Fractional Power NARX Model Identification Using A Harmony Search Algorithm

Hua-Liang Wei, Yifan Zhao, Stephen A. Billings, Jia Zhao
Department of Automatic Control and Systems Engineering
The University of Sheffield
Sheffield, S3 JD, UK
w.hualiang@sheffield.ac.uk, y.zhao@sheffield.ac.uk, s.billings@sheffield.ac.uk

Abstract—**A novel type of discrete-time fractional-power nonlinear autoregressive with exogenous input (FPNARX) model is introduced for system identification, modeling and prediction. Parameter estimation of such a model is a nonlinear optimization problem. A harmony search algorithm is then applied to solve such fractional models. Examples of both simulated and real data are provided.**

Keywords–fractional-power AR/ARX/NARX; harmony search; Parameter estimation; system identifcation; optimization

I. INTRODUCTION

System identification, as a data based modeling approach, has been widely applied in every single area of engineering and science. Traditionally, the target model type to be induced and built from available experimental data for a given dynamic system is commonly chosen to be one form or another of discrete-time representations. For linear dynamic systems, the commonly used model is the well known autoregressive (AR) and autoregressive with exogenous input (ARX) models [1], [2]. For nonlinear dynamic systems, there is a large variety of model representation including neural networks [3],[4], radial basis function networks [5]-[7], fuzzy reasoning and approximation [8],[9], kernels models [10], multiresolution wavelet models [11]-[13], and the well known nonlinear autoregressive moving average with exogenous inputs (NARMAX) model [14],[15] which accommodates many traditional autoregressive types of models such as AR/ARX, ARMA/ARMAX, Voterra, Hammerstein, Wiener and bilinear models as special cases [16].

The autoregressive types of parametric representation such as ARX and NARX models have several advantages, for example such models are transparent and the role of each individual variable (or model term) can clearly be assessed [17] in terms of the contribution they make to the system output, and the interactions with other model variables or terms. Conventionally, the power of each of the individual variables in the models is an integer, like $y(k-1)$, $u^3(k-1)$, $y^2(k-2)u(k-2)$, where the power of the four variables is 1, 3, 2 and 1, respectively. Using variables with integer powers to form model terms and then constructing a system model is a natural way in that most existing mature theoretic results can directly be applied to the model, for example for stability analysis, frequency domain analysis and so on. There are many real-world scenarios, however, where the physical behavior can compactly and efficiently be characterized by a fractional-power model where some system variables take an effect in a form of fractional power [18]-[21].

Whilst there is a growing number of articles on theoretical analysis of continuous-time fractional-order systems and the solution of the forward problem in the literature, the inverse problem, that is, the investigation of methods and algorithms that can be used to reconstruct the original fractional equations or to establish compact equivalent representations for the original dynamic systems, merely from available experimental data, has received little attention and to the best of the authors' knowledge few results have been achieved in this area.

Motivated by these observations, this study aims to investigate and develop a novel fractional-power NARX (FPNARX) modeling approach. Estimation of the FPNARX model can be viewed as a kind of nonlinear optimization problem which can be solved by means of either classical gradient based algorithms or modern meta-heuristic search algorithms. This study, however, introduces a harmony search algorithm [22],[23] to solve the FPNARX model estimation problem. Examples of both simulated and real data are presented to illustrate the efficiency, effectiveness and applicability of newly developed fractional-power NARX modeling approach.

II. PROBLEM FORMULATION

A. The Polynomial NARX Model

It has been shown that under some mild conditions single-input and single-output (SISO) discrete-time or discretized continuous-time dynamical systems can be described by the difference equation model [14],[15]

$$y(k) = f(y(k-1), \cdots \qquad \cdots \qquad) \qquad (1)$$

where $u(k)$, $y(k)$ and $e(k)$ are the system input, output and noise variables, respectively; p and q are the maximum lags in the input and output; f is some unknown linear or nonlinear mapping. It is generally assumed that $e(k)$ is an independent identical distributed noise sequence. A commonly employed form of model (1) is the well-known nonlinear autoregressive with exogenous inputs (NARX) model [14]-[15], which can describe a wide range of nonlinear dynamic systems. In cases where the noise $e(k)$ is coloured, noise modeling should then

be considered to form a NARMAX model [14]-[16].

