Extended Analysis of BPSO Structure Selection of Nonlinear Auto-Regressive Model with Exogenous Inputs (NARX) of Direct Current Motor

Ihsan M. Yassin, Mohd Nasir Taib and Ramli Adnan

Faculty of Electrical Engineering, Universiti Teknologi Mara,
Shah Alam, MALAYSIA

Email: ihsan.yassin@gmail.com

1 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 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 Exogeneous Inputs model (NARX), structure selection, Binary Particle Swarm Optimization algorithm, Direct Current Motor.

2 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, Pedro, & Nyandoro, 2010; Ding, Liu, & Liu, 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 prediction of the system’s future behavior under given inputs and/or outputs (Hong, Chen, & Harris, 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, Lima, & Ebecken, 2011; Gonzales-Olvera & Tang, 2010; Hahn, McCombie, Reisner, Hojman, & Asada, 2010; Matus et al., 2012; Y. Zhao, Westwick, & Kearney, 2011), model-based design of control engineering applications (Dao & Chen, 2011; Fei & Xin, 2012; Gonzales-Olvera & Tang, 2010; Luiz, Perkusich, Lima, Silva, & Almeida, 2012; Manivannan, Singaperumal, & Ramesh, 2011; Oh, Sun, Li, Celkis, & Parsons, 2010), and project monitoring and planning (Evsukoff et al., 2011; Matus et al., 2012; Pierre et al., 2010; YuMin, DaZhen, Hao, & Shu, 2011; Zhu, Guo, Cho, Wang, & Lee, 2012).

Simultaneous structure selection and parameter estimation of NARMAX and derivative models are achievable through the Orthogonal Least Squares (OLS) algorithm (S. A. Billings, Korenberg, & Chen, 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 (L. A. Aguirre & Furtado, 2007; Barbosa et al., 2011; Dimogianopoulos, Hios, & Fassois, 2009; Er, Liu, & Li, 2010; Liao, 2010; Wei & Billings, 2008a; Wong, Seng, & Ang, 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 (L. A. Aguirre & Furtado, 2007; Cheng et al., 2011; Piroddi, 2008; Wei & Billings, 2008a, 2008b). In 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, Taib, Rahim, Salleh, & Abidin, 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 are 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.

3 Review of Structure Selection Methods

Structure selection is defined as the task of selecting a subset of regressors to represent system dynamics (Luis A. 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, Camilleri, & Fabri, 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 criterions to select the most important regressor (Luis A. Aguirre & Letellier, 2009). A review of several structure selection methods are deliberated in the following sections.

3.1 Trial and Error Methods

Two trial and error methods for structure selection were found in literature. The first method, called Zero-and-Refitting (Luis A. 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 (Luis A. Aguirre & Letellier, 2009).

Another method was found in (Z. H. Chen & Ni, 2011) for structure selection of a Bayesian NARX 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.

### 3.2 OLS

The OLS algorithm is a simultaneous structure selection / parameter estimation algorithm introduced by (S. A. 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 (S. Chen, Billings, & Luo, 1989; Cheng et al., 2011; Ning, Jing, & Cheng, 2011). This step decouples the regressors so that their individual contributions are estimated separately from the rest (Balikhin et al., 2011; Kibangou & Favier, 2010; Eduardo M. A. M. Mendes & Steve A. 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 combinations that has the most influence over the variance of the system output (Balikhin et al., 2011; Gandomi, Alavi, & Arjmandi, 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 are 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 (L. A. Aguirre & Furtado, 2007; Barbosa et al., 2011; B. Chen, Zhu, Hu, & Principe, 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 (L. A. 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 (Luis A. Aguirre & Letellier, 2009).

Despite OLS’s effectiveness, several criticisms have been directed towards its tendency to select excessive or sub-optimal terms (L. A. 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 is 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 (Eduardo M. A. M. Mendes & Steve A. Billings, 2001), regularized OLS (Wong et al., 2011), incorporation of information criterions, 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)).

3.3 Clustering Methods

Clustering methods group data based on their similarities and performs 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, Hartmann, Nelles, & Škrjanc, 2011).

