VOL. 10, NO. 22, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

PARTICLE SWARM OPTIMIZATION AND LEAST SQUARES ESTIMATON OF NARMAX

S. M. Abdullah, A. I. M. Yassin and N. M. Tahir
Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
E-Mail: munirohabdullah@gmail.com

ABSTRACT

SI process consist of three steps; structure selection, parameter estimation, model validation. This paper compared method of Particle Swarm Optimization (PSO) and Linear Least Squares solution methods (LLS) (Normal Equation (NE), QR decomposition (QR) and Singular Value Decomposition (SVD)) for parameter estimation using polynomial NARMAX models. The comparison was tested on Flexible Robot Arm (FRA) dataset. Our analysis suggests that the PSO algorithm is comparable to other established algorithms for LLS parameter estimation in terms of model fit accuracy and information criteria (Akaike Information Criterion (AIC), Final Prediction Error (FPE) and Model Descriptor Length (MDL)). Additionally, the PSO algorithm was found to slightly improve the correlation tests relative to other LLS tested algorithms.

Keywords: System identification, linear least squares, particle swarm optimization, NARMAX.

INTRODUCTION

System Identification (SI) is a control engineering discipline that develops the mathematical model to represent the characteristics of physical model [1], [2]. The purpose of identification can be classified in several categorized such as to design and implement the model [2], [3], to analyse the properties of the model [4] and to plan or monitor the engineering projects [5]. Usually, SI is popular in process industries to obtain the better knowledge for improved the control system. However, the application is not limited on industries field only, but it has been used in science [6], marine nature and society [7], financial aspects [8] and numerous different fields.

The SI can be divided into two types, linear and nonlinear model however in real life, all systems are in nonlinear form [2]. Many models exist for modelling nonlinear system. Among them is the NARMAX model. NARMAX is a good for modelling nonlinear system because it is efficient, accurate, powerful, and unified representations [2], [10]–[12]. There are three important steps in NARMAX and its derivatives identification [1]. First is structure selection. It is performed to detect the hidden structure of a dataset. Then, followed by parameter estimation to enhance the target function where it is involving the difference between the identified model and the actual dataset. Finally, the model was validated by using model fit tests and correlation test to ensure the accuracy of model fit and to confirm the whiteness of residuals in the model developed.

This research focused on parameter estimation part and thus validate for NARMAX model. The structure selection part of this experiment was referred to the result based on [2]. There are many approaches used for parameter estimation stage such as least squares method [13]–[17], SVM [8][18], RBF [19], and BPSO [2], RPSO [12]. Among them, Least squares solution are the most popular technique.

The solution of LLS problem is very important in many fields of applied mathematics [20], such as signal processing [21], biomedical engineering [22], communication and others. Typically, the parameter estimation step is performed using various types LLS estimation algorithms because of the excellent characteristics where it is stable and efficient in terms of numerical computation [23]–[25]. However, this method facing a problem where the data have strong outliers because of the squaring effect [26]. Several established algorithms exist to solve LLS [24], [25]; namely Normal Equation, QR Decomposition and Singular Value Decomposition.

In this work, we choose the three establish method of LS to compare with another popular technique which is PSO. The PSO have been chosen as a method to compare because of it is stochastic optimization technique in optimization and its ability to perform in multidimensional data conditions. PSO is highly efficient and powerful algorithm. It is simplicity, stable and the speed and convergence are better than other evolutionary population-based methods [27]–[29]. Additionally, it required less memory when compiled and implemented.
Both method will be compared in parameter estimation NARMAX model process to find the best solution using FRA dataset.

This paper is organized as follows: Literature review and some theoretical background are presented in Section II and III, respectively. This is followed by the methodology (Section IV), results and discussions (Section V) and conclusions (Section VI).

LITERATURE REVIEW

Linear Least Squares (LLS) is a derivative based algorithm derived by minimizes the squares of errors as a cost function to get the linear relationship between input and output [1], [2]. Least squares usually used in estimation part, to estimate the unknown parameter value and also frequently used in the fitting function to find a linear relationship between variable [3], [4].

In modelling process LLS is used to fit a model to the data by lessening the sum of the squared deviations between the data and the model. The LLS method always been chosen in the modelling process because it is a simple method, effective and stable [1], [5].