A generic polynomial representation of the NARX model, with a nonlinear degree of order ℓ, is

$$y(k) = c_0 + \sum_{i=1}^{n} c_0 x_i(k) + \sum_{i=1}^{n} \sum_{j=i}^{n} c_{i,j} x_i(k) x_j(k) + \cdots$$
$$+ \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \cdots \ddot{}_{i_\ell} {}_{i_\ell} \cdots \qquad (2)$$

where $n = p+q$ and

$$x_j(k) = \begin{cases} y(k-j), & 1 \le j \le p \\ u(k-(j-p)), & p+1 \le j \le n \end{cases} \qquad (3)$$

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 autoregressive with exogenous input (ARX) model [1],[2], as a special case of the NARX model (2) where $\ell = 1$ and $c_0 = 0$, is explicitly given by

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

B. The Fractional Power NARX Model

In parallel to the polynomial NARX model (2), the fractional power NARX (FPNARX) model is of the form

$$y(k) = c_0 + \sum_{i=1}^{n} c_i x_i^{\alpha_i}(k) + \sum_{i=1}^{n} \sum_{j=i}^{n} c_{i,j} x_i^{\alpha_i}(k) x_j^{\alpha_j}(k) + \cdots$$
$$+ \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \cdots \ddot{}_{i_\ell} {}_{i_\ell} \cdots \qquad (5)$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are real numbers with either integer or fractional values, the model variables $x_i(k)$ are defined by (2).

The FPNARX model can be implemented through three different approaches, namely,

- Fixed fractional powers. Powers are predetermined to be some fixed values, for example, each of the power parameters α_i ($i = 1,2, n$) in (5) is restricted to a set $P = \{p_1, p_2, \ldots, p_m\}$ where p_j is real. A simple example of such a FPNARX model of two variables $y(k$-1) and $u(k$-1), with $P = \{0.5, 1\}$ and with the inclusion of all possible cross-product terms, could be of the form:

$$\left.\begin{aligned} y(k) &= a_0 + a_1 y(k-1) + a_2 u(k-1) + a_3 y^{1.5}(k-1) \\ &+ a_4 y^{0.5}(k-1) u^{0.5}(k-1) + a_5 y^{0.5}(k-1) u(k-1) \\ &+ a_6 y^2(k-1) + a_7 y(k-1) u^{0.5}(k-1) \\ &+ a_8 y(k-1) u(k-1) + a_9 u^{1.5}(k-1) + a_{10} u^2(k-1) \end{aligned}\right\} \quad (6)$$

- Fixed model structure and variable fractional-power parameters. Below gives a simple example of such a case:

$$y(k) = a_1 y(k-1) + a_2 y(k-2) + a_3 u^{\alpha}(k-1) + e(k) \quad (7)$$

where α is a fractional number.

- Both the model structure and the fractional-power parameters are unknown and need to be identified from given experimental data.

The first case above can easily be converted to a linear-in-the-parameters form which can then be solved with a nonlinear model structure detection and parameter estimation method, for example the well-known orthogonal least squares (OLS) and error reduction ratio (ERR) algorithm [24]-[26]. The second case is a nonlinear-in-the-parameters problem and an optimization method should be employed to solve such a problem. The third case is quite complicated because of the uncertainty of the model structure, which in turn causes the operation on the unknown fractional-power parameters to be quite difficult. It is thus not advocated to employ the FPNARX model of the last case, unless some a priori information is available and the problem can be converted to the first two cases. In fact several effective alternatives to the last case are available and these include neural networks, radial basis function networks, fuzzy reasoning and approximation, kernels models, and multiresolution wavelet models and wavelet networks.

