Original Article

Extended analysis of bpso structure selection of nonlinear auto-regressive model with exogenous inputs (NARX) of direct current motor

Ihsan Mohd Yassin*, Mohd Nasir Taib, and Ramli Adnan

Faculty of Electrical Engineering,
Universiti Teknologi MARA, Shah Alam, Malaysia.

Abstract

System Identification (SI) is a discipline concerned with inference of mathematical models from dynamic systems based on their input and output measurements. Among the many types of SI models, the superior NARMAX model and its derivatives (NARX and NARMA) are powerful, efficient and unified representations of a variety of nonlinear systems. The identification process of NARX/NARMA/NARMAX is typically performed using the established Orthogonal Least Squares (OLS). Weaknesses of the OLS model are known, leading to various alternatives and modifications of the original algorithm. This paper extends the findings of previous research in application of the Binary Particle Swarm Optimization (BPSO) for structure selection of a polynomial NARX model on a DC Motor (DCM) dataset. The contributions of this paper involve the implementation and analysis of a MySQL database to serve as a lookup table for the BPSO optimization process. Additional analysis regarding the frequencies of term selection is also made possible by the database. An analysis of different preprocessing methods was also performed leading to the best model. The results show that the BPSO structure selection method is improved by the presence of the database, while the magnitude scaling approach was the best preprocessing method for NARX identification of the DCM dataset.

Keywords: nonlinear system identification, nonlinear autoregressive with exogenous inputs model (narx), structure selection, binary particle swarm optimization algorithm (BPSO), direct current motor

1. Introduction

System Identification (SI) is a control engineering discipline concerned with the inference of mathematical models from dynamic systems based on their input and/or output measurements (Dahunsi *et al.*, 2010; Ding *et al.*, 2010; Hong *et al.*, 2008; Kolodziej & Mook, 2011; Westwick & Perreault, 2011). It is fundamental for system control, analysis and design (Ding *et al.*, 2010), where the resulting representation of the system can be used for understanding the properties of the system (Hong *et al.*, 2008) as well as predic-

tion of the system's future behavior under given inputs and/or outputs (Hong *et al.*, 2011; Hong *et al.*, 2008).

SI is a significant research area in the field of control and modeling due to its ability to represent and quantify variable interactions in complex systems. Several applications of SI in literature are for understanding of complex natural phenomena (Evsukoff *et al.*, 2011; Gonzales-Olvera & Tang, 2010; Hahn *et al.*, 2010; Matus *et al.*, 2012; Zhao *et al.*, 2011), model-based design of control engineering applications (Dao & Chen, 2011; Fei & Xin, 2012; Gonzales-Olvera & Tang, 2010; Luiz *et al.*, 2012; Manivannan *et al.*, 2011; Oh *et al.*, 2010), and project monitoring and planning (Evsukoff *et al.*, 2011; Matus *et al.*, 2012; Pierre *et al.*, 2010; YuMin *et al.*, 2011; Zhu *et al.*, 2012).

Many SI models exist. Among them, the superior NARMAX (Leontaritis & Billings, 1985) model and its deri-

---

* Corresponding author.
 Email address: ihsan.yassin@gmail.com

vatives (Nonlinear Auto-Regressive with Exogenous Inputs (NARX) and Nonlinear Auto-Regressive Moving Average (NARMA)) are powerful, efficient and unified r epresentations of a variety of nonlinear models (Ahn & Anh, 2010; Cheng *et al.*, 2011; Chiras *et al.*, 2001; Hong *et al.*, 2008; Kukreja *et al.*, 2003; Kukreja *et al.*, 2005; Mendes & Billings, 2001; Rahim *et al.*, 2003; Shafiq & Butt, 2011; Vallverdu *et al.*, 1991). Rich literature is available regarding its success in various electrical, mechanical, medical and natural applications (Ahn & Anh, 2010; Barbosa *et al.*, 2011; Chen & Ni, 2011; Cosenza, 2012).

Simultaneous structure selection and parameter estimation of NARMAX and derivative models are achievable through the Orthogonal Least Squares (OLS) algorithm (Billings *et al.*, 1988; Piroddi, 2008). The OLS algorithm has since been widely accepted as a standard (Hong *et al.*, 2011; Wei & Billings, 2008a, 2008b) and has been used in many works (see (Aguirre & Furtado, 2007; Barbosa *et al.*, 2011; Dimogianopoulos *et al.*, 2009; Er *et al.*, 2010; Liao, 2010; Wei & Billings, 2008a; Wong *et al.*, 2011; Yunna & Zhaomin, 2008)) due to its simplicity, accuracy and effectiveness (Er *et al.*, 2010; Hong *et al.*, 2008).

Despite OLS's effectiveness, several criticisms have been directed towards its tendency to select excessive or sub-optimal terms (Aguirre & Furtado, 2007; Cheng *et al.*, 2011; Piroddi, 2008; Wei & Billings, 2008a, 2008b). A demonstration by Wei & Billings (2008b) proves that OLS selected incorrect terms when the data is contaminated by certain noise sequences, or when the system input is poorly designed (Wei & Billings, 2008b). The suboptimal selection of regressor terms leads to models that are non-parsimonious in nature.

An alternative structure selection method for NARX model using the Binary Particle Swarm (BPSO) algorithm is presented in this paper. It extends the findings by Yassin *et al.* (2010) with algorithm speed-ups and new structure selection analysis methods based on a MySQL database lookup table, as well as expanding the solution to investigate four preprocessing combinations.

The remainder of this paper is organized as follows: a review of current structure selection methods is presented in Section 3, followed by fundamental theories in Section 4. The research methodology is presented in Section 5, and the corresponding results are shown in Section 6. Finally, concluding remarks and future works are presented in Section 7.

2. Review of Structure Selection Methods

Structure selection is defined as the task of selecting a subset of regressors to represent system dynamics (Aguirre & Letellier, 2009; Amrit & Saha, 2007; Piroddi, 2010; Wei & Billings, 2008b). The objective of structure selection is model parsimony – the model should be able to explain the dynamics of the data using the least number of regressor terms (Amrit & Saha, 2007; Wei & Billings, 2008a).

In its simplest form, the structure selection task involves the determination of the optimal lag space (Cassar *et al.*, 2010). Selecting a higher lag space incorporates more history into the model leading to better prediction (Amrit & Saha, 2007). However, this approach increases the computational complexity of the model, as well as reducing its generalization ability (Amrit & Saha, 2007; Cheng *et al.*, 2011; Hong *et al.*, 2008; Wei & Billings, 2008b). Therefore, advanced methods utilize quality measures or criteria to select the most important regressors (Aguirre & Letellier, 2009). A review of several structure selection methods is deliberated in the following sections.

2.1 Trial and error methods

Two trial and error methods for structure selection were found in the literature. The first method, called Zero-and-Refitting (Aguirre & Letellier, 2009) estimates the parameters for a large model and gradually eliminate regressors that have sufficiently small parameter values. This method generally does not work well in practice as the parameter estimates are sensitive to the sampling window, noise variance and non-normalized data (Aguirre & Letellier, 2009).

Another method was found in Chen & Ni (2011) for structure selection of a Bayesian NARX Artificial Neural Network (ANN). The method constructs multiple models with different input lags and hidden units in an effort to find the best ANN and data structure. This method works well if the number of parameter combinations is small, but can be overwhelming when the number of adjustable parameters grows. Similarly, in Piroddi (2008), an iterative two-stage NARX model construction method was presented, which created models by adding and deleting regressors to minimize model simulation error. The authors reported good accuracy with low computational cost.

2.2 OLS

The OLS algorithm is a simultaneous structure selection / parameter estimation algorithm introduced by (Billings *et al.*, 1988; Piroddi, 2008). The algorithm first transforms the original regressors into orthogonal vectors using methods such as Gram-Schmidt, Modified Gram-Schmidt, Householder transform, or Givens Rotation (Chen *et al.*, 1989; Cheng *et al.*, 2011; Ning *et al.*, 2011). This step decouples the regressors so that their individual contributions are estimated separately from the rest (Balikhin *et al.*, 2011; Kibangou & Favier, 2010; Mendes & Billings, 2001; Piroddi, 2008, 2010; Piroddi & Lovera, 2008; Wei & Billings, 2008b).

An Error Reduction Ratio (ERR) measure is then used to rank the contribution of each regressor towards reducing the approximation error of the model (Balikhin *et al.*, 2011; Piroddi, 2008; Wei & Billings, 2008b). The ERR performs this by iteratively adding regressors from an initial empty model (Piroddi & Lovera, 2008), and evaluating regressors combi-

nations that has the most influence over the variance of the system output (Balikhin *et al.*, 2011; Gandomi *et al.*, 2010). Regressors with highest ERR values are deemed most significant and selected to be included in the final model structure (Cheng *et al.*, 2011). The cutoff of selected regressors is determined based on a threshold (Wei & Billings, 2008a). Once the model structure has been determined, the parameters are estimated using Least-Squares (LS) or its variants (Aguirre & Furtado, 2007; Barbosa *et al.*, 2011; Chen *et al.*, 2011; Dimogianopoulos *et al.*, 2009).

The OLS algorithm has since been widely accepted as a standard (Hong *et al.*, 2011; Wei & Billings, 2008a, 2008b) and has been used in many works (see Aguirre & Furtado, 2007; Barbosa *et al.*, 2011; Dimogianopoulos *et al.*, 2009; Wei & Billings, 2008a for examples), and (Er *et al.*, 2010; Liao, 2010; Wong *et al.*, 2011; Yunna & Zhaomin, 2008) for variations of applications) due to its simplicity, accuracy and effectiveness (Er *et al.*, 2010; Hong *et al.*, 2008).

Among the advantages of OLS are that the decoupled and decomposed nature of the orthogonalized regressors allow easy measurement and ranking of each regressor contributions (Wei & Billings, 2008b). Furthermore, OLS can structure selection without a priori knowledge of the system (Bai & Deistler, 2010), a condition common in many real-life modeling scenarios. The algorithm has a successful track record in the field of SI (Aguirre & Letellier, 2009).

Despite OLS's effectiveness, several criticisms have been directed towards its tendency to select excessive or sub-optimal terms (Aguirre & Furtado, 2007; Cheng *et al.*, 2011; Piroddi, 2008; Wei & Billings, 2008a, 2008b), sensitivity to initialization and the order of regressor inclusion (Piroddi, 2008), repetitive orthogonalization procedure (Ning *et al.*, 2011), and bias in regressor selection (Piroddi, 2010; Wei & Billings, 2008b). Furthermore, a demonstration by (Wei & Billings, 2008b) proves that OLS selected incorrect terms when the data are contaminated by certain noise sequences, or when the system input is poorly designed (Wei & Billings, 2008b). The predetermined threshold for selection of regressors also needs to be empirically or heuristically determined (Wei & Billings, 2008a), although it is usually set to a sufficiently high value.