However, LLS suffering from high sensitivity to outlier [3] effect where the squaring will change the magnitude of differences will become bigger. For example, the difference between 30 and 10 is equal to 20 but the difference between 302 and 102 is equal to 800. Additionally, in the literature on [6], LLS method (QR Decomposition) was performed in parameter estimation part. However, the works found a problem of LLS where this method is dependent on the condition of the regressor matrix where it is highly susceptible to ill-conditioning of the matrix and whether it is underdetermined.

Recently, PSO has received increased interest from the researcher. This technique has shown great promise in SI field [7]–[11]. It is an evolutionary population based optimization. It used the concept from swarming behaviours based from the flocks of birds [12], [13]. A set of agent is called particle. These particles tend to be first initialized at random positions, subsequently move within the problem space guided by their individual velocities and positions [14], [13], [15]. The search process is iteration-based. As a result, the search process is repeated until the best solution is found [14], [13], [15].

PSO has been chosen to compare with the LLS solution in this work because it is simple [24] because it has only one operator which is velocity equation thus make it easy to translate into programming code. Thus, it required less memory when compiled and implemented make it less expensive in term of computationally.

There are several previous works shown the successful application of PSO to solve the LS problems. In a study conducted by [9], the paper performed PSO and LS method was used to estimate the inertia parameters of Staubli RX-60 robot and it was shown that PSO is outperform than LS in torque performance estimation.

In paper [8], a comparative study on PSO and Orthogonal Least Squares (OLS) for solving NARX parameter estimation of a DC motor shows the positive results. The comparison between PSO and OLS was made on DC motor NARX model. They found that the PSO is comparable to OLS in solving the LS parameter estimation posed in the NARX model.

However, both of the research are limited into one method of least squares for solving SI problems in parameter estimation part. Paper [9] used Normal Equation to solve the LS problem while [8] used QR factorization for solving LS in parameter estimation part.

Thus, this study will cover all of the method solution of LLS (NE, QR, and SVD) and PSO to solve the least squares problem. Then the result will be compared to find which method gives the best solution for solving least squares for system identification using NARMAX model.

THEORITICAL BACKGROUND

System identification and NARMAX

SI can be defined as a LLS problem:

$$P\theta + \varepsilon = y \tag{1}$$

Where:-

P is $m \times n$ regressor matrix; θ is $n \times 1$ coefficient vector; y is $m \times 1$ vector of actual observations; ε is white noise residuals.

NARMAX model is nonlinear autoregressive moving average model with exogenous inputs where it is good for modelling non linear system because approximate nonlinear systems better as the nonlinearities are embedded into the model [6]. It is efficient and accurate [16], [6], [17]–[19]. NARMAX model output is dependent on its past inputs, outputs and residual. NARMAX equation is shown in Equation. (2).

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

where f^d is the nonlinear function for estimated model, $y(t-1), y(t-2), \dots, y(t-n_y)$ are lagged output terms, $u(t-n_k), u(t-n_k-1), \dots, u(t-n_k-n_u)$ are lagged input terms. n_k is the input signal time delay and normally the value is 1, except when $u(t)$ is required for identification.

Parameter estimation

This part is a process to estimate unknown the parameter value of the system. Two main methods used in this work; LLS and PSO. LLS can be divided into three common techniques [4, 7, 15] which is a Normal Equation (NE), QR Factorization (QR) and Singular Value Decomposition (SVD). Based on Equation. (1), if ε is removed, the estimation problem becomes similar to the form of $Ax = b$ (Equation. (3)):

$$P\theta = y \tag{4}$$

All three have different level computational load and precision, the generally normal equation being the least precise with obvious computational advantages, SVD is the slowest and most precise, while QR decomposition provides a good balance between the two. After the coefficient vector, θ, has been estimated, the model $P\theta$ is compared with the outputs of the original system, y, to produce the residuals, ε.

Normal equation

NE solves LLS problem [15] by transforming the rectangular matrix A into triangular form [7]. See Equation. (5) and Equation. (6).

$$A^T A x = A^T b \tag{5}$$

$$x = (A^T \mathrm{A})^{-1} . A^T b \tag{6}$$

This method is good to use when the matrix is well-conditioned since it is faster [20].

