text{AMDL}^{[k]}(n) \tag{29}$$

where $\text{AMDL}^{[k]}(n)$ is the value for the AMDL criterion associated to the kth dataset. For the case with $K = 1$, the aforementioned multiple orthogonal search algorithm reduces to the well-known orthogonal LS algorithm [26].

REFERENCES

[1] R. L. Grubb, M. E. Raichle, J. O. Eichling, and M. M. Ter-Pergossian, "The effects of changes in PACO2 on cerebral blood volume, blood flow and vascular mean transit time," *Stroke*, vol. 5, pp. 630–639, 1974.

[2] R. B. Buxton, E. C. Wong, and L. R. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: The balloon model," *Magn. Reson. Med.*, vol. 39, pp. 855–864, 1998.

[3] J. B. Mandeville, J. J. Marota, C. Ayata, G. Zaharchuk, M. A. Moskowitz, B. R. Rosen, and R. M. Weisskoff, "Evidence of a cerebrovascular postarteriole Windkessel with delayed compliance," *J. Cereb. Blood Flow Metab.*, vol. 19, pp. 679–689, 1999.

[4] M. Jones, J. Berwick, D. Johnston, and J. Mayhew, "Concurrent optical imaging spectroscopy and laser-Doppler flowmetry: The relationship between blood flow, oxygenation, and volume in rodent barrel cortex," *Neuroimage*, vol. 13, pp. 1002–1015, 2001.

[5] M. Jones, J. Berwick, and J. Mayhew, "Changes in blood flow, oxygenation, and volume following extended stimulation of rodent barrel cortex," *Neuroimage*, vol. 15, pp. 474–487, 2002.

[6] Y. Kong, Y. Zheng, D. Johnston, J. Martindale, M. Jones, S. Billings, and J. Mayhew, "A model of the dynamic relationship between blood flow and volume changes during brain activation," *J. Cereb. Blood Flow Metab.*, vol. 24, pp. 1382–1392, 2004.

[7] Y. Zheng, D. Johnston, J. Berwick, D. Chen, S. Billings, and J. Mayhew, "A three-compartment mode of the hemodynamic respond and oxygen delivery to brain," *Neuroimage*, vol. 28, pp. 925–939, 2005.

[8] K. J. Friston, "Bayesian estimation of dynamical systems: An application to fMRI," *NeuroImage*, vol. 16, pp. 513–530, 2002.

[9] K. J. Friston, A. Mechelli, R. Turner, and C. J. Price, "Nonlinear responses in fMRI: The balloon model, volterra kernels and other hemodynamics," *Neuroimage*, vol. 12, pp. 466–477, 2000.

[10] K. J. Friston, A. P. Holmes, J. B. Poline, P. J. Grasby, S. C. Williams, R. S. Frackowiak, and R. Turner, "Analysis of fMRI time-series revisited," *NeuroImage*, vol. 2, pp. 45–53, 1995.

[11] K. J. Worsley and K. J. Friston, "Analysis of fMRI time-series revisited-again," *NeuroImage*, vol. 2, pp. 173–181, 1995.

[12] K. J. Worsley, J.-B. Poline, K. J. Friston, and A. C. Evans, "Characterizing the response of PET and fMRI data using multivariate linear models," *NeuroImage*, vol. 6, pp. 305–319, 1997.

[13] R. B. Panerai, D. M. Simpson, S. T. Deverson, P. Mahony, P. Hayes, and D. H. Evans, "Multivariate dynamic analysis of cerebral blood flow regulation in humans," *IEEE Trans. Biomed. Eng*, vol. 47, no. 3, pp. 419–423, Mar. 2000.

[14] M. W. Woolrich, B. D. Ripley, M. Brady, and S. M. Smith, "Temporal autocorrelation in univariate linear modeling of fMRI data," *NeuroImage*, vol. 14, pp. 1370–1386, 2001.

[15] G. D. Mitsis, M. J. Poulin, P. A. Robbins, and V. Z. Marmarelis, "Nonlinear modeling of the dynamic effects of arterial pressure and CO2 variations on cerebral blood flow in healthy humans," *IEEE Trans. Biomed. Eng*, vol. 51, no. 11, pp. 1932–1943, Nov. 2004.