III. HARMONY SEARCH

Harmony search (HS) [22],[23] is a new meta-heuristic method for nonlinear function optimization by imitating the musical instrument improvisation process where musicians always try and improvise their instruments performance so as to achieve a preferable state of harmony. Inspired by the mechanism of swarm intelligence and particle swarm optimization (PSO) [27], Omran and Mahdavi [28] proposed an improved version of the simple HS algorithm, called the global-best harmony search (GHS). Following [27] and [28], the GHS algorithm can be outlined as below:

Step 1: Problem presentation. Formulate the optimization problem, for example, min f(x), with $L \le x \le U$, where x is a n-D vector, f(x) is an objective function, and L and U are the lower and upper boundaries (both L and U are vectors); specify relevant parameters (e.g. harmony memory size, harmony memory consideration rate, pitching adjusting rate).

Step 2: Initial condition setting. The initial harmony memory can be generated from a uniform distribution over the interval $[L_i, U_i]$ ($i = 1,2, \ldots, n$). More specifically, this can be achieved by setting $x_i^{(m)} = L_i + r(U_i - L_i)$ for $m = 1,2\ldots,M$ where M is harmony memory size and r is a random number between 0 and 1.

Step 3: Improvisation and new harmony search. Generate new harmony vector $x^{new} = [x_1^{new}, x_2^{new}, \cdots$ by utilizing memory consideration, pitch adjusting and random selection as below (in Matlab format)

```
for i = 1: n   % for each index i in {1,2, …, n}
    if r ≤ HMCR   % harmony memory consideration rate
        x_i^new = x_i^(m); %  m belongs to {1,2, …, M}
        if r ≤ PAR   %  pitch adjusting rate
            x_i^(m) = x_k^best;
            % the best harmony in the harmony memory
            % with k belonging to {1,2, …, n}
        endif
    else
        x_i^new = L_i + r(U_i − L_i);  % random selection
    endif
end
```

Step 4: Harmony memory update. The newly generated harmony vector $x^{new} = [x_1^{new}, x_2^{new}, \cdots$ can be used to replace the old harmony if it can produce a better performance in decreasing the objective function.

Step 5: Termination. The search process can be terminated either if the search iteration number exceeds the prespecified maximum number of improvisations or the value of the objective function drops to a given threshold.

IV. CASE STUDIES

This section provides three examples: the first for static function approximation, the second for fractional-power NARX model identification from simulated data, and the last one for fractional-power AR model identification from a real data set (the annual sunspot numbers).

A. Static Function Approximation

The function is defined by

$$y = f(x) = \begin{cases} 1-x, & 0 \le x \le 1/4 \\ 4x-1, & 1/4 \le x \le 1/2 \\ \sin(\pi x), & 1/2 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases} \quad (8)$$

The values of y for $x = 0, 0.05, 0.1, …, 1$ were recorded, and the 21 data points in total were then used to fit two curves by using two groups of polynomial bases:

- Integer-power polynomial: $\{1, x, x^2, …, x^{10}\}$
- Fractional-power polynomial: $\{1, x^{0.5}, x, x^{1.5}, x^2, …, x^5\}$

The polynomial models, corresponding to the above two groups of bases are:

$$y = g(x) = a_0 + a_1 x + a_2 x^2 + \cdots \quad (9)$$

$$y = h(x) = a_0 + a_1 x^{0.5} + a_2 x + a_3 x^{1.5} + a_4 x^2 + \cdots \quad (10)$$