Several authors have made attempts to address the above issues (addition of generalized cross-validation with hypothesis testing (Wei & Billings, 2008b), OLS with exhaustive searching (Mendes & Billings, 2001), regularized OLS (Wong *et al.*, 2011), incorporation of information criteria, correlation analysis and pruning methods (Cheng *et al.*, 2011; Hong *et al.*, 2011; Piroddi, 2008, 2010; Wei & Billings, 2008b), and hybrid OLS algorithms (Gandomi *et al.*, 2010; Piroddi, 2010)).

2.3 Clustering methods

Clustering methods group data based on their similarities and perform structure selection by adding a cluster to or removing them from the model structure. Clustering methods

are determined by two parameters, namely the number of clusters and the initial location of cluster centers (Tesli *et al.*, 2011).

In Aguirre & Letellier (2009), nearest neighbor clustering was performed to determine the input/output lag space of a NARX model. After clustering, excess clusters can then be deleted from the model to achieve the desired model behavior. In (Juang & Hsieh, 2010), clustering was used to implement the optimal structure of a recurrent fuzzy ANN. The authors reported a reduction of the ANN size with the cluster-based structure determination method. A Gustafsson-Kessel fuzzy clustering algorithm was used in (Tesli *et al.*, 2011) to perform unsupervised data clustering, and the data were then modeled using local linear models. Clustering using Independent Component Analysis (ICA) has also been reported (Talmon *et al.*, 2012).

As reported by Aguirre & Letellier (2009) and Juang & Hsieh (2010), clustering improves the overall model structure either by removing excess terms that are collectively similar or grouping them together to create a compact classifier. However, a small number of references reviewed used clustering, which suggests that this type of method is in its infancy and does not possess a significant track record in current SI methodology.

2.4 Correlation-based methods

Correlation-based methods have been used in Cheng *et al.* (2011), Tan & Cham (2011) and Wei & Billings (2008a, 2008b). Correlation-based methods reveal causal relationships between regressors (Balikhin *et al.*, 2011), which can then be used to determine important terms to include in the model.

Correlation-based analysis guided by OLS was used in Cheng *et al.* (2011) to select significant model inputs for a polynomial model. Correlation analysis was performed to evaluate inputs that make a large contribution towards the output. The candidate inputs are refined further through the use of multi-objective evolutionary optimization to select the optimal model structure.

Works by Wei & Billings (2008b) used a combination of generalized cross-validation, OLS and hypothesis testing to perform model structure selection. An integration between the adaptive orthogonal search algorithm and the Adjustable Prediction Error Sum of Squares (APRESS) was presented in Wei & Billings (2008a) for structure selection. The authors reported that the simple algorithm achieved a good balance between reducing bias and improving variance of the model. Additionally, in Tan & Cham (2011), a gridding method combined with cross-correlation testing was used to determine the lags of an online SI model.

Balikhin *et al.* (2011) cautioned against the use of correlation-based methods because although correlations indicate causal relationship, qualitative analysis based on these methods are inaccurate for nonlinear systems modeling. The authors preferred the ERR analysis instead for their analysis.

2.5 Other structure selection methods

Apart from the methods listed, a structure selection technique for time-series models was presented in Harris & Yu (2012). The method used Monte-Carlo simulation and Analysis of Variance (ANOVA) sensitivity analysis to decompose the variance of regressors. The variance decompositions quantify the effect of variables in the model, and are used as a guide for structure selection.

3. Theoretical Background

3.1 NARX

The NARX model is a generalized version of the Auto-Regressive with Exogenous Inputs (ARX) model (Zhao, Chen & Zheng, 2010). The NARX model takes the form of (Amisigo *et al.*, 2007):

$$y(t) = f^d\left[y(t-1), y(t-2), \ldots, y(t-n_y), u(t-n_k),\right.$$
$$\left. u(t-n_k-1), \ldots, u(t-n_k-n_u)\right] + \varepsilon(t) \tag{1}$$

where f^d is the estimated model, $y(t-1), y(t-2), \ldots y(t-n_y)$ are lagged output terms, $u(t-n_k), u(t-n_k-1), \ldots u(t-n_k-n_u)$ are current and/or lagged input terms and $\varepsilon(t)$ are the white noise residuals. Parameter n_k is the input signal time delay; its value is usually 1 except for cases where the input $u(t)$ is required for identification (in which case, $n_k = 0$) (Amisigo *et al.*, 2007). The lagged input and output terms can also be multiplied with each other to model higher-order dynamics beyond single-term polynomials.

The NARX identification procedure is performed in a non-recursive manner because of the absent residual feedback. The NARX/NARMAX model can be constructed using various methods, such as polynomials (Amisigo *et al.*, 2007; Anderson *et al.*, 2010; Billings & Chen, 1989; Chen *et al.*, 1989), Multilayer Perceptrons (MLP) (Norgaard *et al.*, 2000a; Rahim, 2004; Rahim *et al.*, 2003; Rahiman, 2008), and Wavelet ANNs (WNN) (Billings & Wei, 2005; Wang *et al.*, 2009), although the polynomial approach is the only method that can explicitly define the relationship between the input/output data. The data used for NARMAX and its derivatives can be either be continuous or discrete in nature (Billings & Coca, 2001). For continuous data, the dataset needs to be sampled at regular and fixed intervals in order to discretize it for identification (Billings & Aguirre, 1995; Billings & Coca, 2001).

The identification method for NARMAX and its derivatives are performed in three steps (Balikhin *et al.*, 2011). Structure selection is performed to detect the underlying structure of a dataset. This is followed by parameter estimation to optimize some objective function (typically involving the difference between the identified model and the actual dataset) (Chen & Ni, 2011). Finally, the model is validated using One Step Ahead (OSA) and correlation tests to ensure that it is valid and acceptable.

3.2 BPSO for NARX model structure selection

The use of BPSO for model structure selection is described in this section. Consider a SI problem in Eq. (2):

$$P\theta + \varepsilon = y \tag{2}$$

where P is a $n \times m$ regressor matrix, θ is a $m \times 1$ coefficient vector, and y is the $n \times 1$ vector of actual observations. P is arranged such that its columns represent the m lagged regressors. ε is the white noise residuals.

The BPSO algorithm (Kennedy & Eberhart, 1997) defines a binary string of length $1 \times m$ so that each column in P has a bit assigned to it. A value of 1 indicates that the column is included in the reduced regressor matrix, P_R, while the value of 0 indicates that the regressor column is ignored. The initial binary string is a predefined parameter prior to optimization.

In the swarm, each particle carries a $1 \times m$ vector of solutions, x_{id}. This vector contains the bit change probability. During optimization, the x_{id} values change, and alter which regressor column is selected. The linear least squares solution (θ_R) for the reduced regressor matrix, P_R, can then be estimated using QR factorization method described in Eq. (3) to Eq. (6) (Amisigo *et al.*, 2007):

$$P_R\theta_R + \varepsilon = y \tag{3}$$

$$P_R = Q_R R_R \tag{4}$$

$$g_R = Q_R^T y \tag{5}$$

$$R_R\theta_R = g_R \tag{6}$$

Finally, by rearranging and solving Eq. (7), the value of θ_R can be estimated:

$$\theta_R = R_R^T g_R \tag{7}$$

4. Methodology

4.1 Dataset description

Control of electromechanical devices is an important area of research, as these devices form an integral part of positioning and tracking systems (Tjahjowidodo *et al.*, 2005). The DC motor is such a device, which is widely used in engineering due to its structural simplicity, excellent control performance and minimal cost (Cong *et al.*, 2010). System identification DC motor has enjoyed considerable interest for the purposes of fault diagnosis (Simani, 2002), friction modeling (Tjahjowidodo *et al.*, 2005), parameter estimation (Al-Qassar & Othman, 2008; Cong *et al.*, 2010) and control design (Nasri *et al.*, 2007).

The DCM dataset (Rahim, 2004) is a Single Input Single Output (SISO) nonlinear system relating the angular velocity of a Direct Current (DC) motor in response to the Pseudo-Random Binary Sequence (PRBS) input voltage at its terminals (Figure 1). The Simulink-generated DCM dataset consists 5,000 data points and has been used in Rahim (2004)
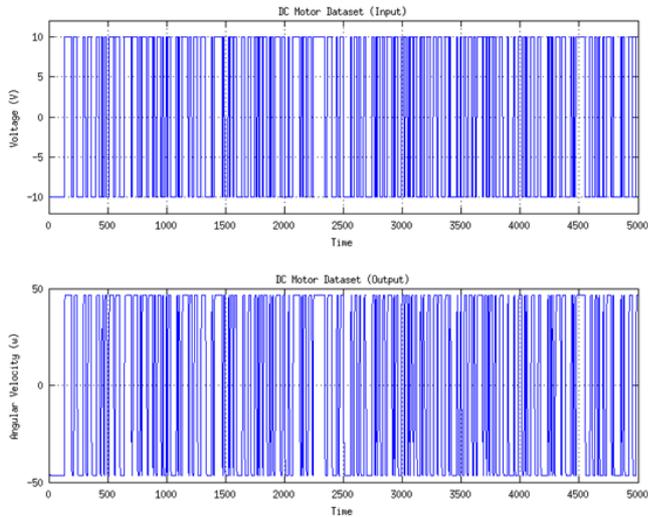
Figure 1. The DCM dataset

Table 1. Preprocessing methods used and their respective codes

Code	Preprocessing Method
00	No magnitude scaling, block division
01	No magnitude scaling, interleaving
10	Magnitude scaling, block division
11	Magnitude scaling, interleaving

for identification of a MLP-based SI model. The system exhibits a model order of two, as determined by experiments conducted by Rahim (2004).

A total of four preprocessing combinations were used on the dataset throughout this paper. The preprocessing methods are shown in Table 1. Magnitude scaling is used to adjust the magnitude of the data to within an acceptable range. Several papers that utilize this preprocessing method are Ahn & Anh (2010); Chen & Ni (2011) and Er *et al.* (2010). Several authors set the range according to arbitrary values (Ahn & Anh, 2010; Er *et al.*, 2010), although Artificial Neural Network (ANN) practitioners tend to set it to between -1 and +1 (Chen & Ni, 2011).