[16] J. Riera, J. Bosch, O. Yamashita, R. Kawashima, N. Sadato, T. Okada, and T. Ozakic, "fMRI activation maps based on the NN-ARx model," *Neuroimage*, vol. 23, pp. 680–697, 2004.

[17] P. Baraldi, A. A. Manginelli, M. Maieron, D. Liberati, and C. A. Porroa, "An ARX model-based approach to trial by trial identification of fMRI-BOLD responses," *Neuroimage*, vol. 37, pp. 189–201, 2007.

[18] S. A. Billings, H. L. Wei, and M. A. Balikhin, "Generalized multiscale radial basis function networks," *Neural Netw.*, vol. 20, pp. 1081–1094, 2007.

[19] G. H. Golub, "Some modified matrix eigenvalue problems," *SIAM Rev*, vol. 15, pp. 318–344, 1973.

[20] G. H. Golub and C. F. Van Loan, "An analysis of the total least squares problem," *SIAM J. Numer. Anal.*, vol. 17, pp. 883–893, 1980.

[21] S. A. Billings and W. S. F. Voon, "Structure detection and model validity tests in the identification of non-linear system," *Proc. Inst. Electron. Eng., D*, vol. 130, pp. 193–199, 1983.

[22] S. A. Billings and W. S. F. Voon, "Correlation based model validity tests for non-linear models," *Int. J. Control*, vol. 44, pp. 235–244, 1986.

[23] S. A. Billings and W. S. F. Voon, "Piecewise linear identification of non-linear systems," *Int. J. Control*, vol. 46, pp. 215–235, 1987.

[24] I. J. Leontaritis and S. A. Billings, "Model selection and validation methods for non-linear systems," *Int. J. Control*, vol. 45, pp. 311–341, 1987.

[25] I. J. Leontaritis and S. A. Billings, "Experimental-design and identifiability for non-linear systems," *Int. J. Syst. Sci.*, vol. 18, pp. 189–202, 1987.

[26] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to nonlinear system identification," *Int. J. Control*, vol. 50, pp. 1873–1896, 1989.

[27] S. A. Billings and C. F. Fung, "Recurrent radial basis function networks for adaptive noise cancellation," *Neural Netw.*, vol. 8, pp. 273–290, 1995.

[28] S. A. Billings and Q. M. Zhu, "Nonlinear model validation using correlation tests," *Int. J. Control*, vol. 60, pp. 1107–1120, 1994.

[29] S. A. Billings and Q. M. Zhu, "Model validation tests for multivariable nonlinear models including neural networks," *Int. J. Control*, vol. 62, pp. 749–766, 1995.

[30] L. A. Aguirre and S. A. Billings, "Dynamical effects of overparametrization in nonlinear models," *Physica D*, vol. 80, pp. 26–40, 1995.

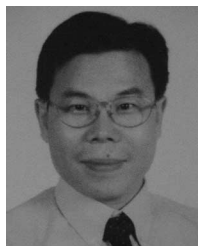
[31] H. L. Wei, S. A. Billings, and J. Liu, "Term and variable selection for non-linear system identification," *Int. J. Control*, vol. 77, pp. 86–110, 2004.

[32] S. A. Billings and H. L. Wei, "An adaptive orthogonal search algorithm for model subset selection and non-linear system identification," *Int. J. Control*, vol. 81, pp. 714–724, 2008.

[33] 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. Identification Control*, vol. 3, pp. 341–356, 2008.

[34] H. L. Wei and S. A. Billings, "Improved model identification for non-linear systems using a random subsampling and multifold modelling (RSMM) approach," *Int. J. Control*, vol. 82, pp. 27–42, 2009.