QR factorization

There are several methods can be used to compute the QR decomposition: Gram-Schmidt, Householder transform, and using Givens rotations. When applied to the Least Squares Problem Equation. (3), becomes equivalent to Equation. (7). The decomposition process is shown in Equation. (7) to Equation. (9):

$$A = QR \tag{7}$$

$$QR.x = b \tag{8}$$

$$x = R^{-1}Q^{T}b \tag{9}$$

Where:-

R is upper triangular, n × n; Q is m×n columns of Q are orthogonal.

Singular value decomposition (SVD)

Singular value decomposition (SVD) is a decomposition of matrix that used orthogonal transformations into a diagonal system [21]. The theorem is presented in Equation. (10) to Equation. (12).

$$A_{m\times n} = U_{m\times n} \sum_{n\times n} V^{T}_{n\times n} \tag{10}$$

$$U_{m\times n} \sum_{n\times n} V^{T}_{n\times n} . x = b \tag{11}$$

$$x = (U_{m\times n} \sum_{n\times n} V^{T}_{n\times n})^{-1}. b \tag{12}$$

SVD is widely used because [21], [20] it gives the most accurate result because it is numerically stable and this method can handle rank deficiency.

Particle swarm optimization (PSO)

The Particle Swarm Optimization (PSO) optimization algorithm was invented by [13] based on observation on the organizational behaviour of swarming animals. The vanilla PSO algorithm searches for the solution based on three major steps [14]:

1. Initialization of particle positions
2. Velocity update of particles
3. Position update of particles

The basic formula of PSO can be described as:

$$V_{id} = \begin{array}{c}\chi[V_{id} + c_1 \times rand_1 \times (p_{best} - x_{id}) + \\ c_2 \times rand_2 \times (g_{best} - x_{id})]\end{array} \tag{13}$$

$$x_{id} = x_{id} + V_{id} \tag{14}$$

Where :

$$\chi = \frac{2k}{||2-\varphi-\sqrt{\varphi^2-4\varphi}||}$$

With

$\varphi = C_1 + C_2, \varphi > 4$

$k \in [0,1]$

χ is constriction factor; c_1 is the self-confidence constant; c_2 = The swarm confidence constant; $rand_1$ and $rand_2$ are distributed random number between 0 and 1; P_{best} is best solution found by the particle ; g_{best} = Best solution found by the swarm; x_{id} = Particle position; V_{id} = Particle velocity.

Thus, this study will cover all of the method solution of LLS (NE, QR, and SVD) and PSO to solve the least squares problem. Then the result will be compared to find which method gives the best solution for solving least squares for system identification using NARMAX model.

METHODOLOGY

Hardware description

All experiments were performed in MATLAB r2014a installed in a computer with the following specifications: 64-bit Microsoft Windows operating system with an installed RAM of 8.0GB and an i5-4200M CPU 2.5GHz processor. Due to the nature of the platform, all numerical computation is limited to 64-bit precision.

Dataset description

The FRA dataset [22][6] is a nonlinear SISO system. It is described by the acceleration of a flexible robot arm in response to its reaction torque input. The FRA dataset consists 1,024 data points and has been used by [22] as a benchmark for comparing seven different of identification algorithms. The dataset is freely available on the internet as part of the Differential Algebra for Identifiability of Systems (DAISY) test suite [23]. The model order of this system has been stated as four [6][22].

Experiment flow

Structure selection

The structure selection part has been done in [6]. This experiment used the model structured based on the result in [6]. There are five terms for NARX model and 9 terms for MA part have been selected. The NARX terms are $u(t-2), y(t-1), y(t-2), y(t-3), y(t-4)$. For MA terms are $e(t-2)$, $e(t-1), e(t-3)$, $e(t-2)e(t-4)$, $e(t-4)$, $e(t-3)e(t-2)$, $e(t-3)e(t-3)$, $e(t-3)e(t-4)$ and $e(t-4)y(t-2)$.

Parameter estimation

This part divided into three parts as well as NARMAX model is recursive in nature, it has three stages which are NARX, MA and NARMAX (the combination of NARX and MA). Four methods were used to estimate the parameter θ; NE, QR, SVD and PSO. This parameter value θ, next will be used to find the fitness and MSE for

VOL. 10, NO. 22, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

the model developed. There are some parameters need to be set in PSO algorithm. Table 1 shows the details the parameter settings used for PSO. The experiments in this section have done by performing optimization under various combinations of parameters swarm size, maximum iterations and several random seeds.