In (Luis A. 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, which were then modeled using local linear models.
Clustering using Independent Component Analysis (ICA) has also been reported (Talmon, Kushnir, Coifman, Cohen, & Gannot, 2012).

As reported by (Luis A. Aguirre & Letellier, 2009; 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.

3.4 Correlation-Based Methods

Correlation-based methods have been used in (Cheng et al., 2011; Tan & Cham, 2011; 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.

3.5 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.

4 Theoretical Background

4.1 NARX

The NARX model is a generalized version of the Auto-Regressive with Exogenous Inputs (ARX) model (W.-X. Zhao, Chen, & Zheng, 2010). The NARX model takes the form
of (Amisigo, Giesen, Rogers, Andah, & Friesen, 2007):

\[ y(t) = f^d[y(t-1), y(t-2), \ldots, y(t-n_y), u(t-n_k), u(t-n_k-1), \ldots, u(t-n_k-n_u)] + \varepsilon(t) \quad (1) \]

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, Lepora, Porrill, & Dean, 2010; S. A. Billings & Chen, 1989; S. Chen et al., 1989), Multilayer Perceptrons (MLP) (Norgaard, Ravn, Poulsen, & Hansen, 2000; Rahim, 2004; Rahim et al., 2003; Rahiman, 2008), and Wavelet ANNs (WNN) (Stephen A. Billings & Wei, 2005; Wang, Shen, Huang, & Zeng, 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 either be continuous or discrete in nature (S. A. Billings & Coca, 2001). For continuous data, it needs to be sampled at regular and fixed intervals in order to discretize it for identification (S. A. Billings & Aguirre, 1995; S. A. 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) (Z. H. Chen & Ni, 2011). Finally, the model is validated using One Step Ahead (OSA) and correlation tests to ensure that it is valid and acceptable.

4.2 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):

\[ P\theta + \varepsilon = y \]  

(2)

where \( P \) is a \( n \times m \) regressor matrix, \( \theta \) 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. \( \varepsilon \) 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 ($\theta_R$) 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):

$$P_R \theta_R + \varepsilon = y$$  

(3)

$$P_R = Q_R R_R$$  

(4)

$$g_R = Q_R^T y$$  

(5)

$$R_R \theta_R = g_R$$  

(6)

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

$$\theta_R = R_R^T g_R$$  

(7)

5  Methodology

5.1  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, Al-Bender, & Brussel, 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, Li, & Feng, 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, Nezambadi-pour, & Maghfoori, 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) 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; Z. H. Chen & Ni, 2011; 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 (Z. H. Chen & Ni, 2011).

It is preferable to separate between 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, Hagan, & Demuth, 2011; Cosenza, 2012; Er et al., 2010; Gandomi et al., 2010; Rahiman, 2008). Block division divides the dataset in 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 $2^{14} = 16,384$.

### 5.2 NARX Structure Selection using BPSO

After the regressors matrices have 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 parameters swarm size, maximum iterations and several random seeds.

The BPSO parameter values are shown in Table 3. The choice of swarm size and maximum iterations were based on preliminary tests to balance between 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 $x_{\text{min}}$ and $x_{\text{max}}$ were set to 0 and 1, respectively. They are within the range of probability values for bit change to occur. $v_{\text{min}}$ and $v_{\text{max}}$
represent the movement range of the particles. Since the value of $x_{id}$ is between the range of 0 and 1 (based on $x_{\text{min}}$ and $x_{\text{max}}$), the values of $v_{\text{min}}$ and $v_{\text{max}}$ 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, Ibrahim, & Youssef, 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 performed to discover regressor selection patterns of the BPSO algorithm.

### 5.3 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 as a lookup table to speed up computations, as well as to verify the execution flow of the lookup table.
6 Results and Discussions

6.1 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 amount (84.72%) of the solutions were 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).

6.2 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 were also observed with small mean. The high maximum values are explained by the initially un-optimized during the initialization phase. The average and fitness values were lower, indicating progressively better solutions were found as optimization progressed. Magnitude scaling appears to have a positive effect, while interleaving has a negative effect on 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 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.