[35] D. M. Sima, S. Van Huffel, and G. H. Golub, "Regularized total least Squares based on quadratic eigenvalue problem solvers," *BIT Numer. Math.*, vol. 44, pp. 793–812, 2004.
[36] J. A. Nelder and P. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.
[37] I. J. Leontaritis and S. A. Billings, "Input–output parametric models for non-linear systems—part I: Deterministic non-linear systems," *Int. J. Control*, vol. 41, pp. 303–328, 1985.
[38] I. J. Leontaritis and S. A. Billings, "Input–output parametric models for non-linear systems—part II: Stochastic non-linear systems," *Int. J. Control*, vol. 41, pp. 329–344, 1985.
[39] S. A. Billings and I. J. Leontaritis, "Identification of nonlinear systems using parametric estimation techniques," in *Proc. IEE Conf. Control Appl.*, Warwick, U.K., 1981, pp. 183–187.
[40] S. Chen and S. A. Billings, "Representation of non-linear systems: The NARMAX model," *Int. J. Control*, vol. 49, pp. 1013–1032, 1989.
[41] R. K. Pearson, *Discrete-Time Dynamic Models*. New York: Oxford Univ. Press, 1999.
[42] K. J. Astrom, *Introduction to Stochastic Control Theory*. New York: Academic, 1970.
[43] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[44] T. Söderström and P. Stoica, *System Identification*. New York: Prentice-Hall, 1989.
[45] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA: SIAM, 1991.
[46] S. Van Huffel, Ed. *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling* (SIAM Proceedings Series). Philadelphia, PA: SIAM, 1997.
[47] S. Van Huffel and P. Lemmerling, Eds., *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*. Dordrecht, The Netherlands: Kluwer, 2002.
[48] I. Markovsky and S. Van Huffel, "Overview of total least squares methods," *Signal Process.*, vol. 87, pp. 2283–2302, 2007.
[49] S. W. Chen, "A two-stage description of cardiac arrhythmias using a total least squares-based prony modeling algorithm," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 10, pp. 1317–1327, Oct. 2000.
[50] G. F. Shou, L. Xia, M. F. Jiang, Q. Wei, F. Liu, and S. Crozier, "Truncated total least squares: A new regulation method for the solution of ECG inverse problems," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 4, pp. 1327–1335, Apr. 2008.
[51] A. N. Tikhonov and V. Arsenin, *Solution of Ill-Posed Problems*. New York: Wiley, 1977.
[52] V. Mesarovic, N. Galatsanos, and A. Katsaggelos, "Regularizad constrained total least squares image restoration," *IEEE Trans. Image Process.*, vol. 4, no. 8, pp. 1096–1108, Aug. 1995.
[53] G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Tikhonov regularization and total least squares," *SIAM J. Matrix Anal. Appl.*, vol. 21, pp. 185–194, 1999.
[54] S. Van Huffel, I. Markovsky, R. J. Vaccaro, T. Soderstrom, and Guest Eds., "Total least squares and errors-in-variables modeling," *Signal Process.*, vol. 87, pp. 2281–2490, 2007.
[55] J. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," *SIAM J. Optim.*, vol. 9, pp. 112–147, 1998.
[56] R. M. Lewis, V. Torczon, and M. W. Trosset, "Direct search methods: Then and now," *J. Comput. Appl. Math.*, vol. 124, pp. 191–207, 2000.
[57] J. Martindale, J. Mayhew, J. Berwick, M. Jones, C. Martin, D. Johnston, P. Redgrave, and Y. Zheng, "The hemodynamic impulse response to a single neural event," *J. Cereb. Blood Flow Metab.*, vol. 23, pp. 546–555, 2003.
[58] J. Mayhew, Y. Zheng, Y. Q. Hou, B. Vuksanovic, J. Berwick, S. Askew, and P. Coffey, "Spectroscopic analysis of changes in remitted illumination: The response to increased neural activity in brain," *Neuroimage*, vol. 10, pp. 304–326, 1999.
[59] H. L. Wei, Z. Q. Lang, and S. A. Billings, "Constructing an overall dynamical model for a system with changing design parameter properties," *Int. J. Modelling, Identification Control*, vol. 5, pp. 93–104, 2008.
[60] I. Daubechies, *Ten Lectures on Wavelets*. Philaelphia, PA: SIAM, 1992.
[61] J. C. Peyton-Jones and S. A. Billings, "Recursive algorithm for computing the frequency response of a class of non-linear difference equation models," *Int. J. Control*, vol. 50, pp. 1925–1940, 1989.
[62] J. C. Peyton-Jones and S. A. Billings, "Interpretation of non-linear frequency response functions," *Int. J. Control*, vol. 52, pp. 319–346, 1990.
[63] Z. Q. Lang and S. A. Billings, "Output frequency characteristics of non-linear systems," *Int. J. Control*, vol. 64, pp. 1049–1067, 1996.
[64] 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.
[65] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," in *Wavelet in Geophysics*, E. Foufoula-Georgiou and P. Kumar, Eds. New York: Academic, 1994, pp. 299–324.
[66] A. Antoniadis, I. Gijbels, and G. Gregoire, "Model selection using wavelet decomposition and applications," *Biometrika*, vol. 84, no. 4, pp. 751–763, 1997.