Table-1. Parameter value for PSO.

Parameter	NAR	MA
Particle dimension	5	9
Swarm size	10, 20, 30, 40, 50	
Maximum iteration	1000	
V_{min}	-40	-40
V_{max}	40	40
x_{min}	-40	-40
x_{max}	40	40
C_1	2.05	
C_2	2.05	
Initial random seed	0, 10000, 20000	
Fitness criterion	Akaike's Information Criterion (AIC), Final Prediction Error (FPE), minimum description length (MDL)	

The particle dimension was determined based on the number of terms to optimize [16], [8]. In this experiment the terms to be optimized are referred to the structure selection experiment done by [6]. The swarm size was measured with 10, 20, 30, 40, and 50 particles. The test was limited to 50 particles because based on literature works [17], [18][24] increasing the number of particles beyond this limit might lead to extend the computational time without any noticeable reduction in the fitness function perform. The velocity value V_{max} and V_{min} are set to control the maximum movement of the particles during each iteration [18] thus its avoid searching in unfeasible regions [6], [8]. Based on literature in [25], V_{max} is set into 10-20% of the dynamic range for each variable dimension. However, in this works, the value of V_{max} is tested in range 10 to 50 and the best value used as in estimation process. x_{min} and x_{max} is a minimum and maximum range of x for all particles. Usually, the range of $x_{min}$and x_{max} is a value of V_{max}. However, in this works, we found that the best fitness achieve when the value set to x_{max} equals to V_{max}. Next, the function of seed is to ensure the same set of random number are generated for every execution of the optimization [6]. Lastly, the criterion AIC, FPE and MDL used in this works to evaluate the fitness in terms of model fit and the complexity of the model.

Model validation

The last part is to check the validity of the model. Two types of test will be used: fitting test and correlation test.

Fitting test was used to measure the best model fit magnitude. It can be tested using three methods which are One-Step-Ahead prediction (OSA), Mean Squared Error (MSE) and R-squared analysis. OSA used previous data to predict the future value of the ability of the model while MSE and R-squared used to indicate the model fit.

Next, to confirm the whiteness of residuals, the correlation test and histogram plot used. The number of Correlation Violations (CRV) was counted based on correlation point outside the 95% confidence limit. The correlation tests used were [26]:

$$\theta_{\varepsilon\varepsilon}(\tau) = E[\varepsilon(t-\tau)\varepsilon(t)] = \delta(\tau) \tag{15}$$

$$\theta_{\varepsilon^2\varepsilon^2}(\tau) = E[\varepsilon^2(t-\tau)\varepsilon^2(t)] = \delta(\tau) \tag{16}$$

$$\theta_{y\varepsilon}(\tau) = E[y(t-\tau)\varepsilon(t)] = 0, \forall\tau \tag{17}$$

$$\theta_{y^2\varepsilon}(\tau) = E[(y^2(t-\tau) - \bar{y}^2(\tau))\varepsilon(t)] = 0, \forall\tau \tag{18}$$

$$\theta_{y^2\varepsilon^2}(\tau) = E[(y^2(t-\tau) - \bar{y}^2(\tau))\varepsilon^2(t)] = 0, \forall\tau \tag{19}$$

Histogram analysis present the distribution of the residuals. The bell-shaped distribution results indicate the exhibit of the white noise is exist because the white noise signal is a random signal[16], [6].

RESULTS AND DISSCUSSION

NARX (PSO and LLS)

This section presented the NARX results for NE, QR, SVD and PSO. The identification results show the same value and achievement for all the three methods of LLS. Thus, it produced the same equation and figure out in the same figure for this three method. The NARX model developed is written as follows. Equation (20) refer to LLS while Equation. (21) refer to PSO algorithm.

$$y(t) = \begin{array}{c} 0.203u(t-2) + 3.1204y(t-1) \\ -4.3509y(t-2) + 3.0420y(t-3) \\ -0.9453y(t-4) + \varepsilon(t) \end{array} \tag{20}$$

$$y(t) = \begin{array}{c} 0.0203u(t-2) + 3.1216y(t-1) \\ -4.3536y(t-2) + 3.0447y(t-3) \\ -0.9464y(t-4) + \varepsilon(t) \end{array} \tag{21}$$

The results summarized in Table-2. As can be seen from the results, all four method used indicated very good and comparable results in terms of model fit, fitness and MSE value. High R^2 value exhibited supported with low MSE value for residuals. However, the correlation violations results indicated that the are many correlation cooficients were out form the confidence limits. Due to the recursive nature of the models, this model will go through MA part in the next section and this may improve the NARX results.