It is preferable to separate training and testing datasets. A model should have good accuracy not only over the training dataset, but also on the independent testing set (Hong *et al.*, 2008). The training dataset is used for updating the parameters of the model. Therefore, the model will naturally fit this data well. Therefore, the testing dataset is important because it serves as an independent measure to evaluate the model's generalization ability. Among others, the division of data can be performed using methods such as block or interleaving division (Beale *et al.*, 2011; Cosenza, 2012; Er *et al.*, 2010; Gandomi *et al.*, 2010; Rahiman, 2008). Block division divides the dataset to blocks according to a predetermined ratio, while interleaving divides the dataset according to the

even and odd data positions in the dataset.

A total of 14 regressors terms were generated as a result of polynomial degree two, as shown in Table 2. Therefore, the size of the solution space (consisting of all combination permutations of the terms) is $2^{14} = 16,384$.

4.2 NARX structure selection using BPSO

After the regressor matrices had been constructed, BPSO was applied to solve the structure selection problem. BPSO convergence depends on several parameters, namely swarm size, maximum iterations and initial random seeds. Therefore, the experiments in this section were done by performing optimization under various combinations of the parameters swarm size, maximum iterations and random seeds.

The BPSO parameter values are shown in Table 3. The choice of swarm size and maximum iterations was based on preliminary tests to balance speed and solution quality. These values were considered optimal given the limited computational performance. Three random seeds were chosen for the Mersenne-Twister pseudorandom number generator to be used as BPSO's initial seed. The values are arbitrary but important to ensure that the experiments are repeatable. The values of x_{min} and x_{max} were set to 0 and 1, respectively. They are within the range of probability values for bit change to occur. v_{min} and v_{max} represent the movement range of the particles. Since the value of x_{id} is between the range of 0 and 1 (based on x_{min} and x_{max}), the values of v_{min} and v_{max} were set to -1 (when x_{id} moves from 1 to 0) and +1 (when x_{id} moves from 0 to 1), respectively. The values of C_1 and C_2 were both set to 2.0. This parameter is well-accepted as optimal based on literature (El-Nagar *et al.*, 2011).

After structure selection has been performed, the resulting candidate models need to be validated and analyzed to determine the best model. Fitting and residual tests were performed to select the best model that fulfills the validation criteria. Analysis of regressor selection frequency was also

Table 2. Terms used in NARX DCM model

No.	Term
1.	$u(t-1)$
2.	$u(t-2)$
3.	$y(t-1)$
4.	$y(t-2)$
5.	$u(t-1) * u(t-1)$
6.	$u(t-1) * u(t-2)$
7.	$u(t-1) * y(t-1)$
8.	$u(t-1) * y(t-2)$
9.	$u(t-2) * u(t-2)$
10.	$u(t-2) * y(t-1)$
11.	$u(t-2) * y(t-2)$
12.	$y(t-1) * y(t-1)$
13.	$y(t-1) * y(t-2)$
14.	$y(t-2) * y(t-2)$

Table 3. BPSO parameter settings for structure selection experiments

Parameter	Value
Swarm Size	10, 20, 30, 40, 50
Fitness Criterion	Akaike Information Criterion (AIC)
	Final Prediction Error (FPE)
	Minimum Descriptor Length (MDL)
Maximum Iterations	250, 500, 750
Initial Seed	0, 10 000, 20 000
x_{min}	0
x_{max}	1
v_{min}	-1
v_{max}	+1
C_1	2.0
C_2	2.0

performed to discover regressor selection patterns of the BPSO algorithm.

4.3 Database description

The Entity-Relationship Diagram (ERD) for the database tables is shown in Figure 2. A total of 11 fields were defined in the table structure. The description for each field is presented in Table 4. The records in the database serve as the lookup table for BPSO during the experiment.

In order to measure the performance of the database during the lookup operation, a simulated structure selection experiment was performed on the DCM dataset. The parameters for the experiment are shown in Table 5. The test was performed using the profile() function in MATLAB. The function measures the times taken by each of the functions in the structure selection program to complete. The result of this experiment is important to justify the use of the database
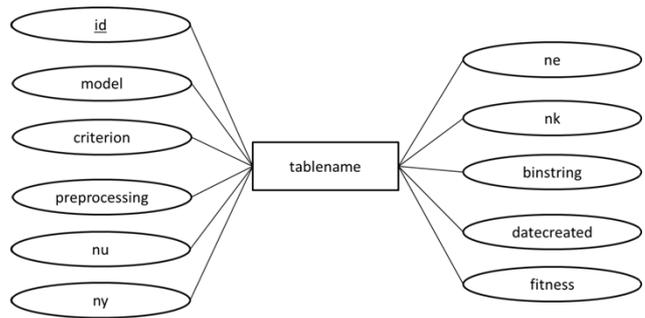
Figure 2. ERD for MySQL lookup table

Table 4. Field descriptions for database tables

Field	Type	Allowed Values	Description
id	integer (max 255-bit), auto-generated, unique, primary key.	Integer between the range of 0 to 2^{225}.	Primary key, a unique reference for each record in the database table.
model	Variable-length string (max 1,000 characters)	'narx', 'narma' or 'narmax'	Indicates the model used.
criterion	Variable-length string (max 1,000 characters)	'aic', 'fpe' or 'mdl'	Indicates which criterion was used to evaluate the structure fitness.
preprocessing	Variable-length string (max 1,000 characters)	'00', '01', '10', '11'	Preprocessing method used. First digit indicates whether magnitude scaling method was used, second digit indicates whether interlacing was used. If interlacing was not used, the value 0 means the default block division.
nu	integer (max 255-bit)	Integer between the range of 0 to 2^{225}.	Maximum lag space for input regressors.
ny	integer (max 255-bit)	Integer between the range of 0 to 2^{225}.	Maximum lag space for output regressors.
ne	integer (max 255-bit)	Integer between the range of 0 to 2^{225}.	Maximum lag space for residual regressors.
nk	integer (max 255-bit)	Integer between the range of 0 to 2^{225}.	Input delay.
binstring	string (max 5,000 characters)	Binary sequence.	Binary sequence depicting structure of reduced psi matrix.
fitness	Double value	Floating point value.	Fitness of the AIC/FPE/MDL criterion for the given binary sequence.
datecreated	timestamp	Date and time	Date and time the record was created.

Table 5.   Parameters for database performance test

Parameter	Value
Dataset	DCM
Criterion	AIC
Swarm Size	10
Maximum Iterations	250
Initial Random Seed	0

as a lookup table to speed up computations, as well as to verify the execution flow of the lookup table.

5. Results and Discussion

5.1 Database performance results

The database performance experiment was done to verify the execution flow and measure the performance and of the database during its search and insertion cycles. The search and insertion process is illustrated in Figure 3. During optimization, the BPSO algorithm searches the database and checks whether a particular structure has been found before. If yes, then the fitness value in the corresponding record is returned to BPSO and the fitness calculation process was skipped (Execution Flow 1). However, if the structure has not been evaluated yet, the fitness value is calculated and stored as a new record inside the database (Execution Flow 2).

During optimization, new records are added to the database. A total of 1,528 unique records were generated from 10,000 evaluations of the fitness function, which accounts for 15.28% of the total evaluations. This observation suggests that a significant number (84.72%) of the solutions was recalled directly from the database instead of being calculated redundantly during optimization.

The results of the MATLAB profile() operation is shown in Figure 4. The fitnessFcn() function took 126.38 seconds to complete. Most of the time spent in fitnessFcn() was during database search (dbSearch) and record insertion (dbInsert). Execution Flow 1 requires only the dbSearch function to search the database, while Execution Flow 2 requires both dbSearch (search records) and dbInsert to complete (calculate and add records). Function dbSearch was called on every fitness function evaluation to search the database, while function dbInsert was called 1,528 times. This corresponds to the number of unique records found by the optimization. These function call findings confirm the execution flow of BPSO on this dataset.

A timing diagram (Figure 5) was constructed based on average execution times of fitnessFcn(), dbSearch() and dbInsert(). The diagram clearly shows that the structure selection process benefited from the retrieval of records (Execution Flow 1) from the database as the process took 82.28% less time to complete compared to the search, fitness calculation and record insertion (Execution Flow 2).

5.2 BPSO database analysis

The full BPSO method generated a total of 42,616 unique solutions for all four preprocessing methods. The breakdown of solutions according to their preprocessing methods is shown in Table 6. The distribution of records indicates that the BPSO has covered approximate 15% to 35% of the solution space with small fitness values indicating good model fit. A large range of fitness values was also observed with small mean. The high maximum values are explained by the initial lack of optimization during the initialization phase. The average and fitness values were lower, indicating that progressively better solutions were found as optimization progressed. Magnitude scaling appears to have a positive effect, while interleaving has a negative effect on
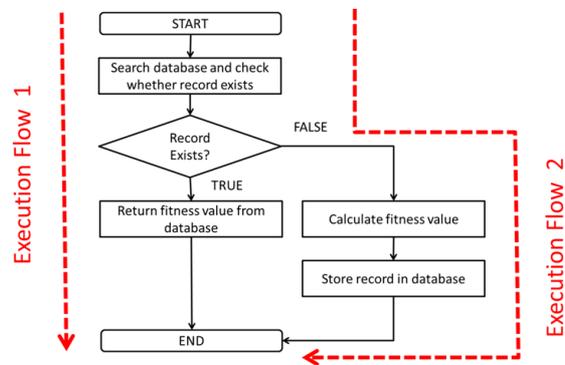

Figure 3.   Database search and insert cycle for BPSO lookup table
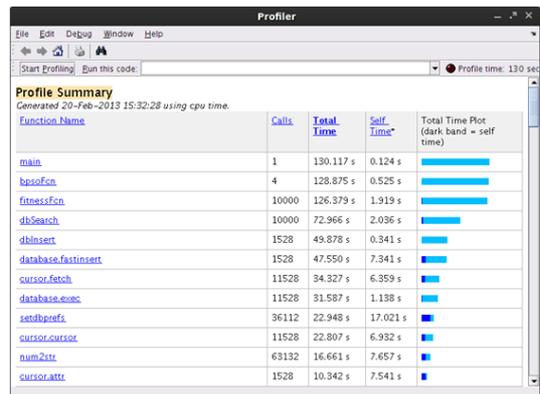

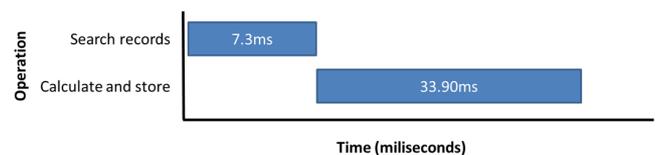Figure 4.   Result of MATLAB profile operation

Figure 5.   Timing diagram of BPSO structure selection of DCM dataset

Table 6.   Breakdown of solutions found by BPSO on DCM NARX model

Pre-processing	Criterion	No. Records (Percent of Solution Space)	Fitness Value		
			Min	Average	Max
00	AIC	2,872 (17.53%)	3.6714e-04	28.8643	963.9526
	FPE	2,553 (15.58%)	3.6811e-04	30.8811	970.4860
	MDL	2,983 (18.21%)	3.6715e-04	28.8560	963.9637
01	AIC	4,465 (27.25%)	4.5055e-04	58.9656	989.9565
	FPE	3,848 (23.49%)	4.5155e-04	67.2147	992.7941
	MDL	4,442 (27.11%)	4.5055e-04	59.0471	989.9679
10	AIC	2,373 (14.48%)	1.6883e-07	1.4252e-02	0.4433
	FPE	2,485 (15.17%)	1.6928e-07	1.4831e-02	0.4463
	MDL	2,562 (15.64%)	1.6883e-07	1.4095e-02	0.4433
11	AIC	5,516 (33.67%)	2.0724e-07	2.6651e-02	0.4550
	FPE	3,076 (18.77%)	2.0770e-07	2.8942e-02	0.4559
	MDL	5,441 (33.21%)	2.0724e-07	2.6829e-02	0.4550

fitness. The best results were obtained using preprocessing method 10.