To mimic a realistic scenario, a noise of mean zero and standard deviation of $\sigma = 0.1$ was added to the measured values for y. The ordinary least squares algorithm was then used to calculate the polynomial parameters for both the noise-free and noisy cases.
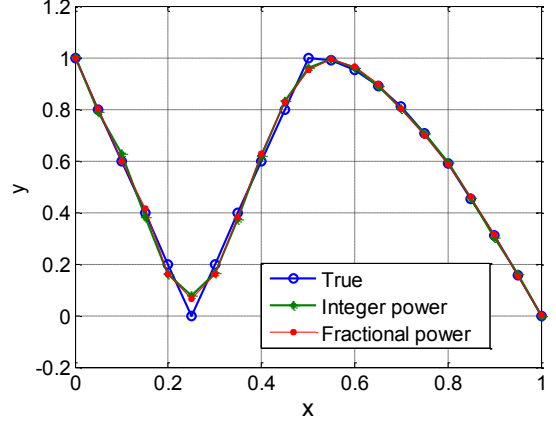

Figure 1. A comparison between the integer-power and fractional-power polynomial fitting for the noise free data.
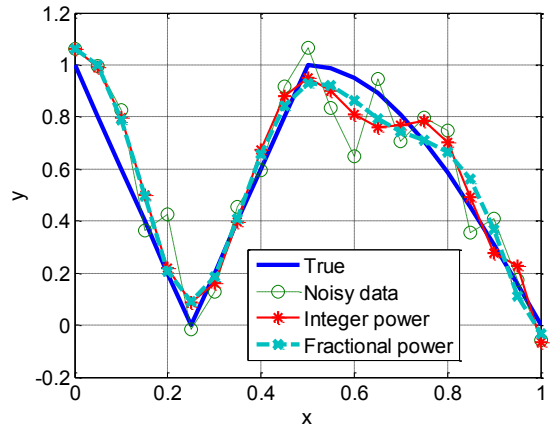

Figure 2. A comparison between the integer-power and fractional-power polynomial fitting for the noisy data.

Fig. 1 shows the results produced by the two types of polynomials fitted from the noise-free data, where the values of the mean squared errors (MSE) for the integer-power and fractional-power polynomials are 6.8307×10^{-4} and 5.7199×10^{-4}, respectively. Similarly, Fig. 2 shows the results of the two types of polynomials fitted from the noisy data, and the values of the MSE values are 0.3439 and 0.3418 respectively. Clearly, the fractional-power polynomial fitting provides a slightly better performance for both the noise-free and noisy data cases.

B. Nonlinear Input-Output Model Identification

A nonlinear Hammerstein system is described by a model structure of (7), where the model parameters $a_1 = 1.8$, $a_2 = -0.825$, $a_3 = 0.1$, and the fractional-power parameter $\alpha = 0.4$. By choosing the noise signal $e(k)$ as a Gaussian random sequence of zero mean and a standard deviation of $\sigma = 0.1$, and the input signal as a stochastic sequence that is evenly distributed in the interval [0,2], this system was simulated and a total of $N = 200$ input-output data points were recorded and were used for model parameter estimation.

The harmony search algorithm described in Section III was applied to the estimation data set of 200 points. The harmony memory size, the harmony memory consideration rate, and the pitch adjusting rate, the maximum number of improvisations were respectively chosen to be 50, 0.8, 0.1, and 1000. The parameter estimates were $\hat{a}_1$ = 1.8243, $\hat{a}_2$ = -0.8450, $\hat{a}_3$ = 0.0827, α = 0.5027. To test the efficiency of the estimated model parameters, both the identified model and the original true system model were simulated by choosing the input signal as a random sequence of a normal distribution over [0,2] (this input signal is different from that used for model parameter estimation). A comparison between the two output signals, generated by the estimated model (model predicted output or shortly MPO) and by the true model, respectively, is shown in Fig. 3, which clearly indicates that the model fit is excellent. Note that for dynamical modelling problems, the resultant models should possess a satisfactory predicative ability in terms of MPO performance, which is an extreme case of long-term prediction and which is the most stringent test for dynamical models [29]-[33].