Table-2. Validation Summary LLS and PSO - FRA NAR Models.

Method	criterion	Training set	Testing set
Least squares (NE, QR, SVD)	AIC	1.3749e-05	1.3865e-05
	FPE	1.3843e-05	1.3959e-05
	MDL	1.3752e-05	1.3867e-05
	R-squared	99.9652	99.9637
	CRV	45	45
	MSE	2.6968e-05	2.7194e-05
PSO	AIC	1.3750e-05	1.3865e-05
	FPE	1.3843e-05	1.3959e-05
	MDL	1.3752e-05	1.3867e-05
	R-squared	99.9652	99.9637
	CRV	44	46
	MSE	2.6968e-05	2.7194e-05

FRA MA (LLS and PSO)

In this section, the prediction error of NARX is used as additional input for MA stage. The results of LLS and PSO identification for MA part presents in Equation. (22) and (23).

$$y(t) = \\ -2.3470e(t-2) + 2.0531e(t-1) + \\ 1.4833e(t-3) - 4.5180e(t-2)e(t-4) - \\ 0.4319e(t-4) - 0.5204\, e(t-3)e(t-2) + \\ 0.7610e(t-3)e(t-3) - 0.5498e(t-3)e(t-4) \\ -0.0528e(t-4)y(t-2) + \varepsilon(t) \tag{22}$$

$$y(t) = \\ -2.3463e(t-2) - 2.3651e(t-1) + \\ 1.4828e(t-3) - 4.5080e(t-2)e(t-4) - \\ 0.4315e(t-4) - 0.5190\, e(t-3)e(t-2) + \\ 0.7593e(t-3)e(t-3) - 0.5488e(t-3)e(t-4) \\ -0.0529e(t-4)y(t-2) + \varepsilon(t) \tag{23}$$

Table-3 shows the summary of identification results for MA stage. The R-squared results show model fit of LLS for training and testing set was very good for all method tested. Furthermore, the low MSE value indicate that the residual were sufficiently small. In this stage, the correlation results improve very well with the decreasing numbers of CRV's. Overall, all the four method tested shows similar results. The combination of FRA NARX and FRA MA will be presented in the next section to create a better NARMAX representation.

Table-3. Validation Summary LLS and PSO - FRA MA Models.

Method	criterion	Training set	Testing set
Least squares (NE, QR, SVD)	AIC	1.1237e-06	1.2552e-06
	FPE	1.1372e-06	1.2703e-06
	MDL	1.1244e-06	1.2559e-06
	R-squared	91.9443	91.0762
	CRV	32	34
	MSE	2.1704e-06	2.4244e-06
PSO	AIC	1.1326e-06	1.2548e-06
	FPE	1.1372e-06	1.2700e-06
	MDL	1.1243e-06	1.2556e-06
	R-squared	91.9446	91.0787
	CRV	34	32
	MSE	2.1704e-06	2.4237e-06

NARMAX (LLS and PSO)

This section presents the results of identification NARMAX model. All three LLS method showed similar results as well as it can be written in the same equation. Equation. (24) shows the identified NARMAX model using NE, QR and SVD and Equation. (25) is the identified model using PSO algorithm.