An analysis of frequency of term selection was also performed for the NARX DCM model. The objectives of the test were: 1) to determine whether the BPSO algorithm considered all terms to be included in the model, and 2) whether any selection patterns were present that indicates preference towards high-performing terms. If achieved, both objectives indicate good optimization qualities: good exploration combined with exploitation of potential solutions. The term selection frequencies for the DCM NARX model are shown in Figure 6 to Figure 9.
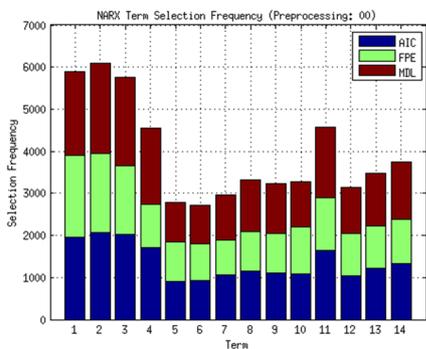
Figure 6.   Histogram of term selection (DCM NARX – preprocessing method 00)
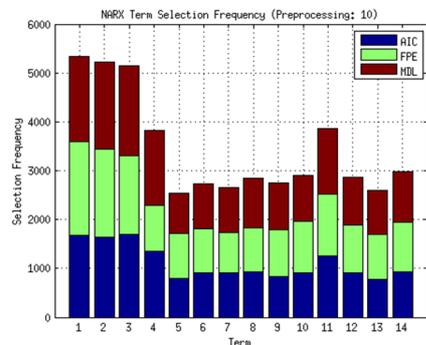
Figure 8.   Histogram of term selection (DCM NARX – preprocessing method 10)
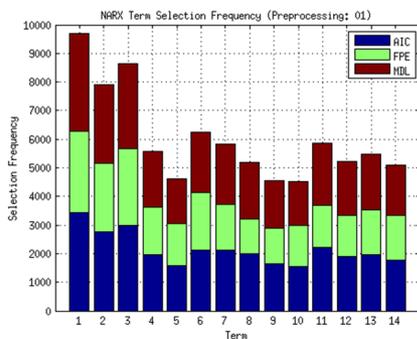
Figure 7.   Histogram of term selection (DCM NARX – preprocessing method 01)
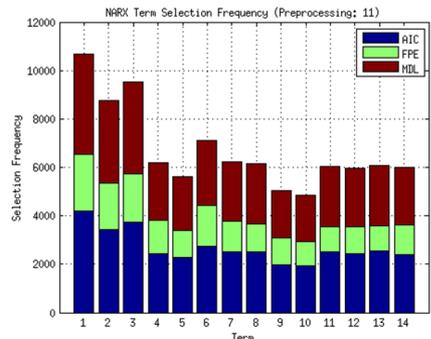

Figure 9.   Histogram of term selection (DCM NARX – preprocessing method 11)

Figure 6 to Figure 9 indicates that BPSO considered all candidate terms during optimization, as all terms are selected at least once by the BPSO algorithm. Additionally, high-performing terms (terms 1, 2, 3, 11) were consistently selected more than the rest. This indicates that the BPSO search concentrates towards high-performing candidate structures with good fitness values.

5.3 BPSO identification results

This section describes the identification results of the DCM NARX models constructed using BPSO structure selection method. The results are focused on preprocessing method 10 as this method obtained the lowest criterion scores for BPSO. BPSO obtained the following model using the AIC and MDL criterions (Eq. (8)):

$$y(t) = 0.4955u(t-1) + 0.0014u(t-2) + 0.4794y(t-1)$$
$$+ 5.8281e^{-3}y(t-2) + 4.5334e^{-5}u(t-2)*y(t-2) + \varepsilon(t)$$
(8)

while the following model was obtained using the FPE criterion (Eq. (9)):

$$y(t) = 0.4995u(t-1) + 0.0083u(t-2) + 0.4911y(t-1)$$
$$+ 4.5329e^{-5}u(t-2)*y(t-2) + \varepsilon(t)$$
(9)

All models constructed using BPSO selected a second-order term $u(t-2)*y(t-2)$. Although the contribution of the additional term was small (indicated by the small coefficient value), it highlights an interaction between second-order terms in the model ignored by OLS (in our preliminary experiments). The models were then validated. A summary of the validation results is shown in Table 7.

The BPSO validation results for the AIC/MDL and FPE models are shown in Figure 10 to Figure 27. Figures 10,

11, 19 and 20 show the One-Step-Ahead (OSA) training and testing set predictions of the best model selected by BPSO using the AIC, FPE and MDL criterions. All the figures show high overlap between the estimated and actual data. Additionally, the Pearson R-Squared test yields a score of 100%.
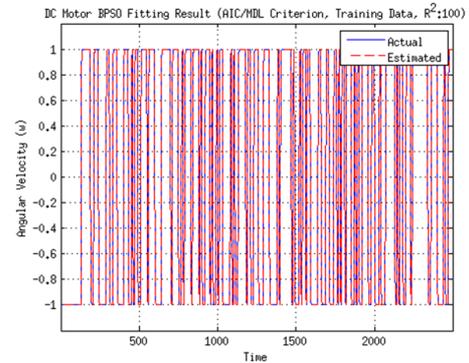

Figure 10. BPSO DCM NARX OSA model fit (training set) (AIC/MDL criterion) – preprocessing method 10
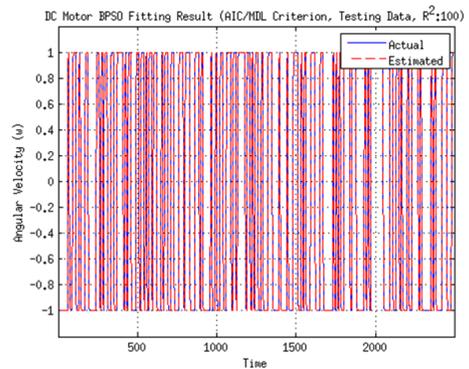

Figure 11. BPSO DCM NARX OSA model fit (testing set) (AIC/MDL criterion) – preprocessing method 10

Table 7. Validation summary of BPSO-identified DCM NARX model (Preprocessing: 10)

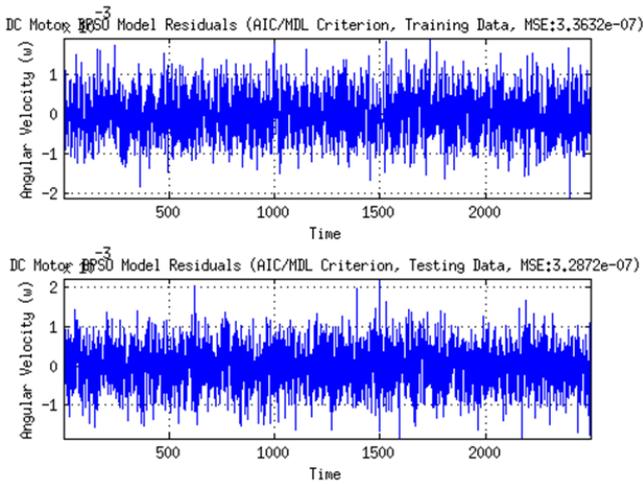| Fitness Criterion	Eval. Criterion	Training Set	Testing Set
AIC/MDL	Times Found	AIC: 33, MDL: 33	
	AIC	1.6883e-07	1.6502e-07
	FPE	1.6930e-07	1.6548e-07
	MDL	1.6883e-07	1.6502e-07
	R-squared (%)	100	100
	CRV	5	1
	MSE	3.3632e-07	3.2872e-07
FPE	Times Found	FPE: 31	
	AIC	1.6890e-07	1.6474e-07
	FPE	1.6928e-07	1.6510e-07
	MDL	1.6890e-07	1.6474e-07
	R-squared (%)	100	100
	CRV	6	3
	MSE	3.3672e-07	3.2842e-07