C. Application to Real Data

The data set used in this example contains 301 observations of the annual sunspot numbers from 1700 to 2000. Denote this time series by $\{y(k)\}$ for k =1,2, …, 301. The first 280 samples for years 1700 to 1979 were used for model identification and the remaining 22 data were used for model performance testing. From our previous studies [17], [29], a total of three most significant model variables, namely, $y(k$-1), $y(t$-2), $y(t$-9), along with the *const* term, were used for the data modeling task. Unlike in conventional modeling practice, the example aims to find a fractional-power AR (FPAR) model of the form

$$y(k) = a_0 + a_1 y^{\beta_1}(k-1) + a_2 y^{\beta_2}(k-2) + a_3 y^{\beta_3}(k-9) + e(k) \quad (11)$$

where a_0, a_1, a_2, a_3 are model parameters, and β_1, β_2, β_3 are fractional-power parameters.

The harmony search algorithm was applied to the 280 estimation data points. The harmony memory size, the harmony memory consideration rate, and the pitch adjusting rate, the maximum number of improvisation were respectively chosen to be 100, 0.8, 0.2, and 5000. The parameter estimates were a_0=2.0132, a_1=3.808, a_2=-3.7935, a_3=1.1030, and β_1=0.7852, β_2=0.6355, β_3=0.6472.

A comparison of one-step-ahead predictions and the observations, over the range from year 1900 to 2000 is shown in Fig. 4. The MSE values calculated from the identified fractional-power model, over the estimation data and the test data were calculated to be 190.4193 and 300.3084, respectively, and the overall MSE value over the all 301 data points is 192.7566. To inspect the performance and efficiency of the identified fractional-power model, a traditional AR model involving exactly the same model variables was also identified by means of the ordinary least squares method. The AR model is given as

$$y(k) = 5.7080 + 1.2211y(k-1) - 0.5252y(k-2) \\ + 0.1935y(k-9) + e(k) \quad (10)$$

The value of MSE calculated from the AR model, over the estimation data and the test data is 218.5723 and 343.207, respectively, and the overall MSE value over all the 301 data points is 220.792. Clearly, the fractional-power AR (FPAR) model is more efficient and can generate a better performance than the traditional AR model in terms of short-term predictive capability. Note the better performance of the FPAR model is achieved by scarifying more time on the nonlinear model parameter estimation. For example, the that the computation time for the FPAR model was 0.58 second, while the time for the normal AR model was only 0.003 second. This implies that the computation load required by a fractional-power model identification could be much more in comparison to that required by a traditional integer-power model identification, simply because the involvement of nonlinear optimization in the FP type model parameter estimation.
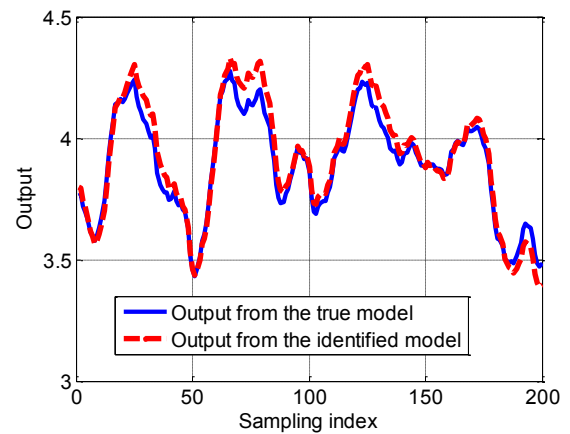

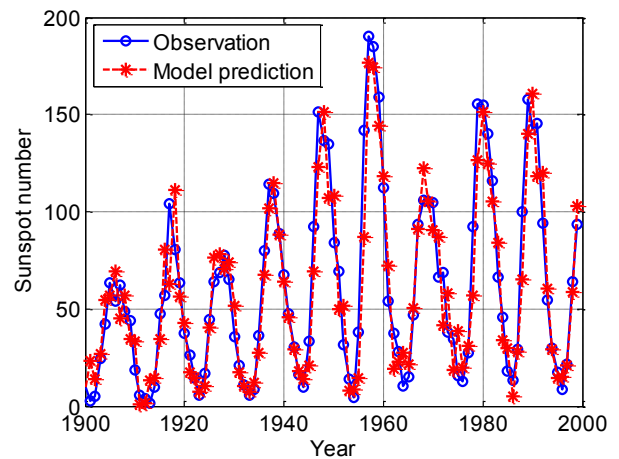Figure 3. The model predicted output and the 'true' values.