6.3 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)):

\[
y(t) = 0.4955u(t - 1) + 0.0014u(t - 2) + 0.4794y(t - 1) \\
+ 5.8281e^{-3}y(t - 2) + 4.5334e^{-5}u(t - 2) * y(t - 2) + \epsilon(t)
\]  

(8)

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

\[
y(t) = 0.4955u(t - 1) + 0.0014u(t - 2) + 0.4794y(t - 1) \\
+ 5.8281e^{-3}y(t - 2) + 4.5334e^{-5}u(t - 2) * y(t - 2) + \epsilon(t)
\]  

(9)
\[ y(t) = 0.4995u(t-1) + 0.0083u(t-2) + 0.4911y(t-1) + 4.5329e^{-5}u(t-2) \cdot y(t-2) + \varepsilon(t) \]  \hspace{1cm} (9)

All models constructed using BPSO selected a second-order term \( u(t-2) \cdot y(t-2) \). 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 are 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%. 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 $1.5 \times 10^{-3}$ and $-1.5 \times 10^{-3}$, 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, Ravn, Poulsen, & Hansen, 2000b) was used to test the randomness of the residuals:

\[
\theta_{\varepsilon\varepsilon}(\tau) = E[\varepsilon(t - \tau)\varepsilon(t)] = \delta(\tau) \quad (10)
\]

\[
\theta_{\varepsilon^2\varepsilon^2}(\tau) = E[\varepsilon^2(t - \tau)\varepsilon^2(t)] = \delta(\tau) \quad (11)
\]

\[
\theta_{y\varepsilon}(\tau) = E[y(t - \tau)\varepsilon(t)] = 0, \forall \tau \quad (12)
\]

\[
\theta_{y^2\varepsilon}(\tau) = E[(y^2(t - \tau) - \tilde{y}^2(\tau))\varepsilon(t)] = 0, \forall \tau \quad (13)
\]

\[
\theta_{y^2\varepsilon^2}(\tau) = E[(y^2(t - \tau) - \tilde{y}^2(\tau))\varepsilon^2(t)] = 0, \forall \tau \quad (14)
\]

where:

- $\theta_{x_1x_2}(\tau) =$ correlation coefficient between signals $x_1$ and $x_2$.
- $E[\square] =$ mathematical expectation of the correlation function.
- $\varepsilon(t) =$ model residuals $= y(t) - \tilde{y}(t)$.
- $\tau =$ lag space.
- $y(t) =$ observed output at time $t$.
- $\delta(\tau) =$ Kronecker delta, defined as:

\[
\delta(\tau) = \begin{cases} 
1, & \tau = 0 \\
0, & \tau \neq 0
\end{cases} \quad (15)
\]

The model is usually accepted if the correlation coefficients lie within the 95% confidence limits, defined as $\pm 1.96/n$, with $n$ is the number of data points in the sequence.
The correlations 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 coefficients 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.

6.4 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 showed 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.

6.5 Comparison with OLS

A comparison of results between OLS and BPSO is shown in Table 8. The inclusion of the second-order term $u(t - 2) \cdot y(t - 2)$ 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 the testing set to be slightly higher than the testing set. This indicates that some over-fitting has occurred when the additional term was included in the model.
7 Conclusions & 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.
8 References


**List of Figures**

*Figure 1: The DCM dataset*
Figure 2: ERD for MySQL lookup table

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
Figure 6: Histogram of term selection (DCM NARX – preprocessing method 00)

Figure 7: Histogram of term selection (DCM NARX – preprocessing method 01)
Figure 8: Histogram of term selection (DCM NARX – preprocessing method 10)

Figure 9: Histogram of term selection (DCM NARX – preprocessing method 11)
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
Figure 12: BPSO DCM NARX residual plot (AIC/MDL criterion) – preprocessing method 10

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

Figure 21: BPSO DCM NARX residual plot (FPE criterion) – preprocessing method 10
Figure 22: BPSO DCM NARX correlation test result (1/5) (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
Figure 28: Swarm size vs. average fitness on DCM NARX – preprocessing method 10