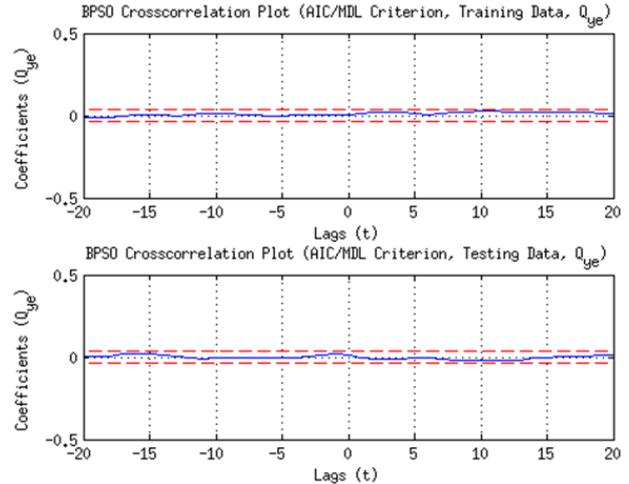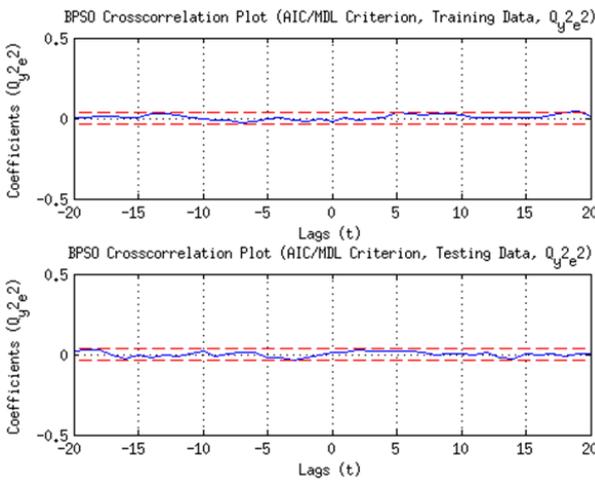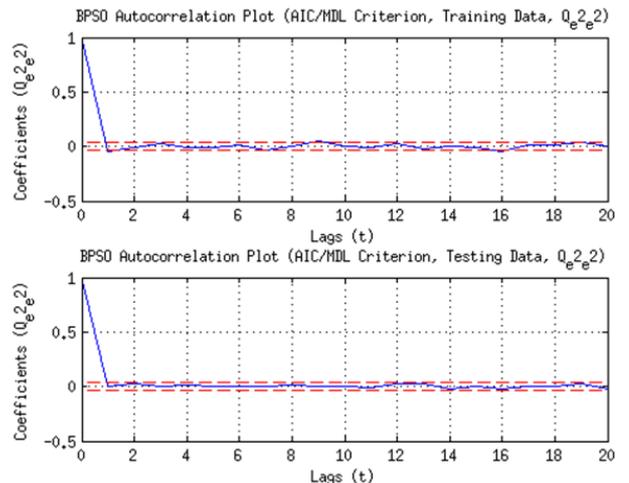Figure 12. BPSO DCM NARX residual plot (AIC/MDL criterion) – preprocessing method

Figure 13. BPSO DCM NARX correlation test result (1/5) (AIC/ MDL criterion) – preprocessing method 10
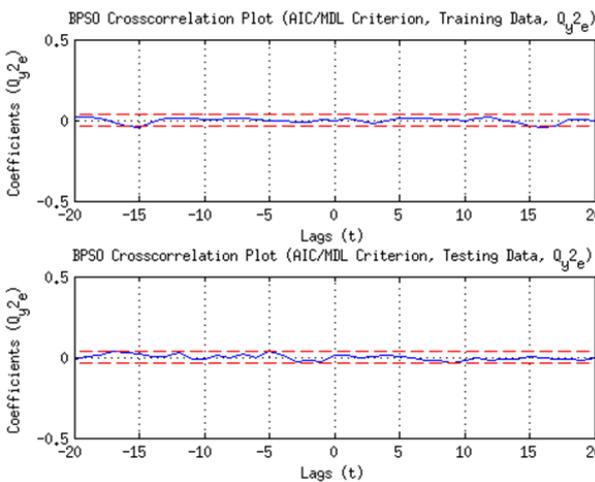

Figure 14. BPSO DCM NARX correlation test result (2/5) (AIC/ MDL criterion) – preprocessing method 10
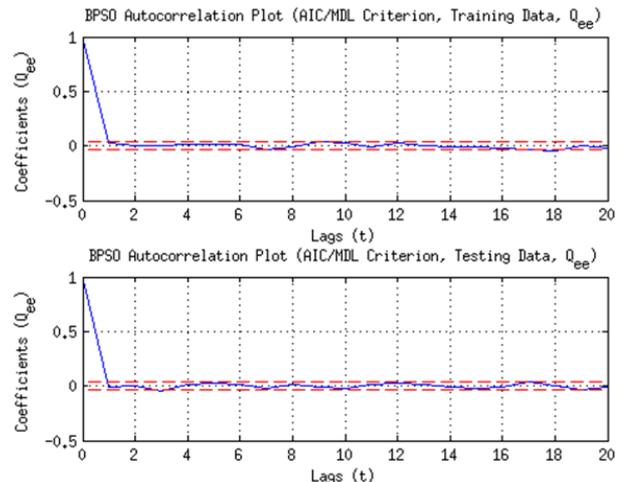

Figure 15. BPSO DCM NARX correlation test result (3/5) (AIC/ MDL criterion) – preprocessing method 10

Figure 16. BPSO DCM NARX correlation test result (4/5) (AIC/ MDL criterion) – preprocessing method 10

Figure 17. BPSO DCM NARX correlation test result (5/5) (AIC/ MDL criterion) – preprocessing method 10
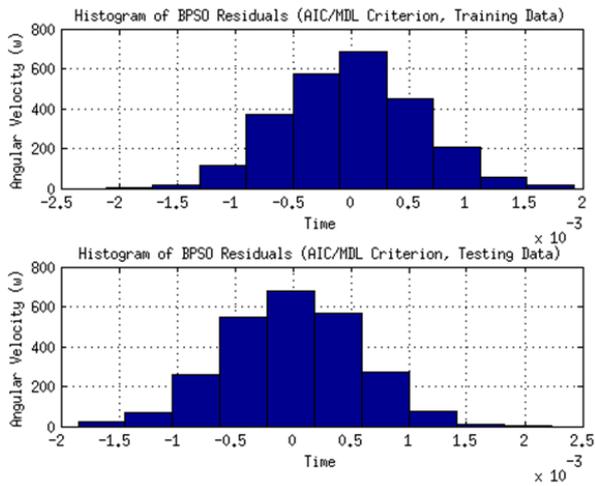
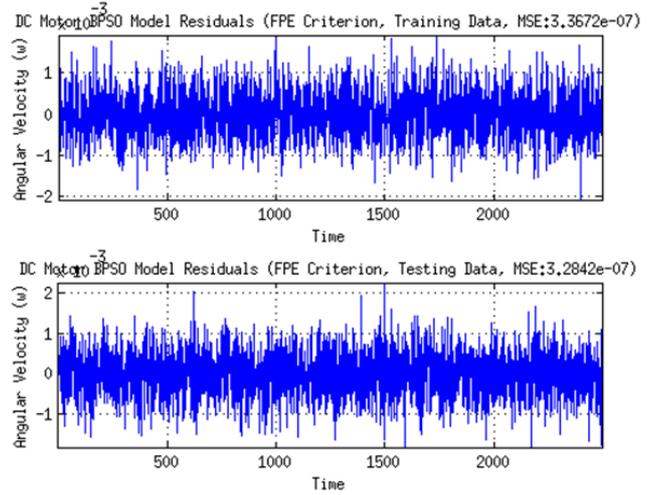Figure 18. BPSO DCM NARX histogram of residuals (AIC/MDL criterion) – preprocessing method 10

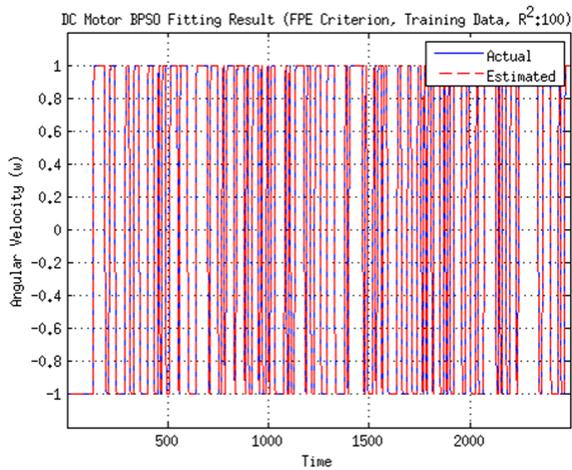Figure 21. BPSO DCM NARX residual plot (FPE criterion) – preprocessing method 10

Figure 19. BPSO DCM NARX OSA model fit (training set) (FPE criterion) – preprocessing method 10
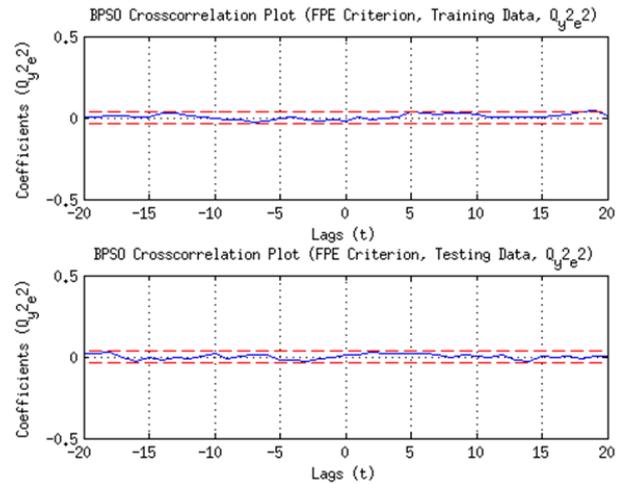

Figure 22. BPSO DCM NARX correlation test result (1/5) (FPE criterion) – preprocessing method 10
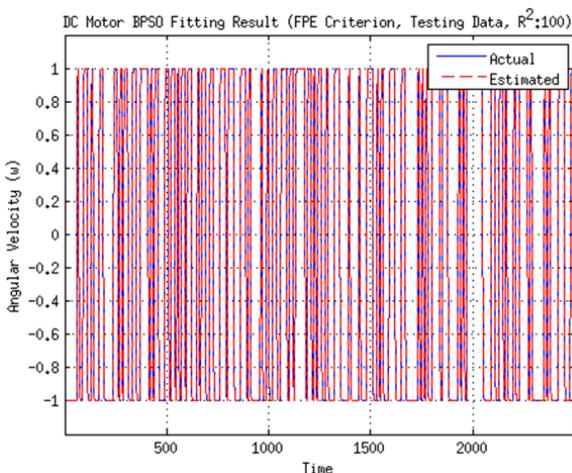

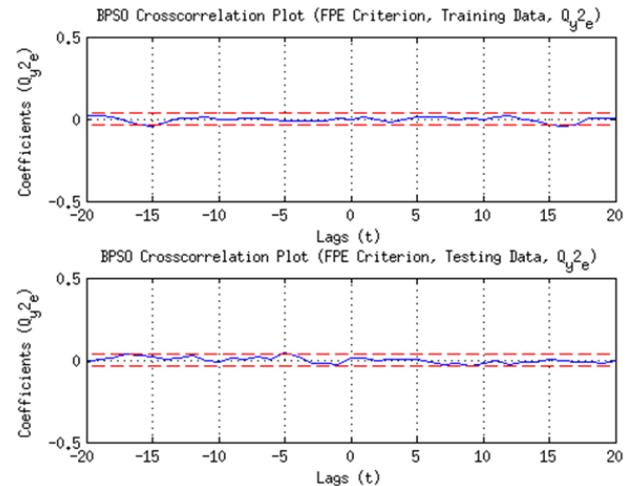Figure 20. BPSO DCM NARX OSA model fit (testing set) (FPE criterion) – preprocessing method 10