Figure 4. One-step-ahead model predictions and the observations.

V. CONCLUSION

The totally new discrete-time dynamical fractional-power autoregressive type of models has been introduced. The efficiency and performance of the newly developed model was investigated through numerical examples in relation to both simulated and real data sets. It has been shown that in cases where a priori information about the model structure is known, the fractional-power model is advantageous over its traditional counterparts in terms of model simplicity, that is, the fractional-power model is relatively compact to describe linear or nonlinear dynamic systems which would otherwise require a relatively larger number of model terms if conventional integer-power models are employed. The main disadvantage of the FPNARX model is that this type of model is relatively more complicated and difficult to operate and implement. While the first two types of FPNARX models (discussed in Section II B) can be solved by means of a model structure detection algorithm or nonlinear optimization methods, the third type of FPNARX model is quite intractable.

Future study in FPNARX modeling would include:

- A systematic and automatic process to carry out the identification of the first type of FPNARX model (Section II B). This type of FPNARX model can be treated as a natural extension of the traditional integer-power NARX model, and existing nonlinear model structure detection methods can be adapted and developed for FPNARX model structure identification and model parameter estimation.

- An investigation of methods to detect the varying range of the relevant fractional-power parameters, this would help reduce the reliance on nonlinear optimization methods.

- Model variable selection. The inclusion of redundant variables would significantly increase the computational load relevant to nonlinear optimization, which in turn would affect the performance of the identified model.

REFERENCES

[1] L. Ljung, *System identification: theory for the user*. Prentice-Hall Englewood Cliffs, NJ, 1987.

[2] T. Söderström and P. Stoica, *System Identification*. Prentice-Hall Englewood Cliffs, NJ, 1989.

[3] S. Haykin, *Neural Networks: A Comprehensive Foundation* (2nd edit). New Jersey: Princeton-Hall, 1999.

[4] G. P. Liu, *Nonlinear Identification and Control: A Neural Network Approach*. London: Springer-Verlag, 2001.

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

[6] S. A. Billings, H. B. Jamaluddin, and S. Chen, "Properties of neural networks with applications to modelling nonlinear dynamic systems," *Int. J. Control*, vol. 55, no. 1, pp. 193-224, 1992.

[7] H. L. Wei, D. Q. Zhu, S. A. Billings, and M. A. Balikhin, "Forecasting the geomagnetic activity of the *Dst* index using multiscale radial basis function networks", *Advances in Space Research*, vol. 40, no. 12, pp.1863-1870, 2007.

[8] L. X. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning," *IEEE Trans. Neural Networks*, vol. 3, no. 5, pp. 807-814, 1992.

[9] M. Mahfouf, D. A. Linkens, and D. Xue, "A new approach to model reduction for complex physiologically based drug models," Control Engineering Practice, vol. 10, no. 1, pp. 67-81, 2002.

[10] C. Campbell, "Kernel methods: a survey of current techniques," Neuralcomputing, vol. 48, pp. 63–84, 2002.

[11] H. L. Wei, S. A. Billings, and M. Balikhin, "Prediction of the Dst index using multiresolution wavelet models", *J. Geophysical Research-Space Physics*, 109(A7), Art. No. A07212, 2004.

[12] H. L. Wei and S. A. Billings, "A unified wavelet-based modelling framework for nonlinear system identification: the WANARX model structure", *Int. J. Control*, vol. 77, no. 4, pp.351-366, 2004.