Figure 29: Maximum iterations vs. average fitness on DCM NARX – preprocessing method 10
Figure 30: Average number of iterations to convergence on DCM NARX – preprocessing method 10

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

Table 1: Preprocessing methods used and their respective codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Preprocessing Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>No magnitude scaling, block division</td>
</tr>
<tr>
<td>01</td>
<td>No magnitude scaling, interleaving</td>
</tr>
<tr>
<td>10</td>
<td>Magnitude scaling, block division</td>
</tr>
<tr>
<td>11</td>
<td>Magnitude scaling, interleaving</td>
</tr>
</tbody>
</table>

Table 2: Terms used in NARX DCM model

<table>
<thead>
<tr>
<th>No.</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$u(t - 1)$</td>
</tr>
<tr>
<td>2</td>
<td>$u(t - 2)$</td>
</tr>
<tr>
<td>3</td>
<td>$y(t - 1)$</td>
</tr>
<tr>
<td>4</td>
<td>$y(t - 2)$</td>
</tr>
<tr>
<td>5</td>
<td>$u(t - 1) \times u(t - 1)$</td>
</tr>
<tr>
<td>6</td>
<td>$u(t - 1) \times u(t - 2)$</td>
</tr>
<tr>
<td>7</td>
<td>$u(t - 1) \times y(t - 1)$</td>
</tr>
<tr>
<td>8</td>
<td>$u(t - 1) \times y(t - 2)$</td>
</tr>
<tr>
<td>9</td>
<td>$u(t - 2) \times u(t - 2)$</td>
</tr>
<tr>
<td>10</td>
<td>$u(t - 2) \times y(t - 1)$</td>
</tr>
<tr>
<td>11</td>
<td>$u(t - 2) \times y(t - 2)$</td>
</tr>
<tr>
<td>12</td>
<td>$y(t - 1) \times y(t - 1)$</td>
</tr>
<tr>
<td>13</td>
<td>$y(t - 1) \times y(t - 2)$</td>
</tr>
<tr>
<td>14</td>
<td>$y(t - 2) \times y(t - 2)$</td>
</tr>
</tbody>
</table>
**Table 3: BPSO parameter settings for structure selection experiments**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swarm Size</td>
<td>10, 20, 30, 40, 50</td>
</tr>
<tr>
<td>Fitness Criterion</td>
<td>Akaike Information Criterion (AIC) Final</td>
</tr>
<tr>
<td></td>
<td>Prediction Error (FPE)</td>
</tr>
<tr>
<td></td>
<td>Minimum Descriptor Length (MDL)</td>
</tr>
<tr>
<td>Maximum Iterations</td>
<td>250, 500, 750</td>
</tr>
<tr>
<td>Initial Seed</td>
<td>0, 10 000, 20 000</td>
</tr>
<tr>
<td>$x_{min}$</td>
<td>0</td>
</tr>
<tr>
<td>$x_{max}$</td>
<td>1</td>
</tr>
<tr>
<td>$v_{min}$</td>
<td>-1</td>
</tr>
<tr>
<td>$v_{max}$</td>
<td>+1</td>
</tr>
<tr>
<td>$C_1$</td>
<td>2.0</td>
</tr>
<tr>
<td>$C_2$</td>
<td>2.0</td>
</tr>
<tr>
<td>Field</td>
<td>Type</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>id</td>
<td>integer (max 255-bit), auto-generated, unique, primary key.</td>
</tr>
<tr>
<td>model</td>
<td>Variable-length string (max 1,000 characters)</td>
</tr>
<tr>
<td>criterion</td>
<td>Variable-length string (max 1,000 characters)</td>
</tr>
<tr>
<td>preprocessing</td>
<td>Variable-length string (max 1,000 characters)</td>
</tr>
<tr>
<td>nu</td>
<td>integer (max 255-bit)</td>
</tr>
<tr>
<td>ny</td>
<td>integer (max 255-bit)</td>
</tr>
<tr>
<td>ne</td>
<td>integer (max 255-bit)</td>
</tr>
<tr>
<td>nk</td>
<td>integer (max 255-bit)</td>
</tr>
<tr>
<td>binstring</td>
<td>string (max 5,000 characters)</td>
</tr>
<tr>
<td>fitness</td>
<td>Double value</td>
</tr>
<tr>
<td>datecreated</td>
<td>timestamp</td>
</tr>
</tbody>
</table>
Table 5: Parameters for database performance test