$$y(t) \\ = 0.203u(t-2) + 3.1204y(t-1) - 4.3509y(t-2) \\ +3.0420y(t-3) - 0.9453y(t-4) - 2.3470e(t-2) \\ +2.0531e(t-1) + 1.4833e(t-3) - 4.5180e(t-2) \\ e(t-4) - 0.4319e(t-4) - 0.5204\, e(t-3)e(t-2) \\ +0.7610e(t-3)e(t-3) - 0.5498e(t-3)e(t-4) \\ -0.0528e(t-4)y(t-2) + \varepsilon(t) \tag{24}$$

$$y(t) = \\ 0.0203u(t-2) + 3.1216y(t-1) - 4.3536y(t-2) \\ +3.0447y(t-3) - 0.9464y(t-4) - 2.3463e(t-2) \\ -2.3651e(t-1) + 1.4828e(t-3) - 4.5080e(t-2) \\ e(t-4)\ ` - 0.4315e(t-4) - 0.5190\, e(t-3)e(t-2) \\ +0.7593e(t-3)e(t-3) - 0.5488e(t-3)e(t-4) \\ e(t-4) - 0.0529e(t-4)y(t-2) + \varepsilon(t) \tag{25}$$

Based on the R-squared results in Table-4, all four method used shows the model is well-fitted with the R-squared value near to 100%. Compared to the NARX model, the NARMAX model improve the MSE , fitness and model fit of the model which can be proved with R^2 near to 100%. This is supported a small MSE of residuals for all four method tested. The low CRV's coefficient value show the small violation presents in this model. Therefore, all four model tested considered valid and acceptable for FRA NARMAX model.

Overall, all for method tested yielded similar results. However, in term of correlation and fitness PSO algorithm is outperform than NE, QR and SVD. The more correlation violation presents in LLS method because of the squaring effect of NE, QR and SVD.

In identification of real-life system, not all of the dynamics in the real system can be modelled fully as there are elements in the real system that are not fully understood. However, the effect of un-modeled dynamics can be minimized by choosing suitable parameters for identification and recursive introduction of residuals to help improve the model. Based on our experiments, the optimal parameters were obtained from the description of the dataset provider [22], [23], thus this has been addressed and the correlation violations minimized based on the available information on the system.

The results suggest that PSO's swarm-based searching method and numerical-based methods (NE, QR and SVD) are quite similar with minor improvements in correlation violations and MSE values. Therefore, it can be concluded that both methods can achieve superior performance for the FRA problem. However, in cases where the regressor matrix is ill-conditioned, the PSO algorithm may achieve significant advantage compared to the numerical methods as the search process is stochastic in nature and independent of problems present in the regressors. However, this is beyond the scope of this paper and will be the objective of our future studies.

Table-4. Validation summary LLS and PSO - FRA NARMAX models.

Method	criterion	Training set	Testing set
Least squares (NE, QR, SVD)	AIC	1.1450e-06	1.2790e-06
	FPE	1.1661e-06	1.3026e-06
	MDL	1.1467e-06	1.2809e-06
	R-squared	99.9972	99.9968
	CRV	26	24
	MSE	2.2128e-06	2.2968e-06
PSO	AIC	1.1450e-06	1.2787e-06
	FPE	1.1661e-06	1.3022e-06
	MDL	1.1467e-06	1.2805e-06
	R-squared	99.9972	99.9968
	CRV	25	24
	MSE	2.1704e-06	2.4237e-06

CONCLUSIONS

The comparison of PSO and least squres solution (NE,QR,SVD) for parameter estimation using NARMAX model was succesfully made. The MA part leads to a better prediction compared to NARX model. The comparison made found that the LLS method have more correlation bias compared to PSO method. Based on the results, PSO is less sensitive to outliers compared to LLS method. In this experiment, it can be conclude that PSO method is the best way in solving parameter estimation of FRA NARMAX model compared to another three method approach (established least squares solution, NE, QR and SVD) for NARMAX parameter estimation using FRA dataset because it is less bias than LLS method. PSO can be another method in solving parameter estimation posed in NARMAX model.

REFERENCES

[1] Blais, J. A. 2010. Least squares for practitioners. Mathematical Problems in Engineering, 2010.

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

[3] Abdi, H. 2007. The method of least squares. Encyclopedia of Measurement and Statistics. CA, USA: Thousand Oaks.

[4] Miller, S. J. 2006. The method of least squares. Mathematics Department Brown University, 1-7.

[5] Golub, G. H., & Van Loan, C. F. 2012. Matrix computations (Vol. 3). JHU Press.

[6] M. Y. Ahmad Ihsan. 2014. "Nonlinear Auto-Regressive Model Structure Selection using Binary Particle Swarm Optimization Algorithm," Universiti Teknologi MARA.

[7] Wen-chuan, W., Dong-mei, X., Kwok-wing, C., & Shouyu, C. 2013. Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD.