[13] S. A. Billings and H. L. Wei, "The wavelet-NARMAX representation: A hybrid model structure combining polynomial models with multiresolution wavelet decompositions", *Int. J. Systems Science*, vol. 36, no. 3, pp. 137-152, 2005.

[14] I. J. Leontaritis and S. A. Billings, "Input–output parametric models for non-linear systems—part I: Deterministic non-linear systems," *Int. J. Control*, 41, pp.303–328, 1985.

[15] I. J. Leontaritis and S. A. Billings, "Input–output parametric models for non-linear systems—part II: Stochastic non-linear systems," *Int. J. Control*, 41, pp.329–344, 1985.

[16] S. Chen and S. A. Billings, "Representation of non-linear systems: the NARMAX model," *Int. J. Control*, 49, pp.1013-1032, 1989.

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

[18] T. T. Hartley, C. F. Lorenzo, H. K. Qammer, "Chaos in a fractional order Chua's system," *IEEE Trans Circuits and Systems I: Fundamental Theory and Applications*, vol. 42, no. 8, pp. 485-490, 1995.

[19] K. S. Miller and B. Ross, *An Introduction to the Fractional Calculus and Fractional Differential Equations*. New York: Wiley, 1993.

[20] R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. Van Hamme, "Parametric identification of transfer functions in the frequency domain - a survey," *IEEE Trans Automatic Control*, vol. 39, no. 11, pp. 2245-2260, 1994.

[21] I. Podlubny, "Fractional-order systems and $PI^\lambda D^\mu$-controllers," *IEEE Trans Automatic Control*, vol. 44, no. 1, pp. 208-214, 1999.

[22] Z. W. Geem, J. H. Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: harmony search," *Simulation*, vol. 76, no. 2, pp.60-68, 2001.

[23] K. S. Lee and Z. W. Geem, "A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice," *Computer Methods in Applied Mechanics and Engineering*, vol. 194, pp. 3902-3933, 2005.

[24] S. A. Billings, S. Chen, and M. J. Korenberg, "Identification of MIMO non-linear systems using a forward regression orthogonal estimator," *Int. J. Control*, 49(6), pp. 2157-2189, June 1989.

[25] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, 50(5), pp. 1873-1896, Nov. 1989.

[26] H. L. Wei, S. A. Billings, and J. Liu, "Term and variable selection for nonlinear system identification," *Int. J. Control*, 77(1), pp. 86-110, Jan. 2004.

[27] J. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm Intelligence*. San Francisco, CA: Morgan Kaufmann, 2001.

[28] M. G. H. Omran and M. Mahdavi, "Global-best harmony search," Applied Mathematics and Computation, vol. 198, no. 2, pp. 643-656, 2008.

[29] H. L. Wei, S. A. Billings, and M. Balikhin, "Prediction of the Dst index using multiresolution wavelet models," *J. Geophysical Research-Space Physics*, 109(A7), Art. No. A07212, 2004.

[30] H. L. Wei, D. Q. Zhu, S. A. Billings, and M. Balikhin, "Forecasting the geomagnetic activity of the *Dst* index using multiscale radial basis function networks," *J. Geophysical Research-Space Physics*, *Advances in Space Research*, 40(12), 1863-1870, 2007.

[31] 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. Modelling, Identification and Control*, 3(4), pp. 341-356, 2008.

[32] H. L. Wei, Y. Zheng, Y. Pan, D. Coca, L. M. Li, J. E. W. Mayhew, and S. A. Billings, "Model estimation of cerebral hemodynamics between blood flow and volume changes: A data-based modeling approach", *IEEE Trans Biomedical Engineering*, 56(6), pp. 1606-1616, 2009.

[33] H. L. Wei, S. A. Billings, A. S. Sharma, S., Wing, R. J. Boynton, and S. N. Walker, "Forecasting relativistic electron flux using dynamic multiple regression models," *Annales Geophysicae*, 29, 415-420, 2011.