Figure 23. BPSO DCM NARX correlation test result (2/5) (FPE criterion) – preprocessing method 10
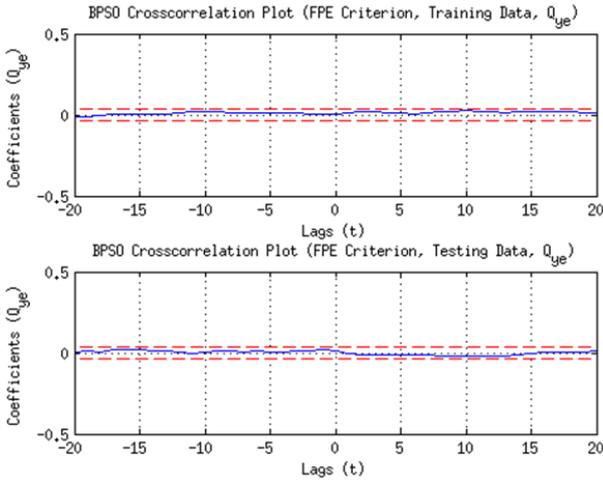
Figure 24. BPSO DCM NARX correlation test result (3/5) (FPE criterion) – preprocessing method 10

Figure 25. BPSO DCM NARX correlation test result (4/5) (FPE criterion) – preprocessing method 10

Figure 26. BPSO DCM NARX correlation test result (5/5) (FPE criterion) – preprocessing method 10
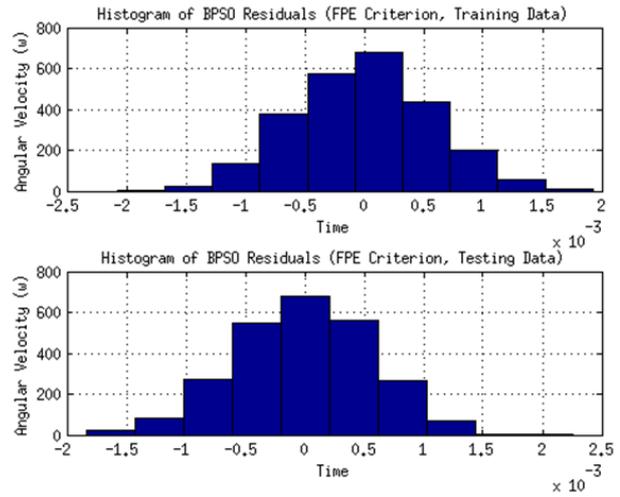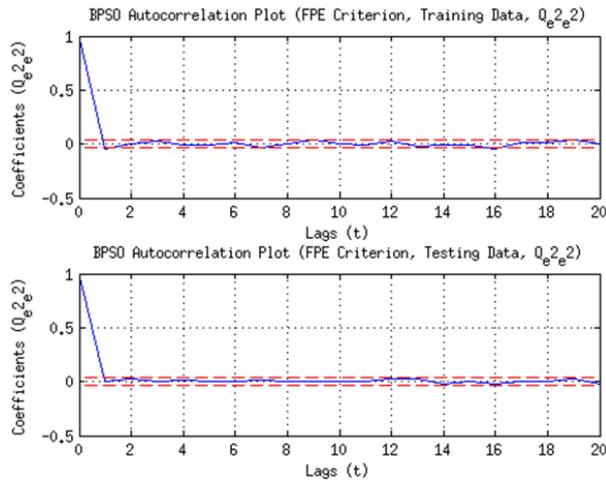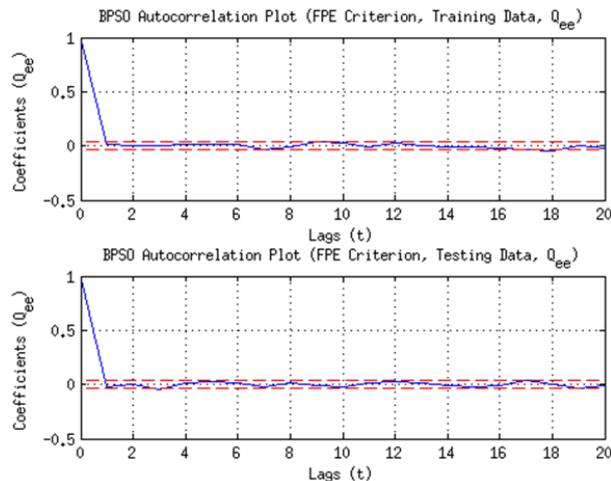

Figure 27. BPSO DCM NARX histogram of residuals (FPE criterion) – preprocessing method 10

Both observations are indicative of a good model fit, and that the model was representative of the system in question.

The residuals plot for the training and testing data is shown in Figure 12 and Figure 21. For a model to be accepted as a valid representation of the original system, the residuals need to exhibit characteristics similar to white noise (small and uncorrelated residuals). The magnitude of the residuals was measured by observing the peak-to-peak magnitude and Mean Squared Error (MSE), while the uncorrelatedness was measured by performing autocorrelation, cross-correlation and histogram tests on the residuals.

As can be seen from Figure 12 and Figure 21, the peak-to-peak magnitude generally ranged from 1.5 x 10^{-3} and -1.5 x 10^{-3}, while the MSE was very small. These observations fulfill a part of the validity requirement – namely the small magnitude of residuals.

A set of autocorrelation and cross-correlation tests by (Norgaard *et al.*, 2000b) was used to test the randomness of the residuals:

$$\theta_{\varepsilon\varepsilon}(\tau) = E\left[\varepsilon(t-\tau)\varepsilon(t)\right] = \delta(\tau) \tag{10}$$

$$\theta_{\varepsilon^2\varepsilon^2}(\tau) = E\left[\varepsilon^2(t-\tau)\varepsilon^2(t)\right] = \delta(\tau) \tag{11}$$

$$\theta_{y\varepsilon}(\tau) = E\left[y(t-\tau)\varepsilon(t)\right] = 0, \ \forall \tau \tag{12}$$

$$\theta_{y^2\varepsilon}(\tau) = E\left[(y^2(t-\tau)-\bar{y}^2(\tau))\varepsilon(t)\right] = 0, \ \forall \tau \tag{13}$$

$$\theta_{y^2\varepsilon^2}(\tau) = E\left[(y^2(t-\tau)-\bar{y}^2(\tau))\varepsilon^2(t)\right] = 0, \ \forall \tau \tag{14}$$

where:

$\theta_{x_1x_2}(\tau)$ = correlation coefficient between signals x_1 and x_2.

E[■]   = mathematical expectation of the correlation function.

$\varepsilon(t)$ = model residuals = $y(t) - \hat{y}(t)$.

τ = lag space.

$y(t)$ = observed output at time t.

$\delta(\tau)$ = Kronecker delta, defined as:

$$\delta(\tau) = \begin{cases} 1, & \tau = 0 \\ 0, & \tau \neq 0 \end{cases} \qquad (15)$$

The model is usually accepted if the correlation co-efficients lie within the 95% confidence limits, defined as $\pm 1.96/n$, with n is the number of data points in the sequence. The correlation test results are shown in Figure 13 to Figure 17. As can be seen from the figures, the residuals have passed all of the tests performed. This is because the correlation co-efficients were all within the boundaries defined by Eq. (10) to Eq. (15). Additionally, histogram tests performed on the data (Figure 18) showed a Gaussian curve, describing a random distribution. Based on the observations, it was concluded that the residuals were random and uncorrelated. Therefore, all models generated using BPSO were considered valid and accurate representations of the system.

5.4 Effect of BPSO parameter adjustment

As mentioned previously, convergence of the BPSO algorithm is affected by three parameters: swarm size, maximum iterations and initial random seed. The effect of the parameters on convergence is shown in Figure 28 to Figure 31. Small fitness variations were observed for swarm sizes between 10 and 50, with swarm size 30 showing slightly better fitness values. The maximum iterations parameter did not have a significant effect on the fitness, as all the training runs converged in less than 5 iterations. The initial random seeds caused the average fitness values to be varied. This is expected since different initialization seeds contribute differently to the search process of the solution space.

5.5 Comparison with OLS

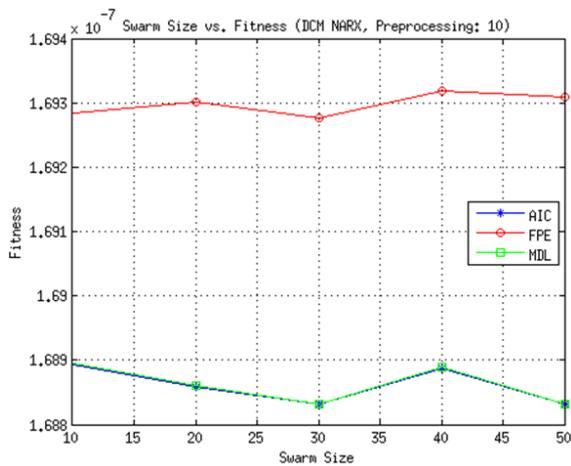A comparison of results between OLS and BPSO is

Figure 28. Swarm size vs. average fitness on DCM NARX – preprocessing method 10
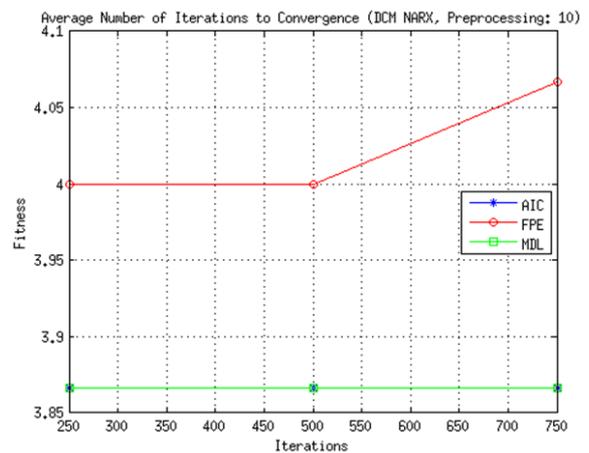