Hua-Liang Wei received the B.Sc. degree in mathematics from Liaocheng University, Shangdong Province, China, in 1989, the M.Sc. degree in automatic control theory and applications from Beijing Institute of Technology, Beijing, China, in 1992, and the Ph.D. degree in signal processing and control engineering from the University of Sheffield, Sheffield, U.K., in 2004.

He previously held academic appointments (Lecturer and Associate Professorship) from 1992 to 2000 at Beijing Institute of Technology, Beijing, China. He joined the Department of Automatic Control and Systems Engineering, the University of Sheffield in 2004, where he is currently a Lecturer. His current research interests include modeling and identification methods for nonlinear systems; NARMAX methodology and its applications; signal and image processing; pattern recognition; spatiotemporal systems; neuroimaging, data modeling and analysis; wavelets and neural networks; applications of signal processing, system identification, and modeling in complex system analysis including systems/synthetic biology and biomedical engineering; forecasting of stochastic and dynamical processes; regression analysis; and linear and nonlinear optimization.

Ying Zheng received the B.Eng. degree (with first class honors) in automatic control and systems engineering and the Ph.D. degree in control systems engineering from the University of Sheffield, Sheffield, U.K., in 1983 and 1988, respectively.

Her current research interests include the development of mathematical modeling and system identification methodologies with applications to neuroimaging and biophysical modeling of brain function.

Yi Pan received the B.Sc. and M.Sc. degrees in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 2000 and 2002, respectively, and the Ph.D. degree from the University of the Sheffield, Sheffield, U.K., in 2007.

He is currently a Research Associate in the Department of Psychology, University of Sheffield. His current research interests include statistics, and identification and model validation of complex nonlinear systems.

Daniel Coca received the M.Eng. degree in electrical engineering from the "Transilvania" University of Brasov, Brasov, Romania, in 1993, and the Ph.D. degree in control systems engineering from the University of Sheffield, Sheffield, U.K., in 1997.

Since 1997, he has been a Research Associate in the Department of Automatic Control and Systems Engineering, University of Sheffield. From 2002 to 2004, he was a Lecturer in the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U.K. Since June 2004, he has been with the Department of Automatic Control and Systems Engineering at the University of Sheffield where he currently holds a Senior Lecturer position. His current research interests include modeling, identification, and control of complex systems, bioimaging, and biological data analysis using reconfigurable computers.

Dr. Coca is a Chartered Engineer (CEng) and a member of the Institution of Engineering and Technology (IET).

Liang-Min Li received the B.Eng. and M.Sc. degrees from Northeastern University, Shenyang, China, in 1990 and 1993, respectively., and the Ph.D. degree from the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, U.K., in 2001.

He is currently a Research Associate in the Signal Processing and Complex Systems Research Group, University of Sheffield. His current research interests include continuous nonlinear system identification, repetitive control, severely nonlinear system identification, and analysis in the time and frequency domain.

J. E. W. Mayhew, photograph and biography not available at the time of publication.

Stephen A. Billings received the B.Eng. degree (with first class honours) in electrical engineering from the University of Liverpool, Liverpool, U.K., in 1972, the Ph.D. degree in control systems engineering from the University of Sheffield, Sheffield, U.K., in 1976, and the degree of D.Eng. from the University of Liverpool in 1990.

He was appointed as Professor in the Department of Automatic Control and Systems Engineering, University of Sheffield in 1990 and leads the Signal Processing and Complex Systems Research Group. His current research interests include system identification and information processing 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, spatiotemporal systems, EEG, fMRI, and optical imagery of the brain, metabolic systems engineering, systems biology, and related fields.

Prof. Billings is a Chartered Engineer [CEng], Chartered Mathematician [CMath], Chartered Scientist [CSci], Fellow of the IET [U.K], and Fellow of the Institute of Mathematics and its Applications.