[8] Mohamad, M. S. A., Yassin, I. M., Zabidi, A., Taib, M. N., & Adnan, R. 2013, September. Comparison between PSO and OLS for NARX parameter estimation of a DC motor. In Industrial Electronics and Applications (ISIEA), 2013 IEEE Symposium on (pp. 27-32). IEEE.

[9] Bingül, Z., & Karahan, O. 2011. Dynamic identification of Staubli RX-60 robot using PSO and LS methods. Expert Systems with Applications, 38(4), 4136-4149.

[10] Shuzhi, G., Xing, D., & Xianwen, G. 2012, May. The temperature system identification of the PVC stripper tower top based on PSO-FCM optimized TS model. In Control and Decision Conference

(CCDC), 2012 24th Chinese (pp. 2529-2532). IEEE.

[11] Kang, D., Lee, B., & Won, S. 2007, November. Nonlinear system identification using ARX and SVM with advanced PSO. In Industrial Electronics Society, 2007. IECON 2007. 33rd Annual Conference of the IEEE (pp. 598-603). IEEE.

[12] Yassin, I. M., Taib, M. N., Hassan, H. A., Zabidi, A., & Tahir, N. M. 2010, December. Heat exchanger modeling using NARX model with binary PSO-based structure selection method. In Computer Applications and Industrial Electronics (ICCAIE), 2010 International Conference on (pp. 368-373). IEEE.

[13] Eberhart, R. C., & Kennedy, J. 1995, October. A new optimizer using particle swarm theory. In Proceedings of the sixth international symposium on micro machine and human science (Vol. 1, pp. 39-43).

[14] Hassan, R., Cohanim, B., De Weck, O., & Venter, G. 2005, April. A comparison of particle swarm optimization and the genetic algorithm. InProceedings of the 1st AIAA multidisciplinary design optimization specialist conference (pp. 1-13).

[15] Rini, D. P., Shamsuddin, S. M., & Yuhaniz, S. S. 2011. Particle swarm optimization: technique, system and challenges. International Journal of Computer Applications, 14(1), 19-26.

[16] Yassin, I. M., Rahim, N. A., & Taib, M. N. 2008. NARMAX identification of DC motor model using particle swarm optimization. In 4th Int. Colloquium on Signal Processing & Its Applications (CSPA 2008) (pp. 1-7).

[17] Yassin, I. M., Taib, M. N., & Adnan, R. 2013. Recent Advancements & Methodologies in System Identification: A Review.

[18] Rahrooh, A., & Shepard, S. 2009. Identification of nonlinear systems using NARMAX model. Nonlinear Analysis: Theory, Methods & Applications, 71(12), e1198-e1202.

[19] Rahim, N. A., Taib, M. N., & Yusof, M. I. 2003. Nonlinear system identification for a DC motor using NARMAX Approach. In Sensors, 2003. AsiaSense 2003. Asian Conference on (pp. 305-311). IEEE.

[20] Sewell, G. 2005. Computational methods of linear algebra (Vol. 74). John Wiley & Sons.

[21] DO Q, L. E. E. 2012. Numerically Efficient Methods for Solving Least Squares Problems.

[22] Markovsky, I., Willems, J. C., & De Moor, B. 2005. Comparison of identification algorithms on the database for system identification DAISY. InProc. 17th Symp. on Math. Theory of Networks and Systems (pp. 2858-2869).

[23] B. De Moor. Accessed: 03-Jan-2013. “Database for the identification of systems,” Dept. EE and K.U.Leuven. [Online]. Available: www.esat.kuleuven.be/sista/daisy/

[24] Talukder, S. 2011. Mathematical modelling and applications of particle swarm optimization (Doctoral dissertation, Blekinge Institute of Technology).

[25] Lin, W. X., Gu, Q. T., & Chen, Y. H. 2009, August. Multi-rate system identification based on improved PSO. In Electronic Measurement & Instruments, 2009. ICEMI'09. 9th International Conference on (pp. 3-616). IEEE.

[26] Rahim, N. A., Taib, M. N., & Yusof, M. I. 2003. Nonlinear system identification for a DC motor using NARMAX Approach. In Sensors, 2003. AsiaSense 2003. Asian Conference on (pp. 305-311). IEEE.