Figure 30. Average number of iterations to convergence on DCM NARX – preprocessing method 10
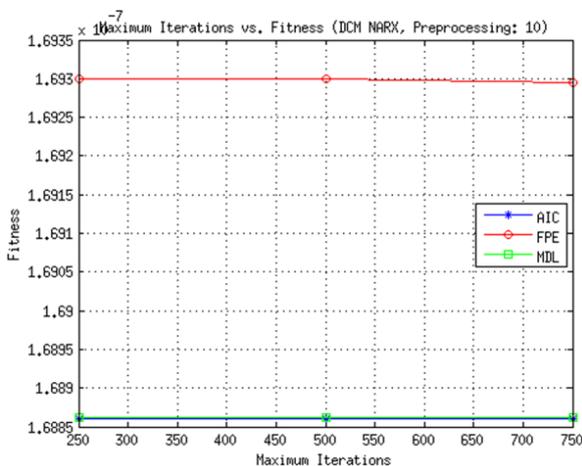

Figure 29. Maximum iterations vs. average fitness on DCM NARX – preprocessing method 10
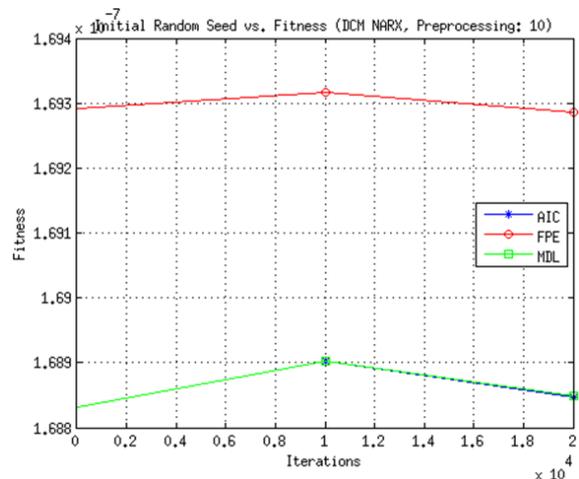

Figure 31. Initial random seed vs. average fitness on DCM NARX – preprocessing method 10

Table 8.   Comparison between OLS and BPSO on DCM NARX model (Preprocessing: 10)

Criteria	OLS		BPSO (AIC/MDL)	
	Training	Testing	Training	Testing
Terms selected	3		4	
AIC	1.6967e-07	1.6471e-07	1.6883e-07	1.6502e-07
FPE	1.6996e-07	1.6499e-07	1.6930e-07	1.6548e-07
MDL	1.6967e-07	1.6471e-07	1.6883e-07	1.6502e-07
R-squared (%)	100.00%	100.00%	100.00%	100.00%
CRV	7	2	5	1
MSE	3.3853e-07	3.2863e-07	3.3632e-07	3.2872e-07

shown in Table 8. The inclusion of the second-order term $u(t-2)*y(t-2)$ by BPSO had a positive effect on the number of correlation violations (CRV) with small differences in criterion and MSE scores. However, the inclusion of this term caused the fitting results of BPSO's testing set to be slightly higher than OLS's testing set. This indicates that some over-fitting has occurred when the additional term was included in the model.

6. Conclusions and Future Works

A structure selection method for the NARX model is presented. The focus of the paper was to analyze the implementation of the MySQL lookup table, as well as to compare between different preprocessing approaches. The database performance experiments confirm the execution flow and the performance of the table. Additional structure selection frequency analysis was possible as a result of the lookup table implementation. They reveal that BPSO had the tendency to select best-performing terms more frequently as a result of exploitation of the problem space.

Preprocessing method 10 was found to be the most suitable based on tests conducted on the four preprocessing methods. The magnitude scaling approach helped improve the MSE scores as the scaling approach reduced the MSE value. Both OLS and BPSO were capable of approximating the model well. OLS selected three single-order terms for preprocessing method 10, while the best BPSO identification result selected an additional second-order term. The additional term helped improve the model fit and reduce the number of CRVs relative to OLS. However, there appears to be some over-fitting of the testing dataset.

This research is extendable in several directions. Based on the results of this paper, the BPSO algorithm can also be applied to NARMA and NARMAX by employing a two-stage identification procedure for the additional residual terms. Furthermore, we are looking for ways to improve calculation speeds. A possible avenue for this would be high-performance distributed clusters to improve the search capability of BPSO.

References

Aguirre, L. A., and Furtado, E. C. 2007. Building dynamical models from data and prior knowledge: The case of the first period-doubling bifurcation. Physical Review. 76, 046219-046211 - 046219-046213.

Aguirre, L. A., and Letellier, C. 2009. Modeling Nonlinear Dynamics and Chaos: A Review. Mathematical Problems in Engineering. 2009, 1-35.

Ahn, K. K., and Anh, H. P. H. 2010. Inverse Double NARX Fuzzy Modeling for System Identification. IEEE/ ASME Transactions on Mechatronics. 15(1), 136-148.

Al-Qassar, A. A. and Othman, M. Z. 2008. Experimental Determination of Electrical and Mechanical Parameters of DC Motor using Genetic Elman Neural Network. Journal of Engineering Science and Technology. 3(2), 190-196.

Amisigo, B. A., Giesen, N. v. d., Rogers, C., Andah, W. E. I., and Friesen, J. 2007. Monthly streamflow prediction in the Volta Basin of West Africa: A SISO NARMAX polynomial modelling. Physics and Chemistry of the Earth. 33(1-2), 141-150.

Amrit, R., and Saha, P. 2007. Stochastic identification of bioreactor process exhibiting input multiplicity. Bioprocess and Biosystems Engineering. 30, 165-172.

Anderson, S. R., Lepora, N. F., Porrill, J., and Dean, P. 2010. Nonlinear Dynamic Modeling of Isometric Force Production in Primate Eye Muscle. IEEE Transactions Biomedical Engineering. 57(7), 1554-1567.

Bai, E.-W. and Deistler, M. 2010. An Interactive Term Approach to Non-Parametric FIR Nonlinear System Identification. IEEE Transactions Automatic Control. 55(8), 1952-1957.

Balikhin, M. A., Boynton, R. J., Walker, S. N., Borovsky, J. E., Billings, S. A., and Wei, H. L. 2011. Using the NARMAX approach to model the evolution of energetic electrons fluxes at geostationary orbit. Geophysical Research Letters. 38(L18105), 1-5.

Barbosa, B. H. G., Aguirre, L. A., Martinez, C. B., and Braga, A. P. 2011. Black and Gray-Box Identification of a Hydraulic Pumping System. IEEE Transactions Control Systems Technology. 19(2), 398-406.

Beale, M. H., Hagan, M. T., and Demuth, H. B. 2011. Neural Network Toolbox User's Guide. Natick, MA: Mathworks Inc.

Billings, S. A. and Aguirre, L. A. 1995. Effects of the Sampling Time on the Dynamics and Identification of Nonlinear Models. International Journal of Bifurcation and Chaos. 5(6), 1541-1556.

Billings, S. A. and Chen, S. 1989. Extended model set, global data and threshold model identification of severely non-linear systems. International Journal of Control. 50(5), 1897-1923.

Billings, S. A. and Coca, D. 2001. Identification of NARMAX and Related Models. In H. Unbehauen (Ed.), Control, Systems, Robotics and Automation (Vol. 1): EOLSS Publishers Co Ltd.

Billings, S. A., Korenberg, M., and Chen, S. 1988. Identification of nonlinear output-affine systems using an orthogonal least-squares algorithm. Journal of System Science. 19(8), 1559-1568.

Billings, S. A., and Wei, H.-L. 2005. A New Class of Wavelet Networks for Nonlinear System Identication. IEEE Transactions Neural Networks. 16(4), 862-874.

Cassar, T., Camilleri, K. P., and Fabri, S. G. 2010. Order Estimation of Multivariate ARMA Models. IEEE Journal of Selected Topics in Signal Processing. 4(3), 494-503.

Chen, B., Zhu, Y., Hu, J., & Prýncipe, J. C. 2011. Stochastic gradient identification of Wiener system with maximum mutual information criterion. IET Signal Processsing. 5(6), 589- 597.

Chen, S., Billings, S. A., and Luo, W. 1989. Orthogonal least squares methods and their application to non-linear system identification. International Journal of Control. 50(5), 1873-1896.

Chen, Z. H. and Ni, Y. Q. 2011. On-board Identification and Control Performance Verification of an MR Damper Incorporated with Structure. Journal of Intelligent Material Systems and Structures. 22, 1551-1565.

Cheng, Y., Wang, L., Yu, M., and Hu, J. 2011. An efficient identification scheme for a nonlinear polynomial NARX model. Artificial Life Robotics. 16, 70-73.

Chiras, N., Evans, C., and Rees, D. 2001. Nonlinear gas turbine modeling using NARMAX structures. IEEE Trans. Instrumentation and Measurement. 50(4), 893-898.

Cong, S., Li, G., and Feng, X. 2010. Parameters Identification of Nonlinear DC Motor Model using Compound Evolution Algorithms. Paper presented at the Proceeding World Congress on Engineering, London, UK.

Cosenza, B. 2012. Development of a neural network for glucose concentration prevision in patients affected by type 1 diabetes. Bioprocess and Biosystems Engineering. 1-9.

Dahunsi, O. A., Pedro, J. O., and Nyandoro, O. T. 2010. System Identification and Neural Network Based PID Control of Servo-Hydraulic Vehicle Suspension System. SAIEE Africa Research Journal. 101(3), 93-105.

Dao, T.-K. and Chen, C.-K. 2011. Path Tracking Control of a Motorcycle Based on System Identification. IEEE Trans. Vehicular Technology. 60(7), 2927-2935.

Dimogianopoulos, D. G., Hios, J. D., and Fassois, S. D. 2009. FDI for Aircraft Systems Using Stochastic Pooled-NARMAX Representations: Design and Assessment. IEEE Trans. Control Systems Technology. 17(6), 1385-1397.

Ding, F., Liu, P. X., and Liu, G. 2010. Multiinnovation Least-Squares Identification for System Modeling. IEEE Trans. Systems, Man and Cybernetics - Part B: Cybernetics. 40(3), 767-778.

El-Nagar, M. M. K., Ibrahim, D. K., and Youssef, H. K. M. 2011. Optimal Harmonic Meter Placement Using Particle Swarm Optimization Technique. Online Journal of Power and Energy Engineering. 2(2), 204-209.

Er, M. J., Liu, F., and Li, M. B. 2010. Self-constructing Fuzzy Neural Networks with Extended Kalman Filter. International Journal of Fuzzy Systems. 12(1), 66-72.

Evsukoff, A. G., Lima, B. S. L. P. d., and Ebecken, N. F. F. 2011. Long-Term Runoff Modeling Using Rainfall Forecasts with Application to the Iguaçu River Basin. Water Resource Management. 25, 963-985.

Fei, J. and Xin, M. 2012. An adaptive fuzzy sliding mode controller for MEMS triaxial gyroscope with angular velocity estimation. Nonlinear Dynamics. 1-13.