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>DCM</td>
</tr>
<tr>
<td>Criterion</td>
<td>AIC</td>
</tr>
<tr>
<td>Swarm Size</td>
<td>10</td>
</tr>
<tr>
<td>Maximum Iterations</td>
<td>250</td>
</tr>
<tr>
<td>Initial Random Seed</td>
<td>0</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Pre-processing</th>
<th>Criterion</th>
<th>No. Records (Percent of Solution Space)</th>
<th>Fitness Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Min</td>
</tr>
<tr>
<td>00</td>
<td>AIC</td>
<td>2,872 (17.53%)</td>
<td>3.6714e-04</td>
</tr>
<tr>
<td></td>
<td>FPE</td>
<td>2,553 (15.58%)</td>
<td>3.6811e-04</td>
</tr>
<tr>
<td></td>
<td>MDL</td>
<td>2,983 (18.21%)</td>
<td>3.6715e-04</td>
</tr>
<tr>
<td>01</td>
<td>AIC</td>
<td>4,465 (27.25%)</td>
<td>4.5055e-04</td>
</tr>
<tr>
<td></td>
<td>FPE</td>
<td>3,848 (23.49%)</td>
<td>4.5155e-04</td>
</tr>
<tr>
<td></td>
<td>MDL</td>
<td>4,442 (27.11%)</td>
<td>4.5055e-04</td>
</tr>
<tr>
<td>10</td>
<td>AIC</td>
<td>2,373 (14.48%)</td>
<td>1.6883e-07</td>
</tr>
<tr>
<td></td>
<td>FPE</td>
<td>2,485 (15.17%)</td>
<td>1.6928e-07</td>
</tr>
<tr>
<td></td>
<td>MDL</td>
<td>2,562 (15.64%)</td>
<td>1.6883e-07</td>
</tr>
<tr>
<td>11</td>
<td>AIC</td>
<td>5,516 (33.67%)</td>
<td>2.0724e-07</td>
</tr>
<tr>
<td></td>
<td>FPE</td>
<td>3,076 (18.77%)</td>
<td>2.0770e-07</td>
</tr>
<tr>
<td></td>
<td>MDL</td>
<td>5,441 (33.21%)</td>
<td>2.0724e-07</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Fitness Criterion</th>
<th>Eval. Criterion</th>
<th>Training Set</th>
<th>Testing Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC/MDL</td>
<td>Times Found</td>
<td>AIC: 33, MDL: 33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>1.6883e-07</td>
<td>1.6502e-07</td>
</tr>
<tr>
<td></td>
<td>FPE</td>
<td>1.6930e-07</td>
<td>1.6548e-07</td>
</tr>
<tr>
<td></td>
<td>MDL</td>
<td>1.6883e-07</td>
<td>1.6502e-07</td>
</tr>
<tr>
<td></td>
<td>R-squared (%)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>CRV</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>3.3632e-07</td>
<td>3.2872e-07</td>
</tr>
<tr>
<td>FPE</td>
<td>Times Found</td>
<td>FPE: 31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>1.6890e-07</td>
<td>1.6474e-07</td>
</tr>
<tr>
<td></td>
<td>FPE</td>
<td>1.6928e-07</td>
<td>1.6510e-07</td>
</tr>
<tr>
<td></td>
<td>MDL</td>
<td>1.6890e-07</td>
<td>1.6474e-07</td>
</tr>
<tr>
<td></td>
<td>R-squared (%)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>CRV</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>3.3672e-07</td>
<td>3.2842e-07</td>
</tr>
</tbody>
</table>
Table 8: Comparison between OLS and BPSO on DCM NARX model
(Preprocessing: 10)

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