Gandomi, A. H., Alavi, A. H., and Arjmandi, P. 2010. Genetic Programming and Orthogonal Least Squares: A Hybrid Approach to Compressive Strength of CFRP-confined Concrete Cylinders Modeling. Journal of Mechanics of Materials and Structures. 5(5), 735–753.

Gonzales-Olvera, M. A., and Tang, Y. 2010. Black-Box Identification of a Class of Nonlinear Systems by a Recurrent Neurofuzzy Network. IEEE Transaction on Neural Networks. 21(4), 672-679.

Hahn, J.-O., McCombie, D. B., Reisner, A. T., Hojman, H. M., and Asada, H. H. 2010. Identification of Multichannel Cardiovascular Dynamics Using Dual Laguerre Basis Functions for Noninvasive Cardiovascular Monitoring. IEEE Transaction on Automatic Control Technology. 18(1), 170-176.

Harris, T. J. and Yu, W. 2012. Variance decompositions of nonlinear time series using stochastic simulation and sensitivity analysis. Statistical Computing. 22, 387-396.

Hong, X., Chen, S., and Harris, C. J. 2011. Construction of Radial Basis Function Networks with Diversified Topologies. In L. Wang and H. Garnier, editors. System identification, environmental modelling, and control system design, Springer, London, pp 251-270.

Hong, X., Mitchell, R. J., Chen, S., Harris, C. J., K.Li, and Irwin, G. W. 2008. Model selection approaches for non-linear system identification: a review. International Journal of Systems Science. 39(10), 925-946.

Juang, C.-F. and Hsieh, C.-D. 2010. A Locally Recurrent Fuzzy Neural Network With Support Vector Regression for Dynamic-System Modeling. IEEE Transaction on Fuzzy Systems. 18(2), 261-273.

Kennedy, J. and Eberhart, R. C. 1997. A discrete binary version of the particle swarm algorithm. Paper presented at the Proceeding of the 1997 Conference on Systems, Man, Cybernetics, Piscataway, New Jersey, U.S.A.

Kibangou, A. Y., and Favier, G. 2010. Identification of fifth-order Volterra systems using i.i.d. inputs. IET Signal Processing. 4(1), 30-44.

Kolodziej, J. R., and Mook, D. J. 2011. Model determination for nonlinear state-based system identification. Nonlinear Dynamics. 2011, 735-753.

Kukreja, S. L., Galiana, H. L., and Kearney, R. E. 2003. NARMAX representation and identification of ankle dynamics. IEEE Transaction on Biomedical Engineering. 50(1), 70-81.

Kukreja, S. L., Kearney, R. E., and Galiana, H. L. 2005. A least squares parameter estimation algorithm for switched hammerstein systems with applications to the VOR. IEEE Transaction on Biomedical Enginereing. 52(3), 431-444.

Leontaritis, I. J., and Billings, S. A. 1985. Input-output parametric models for nonlinear systems. International Journal of Control. 41(2), 311-341.

Liao, C.-C. 2010. Enhanced RBF Network for Recognizing Noise-Riding Power Quality Events. IEEE Transaction on Instrumentation and Measurement. 59(6), 1550-1561.

Luiz, S. O. D., Perkusich, A., Lima, A. M. N., Silva, J. J., and Almeida, H. 2012. System Identification and Energy-aware Processor Utilization Control. IEEE Transaction on Consumer Electronics. 58(1), 32-37.

Manivannan, P. V., Singaperumal, M., and Ramesh, A. 2011. Development of an Idle Speed Engine Model using In-Cylinder Pressure Data and an Idle Speed Controller for a Small Capacity Port Fuel Injected SI Engine. International Journal of Automotive Technology. 12 (1), 11-20.

Matus, M., Sáez, D., Favley, M., Suazo-Martínez, C., Moya, J., Jiménez-Estévez, G., and Jorquera, P. 2012. Identification of Critical Spans for Monitoring Systems in Dynamic Thermal Rating. IEEE Transaction on Power Delivery. 27(2), 1002-1009.

Mendes, E., and Billings, S. A. 2001. An Alternative Solution to the Model Structure Selection Problem. IEEE Transaction on Systems, Man and Cybernetics - Part A: Systems and Humans. 31(6), 597-608.

Nasri, M., Nezambadi-pour, H., and Maghfoori, M. 2007. A PSO-based Optimum Design of PID Controller for a Linear Brushless DC Motor. Paper presented at the Proceeding on World Academy of Science, Engineering and Technology.

Ning, H., Jing, X., and Cheng, L. 2011. Online Identification of Nonlinear Spatiotemporal Systems Using Kernel Learning Approach. IEEE Transaction on Neural Networks. 22(9), 1381-1394.

Norgaard, M., Ravn, O., Poulsen, N. K., and Hansen, L. K. 2000a. Neural networks for modeling and control of dynamic systems: A practitioner's handbook. Springer, London, U.S.A.

Norgaard, M., Ravn, O., Poulsen, N. K., and Hansen, L. K. 2000b. Neural Networks for modelling and control of dynamic systems. Springer-Verlag London Ltd., U.S.A.

Oh, S.-R., Sun, J., Li, Z., Celkis, E. A., and Parsons, D. 2010. System Identification of a Model Ship Using a Mechatronic System. IEEE/ASME Transaction on Mechatronics. 15(2), 316-320.

Pierre, J. W., Zhou, N., Tuffner, F. K., Hauer, J. F., Trudnowski, D. J., and Mittelstadt, W. A. 2010. Probing Signal Design for Power System Identication. IEEE Transaction on Power Systems. 25(2), 835-843.

Piroddi, L. 2008. Simulation error minimisation methods for NARX model identification. International Journal of Modelling, Identification and Control. 3(4), 392-403.

Piroddi, L. (2010). Adaptive model selection for polynomial NARX models IET Control Theory and Applications. 4(12), 2693-2706.

Piroddi, L., and Lovera, M. 2008. NARX model identification with error filtering. Paper presented at the Proceeding 17th World Congress International Federation of Automatic Control, Seoul, Korea.

Rahim, N. A. 2004. The design of a non-linear autoregressive moving average with exegenous input (NARMAX) for a DC motor. (MSc), Universiti Teknologi MARA, Shah Alam, Malaysis.

Rahim, N. A., Taib, M. N., and Yusof, M. I. 2003. Nonlinear system identification for a DC motor using NARMAX Approach. Paper presented at the Asian Conference on Sensors (AsiaSense).

Rahiman, M. H. F. 2008. System Identification of Essential Oil Extraction System. (Ph. D. Dissertation), Universiti Teknologi MARA, Shah Alam, Malaysia.

Shafiq, M. and Butt, N. R. 2011. Utilizing Higher-Order Neural Networks in U-model Based Controllers for Stable Nonlinear Plants. International Journal of Control, Automation and Systems. 9(3), 489-496.

Simani, S. 2002. Model-base Fault Diagnosis in Dynamic Systems using Identification Techniques: Springer, U.S.A

Talmon, R., Kushnir, D., Coifman, R. R., Cohen, I., and Gannot, S. 2012. Parametrization of Linear Systems Using Diffusion Kernels. IEEE Transaction Signal Processing. 60(3).

Tan, A. H. and Cham, C. L. 2011. Continuous-time model identification of a cooling system with variable delay. IET Control Theory Applications. 5(7), 913–922.

Tesli, L., Hartmann, B., Nelles, O., and Škrjanc, I. 2011. Nonlinear System Identification by Gustafson–Kessel Fuzzy Clustering and Supervised Local Model Network Learning for the Drug Absorption Spectra Process. IEEE Transaction on Neural Networks. 22(12), 1941-1951.

Tjahjowidodo, T., Al-Bender, F., and Brussel, H. V. 2005. Friction identification and compensation in a DC motor. Paper presented at the Proceeding 16th IFAC World Congress, Prague, Czech Republic.

Vallverdu, M., Korenberg, M. J., and Caminal, P. 1991. Model identification of the neural control of the cardiovascular system using NARMAX models. Paper presented at the Proceeding of Computer in Cardiology.

Wang, H., Shen, Y., Huang, T., and Zeng, Z. 2009. Nonlinear System Identification based on Recurrent Wavelet Neural Network Advances in Soft Computing (Vol. 56): Springer, U.S.A.

Wei, H. L. and Billings, S. A. 2008a. An adaptive orthogonal search algorithm for model subset selection and nonlinear system identification. International Journal of Control. 81(5), 714-724.

Wei, H. L. and Billings, S. A. 2008b. Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information. International Journal of Modeling, Identification and Control, 3(4), 341-356.

Westwick, D. T. and Perreault, E. J. 2011. Closed-Loop Identification: Application to the Estimation of Limb Impedance in a Compliant Environment. IEEE Transaction Biomedical Engineering. 58(3), 521-530.

Wong, Y. W., Seng, K. P., and Ang, L.-M. 2011. Radial Basis Function Neural Network With Incremental Learning for Face Recognition. IEEE Transaction on Systems, Man and Cybernetics - Part B:Cybernetics. 41(4), 940-949.

Yassin, I. M., Taib, M. N., Rahim, N. A., Salleh, M. K. M., and Abidin, H. Z. 2010. Particle Swarm Optimization for NARX structure selection—Application on DC motor model. Paper presented at the IEEE Symposium on Industrial Electronics and Applications 2010 (ISIEA 2010), Penang, Malaysia.

YuMin, L., DaZhen, T., Hao, X., and Shu, T. 2011. Productivity matching and quantitative prediction of coalbed methane wells based on BP neural network. Science China Technological Sciences. 54(5), 1281-1286.

Yunna, W. and Zhaomin, S. 2008. Application of Radial Basis Function Neural Network Based on Ant Colony Algorithm in Credit Evaluation of Real Estate Enterprises. Paper presented at the Proceeding of 2008 International Conference on Management Science and Engineering.

Zhao, W.-X., Chen, H.-F., and Zheng, W. X. 2010. Recursive Identification for Nonlinear ARX Systems Based on Stochastic Approximation Algorithm. IEEE Transaction Automatic Control. 55(6), 1287-1299.

Zhao, Y., Westwick, D. T., and Kearney, R. E. 2011. Subspace Methods for Identification of Human Ankle Joint Stiffness. IEEE Transaction Biomedical Engineering. 58(11), 3039-3048.

Zhu, D., Guo, J., Cho, C., Wang, Y., and Lee, K.-M. 2012. Wireless Mobile Sensor Network for the System Identification of a Space Frame Bridge. IEEE/ASME Transaction on Mechatronics. 17(3), 499-507.